Mechanisms of Music Perception through Cochlear Implants

Dissertation*

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Sherif Abdellatif Omran

aus der

Arabischen Republik Ägypten

Promotionskomitee

Prof. Dr. Rodney Douglas (Vorsitz)

Prof. Dr. Norbert Dillier (Leitung der Dissertation)

Prof. Dr. Richard Hahnloser

Zürich, 2011

Kept blank intentionally

To the spirit of my father, peace may be upon him, who passed away two month after starting the thesis and to my mother, brother and my family.

"Bless you dad, you was so precious for me, you was the best, one can be
Why to complain and god helps me the best is to ask the almighty, he,
the creator is the one who guides me as Abraham and dad, I say; Thy
God is the only one who feeds me and when I'm sick, he cures me",
Sherif Omran

Thanks to my father, mother and my brother for supporting me continuously and for their deep care and hopes, as well as my uncles, aunts and all my merciful family members.

Thanks to my near friends (Driss Alaoui, Gabriele Gahler, Shaykh Abozar, Abdulilah Guendouz, Ahmed Tayara, Mohamed Assimi, Hassan Djamby, Ahmed Malek, Khaled Eljoundi and Zakareya Ramdana) who helpd me a lot during my stay in Germany and Switzerland.

أهدي رسالتي هذة إلى روح أبي المحبوب رحمهُ اللهُ تعالى رحمةً واسعةً المتوفى بعد شهرين من بداية دراستي للدكتوراه غفر اللهُ له و لوالديه و إلى أمي الغالية أطال اللهُ في عمرها و إلى أخي الحبيب مُحمَّد. كما اشكرهم على دعمِهم المستمر و لرعايتهم الحارة و أمنياتهم و كذلك إلى أخوالي و خالاتي و أعمامي و عماتي و كل عائلتي الطيبين.

رَحَماتٍ يا صلاحُ كُنْتَ وما دُومْتَ وقَدْ كُنْتَ لي خَيرَ مَن كُنْتَ
إلامَ أشْكو، واللهُ يَحمينْ وخَيْرَ الدُعاءِ "يارَبُ العالمين"
أقولُ ورائِكَ و الخليلَ، أمين اللّهُ خَالَقَي اَلَّذِي يَهْدِنِ
وَالَّذِي يُطْعِمُنِي وَيَسْقِينِ وَإِذَا مَرِضْتُ فَهُوَ يَشْفِينِ

شريف عمران

أشكر أعز الأصدقاء ﴿ إدريس عَلوي، جَبرْا ئيلّا جالر، الشيخ أبوذَرْ، عبدُالإله جندوز، أحمد طيّارة، محمد العاصمي، حسّان دْجامْبي، أحمد مالِك، خالد الجُنْدي و زكريا رَمْضانه﴾ الذين ساعدوني كثيرا خلال إقامتي في ألمانيا و سويسرا.

Kept blank intentionally

Abstract

Cochlear Implants (CIs) are devices aimed at restoring hearing with electrical stimulation of the hearing nerve. The average unrolled length of the cochlea is ~33 mm with a spiral shape of 2½ turns (~ ½ cm Ø) and a tonotopical arrangement. The Nucleus CI's electrode array typically occupies the basal 1½ turns, corresponding to characteristic frequencies of about 500-1000 Hz. Deeper insertion of the array to reach lower characteristic frequencies is hindered partly by the array's dimensions. Additionally, the number of electrodes is limited. As a consequence the standard (Std) Nucleus ACE (Advanced Combination Encoder) frequency mapping presents acoustically relevant information (188-7930 Hz) on upto 22 electrodes, mapping these frequencies to locations with tonotopical characteristic frequencies between ~500 - 1000 to ~ 6000 Hz. Stimulating an electrode produces an electric field that excites neighbor positions of the auditory nerve. Stimulating two electrodes simultaneously produces a virtual channel (VC) leading to the perception of an intermediate pitch. The VCs concept has been integrated into an acoustic model (AMO) of CI stimulation. New frequency to electrode mappings such as Greenwood and Semitone (Smt) mapping have been proposed and implemented. Basic parameters of the AMO model were optimized by matching results from a sentence recognition test with normal hearing (NH) subjects to results from CI recipients. Pitch discrimination tests using pure and complex tones with the Std mapping processed by the AMO were conducted with NH subjects to examine the effect of VCs on the score. The results showed a trend towards enhanced pitch discrimination scores with VCs. In a further step simple synthetic musical tones processed with the Std mapping and resynthesized using the AMO were inspected with a three dimensional (time, critical band and loudness) psychoacoustic spectral-contrast analysis. The output showed deterioration in the harmonic analysis structure of overtones, especially at lower frequencies with the Std mapping. The spectral-contrast analysis emphasized the results of the pitch discrimination test with pure tones where wider semitone intervals in higher octaves were easier distinguishable whereas complex tones were poorly resolved. This may coincide with the description of perceived music which many patients characterized as being an unpleasant noise. Technically, music can be described as a series of complex acoustic sounds composed of tones with fundamentals and overtones that are harmonically related. Music has many aspects such as pitch, melody, harmony, rhythm and timbre. Preserving the harmonic structure enhances melody representation, even if the fundamental frequency component is not explicitly encoded. The two proposed Smt mapping ranges (Smt-LF [130-1502] Hz and Smt-MF [440-5040] Hz therefore were designed to preserve harmonic structures in musical sounds. Electrodograms of a sequence of synthetic piano tones processed with different mappings (Std, Smt-LF and Smt-MF) were compared using an algorithm employing a newly proposed distance matrix. The results showed that at least one of the Smt mapping ranges preserves the harmonic structure better than the Std mapping. In a further examination, music tones were processed with the three mappings (Std, Smt-LF and Smt-MF) and a proposed harmonicity index was calculated. The analysis showed similar results. The Smt mapping was also evaluated by CI recipients. Testing was conducted using a melody contour identification (MCI) and instrument recognition (IR) tests by directly streaming offline-processed data to the implant. MCI results showed that the Smt mapping may improve the scores or at least produces similar results to the Std mapping. However, the average scores from the IR test were decreased suggesting a general deficiency in coding of timbre.

Zusammenfassung

Cochlea-Implantate (CI) sind Geräte, die das Hören mit Hilfe von elektrischer Stimulation des Hörnervs wiederherstellt . Die durchschnittliche Länge der ausgerollten Cochlea ist ~ 33 mm mit einer Spiralform von 2 ½ Windungen (~ ½ cm Ø) und einer tonotopischen Anordnung. Das Elektroden-Array des Nucleus CI deckt die tiefsten 1 ½ Windungen ab, das entspricht charakteristischen Frequenzen von ca. 500 - 1000 Hz. Ein tieferes Einsetzen des Arrays, um tiefe Frequenzen zu erreichen, ist teilweise durch die Grösse des Arrays verhindert. Außerdem ist die Zahl der Elektroden begrenzt. Als Ergebnis verteilt die Standard (Std) Nucleus ACE (Advanced Combination Encoder) Frequenz-Zuordnung akustisch relevante Informationen (zwischen 188-7930 Hz) auf 22 Elektroden. Die Frequenzen sind auf Elektroden verteilt, die tonotopisch charakteristische Frequenzen zwischen ~ 500 - 1000 bis 6000 Hz haben. Die Stimulation einer Elektrode erzeugt ein elektrisches Feld, das benachbarte Positionen des Hörnervs erregt. Die Erregung von zwei Elektroden gleichzeitig produziert einen virtuellen Kanal (VC), welcher zu der Wahrnehmung einer Zwischen-Tonhöhe führt. Das VC Konzept wurde in einem akustischen Modell (AMO) einer CI-Simulation integriert. Neue Frequenz-Elektroden-Verteilung (z.B Greenwood Mapping und Semitone Mapping) sind vorgeschlagen und implementiert worden. Die grundlegenden Parameter des AMOs wurden optimiert durch den Vergleich von Ergebnissen aus dem Satztest mit Normalhörenden (NH) und den Ergebnissen von CI-Trägern. Tonhöhenunterscheidungs-Tests mit reinen und komplexen Tönen wurden mit NH Probanden durchgeführt, wobei die Töne mit dem Std Mapping durch AMO verarbeitet wurden, um die Wirkung der VC auf das Ergebnis zu untersuchen. Das Ergebnis zeigte einen Verbesserungstrend für den Tönhöhen-Unterscheidungs-Tests. In einem weiteren Schritt wurden rein synthetische musikalische Töne, die mit dem Std Mapping verarbeitet sind und mit dem AMO resynthetisiert wurden, mit einer dreidimensionalen (Zeit, Kritisch Bänder und Lautheit) psychoakustischen Spektral-Kontrast-Analyse überprüft. Die Ergebnisse zeigten eine Verschlechterung in der harmonischen Struktur der Obertöne, besonders bei niedrigeren Frequenzen mit der Std-Frequenzzuordnung. Die Spektral-Kontrast Analyse bestätigt die Ergebnisse des Tönhöhen-Unterscheidungs-Tests mit reinen Tönen. Ein grösseres Semitone-Intervall in höheren Oktaven ist mit reinen Tönen einfacher zu unterscheiden als mit komplexen Tönen. Dies kann mit der Beschreibung vieler CI-Träger von wahrgenommener Musik als einem unangenehmen Geräusch übereinstimmen. Technisch gesehen ist Musik eine Folge von komplexen akustischen Tönen, die eine Grundfrequenz und harmonische Obertöne haben. Musik hat viele Aspekte wie Tonhöhe, Melodie, Harmonie, Rhythmus und Klangfarbe. Die Bewahrung der harmonischen Struktur verbessert die Melodieerkennung, auch wenn die Grundfrequenz nicht ausdrücklich kodiert wird. Die zwei vorgeschlagenen Smt Frequenz-Zuordnung (SMT-LF [130-1502] Hz und SMT-MF [440-5040] Hz) sind aufgebaut, um die harmonische Struktur der Melodie zu bewahren. Elektrodogramme einer Sequenz von synthetischen Klavierklängen mit verschiedenen Frequenz-Frequenzzuordnung (Std, Smt-LF-und Smt-MF) wurden mit einem Algorithmus verglichen, welcher eine Spezialmatrix zur Erkennung von Diagonalstrukturen in Spektrogrammen verwendet. Die Ergebnisse zeigten, dass mindestens eine der Smt-Mappings die harmonische Struktur besser bewahrt als die Std Frequenz-Zuordnung. In einer weiteren Untersuchung wurden die Musiktöne verglichen, die mit den drei Mappings bearbeitet sowie mit einen neu vorgeschlagenen Harmonizität-Index evaluiert wurden. Die Analyse bestätigt die bisherigen Ergebnisse. Die Smt-Frequenzzuordnungen wurde auch von CI-Trägern evaluiert. Es wurden Tests zur Melodiekonturidentifikation (MCI) und Instrumenterkennung (IR) mit CI-Trägern durch direkte Stimulation des Implantats mit bereits verarbeiteten Daten durchgeführt. Die Ergebnisse des MCI-Test, zeigen dass die Semitone-Zuordnung eine Verbesserung oder mindestens ähnliche Ergebnisse im Vergleich zum Std- Zuordnung bringen kann. Die Ergebnisse der IR hingegen waren im Mittel mit Smt schlechter, was vermuten lässt, dass die Klangfarbencodierung (Timbre) ungenügend war.

الملخّــص

أجهزة زراعة قوقعة الأذن (أ..ز.ق.ن) هي أجهزة تأمَل في إستعادة فَقد السمع عن طريق التحفيز الكهربي للعصب السمعي. قوقعة الاذن صغيرة الحجم حلزونية الشكل لها 2½ دوران (قطر ~ ½ سم) و متوسط طولها مفرودةً ~ 33 مم، و فيها توزيع استشعار قرار النغمات (الترددات) مكاني تنازلي باتجاه القمه. أ.ز.ق.ن من طراز نيوكليوس تبطن 1½ دوران و هذا يقابل اقل تردد مكاني تقريبا بين 500 - 1000 هرتز. حجم حامل الإلكترودات يعيق جزئيا زيادة عمق الدخول للوصول لمكان الترددات المنخفضة بالقوقعة. اضافة الى هذة المعضلة، يوجد عدد محدود من الإلكترودات. تُستخدم الخريطة التقليدية في ا.ز.ق.ن من طراز نيوكليوس في توجيه المعلومات الصوتيه الموجودة في الترددات من 188 الى 7930 هرتز الى 22 إلكترودات بحد أقصي و عليه تُوجّه تلك المعلومات الى اماكن ذات تردد مكاني من تقريبا 500 – 1000 هرتز حتى 6000 هرتز. تحفيز إلكترود كهربائيا يولّد مجال كهرُبائي يثير نقاط مجاورة عالعصب السمعي مقابلة للإلكترود. أما تحفيز الكترودين متزامنين يولّد قناة تخيليه تُسمع كنغمة متوسطة القرار. أضفنا مفهوم القنوات التخيلية علي برنامج لمحاكات تحفيز أ.ز.ق.ن. اقترحنا و طبقنا خرائط جديدة لتوجيه الترددات الصوتية الى الإلكترودات مثل خريطة جرين وود بالإضافة الى خريطة نصف النغمات. ضُبطت المعاملات الأساسية في برنامج المُحاكاة عن طريق مقارنة نتائج تعرف أشخاص أصحاء على جُمل منطُوقة عبر المحاكي بنتائج تعرف المرضى على نفس الجُمل. اُختُبر تعيين قرار نغمات أُحادية التردد (بسيطة) و متعددة الترددات (معقدة) بالخريطة التقليدية لأُناس أصحاء عبر المُحاكي لاختبار تأثير استخدام القنوات التخيلية عليها. أظهرت النتائج إتجاه في تحسن تعيين القرار باستخدام القنوات التخيلية. في خطوة تالية فُحصت نغمات موسيقية بسيطة إصطناعية مُعالجة بالخريطة التقليدية عن طريق نموذج سمعي بصري نَفسي ثلاثي الأبعاد (الزمن، أمادي ترددات الاعصاب السمعية و علوا الصوت المسموع) لتوصيف الصوت. الفحص أوضح تدهور في إنسجام ترددات النغمات المُعالجة بالخريطة التقليدية و بالأحرى في الترددات المنخفضه. التوصيف البصري أكد النتائج السابقة لإختبار تعيين قرار النغمات البسيطة ؛ فزيادة مسافة نصف النغمات بين نغمتين في المازورات العليا يُسهّل تعيينها و هذا أشار الى سبب انخفاض نتائج تعيين قرار النغمات المعقدة بالخريطة التقليدية. يتماشى ذلك مع وصف كثير من المرضى للموسيقى المسموعة على انها ضوضاءٌ مملةٌ. تقنياً الموسيقى يمكن وصفها كمتواليةٍ من النغمات المعقدة، كلٌ منهم له تردد أساسي و ترددات انسجامية أخرى لها علاقه بالتردد الأساسي. الموسيقى لها خصائص عديدة مثل القرار، اللحن، الإنسجام، الريتم و لون النغمات. الحفاظ علي إنسجام مكونات النغمات يُحسّن إستماع اللحن حتى ولو لم يُرسل التردد الأساسي للإلترودات. بِناءً عليه فان خرايطتين نصف النغمات المُقترحتين (ذات المدى السفلي من 130 – 1502 هرتز والمدى المتوسط من 400 – 5040 هرتز) تم تصميمَهُما للحفاظ على الشكل الإنسجامي لمُكونات الصوت الموسيقي. تم مقارنة التغير الزمني لتوزيع شدة التيار على الإلكترودات الناتجة من نغمات إصطناعية لبيانو مُحاكاه بالخرائط المختلفة (التقليدية، نصف النغمات بالمدى السفلي و المتوسط) باستخدام لوغاريتم يعتمد على مصفوفة جديدة مُقترحة للقياس. النتائج أظهرت انه على الأقل أحد أمداد خريطة نصف النغمات يحافظ على إنسجام الترددات أفضل من الخريطة التقليدية. في إختبار تالي عُولجت نغمات موسيقية بالخرائط الثلاثة (التقليدية و نصف النغمات بمدييهما) و تم تعيين مقياس لإنسجام مكونات الصوت الناتج عن طريق لوغاريتم مُقترح و المقياس أوضح نتائج مُشابهة. أُختُبرت خرائط نصف النغمات مع المرضى في إختبار توصيف إتجاه اللحن و التعرف على الآلات الموسيقية عن طريق تحفيز الإلكترودات مُباشرة من ملفات مُجهّزة مُسبقا. نتائج التعرف على اتجاه اللحن أوضحت أنّ خريطة نصف النغمات تُحسّن النتائج أو على أقل تقدير تُعطي نتيجة مُشابهة للخريطة التقليدية. أما متوسط نتائج التعرف على الآلات الموسيقية فانخفضت مُقترحةً وجود قصور بشكل عام في معالجة لون النغمات.

Contents

Kept blank intentionally

List of abbreviations

Abbreviation	Meaning	Page
A/D	analogue to digital converter	6
ACE	advanced combination encoder	7
AIS	asynchronous interleaved sampling	18
AM	amplitude modulation	24
AMO	acoustic model	26
ANSI	American national standards institute	5
BF	best frequency	15
BM	basilar membrane	4
BPF	band pass filter	11
BQFFT	bounded-Q fast filter bank transform	62
BW	band width	12
C-Level	comfort level	6
CA	compressed analogue	18
CB	critical band	11
CF	characteristic frequency	18
cf	center frequency	31
CHAR	characteristic frequency mapping strategy	22
CI	cochlear implant	2
CIC	channel interaction compensation	17
CIS	continuous interleaved sampling	16
CTM	channel time matrix	30
dB	decibel	10
DC	direct current	2
DDM	diagonal detecting matrix	53
dll	dynamic link library	48
EECIS	envelope enhanced continuous interleaved sampling	17
ERB	equivalent rectangular bandwidth	12
f0Sync	f0 synchronized ACE	20
FAME	frequency amplitude modulation encoding	24
FFT	fast Fourier transform	19
FM	frequency modulation	24
FS	fine structure	24
FSP	fine structure presentation	25
FTM	frequency time matrix	37
GTF	Gamma tone filter bank	23
HAP	hybrid analog pulsatile	18
HF	high frequency	36
HiRes	HiResolution	18
HPI	Harmonicity probabilistic inequality	61
HT	Hilbert transform	17
IR	instrument recognition	27
ISO	international standard organization	11
IWAIF	intensity weighted average of instantaneous frequency	24
jnd	just-noticeable differences	19
LF	low frequency	24
LGF	loudness growth function	30

LPF	low pass filter	20
MCI	melody contour identification	16
MDE	modulation depth enhancement	24
MEM	multi channel envelope modulation	20
MPEAK	Multi-Peak	15
MPEG	moving picture expert group	28
MPS	multi pulsatile sampling	18
NH	normal hearing	6
NIC	Nucleus implant communicator	48
NMT	Nucleus Matlab toolbox	43
OHC	outer hair cell	4
PACE	psychoacoustic advanced combination encoder	22
PDT	Peak derived timing	24
pps	pulses per second	15
PPS	paired pulsatile sampler	17
PSD	power spectral density	7
RF	radio frequency	7
RMS	root mean square	10
SAS	simultaneous analog stimulation	17
SMSP	spectral maxima sound processor	15
Smt	semitone	16
Smt-LF	semitone mapping LF	21
Smt-MF	semitone mapping MF	21
SNR	signal to noise ratio	23
SOE	spread of excitation	27
SPEAK	spectral peak	16
SPL	sound pressure level	10
STAR	spike-based temporal auditory representation	23
Std	standard	16
STFT	short time Fourier transform	19
T-Level	threshold level	6
VC	virtual electrode channel	5

Chapter 1- Introduction

Chapter 1 – Introduction

1.1 Background

A Cochlear Implant (CI) is a device which partially restores hearing in a deaf or hard of hearing person. Intensive research and development took place since several decades. A method attempting to provide hearing by stimulating the auditory system electrically dates back 200 years (for a review see Simons 1966). Signal processing algorithms in the external hardware of CI devices are usually called „speech processing strategies" because they were mainly developed to provide recognition of spoken words. Currently, 80% or higher word recognition can be restored by CI devices incorporating sophisticated processing strategies (Mueller-Deile 2009), but music perception remains generally rather poor. In particular, many CI recipients report that music is experienced as unpleasant noise despite of the substantial progress in designing approaches for music perception. CI devices from different manufacturers utilize different algorithms; each improves a certain aspect of music and may deteriorate others. This thesis studies mechanisms of music perception using CI devices from a signal processing point of view and aims at introducing a new algorithm that may enhance music representation, using specific measuring tools and psychoacoustic tests to examine music appreciation.

1.1.1 History of Electrical Hearing

Interest in electrical methods of stimulating the hearing began in the late 18th century when Alessandro Volta (1745-1825) discovered the electrolytic cell. Volta was the first to stimulate the auditory system electrically, by connecting a battery of 30 or 40 'couples' (approximately 50V) to two metal rods that were inserted into his ears. When the circuit was closed, he received the sensation described as "a boom in the head", followed by a sound similar to that of boiling of thick soup. Volta's observation sparked sporadic attempts to investigate the phenomena over the next 50 years, but the sensation patients described was always momentary and lacked tonal quality (Pancaldi 2003).

Investigation of electrical stimulation was described through the 18th and the 19th century using direct current (DC). In 1855, Duchenne de Boulogne stimulated the ear with an alternating current that he produced by inserting a vibrator into a circuit containing a condensator and a coil. This resulted in a sound he described as 'the beating of a fly's wings between a pane of glass and a curtain' (Boulogne 1855).

In 1868, Brenner published a more extensive investigation of these effects. The perception of varying polarity, rate, intensity of stimulus, and placement of the electrodes was studied (Simmons 1966). He found that hearing was better with an electrical stimulus that created a negative polarity in the ear, and that correct placement of the electrodes could reduce the unpleasant side effects. Brenner used monopolar stimulation, meaning that one electrode was placed in saline in the external auditory meatus, and the other was placed on a more distant part of the body. This electrode is now referred to as the Brenner electrode (Clark 2003).

Wever and Bray (1930) demonstrated that the electrical response recorded from the auditory nerve of a cat was similar in frequency and amplitude to the sounds which the ear had been exposed to (Hallpike and Rawdon-Smith 1934). Meanwhile, the Russian investigators Gersuni and Volokhov (1936) examined the effects of an alternating electrical stimulus on the

hearing (Arapova et al. 1936). They found that the hearing could persist following surgical removal of the tympanic membrane and ossicles, and thus hypothesized that the cochlea was the site of stimulation.

Another set of researchers, Stevens and Jones (1939), thought that electrical current could be transduced into sound vibrations before it reached the inner ear. Hearing induced in this way was termed the electrophonic effect. They were able to determine whether a linear or non-linear transducer was involved from presence and strength of overtones, which were detected when subjects heard beats (Stevens and Jones 1939). Studies by Stevens and Jones (1939), as well as Jones et al.(1940) indicated that when the cochlea was electrically stimulated, there were three mechanisms which produced the hearing (Jones et al. 1940):

1. The middle ear could act as a transducer, which obeys the 'square law' and convert alternations in strength of an electrical field into mechanical vibrations that produce the sound.

2. Electrical energy could be converted into sound by a direct effect on the BM, which would then vibrate maximally at a point determined by the frequency and these vibrations would stimulate the hair cells.

3. Direct stimulation of the auditory nerve produced a crude hearing sensation.

Their conclusions were basically correct, although now other body tissues have been shown to act as transducers.

A wealth of research in the 1940's and the 1950's about mechanisms involved in electrophonic hearing indicated that the hearing is produced by transducing electrical energy into sound vibrations and that residual cochlear function is also required (Flottorp 1952). It became apparent that total perceptive deafness could not be corrected by inducing a widespread electrical field in the region of cochlea. Instead, a more localized stimulation of the auditory nerve fibers is required (Clark 2003).

The first direct stimulation of the auditory nerve in a human was performed during an operation by Lundberg in 1950 - the patient became aware of noise. In 1957, Djourno and Eyries implanted an electrode attached to an inductive coil in the head of a deaf patient (Eisen 2003). They were able to transmit a signal to the electrode via a radio antenna on the outside of the body. The implanted patient heard sounds resembling chirping of a grasshopper or cricket. He was also able to recognize simple words like mama, papa, and hello. This experiment inspired many investigators about possibilities of using implanted prostheses to enable deaf people to hear.

In 1964, Doyle et al., reported inserting an array of electrodes into the cochlea of a patient with total perceptive deafness. The electrodes were designed to limit the spread of the electrical field and were stimulated in sequence with threshold square waves that were superimposed with speech signals. The four electrodes were not especially implanted to take advantage of the spatial distribution of the auditory nerve fibers responding to different frequencies, and the result obtained was only satisfactory. However, it was significant that the patient was able to repeat phrases (Doyle et al. 1963).

Simmons (1966) provided a more extensive study in which electrodes were placed through the promontory and vestibule directly into the modular segment of the auditory nerves. The nerve fibers representing different frequencies could be stimulated. The patient was tested to

assess the effect of alterations in frequency and intensity of the signal. The subject demonstrated that in addition to being able to discern the length of signal duration, some degree of tonality could be achieved (Simmons 1966).

The clinical applications of electrical stimulation of the auditory nerve were refined by Michelson (1971) and House (1976) through scala tympani implantation of electrodes driven by implantable receiver-stimulators. House observed the percepts of patients when small electric currents were introduced to the promontory during middle ear procedures under local anesthesia. During the early sixties, House implanted several devices in totally deaf volunteer patients. Although these were rejected due to lack of biocompatibility of the insulating material, they worked for a short time and provided optimism toward a solution for sensorineural deafness. House teamed up with Jack Urban, a mechanical engineer, to ultimately make Cochlear Implants a clinical reality.

In 1972, a speech processor for a single-electrode implant was developed and implanted later that became the first to be commercially marketed as the 3M House system. During the same period, work outside the United States was progressing, most notably in Australia when Clark and colleagues were developing a multi-channel Cochlear Implant that, in the eighties, was to become the single-most used implant in the world under the name "Nucleus Multi-channel Cochlear Implant". Multiple channel devices provided enhanced spectral perception and speech recognition capabilities compared to the single-channel device, as reported in clinical tests.

1.2 Cochlear Implants and Hearing

CI devices have been used in the last years with great success. This section illustrates how the hearing in animals functions and provides a basic background for electric hearing with CIs. In the next sections more details about CIs and speech processing strategies are presented.

1.2.1 Hearing Physiology of Mammalian

In the mammalian auditory system, sound enters through the external ear canal till it hits the ear drum. This causes small vibrations on the drum's membrane. The vibration is transduced by the middle ear which consists of three ossicles (Malleus, Incus and Stapes) into pressure changes in the inner ear. The inner ear has a snail shape called "C*ochlea*" of 2½ spiral turns having a height of 0.9 mm and ~30 mm length in the unrolled form. It has two inlets called the oval and the round windows. Pressure produces a displacement of the membrane which the inner ear forms on one of the two openings. The cochlea consists of three layers; scala tympany at the bottom with a tube diameter of ~0.3 mm, scala vestibule at the top and in between scala media. Basilar membrane (BM) has inner and outer hair cells (OHCs). Pressure waves vibrate BM and in turn move hair cells. The cochlea has a tonotopical representation due generating standing pressure waves from fluid motion. If it is unrolled, low frequencies are located at the apical tip and high frequencies at the basal end (Greenwood 1961).

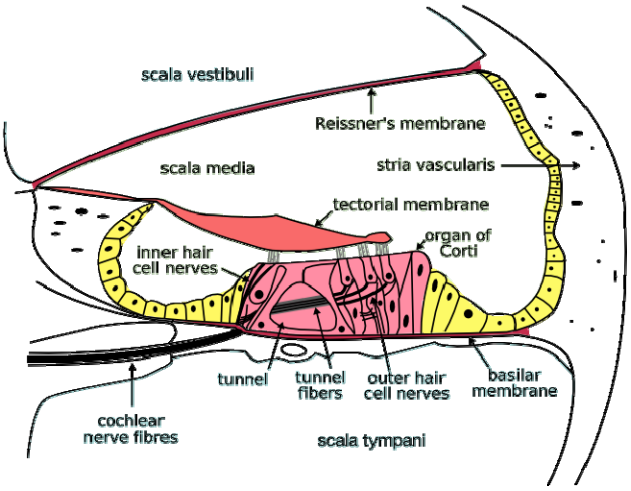


Figure (1.1): Cross section segment in the cochlea.

1.2.2 Introduction to Cochlear Implants

CI devices were originally aimed at restoring speech perception (Gfeller et al. 2000b). Electrodes are implanted in the cochlea and connected to a receiver which contains electric circuit to generate current stimuli. A speech processor picks up sound signal through microphone then encodes these signals and transmits them. Stimulating an electrode causes current to flow through the nerve to a reference electrode. This depends on a time varying usually biphasic current pulses are generated which produce a reference varying electric field. All CIs share the same principle. An incoming audio signal is processed with a speech processor, and the corresponding electrodes are stimulated using an underlying processing strategy (Miller and Spelman 1989). The implant has to replace the effect of a huge number of hair cells with a limited number of electrodes. Exciting an electrode stimulates a range of tonotopical positions which is perceived as a plane pitch sensation. Pitch, according to the American national standards institute (ANSI) definition of psychoacoustical terminology, is the auditory attribute of sound by which sounds can be ordered on a scale from low to high. Although the electrode carrier is relatively thin, it usually is not pushed as far as the cochlear apex typically after 1½ turns. The used angles are about 400 degrees which corresponds to less than 1½ turns. Thus positions corresponding to low frequencies below 500±100 Hz can not be directly stimulated. Standard CI processors use a bandwidth compression approximation to present a band of frequencies greater than the available tonotopical range.

1.2.2.1 Electrode Stimulation

CI processing strategies were optimized to enhance speech comprehension and large improvements in speech recognition performance have been documented. However, many patients still describe the perceived musical sounds as unpleasant. The implant does not transmit replica of musical sound as it is perceived by a normal ear; though it gives fairly good representation of rhythm and other aspects of music are generally encoded (Gfeller and Knutson 2003). Music has different aspects (e.g. pitch, melody, harmony … etc). Ameliorating pitch perception requests an increase in the implant's frequency resolution. Due to physical limitations in the available number of electrodes, the use of virtual electrode channel (VC) is investigated. VCs are produced by simultaneous or non-simultaneous dual-electrode stimulation, which excites intermediate positions and causes the perception of an intermediate complex pitch (McDermott and McKay 1994; Busby and Plant 2005). In the

nonsimultaneous dual electrode stimulation; two electrodes are stimulated sequentially with an intermediate stop cycle that is smaller than the neural refractory time. This in turn causes the perception of an intermediate complex pitch. This concept was tested on a low to high scale in a pitch ranking test by asking patients to mark the perceived pitches on a scale. It was found that the patients tended to have more difficulty distinguishing pitches at the basal end of the electrode array (McDermott and McKay 1994). In the simultaneous dual electrode stimulation; two adjacent nearby electrodes are stimulated simultaneously, with the same current level in the Nucleus implant. It was found that this causes a reduction of 38% in the average electrode impedance, which is a potential parameter that limits the amount of electric current delivered to the cochlea and decreases the power consumption. If the applied impedances are too high, there is a possibility that the implant voltage will not be sufficient to generate the programmed electric current level and in such a case the perceptual loudness does not increase with increase in current level. Many implanted subjects could discriminate between pitch differences of single and double electrodes at least in part of the electrode's array (Busby and Plant 2005). Poorer performance of some subjects may be due to reduction in spiral ganglion cells surviving at such positions in the cochlea. Stimulation of the closely spaced electrodes essentially stimulates the same group of ganglion cells. The probability of success is expected to increase, if the number of ganglion cells is high (Busby and Plant 2005). There is another type of simultaneous dual electrode stimulation with different current levels presented on two electrodes called *current steering*. In current steering, two adjacent electrodes are stimulated at different levels, creating an imbalance to stimulate a certain location in between them. This creates an intermediate VC and causes the perception of an intermediate pitch that depends on the current ratio in the two electrodes (Firszt et al. 2007).

The cochlea has a tonotopical representation and each electrode stimulates a certain position. Each pulse stimulates a group of nerve fibers located at a place on the BM using a stimulation rate adjusted for each patient subjectively. Increasing the rate, increases the perceived pitch upto an upper limit around 300 Hz; between 400 and 1000Hz the perceived pitch does not change anymore. The perceived pitch in CI recipients does not depend only on the tonotopical place as for normal hearing (NH) subjects but there are also other factors such as stimulation rate, number of hair cells available, neural refractory period and exact position of electrodes.

1.2.2.2 Loudness Function and Current Levels

Pulses stimulate the electrodes and each pulse transfers a certain amount of energy (charge). A higher number of pulses transfers higher energy and this causes the perception of a louder sound. Loudness of a tone is correlated to the current level at an electrode. Current level for the Nucleus device is described in units that vary from 0 to 255. Generally, two levels are the most important; threshold level (T-Level) at which the patient just notices a sound and comfort level (C-Level) at which the sound is comfortable. Loudness is calculated as the summation of overlapping neural populations. Moore and Glasberg (1997) proposed a loudness model that is later used to formulate a relationship between the induced current level and loudness perceived (McKay et al. 2003). It was noticed that changing stimulation rate, changes T and C-Levels current source characteristics of Nucleus device (Hong and Rubinstein 2003; McDermott et al. 2003a). McKay et al., (2003) proposed Equation (1.1) for loudness calculation, where 'i' is the induced current in the cochlea and the number of steps 'c' of the current is described by an analogue to digital converter (A/D) with a maximum of 255 levels.

$$i = 10 * 175^{c/255} \tag{1.1}$$

In non-overlapping neural excitation, patterns contribute independently to loudness and the overall loudness doesn't change with the degree of overlapping, thus electrical signals always behave as if loudness contributions from different electrodes were independent, regardless of the electrode positions or the overlap. This finding was used in dual electrode stimulation (McKay et al. 2003).

The loudness function (L) is linear at low values as described empirically by Equation (1.2) in dB in terms of the current steps per pulse within a 2 ms time interval with Nucleus devices only (McKay et al. 2003).

$$\log(L) = a * c + 0.03 * b * e^{(c-c0)/b} + k \qquad (1.2)$$

where c is the current levels, k is an electrode dependent constant, a, b and c are patient dependent constants and should be pre-calculated for each patient. However several psychoacoustic studies provided evidence that this simple conversion of acoustic to electric levels may not be the optimal because complex sounds vary in bandwidth. This model was based on an approximation that individual pulses in a set of time interval act as if their contributions to the overall loudness were independent (McKay et al. 2003; McDermott et al. 2005).

1.2.2.3 Speech Processors

Processing algorithms in CI devices are important system components. In the standard Nucleus CI, sound is captured through a microphone behind the ear and analyzed in a sound processor which calculates the corresponding power spectral density (PSD) per time sequence. A number of maxima are selected from the envelopes of the frequency loudness and stimulus parameters are determined. These parameters are transferred through a coil placed on the head above the skin to a receiving coil beneath the skin by radio frequency (RF) waves. An electrode is stimulated for a short time interval known as the stimulation time per electrode (Swanson 2005). Figure (1.1) shows a block diagram of the Freedom Cochlear Implant's speech processor.

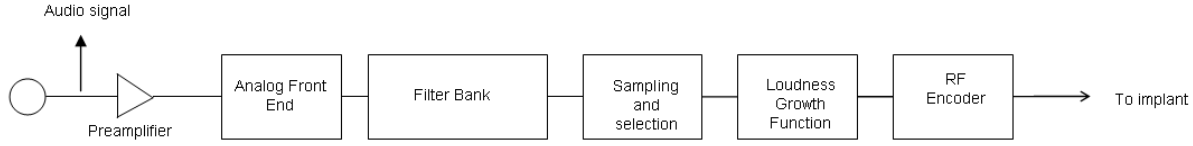


Figure (1.2): Block diagram of the Freedom Cochlear Implant's speech processor (Swanson 2001).

There are many different processing strategies and this will be described in detail in section 1.3. Many of them depend on selecting a number of maxima "n" from a total number of maxima "m" like the advanced combination encoder (ACE) and are called "n-of-m" strategies, others strategies stimulate the full set of electrodes for each cycle and are called "m-of-m".

1.2.3 Musical Sound Psychoacoustics

Music has many descriptions; Heinrich Hueschen (1961) described it among the artistic disciplines and said "it is the one whose material consists of tones. Of the raw materials available in nature, only a small proportion is actually used in music. The finite number of tones selected for musical use from the infinity available in nature is organized into specific tone systems through defined rational processes" (Blume 1961). Hans Heinrich Eggebrecht (1967) provided a characterizing definition and said "music is – in the area in which the concept is relevant, western culture – the artistic formation of those sounds that represent the world and the spirit in the form of a voice of nature and emotion in the realm of hearing, concretely conceived, and which achieves significance as an art, become both meaningful and meaning-creating material through reflected and ordered cognition and theory. For the basic element of music, the tone is on one hand the bearer of meaning, while on the other hand it is the vehicle of meaning as the beneficiary of the tonal order. These lend to the unit of music, tone, its specifically cultural forms, meanings and conceptions and at the same time, as a natural phenomenon, it remains accountable to the laws of nature" (Riemann 2000). Upto the time being there is no complete definition of music. Rather, various descriptions of the concept that appear in the literature and emphasize a particular aspect of the total phenomenon (Blume 1961; Sadie and Grove 1995).

Music consists of tones; where a *tone* is a sound played or sung with a particular pitch and duration. It is one of the important aspects of music. Melody is an agreeable succession of tones that represents a link between the individual musical tone and tonal music as a whole (Terhardt 1998). The sequence of tones in a melody forms its melodic line. Within each melodic line a sequence of oriented intervals defined by neighboring tones can be traced. This sequence demonstrates the movement of the melodic line (Hofman-Jablan 1995). Melodies are often described as being made up of phrases. They may also be perceived in short musical ideas (melodic, harmonic, rhythmic or any combination of these three) called *motifs* (Sadie and Grove 1995). A motif is the smallest meaningful unit, a perceivable or salient recurring fragment or succession of notes that may be used to construct parts of a melody (Scruton 1997). Long melodies that reappear in a music piece are called *themes*.

Another two important aspects of music are rhythm and meter. A series of events (whether musical notes or blows of a hammer) is commonly characterized as 'rhythmic' if some or all of those events occur at regular time intervals. Being 'rhythmic' is not the same thing as being 'Rhythm'. *Rhythm* is a major aspect of music, dance, and most poetry; it involves the pattern of both regular and irregular durations that are present in music. *Meter* involves the perception and anticipation of rhythm patters (Houle 1987). Meter is a structured attending to time which allows listeners to have precise expectations as to when subsequent musical events are going to occur. It plays a crucial role in the determination of relative durations. Not all durations are perceived the same, as there are a number of psycho-physical limits on the ability to perceive durations and durational succession. Divenyi (1974) demonstrated that two onsets must be separated by at least 2 ms in order to be distinctly perceived, and that at least 15-20 ms are required to determine which onset came first and 100 ms seems to be the threshold for reliable judgments of the length (Divenyi and Hirsh 1974). In rhythm, any periodic transient signal that marks it is called a *beat*. Beat may be described as a sudden change in energy with respect to the preceding history. Much music is characterized by the sequence of stressed and unstressed beats (Neukom 2005). The perception of a beat is not only necessary for a sense of 'connectedness' among successive events; it may also be necessary for a sense of motion (Bregman 1990). A sense of beat, while necessary, is not sufficient to engender a sense of meter.

Harmony is another important aspect of music. The term harmony is derived from the Greek word 'harmonia' and means 'fit or join together'. The term '*harmony*' refers to the combination of notes simultaneously to produce chords and successively to produce chord progressions (Sadie and Grove 1995). *Chords* are groups of notes built on major or minor triads and sounded simultaneously. A sound appears pleasant to the ear if a musical relationship exists between the overtones of a chord. From an engineering point of view, music can be described as a series of complex acoustic sounds composed of tones. Such tones have a harmonic structure of overtones (Pierce 1983). Harmony can also be produced from an inharmonic spectrum (Neukom 2005). An inharmonic spectrum is formed from aperiodic tones; such as frequencies with the "Golden ratio" 1.61803 (Walser 1993) where the signals in the time domain do not overlap except at t=0.

A human being can hear sounds between 20 and 20000 Hz. Doubling the frequency of a tone, increases the octave by one. Accordingly, one hears exponentially increasing frequencies as a linearly increasing pitch and vice versa a linear increase in frequencies, causes a logarithmic increase in the perceived pitch (Zwicker and Fastl 1990; Terhardt 1998; Neukom 2005).

Figure (1.3): Relation between frequency (left) and pitch (right) (Neukom 2005).

Pitch intervals in music are defined according to the logarithmic relationship between fundamental frequencies of tones in a chord (Pierce 1983) measured in cents as described by Equation (1.3). The overtones ratio is the basis of many natural elements in the traditional music theory such as octave identity and major triad.

$$x(Cents) = 1200.\log_2\left(\frac{f1}{f2}\right)^1 \tag{1.3}$$

Interval name	Octave	Fifth	Fourth	Major third	Minor third
Frequency	1:2	2:3	3:4	4:5	5:6

Table (1.1): Relation between interval and frequency.

Timbre is a characteristic feature of musical instruments. Helmholtz was the first author to describe the timbre as a property of the spectral components of the sound (Helmholtz 1954). Helmholtz view of timbre was that the perceptual cues came from the Fourier series coefficients. Unlike pitch or rhythm, timbre is developed from many physical features and can not be measured with a scale (Neukom 2005). Timbre is defined physically by the temporal envelope (particularly the onset) and the spectral shape of the acoustic sound (Handel 1995). The characteristics of timbre do not come particularly from the ratio of the strengths of partials, but from the *formants*.

[1] Instrument intonation is usually not exact with this formula

Formants are frequency ranges (energy bands) in which partials are particularly strong. The *formant characteristics* is a relationship between amplification, or degree of resonance, and frequency for any device that modifies, transmits or radiates sound (Sadie and Grove 1995). Formant detecting has its disadvantages because a shift in the estimated formant values due to different reasons such as quantization errors or noise, may introduce a shift in the corresponding sample sequence and therefore could change the reconstructed envelope as illustrated in Figure (1.4).

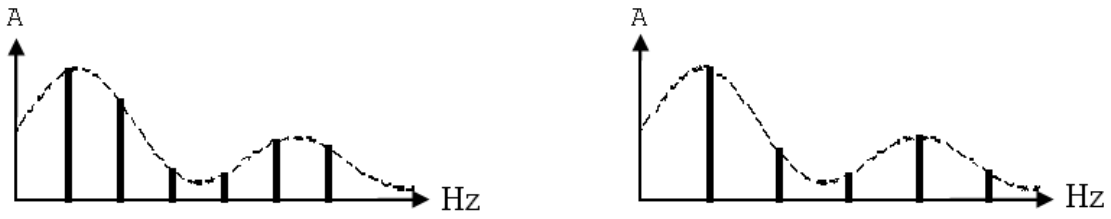


Figure (1.4): Selected formants change the power spectrum envelope

Simple bodies like strings and air columns produce sounds with an approximately harmonic spectrum. Thus, for example, frequencies of high tones in a string increase as it gets thinner because the stiffness decreases. In pipes, frequency ratios are not integers, because the length of the effective vibrating air column is not exactly the length of the pipe and the length is not the same for all tones. Bodies from which the sound do not propagate in one direction only, like plates and bells, have natural vibrations, whose frequencies can have any relationships.

Sound waves transport energy. The total energy produced by a sound source per second is called *sound power* or *acoustic power*, measured in Watt. A normal conversation produces around 10 μW, a violin produces approximately 1 mW while a bird produces about 0.2 W (Hans 1980). These relatively small energies are distributed in space, and a small portion reaches the eardrum called *sound power density* or *sound intensity* (*I*). It is described in Watt/m^2. Sound intensity of approximately 10^{-12} W/m^2 is the *hearing threshold* (I_0). The sound intensity is directly proportional to the square of the sound amplitude (*A*) and the square of the frequency (*f*). The acoustic impedance (*Z*) is a property of the transport medium. For air Z = 413 N·s/m³.

$$I = A^2.f^2.4.\pi^2.Z \tag{1.4}$$

To describe sound intensities, it is generally required to compare values. Decibel (dB) is a unit used to measure ratios. The ratio between *I* and I_0 is termed *sound intensity level*.

$$L=10log_{10}\,(I/I_0)\ dB \tag{1.5}$$

Sound pressure is the local change in pressure from the equilibrium caused by a sound wave at a given location and a given instant in time. Sound pressure level (SPL) or sound level (*Lp*) is a logarithmic measure of the root mean square (RMS) of a sound pressure relative to a reference (P_{ref}) threshold value 2x10^5 Pascal.

$$L_p = 20\log\left(\frac{P_{rms}}{P_{ref}}\right)\ \text{dB} \tag{1.6}$$

The ear sensitivity varies with frequency. The hearing threshold of a sound below 20 Hz is high enough to be inaudible. For an increasing frequency from 20 Hz to ~3.3 kHz, the threshold decreases and the sound is perceived as louder, then the threshold increases for frequencies higher than ~3.3 kHz (see Figure 1.5). This was measured psycho-acoustically

using single harmonic tones that have the same sound intensity and different frequencies, where subjects had to choose the one higher in pitch. That is, if two sinusoidal sounds of different frequencies and sound equally loud are said to lying on the same intensity level curve (isophone). A single harmonic tone of 1 kHz at 40 dB SPL is considered to have 40 Phon. The same loudness level is perceived at other frequencies with different SPLs as shown in equal loudness curves from the international standard organization (ISO) as in Figure (1.5) (Robinson and Dadson 1956; ISO-226:2003 revision). Loudness itself is measured in *sones*. A sound with a loudness of 20 sones sounds twice as loud as a sound with a loudness of 10 sones. It has a linear relation starting from 1 sone as shown in Figure (1.6).

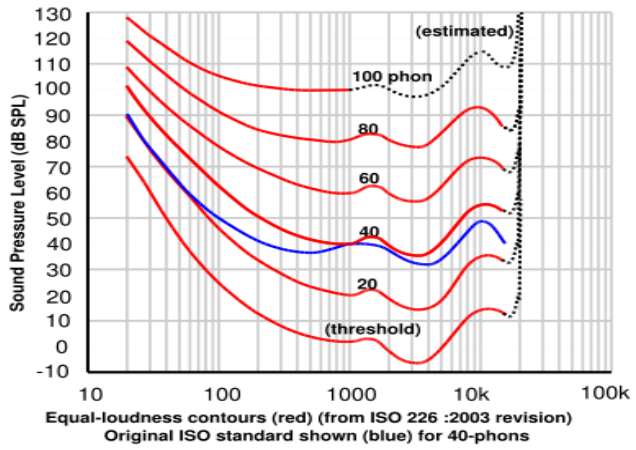


Figure (1.5): Equal loudness contours (ISO 226).

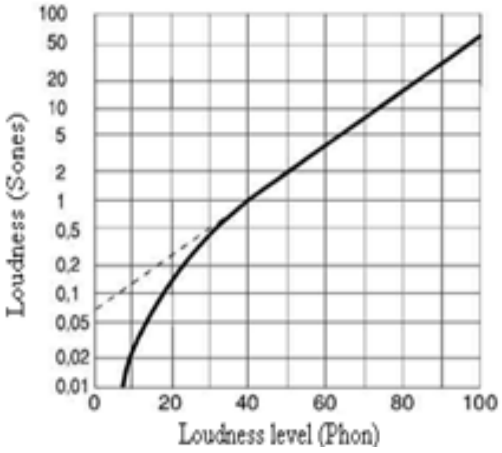


Figure (1.6): loudness levels relation.

Orchestra sound covers a range of intensity levels from 40-100 phon that corresponds to a range from about 1 to 50 sones. The human ear can detect small changes in loudness. The most common music loudness levels are pianissimo, piano, mezzopiano, mezzoforte, forte and fortissimo (Pierce 1983). Table (1.2) shows music different levels and their corresponding A-filtered loudness levels (Hall 2001). Since SPL is frequency dependent as shown in Figure (1.5), thus the effect of different frequencies do not represent the total equivalent ear perception. Therefore, it is common in acoustics to use a certain filter called "*A-Filter*" to calculate the equivalent effect of all sound components on the ear and the output is represented in dBA.

Music term	Perceived loudness	Sound pressure (dBA)
pianissimo	Very soft	50
piano	Soft	60
mezzopiano	Moderate soft	66
mezzoforte	Moderate loud	76
forte	Loud	80
fortissimo	Very loud	90

Table (1.2): Main loudness levels in music (Hall 2001).

Critical band (CB) is a psychoacoustic aspect of hearing. It results from the way the ear uses to resolve frequencies. The auditory system is thought to contain an array of overlapping band pass filters (BPFs) known as 'auditory filters' (Fletcher 1940, cited and calculated in (Zwicker

1961)). The band width (BW) of the auditory filters is called the *critical bandwidth*. They are non-linear, level-dependent and the BW increases from the apex to the base of the cochlea. Two simultaneous pure tones that are close enough in frequency and fall in the same CB give rise to audible beats. If two tones are too far apart to produce beats, they produce a sensation of roughness. If the frequency interval increases more than 1 CB, tones are heard separately. At low and moderate sound levels, frequency components lying farther apart than 1 CB send signals to the brain over separate groups of nerve fibers, but frequency components lying in the same CB send a mixed signal over the same fibers. To calculate the effective intensity level of overtones existing within a CB, the actual intensities of each overtone are added before converting the intensity level into dB (Pierce 1983). The BW of CBs is roughly directly proportional to frequency starting from 500 Hz with a slope *0.2f* (Zwicker and Fastl 1990). If two tones are separated by a minor third (ratio: 6/5) or more they will sound consonant together. For tones made up of many (e.g. six) harmonic overtones, one can avoid serious beats or roughness by playing chords whose fundamental frequencies have integer ratios shown in Table (1.1) (Pierce 1983). CBs are measured on a Bark scale named after Barkhausen (1954) who first proposed a subjective measurement of loudness (Barkhausen 1954). The Bark scale is a psychoacoustical scale proposed by Zwicker in 1961 (Zwicker 1961). The scale ranges from 1 to 24 and corresponds to the first 24 critical bands of hearing. It has a maximum value of 24 at 15 kHz (Terhardt 1998). Each critical band has a corresponding bark value calculated from psychoacoustic measurements. The bark scale has a quadratic representation with respect to frequency. The maximum Bark value is 24 at 15 kHz (Terhardt 1998). Bark scale is calculated from Equation (1.7) (see Figure (1.7)). Another frequently used scale to describe the psychoacoustically relevant filter bandwidth is the equivalent rectangular bandwidth (ERB)(Glasberg and Moore 1990).

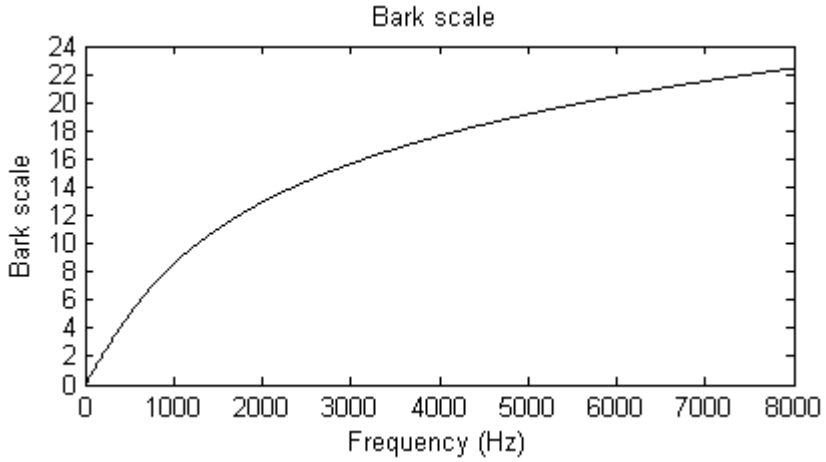


Figure (1.7): Bark scale versus frequency.

$$Bark = 7.\ln\left(\frac{f}{650} + \sqrt{1+\left(\frac{f}{650}\right)^2}\right) \tag{1.7}$$

Acoustic sound in general is described in three physical dimensions (intensity, frequency and time) that correspond to perceptual dimensions (loudness, pitch variation and duration). “Music is made up of sounds that can be organized into three elements: melody, rhythm and harmony” (Wyatt and Schroeder 1998). Music has many important aspects and some of them are represented in Figure (1.8).

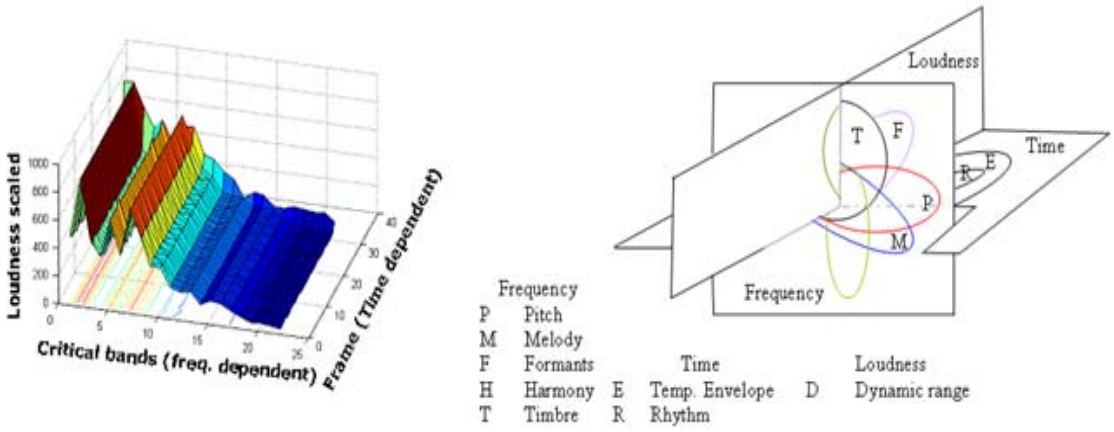


Figure (1.8): Representation of some important music aspects

1.3 Processing Strategies (Literature Review)

CI devices are programmed differently using a variation of algorithms. They were developed over the last decades. Signal processing algorithms in the external hardware of CIs are mostly based on „speech coding strategies" because they were historically developed to improve speech recognition. Earliest devices used a feature extraction approach and reduced the stimulation patterns afterwards to electrodes. Since then, many speech processing strategies are based on extracting speech features as formants or splitting the frequency spectrum into bands, whose sizes are chosen to maximize speech comprehension. This section provides a review of previously developed processing strategies. Not all of them were tried for music perception. Results will be described if available.

In this section, strategies are categorized according to one of three domains in which the main emphasis was put. Figure (1.8) shows a structural organization of different strategies. Some strategies are designed for both frequency and time domain processing.

1.3.1 Frequency Domain

Pythagoras was the first to correlate between pitch and frequency, when he noticed that shorter strings vibrate faster (Wightman 1981; Pierce 1983). Ohm formulated the definition of tone, stating that a tone of frequency (*f*) is heard when a complex sound which contains $2a\sin(2\pi(f+\theta))$ as a component is played (Plomp 1964). Helmholtz suggested that the BM contains transversely stretched fibers where by each resonates at different frequencies and thus the cochlea acts as a mechanical spectral analyzer. He suggested that the ear acts as a fairly broad spectrum analyzer and that the eardrum and middle ear are nonlinear and therefore would produce distortions. He proposed the place pitch theory of frequency analyzers (Helmholtz 1863). The psychoacoustic phenomenon known as "missing fundamental" involves tonal perception of complex waveforms when a listener can perceive a clear sensation of the fundamental frequency even if no energy is present at the fundamental frequency (Wightman 1981). In 1928, Békésy inspected traveling wave patterns of mechanical motion in the inner ear using physical models of the cochlea and found that the maximal excitation was related to the applied frequency (Bekesy 1960). Greenwood (1961) proposed a cochlear frequency to position tonotopical function. The function covered 35 mm of the cochlea, proposed 35 critical bands of 1 mm each, where spectral components can be heard if they are wider than a CB. In the NH cochlea the Greenwood function calculates the frequency of the tonotopically perceived pitch as follows: $Pitch = 165.4 * \left(10^{0.06\ell} - 1\right)$ Hz (Greenwood 1990), where *l* is the distance in mm from the apex.

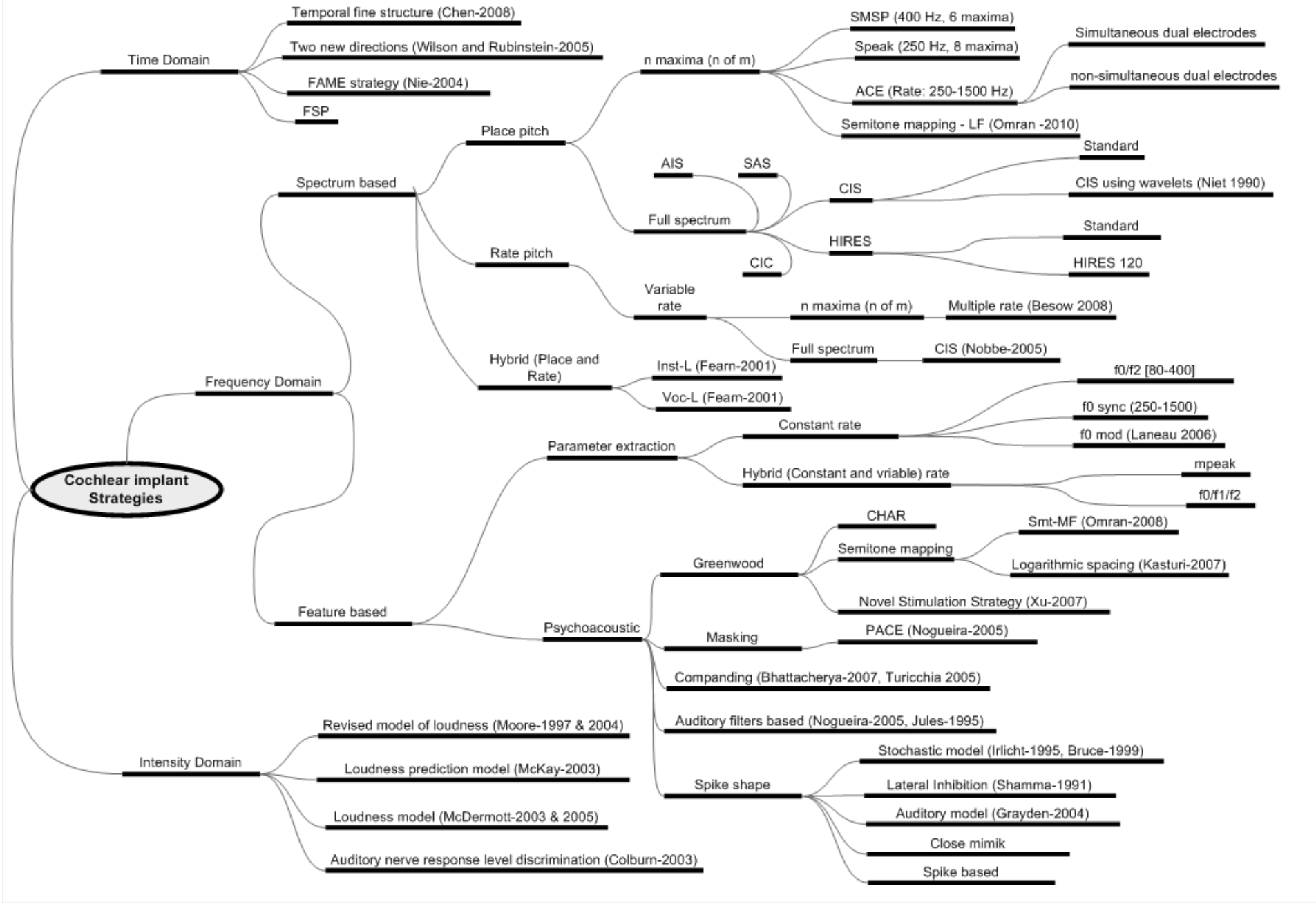


Figure (1.9): Overview of most of the available cochlea implant strategies

Greenwood plotted data from Mayer (1894) against the estimated size of CBs. Mayer collected the data when the listeners were asked to judge smallest consonant intervals between two pure tones. Greenwood pointed out that dissonance was judged as absent when the distance between two pure tones was roughly the size of a CB (Greenwood 1991). In studies with CI recipients a mismatch was found between pitches calculated according to Greenwood and pitches CI recipients perceive (Sridhar et al. 2006). The place theory has difficulty in explaining some type of pitch perception such as the missing fundamental phenomena. Experiments of this phenomenon by Schouten (1938) and later by Licklider (1954) showed that the perceived pitch corresponds to a fundamental that was not present at the receptor level, suggesting another mechanism that must exist for pitch perception of missing fundamentals (Schouten 1938; Licklider 1951). Goldstein (1973) suggested that there exists internalized harmonic templates against which incoming resolved spectra can be compared (Goldstein 1973). A biologically plausible model for how such templates can form in the early stages of the auditory system was presented by (Shamma and Klein 2000). The frequency alone is not sufficient to describe the pitch (Zwicker and Fastl 1990; Hofman-Jablan 1995). Loeb et al. (1983) proposed a spatial cross correlation as a mechanism for pitch perception. They suggest there may be three different mechanisms for pitch extraction operating over different frequency ranges. The first using place pitch in the 5-20 kHz range, the second using periodicity rate pitch in the 500-5000 Hz range, tracking the stimulation frequency by utilizing discharging information while the third mechanism uses pure rate pitch, ranging from 20 to 400±100 Hz; in this range the nerves are capable of firing on every cycle (Loeb et al. 1983). The rate theory mainly depends on the idea that energy peaks in the spectrum (e.g., formats) produce highest BM displacement at corresponding best frequency (BF) locations. Two measures have been used to characterize responses in the auditory nerve: average firing rate and phase-locked rate (Delgutte and Kiang 1984; Carney and Geiseler 1986). The rate of neural firing is a direct relation to the rate of stimulation (Langner 1992; Fearn 2001) and an increase in the stimulation rate on a single electrode yield changes in pitch perception until a critical value of about 300 pulses per second (pps), after which changes in rate are only perceived as changes in loudness; Landsberger (2004) reported that the critical rates between 300 and 1500 pps are indistinguishable from each other (Landsberger and McKay 2004).

1.3.1.1 Place Pitch Strategies

Maxima selection strategies are categorized according to the number of maxima extracted from the sound spectrum. There are two categories; n-of-m and m-of-m strategies. In the first, a certain number of maxima (*n*) is extracted from the full set of all spectral maxima (*m*) while in the second all maxima are utilized respectively. The difference between the two categories and their underlying processing strategies is discussed in this section.

(A) n-of-m Strategies

1- Spectral Maxima Sound Processor and Spectral Peak

In the early 1990s a strategy called spectral maxima sound processor (SMSP) was developed at the University of Melbourne for the Nucleus multi-electrode CI (McDermott et al. 1992). In this strategy the six largest maxima from a filter bank, containing 16 BPFs in the range [200-5000] Hz, were used to stimulate the electrodes with a constant stimulation rate 250 Hz. The SMSP processor did not use formant extraction techniques as the preceding Multi-Peak (MPEAK) strategy still did. The used electrodes were the 16 most apical electrodes available in the Nucleus CI 22 array. This speech processing strategy was compared to the MPEAK strategy in which four electrodes were selected for every glottal

pulse period. A significant increase in performance with vowel recognition was reported (McKay et al. 1992). The intensity levels of the stimuli followed an approximately logarithmic function (McDermott and al. 1993). A commercial implementation of the SMSP strategy was developed by increasing the number of BPFs upto 20 with the same rate and was introduced as the spectral peak (SPEAK) strategy. SPEAK calculated the energy in each BPF and selected upto 10 maxima with largest energy amplitudes (Loizou 1998). MPEAK and SPEAK were compared using a word and sentence tests. A significant improvement with SPEAK was reported (Skinner et al. 1994; Dillier et al. 1995).

2- Advanced Combination Encoders

The ACE strategy implemented also for the Nucleus CIs, is an extension of the SPEAK strategy with the option to stimulate at higher rates. The rate of stimulation and the number of channels could be optimized subjectively to maximize the information transfer (Loizou et al. 1999). This strategy allows to combine the rate and maxima variation. The total pulse rate was limited to 14400 pps originally and extended upto 35000 pps in the Freedom implant. Pansini (2002) compared ACE to SPEAK strategy with nine deaf children who received a Nucleus “CI-24M” CI and were fitted with a SPrint speech processor. Words and common phrase speech recognition tests in noise and quite showed a significant improvement with ACE for open-set words and sentence recognition in quiet and noise (Pasanisi et al. 2002). Music quality was tested using the two strategies (ACE and SPEAK) and the results suggested that neither of them enables highly satisfactory music appreciation; all CI recipients appraised music that involved multiple instruments to sound less pleasant on average than music played by a single instrument (Looi et al. 2007). Gfeller (2005) tested recognition of "real-world" musical excerpts with CI recipients using SPEAK, ACE and continuous interleaved sampling (CIS) strategies and concluded that they are not effective in transmitting several musical aspects (i.e., pitch, harmony and timbre), if compared to recognition of well-known musical selections (Gfeller et al. 2005). The three (SPEAK, ACE and CIS) strategies do not differ so much in response to musical activities (Brockmeier et al. 2007). However, the ACE strategy was recommended for speech as an initial choice specifically for the Nucleus “CI-24M” CI (Kiefer et al. 2001).

3- ACE with Semitone Mapping

The Nucleus CI electrode array has a length of 25 mm and the insertion depth is around 20 mm which corresponds to a tonotopical frequency around 500±100 - 6000 Hz (Whitford et al. 1993). However, some speech processors map the acoustic frequency range 188 Hz to 8 kHz onto these electrodes. Although this mapping preserves the entire range of acoustic frequency information, it results in a compression of frequency components delivered to the brain (Baskent and Shannon 2003). Baskenet et al. (2003) tested the effect of frequency compression and suggested that speech recognition, at least without training, is dependent on mapping of acoustic frequency information onto the appropriate cochlear positions. Kasturi et al. (2007) first suggested a semitone mapping for Clarion implants. Omran et al. (2008) extended his idea and presented two semitone mapping ranges (130 Hz to 1502 Hz and 440 Hz to 5009 Hz) for Nucleus CI. Semitone (Smt) mapping is believed to preserve the representation of harmonic structure of the overtones because it stimulates different group of electrodes with different tones. Pitch discrimination in addition to melody contour identification (MCI) was tested with NH subjects and a significant improvement over standard (Std) mapping was found (Omran et al. 2008; Omran et al. 2010) (see section 2.4 for details about Smt mapping)

(B)- m-of-m Strategies

Many speech processing algorithms generate stimulating pulses for all frequency channels of the sound spectrum irrespective of their instantaneous amplitude to mimic electro-mechanical functions of the normal cochlea and are therefore called m-of-m strategies in contrast to the n-of-m strategies described in the previous section.

1- Continuous Interleaved Sampling

CIS is a widely used strategy (Wilson et al. 1991) in which brief pulses are presented onto electrodes at a rate of 1000 pps or higher to decrease channel interaction. Pulse trains are modulated with the signal envelope in different frequency bands. Guerts (1999) proposed an envelope enhanced continuous interleaved sampling (EECIS) strategy to emphasize the rapid adaptation seen in the response of auditory nerves to sound stimuli. It was observed that EECIS produced a significant improvement in speech intelligibility in some conditions, while there was no significant decrease in the transmission of speech features with respect to CIS (Guerts and Wouters 1999). In 1998 an implementation of CIS strategy using wavelets was presented. The proposed algorithm was claimed to be faster than the standard CIS algorithm (Niet et al. 1998).

2- Continuous Interleaved Sampling with Logarithmic Spacing

Kasturi & Loizou (2007) investigated a basic harmonic spacing map with 12-electrodes over a limited frequency range using Clarion CII implants. The results were compared at Medel to CIS on the Hilbert transform (HT)to calculate better representation of envelope to NH subjects by hearing simulations from an acoustic model. They reported that melodies processed with a semitone filter spacing were preferred over melodies processed by the conventional logarithmic filter spacing (Kasturi and Loizou 2007).

3- Channel Interaction Compensation (CIC)

MED-EL developed a coding strategy that stimulates multiple electrodes simultaneously called channel interaction compensation (CIC) to overcome potential channel interaction that can arise from parallel stimulation. The CIC strategy calculates the SOE in in real-time and automatically reduces the potential for channel interaction when electrodes are fired at the same time (Nobbe et al. 2007).

4- Simultaneous Analog Stimulation

This strategy activates all electrodes simultaneously using analog (sinusoidal) waveform rather than pulsatile stimuli in order from the basal to the apical electrodes. The advantage of this stimulation mode is that it resembles more closely sound transduction in the inner ear and eventually provides more sound information within each time sequence. The disadvantage is the channel interaction; in which electric fields from one electrode interfere with adjacent fields. The Clarion Hi-Focus electrode was designed to minimize channel interaction, but not eliminate it. SAS uses bipolar stimulation with 8 simultaneous channels activated (Mishra 2000). Desa Souza (2003) tested the following strategies SPEAK, ACE, CIS and simultaneous analog stimulation (SAS) and found no significant difference (Souza et al. 2003). Paired pulsatile sampler (PPS) is an intermediate strategy between CIS and SAS; in which two electrodes apart by four electrodes are simultaneously simulated (1+5, 2+6, 3+7 or 4+8) (Büchner et al. 2005). It uses a monopolar pulsatile stimulus at the rate of

1600 Hz. Quadruple pulsatile stimulation is similar to PPS in idea but four odd or even channels are stimulated simultaneously using a rate of 3300 Hz. Hybrid analog pulsatile (HAP) strategy uses a mix between SAS and CIS; in which electrodes from 1-5 follow SAS strategy and electrodes 6-7 follow CIS. It uses a mix of analogue and pulsatile stimuli for monopolar and bipolar pulses (Loizou et al. 2003a). Results in young and old prelinguals were superior to multi electrode implants, and best with SAS and PPS strategies (Souza et al. 2003).

5- Compressed Analogue

Compressed analogue (CA) is one of the earliest strategies that were used with the Clarion implant. It provided a purely analog waveform for electrical stimulation, whereas the SAS strategy approximates the analog waveform in a staircase-type fashion using a 75 μs temporal resolution. The SAS scheme is similar to that used in the CIS strategy with two notable differences; only seven channels are stimulated and without employing an envelope detector (Loizou et al. 2003b).

6- Multi Pulsatile Sampling

Multi pulsatile sampling (MPS) or quadruple pulsatile sampling is a strategy used with Clarion devices only; it is a hybrid strategy between CIS and SAS, in an attempt to gain some of the theoretical benefits of SAS without the drawback of channel interaction. In MPS, two electrodes or more are stimulated simultaneously, but they are far apart so as not to experience channel interaction. Like SAS and MPS it uses bipolar stimulation and 16 electrodes to activate 8 channels simultaneously. The results from a speech test with MPS, CIS and HiResolution (HiRes) strategies were compared together using a Clarion II implant and it was reported that the performance increases significantly with the HiRes strategy (Ostroff et al. 2003).

7- Asynchronous Interleaved Sampling (AIS)

Sit (2007) proposed a way to stimulate phase information using CIS strategies. The input signal was multiplied with $\sin(\omega_0)$ and $\cos(\omega_0)$, where ω_0 is the characteristic frequency (CF) of each channel. Bandpass filtering each of them and multiplying the output with the same previous modulator ($\sin(\omega_0)$ and $\cos(\omega_0)$ respectively) with the additional processing steps to modulate channel activities in CIS (Sit et al. 2007). This was based on suggestions from Nie et al.(2005). AIS was reported to encode fine phase timing information (Ji-Jon et al. 2007).

8- HiResolution

This strategy was developed by Advanced Bionics Corporation for Clarion implants (16 electrodes). It stimulates the entire spectrum on all channels in an increasing order with a high rate that can increase upto a total of 83,000 pps. In addition, current steering is used; in which two electrodes are stimulated simultaneously with different current levels causing the perception of an intermediate pitch. In the first generation of HiRes, incoming sound is filtered into 16 spectral bands. The energy of each band is extracted, and its envelope was used to modulate a high-rate pulse train that was transmitted to corresponding electrodes. In HiRes 120, the input signal is analyzed in greater spectral detail (than with conventional processing) to achieve a maximum of 120 spectral bands. The HiRes was compared to other strategies from Advanced Bionics such as SAS and CIS. Bosco (2005) reported that

there is a significant improvement with HiRes compared to SAS and CIS (Bosco et al. 2005). However, other studies did not show the same sort of improvements.

1.3.1.2 Rate Pitch Strategies

Townshend et al. (1987) investigated the influence of level change, stimulation rate and location on pitch perception with CI recipients (Townshend et al. 1987). It was found that rate variation has a strong effect upto a saturation level between 200 and 300 pps (for some recipients it may rise upto 1000 pps). It is suggested that the upper limit for temporal coding with implants should range between 200 Hz and 1 kHz (Pijl and Schwarz 1995b; Fearn and Wolfe 2000). In this range, temporal coding alone is sufficient for the perception of musical pitch (Pijl and Schwarz 1995a; Pijl 1997). Although the boundary of using temporal cues for pitch estimation was about 250 Hz, it could be used to recognize melodies upto 800 Hz for some subjects (Chen 2003).

Nobbe (2004) developed a coding strategy where six electrodes were switched adaptively between a high (1515 pps) and a low (252 pps) rate, with the aim to produce two intermediate pitch percepts at one electrode. Although the strategy could only be tried out in the laboratory and not with wearable devices, it was preferred by most test subjects for music. It was claimed that the strategy could be improved by fitting varying stimulation rate in each channel individually to create a type of tonal scale for all combinations of place and rate of stimulation (Nobbe 2004). Another variable rates strategy for the Nucleus CI was developed in (Besouw 2008).

1.3.1.3 Hybrid Place and Rate Pitch Strategies

Fearn (2001) developed two strategies (Instrument-L and Voc-L) that make use of constant and variable rates. Instrument-L used a sampling rate 12 kHz, fast Fourier transform (FFT) bin width of 93.75 Hz and 10 channels using variable rates. In general, ten channels were activated with frequencies below 1 kHz. Instrument-L strategy intended to improve music perception but may degrade speech perception. Voc-L used a sampling rate of 16 kHz, FFT bin width of 125 Hz and five channels for variable rates. It used also ten channels for frequencies below 1 kHz. Voc-L strategy is intended to improve music perception while not degrading speech perception (Fearn 2001).

In Music-L strategy the algorithm used a short time Fourier transform (STFT) and the output was splitted into two parallel pathways. The first pathway used a phase vocoder; and the phase difference was converted into a change in rate activity which was limited with a rate limiting block. The second pathway calculated the amplitude and selected the maxima from the power spectrum. The information from both pathways was collected and synchronized using a timing block. Pitch ranking and just-noticeable differences (jnd) pilot tests reported that Voc-L improved music representation with vocals while Instrument-L improved instrumental music representation and sounded more natural (Fearn 2001).

1.3.2 Feature Extraction Strategies

Speech is produced due to vibrations of vocal chords and is shaped by moving the tongue which changes the resonance cavity in the throat and the mouth. Since the early implant that used a formant based processing strategy, there were different trends to extract sound features to enhance speech comprehension. This started by extracting important formants that were stimulated using place pitch and in a later stage different modulation schemes were applied to benefit from rate variation in pitch enhancement. In more complex strategies psychoacoustic

features were implemented (e.g. masking). This category of strategies includes three sub-groups:- place pitch, both rate and place pitch and psychoacoustic feature extraction.

1.3.2.1- Place Pitch

1- *f0* synchronized ACE

The *f0* synchronized ACE (*f0*Sync) strategy was developed by the CI research group at the University of Leuven (Milczynski et al. 2008). The *f0*-extractor is based on an autocorrelation function applied to the input sound. It uses filter banks similar to ACE and then a low pass filter (LPF) of 50 Hz in analyzing the signal.

2- *f0* mod

In the *f0*mod scheme, slowly varying channel envelopes were amplitude modulated sinusoidally with the fundamental frequency (*f0*) of the input signal, using 100% modulation depth in phase between channels. The results showed that *f0* discrimination enhances pitch ranking and melody recognition of Flemish songs with CI recipients but the results were not significantly different than with the ACE strategy (Laneau et al. 2006b).

3- Multi-Channel Envelope Modulation

Another strategy called multi channel envelope modulation (MEM) followed ACE and *f0*Sync strategies in the concept. It provided *f0* periodicity information synchronously across all activated electrodes. However, rather than employing an *f0* estimator, MEM utilized the envelope of the input sound (which inherently contains *f0* periodicity information) to modulate the electrode levels derived from the ACE strategy. The envelope of the signal was extracted with a full wave rectifier followed by a 300-Hz, fourth-order LPF. It provided deeper modulation cues of f0 than ACE strategy. MEM was reported to provide a significant enhancement in pitch ranking tests with CI recipients. However, speech perception results were equivalent to the ACE strategy (Vandali et al. 2005).

1.3.2.2- Hybrid Place and Rate Pitch

1- f0/f2

This was the first implemented strategy in CIs in 1980 (Seligman et al. 1984; Clark 1987); it mainly depended on extracting the fundamental frequency *f0* (pitch) and the second formant *f2* from the sound. Rate of stimulation varied according to the energy found between [80-400 Hz] where the fundamental frequency of the voice was expected. The second formant *f2* was extracted with a BPF [800-2.3kHz]. Initial results were promising (Dowell et al. 1986). If voicing information was present, the stimulation rate was at *f0*; otherwise the rate fluctuates with the envelope of the signal (Dowell et al. 1987; Seligman 1987). This strategy switched all electrodes to stimulate at *f0* rate upon detection of voiced energy and provided the best male/female identification whereas a standard CIS strategy produced chance identification of male and female voices. This suggested that temporal presentation of voice pitch was important for detection of pitch related speech feature. It was found that stimuli with two component pulse rates were perceived to have two perceptual dimensions. The information presented on the more apical electrode pairs was more likely to be received better than if it was presented on a basal electrode. Information transmission was better encoded when pulse rates were below 200 pps (Lai 1990).

2- f0/f1/f2

This was a next step after the f0/f2 strategy by including a BPF [280-1000 Hz] and using a zero crossing algorithm to capture the first formant (*f1)* (Blamey et al. 1987). This strategy selected the five most apical electrodes to stimulate *f1* and the remaining electrodes to stimulate *f2*. The stimulation rate was controlled by the *f0* value for voiced signals while the unvoiced signals were stimulated at a rate of 100 pps. The pulses used were biphasic with 200 μs pulse width and with 800 μs interpulse interval to avoid interference between stimulated electrodes. The addition of *f1* component improved speech recognition with both NH subjects listening to a simulation reconstructed from encoded signals and CI recipients. Average correct word recognition increased from 30% using *f0*/*f2* to 63% using *f0*/*f1*/*f2* (Dowell et al. 1987) and improved for monosyllabic words from 8% using *f0*/*f2* to 28% using *f0*/*f1*/*f2* (Tye-Murray et al. 1990). It was found also that it improved vowel recognition and did not improve consonant recognition.

3- Multi-Peak

MPEAK differed slightly from *f0*/*f1*/*f2* strategy by including high frequencies that contribute to consonant perception. Electrodes had a pulsing rate depending on *f0* and there were 3 Electrodes for the first Formant (*f1)* and the rest for *f2*. Electrodes 1, 4, 7 were stimulated with the envelope of frequency ranges [4-6k], [2.8-4k] and [2-2.8k] respectively. The formants were extracted the same way as in *f0*/*f1*/*f2* strategy (Patrick and Clark 1991). This strategy was tested with consonant perception and a mean improvement in consonant identification and open-set sentence recognition was reported (Dowell et al. 1991; Skinner et al. 1991; Wallenberger and Battmer 1991). In another study that compared music perception in adult Nucleus CI recipients using *f0*/*f1*/*f2* and MPEAK, it was found that neither one of them seems clearly superior for sequential pitch or rhythmic patterns recognition (Gfeller et al. 1997b).

1.3.2.3 Psychoacoustic Feature Extraction

Many strategies make use of psychoacoustic features such as Greenwood, Masking and mimicking spike shapes. Greenwood proposed a frequency to place function of the cochlea in the normal ear, and this was used in some strategies to align CF of electrodes with their tonotopical frequencies. Different strategies in this group are addressed in this section.

A- Greenwood

1- Novel Stimulation Strategy

In 2007, a novel stimulation strategy was presented. It is generally similar to the ACE strategy but was processed differently with a Greenwood like harmonic spacing map. The method extracts 3rd, 4th, 5th, and 6th harmonics of a tone and stimulated the corresponding tonotopical position (Xu and Wei 2007). The strategy was tested with NH subjects acoustically and the results claimed an improvement in melody recognition. There is a similarly to Smt mapping; it corresponds to semitone mapping (Smt-LF) from the analysis point of view and to semitone mapping (Smt-MF) from the stimulation point of view where Greenwood function is used (see section 2.4 for more details regarding Smt mapping).

2- Semitone Mapping for Mid-frequency (Smt-MF)

Omran et al. (2008) presented a harmonic spacing strategy called Smt-MF, in which a frequency band (440-5009 Hz) is analyzed and mapped to corresponding Greenwood electrodes (Omran et al. 2008; Omran et al. 2010). This strategy filtered out frequency components below 400 Hz assuming the implant insertion depth is between 21 and 26 mm. Baskent and Shannon (2004) reported an increase in speech recognition when the center frequencies of electrodes were aligned with Greenwood frequencies (Baskent and Shannon 2004). Smt-MF was tested using MCI (Galvin et al. 2007) with NH and with CI recipients. An improvement in melody recognition compared to standard ACE mapping in high frequencies was found while there were pitch reversals at low frequencies due to filtering out low frequencies including the fundamental. They reported that this may not be enough for musical instrument identification (Omran et al. 2008). Oldenburg sentences test (Thornton and Raffin 1978) was used to evaluate this strategy with NH subjects and no decrease in word identification compared to ACE was observed.

3- Characteristic Frequency Mapping Strategy

With the standard frequency to electrode mapping, a frequency range upto 8 kHz is assigned to 22 electrodes in a Nucleus implant. However, the actual electrode positions have different tonotopical frequencies starting from 500±100 to 6000 Hz (Whitford et al. 1993) assuming an average human BM length of 32 mm and an electrode insertion depth of 25 mm ranges according to Greenwoods formula. This compression was examined and it was found that it had a side effect on speech comprehension whereby the best results were obtained when the input acoustic information was delivered to the matching tonotopic place in the cochlea with a minimum frequency-place distortion (Baskent and Shannon 2003; Baskent and Shannon 2004). The characteristic frequency mapping strategy (CHAR) mapping was developed at the University of Melbourne, using the MPEAK strategy and altering frequency boundaries of the electrodes. In this strategy, the electrode with a tonotopical CF equal to ~3 kHz was fixed and the more apical electrodes were re-allocated to cover the whole *f1* and *f2* frequency ranges. Thus, the *f2* frequency range is shifted closer to the CF region. It was reported that patient results slightly improved in sentence tests (Whitford et al. 1993).

B- Psychoacoustic Advanced Combination Encoders

Nogueira et al. (2005) presented a variation of the ACE strategy to include psychoacoustic masking curves. The strategy was similar to the processing algorithm in ACE, where a number of maxima (n) was selected and the corresponding channels were stimulated. The strategy was called psychoacoustic advanced combination encoder (PACE). PACE used masking curves to determine essential spectral components in a way to increase the temporal resolution. PACE was tested with CI recipients using 4 and 8 channels stimulated per cycle. There was a mean improvement over ACE in the 4 channels case but the difference was not significant (Nogueira et al. 2005). Other studies reported no advantage of PACE over ACE in terms of music perception, although familiarity of subjects with ACE may have been an important factor (Lai and Dillier 2008).

C- Spike Shape

Some neurons in the central auditory system can inhibit spike activity generated by ascending pathway cells (Paolini et al. 2004). Such a topographic representation improves spectral

contrast. In 1977 a BM model was presented that was calculated from the response with pure tones (Deutsch 1977). This work was extended and in 1985 a uniform lateral inhibition network's model was described to be capable of extracting complex spatiotemporal firing patterns from the cat's auditory nerve (Shamma 1985b; Shamma 1985a). A couple of years later, the lateral inhibition concept was employed in an auditory neural network model for speech recognition in noise and the results reported an enhancements (Yan et al. 1998; James et al. 2004). Unfortunately, the lateral inhibition as a separate block is still not implemented in CIs. However, there are other approaches to mimic these effects such as "*better mimicking"* (Wilson et al. 2005) and spike-based temporal auditory representation (STAR) (Grayden et al. 2004). The STAR was developed to mimic the spiking activity generated in the brain to become closer to that in NH with the MPEAK strategy.

D- Auditory Filter

Helmholz hypothesis of 1877 that the ear uses a filter bank to analyze the spectrum was quite fruitful. Johannesma (1972) presented a Gamma tone filter (GTF) that characterized physiological impulse-response data gathered from primary auditory nerve fibers in the cat. Later the GTF was established (de Boer 1975). The bandwidth of the auditory BPFs depends on the CF of the filters and is often described by an ERB. For moderate sound levels, the relation between ERB and the CF (*fc)* in Hz is described in (Glasberg and Moore 1990). Current speech processing strategies for CIs use a filter bank which decomposes audio signals into multiple frequency bands associated with corresponding electrodes. Nougeira (2006) proposed three filter banks, based on a Wavelet structure transform with different basis functions and reported a significant enhancement in speech recognition (Nogueira et al. 2006).

E- Companding

The auditory system uses nonlinear processing to provide the necessary spectral and temporal resolution. In the cochlea, the OHC are responsible for the active mechanisms observed in the peripheral auditory system. The OHCs employ nonlinear adaptive gain processing to encode a large dynamic range with a relatively narrow physiological range of auditory nerve fibers. The same OHCs are also responsible for the sharp frequency selectivity in the auditory system. Another nonlinear phenomenon arising from complex interactions between OHCs and the BM is the two-tone suppression (Rhode 1974). It is characterized by a decrease in the evoked response to a tone in the presence of a second tone (Sachs and Kiang 1968). Two-tone suppression is considered to be the primary mechanism underlying spectral enhancement and is thought to improve the signal to noise ratio (SNR) of stronger components (Rhode et al. 1978; Sachs et al. 1983 ; Stoop and Kern 2004). Spectral enhancement or "spectral sharpening" is described as an increase in peak-to-valley ratio (Bhattacharya and Zeng 2007).

Some algorithms were developed to improve speech recognition like the bio inspired companding strategy for speech enhancement (Turicchia and Sarpeshkar 2005) that was later tested by (Bhattacharya and Zeng 2007). A companding algorithm combined two-tone suppression (Ruggero et al. 1992) and dynamic gain control to increase the spectral contrast. One specific goal in that study was to improve speech recognition in noise with CIs. Companding may be present along the auditory pathway since both the cochlea and the cochlear nucleus perform logarithmic compression on the input signals, while the brain performs exponential expansion (Zeng and Shannon 1999). The companding algorithm was composed of compression and expansions functions. The envelope was detected by a half wave rectifier and then was used to modulate the output from the compression and expansion blocks.

1.3.3 Temporal Domain

Smith et al. (2002) divided acoustic sounds into 64 frequency bands and extracted the temporal envelope and fine structure (FS) using the HT. They switched the envelope of one sound with the FS from another sound and presented chimaeric sounds. By studying the perception of chimaeric sounds, they found that envelope cues strongly influence speech recognition while FS cues influence pitch perception. The envelope alone can provide sufficient information for speech recognition in quiet. But it is not sufficient for perception of tonal languages, speech in noise, music and localization (Nie and Zeng 2004). The importance of FS information was reported to be important for speech recognition using less than 8 channels and for music appreciation using less than 40 channels (Smith et al. 2002; Wilson et al. 2005). FS information in low frequency (LF) bands, when combined with electric stimuli for a coarse representation of high frequencies, may support high levels of speech recognition in noise (Wilson et al. 2002; Gantz et al. 2004; Turner et al. 2004; Kiefer et al. 2005). However, it was suggested that frequency resolution in the LF range can support high levels of music appreciation (Gfeller et al. 2002a; Gantz et al. 2004). But frequency resolution is limited by the available number of channels.

Several methods were studied to encode FS information from acoustic sounds. One method was the phase vocoder, in which the envelope and the phase were extracted from the acoustic sound (Flanagan and Golden 1966). Flanagan later examined various coding schemes and established amplitude modulation (AM) and frequency modulation (FM) methods used in various CI processing strategies (Flanagan 1980). Maragos developed a mathematical model called the energy operator to separate AM and FM components in speech sounds (Maragos et al. 1993).

Nie and Zeng (2004) used the AM and the FM idea and presented a strategy called frequency amplitude modulation encoding (FAME). It extracts the FM parameters using the intensity weighted average of instantaneous frequency (IWAIF) model to extract pitch information from complex tones (Feth et al. 1982). It aimed at encoding slowly-varying AMs and FMs within a frequency analysis band. A significant increase in vowel recognition was reported with increase of FM components (Nie and Zeng 2004).

Modulation depth enhancement (MDE) is another strategy designed to include FS components by increasing the AM depth of signals and using the ACE strategy. The algorithm extracted the modulating frequencies in the range 80–300 Hz encompassing *f0* for nearly all adult males and many females and children. For input signals in which the modulation depth was small and perhaps difficult to utilize as a cue, the algorithm expanded it (Vandali et al. 2005). Vandali et al. (2005) compared MDE with ACE strategy and reported an enhancement in pitch ranking with sung male vowels.

Peak derived timing (PDT) is a testing strategy that was designed primarily to preserve temporal FS. It employed a filter bank with 19 BPFs covering frequency range 200-7100 Hz (Van-Hoesel 2001). The average stimulation rate on each electrode varied with its CF with a maximum of 1400 pps/channel (Vandali et al. 2005). In this strategy positive peaks in the output from each BPF were stimulated onto corresponding electrodes at their corresponding timings.

Chen and Zang (2008) presented a speech synthesis CI model based on FS extraction. Their model used filter banks to analyze a sound signal, the output from each filter bank was band-passed and underwent HT. The peaks of the FS component from the output of HT were extracted and multiplied by cosine wave carriers. The modulated output was then used to

modulate the envelope in the output from the HT. All modulated envelopes were summed to produce the synthesized speech signal. This model was used to test NH subjects and showed an increase in tone recognition of mandarin language (Chen and Zhang 2008).

The AIS strategy (see section 1.3.1.1.B.5) was developed to encode temporal phase information as well. It derives stimulation parameters from the full spectrum and is included with the m-of-m strategies.

MED-EL developed a coding strategy to enhance the fine structure processing (FSP) which is may provide better temporal coding of sounds in the low frequencies that is similar to normal hearing by using a stimulation paradigm that is derived from a sound's fine structure in that frequency range. A significant enhancement with music and speech against the CIS strategz was reported (Arnoldner et al. 2007).

1.3.4 Sound Intensity Domain

The normal ear has a wider dynamic range for sound levels (~ 110 dBA) than what is usually available for CI recipients in current speech processors (~45 dBA). CI devices use a loudness map to convert the real world loudness into the available dynamic range. The dynamic range is adjusted in a systematic way during the rehabilitation phase. Computational models estimate the loudness of incoming sounds, and generate appropriate patterns of electric stimulation. The effect of increasing current level on loudness results more from recruitment of a higher number of neurons than from an increase in the average spike probability of simple neurons (McKay and McDermott 1998).

Changing loudness levels influences pitch estimation reversively. Although for CI recipients this relation is highly subjective (Townshend et al. 1987), similar a relationship between pitch variation with loudness change was reported in (Zwicker and Fastl 1990) for NH. Changing the stimulation rate changes the loudness. This is influenced by a central temporal integration (a time window of several ms). As the stimulation rate increases, the number of stimulus pulses included in the integration window increases and the amount of neural activity evoked by each stimulus pulse decreases due to refractory effects (McKay et al. 2001). The refractory effect is likely to underlie much of the variability between implantees in the way that loudness changes when a fixed-current pulse train changes in rate.

McDermott et al., (2003) developed a loudness model for CIs called "*Spel*". It is based on real-time numerical estimations of loudness. The perceptual performance of an initial version of SpeL scheme was compared to that of SPEAK and ACE using word recognition tests with 5 subjects. Preliminary results showed that SpeL enhanced consonant recognition, at least at a relatively low speech level, but lowered vowel recognition, especially at a higher speech level. An overall increase in phoneme recognition with SpeL at a speech level of 50 dBA was found to be statistically significant. A significant decrease in scores was observed at a level of 60 dBA and was related mainly to the way acoustic frequencies were distributed across the available electrode positions in this version of SpeL (McDermott et al. 2003b). Moore and Glasberg (2004) revised a previous loudness model applicable for both normal and impaired hearing.

1.4 Music Perception with Cochlear implants

Although speech recognition in quite with CI devices can reach levels of more than 80 % (Mueller-Deile 2009) because CI devices were designed primarily to restore speech perception (Gfeller et al. 2000b; Clark 2003) many CI recipients describe music as unpleasant

noise. Various sound aspects are important in music appreciation such as pitch, melody, frequency resolution, timbre and loudness.

Pitch is the perceived property of a sound that can convey a melody. Place and rate pitch are two mechanisms for pitch perception; the first involves sending a pulse to a segment on the BM in the cochlea, which will cause perception of a certain pitch according to its tonotopical positon (Greenwood 1961) while the second involves varying the stimulus rate that influences the perceived pitch (Bregman 1990; Pierce 1991; Warren 1999). The implant may introduce information loss during signal processing such as limitations in frequency resolution beside loss of phase information.

Rhythm in sounds is perceived as more appealing by CI recipients than sounds without a distinguishable rhythm (Gfeller and Knutson 2003). While recognizing musical instruments is a difficult task, this may be due to corrupted temporal information transmitted. Timbre is a subjective perceived attribute, which depends on physical variables such as the frequency contents, spectral profile and FS of the sound (Handel 1995) by which different instruments are characterized (Helmholtz 1954). It is not well represented and the implant does not transmit a replica of musical sound in addition to distorting FS components.

Although listening habits prior to implantation may be one factor to influence music appreciation (Gfeller et al. 2000a) many CI recipients are generally not satisfied with the way music is presented.

1.5 Objectives

Music has many aspects such as rhythm, melody, harmony, beats, meter … etc. However, it can be organized into three main elements (melody, harmony and rhythm) (Wyatt and Schroeder 1998). Since melody is described as a group of pitches that are perceived as single entity (Terhardt 1998), enhancing pitch perception may improve melody representation. Music encoding in the CIs may need a more accurate frequency representation than what is currently provided. One way to address it is using VCs. New frequency-to-channel mappings for CIs are studied in this thesis because other common calling devices were able to successfully transmit music and speech via channels with limited BWs. As examples are transistor radios and telephone line transmission of speech which present fairly appreciable music. Even telephone lines that has a BPF [300-3000 Hz] maintains speech comprehension (Möller 2005). This may mean that stressing on a selected part of the input BW may be beneficial.

1.5.1 Place Pitch and Virtual Channels

Nucleus CI devices provide 22 intra-cochlear stimulation electrodes which are used as 22 stimulation channels. There are physical limitations preventing a substantial increase in number of electrodes although theoretically a higher number of stimulation channels is expected to improve musical note discrimination. One way to do this despite of the physical constraints of currently available electrode arrays would be to use VC. These channels can result in the perception of an intermediate pitch when two electrodes are stimulated simultaneously (Busby and Plant 2005). The current Nucleus CI device could stimulate VC but does not allow to set different current levels of the two electrodes independently in order to achieve so called "*current steering*". Employing VCs increases number of available channels from 22 to 43. This is regarded to improve frequency representation. Resynthesis of channel information into an audio signal using an acoustic model (AMO) which includes VCs allows testing different strategies with NH subjects.

1.5.2 Coding Strategies

The Standard (frequency to channel) mapping in Nucleus CI processors assigns the audible frequency range 250 Hz – 8 kHz onto available electrodes. Due to physical constraints, the implant usually is inserted almost 1½ turns into the cochlea. Since the cochlea is tonotopically organized (Greenwood 1990), the most apical electrode is located at a position that corresponds to a pitch generally between 500 and 1000 Hz in NH subjects (Whitford et al. 1993). This value depends on the insertion depth of the implant (~ 20 mm) and the average cochlear length (~ 33 mm). Such constraints together with the Std mapping produce a frequency compression that may reduce the harmonic representation in music (Baskent and Shannon 2004). Two Smt mappings in different frequency ranges were investigated. The first, Smt-LF, covers a range from 130 to 1502 Hz which includes the fundamental frequencies of most musical instruments. The second, Smt-MF, covers a range from 440 to 5040 Hz, allocating frequency bands of sound close to their characteristic tonotopical sites according to Greenwood's function. Smt-LF, in contrast, transposes input frequencies onto locations with higher characteristic frequencies. It is expected that Smt mapping may ameliorate melody recognition with CI recipients.

1.5.3 Psychoacoustic and Speech Recognition Tests

Psychoacoustics is the scientific discipline which investigates and describes the interactions and the perception of the world of sound and between humans or animals. It encompasses studies of sound perception (Katz et al. 2002). All psychoacoustic tests share the same principle; a sound is presented to a subject (or listener) who has to recognize it from multiple choices with or without a background noise. Pitch ranking tests using pure and complex tones were utilized to evaluate VCs. Pitch ranking, MCI (Galvin et al. 2007) and instrument recognition (IR) tests examined discrimination and identification limits in addition to psychoacoustic tests, speech recognition tests in quite and noise cases condition using Std and Smt mappings with NH subjects. MCI and IR tests were also conducted with CI recipients. In this thesis, a speech recognition test such as sentence test (Thornton and Raffin 1978; Wagener et al. 1999b; Wagener et al. 1999a; Wagener et al. 1999c) was conducted to identify comprehension enhancement quality.

1.5.4 Auditory Filters

NH subjects perceive sound due to displacement of the BM that causes a motion to the outer and inner hair cells. This motion along different positions in the cochlea acts as an auditory filter bank. Glasberg and Moore (1990) derived an approximation of this filter bank referred to as GTF. On the other hand, CI recipients hear sounds due to exponentially decaying electrical fields that stimulate the auditory nerve (8^{th} nerve). The field distribution can be approximated by a spread of excitation function (SOE). An analytic comparison between NH subjects auditory filter banks and CI recipients exponential decay filters may lead to a new way of addressing limitations of electric hearing.

1.5.5 Tools

Many tools were developed in this project from scratch and some were extended from previous work. The AMO is an application used to process sounds using many implemented CI mapping strategies to simulate channel activity. It synthesizes channel activities using sinusoidal and noise approaches into audible sounds. Outputs of the AMO were used to test NH subjects. A more advanced tool is the Checker, which is used to processes sounds with a

selected mapping algorithm according to patient data loaded from a clinical database. It generates stimulus files that can be processed offline and checks them for various types of possible errors before streaming then to a CI recipient. To compare the resynthesized sounds, a Spectral-Contrast tool was developed to illustrate 3d (time frame, critical band, loudness) comparisons. This tool depends on computing psychoacoustic parameters. Judging different simulations analytically was aided by calculations combined in the special test signal sequences and this to compare between amounts of harmonic structure preservation in musical tones using different mappings. As a second analytical index to judge harmonicity preservation, a harmonicity index was developed to estimate the amount of harmonic structures existing in a processed sound with different mappings. It depends on psychoacoustic features such as masking and is based on ISO standards for moving picture expert group (MPEG) coding (ISO-13818-3 1994; ISO-226:2003 revision).

1.6 Thesis Outline

The thesis consists of six chapters and is organized as follows:

Chapter 1 is a general background about the history of electric hearing, explaining important psychoacoustic features in sound and illustrating many aspects of music. It provides a literature review of previous processing strategies used in CI devices.

Chapter 2 presents the current standard mapping in a comparison to newer implemented mapping algorithms such as Greenwood and Semitone mapping together with VCs. Theoretical background and technical details of each mapping are illustrated in separate sections.

Chapter 3 describes some tools in detail that were developed. Among them is the acoustic model which resynthesizes sounds according to a CI model and the “Checker” which is an application used to query patient data from a clinical database, generate electrode stimuli and check for possible errors. Spectral-Contrast is another tool used to determine differences in sounds according to psychoacoustic analysis. Harmonicity index is an index to calculate the probability that overtones in a sound have a harmonic structure. This index was used to compare between simulations of different mappings. Finally, a test and measurement procedure is described and applied to evaluate harmonicity in musical tones using different mapping strategies based on matrix calculation for specific sound sequence.

Chapter 4 illustrates psychoacoustic tests that were conducted; e.g. sentence tests, pitch discrimination, MCI and instrument recognition tests. Experimental results are in separate sections.

Chapter 5 is a general discussion of the presented work and the conclusion.

Chapter 6 represents an analytical approach to study spread of excitation that may support future design of CIs.

Chapter 2 - Mappings

Chapter 2 – Frequency Mappings

2.1 Introduction

Many CI recipients reported that music sounds unpleasant with the standard ACE strategy. One reason is the limitation in frequency representation and distortion in the harmonic structure of music due to compression in LF components. Enhancing music appreciation may be achieved by improving frequency representation using VCs and using different frequency to electrode mapping with CIs. In this chapter VC, Greenwood mapping and Smt mapping are investigated. VCs are produced in Nucleus CIs when two adjacent electrodes are stimulated simultaneously with the same current level; this is perceived as an intermediate pitch. VCs potentially enhance the frequency representation of a sound; they increase the available number of channels from 22 to 43. VCs were implemented in the Std, Greenwood and Smt mappings. Greenwood mapping is a frequency to electrode mapping where frequency bands of an acoustic signal are mapped to electrodes with corresponding tonotopical CFs. The Smt mapping is a technique to preserve harmonic structure representation in music with CIs by mapping fundamental frequencies of semitones to separate channels. Both VCs and Smt mapping were implemented into an AMO. In this chapter different frequency to electrode mappings and their implementation are described.

2.2 Standard Mapping

The Std ACE mapping is the most frequently used coding strategy in the Nucleus CI devices. It is an "n-of-m" strategy, in which a number of maxima (n) is selected out of (m) values to represent frequency bands which contains the highest signal energies.

Figure (2.1) shows a block diagram for the Std ACE mapping. An acoustic input sound sampled at 16 kHz is buffered and undergoes a STFT with 128 FFT points per frame. The output is used to calculate the PSD excluding the DC component. The output bins (125 Hz each) are grouped into bands that equal in number to the number of safe channels; channels that do produce a side effect when stimulated such as simultaneous stimulation of the facial (5^{th}) cranial nerve are disabled. The highest n maxima are selected; where n is a subjective parameter that may differ in each CI recipient's map. Signals in these bands are processed with a loudness growth function (LGF) to adjust the stimulation level for each channel and time frame. A channel time matrix (CTM) as shown in Figure (2.1) is obtained which represents the stimuli in different channels over time.

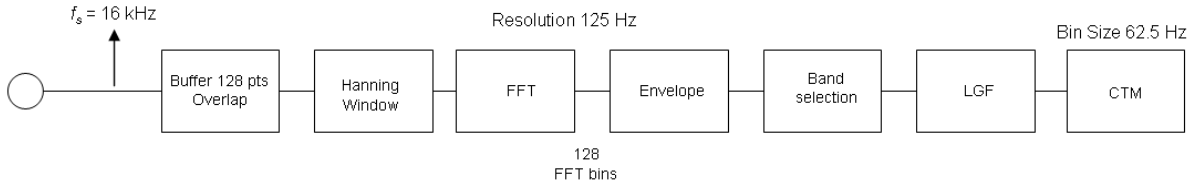


Figure (2.1): Processing block diagram of the standard ACE coding strategy for the Nucleus Cochlear Implant.

In the Std ACE CI processing, the frequency range from 188 to 7980 Hz is mapped to the electrodes. The electrodes have a narrow tonotopical frequency range starting from ~ 500–

1000 Hz to about 6000 Hz in the normal ear according to Greenwood (1960). This algorithm is associated with 22 channels in standard devices. The following section illustrates VCs implementation with the Std mapping.

2.2.1 Virtual Channels

VCs are effective channels that physically do not exist. They are formed when stimulating two adjacent electrodes simultaneously. An intermediate pitch sensation is produced which lies between the pitch sensations produced by the two physical neighboring electrodes (Busby and Plant 2005). Using VCs with the standard Nucleus CI, it is possible to stimulate 43 channels from 22 electrodes. Stimulating an electrode produces an electric field that excites the auditory nerve and leads to a pitch perception. This pitch is generally described as a complex tone; because the field has a SOE that excites a range of tonotopical positions along the hearing nerve. The frequency ranges using 22 and 43 channels with the Std mapping covers are (188 -7980 Hz) and (125 - 7980 Hz) respectively. The following sections describe the VC implementation in detail.

In the 22 channels mode with Std ACE mapping, the available BW is distributed over available channels. Each frequency band with a center frequency (cf) termed *CF* is mapped to a physical electrode. The CF of different channels are plotted in Figure (2.2).

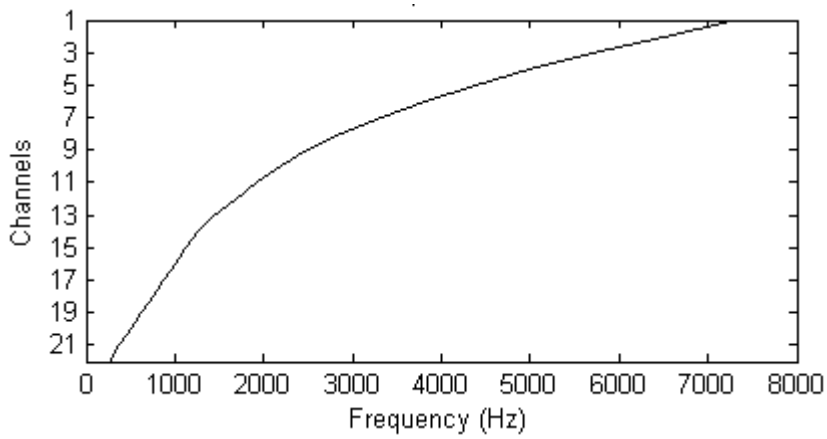


Figure (2.2): Center frequencies of channels in the 22 channels mode. Channel 22 represents the most apical channel (lowest frequency).

Figure (2.2) shows the CF of each channel in the 22 channels mode. Channel number 22 is the most apical one that corresponds to the lowest frequency. In the 43 channels mode, the BW is slightly stretched (125 to 7980 Hz). Assuming that VCs stimulate exactly intermediate midway points and that the electrical impedances in both directions from the mid point between two electrodes are equal, CFs of VCs are assumed to be the mean CF values of neighboring physical electrodes.

In the Bark scale (see 1.2.3), the frequency range (20-15500 Hz) is divided into 24 bands called CBs. Each CB has a BW calculated from psychoacoustic measurements. If two sounds of equal loudness which are separated by more than one CB (about 1/3 of an octave) are played simultaneously, they will be perceived as two sounds and the overall loudness is twice as loud as when each is played separately (Pierce 1983; Terhardt 1998). Band-limited sounds with a sampling rate of 16 kHz maximum frequency, therefore 8 kHz according to Nyquest criteria can reach a maximum of ~ 22 Bark (Zwicker 1961). Figure (1.7) presents the bark scale versus frequency in the ear coordinates.

According to the design of the Std ACE mapping, the input signal has a maximum of 8 kHz (Swanson et al. 2007) and corresponds to the first 22 CBs that are mapped to the 22 physical channels. Theoretically, this produces a linear representation of Barks with respect to channels. This linear representation is extrapolated into 43 channels as shown in Figure (2.3). However, practically speaking, it is not a perfect linear relationship because of discretization errors in calculating the frequency band of each channel since each band contains an integer number of FFT bins.

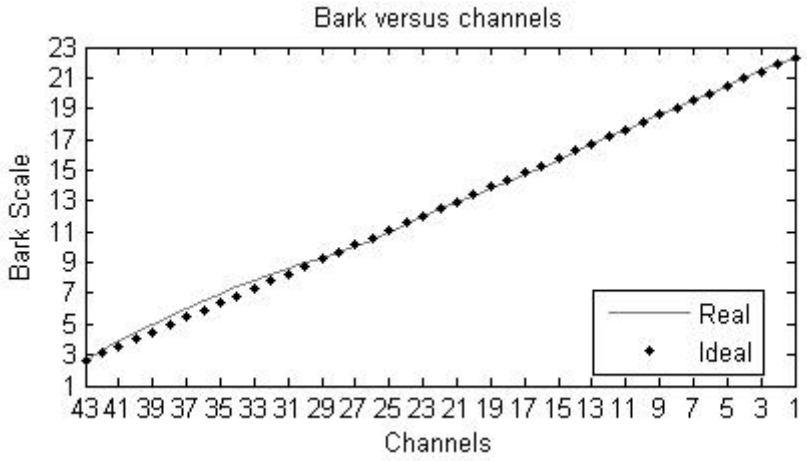


Figure (2.3): Ideal (diamond) and real (dotted) relationship between VC-ACE channels and Bark scale.

Figure (2.3) shows the relationship between channels and Bark scale values. It is obvious that the real relationship is not perfectly linear due to discretization errors in calculating the frequency band of each channel which is composed of an integer number of FFT bins. From Figures (1.7 and 2.3) CFs of channels in the 43 channels mode are reassigned as presented in Figure (2.4). The new distribution of FFT bins uses 256 samples to calculate the FFT which increases the frequency resolution and decreases the FFT bin size to 62.5 Hz. Thus a total of 127 bins without the DC bin are available for further processing.

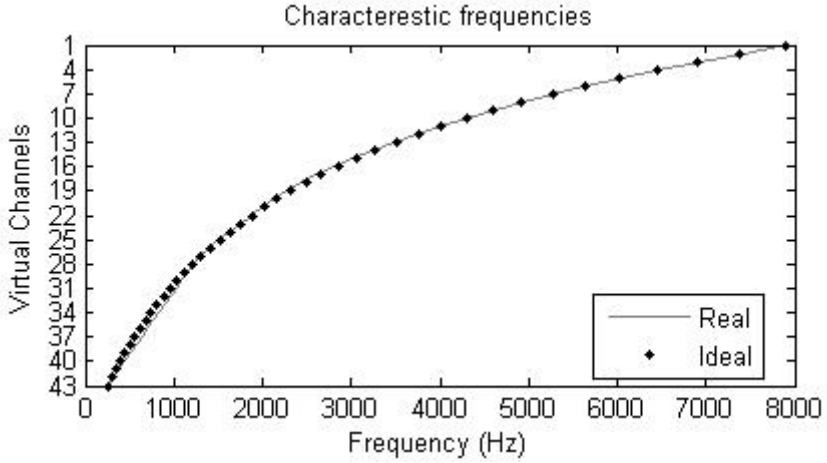


Figure (2.4): Ideal (diamond) and real (dotted) center frequencies of channels in the 43 channels mode. Channel 43 represents the most apical channel (lowest frequency).

Figure (2.4) presents both the ideal and real CFs in the 43 channels mode. There is a slight difference at the lower end between both of them due to FFT discretization error. After calculating the CFs of the VCs, an optimization search algorithm was employed to find the optimum bin to band distribution with minimum discretization. The frequency band of each

channel is represented by an integer number of FFT bins that starts with 1. In some patients one or more electrodes may be disabled due to clinical reasons such as pain or facial nerve stimulation. A technical consequence is a reduction in the number of available channels. Since almost the same BW is covered using 22 and 43 channels, the discretization may result in using fewer number of bins which is then optimized.

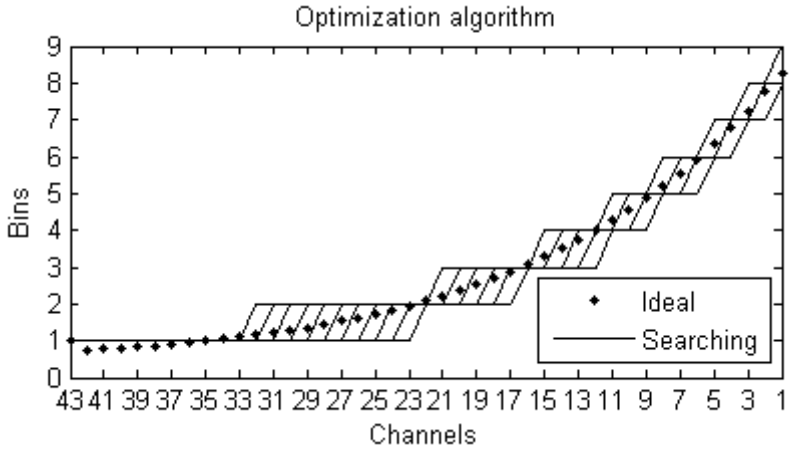


Figure (2.5): Possible distributions (solid), compared with the minimum discretization error. The optimization determins the ideal theoretical curve (diamonds).

Figure (2.5) illustrates the optimization procedure compared with the available number of FFT bins to available channels. The process involves avoiding unused bins due to a discretization. Figure (2.6) shows an output assuming the availability of all 43 channels.

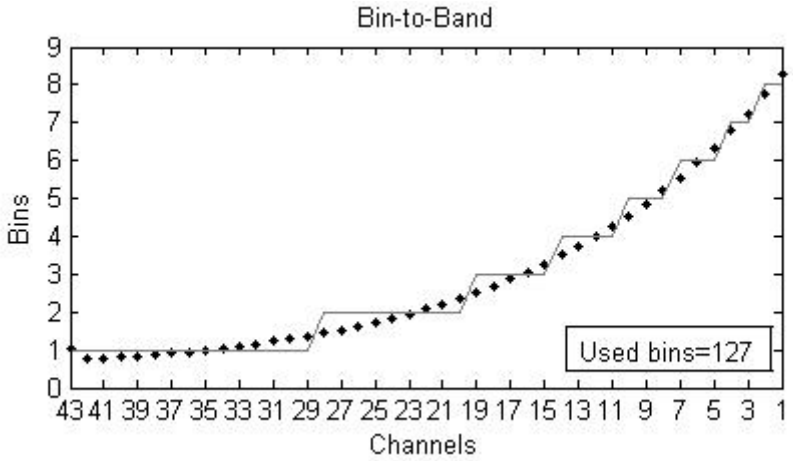


Figure (2.6): distribution of FFT bins on 43 channels.
(a) The ideal curve (diamond). (b) The optimum result (solid).

Figure (2.6) illustrates the best fit output of the minimization algorithm. On the lower right corner, the total number of used bins is indicated. In this case, with 43 channels the number of unused bins was minimized to zero.

Newly implanted CI recipients undergo many subjective tests and measurements. In many cases it may be necessary to disable one or many electrodes due to clinical reasons as described above. For a patient with the last 8 channels disabled, the frequency range is distributed onto the remaining 35 channels. The optimization algorithm minimized the number of unused bins due to discretization error to 1 bin as shown in Figure (2.7).

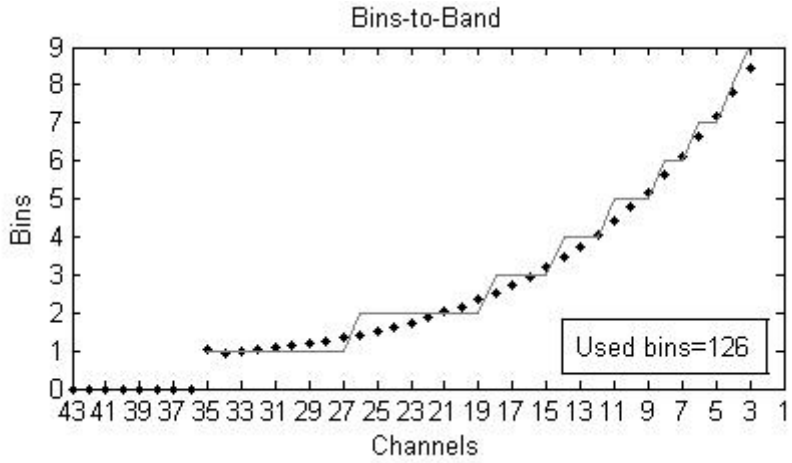


Figure (2.7): distribution of frequency bands onto different channels with Std mapping when 8 electrodes are disabled. (a) The ideal curve (diamond). (b) The optimum result (solid).

This algorithm has been compared to the actual bin to band distribution used in Sprint CI devices for 22 channels as shown in Figure (2.8).

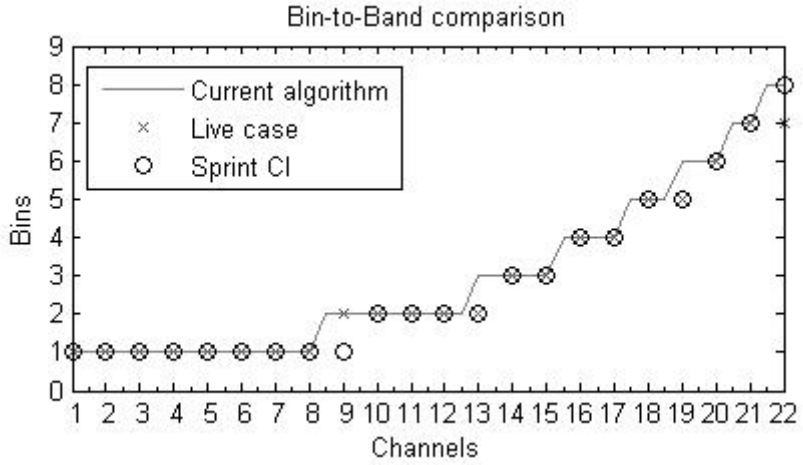


Figure (2.8): Comparison between bin-to-band distributions in the live case (star), current algorithm used (dotted line) and sprint devices (circle).

Figure (2.8) compares the bin-to-band distributions used in the live case (star), current implementation (dotted) and sprint devices (circle). The graph shows differences at channels 9, 13, 19 and 22. The current best fit search algorithm and Sprint algorithm both have two odd points that do not lie on the ideal curve. However, the best fit algorithm minimizes the number of unused bins. The calculations were processed with Matlab function called "*BarkGUI"* (see appendix B).

2.3 Greenwood Mapping

A Greenwood mapping, in which CFs of different channels were set to their tonotopical frequencies according to their positions in the cochlea, had been believed to produce more natural sounds than the Std ACE. It adopted a concept similar to that used in telephone transmission which acts as a BPF and cuts off frequencies outside this range [300 - 3400 Hz] and nevertheless maintains speech intelligibility (Möller 2005). Since the Nucleas implant does not reach the far end of the BM due to physical constraints, the lowest CF was set to 500

Hz assuming an average cochlear length of 32 mm and an insertion depth of 22 mm. In this mapping, low frequencies below 500 Hz were filtered out. This eliminates frequency compression existing in the Std ACE mapping but also removes the fundamental frequency and may cause pitch reversals of sounds. Eliminating the frequency compression produced better results with speech (Baskent and Shannon 2003). This mapping function may be assumed to be beneficial also with music. It was implemented in the AMO and the resynthesized outputs were analyzed. Spectral analysis of music samples which were generated through the MIDI protocol showed that music requires high degrees of preserving the harmonic structure representation of the overtones. This finding together with the experimental results of the Greenwood mapping led to the development of the Smt mapping. Smt mapping mainly preserves the representation of harmonic structure of overtones. Investigations of Greenwood mapping were not pursuited anymore and diverted towards the Smt mapping development.

2.4 Semitone Mapping

CIs were originally aimed at restoring speech perception for patients with profound hearing loss (Gfeller et al. 2000b; Clark 2003). The standard ACE speech coding strategy used in the Nucleus CIs typically encodes signals between 188 to 7980 Hz onto maximally 22 intracochlear electrodes. Additionally, the frequency range upto 1 kHz is represented by only upto 8 electrodes in the Std ACE frequency to electrode mapping. This is insufficient to preserve the representation of the harmonic structure of musical tones, because the fundamental frequencies as well as overtones of adjacent musical tones will often be mapped onto the same electrode, especially for frequencies below 500Hz. It can be hypothesized therefore that this coding strategy will not be optimal for music melody representation.

The majority of musical instruments generate fundamental frequencies below 1 kHz (Pierce 1983). Preserving the harmonic structure of musical contours is expected to ameliorate melody representation (Idson and Massaro 1978). An important aspect of music is melody (Sadie and Grove 1995) which can be defined as a sequence of individual tones that are perceived as a single entity (Terhardt 1998). Preserving the harmonic structure of individual tones is assumed to be necessary for preserving the melody perception and enhance musical sound representation.

One way to improve melody representation would be to ensure that the fundamental frequencies of adjacent tones on the musical scale are assigned to separate electrodes. Such an approach involves mapping fundamental frequencies of musical tones to electrodes based on a semitone scale. The idea was initially investigated in a study by Kasturi and Loizou (Kasturi and Loizou 2007), using the 12 electrode Clarion CII (Advance Bionics) implant with a limited range of semitone frequencies. They concluded that semitone spacing improved melody recognition with CI recipients. Additionally, music could be enhanced by increasing the frequency representation in CIs. This may be possible using VCs (Busby and Plant 2005). However there is no increase in frequency resolution because SOE is expected to be effectively similar with VCs. VCs on an array of 22 electrodes would yield a total number of 43 channels, which would cover three and a half octaves with Smt mapping with one-semitone intervals between the characteristic frequencies of successive channels. Note that the middle VC between two adjacent electrodes is the only VC that can be created in current Nucleus CI devices, because there is only one current source.

In this section, Kasturi and Loizou's idea was extended and a Smt mapping algorithm incorporating VCs is presented. Two different Smt mapping ranges were considered. The first one, Smt-LF, is in the low and mid frequency range [130 – 1502 Hz] and the second, Smt-MF, is in the mid and high frequency (HF) range [440 – 5009 Hz]. The ranges of Smt-LF and Smt-MF mappings in relation to a piano keyboard are illustrated in

Figure (2.9). Note that at the lower end of the piano scale, the fundamental frequencies of successive tones differ by as little as 3 Hz at A0 (f_0 = 27 Hz) and ~8 Hz at C3 (f_0 = 130 Hz). This difference increases as the fundamental frequency of the tone is also increased.

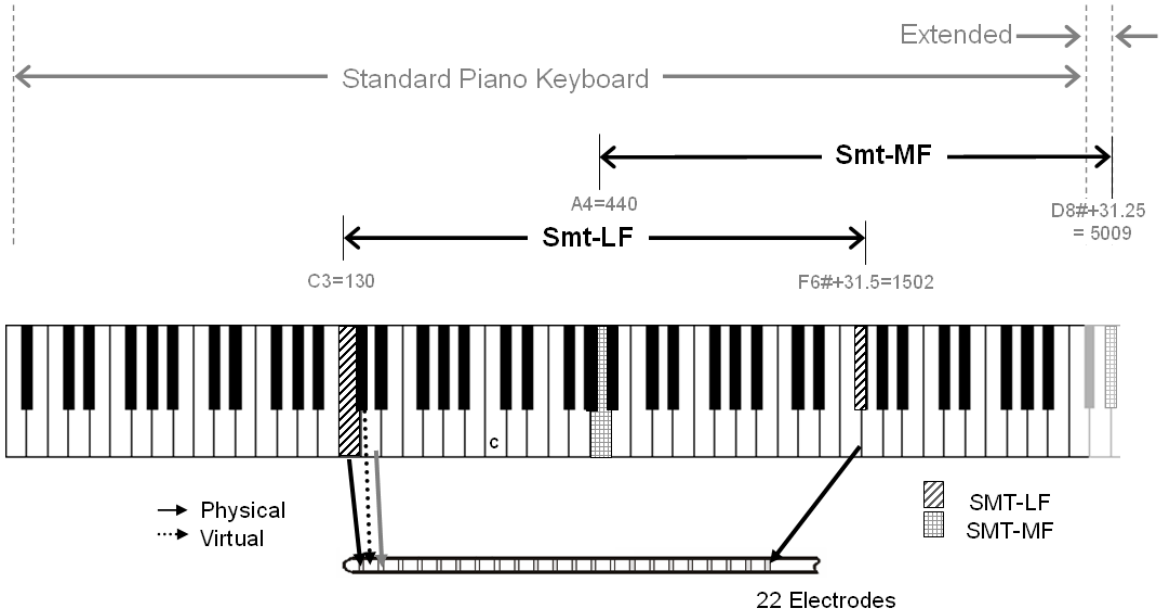


Figure (2.9): With Semitone mapping, consecutive tones on the musical scale (illustrated here using a piano keyboard) are assigned to adjacent channels corresponding to either physical electrodes (solid lines) or virtual channels (dotted lines). An array of 22 physical electrodes yields a total of 43 channels. Smt-LF and Smt-MF respectively map the tones from C3 to F6# and from A4 to D8#, to these channels. Note that the tones C8# to D8# (shaded keys) have been added for illustration purposes only and do not exist on the standard piano keyboard.

The frequency range of the Smt-LF mapping covers low and mid frequencies. These frequencies are common to most musical instruments (Pierce 1983). The Smt-LF mapping has a BPF that filters out the fundamental frequencies and the lower partials lying within the 1st and 2nd piano octaves (less than 130 Hz) as well partials in the 6th piano octave and above (greater than 1502 Hz). The range of the Smt-MF mapping covers part of the mid and high frequencies that are common between music and comprehensible speech bands used in telephone lines (Möller 2005). The Smt-MF mapping band-pass filters out frequencies lower than 440 Hz (A4) and higher than 5009 Hz. Thus fundamental frequencies and partials in most of the 4th piano octave (261-493 Hz) and below will not be represented. Smt-MF mapping also allocates frequency bands of audible sounds to electrodes with similar CF according to Greenwood function (Greenwood 1990; Greenwood 1991) assuming an average cochlear length of 33 mm and electrode insertion depth of 22 mm.

This section is organized as follows. Section 2.4.1 provides the theoretical basis of Smt mapping and emphasizes why Smt mapping was chosen. Section 2.4.2 presents a brief description of the processing technique. Section 2.4.3 describes in detail the two Smt-LF and

Smt-MF ranges investigated. Sections 2.4.4 - 2.4.6 show sub-bands to enhance the frequency resolution at low frequencies with Smt-LF and illustrates the frequency time matrix (FTM)s in both ranges. Mapping frequency bands in the FTMs to their corresponding channels involves using CTMs. Section 2.4.7 is a description of how the AMO is implemented to resynthesize channel activities into an acoustic sound that will be used in psychoacoustic tests with NH subjects in a later stage. The last section shows an analysis for the resynthesized sounds.

2.4.1 Theoretical Basis of Semitone Mapping

Smt mapping assigns fundamental frequencies of successive semitones on the musical scale to individual channels. Note that the harmonic overtones, which are integer multiples of the fundamental frequency, of each musical tone will also be mapped to the cf of separate channels with Smt mapping. Therefore, different musical tones will correspond to different sets of channels.

The relationship between the fundamental frequencies f_n and f_r of two musical tones k semitones apart is described by Equation (2.1) below.

$$f_n = f_r \cdot 2^{k/12} \tag{2.1}$$

$$\Rightarrow \quad \log_{10}\left(\frac{f_n}{f_r}\right) = 0.025\,k \tag{2.2}$$

where f_r is the fundamental frequency of the lower tone and f_n is the fundamental frequency of a tone that is n semitones apart. Equation (2.1) represents the ratio of characteristic frequencies of channels in Smt mapping. Substituting $k = 1$ gives frequency ratios for one-semitone steps.

The characteristic frequencies of Smt mapping for 43 channels each 1 semitone apart ($k = 1$) for the two Smt-LF and Smt-MF ranges (filled circles and squares respectively) are plotted in Figure (2.10). Note that higher channel numbers correspond to lower frequencies, to be consistent with the numbering used for Nucleus CIs. The characteristic frequencies of the Std mapping with 43 channels (i.e. including VCs) is also shown in Figure 2.10 (open circles).

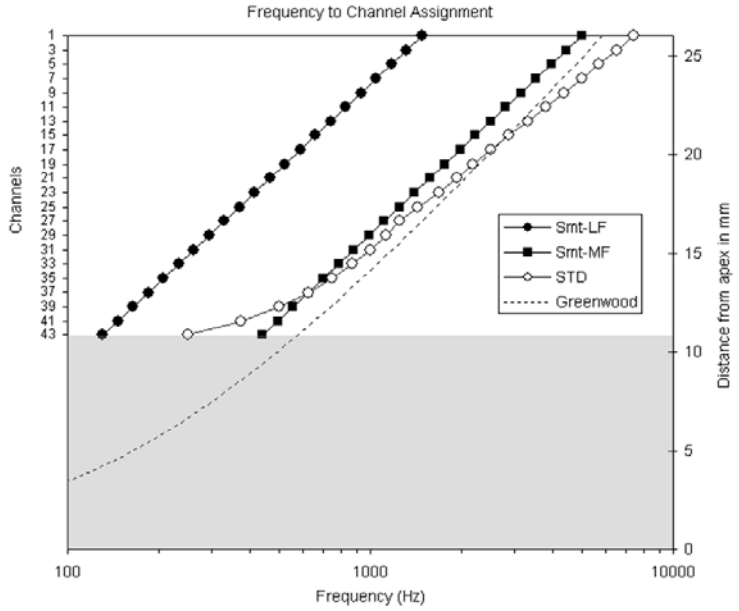


Figure (2.10): The frequency to channel assignments for the Smt-LF (filled circles), Smt-MF (filled squares) and Std (open circles) mappings are shown here together with the Greenwood

function (dashed line, secondary y-axis). The figure assumes a cochlea length (unrolled) of 33 mm, and illustrates an insertion depth of 23 mm for a Nucleus straight array with 22 equally spaced electrodes. The channel location within the cochlea can be derived from the two y-axes.

The two Smt mapping functions yield straight lines with a slope corresponding to the value of 0.025 as given in Equation (2.2). This value is required to map consecutive semitones to consecutive individual channels. Shallower slopes would result in more than one semitone being mapped to the same channel, distorting the original harmonic structure of the overtones. This would be the case for the Std mapping function, particularly with the first 8 channels in the lower frequency range. This distortion decreases at higher frequencies as the slope approaches a value corresponding to 0.025.

Since the inner ear resolves frequencies mainly based on a logarithmic function, harmonic overtones with the Smt mapping will be regularly spaced along the BM as described by the following equations.

Equation (2.3) below describes the characteristic frequencies at distance x mm from the cochlea's apex according to Greenwood's empirically derived function which was empirically verified against data that correspond to a range of x from 1 to 26 mm (Greenwood 1961)

$$f(x) = 165.4(10^{0.06.x} - 1) \tag{2.3}$$

$$\Rightarrow \quad x(f) = \frac{1}{0.06}\log\left(\frac{f}{165.4} + 1\right) \tag{2.4}$$

The distance (in mm) between two locations with different characteristic frequencies f_1 and f_2 is given by Equation (2.5)

$$\Delta x = x_2 - x_1 = \frac{1}{0.06}\log\left(\frac{f_2 + 165.4}{f_1 + 165.4}\right) \tag{2.5}$$

Substituting f_2 and f_1 by f_n and f_r respectively from Equation (2.3) yields:

$$\Delta x = \frac{1}{0.06}\log\left(\frac{2^{k/12} + 165.4/f_r}{1 + 165.4/f_r}\right) \tag{2.6}$$

Equation (2.6) shows that the spacing along the BM between two successive semitones (substitute $k = 1$ and f_r with the fundamental frequency of the lower tone) will vary depending on the frequency range, and is smaller at low frequencies, asymptotically approaching 0.4 mm with higher frequencies. For C3 (f_0 = 130.8 Hz), the spacing would be about 0.19 mm, while at C8 (f_0 = 4186 Hz) about 0.4 mm. The electrode spacing between successive electrodes in the Nucleus 24 implant Contour Advance electrode array varies gradually between 0.1 mm at apical locations to 0.48 mm at basal locations (Cohen et al. 2002). As the width of each electrode contact is 0.3 mm the center-to-center distance therefore varies from 0.4 to 0.8 mm (Xu et al. 2001). For VCs, assuming that the centre of stimulation is halfway between the two physical electrodes, the channel spacing would vary between 0.2 mm to 0.4 mm. This would correspond roughly to the tonotopical spacing for the tones involved in the two Smt-mappings.

2.4.2 Processing and Implementation

The block diagram in Figure (2.1) shows the standard ACE processing algorithm. An acoustic signal undergoes STFT, from which the PSD is calculated. The frequency range of the PSD is divided into different bands. The *n* bands with the highest energies (maximas) are then selected for presentation, where *n* is a parameter that can be defined for each CI recipient's map. The resulting FTM is then processed as follows: The energy within each selected band is used to determine the corresponding stimulation level according to a LGF. Using a mapping function, the respective bands are then assigned to channels, which can be physical electrodes or VCs, to produce the CTM.

2.4.3 Semitone Mapping Frequency Ranges Setup

The fundamental frequencies of the musical tones from the piano keyboard vary between 27.5 Hz (A0) and 4186 Hz (C8) (Pierce 1983). Two ranges were investigated for the Smt mapping:

Smt-LF [125 Hz - 1568 Hz] (C3-F6#)

The minimum required frequency resolution for the Smt-LF mapping is ~8 Hz at C3 (f_0 = 130 Hz). Analyzing a signal that has a sampling rate of 16 kHz with 2048 FFT points provides a 7.8 Hz resolution between successive frequency bins. The lowest acoustic frequency of 130 Hz for Smt-LF will be mapped to the most apical electrode location, which would correspond to a characteristic tonotopical frequency of approximately 571 Hz estimated according to Greenwood's equation (Greenwood 1990), assuming an average cochlear length of 33 mm and an electrode array insertion depth of 22 mm. This will cause sounds to be perceived higher in pitch. However, as the frequency shift is expected to be the same for all partials with Smt mapping, this would be equivalent to a transposition.

Smt-MF [440 Hz - 5918 Hz] (A4-D8#)

A reduced frequency bandwidth is enough to maintain speech comprehension as in telephone transmission, where the bandwidth used is [300 – 3400 Hz] (Möller 2005). The Smt-MF mapping covers part of the bandwidth that is common between speech and music [440 – 5009 Hz]. Note that transposing the Smt-MF range three semitones higher to cover a range from 523 Hz (C5) to 5919 Hz (F8#) would minimize the difference between characteristic and tonotopical frequencies of electrodes according to Greenwood (Greenwood 1990) (see Figure (2.10)) assuming an average cochlear length of 33 mm and an insertion depth of 22 mm. However, it is impossible to precisely match the tonotopical CF for any given individual through Greenwood's function as the latter is empirical in nature and is also supposed to only represent the average NH listener. Also cochlear length and electrode insertion depth vary individually. Thus, some discrepancy is always to be expected.

2.4.4 Frequency Time Matrix

Frequency components at different time frames are analyzed using STFT and are organized into a FTM. A typical CI processor like the Nucleus Freedom uses a sampling rate f_s of 16 kHz to produce the FTM with a 128 points FFT (Swanson et al. 2007; Swanson 2008), giving a frequency resolution Δf of 125 Hz (Fourakis et al. 2004; Carroll and Zeng 2007; Fourakis et al. 2007). However, Smt-LF mapping needs a higher resolution (Δf of ~8 Hz) at low frequencies (~130Hz). Increasing the number of points $N = fs / \Delta f$ increases the frequency resolution but at the same time will produce smearing in the time signal due to the reduction

in the temporal resolution. In order to increase the frequency resolution at low frequencies and retain some of the time resolution at higher frequencies, frequency sub band decomposition (Crochiere et al. 1976; Vary et al. 1998; Wyrsch 2000) is used to generate the FTM.

2.4.5 Frequency Resolution and Subband Decomposition

First, the input signal is sampled at 16 kHz. Then the sampled signal is processed in 2 frequency subbands (see Figure (2.11)) to yield two different frequency resolutions.

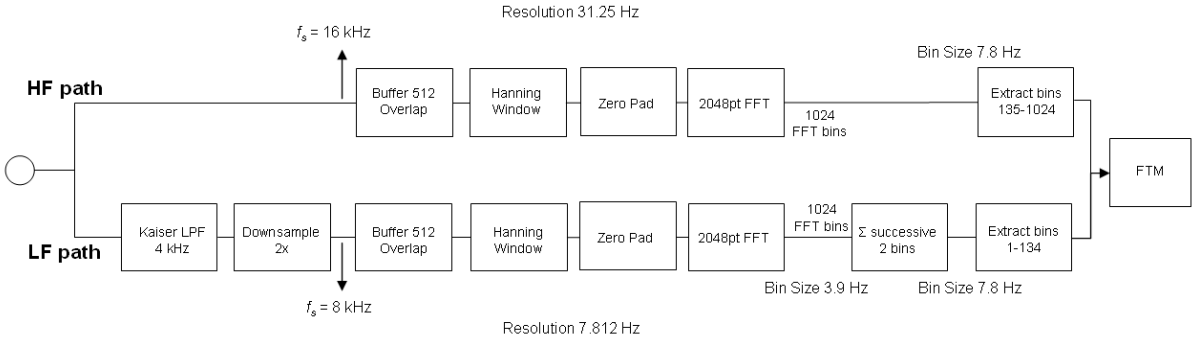


Figure (2.11): Implementation of frequency subband decomposition for Smt-LF, producing a frequency resolution of 7.8 Hz at frequencies below 1054 Hz (low frequency path), and a frequency resolution of 31.25 Hz at higher frequencies (HF path). The number of overlapping points between frames is calculated as follows: Overlap = 512 – *f*s / (stimulation rate). Note that the analysis frame rate is the same as the stimulation rate.

Figure (2.11) shows the input signal flows into two parallel pathways; one for the LF and the other for the HF. The LF pathway uses 512 (*N*) samples which are split into overlapping time frames and analyzed. The amount of overlap depends on the stimulation rate such that at the end of each stimulation period, as much new data (sampled at 16 kHz) as possible is added to the data buffer. For instance, with a stimulation rate of 500 Hz, 32 new samples are added every stimulation period to the data buffer of length 512 samples, resulting in an overlap of 480 samples. The signal is first filtered using a Kaiser LPF with a cutoff at 4 kHz, and then decimated by a factor of two (d = 2) which increases the frequency resolution by the same factor while keeping the buffer length 512 points. Each time frame with 512 points is processed with a Hanning filter, and then zero padded before undergoing a 2048 (*m*) point FFT. Notice that after zero padding each bin represents a frequency band of 3.9 Hz (fs/m=8k/2048). Every two successive bins are then summed to preserve the power and decrease the overall minimum detectable frequency difference ($\Delta f = fs/(d \cdot m/2)$) in the LF branch within successive bands to 7.8 Hz.

The HF pathway uses the same number of points (*N* = 512) was used in the LF pathway, producing a frequency resolution of 31.25. The signal is split into overlapping time frames in the same manner as in the LF pathway. Each time frame is processed with a Hanning filter of the same number of points; zero padded and undergoes a 2048 point FFT.

The output bins from both pathways are combined to form the FTM which has a bin resolution of 7.8. The boundary between the LF and the HF pathways was set to 1054 Hz (between C6 and C6#), where the difference in frequency between successive semitones starts to exceed by a factor of 2 the HF resolution and this ensures that any successive semitones will at least lie on successive electrodes. The lower 134 bins are from the LF pathway, while the higher bins are from the HF pathway.

An example of a FTM produced using frequency sub band decomposition for a signal with 4 sinusoidal components with 900, 936, 1200 and 1295 Hz is shown in Figure (2.12). The difference in frequency resolutions can be clearly seen in the narrower bands at lower frequencies and wider bands at higher frequencies.

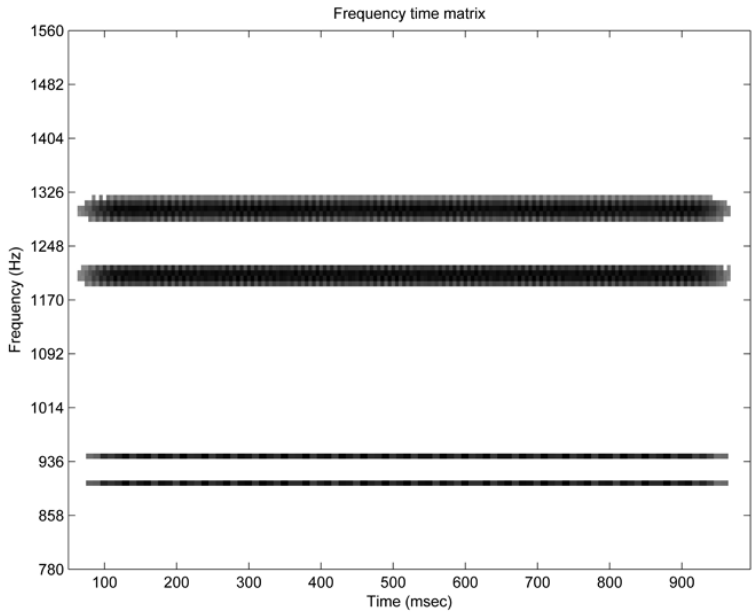


Figure (2.12): FTM of a complex tone consisting of 4 components at 900, 936, 1205 and 1285 Hz processed with frequency sub band decomposition (Smt-LF). The different frequency resolutions of 7.8 Hz and 31.25 Hz above and below the threshold value of 1054 Hz respectively is illustrated by the different track widths of the resolved components.

The above frequency sub band decomposition only applies to the Smt-LF mapping. For Smt-MF mapping, the minimum frequency resolution required is ~26.6 Hz at A4 (f_0 = 440 Hz). Using N = 512 provides a minimum resolution of 31.25 Hz which is slightly larger than the required resolution for the lowest semitone frequencies. Note that in the present implementation of the Smt-MF, the first two tones (A4 & A4#) will not be adequately resolved and therefore fall within a single FFT bin which will in turn be mapped to two adjacent channels because the difference between them is less than the LF resolution (7.8 Hz). To preserve the starting frequency and approach characteristic frequencies (CF)s of electrodes to Greenwood frequencies without having an empty channel, it is suggested to activate the first two electrodes to frequencies of (A4 & A4#). The starting frequency could have been made slightly lower, but a drawback is that this would increase the difference between CFs and Greenwood frequencies. The remaining tones are adequately resolved. Subband decomposition is not used with Smt-MF mapping. The processing block diagram for Smt-MF is similar to the one described in the HF pathway shown in Figure (2.11), with N = 512 without zero padding and without the frequency scaling block. Note that the frequency resolution of Smt-MF is 31.25 Hz, compared to 7.8 Hz (for frequencies below 1054 Hz) and 31.25 Hz (for frequencies above 1054 Hz) in Smt-LF.

2.4.6 Channel Time Matrix

Depending on the frequency range of interest (e.g. Smt-LF [130 – 1502 Hz] and Smt-MF [440 – 5009 Hz]), different bins in the FTMs are combined into frequency channels to produce CTMs. A mapping matrix (M) is introduced to define which FFT bins should be mapped to which corresponding channels. The mapping matrix attempts to map the cf of the channels and FFT bins as close as possible to the fundamental frequencies of each corresponding semitone.

$$CTM = M \cdot FTM \tag{2.7}$$

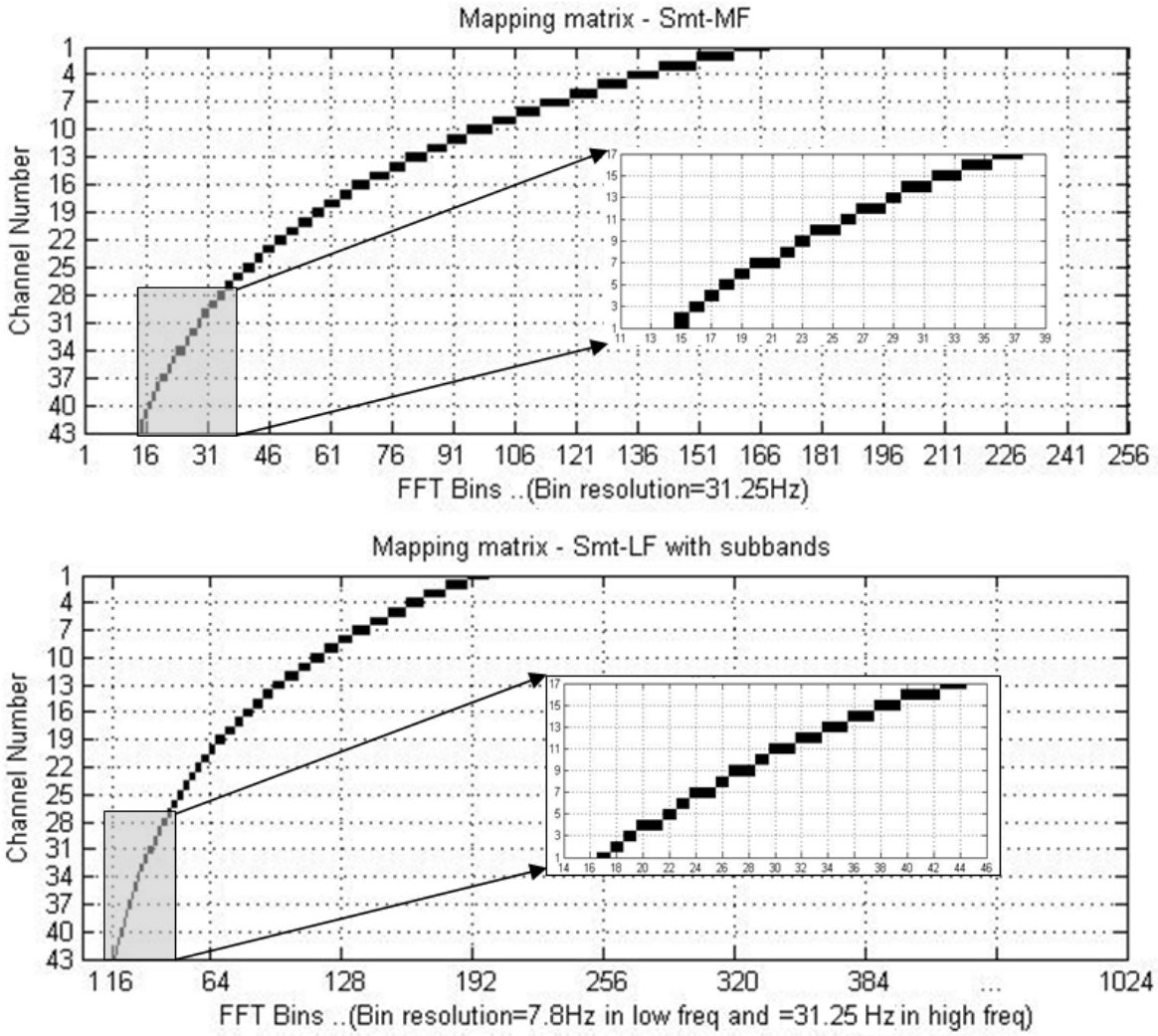


Figure (2.13): Mapping matrices for Smt-LF (a-upper) and Smt-MF (b-lower) with 43 channels.

Figures (2.13a and 2.13b) illustrate the mapping matrices for both the Smt-MF and the Smt-LF mapping respectively for 43 channels. The Smt-LF mapping covers frequency band [130 – 1502 Hz] which corresponds to bin numbers [17–200], with a frequency resolution of 7.8 Hz. The Smt-MF mapping covers frequency band [440 – 5009 Hz] which corresponds to bin numbers [15–169] where the frequency resolution is 31.25 Hz. Smt-MF mapping does not incorporate sub bands and accordingly bin 15 (corresponding to 440 Hz) is mapped to channel 1 and 2.

2.4.7 Signal Resynthesis

To resynthesize a CTM into an audible sound, each column in the CTM is converted into 256 samples in the time domain. Each column corresponds to 16 ms if there was no overlap performed in the FFT analysis between successive time frames, otherwise Equation (2.8) should be used. Either the rows in the CTM should be up-sampled or the audio playback sampling rate should be reduced by the *UpSampling* factor. Equation (2.8) uses the rate of stimulating the electrodes from the patient's map and is fed to the AMO, which is a CI model used to acoustically simulate the sound perceived by a CI recipient. It uses the Nucleus Matlab toolbox (NMT) from Cochlear Corporation to describe the implant, and noise band vocoders to resynthesize the sound.

$$UpSampling = \frac{(WindowSize - Overlap) * StimulationRate}{SamplingRate} \tag{2.8}$$

Where,
Stimulation rate: is the CI simulation rate
Window Size: Number of points used to analyze the HF Branch
Overlap: Number of points used to overlap successive frames.

2.4.8 Nucleus Matlab Toolbox

The Smt mapping follows the ACE strategy in selecting the highest n channels. It was implemented in Matlab and incorporated into the NMT framework (Swanson 2008). The AMO was based on noise band vocoders (Laneau et al. 2006a). The activity in each channel is simulated as a white Gaussian noise convolved with an exponentially decaying filter, whose center frequency is the CF of the channel. Channel interactions arising from the spread of the electric field from its center at the stimulation site can be set by a SOE (*"width of stimulation"*) parameter. The resultant stimulation of the auditory nerve, causing also the perception of adjacent pitches, are simulated with the AMO. In the Smt-LF mapping, the AMO simulated the frequency transposition.

2.4.9 Analysis

Following the definition of (Terhardt 1998) for melody a better representation of individual musical tones is expected to ameliorate melody recognition, or in other words, melody is poorly resolved if individual musical tones are poorly represented. Musical tones are characterized by their harmonic structure. To compare the harmonic structure representation for the three different mappings (Std, Smt-MF and Smt-LF), a sound sequence consisting of 36 consecutive synthetic musical tones was constructed. Each tone consisted of 5 partials with successive 20% decrease in amplitude and lasting for 150 ms. The fundamental frequency of each tone increased from 130 Hz (C3) to 987 Hz (B5) with 1-semitone intervals.

a)

b)

c)

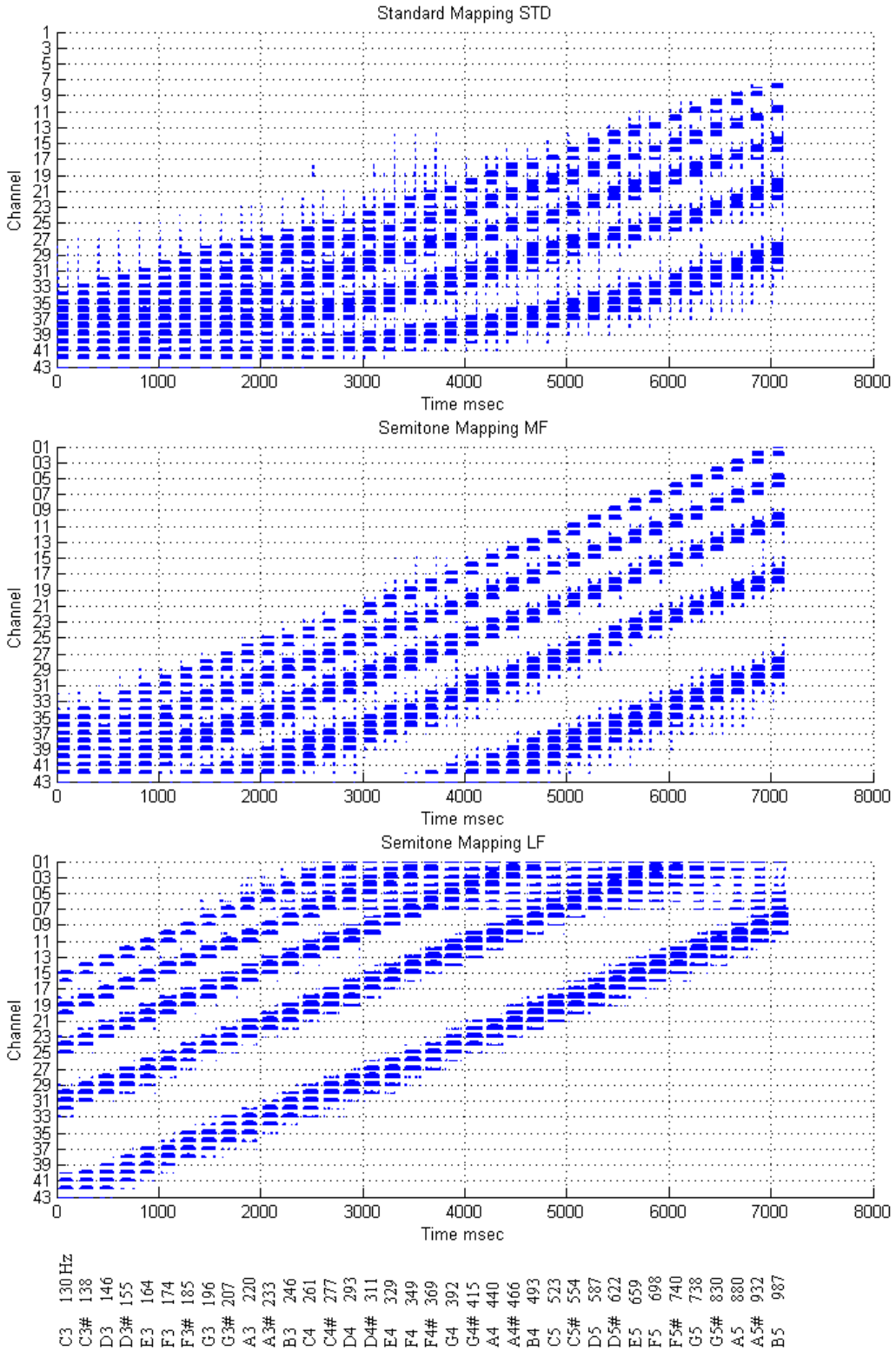


Figure (2.14): Analysis using Std map (a), Smt-MF mapping (b) and using Smt-LF (c) mapping.

Figure (2.14) shows the harmonic structure being preserved with both Smt-MF and Smt-LF mappings, where the spacing of the partials remains uniform across tones. With the Std mapping (see Figure (2.14-a)) at low frequencies, partials are not resolved. With Smt-MF, frequency components below 440 Hz are filtered out (as indicated by arrows in (see Figure (2.14-b)), while with Smt-LF (see Figure (2.14-c)), the HF partials greater than 1.6 kHz are filtered out.

Chapter 3 – Tools

Chapter 3 – Tools

Evaluating music perception with CIs requires utilities. Most of the utilities were programmed in Matlab using specific libraries. AMO is a CI analysis and resynthesis application which processes sounds generally like a CI processor with Std, Greenwood and Smt mappings. Channel activities can be resynthesized using either noise bands or sinusoidal vocoders with a preset SOE using 22 and 43 channels. Checker is a utility to process sounds according to patient maps loaded from a clinical database and to produce offline stimulus sequence files. The Checker scrutinizes the stimulus files for various errors before streaming to a CI recipient. Another utility called "Spectral-Contrast" was used to display differences between two sounds by converting them into a loudness and critical band domain. A matrix was used to identify regular diagonal activities in matrices or images. It was applied on electrodograms from Std and Smt mappings to compare harmonic structure preservation of overtones using both mappings. Finally, this chapter proposes a method to estimate harmonicity probability in sounds with and without a reference. Sounds are processed with a CI model using different mappings and the harmonicity indices of musical sounds are measured before and after processing.

3.1 Acoustic Model (AMO)

The AMO is a CI analysis and resynthesis vocoder application. The AMO consists mainly of two sections; processing and resynthesizing sections. The AMO employed modules from the NMT from Cochlear Corporation (Swanson 2008).

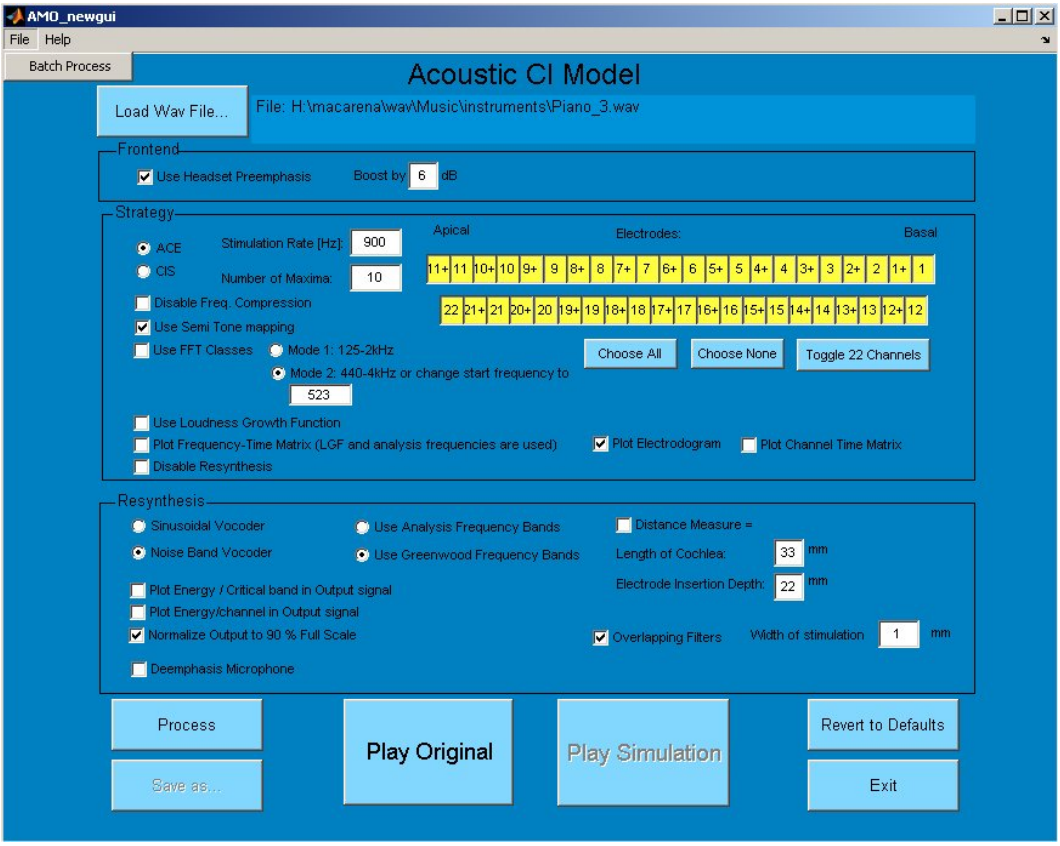


Figure (3.1): Graphical user interface of the acoustic model (for more detail see Appendix C).

The first section processes signals with the same algorithms as used by Nucleus cochlear speech processors. The processing Std mapping algorithm was packed in the NMT produced by Cochlear PTY. Extension of the Std using VCs as well as Smt mapping were written from scratch and then packed in the NMT. Figure (2.1) shows a block diagram of the processing section. The FFT calculation function was enhanced; first to avoid out-of-memory limitations using lighter and more efficient algorithm with sparse matrices and second to add the 43 channels mode, Greenwood and Smt mapping algorithms. Once a wave file is loaded, it is buffered into overlapping time frames. The amount of overlap follows Equation (3.1).

$$Overlap = \frac{fs}{AnalysisRate} \tag{3.1}$$

where,
fs: Sampling frequency
AnalysisRate: Stimulus rate in pulses per second

Each window is transferred into the frequency domain and the output of the FFTs is used to calculate the PSD in each window. A predefined number of maxima 'n' in patient's map is selected that represents the bands with the highest energies in the spectrum. The selected bands from all frames are used to construct a FTM. Frequency bins in the FTM are grouped into bands and each band represents a channel. Frequency intensities in the FTM are mapped into a current level using a LGF. The outcome is a FTM with adjusted loudness equivalent levels. This produces the CTM which is used in generating stimuli data. The block diagram is illustrated in Figure (2.1).

The second part of the AMO is a resynthesizer. As discussed previously in section (2.4.7), the resynthesizer converts CTMs into audible acoustic sounds. Figure (3.2) shows a block diagram of the resynthesizer. It uses either a sinusoidal or noise band vocoder as the carrier generator. The temporal activities allocated to each channel are convolved with the carrier that has a CF according to the tonotopical representation of its channel. In case of using noise vocoders, a white Gaussian noise is convolved with an exponential rising decay filter with a predefined SOE (1 mm by default). The vocoder outputs for each channel are then summed and normalized to a value between]-1, 1[to avoid clipping.

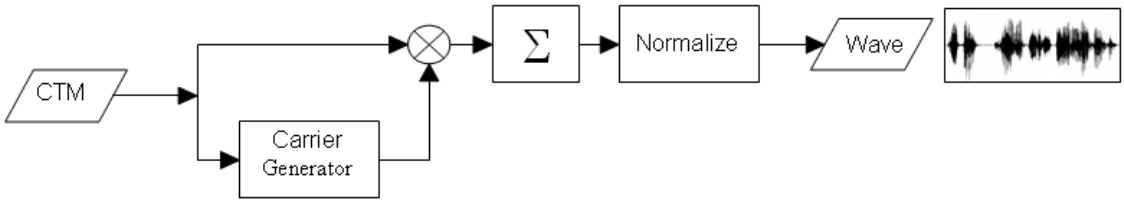


Figure (3.2): Block diagram of the resynthesizing section in the acoustic model.

3.2 Checker

Clinical testing with CI recipients is subjected to strong regulations and rules. Before examining a new strategy by direct streaming to a speech processor, stimuli patterns must be checked for all possible types of errors. Small errors in stimulation may lead to overstimulation of ganglion cells that may be uncomfortable for a recipient. To avoid erroneous stimulations, the Checker application was developed.

The Checker is a Matlab application based on the AMO. It is used to process sound files according to patients' maps that are loaded from a clinical database. The Checker generates sequences of pulses using the NMT from Cochlear PTY (Swanson 2008). These pulses are stored in a stimuli file which is verified before being streamed to a speech processor. After the checking process finishes, it produces a detailed report for each stimuli file. This application was enriched with many features such as batch processing files for different mappings. Processing and generating files is time consuming and usually done over night. The output wave file is calibrated according to the required output level. The calibration was done by comparing electrodograms of the speech processor when a sound was played at a predefined SPL via a loudspeaker to direct electrical output generated and the electrodograms at different arbitrary loudness values. For more details see Appendix C.

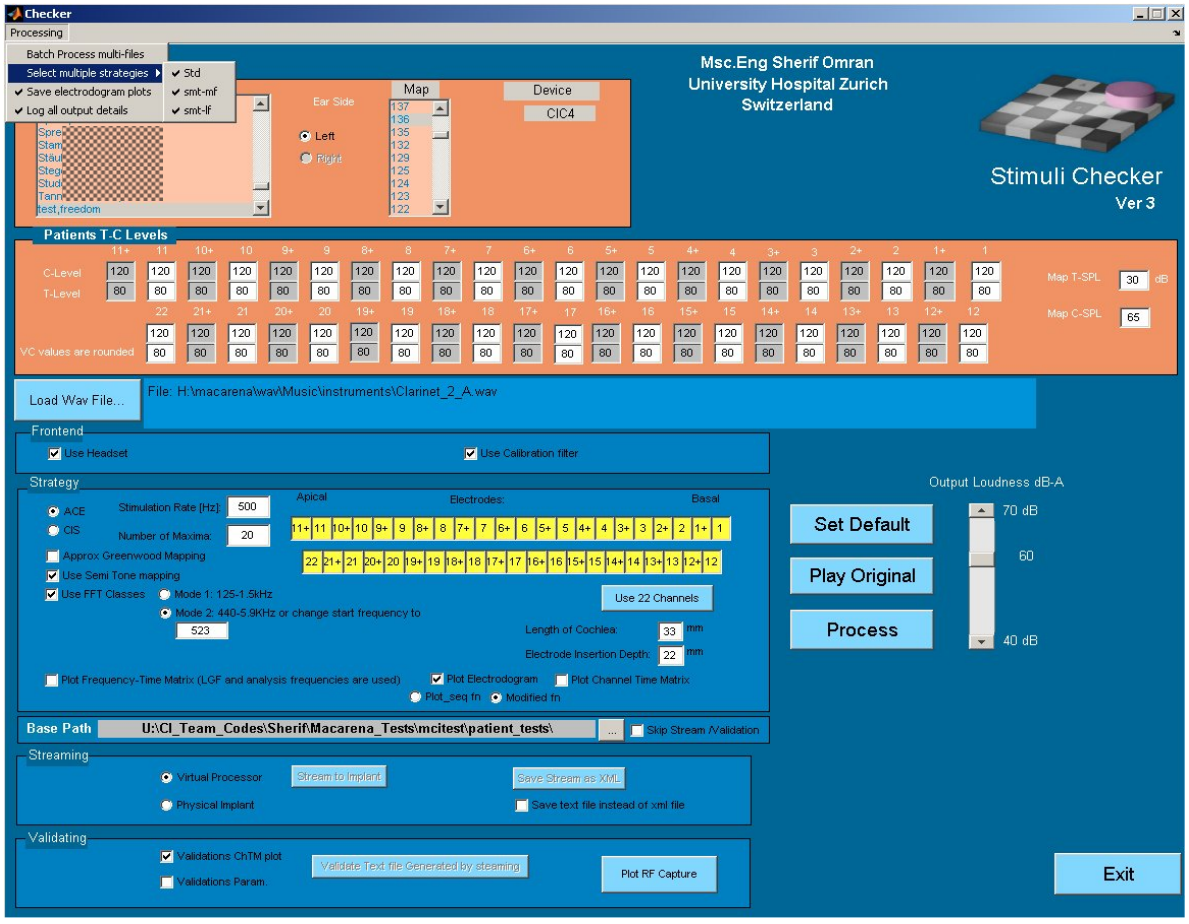


Figure (3.3): Graphical interface of the Checker program (see Appendix C for more details).

The Checker loads patients' data from a clinical database. The patient name, side of implantation (left or right) and the electrode map have to be selected. Different mappings (Std, Smt-MF and Smt-LF) are implemented in the Checker. Signals are processed and stimuli files are generated. The generated stimuli file is examined according to a setup rules as described in the next section. If the validation is error-free, an XML file is generated with a unique name and a special folder structure.

MACarena is a psychoacoustic testing environment (Lai and Dillier 2002). It interfaces with the Nucleus implant communicator (NIC), which is a dynamic link library (dll) from Cochlear Corporation to communicate with the implant. Xml files are streamed to a speech processor of a patient who sits in front of a touch screen and has to characterize the perceived sound.

3.2.1 Checker Procedures

When the Checker loads, it connects to a clinical database and loads patient names if online connection exists otherwise it uses locally stored maps. Some functions were extended in the NMT to include additional database calls (see appendix C). The user is asked to select a patient name, the implanted side and the corresponding map. This map holds different electrode setting parameters such as T-Levels and C-Levels, enabled and disabled electrodes and rate of stimulation of each electrode for the selected patient. The T and C-Level values that were measured and stored in the database are for 22 electrodes. Values of T-levels and C-levels were interpolated to determine intermediate values for VCs assuming that T-Levels and C-Levels for VCs are the averages of neighboring electrodes' T-Levels and C-Levels respectively. VCs are excited by stimulating two neighbor physical electrodes with the same current level. A disabled physical electrode causes the neighbor virtual channel to be disabled as well.

If all settings are adjusted, one can load and process a sound file. The outcome described as CTM is plotted and the stimuli sequence is stored into a stimuli file. This stimuli file is checked for errors. If it is validated, it will be converted into an XML file called "*streaming file"*. This xml file can be later uploaded to the speech processor.

3.2.2 Reconstructing CTMs

In an error tracing subroutine in the Checker, the CTM is reconstructed from the stimuli file and is plotted side by side against the CTM used in processing. The two versions are also compared pulse by pulse. The relative time index of each pulse is stored in the stimuli file. To reconstruct the CTM correctly, one needs to calculate the absolute time index for each pulse. Additionally, each point in the CTM depends on the number of points (n) used in the FFT of each time window with an overlap (m) between successive windows. The total time duration in ms of the reconstructed CTM could be calculated by this Equation (3.2).

$$Time\_duration = \text{int}\left(\frac{m}{fs} * n * 100\right) * 10 \text{ ms} \qquad (3.2)$$

Where,

n : number of points used in FFT of each time window
m : number of points used in overlap between FFT windows
fs : sampling rate of the processed wave file (default = 16 kHz)

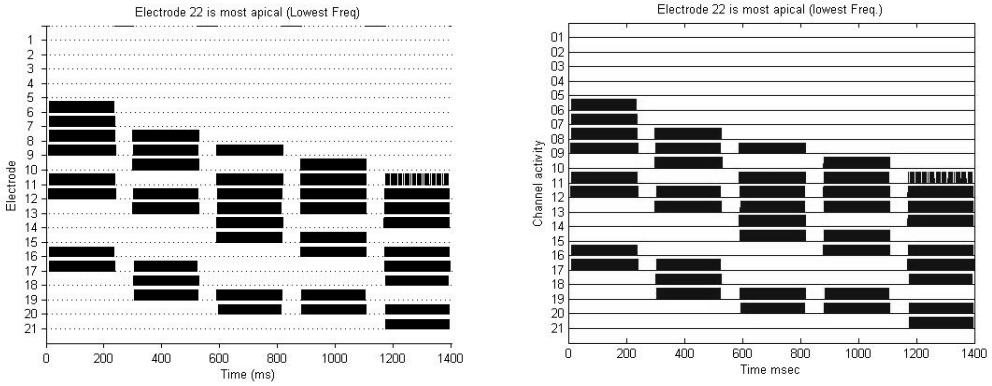


Figure (3.4): An example for a generated CTM during signal processing (left) and the reconstructed one (right) for a test signal.

Figure (3.4) illustrates a comparison between the CTM generated during processing and the reconstructed one.

3.3 Spectral-Contrast model

Sound can be represented in three physical domains; frequency, intensity and temporal domain. Zwicker (1960) showed that the ear resolves sounds according to a filter bank and proposed 24 CBs from 50 to 15500 Hz. The frequency domain in the proposed Spectral-Contrast measure uses CBs upto 8 kHz. Loudness is a perceptual measure of sound intensities measured in Sones. Loudness of two sounds that are simultaneously played appears to be louder as if they were played separately (Pierce 1983; Terhardt 1998). This Spectral-Contrast model compares the contrast in loudness for all CBs in each time frame of two sounds. The third dimension is the time domain; acoustic signals sampled at 16 kHz are split into 256 points non-overlapping frames. This algorithm was used to investigate the difference between unprocessed tones and tones processed with the Std mapping.

3.3.1 Description

An audible sound with maximum frequency 8 kHz is decomposed into temporal non-overlapping window frames and the FFT is calculated for each frame. The frequency domain is divided into CBs from 1 to 21 which corresponds to frequencies from 50 to 7700 Hz (Zwicker 1961).

For an input sound *h(t)*, its FFT *H(f)* is calculated. The FFT uses N sample points (N=256). The FFT divides the frequency spectrum to k bins with a bin size (B) in Hz according to the sampling rate *fs* (16 kHz) as in Equation (3.3)

$$B = \frac{fs}{N} \tag{3.3}$$

The starting frequency of each bin is described by $f(k)$ where *k* varies from *0-N /2*.

$$f(k) = \frac{fs}{2}.\frac{k}{k+1} \tag{3.4}$$

The Bark scale assigns an integer value to each CB. Values extend upto 21 barks for frequencies upto 7.7 kHz. The bark scale is related to the frequency by Equation (3.5).

$$Bark(f) = 7\ln\left(\frac{f}{650} + \sqrt{1+\left(\frac{f}{650}\right)^2}\right) \tag{3.5}$$

If several partials lie within the same CB, the actual intensities of all partials are added to get the total intensity in this CB. If two sounds of equal loudness are separated by more than a CB when they are played together, they produce upto twice the loudness as when they are played separately (Pierce 1983).

The intensities *S(k)* of all overtones lying in the same CB are summed as described by Equations (3.6 and 3.7).

$$S(k) = \sum_{f(k)}^{f(k+1)} H(f) \tag{3.6}$$

$$\Rightarrow S_{dB}(k) = 10.\log\left(\frac{S(k)}{10^{-12}}\right) \tag{3.7}$$

The intensity summation *S(k)* is converted into dB with respect to the minimum audible value (10exp-12) as described in Equation (3.7). This has in turn to be converted from dB-SPL at the center frequency of the CB into loudness level in Phons. The equal-loudness curve defined by function C_L in Equation (3.8) is a function of given dB-SPL and frequency. The equal loudness curves are plotted in Figure (1.5), and used to convert the loudness level of the summed intensities of partials in a CB frequency ranges from dB to Phons. To implement the equal loudness curves tables from the ISO-226 Standard were used.

$$S_{phons}(k) = C_L(S_{dB}(k), \frac{f(k)+B}{2}) \tag{3.8}$$

Phons are converted into Sones using a conversion curve shown in Figure (1.6) and described by Equation (3.9). Figure (1.5) illustrates a direct relationship between Phons and Sones, where 40 Phones = 1 Sones at 1 kHz by definition. A 10 Phons increase doubles the loudness value (Pierce 1983). There is nonlinearity below 40 Phons as shown in Figure (1.6). The output $S_{sones}(k)$ is calculated in each CB for each FFT frame.

$$S_{sones}(k) = P_{sones}(S_{phons}(k)) \tag{3.9}$$

The function $S_{sones}(k)$ is calculated for two sounds to produce $S_{sones}(k)$ and $\widehat{S}_{sones}(k)$. Both are subtracted as in Equation (3.10).

$$D(k,t) = S_{sones}(k,t) - \hat{S}_{sones}(k,t) \tag{3.10}$$

The function $D(k,t)$ is a 3-D matrix that holds differences between the two signals in each CB for each temporal frame. A 3D clarinet tone taken from the RWC database (Goto et al. 2001) was processed using the AMO with 22 channels Std mapping and compared to its unprocessed version.

The processed and the unprocessed signals are plotted in Figure (3.5). This figure shows that the intensities of the partials in the unprocessed tone vary smoothly and the harmonic structure is clearly seen while in the processed tone the harmonic structure is not preserved with Std mapping.

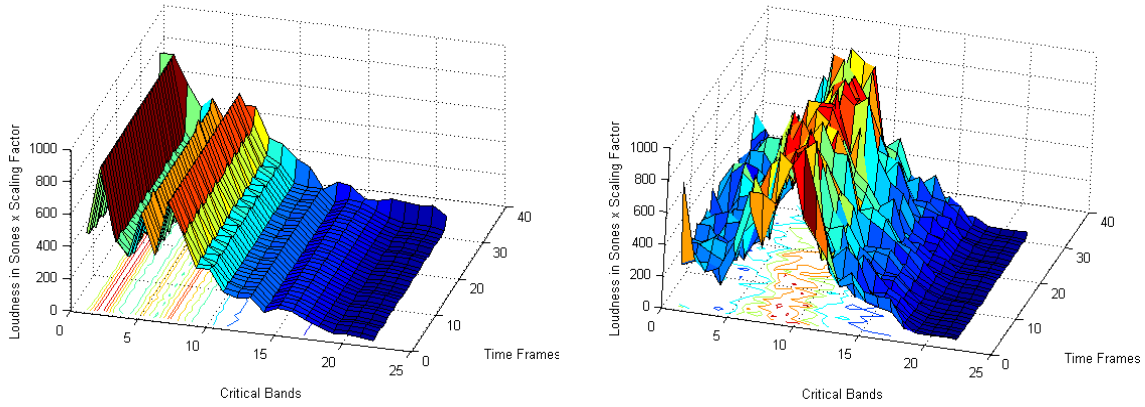


Figure (3.5): Comparison between unprocessed (left) and processed (right) clarinet tone. The tone was processed with AMO using Std mapping with 22 channels. The vertical axis is the loudness in sones multiplied by an arbitrary scaling factor for illustration.

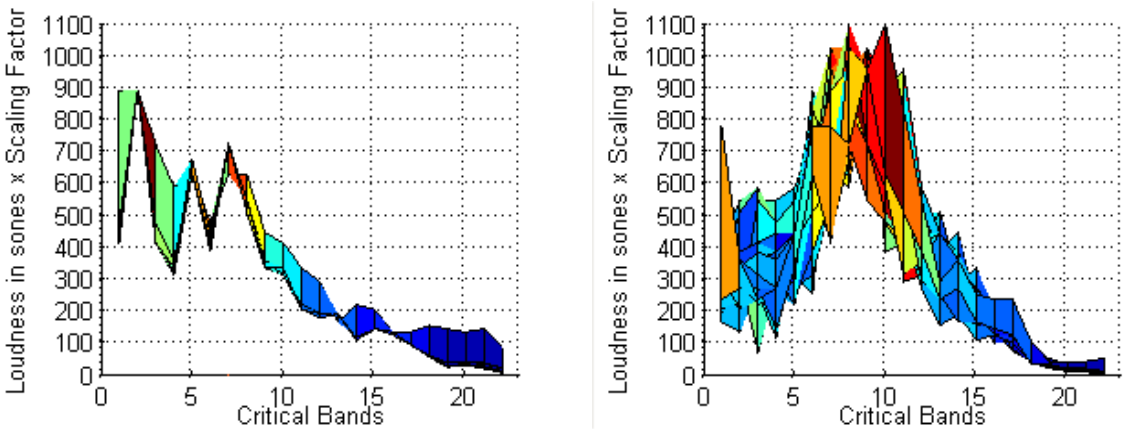


Figure (3.6): Side view comparison between unprocessed (left) and processed (right) clarinet tone. The tone was processed with the AMO using Std mapping with 22 channels. The vertical axis is the loudness in sones multiplied by an arbitrary scaling factor for illustration.

Figure (3.6) illustrates a comparison between the projection on the CB dimension for an unprocessed (left) and processed D tone (right) from the 3rd octave of a Clarinet. The left graph shows a sharper and thinner envelope than the figure on the right side. Peaks represent the existing harmonics. It shows also that some harmonics are dominating.

To compare the two signals the upper and lower trajectories of each plot are traced. They are then compared to each other as calculated as in Equation (3.11).

$$\text{Index}_{CB}=|\text{Upper Edge}_{\text{Signal 1}} - \text{Lower Edge}_{\text{Signal 1}}| + |\text{Upper Edge}_{\text{Signal 2}} - \text{Lower Edge}_{\text{Signal 2}}| \quad (3.11)$$

$$Index_{CB} = \sum_{CB} Measure(CB) \quad (3.12)$$

The index is calculated for each CB as represented in Equation (3.12). Pure and complex synthetic tones were compared to each other using 22 and 43 channels after being processed and resynthesized with the AMO and presented in Appendix E, where the results of the proposed index matche with the results from a psychoacoustic pitch discrimination test using

pure tones with NH subjects (see section 4.2.1 and Figure (4.1)). However, the measure could not explain results from a psychoacoustic pitch discrimination test with complex tones (see Figure (4.2) in section 4.2.1 and appendix G.2.3) (Omran et al. 2007). Thus, further investigation of this measure was stopped. The measure in its current form may be used for simple pitch discrimination tests. Integrating an acoustic masking model may enhance it in the future.

3.4 Diagonal Detecting Matrix

This section aims at quantifying the degree Smt mapping algorithms can reach in preserving harmonic structure of overtones in musical tones. Music can be described as a series of complex acoustic sounds composed of tones with fundamentals and overtones that are harmonically related to each other (Salzer 1982). Detecting active diagonals in matrices generated in response to tone sequences with logarithmically increasing fundamental frequencies (see tone sequence) may be one way to quantify harmonic structures preservation for a range of successive tones. The index of the detected diagonals in frequency time matrices is used to calculate the ratio between different partials for sequential tones. Recognition of patterns inside matrices may be time consuming, but using a diagonal detecting matrix (DDM) (Omran 2011) together with a matrix manipulation may be an easier way to estimate diagonal activities. This section describes the DDM in detail in addition to some examples of special cases. Later, the algorithm is summarized in simple steps. A special acoustic sound, representing synthetic musical notes, was processed with the Std, Smt-MF and Smt-LF mappings. The output was preprocessed and was multiplied with the DDM to quantify the amount of harmonic structure preservation. Results are shown at the end of this section.

3.4.1 Description

Definition: DDM is a vertical rectangular or square matrix O[i,j] described by Equation (3.13). A matrix 'A' is singular (not invertible) if the system AO=0 has non trifle solutions (the determinant of O ≠ 0) (Gantmacher 1990). Since the vertical rectangle form of DDM is singular, therefore the Moore-Penrose pseudo inverse is calculated. The first square B of the Moore-Penrose pseudo inverse of DDMs can be singular as well (Neubauer 2003), where B matrix has dimensions equal to the rank of the DDM.

$$\begin{aligned} \text{Matrix } O[i,j] &= 0, && i=m,\ j=n \text{ for } i \neq j \text{ and } i \geq j \\ &= m-n+1, && i=m-n+1,\ j=c \text{ where } 1 \leq c \leq m-1 \text{ and } i \geq j \end{aligned} \tag{3.13}$$

Each diagonal in the matrix is given a representative unique incrementing index number as shown in Figure (3.7).The first diagonal starts with index =1 and increments upto the main diagonal of the matrix. For appropriate recognition, each diagonal has to be fully active otherwise shifting the matrix one row upwards and one column to the left or matrix rotation may be mandatory. Lower triangles cannot be directly detected unless with matrix shifting.

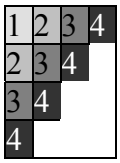


Figure (3.7): Diagonal indices of a matrix are shown in the same color. Each diagonal in the upper triangle of the matrix has a unique index number.

DDM is a newly proposed matrix with zeros everywhere except the main diagonal of the lowest square as described by Equation (3.13) and illustrated in Figure (3.8). All values along the main diagonal increment by one and have their maximum equal to the number of rows (*i*) at the lowest right corner. Figure (3.8) shows two DDMs with 4 rows (i=4) (square and rectangular shapes).

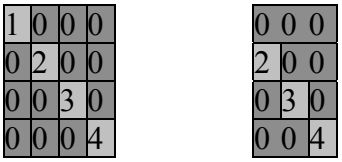


Figure (3.8): An example of a square (left) and a vertical rectangular (right) shape of DDM with i=4. DDM can not be horizontal by definition**.**

In the following section, examples of special cases are selected to describe the idea. To understand the concept, it is easier to start with a simple example for a non-singular DDM and a black and white image matrix represented with one bit (black=0 and white=1).

Proof

Assume an image B (4x4) with 2 diagonals active and the corresponding DDM (4x4) (O)

$$B = \begin{bmatrix} 0 & a & b & 0 \\ c & d & 0 & 0 \\ e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad O = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

Result = B x O

$$= \begin{bmatrix} 0 & 2a & 3b & 0 \\ 1c & 2d & 0 & 0 \\ 1e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

R(i,j) = B(i,j).O(j,j)

If the B image is binary and B(i,j) = 1 are elements in an active diagonal (see Figure (3.9)).

$$R(n,m) = 1.\ O(m,m) \tag{3.14}$$

Equation (3.14) shows that the elements in the resulting matrix (R) that represent the active diagonals do not depend on the image, however the output will represent DDM values in an orderly sequence. The following section illustrates the idea with some examples.

To understand the concept, lets start with a simple example for a non-singular DDM and a binary (black=0 and white=1) image matrix image with four rows and four columns (16 pixels). The lowest right element at i = m and j = n should be replaced by unity and divided by the number of rows (i) in the DDM prior to being multiplied with DDM. This is illustrated in the following examples:-

Example 1

Output = Image x DDM mat

$$\begin{bmatrix} 0 & 2 & 3 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tfrac{1}{4} \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

Figure (3.9): An example to detect diagonal activities in an image using a non singular DDM.

$$|\text{Determinant (Output)}| = |-6| = 3 \times 2 = 6$$

$$Weight = \prod_{1}^{n} output(1,n) \text{ where } , output(1,n) \neq 0 \tag{3.15}$$

$$= 3 \times 2 = 6$$

The weight value is calculated by multiplying all non zero elements in the first row of the output matrix as described by Equation (3.15). The idea behind this equation is to have a unique representing coded output for upto 5x5 matrices which can be used as elements in larger ones. In this example, the determinant of the output equals its weight.

Example 2

Output = Image x DDM mat

$$\begin{bmatrix} 2 & 3 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & \tfrac{1}{4} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Figure (3.10): An example to detect diagonal activities in an image using DDM, whose first square of its pseudo inverse is singular while the output is a non-singular square matrix.

| Determinant (Output) | = | -6 | = 3 x 2

$$Weight = \prod_{1}^{n} output(1,n) \text{ where } , output(1,n) \neq 0$$

= 3 x 2 = 6

Both examples illustrated that the weight value is a multiplication of the diagonal indices in the images that were fully active.

Example 3

The following examples illustrate the case when the image is singular.

Output = Image x DDM mat

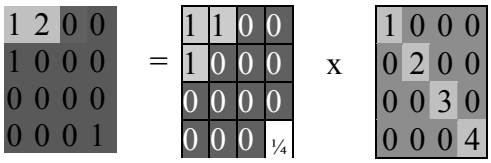


Figure (3.11): An example to detect diagonal activities in an image represented by a singular square matrix. One drawback of DDM its inability to detect activity in a diagonal with index =1. However, this can be programmatically solved with one conditional statement.

| Determinant (Image) | = 0

$$Weight = \prod_{1}^{n} output(1,n) \text{ where } , output(1,n) \neq 0$$

= 2 x 1 = 2

The weight value proposed is not reversally decodable for matrices with row or columns >5. However for large matrices either image decomposition should be employed or another weighting function may be utilized as proposed as in Equation (3.16):

$$Estimated\ Weight = \sum_{1}^{n} 2^{output(1,n)} \qquad (3.16)$$

Steps to detect a diagonal activity:-

1 - Construct a zeros vertical rectangular or square matrix $_{(i,j)}$, where i≥j

2 - Fill the diagonal of the lowest square with incrementing numbers such that element (i=n, j=n) equals n

3 - Change the last (i,j) element in the image matrix into unity divided by n

4 - Post multiply DDM with an image matrix whose diagonal activity is sought.

5 - Multiply all non-zero elements in the first row vector of the output matrix by each other to get the weight.

6- The indices of the active diagonals are to be found in the first row vector of the output matrix.

7- Check the activity of the first elements (1,j) of the input image manually.

```
function x=omran_matrix(i,j)
% Input : Required matrix dimension (i,j)
% Output: Omran Matrix
if nargin < 2,
    x=[];
    disp('please give matrix dimensions');
    return
end
if i<j,
    x=[];
    disp('Error: dimensions are not correct. Matrix can be vertical or square i>=j');
    return
end

x=zeros(i,j);
counter=1;
for q=i-j+1:i
    x(i,counter)=q;
    counter=counter+1;
end

end
```

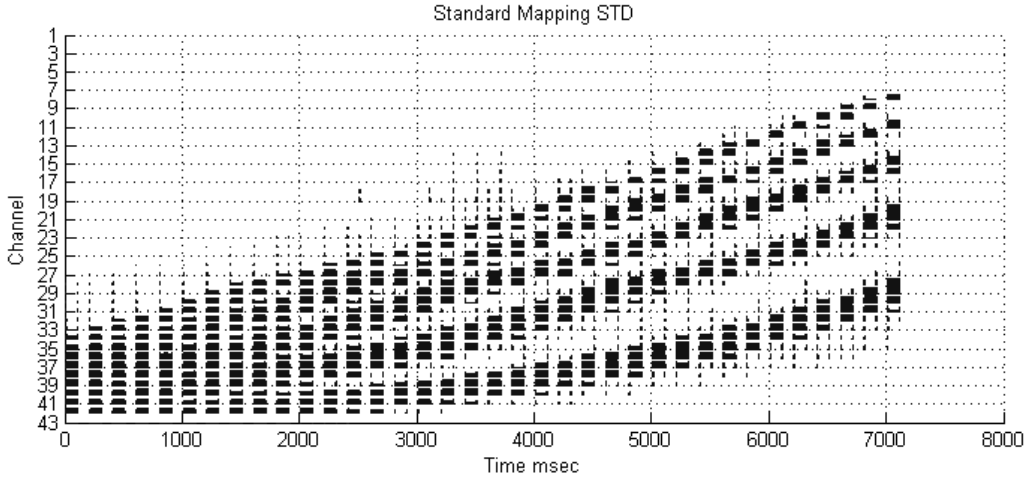
Figure (3.12): A function to construct DDM written in Matlab.

3.4.2 Application

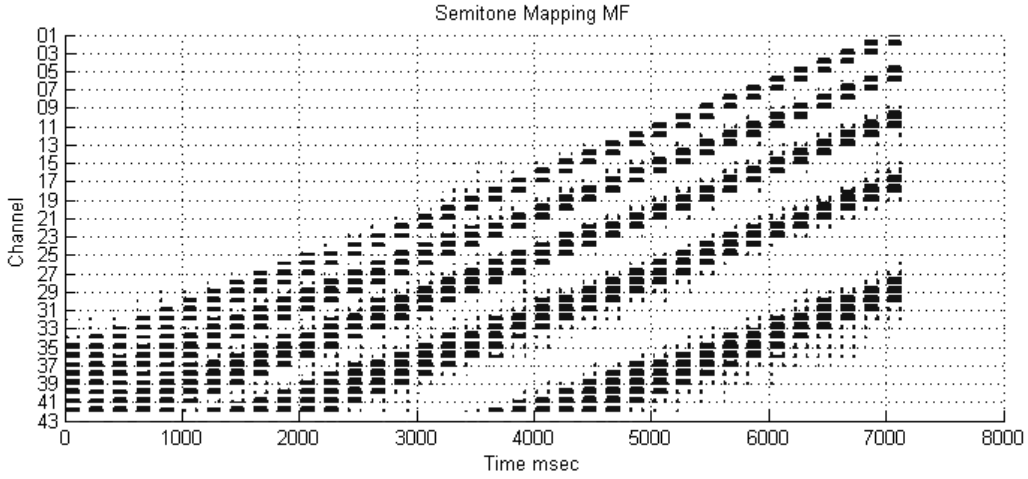
One way to improve melody representation in CIs would be to ensure that the fundamental frequencies of adjacent tones on the musical scale are assigned to separate electrodes as described previously when describing the Smt mapping in section 2.4. Such an approach involves mapping fundamental frequencies of musical tones to electrodes based on a semitone scale. Also, music may be enhanced by increasing the frequency representation using VCs formed by stimulating two adjacent electrodes simultaneously with the same current level. VCs on an array of 22 electrodes would yield a total number of 43 channels, which would cover three and a half octaves with Smt mapping with one-semitone intervals between the characteristic frequencies of successive channels. A comparison between Std and Smt mapping with NH subjects and CI recipients is presented in chapter 4. The amount of harmonic structure preservation is calculated quantitatively with DDM utilizing a special sound sequence processed with Std, Smt-MF and Smt-LF mappings and then resynthesized with a noise band vocoder using AMO (Laneau et al. 2006a).

A sequence of 36 harmonic complex tones was prepared; each tone had the fundamental frequencies of a piano note for a period of 100 ms with 4 overtones. Smt mapping preserved the harmonic structure of overtones as shown in Figures (3.13 and 3.14), where the ratio between the overtones with Smt mapping is almost constant. With the Std mapping there is a distortion in the harmonic structure at low frequencies visible. The order of the signals could also have been chosen from higher to lower pitched signals, but then the preprocessing block would have needed horizontal mirror operation before matrix manipulation. A Matlab program was developed to perform preprocessing and matrix manipulation before applying the DDM algorithm. The flow chart of this program is presented in Figure (3.15).

a)



b)



c)

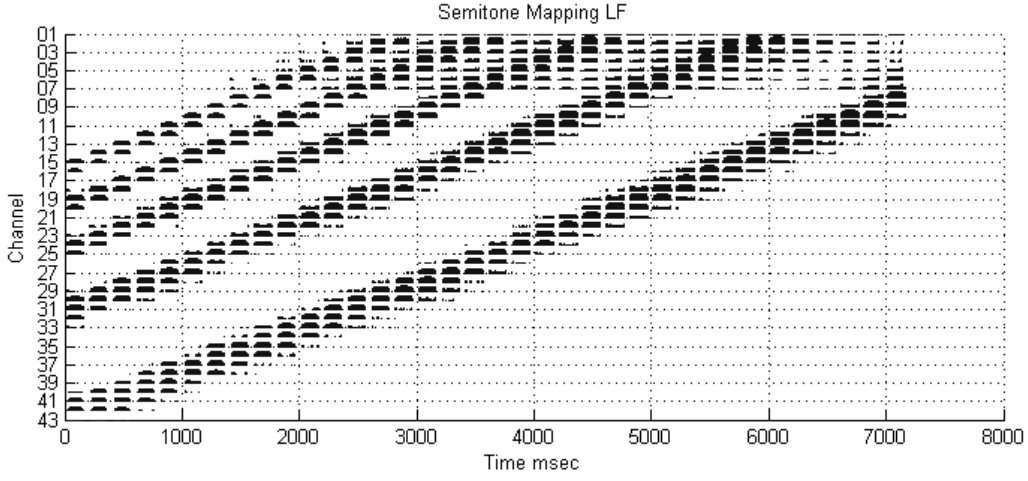


Figure (3.13): CTM outputs from the AMO for the Std (left), Smt-MF (middle) and Smt-LF (right) mapping using 43 channels. Harmonic structure preservation is demonstrated by a linear frequency to channel relationship, as can be seen for Smt-MF and SMT-LF for each of the partials. With the Std mapping, the linearity is seen only for the higher frequencies. At lower frequencies, the partials cannot even be resolved. With Smt-MF components below 440 Hz are filtered out and with Smt-LF the HF partials greater than 1.6 kHz are filtered out (as indicated by arrows).

a)

b)

c)

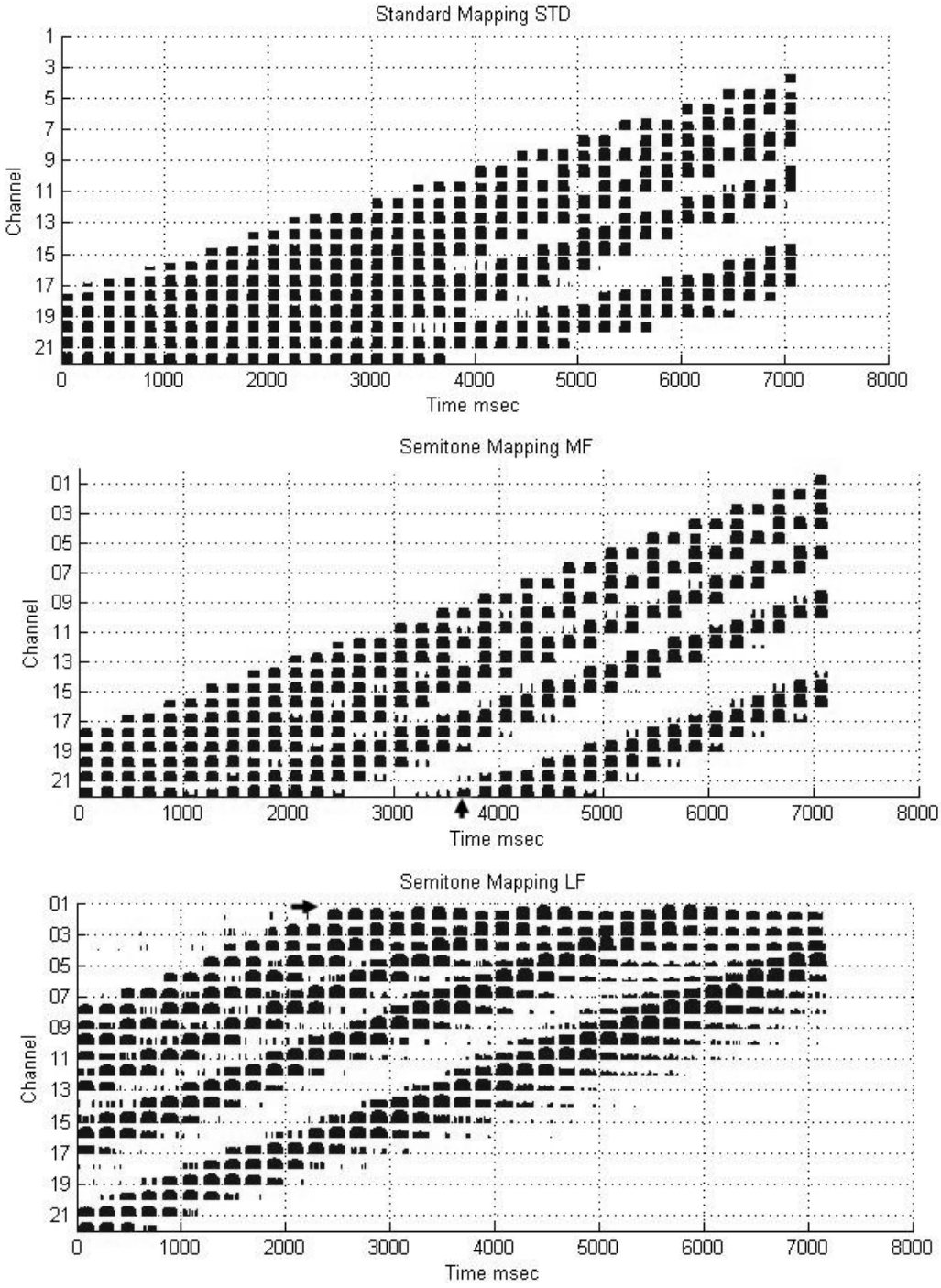


Figure (3.14): CTM outputs from the AMO for the Std (left), Smt-MF (middle) and Smt-LF (right) mapping using 22 channels. Harmonic structure preservation is demonstrated by a linear frequency to channel relationship, as can be seen for Smt-MF and SMT-LF for each of the partials. With the Std mapping, the linearity is seen only for the higher frequencies. At lower frequencies, the partials cannot even be resolved. With Smt-MF components below 440 Hz are filtered out and with Smt-LF the HF partials greater than 1.6 kHz are filtered out (as indicated by arrows).

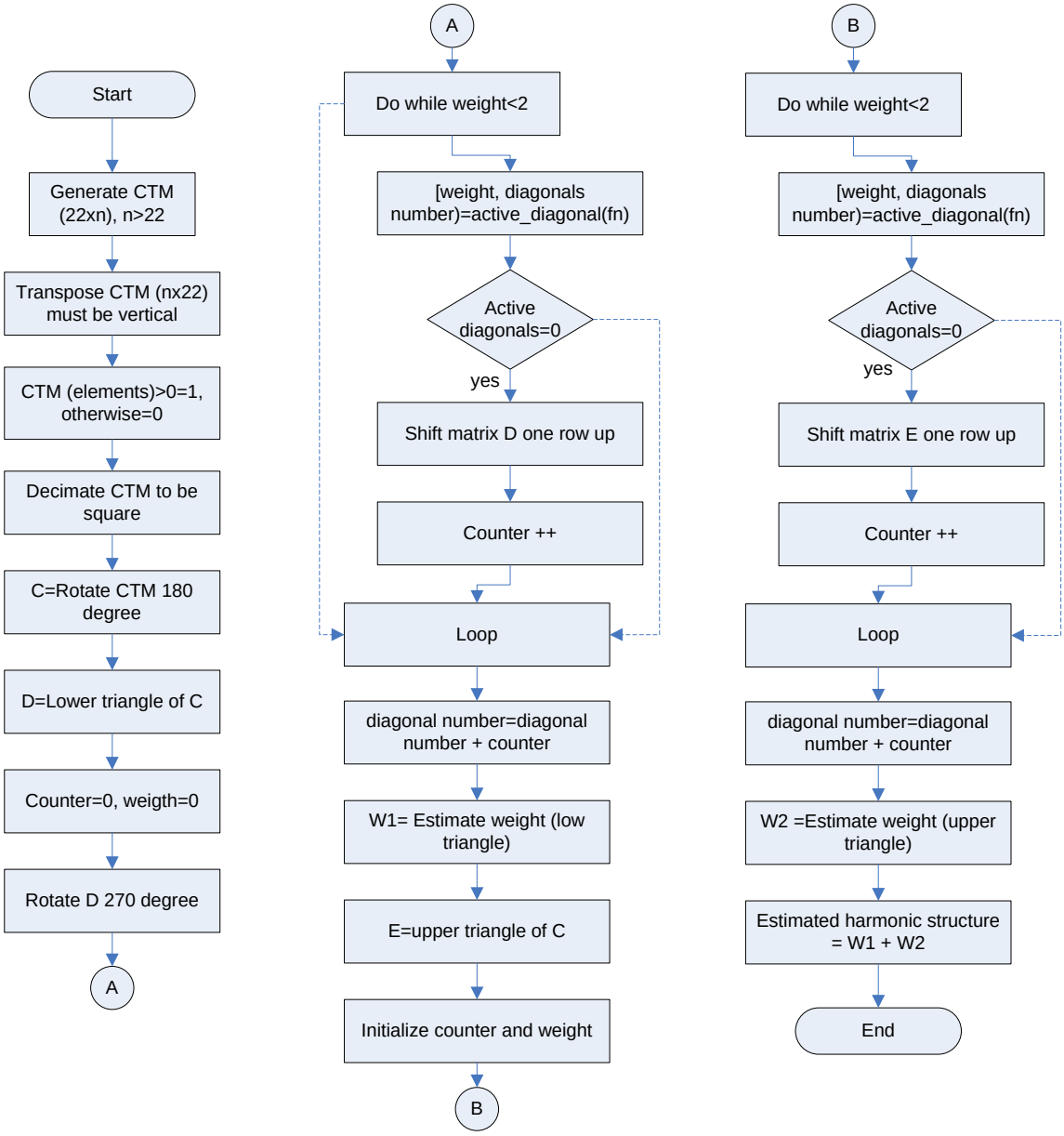


Figure (3.15): A program flow chart to preprocess matrices and then apply DDM to quantify the degree of harmonic structure preservation.

3.4.3 Analysis

The degrees of harmonic preservation in the AMO resynthesized signals were measured using a program described by the flow chart in Figure (3.15). Signals were resynthesized using noise band vocoders with 1 mm SOE using 22 and 43 channels with Std, Smt-LF and Smt-MF mappings.

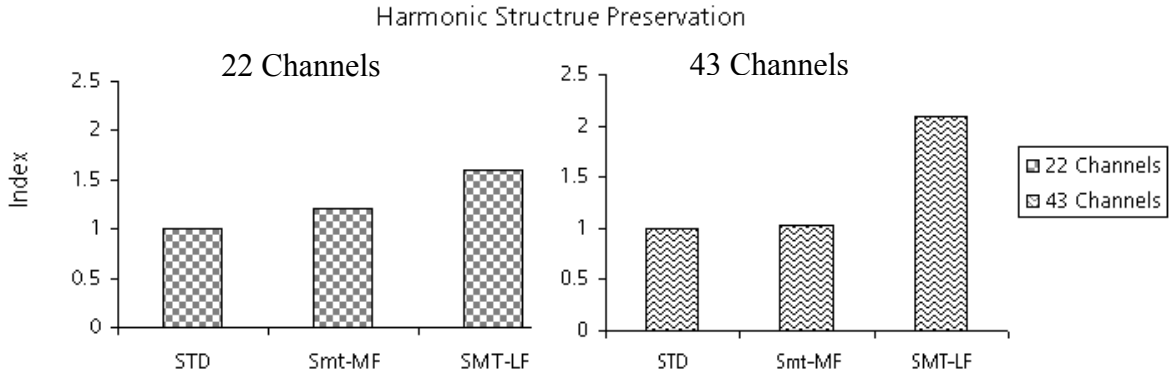


Figure (3.16): Index of preserved harmonic structure with respect to Std mapping for 22 channels (left) and 43 channels (right).

Since the program flowchart depends on equal matrices in form of images representing electrodograms processed using the AMO with a specified mapping, there is no electrodogram for the unprocessed tones. The unprocessed tones could have been plotted in the time frequency domain to form similar matrices but the size will differ and this is expected to introduce other varying factors. Accordingly, the results shown in Figure (3.16) were compared with respect to the Std mapping (index = 1) to illustrate the relative ratio of harmonic structure preservation for both 22 and 43 channels.

Smt-LF mapping enhances the amount of preserved harmonic structure for both 22 and 43 channels. The Smt-MF produces a slight increase using 22 channels and almost no increases using 43 channels with respect to the Std mapping. This is because increasing the resolution of the Std mapping to 43 channels, enhances the visual representation of electrode activity with respect to 22 channels and this decreases the difference to Smt-MF for tones from 2000 to 4000 ms at channels from 37 to 43. In Smt-LF the index increases with 22 channels and even more with 43.

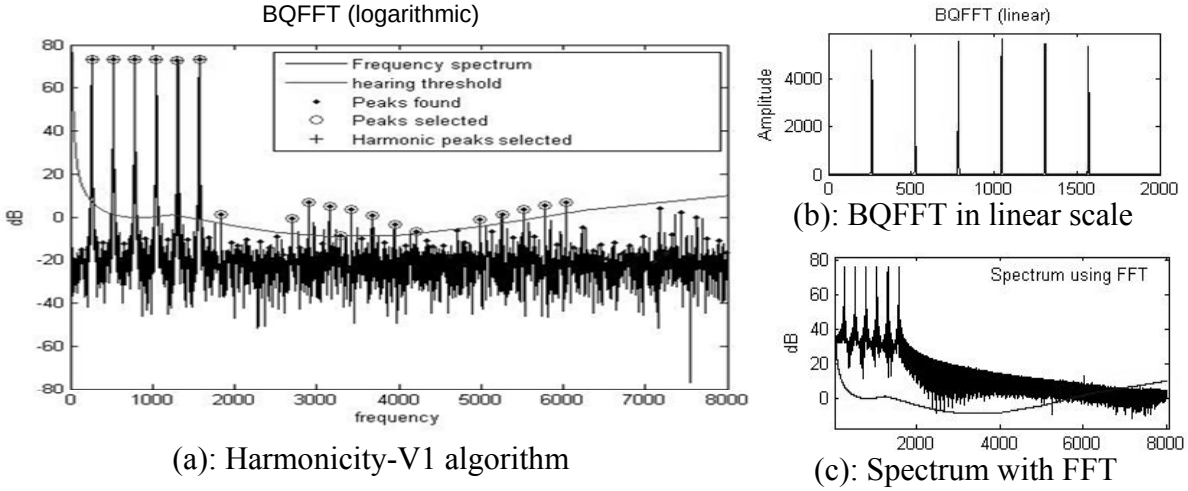
3.5 Harmonicity Estimation

3.5.1 Introduction

The origin of the English word harmony is the Greek word, “harmos”, which means to join. Harmony generally means fitting and acceptable joining of diversities. Similarly, in the real world each musical sound has harmonic overtones. Playing two tones simultaneously forms a chord. In a chord, the overtone structures of the played tones mix producing either a consonance or dissonance chord. A perfect harmonic structure would exist in chords produced from similar notes (Pierce 1983). This section estimates the amount of preserved harmonic structure in sound through three different algorithms. The first two algorithms (“Harmonicity-V1” and “Harmonicity-V2”) depend mainly on detecting the fundamental frequency through peak and formant detectors. If the fundamental is not correctly detected, the algorithm sensitivity decreases. The third algorithm called “harmonicity probabilistic inequality (HPI)” estimates the harmonicity with an approximation independent of the fundamental frequency. The masking model (ISO-13818-3 1994) is incorporated in both the ‘Harmonicity-V2’ and the HPI algorithms. The application of such algorithms is to estimate the amount of relative frequency preservation in simulated sounds produced with a Cochlear Implant acoustic model using standard ACE (Std) and Smt mapping algorithms mappings (Omran et al. 2010; Omran et al. 2011). This section is organized as follows: - the first subsection describes the algorithms and the next subsection describes tests of different musical tones with the proposed algorithms.

3.5.2 Harmonicity-V1 Algorithm

In this algorithm, signals travel through a partial and formant detector, from which the fundamental frequency is estimated. The algorithm extracts formants and partials above the hearing threshold assuming the given wave file is played at 96 dBA. The smallest expected amplitude in the wave file will then be at 3 kHz where the ear is most sensitive (ISO-13818-3 1994).



(a): Harmonicity-V1 algorithm

(b): BQFFT in linear scale

(c): Spectrum with FFT

Figure (3.17): Showing output of harmonicity-v1 algorithm using a synthetic musical tone with 5 overtones (a), illustrating the determined peaks and the selected overtones above the hearing threshold (circles with pluses). Peaks (dots and open circles) above the hearing threshold and below 20 dB represent a noise floor distortion magnified in the logarithmic scale (see (b) in the linear scale). This is a disadvantage of the used BQFFT algorithm against the advantage of low computational cost and higher dynamic range of peaks (~ 20 dB more) than using FFT (see (c)). Such peaks were ignored in the harmonicity algorithm.

All the partial intervals are then related to the first musical note (a0 = 27.5 Hz) and the number of semitones in between is computed. The semitone distances are then rounded to the nearest integer and used to estimate the approximated partial value. If the approximated partial value lies within the ERBs (Moore 2003) located at the frequency of the given partial, the partial is accepted; otherwise is rejected. For an overtone to be harmonic, its frequency has to be an integer multiple of the estimated fundamental frequency. The multiplier value of each partial is rounded and the approximation error is checked for consistency. If using the approximated multiplier value and the approximated partial lies inside the ERB, the partial is again selected or otherwise rejected. The algorithm searches for the harmonic partials among all selected partials. It uses a reciprocal distribution function ($f(x) = 1/(x(\ln 10 - \ln 1))$) with boundary parameter ($1 \leq x \leq 10$) to estimate the probability for the full signal to be harmonic if the preceding harmonic tones exist. The overall mean value is the expected harmonicity value for each time frame. Figures (3.17 and 3.18) show the peak detection algorithm applied to the spectrum of a synthetic harmonic tone with 5 overtones estimated using the bounded-Q fast filter bank transform (BQFFT) (Diniz et al. 2007) algorithm and the flow-chart of the used algorithm respectively.

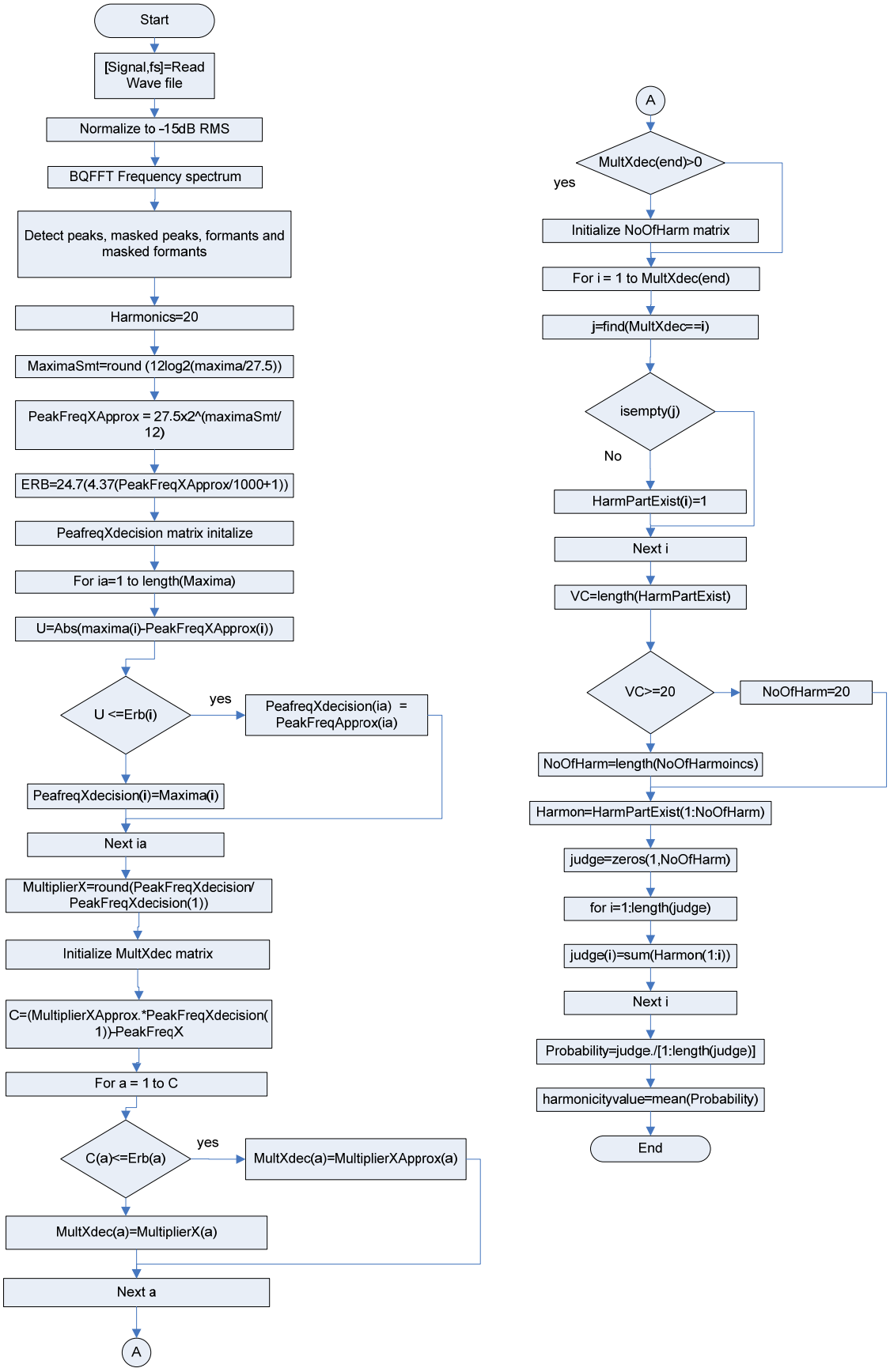


Figure (3.18): Flowchart of the harmonicity v1 algorithm. This algorithm depends on the correct estimation of the fundamental frequency.

3.5.3 Harmonicity-V2 Algorithm

In this section a second algorithm called "harmonicity-V2" is presented which is more advanced than the previous one. In this algorithm the input signal is normalized to a common RMS value at -15 dB to avoid clipping. It is then processed with the BQFFT (Diniz et al. 2006) to produce the frequency spectrum. This algorithm undergoes a simultaneous heuristic masking model of partials based on the limited ability of the human ear to distinguish changes in the BM excitation (Bosi and Goldberg 2003). The masking model is inherited from the standard MPEG algorithm (ISO-13818-3 1994). The "harmonicity-V2" algorithm uses a Matlab implementation for the MPEG layer I model (Petitcolas 1998) with special changes to compute the masking effect. In the masking model, a signal is buffered into nonoverlapping 8 ms time frames and each frame is convolved with a Hanning filter. The output is transferred to the frequency domain using 512 FFT points and the signal level (L_k) is computed for each spectral line k as follows (quoted from Bosi and Goldberg 2003)

"$L_k = 96dB + 10\log_{10}\left(\frac{4}{N^2}|X(k)|^2 \,/\, 8/3\right)$ for k=0,…, N/2-1

where X[k] represents the FFT output of the time frame and N equals 512 for Layer I. The signal level is normalized so that the level of an input sine wave that just overloads the quantizers, here defined as being at x[n]=±1.0, has a level of 96 dB when integrated over the peak. In this equation, the factor of $1/N^2$ comes from Parseval's theorem, one factor of 2 comes from the power of a unit amplitude sinusoid being equal to ½, and the factor of 8/3 comes from the reduction in gain from the Hanning window" (Bosi and Goldberg 2003)

The frequency domain is divided into 32 sub bands as defined in (ISO-13818-3 1994) and the SPL is calculated for each sub band. The algorithm then searches for local maxima at which the derivative direction changes or becomes 0.

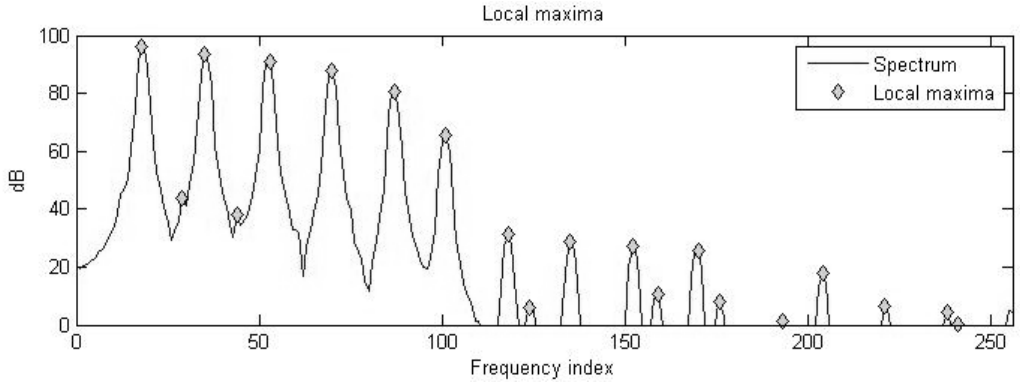


Figure (3.19): Showing the spectral magnitude and local maxima (diamond) found in a randomly selected time frame of a synthetic musical tone with 5 overtones using 512 point FFT points and sampling rate 16 kHz. Maxima below 40 dB are distortion arising from the BQFFT used (see Figure (3.17) for illustration).

Tonal peaks are peaks higher than the neighboring samples with 7 dB according to the MPEG ISO standard (ISO-13818-3 1994). The number of neighboring samples is dependent on the frequency index and sampling frequency. It searches for tonal peaks among the local maxima.

“Since noise is a better masker than tones, a search for tonal maskers in the signal is performed in Model l. This evaluation is based upon the assumption that local maxima within a critical band represent the tonal components of the signal” (Bosi and Goldberg 2003)

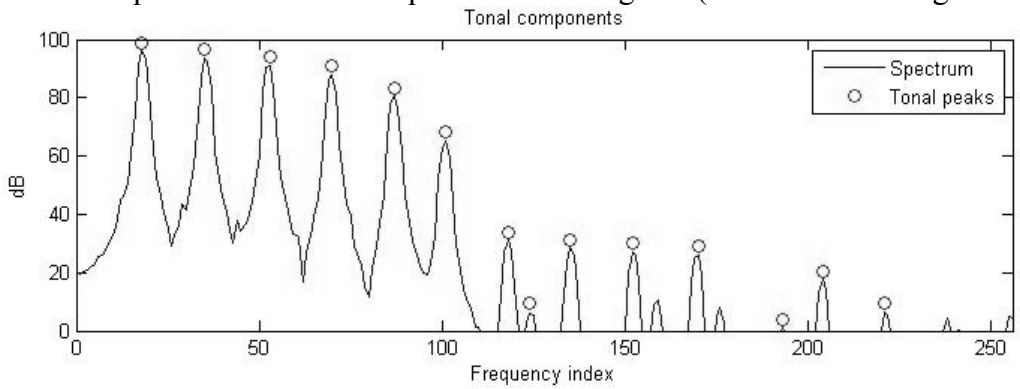


Figure (3.20): Tonal peaks (open circles) are peaks higher than the neighboring samples by 7 dB. Tonal peaks are selected from the determined local maxima.

Figure (3.20) shows the position of tonal peaks illustrated with open circles. The amplitude of the non-tonal peaks is the power in each CB and the frequency index number for the non tonal component is the nearest index to the geometric mean of the CB (Petitcolas 1998).

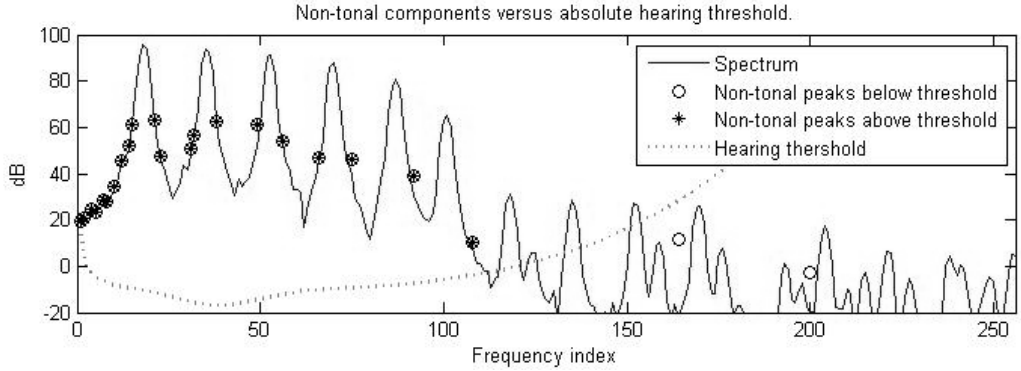


Figure (3.21): Spectrum (solid), non-tonal spectral amplitude values below (circle) and above (filled circles and pluses) the hearing threshold (dotted).

The hearing threshold is the absolute hearing threshold in quiet for NH subjects. It is computed and used to select tonal peaks that exceed this threshold as shown in Figure (3.22). Only tonal peak is allowed within a frequency range of ½ Bark; if more than one peak is found, the highest is selected.

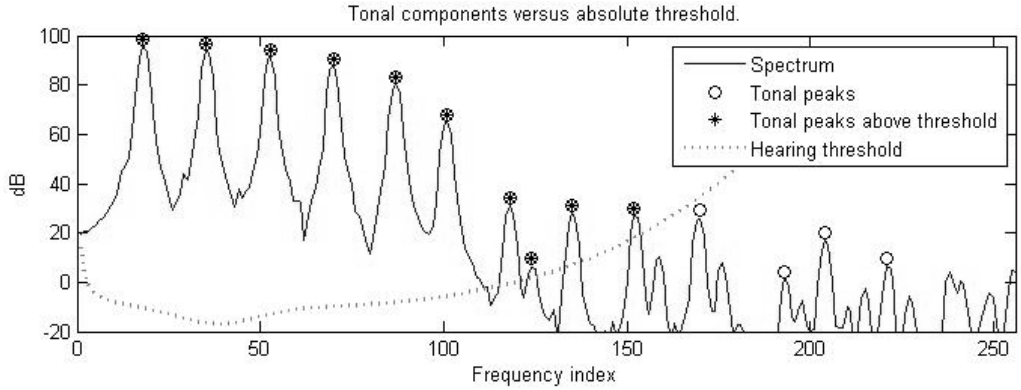


Figure (3.22): Spectrum magnitude (solid) and tonal peaks above (filled circles and pluses) and below (open circles) the absolute threshold (dotted).

For each peak the masking curve located at its frequency is computed using ISO/IEC MPEG Psychoacoustic Model 1 spreading function (Bosi and Goldberg 2003) as defined by Equation (3.17)

$$10\log_{10}(F(dz,L_M)) = \begin{cases} -17\,dz + 0.15\,L_M\,(dz-1)\,\theta(dz-1) & \text{for } dz \geq 0 \\ -(6+0.4\,L_M)\,|dz| - (11+0.4\,L_M)\,(|dz|-1)\,\theta(|dz|-1) & \text{for } dz < 0 \end{cases} \tag{3.17}$$

where Lm is the loudness level in each CB and *dz* is Bark frequency located at frequency f of the tonal and non-tonal peaks defined by Equation (3.18)

$$dz = 13\tan^{-1}(0.76f) + 3.5\tan^{-1}\left(\left(\frac{f}{7.5}\right)^2\right) \tag{3.18}$$

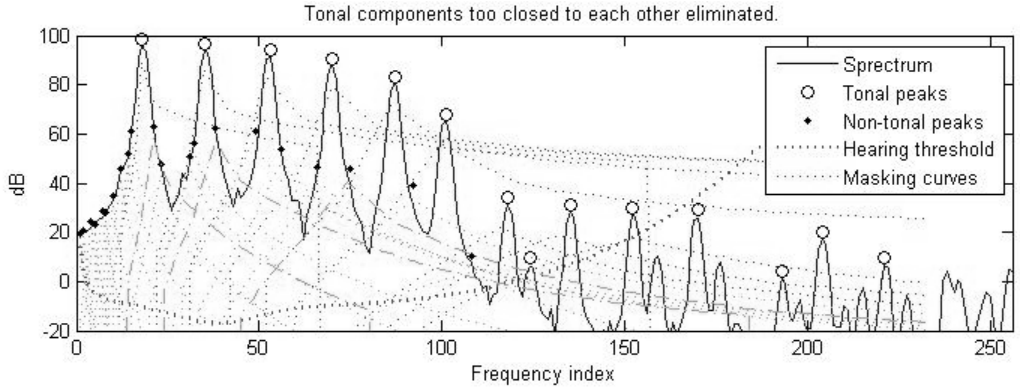


Figure (3.23): Showing masking curves for all tonal (open circles) and non-tonal components (filled circles).

All the excitation spreads (I_n) are combined together according to Equation (3.19) to produce the overall masking level as shown in Figure (3.24).

$$I_N = \left(\sum_{n=0}^{N-1} I_n^{\alpha}\right)^{\frac{1}{\alpha}} \qquad 1 \leq \alpha \leq \infty \tag{3.19}$$

where α in Equation (3.19) is a parameter that defines the way the curves add, setting α = 1 corresponds to intensity addition while taking the limit $L_{\alpha\to\infty}$ corresponds to using the highest masking curve. Setting α to values between 1 and infinity gives results intermediate to these two cases. In the MPEG layer I and II coder α = 1 is used.

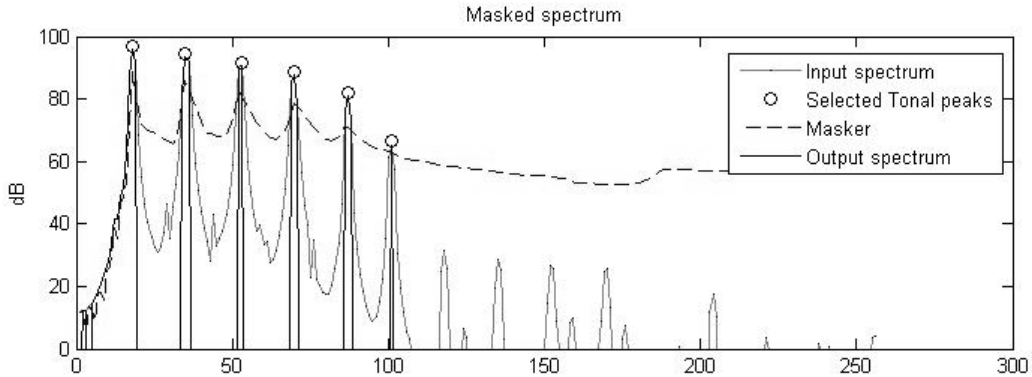


Figure (3.24): Showing the input spectrum (gray), the overall masker (dashed), part of the spectrum that exceeds the masker (solid) and the selected tonal peaks (circle).

Once the tonal unmasked components are extracted, these values are passed to the harmonicity estimation routine. The estimation routine searches for formants in the input frame and by comparing the detected formants with the extracted tonal peaks, formants within the range of ±75 Hz from the tonal components are selected and called the MPEG formants. The convention MPEG is used here to indicate that the computation depends on the masking model.

The harmonicity-V2 algorithm processes each buffered frame twice; in the first and second round it calculates the harmonicity using formants, and the MPEG formants respectively. Both of them are held in a variable matrix called "Maxima" in each round. Assuming that the first determined maximum represents the fundamental frequency with an arbitrary tolerance level of ± 100 Hz. The first 19 harmonic overtones are computed with respect to the assumed fundamental. A problem occurs when the fundamental is missing; this causes the algorithm to malfunction and abort.

A GTF bank with 64 filters is constructed using the Matlab auditory toolbox (Slaney 1998). The estimation routine searches for filter indices with nearest CF to the expected harmonic overtones. The ERBs are also calculated (Glasberg and Moore 1990). The input time frame is convolved with the GTFs and the power in each Gammatone and ERB filter is then computed. The computed values are divided with the total frame's power to setup internal parameters (e.g. thresholds and relative error in each ERB filter with respect to its CF).

The algorithm refines the Maxima and selects those falling within ERBs ranges in which the expected harmonic overtones are expected. A matrix representing the existence of the expected harmonic overtones is constructed and the corresponding maxima are represented by flags equal to 1 (present) or 0 (missing).

The expected harmonicity for the maxima is the mean value of the flags in the preceding harmonic overtones. The parameters computed above are utilized in computing the mean value using different distributions (equal, Binomial, triangular and Gaussian); adopted here an equal distribution assuming that all CBs in the linear range may represent approximately equal chances for a sound to be fully harmonic. The mean value of harmonicity values for the maxima is calculated and represents the overall expected harmonicity. Figure (3.25) presents a flowchart of this algorithm.

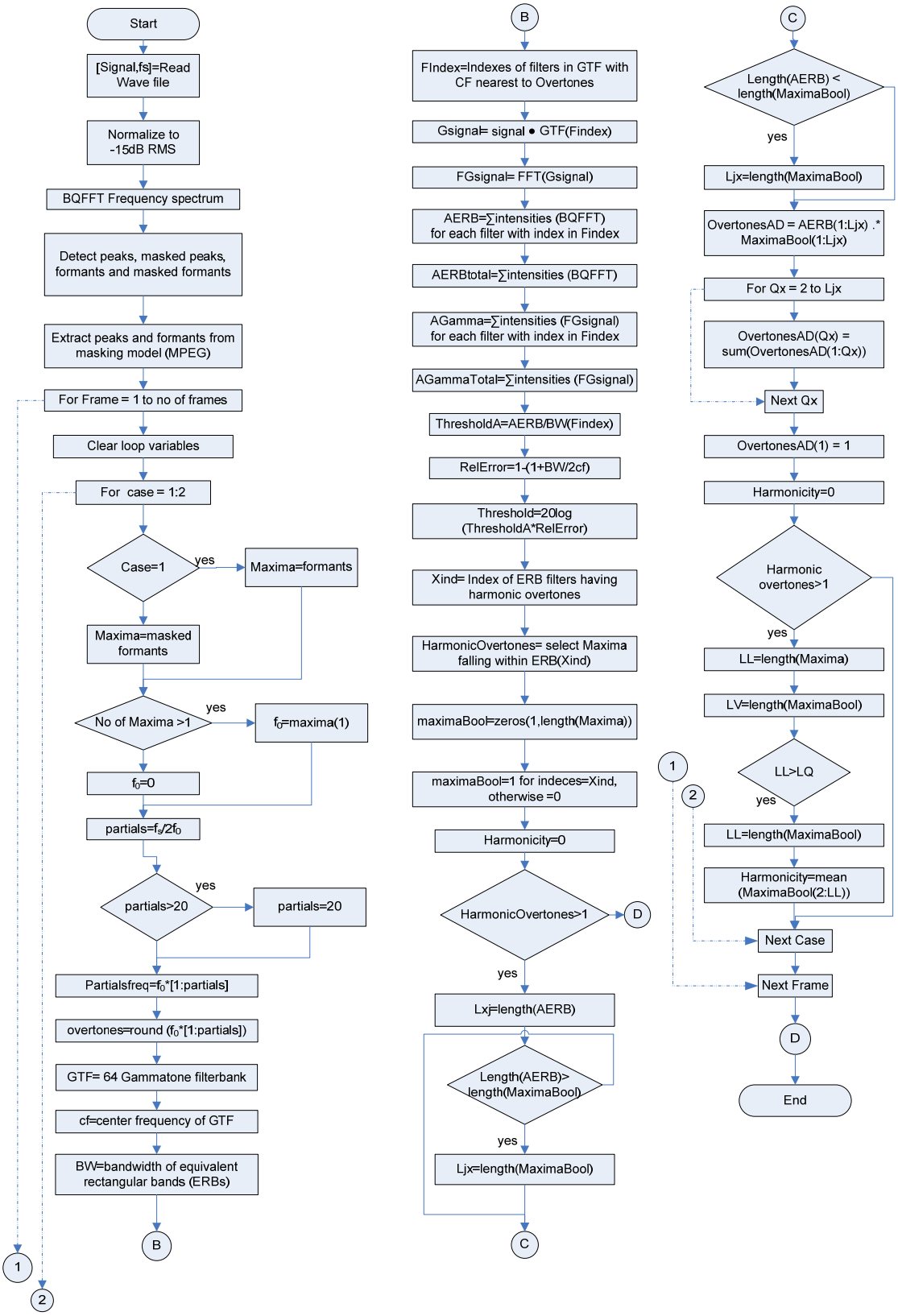


Figure (3.25): Flowchart for "Harmonicity estimation-v2" algorithm. This algorithm mainly depends on the correct estimation of the fundamental frequency.

3.5.4 Harmonicity-V3 (HPI) Algorithm

The previous two algorithms ("Harmonicity-V1" and "Harmonicity-V2") depended on the correct estimation of the fundamental frequency. However in many sounds the fundamental may be missing or difficult to estimate due to various noise sources. Thus this algorithm called HPI was developed with a different approach to overcome such a problem using probability. It mainly depends on estimating the distance from given frequency components f_1 and f_2 to their nearest harmonic overtones with a tolerance of ½ Bark at each direction.

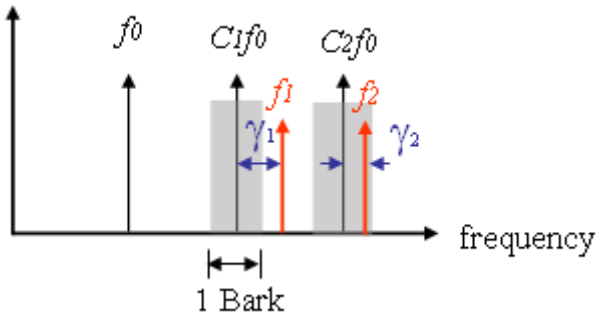


Figure (3.26): Example showing two detected frequency components (f_1 and f_2) between harmonic overtones (C_nf_0). The distance between the two components and the nearest harmonic overtone is γ_n, where n is the harmonic number. The algorithm uses a window tolerance of 1 Bark.

$$\frac{f_2}{f_1} = \frac{C_2 f_0 \pm \gamma_2}{C_1 f_0 \pm \gamma_1}$$

$$\delta_2 = \frac{\gamma_2}{C_2} \tag{3.20}$$

$$\delta_1 = \frac{\gamma_1}{C_1} \tag{3.21}$$

δ_1 and δ_2 are the relative difference of a given frequency from the expected overtone with respect to overtone index.

$$\frac{f_2}{f_1} = \frac{C_2(f_0 \pm \delta_2)}{C_1(f_0 \pm \delta_1)} = \frac{C_2}{C_1}\left(\frac{f_0}{f_0 \pm \delta_1}\right) \pm \frac{C_2}{C_1}\left(\frac{\delta_2}{f_0 \pm \delta_1}\right)$$

Assuming $f_0 >> \delta_1$, example: $f_0 = 500$ Hz, $f_1 = 520$ Hz, $f_5 = 2600$ Hz.

$\gamma_1 = 5x500 - 2600 = 100$ Hz

$\delta_1 = \frac{100}{5} = 20$ Hz/overtone number

$\therefore 520 >> 20$

$$\frac{f_2}{f_1} = \frac{C_2}{C_1} \pm \frac{C_2}{C_1}\left(\frac{\delta_2}{f_0 \pm \delta_1}\right)$$

$$\chi_2 = \frac{f_0}{\delta_2} \tag{3.22}$$

$$\chi_1 = \frac{f_0}{\delta_1} \tag{3.23}$$

$$\psi = \frac{\chi_2}{1+\chi_1} \tag{3.24}$$

$$\frac{f_2}{f_1} = \frac{C_2}{C_1} \pm \frac{C_2}{C_1}\left(\frac{\chi_2 . f_0}{f_0(1 \pm \chi_1)}\right)$$

$$\frac{f_2}{f_1} = \frac{C_2}{C_1} \pm \frac{C_2}{C_1}\left(\frac{\chi_2}{1 \pm \chi_1}\right)$$

There are two cases for positive and negative frequency shifts,

$$\frac{f_2}{f_1} = \frac{C_2}{C_1} + \frac{C_2}{C_1}\left(\frac{\chi_2}{1+\chi_1}\right) \text{ and } \frac{f_2}{f_1} = \frac{C_2}{C_1} - \frac{C_2}{C_1}\left(\frac{\chi_2}{1-\chi_1}\right)$$

$\because$ C_n is an integer

$\therefore \frac{f_2}{f_1} = fix\left(\frac{f_2}{f_1}\right) \pm fix\left(\frac{f_2}{f_1}\right).\psi$, where fix is a Matlab function which rounds to the nearest integer

Case 1:

$$\psi = \frac{\chi_2}{1+\chi_1} = \frac{\delta_2}{f_0}\left(\frac{1}{1+\frac{\delta_1}{f_0}}\right) = \frac{\delta_2}{f_0+\delta_1}$$

$$\psi = \frac{\gamma_2}{C_2} . \frac{1}{f_0 + \frac{\gamma_1}{C_1}}$$

Let $\varphi = \frac{C_2}{C_1}$

$$\psi = \frac{\gamma_2}{\varphi} . \frac{1}{f_0 C_1 + \gamma_1}$$

$$\psi . \varphi = \frac{\gamma_2}{f_0 C_1 + \gamma_1} = \frac{\gamma_2}{f_1}$$

$$\gamma_2 = \psi.\varphi.f_1 \tag{3.25}$$

$f_2 - \gamma_2 = C_2.f_0$, this value is a harmonic overtone

Check whether ½ bark tolerance width is met

$$\gamma_2 \le f(Bark(C_2 f_0 + \frac{1}{2}) - (C_2 f_0) \tag{3.26}$$

where, Bark (.) is a function to calculate bark value for a given frequency while f(Bark(.)) is the frequency for a given bark value.

If the inequality given by Equation (3.26) is true, the two given frequencies f_1 and f_2 are expected to be harmonic.

Case 2:

$$\psi = \frac{\chi_2}{1-\chi_1} = \frac{\delta_2}{f_0}\left(\frac{1}{1-\frac{\delta_1}{f_0}}\right) = \frac{\delta_2}{f_0 - \delta_1}$$

$$\psi = \frac{\gamma_2}{C_2}.\frac{1}{f_0 - \frac{\gamma_1}{C_1}}$$

$$\psi = \frac{\gamma_2}{\varphi}.\frac{1}{f_0 C_1 - \gamma_1}$$

$$\gamma_2 = \psi.\varphi.f_1$$

$$\gamma_2 \le (C_2 f_0) - f(Bark(C_2 f_0 - \frac{1}{2}) \tag{3.27}$$

$$\gamma_1 = f_1 - C_1 f_0 = f1 - \frac{(f_2 - \gamma_2)}{fix(f_2 / f_1)}$$

If the inequality given by Equation (3.27) is true, the detected two frequency components are very likely to be in a harmonic ratio with a tolerance of ½ Bark and their corresponding flag is set to high.

As a conclusion, for two frequency components to be harmonic, the inequalities described by Equations (3.26 and 3.27) must be satisfied.

Note that if f_2/f_1 is an integer value, they are definitely in a harmonic relationship and there is no need to perform the calculation. Figure (3.27) shows a flowchart for this algorithm being implemented in Matlab.

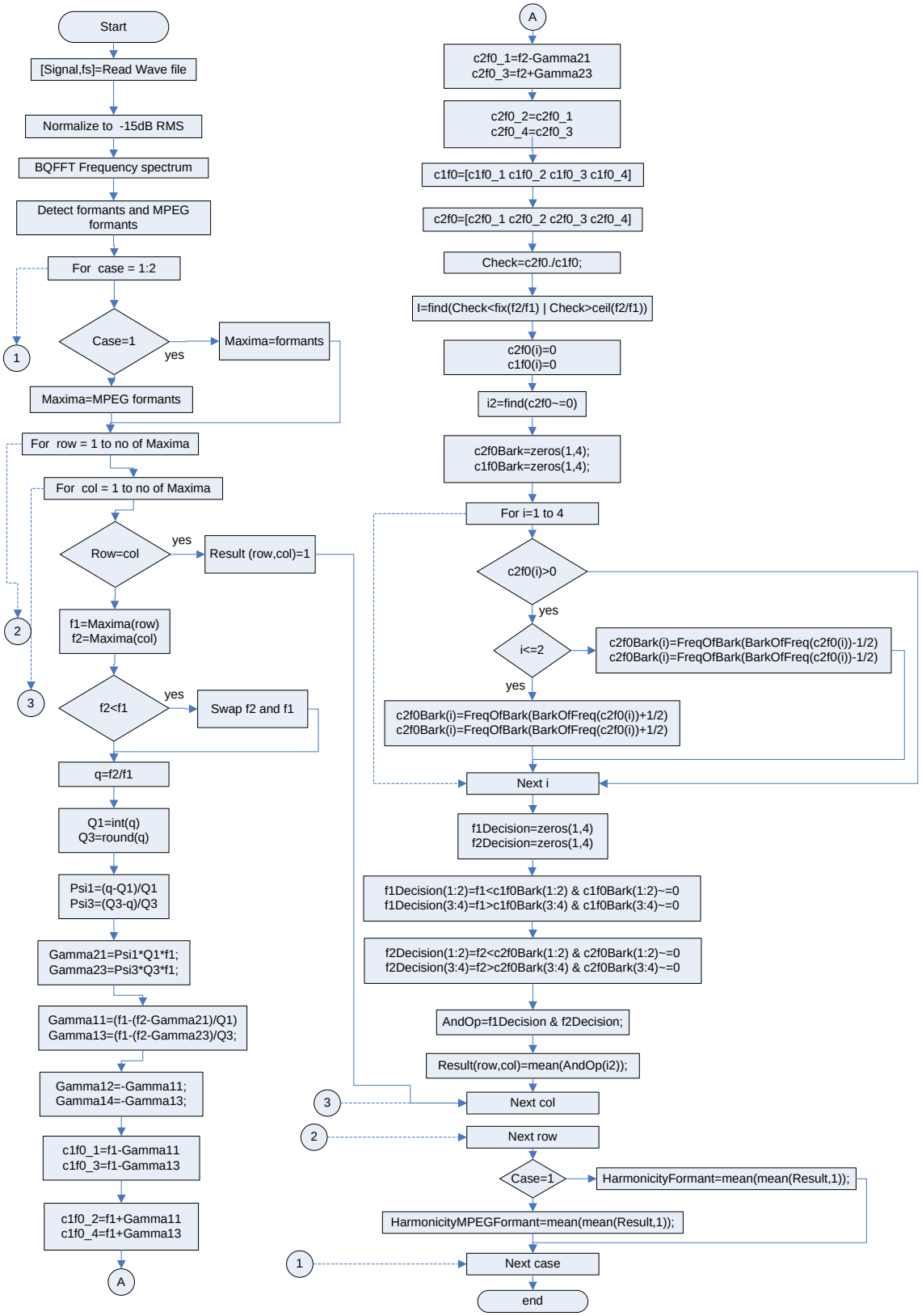


Figure (3.27): Flowchart for “Might-be-harmonic” algorithm. This algorithm does not depend on the fundamental frequency but uses an approximation.

3.5.5 Analysis

The three harmonicity algorithms (V1, V2 and V3) were used to analyze musical tones from the RWC music library (Goto et al. 2001) for different musical instruments. The musical instruments are divided into two groups; one group contains notes in octave 4 and 5 from Piano, Cello, Clarinet and Trombone while the second group contains notes in the same octaves from Guitar, Violin, Flute and Trumpet. Bell and white Gaussian noise tones were added for reference. Algorithms V1, V2 and V3 were used to analyze tones from group 1, while tones from both groups were analyzed with algorithm V3 in a further step. Each sound was buffered into non-overlapping 8 ms frames and each frame was processed by the three proposed algorithms using peaks, formants and their corresponding MPEG variations; these variations were generated from a masking model. Figures (3.28-3.30) show the results of the analysis using the three algorithms, the harmonicity value varied between 1 (fully harmonic) and 0 (inharmonic).

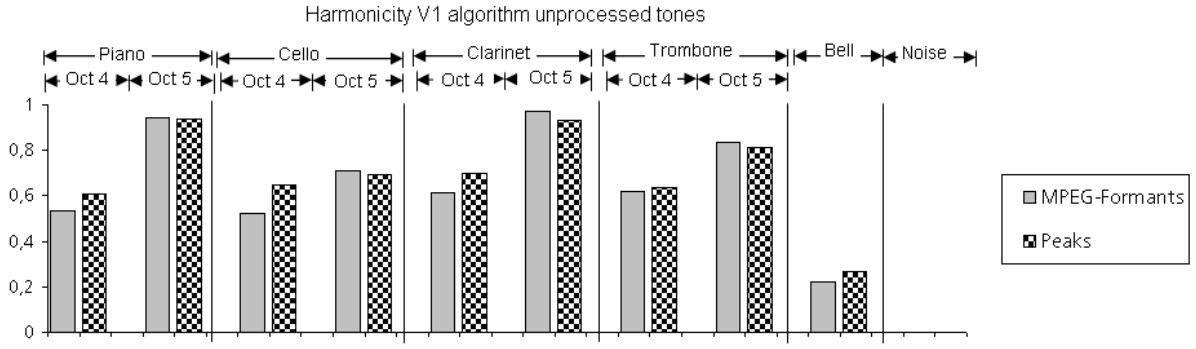


Figure (3.28): Average harmonicity index of time frames with the harmonicity V1 algorithm for tones from Piano, Cello, Clarinet and Trombone in octaves 4 and 5, in addition to bell and white Gaussian noise. The algorithm uses peaks detected in acoustic sound (bricked) and the unmasked formants (gray) according to the MPEG standard.

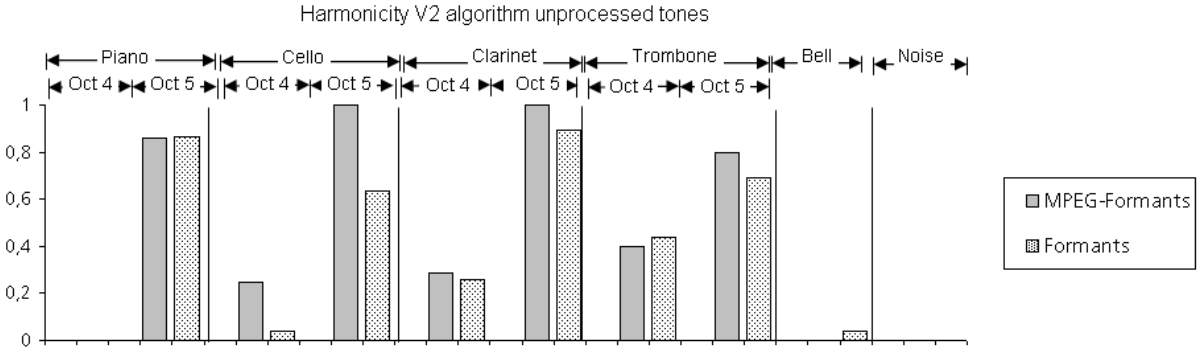


Figure (3.29): Average harmonicity index of time frames with the harmonicity V2 for tones from Piano, Cello, Clarinet and Trombone in octaves 4 and 5, in addition to bell and white Gaussian noise. The algorithm uses formants (dots) and the unmasked formants (gray) according to the MPEG standard.

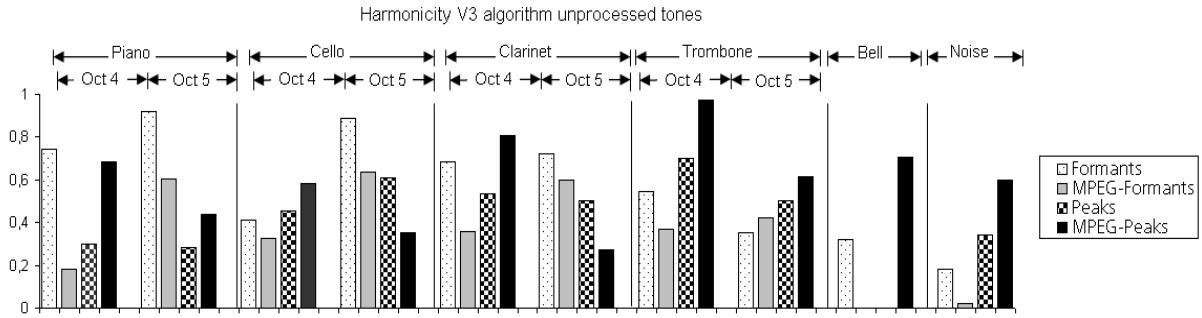


Figure (3.30): Average harmonicity index of time frames with the harmonicity V3 for tones from Piano, Cello, Clarinet and Trombone in octaves 4 and 5, in addition to bell and white Gaussian noise. The algorithm was analyzed using formants (dot), masked formants according to the MPEG standard (gray), peaks detected in tones (bricked) and the masked peaks from MPEG standard (black).

Figures (3.28-3.30) show that fundamental frequencies in the clarinet and piano tones could be clearly estimated in octaves 4 and 5, and therefore algorithm V1 produced high results approaching unity while algorithm V2, showed more octave sensitivity than V1. V2 depends on the energy distribution per CB and since CBs at low frequencies are narrower, this may explain the poorer performance to estimate the fundamental frequency in octave 4. Thus, the results are greater than 0.7 in octave 5 and sharply drop in octave 4. In a piano, pressing a key produces a delta like function and invokes additional partials that mix with the harmonic structure of the tone. This falsifies the estimation of the fundamental frequency with algorithm V1 especially at low frequencies such as in octave 4 with piano tones and hence the harmonicity index decreases while in octave 5 the fundamental frequency is relatively clearer than other partials. Similar consequences occur with clarinet and trombone due to the vibrato partials. Bell sounds produce relatively lower or equal indices than musical tones using algorithm V1 and V2. White Gaussian noise is considered an inharmonic signal and thus produces low values approaching zero using V1 and V2 algorithms.

In general, the V1 algorithm is more stable than V2 but it may be octave dependent which is not a desired feature. On the other hand, algorithm V3 looks more stable than the V1 and V2 algorithms and is designed not to be octave dependent. Figure (3.30) illustrates the general analysis with algorithm V3. It shows that index values for musical instruments using Formants are relatively higher than for Bell and noise. Bell sounds are less harmonically than tones from musical instruments. However, there may be ranges of harmonicity; some instruments are more harmonically than others. MPEG-Formants represent existing formants that are not masked with the MPEG masking model. The masking model is frequency dependent and this may produce octave dependent harmonicity indices using MPEG-Formants, which is illustrated in Figure (3.30) where results from octave 5 are higher than for octave 4 for all musical instruments using MPEG-Formants. In general, the harmonicity index of musical tones with MPEG-Formants is higher than for Bell and Noise. Algorithm V3 uses two different modules for the detection of peaks and unmasked MPEG-peaks. Index values from musical instruments using peaks are higher or equal to noise which may be misleading, because in the frequency spectrum of a white Gaussian noise many peaks are found in a random form rather than having an equal amplitude. But this nevertheless increases the harmonicity index of noise. A similar case occurs with the MPEG-peaks version. This illustrates the need for a weighting function that changes the dependence on these estimating parameters (Formants, MPEG-Formants, Peaks and MPEG-Peaks) with different portions to fulfill a certain target criterion.

The harmonicity index in many tones with V2 could not be calculated due to improper estimation of the fundamental frequency. On the other hand, V1 produced frequency dependent results which may not be desired. This analysis suggests that the third algorithm, V3, may be generally recommended because it is more stable than V1 and V2 algorithms. A weighting function is proposed to combine the results from different estimating parameters and the output is divided by the total sum of the weights. The optimum weights to be used with the V3 algorithm are as follows: 1.5, 0.1, 0.5 and 0.2 for formants, MPEG formants, MPEG peaks and peaks respectively as shown in Figure (3.31). These values are found to satisfy the following criteria:

1- Maximize results from unprocessed and processed tones with respect to the harmonicity of an unprocessed white Gaussian noise.
2- Minimize difference between unprocessed tones from similar instruments at different octaves.
3- Maximize significant differences between tones from all strategies with respect to unprocessed tones.

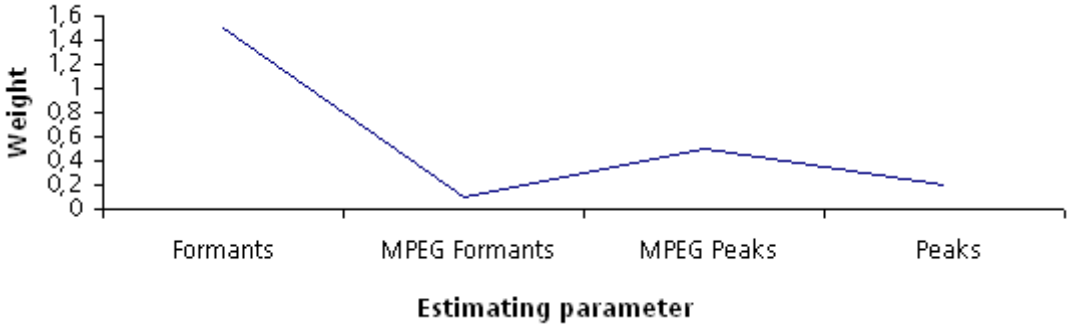


Figure (3.31): Weighting values for estimation parameters used in algorithm V3.

The V3 harmonicity index was validated with two tones; each tone had 6 harmonic overtones. The fundamental frequency of the lower tone was fixed at 250 Hz and that of the upper tone varied from a little above 250 Hz to a little below 500 Hz (an octave above 250 Hz). The harmonicity index measured for each chord was compared to data plotted from (Pierce 1983).

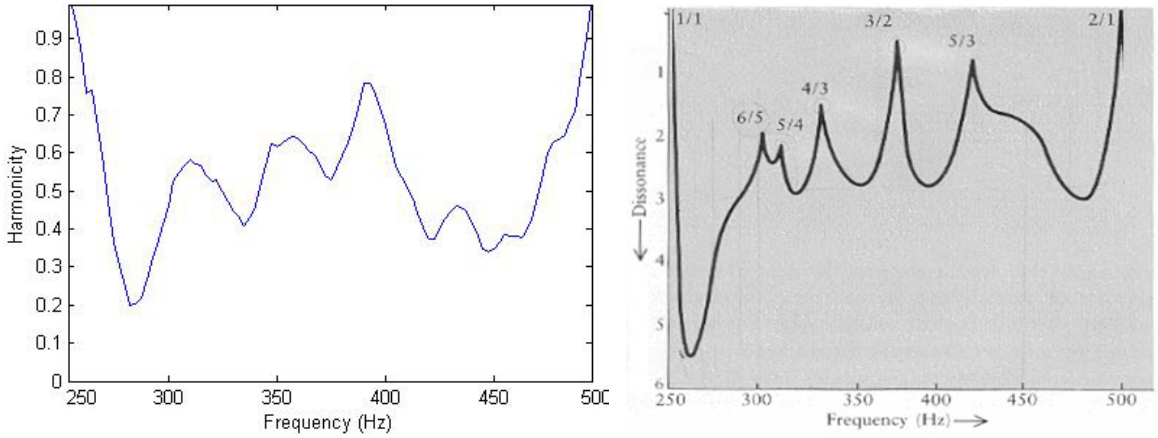


Figure (3.32): Harmonicity produced by V3 index from two tones; the fundamental frequency of the first is fixed at 250 Hz while the second is slightly shifting with respect to the first (left) compared to results from a consonance-dissonance study (right)(Pierce 1983).

Figure (3.32) illustrates that the results with the harmonicity index V3 are similar to those of a consonance-dissonance study by (Pierce 1983). However, there are some differences. The

tolerance used in the harmonicity index V3 is ± 0.5 Bark and this may be the cause of different lobes sharpness. Additionally, there is a slight shift with a relative error less than 4 % between the two plots; this may be because the harmonicity index uses an approximation which is relatively high at low frequencies. But generally the main lobes exist.

Examining the harmonicity algorithm V3 systematically with chords was done by constructing two pure synthetic tones (reference and probe). Each tone has a fundamental frequency and five harmonic overtones. The fundamental frequency of the reference tone varies in a frequency range from 32 to 2217 Hz with one semitone interval step. The fundamental frequency of the probe was shifted with respect to the reference tone in a range from 0 to 110 percent. Both tones were added in the time domain and the harmonicity index was measured for different chords along the fundamental frequency and shift dimensions.

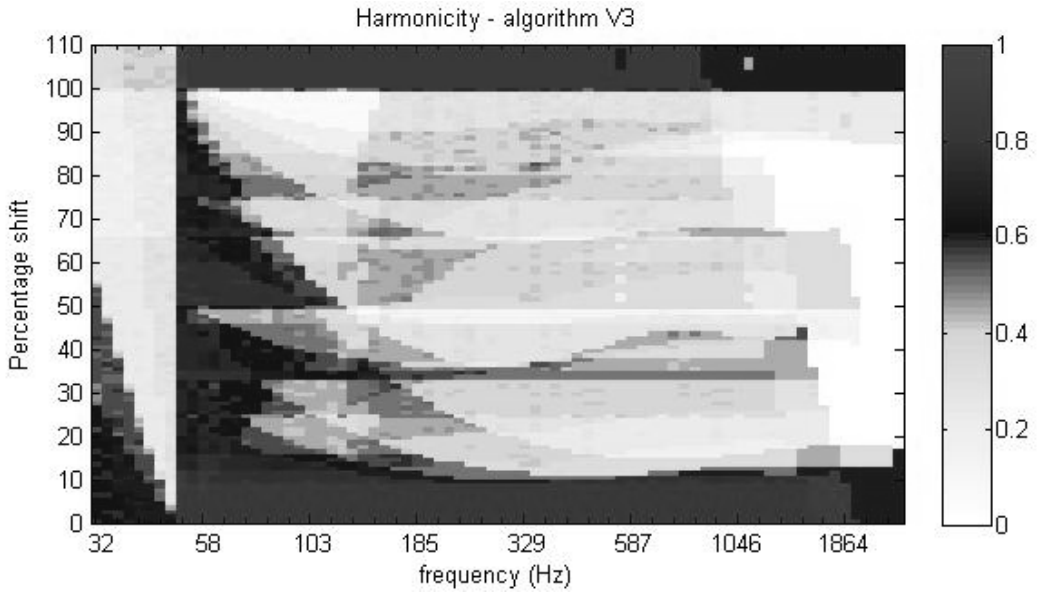


Figure (3.33): Illustrating harmonicity index V3 for chords composed of two tones with fundamental frequency from 32 to 2217 Hz with one semitone interval step for different percentage shifts (from 0 to 110 %) from the fundamental of the second tone. Values are represented in a colored scale on the right side (red for highly harmonic, blue for moderate harmonic, green for less harmonic and white for inharmonic).

Figure (3.33) shows that tones with fundamentals from 55 to 1864 Hz with a zero shift produce highly harmonic values (dark) which is expected because similar tones are perfectly harmonic. Increasing the shift upto 9 % in the same range does not reduce the index a lot because the algorithm uses one Bark tolerance. For frequencies below 55 Hz, the index does not approach unity value because algorithm V3 uses an approximation that relatively increases with the decrease in the fundamental frequency and at frequencies less than 55 Hz, the approximation error is relatively high. A consequence is that the algorithm may not function properly at frequencies below 55 Hz. The figure illustrates some dark zones (e.g. starting at 26%, 34% and 50% shifts) and this corresponds to the lobes in Figure (3.32 - left) and by increasing the shift to 100 % (one octave higher), the index starts approaching unity.

Figures (3.32 and 3.33) are promising examples illustrating the ability of the harmonicity index V3 for musical tones. Additional analysis of musical instruments processed with Std, Smt-LF and Smt-MF mappings using the AMO are shown in Figures (3.36-3.37).

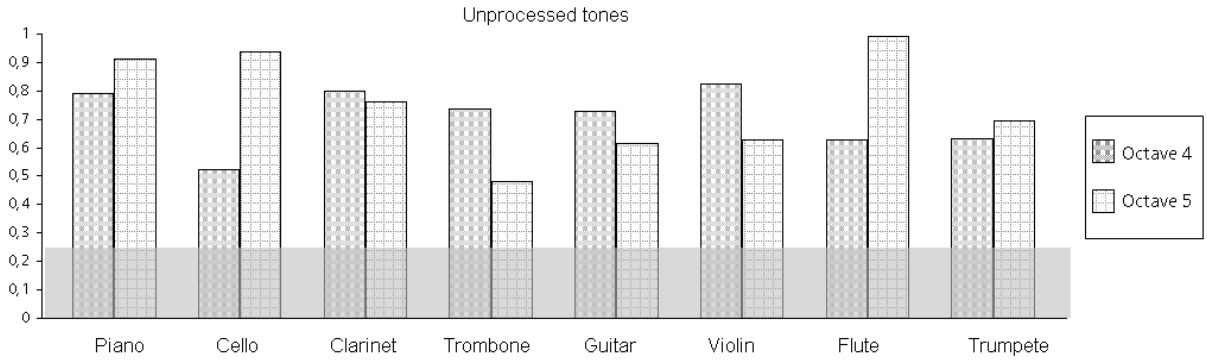


Figure (3.34): Harmonicity measured for tones played by different musical instruments (Piano, Cello, Clarinet, Trombone, Guitar, Violin, Flute and Trumpet) in octaves 4 (dark bricks) and 5 (light bricks). Shaded area is an inharmonic area; its threshold was determined using a white Gaussian noise signal.

Figure (3.34) displays the harmonicity index using the V3 algorithm for tones played by 8 different musical instruments in octave 4 and octave 5. The shaded area represents the noise level (inharmonic) whose threshold was measured with a white Gaussian noise signal. All instruments produced relatively high values taking into consideration that a high harmonicity index is in the range from 0.7-1 and the noise level is from 0-0.2 as illustrated by the red and white zones in Figure (3.33) respectively.

A T-test analysis showed that there was no significant difference in the unprocessed tone between results from octave 4 and octave 5 ($P<0.05$). Thus the two octaves were combined in a wider set, and the results from the two octaves were averaged as shown in Figure (3.35). However, there may be significant differences between octaves for the processed signals.

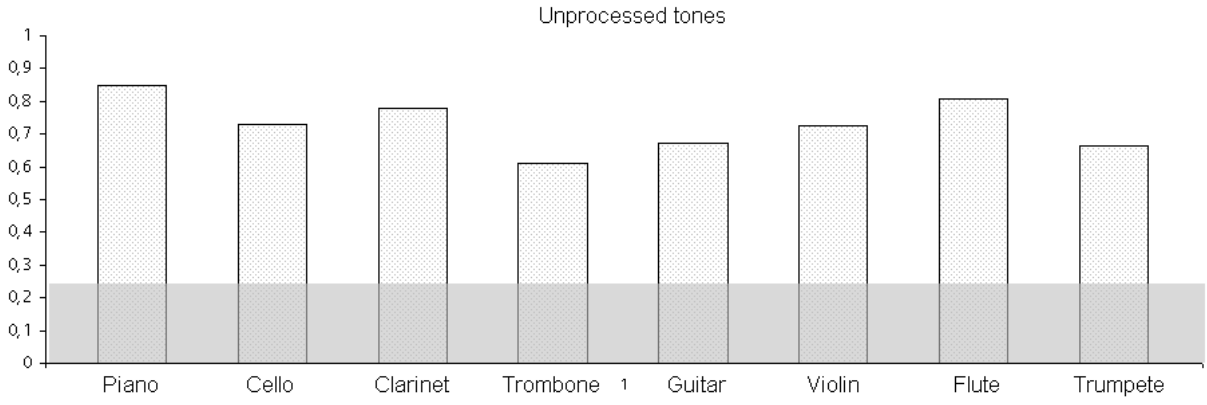


Figure (3.35): Average of the harmonicity index for tones in octaves 4 and 5 from different musical instruments (Piano, Cello, Clarinet, Trombone, Guitar, Violin, Flute and Trumpet). Shaded area is an inharmonic area; its threshold was measured a white Gaussian noise signal.

These tones were processed with Std, Smt-LF and Smt-MF mappings and resynthesized using sinusoidal and noise band vocoders with the AMO. Noise band vocoders are set to use a 0.5 and 1 mm SOE and the results are shown in Figures (3.36-3.37) illustrating that :-

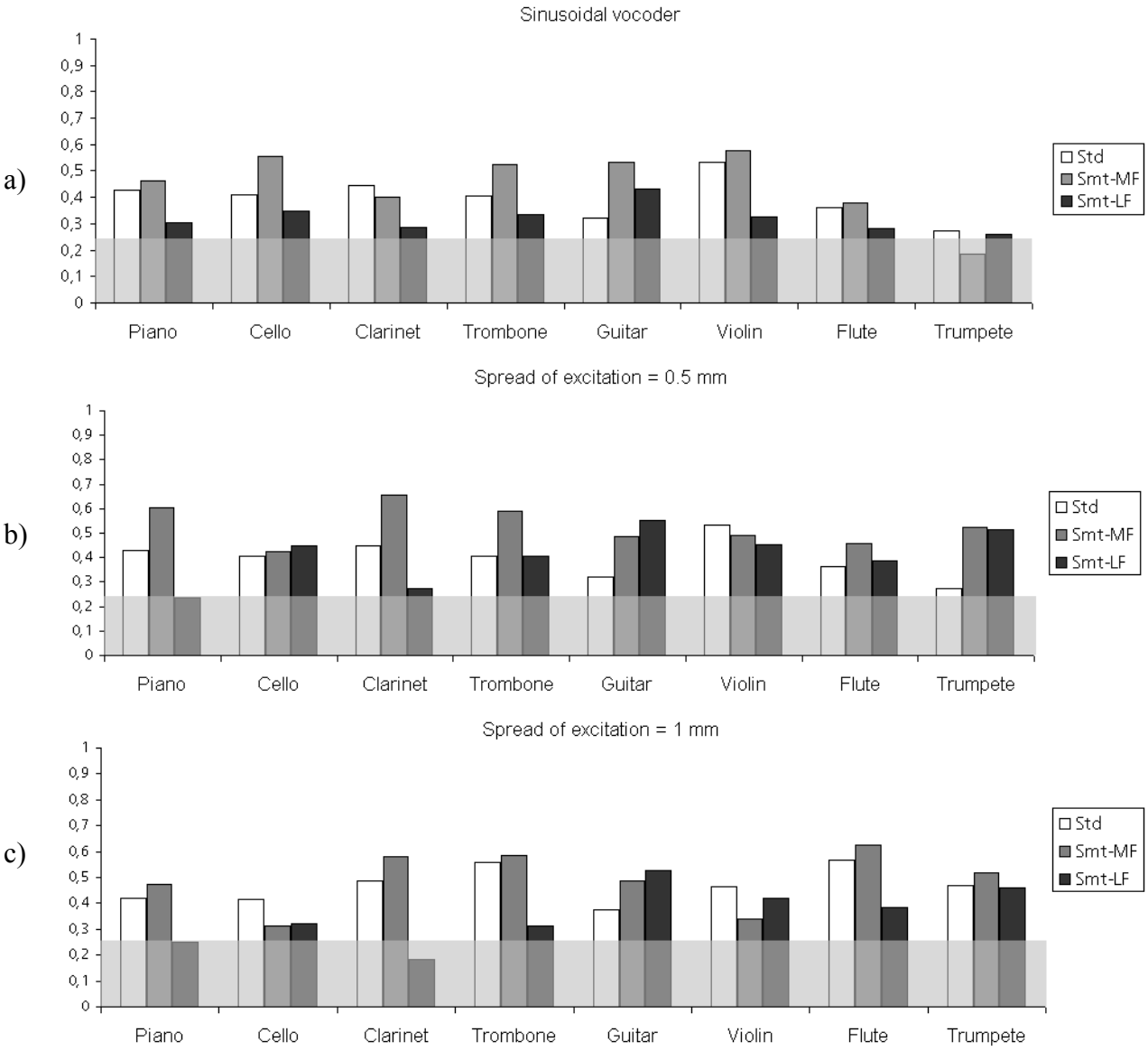


Figure (3.36): Average of the harmonicity index for tones in octaves 4 and 5 from different musical instruments (Piano, Cello, Clarinet, Trombone, Guitar, Violin, Flute and Trumpet) processed with standard and semitone mapping (Smt-LF and Smt-MF) and resynthesized using (a) sinusoidal vocoder, (b) noise vocoder with spread of excitation = 0.5 mm and (c) noise vocoder with spread of excitation = 0.5 mm. Shaded area represents the noise level (inharmonic area); its threshold was determined using an unprocessed white Gaussian noise signal.

(a) Although musical tones were processed and resynthesized with sinusoidal tones the harmonicity index was generally between 0.3 and 0.4. This illustrates a reduction after being processed with all mappings. However the harmonicity index with Smt-MF was higher than with Std mapping except for the Trumpet tone. The Trumpet instrument produces vibrato acoustic signals together with the fundamental and the harmonic overtones. The frequency intensity of the vibrato is less than the harmonic overtones but it is relatively high. This causes the Smt-MF algorithm to pick many peaks from the vibrato as well. A consequence is that the resynthesized sound has a mix of overtones from the harmonic structure superimposed with overtones from the vibrato and in general the

overtone structure of the resynthesized signal is less harmonic than the white Gaussian noise.

(b) Using a SOE = 0.5 mm, the harmonicity index with Smt-MF generally increases over that using Std; the increase varies from being relatively small to higher differences, except with Violin where there is a relative small decrease. Using a noise vocoder together with a SOE may increases the likelihood of signals to be harmonic, since algorithm-V3 searches for harmonic overtones using a tolerance of ± 0.5 Bark.

(c) In general, increasing SOE from 0.5 to 1 mm does not change the relative values of Std compared to Smt-LF and Smt-MF except with Cello, Clarinet, Violin and Trumpet. However such a relative change may be more significant in Smt-LF and Smt-MF values than in Std. One reason may be the transposition which both Smt-MF and Smt-LF cause in the signal, since both of them use certain bandwidths (440-5009 Hz and 130-1550 Hz respectively) that are mapped to tonotopical frequencies upto 8 kHz. A T-test showed no statistically significant effect of SOE on the harmonic index result compared to the sinusoidal vocoder from the same strategy, except with Smt-LF using SOE = 0.5 mm.

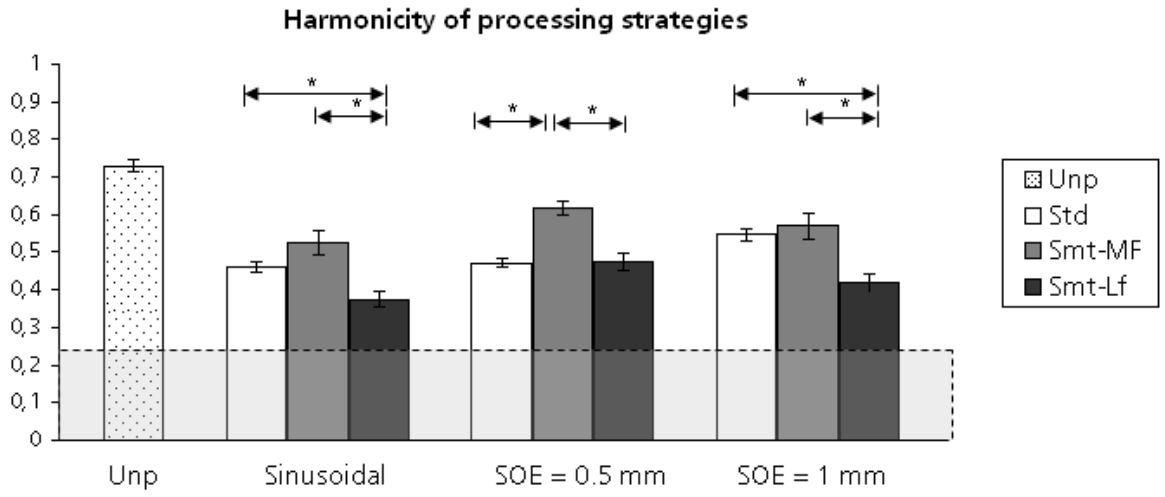


Figure (3.37): Average harmonicity index of tones from eight musical instruments using sinusoidal and noise vocoder with SOE = 0.5 and 1 mm for different mappings (Std, Smt-MF and Smt-LF) compared to unprocessed tones.

Processing musical tones by the three mappings produces a harmonicity index that is higher with the unprocessed condition than processed tones as shown in Figure (3.37). It illustrates also that the harmonicity index with Smt-MF is significantly different from the Smt-LF for all conditions (sinusoidal, SOE = 0.5 and SOE = 1 mm) and is significantly different from the Std with the SOE = 0.5 condition. In addition, the index for Smt-LF is significantly lower than for the Std condition with both sinusoidal and SOE = 1 mm conditions. The harmonicity index for Smt-MF on the other hand is significantly higher than for Smt-LF. In general, it is also higher than for the Std condition with SOE = 0.5 mm.

Chapter 4 – Psychoacoustic Tests and Results

Chapter 4 – Psychoacoustic Tests and Results

4.1 Introduction

Many post lingually deafened CI patients report that music is not well perceived. Music consists of complex acoustic sounds composed of tones with harmonically related overtones. Most musical instruments generate fundamental frequencies below 1kHz (Pierce 1983) which points to the importance of preserving low frequency sound components for music perception. In Chapter 2, two Smt frequency mappings were proposed to improve melody representation with CI patients (Omran et al. 2008). Smt mapping essentially involves assigning the fundamental frequencies of adjacent tones on the musical scale to corresponding adjacent electrodes or channels. This also requires that the frequency to electrode/channel mapping is based on a semitone scale. The idea was initially investigated in a study by (Kasturi and Loizou 2007), using the 12 electrode Clarion CII (Advance Bionics) implant with a limited range of semitone frequencies. The Smt mappings investigated in this study, Smt-LF and Smt-MF, cover the frequency ranges from 130 to 1502 Hz and from 440 to 5009 Hz respectively. Smt mappings preserve the representation of harmonic structure of musical tones for the CI. This may help to improve music perception.

Psychoacoustic tests can be carried out to evaluate these various dimensions of music perception such as pitch, melody and timbre. Frequency representation, loudness and temporal resolution are important aspects that affect music perception. To examine music perception with Smt mapping in this study; three psychoacoustic tests (pitch discrimination, MCI and IR) were conducted with three processing conditions (Std ACE, Smt-LF and Smt-MF mappings). Pitch discrimination and MCI tests were carried out with NH subjects listening to noise band vocoded representations of the test sounds while MCI and IR tests were carried out with CI recipients. The pitch discrimination test was conducted in two steps; in the first step it was tested using Std mapping with and without VCs an in the second step with pure and complex tones using Std, Smt-LF and Smt-MF mappings.

An improved representation of the harmonic structure through Smt mapping is expected to also yield better resolution of individual tones on the musical scale, particularly towards the lower frequencies. The pitch discrimination experiment was employed to explore the VC effect and to find out whether Smt mapping produces the expected improvement in resolution over Std mapping. The test involved processed clarinet tones in the 1st step, and in the 2nd step synthetic complex tones with a harmonic structure, similar to musical tones, rather than signals that only excite single electrodes. The later test was mainly intended to check whether Smt mapping is viable, and it was decided that conducting these tests with NH subjects only would help expedite the testing. Testing with NH subjects requires that the processed signals of Std or either Smt mappings, originally meant for presentation to CI recipients, be made audible. This was achieved by additional processing of these CI signals with the AMO (see section 3.1) which resynthesizes and simulates the sound of a CI (Laneau et al. 2006a). The AMO outputs are then presented to NH subjects.

Melody is an important aspect of music (Sadie and Grove 1995) which can be described as a group of tones perceived as a single entity (Terhardt 1998). Each tone has a harmonic structure of overtones and preserving this structure (as with Smt mapping) is expected to improve melody perception. The Pitch discrimination tests involve only single tones and yield little direct information about melody perception. A more complex task that would reflect

melody perception necessarily involves a sequence of tones. Galvin et al. (Galvin et al. 2007) provides a very good overview of the shortcomings of many existing tests that attempt to measure melody perception. The MCI test which they developed is suitable for this study. The MCI test was carried out with the three mapping conditions, first with NH subjects and then with CI recipients.

Timbre (tone color) is another aspect of music, by which different instruments are characterized (Helmholtz 1954). Timbre depends on the relationship between intensities of different partials as well as the presentation of the temporal FS. In the IR test, sounds from different musical instruments encoded using the different mappings were presented to subjects. As the mappings in this study do not explicitly present any fine temporal structure information, this test investigates whether the expected improvement in the representation of the harmonic structure using Smt mapping would be beneficial for timbre recognition. This test was only conducted with CI recipients.

4.2 Pitch Discrimination

4.2.1 Estimating Virtual Channels (Pilot test)

Nucleus CI devices provide 22 stimulation channels that stimulate 22 intracochlear positions. Increasing the number of stimulation channels is expected to improve perception of musical notes. One way to increase the number of channels on the currently available electrode arrays is to use VCs formed by stimulating two adjacent electrodes simultaneously. It was shown in (Busby and Plant 2005) that VC stimulation can result in the perception of an intermediate pitch sensation between adjacent electrodes. One approach to study the effect of VC on music perception is to examine tone discrimination for 43 and 22 Channels with complex tones composed of several overtones and pure tones without overtones.

Hypothesis

43 Channel mode increases frequency representation. This may help in musical note discrimination, leading to better discrimination of smaller semitone distances than with 22 channels.

Materials and Method

A CI AMO is used to simulate auditory percept of CI patients for NH subjects. The AMO assumes that there is no change in effective stimulation width between 43 and 22 channels.

Test tones were initially taken from the RWC Music Database (Goto 2004): Instrument no. 31 (clarinet), variation 1, normal articulation and mezzo dynamics were chosen for this experiment. Initial analysis showed that the temporal pattern of the partials in the clarinet tones varies a lot from certain tones to others, providing additional cues which the test subjects might use instead of the pitch to identify the different tones. To minimize these additional cues, all tones were altered to have the same envelope, and duration of 0.5 s. The starting and ending of all tones were faded with 30 ms attack and release times.

Sound samples were prepared using clarinet and pure tones. Pure tones were used to determine the effect of fundamentals. Three tones (D#, F and G#) 1, 3 and 6 semitones higher

than the reference tone (D) from octaves 3, 4 and 5 were processed using the AMO, whose resynthesis consisted of the superposition of modulated narrowband noise signals with bin to channel allocation used in CI processors. A pair of processed notes (1 test and 1 reference) was then presented and the subject was asked to state which one was higher in pitch. The test consisted of presenting a total of 9 pairs of sounds, each pair being repeated 8 times and randomized. The test was repeated 3 times for each subject. 4 NH subjects took part in this pilot test.

All tones were processed using 22 and 43 channels with the acoustic model using a SOE of 1 mm as recommended in (Laneau and Wouters 2004; Laneau et al. 2006a). Sound samples were then normalized to have equal loudness. NH subjects were seated in front of a loudspeaker at a distance of 1.5 m. Sounds were presented at a level of 70 dBA. MACarena (Lai and Dillier 2002) software was used to play a set of two notes that were randomly chosen from the 3 octave groups. The group was also randomly selected to minimize learning effects of notes sequence. For each group the same number of repetitions was presented. The tone pairs were presented sequentially with a pause in between of 0.5 s. The same reference note (D) was used for different groups. Levels were roved ± 6 dB to avoid loudness cues from being used.

Results

Experiment 1 used pure tones processed with the AMO. Four subjects participated in the experiment. A test of variance (ANOVA) showed that there was no statistically significant difference between 22 and 43 channels (P>0.05).

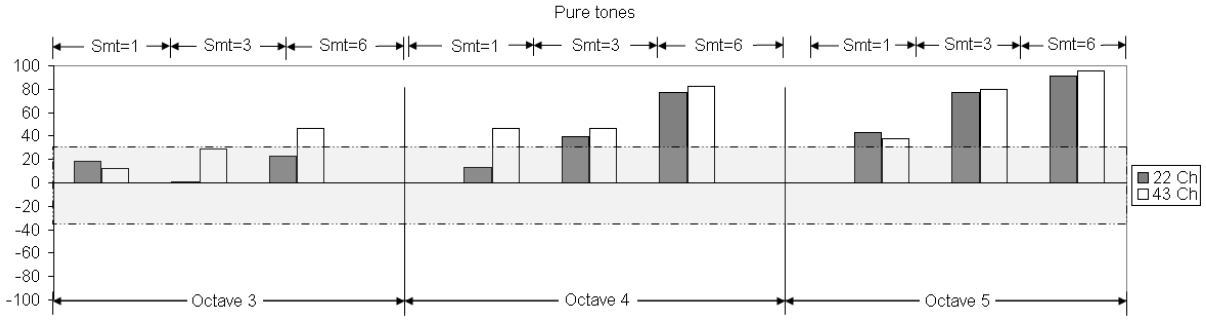


Figure (4.1): Results of 43 (gray with dot) and 22 channels (black with dot) for 1, 3 and 6 semitones intervals in octaves 3, 4 and 5 for pure tones. Shaded area is a chance level.

The results showed that all subjects showed the expected trend in pitch discrimination performance: 6 semitone differences are easier than 3 and easier than 1 semitone.

Experiment 2 used complex tones processed with the AMO. Four NH subjects participated in the experiment. A test of variance (ANOVA) showed that there is no statistically difference between 22 and 43 channels with a significance level of 95%.

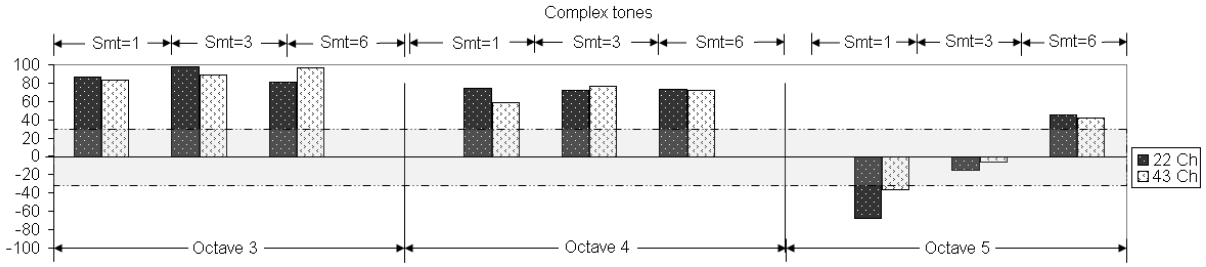


Figure (4.2): Results of 43 (gray with dot) and 22 channels (black with dot) for 1, 3 and 6 semitones intervals in octaves 3, 4 and 5 for complex tones. Shaded area is a chance level.

The results showed that:

a) 1-Semitone differences for lower octaves are easier to discriminate.
b) Semitone discrimination for pure tones was more difficult than for complex tones in the lower range.
c) Different notes have different pitch and timbre which sometimes may mislead the subject.
d) 2-way ANOVA (mode and subject) showed that the results are subject dependent at the 95% significance level for all modes.

4.2.2 Estimating Different Mappings

This pitch discrimination test was intended to examine whether the Smt mappings would produce better resolution of the pitch scale for the complex tones than the Std mapping. NH subjects listened to sound representations processed by the AMO with Std, Smt-MF and Smt-LF mappings. Enhancing pitch discrimination may ameliorate music perception (Baumann and Nobbe 2004).

Hypothesis

The discriminability of two complex tones separated by only a few semitones will improve with Smt mapping compared to Std mapping due to better preservation of the harmonic structure.

Method

The signals used for the test were synthetic complex tones which had the same fundamental frequencies as the corresponding musical tones. Each tone had four harmonic overtones with successive 20% decrease in amplitude. To avoid temporal cues, all tones were designed to have the same temporal envelope, namely duration of 500 ms including 30 ms fading in/out at the beginning and the end respectively. The RMS energy of the signals (in digital form: WAV file format) was set to –15 dB to prevent clipping, where 0 dB corresponded to the RMS signal energy of the maximum peak-to-peak waveform.

Subjects were presented with pairs of synthetic complex tones processed by the AMO and were asked to indicate the one higher in pitch. Each presentation consisted of a probe tone and a reference tone. The fundamental frequency of the probe was higher than that of the reference by 1, 3 or 6 semitones. Two reference tones D and G# in octaves 3, 4 and 5 were used and the full set of tone pairs tested is summarized in Table (4.1).

		Groups					
Semitone Intervals	1	D3,D3#	D4,D4#	D5,D5#	G3#,A3	G4#,A4	G5#,A5
	3	D3,F3	D4,F4	D5,F5	G3#,B3	G4#,B4	G5#,B5
	6	D3,G3#	D4,G4#	D5,G5#	G3#,D4	G4#,D5	G5#,D6

Table (4.1): The signals used in each presentation can be separated into three groups with different interval sizes of 1, 3 or 6. Each group in turn consists then of 6 tone pairs with two references D and G# in octaves 3, 4 and 5.

The above signals were processed by the AMO with the Std, Smt-MF and Smt-LF mappings before being presented via loudspeaker to the NH subjects. For this test, the AMO was set to simulate CI stimuli that had a SOE (stimulation width) of 1 mm as recommended in (Laneau and Wouters 2004; Laneau et al. 2006a). The AMO also incorporated VCs, produced by stimulating two adjacent electrodes simultaneously.

In each presentation, the reference and probe tones were presented in a random order, separated by a silent interval of 500 ms between each tone. A single test session involved presenting randomly each of the 18 tone pairs, summarized in Table (4.1), with 4 repetitions. The tone pairs were presented from a calibrated loudspeaker (Genelec 1029A) at 65 dB(A) located 1.5 m in front of the subject. The presentation level was roved by ± 6 dB to minimize the effects of loudness cues on the pitch discrimination task.

Initially, the original unprocessed tones were presented and tested to familiarize the subjects with the task. For this condition, the test was conducted once, i.e. each tone pair was repeated 4 times. Testing the unprocessed tones also served to establish that the test material was not too difficult to begin with. Thereafter, testing proceeded with the AMO outputs for the Std, Smt-MF and Smt-LF mappings. The order of testing of the three mappings was randomized. For each mapping condition, a training session with correct/wrong feedback was first carried out. 2 test sessions without feedback were then carried out, and the results from these 2 sessions were collected for the final results. Thus, the results consisted of a total of 8 presentations of each tone pair for each subject. A total of 8 NH subjects were evaluated for this test. A custom test software (MACarena) (Lai and Dillier 2002) was used to playback sound files and record the responses.

Results

The results from the familiarization test are shown in Figure (4.3), and confirm that the unprocessed tone pairs are generally easy to rank correctly, yielding scores that are significantly above chance. As expected, the scores also tended to be lower with smaller interval sizes.

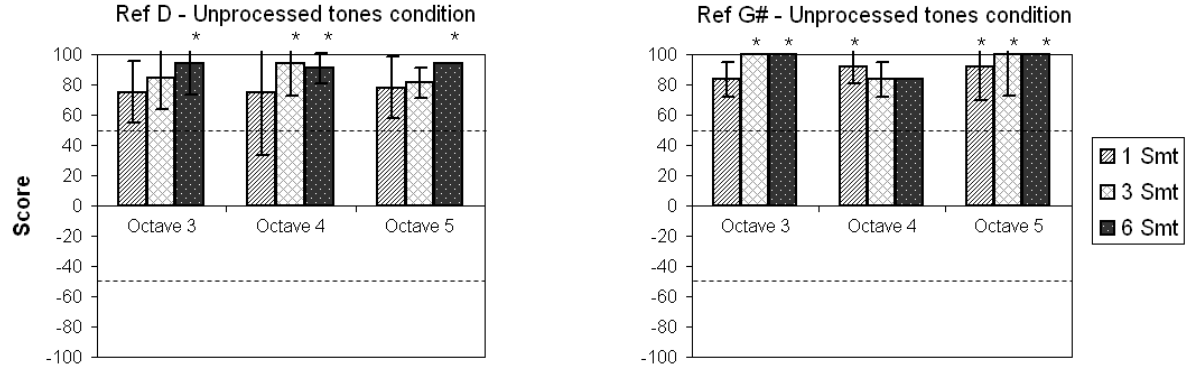


Figure (4.3): Mean results for unprocessed tones with both references D (left) and G# (right) in octaves 3, 4 and 5 with 1, 3 and 6 semitone intervals between the probe and reference tones. Pitch reversals, which would be indicated by negative scores, were not observed at all. Columns marked with an asterisk are significantly above chance (p = 0.05) according to the cumulative binomial distribution of mutually exclusive events; at least 7/8 correct answers are considered significant. Chance level is indicated by the dashed line.

The results with sounds processed by the AMO for the Std, Smt-MF and Smt-LF mappings are summarized in Figure (4.4). Scores in this test were calculated in percentage from 0-100%, biased to -50% and normalized to be between ± 100. The negative side indicates pitch reversals and -100% is complete pitch reversal. With the Std mapping (white filled bars), pitch discrimination of tone pairs separated by larger intervals was easier than of tone pairs with smaller intervals (e.g. 6-semitones interval is easier than 3 and 1 semitone interval). The score with 1-semitone interval in octave 3 was close to chance level with reference D, but was higher with reference G#. This could be due to the Std mapping compressing the input frequency range, especially towards the lower frequencies. As a result, the partials of tones at the lower end of the musical scale are more likely to be compressed than those higher up on the musical scale. This would cause tone pairs close to one another to be more difficult to resolve.

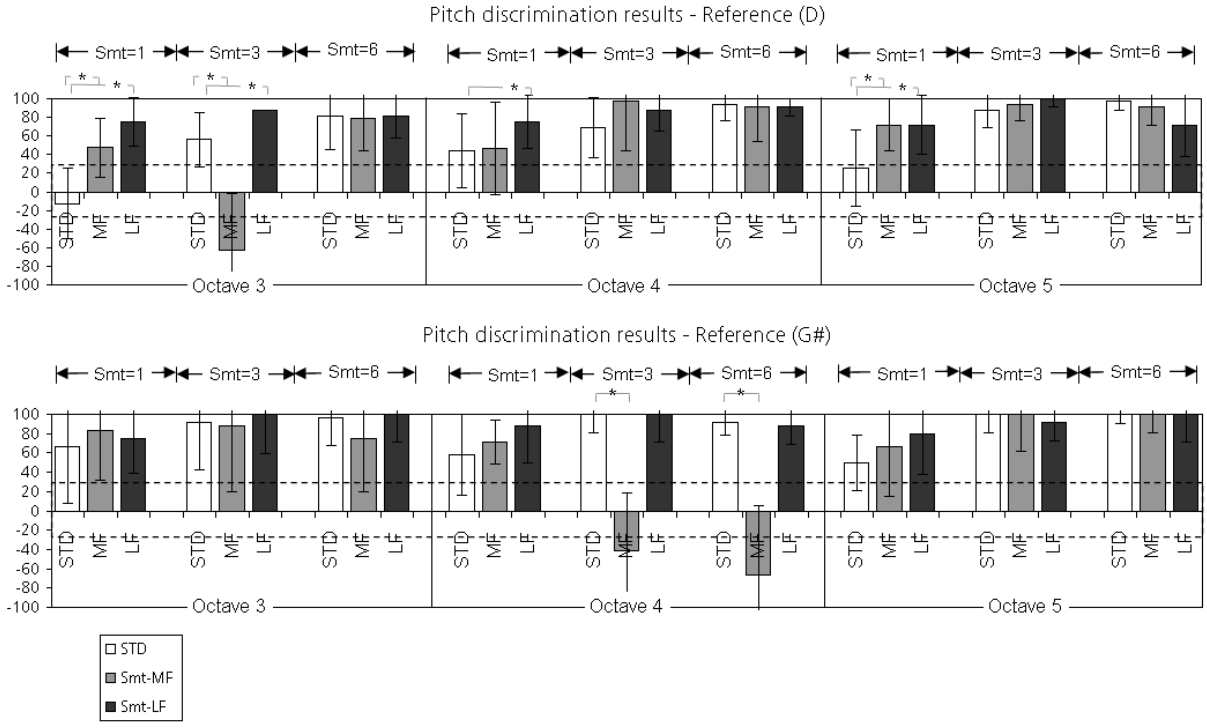


Figure (4.4): Showing results with Std mapping (white), semitone mapping Smt-MF (gray) and semitone mapping Smt-LF (black) with reference tones D (upper graph) and G# (lower

graph) using semitone intervals (1, 3 and 6) in octaves range from 3 to 5. Chance level is indicated by the dashed line. An asterisk between two columns indicates that the corresponding scores are significantly different ($p = 0.05$) from one another (t-test). When pitch reversals occur, which are indicated by negative scores, the significance test was calculated using the absolute values.

Figure (4.4) also shows the results with Smt-MF (gray bars) and Smt-LF (black bars) mappings. Smt-LF generally performs significantly better in Octaves 3 and 4 than Smt-MF and Std, particularly with Reference D and smaller intervals. Smt-MF, apart from the pitch reversals observed, also performs better than Std, especially at small (1-semitone) intervals (Octave 3 and 5 with Reference D). With Reference G#, notwithstanding the pitch reversals with Smt-MF, there were no significant differences observed between the three mappings. The pitch reversals with Smt-MF are most likely due to filtering out of partials below 440 Hz. The reference G4# (415 Hz) will have its fundamental filtered out, leaving the first harmonic overtone as its lowest tone. Notice that there is no evidence that CI recipients can perceive missing fundamentals (Oxenham 2008), which may be due to the SOE at electrodes. This can lead to pitch reversals when the probe tone has an unfiltered fundamental that has a lower frequency than G4#'s first harmonic. In octave 3, the reference tone G3# (207 Hz) and the probe tones all have their fundamental filtered out, and pitch discrimination can apparently still be reliably carried out with the remaining unfiltered overtones.

Smt-LF also appears to perform better than Smt-MF. One possible reason for this could be that it preserves LF components, transposing them into a higher perceptual range, whereas Smt-MF tends to cut-off frequencies below 440 Hz (A4) and therefore has poorer representation of the partials of tones, particularly in the lower octaves. Note that the frequency transposition that occurs with Smt-LF tends to also make the sounds unnaturally higher in pitch than with Smt-MF, which has a frequency mapping which is closer to the natural tonotopic characteristic frequency. In general, the pitch discrimination is improved with Smt mapping compared to Std mapping.

4.3 Melody Contour Identification

Melody is one of the important aspects of music (Sadie and Grove 1995). It is described as a group of tones that are perceived as a single entity (Terhardt 1998). Each tone has a harmonic structure of overtones and preserving this harmonic structure may improve melody identification. One way to improve melody representation would be to ensure that the fundamental frequencies of individual tones on the musical scale are assigned to separate electrodes. Such an approach involves mapping fundamental frequencies of musical tones to electrodes based on a semitone scale. In this study the two Smt mapping ranges (see section 2.3) were investigated. As described before in Chapter 2, the first mapping, Smt-LF, is in the low and mid frequency range [130 – 1502 Hz] using a buffer of 512 points which is zero padded before undergoing a 2048-point FFT. Smt-LF yields a resolution of 7.8 Hz for frequencies below 1054 Hz, and 31.25 Hz for higher frequencies. The second mapping, Smt-MF, is in the mid and HF range [440 – 5009 Hz] and involves a 512-point, giving a resolution of 31.25 Hz. The Std mapping uses 128-point FFT with a resolution of 125 Hz. All three mappings use overlapping data buffers, the amount of overlap depending on the stimulation rate such that at the end of each stimulation period, as much new data (sampled at 16 kHz) as possible is added to the data buffer.

4.3.1 Hypothesis

Smt mapping will yield higher MCI scores than Std mapping. Ambiguities may occur with Smt-MF mapping at low frequencies due to filtering out partials below 440 Hz and the performance may decrease with Smt-LF mapping because frequencies are transposed to higher ranges.

4.3.2Method

The MCI was originally designed and proposed by (Galvin et al. 2007). Subjects are presented with a sequence of tones and have to identify the corresponding contour pattern. For each contour pattern, the lowest note is the root note, which is kept the same for all nine patterns (rise, rise-flat, rise-fall, flat-rise, flat, flat-fall, fall-rise, fall-flat, fall) as shown in Figure (4.5).

In this study, each pattern consists of a sequence of five synthetic complex tones. Each tone in turn consisted of five harmonic partials. The fundamental frequency of each synthetic complex tone was the same as its corresponding musical tone. The amplitude of each partial was reduced successively by 20 % compared to the previous one. To avoid envelope cues, all tones were designed to have similar temporal envelope structure and the RMS energy of each pattern was normalized to –15 dB, where 0 dB corresponded to the RMS signal energy of the waveform with maximum amplitude. However, there may still be periodicity cues in the temporal domain. Each tone in the pattern had a duration of 250 ms with a 50 ms pause in between tones. Tones were faded in/out with a 10 ms Hanning window at the beginning and the end respectively. A root note of "A" was used for all the contour patterns, the same as was used by (Galvin et al. 2007).

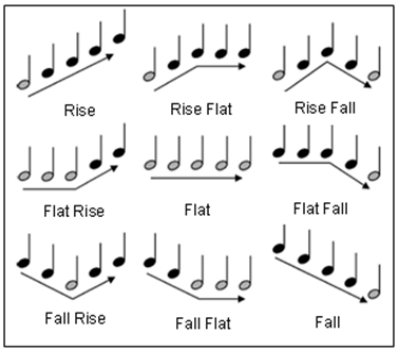


Figure (4.5): The nine different melody contour patterns used in the MCI test with NH subjects. The root notes are indicated with gray filling.

The MCI test was carried out firstly with NH subjects. The interval size was varied between 1 to 5 semitones in octave 3, between 1 to 3 semitones in octave 4 and between 1 to 2 semitones in octave 5, as summarized in Table (4.2).

	NH Subjects					CI Recipients		
Interval size	1	2	3	4	5	1	2	3
Octave 3	x	x	x	x	x	x	x	x
Octave 4	x	x	x			x	x	x
Octave 5	x	x						

Table (4.2): Summary of the semitone interval sizes between successive tones in the contour patterns as well as the octave ranges that were investigated for NH subjects and CI recipients.

For NH subjects, the different patterns were processed by the AMO with the Std, Smt-LF and Smt-MF mappings using a 1 mm stimulation width and 22 channels. The patterns were presented at a level of 65 dB(A) at a distance of 1.5 m in front of a calibrated loudspeaker (Genelec 1029A). Test subjects responded via a touch screen by indicating the corresponding button containing a sketch of the corresponding MCI pattern as shown in Figure (4.5). At the start of a test, the subjects were allowed to first familiarize themselves with the MCI contours in a condition expected to be easy: for instance, octave 4 with 3 semitone intervals. In this testing phase, pressing a button on the touch screen would present the corresponding sound over the loudspeaker. After they had heard each pattern at least once, a training session with correct/wrong response feedback was conducted. A single test session involved presenting each of the 9 contour patterns with each of the 10 interval-size/octave conditions once. After 1 training session (with feedback), 2 test sessions (without feedback) were conducted. A total of 8 NH subjects were evaluated for this part of the MCI test.

The nine patterns designed by Galvin et al. (2007) were utilized to test the NH subjects. However, the large number of response choices proved to be too demanding for some CI recipients in initial testing. Therefore, in order to simplify the test, only five patterns were subsequently utilized to test CI recipients as shown in Figure (4.6). For the CI recipients, octave 3 and 4 with interval size from 1 to 3 semitones were tested. Testing in Octave 5 was eliminated (see Figure (4.6)). This reduction of test conditions was due to analysis of NH which showed that tones with one part being flat are likely to be misperceived with Smt mapping in cases when the fundamental is filtered. To simplify the test with CI subjects, all such tones were eliminated. Conditions with one-semitone intervals were processed with 22 channels and represent effectively a resolution of two semitones. Another pitch discrimination study with NH using 22 and 43 channels showed no significant differences. Therefore it is assumed that results from CI recipients with 22 channels are similar to those with 43 channels. Testing was done using the MACarena (Lai and Dillier 2002) software which allowed randomized sound presentation and automatic recording of subjects' responses.

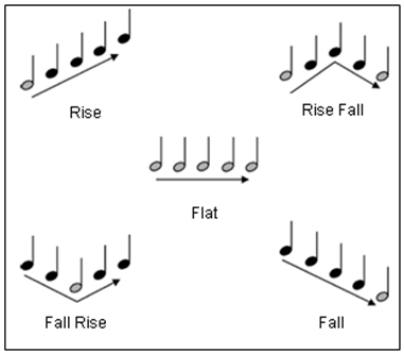


Figure (4.6): The five different melody contour patterns used in the MCI test with CI recipients. The root notes are indicated with gray filling.

Testing with CI recipients involved stimuli being streamed directly to the implant using the NIC research software from Cochlear Corporation (Swanson 2008). Stimuli were first prepared offline using a custom Matlab utility called "*Checker*" (see section 3.2) in which the Std, Smt-LF and Smt-MF mappings were implemented. The Std mapping is the default implementation in the NMT from Cochlear Corporation, whereas the Smt-LF and Smt-MF mappings are custom implementations. First the latest speech processor map for each CI recipient was loaded from a clinical database. The WAV files for the different MCI patterns were then loaded and processed for all the three mappings. For this test, the "*Checker*" program was set for 22-channel output. Testing 43-channels with CIs was omitted due to technical constraints and time limitations. The resulting output was verified to ensure that the

stimuli were calibrated to correspond to an equivalent acoustic level of 65 dB(A). The resulting output was a sequence of parameters that when streamed to the CI, would produce a corresponding sequence of stimulation. To meet safety requirements, the entire output sequence was verified to ensure that none of the parameters exceeded the limits set by the corresponding CI recipient's individual speech processor settings. Once the sequences were verified, the "*Checker*" program stored them offline as XML files. During a test, the corresponding XML files for the selected CI recipient were streamed to the L34 speech processor. The MACarena test software had been enhanced with an additional output option which allowed direct streaming of CI stimulation sequences from XML files via the L34 speech processor. As with the NH subjects, a test was started with a CI recipient familiarization session using the MCI signals in a higher octave (octave 4) and large interval size (e.g. octave 4 with 3-semitone intervals) for the three mappings used. This was then followed by a training session with correct/wrong response feedback using test signals. A single test session involved presenting each of the 5 contour patterns with each of the 6 interval-size/octave conditions twice. After one training session (with feedback), two test sessions (without feedback) were conducted. A total of 8 CI recipients were evaluated. All subjects had at least 1 year's experience using a CI device. All of them used the Nucleus CI and the Std mapping.

4.3.3 Results

In the MCI test, different contour patterns were presented to NH subjects and CI recipients. The mean correct identification scores of the MCI test was evaluated for different octaves and different semitone intervals using Std, Smt-MF and Smt-LF mappings.

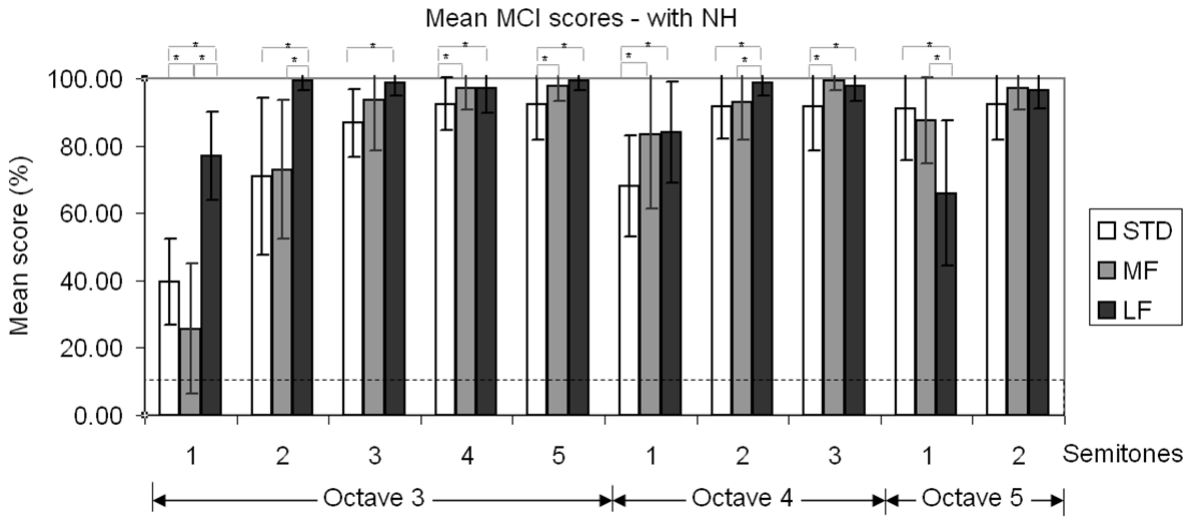


Figure (4.7): Results with standard mapping (white), semitone mapping Smt-MF (gray) and semitone mapping Smt-LF (black) for NH subjects with AMO output. Three octave ranges (3, 4 and 5) were tested with different semitone intervals. Chance level is indicated by the dashed line. An asterisk between two columns indicates that the corresponding scores are significantly different ($p = 0.05$) from one another.

The results for NH subjects listening to the AMO outputs are summarized in Figure (4.7) and generally showed that the MCI scores improve with increasing interval size. With Smt-MF mapping, the scores were significantly better than those with Std mapping in octave 3 with 4 and 5 semitone intervals, as well as in octave 4 with 1 and 3 semitone intervals. In octave 3 with 1-semitone intervals a significant decrease was found, most probably due to Smt-MF

filtering out partials below 440 Hz, which can result in pitch reversals with the Smt-MF mapping at low frequencies due to strong confusion between rise-fall, fall-rise, fall-flat and flat-fall in octave 3.

Smt-LF mapping generally yielded significant improvements over Std mapping, with the exception that a significant decrease in the recognition score was found at octave 5 with 1 semitone interval. For tones in octave 5, Smt-LF filters out all overtones above 1502 Hz, leaving only the fundamental in the melody contours. With only a single component which is at the same time spread out over several adjacent critical bands, the melody contour patterns with 1 semitone intervals become difficult to resolve, as illustrated in Figure (4.8).

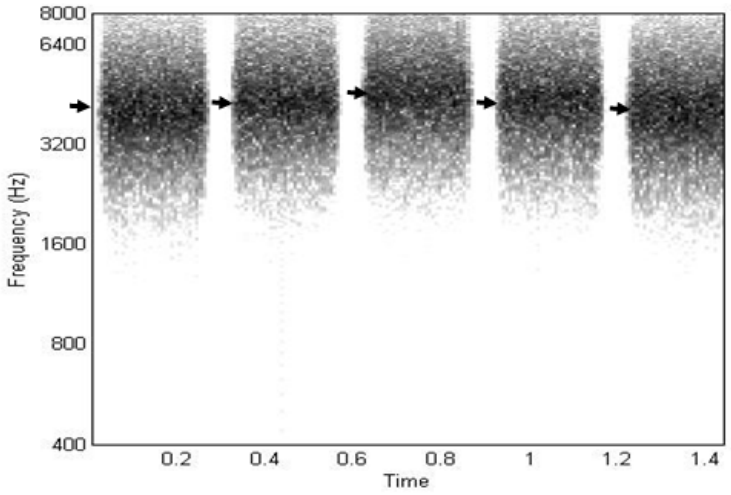


Figure (4.8): Spectrogram of the AMO output for the MCI rise-fall pattern in octave 5 with 1-semitone intervals and fundamental frequency of the root note equal to 880 Hz, processed with Smt-LF mapping. Only the fundamental frequencies are left after Smt-LF has filtered out partials above 1502 Hz. The Smt-LF output is then resynthesized in the AMO using the tonotopical frequencies at the corresponding electrode positions, which results in a transposition of the center activity to around 4000 Hz.

There was also a significant difference between Smt-LF and Smt-MF in octaves 3 and 4 with two semitone intervals.

The inability or failure to resolve a melody contour is indicated by “flat” responses when the presented contour was not “flat”. Figure (4.9) shows the mean number of occurrences of such failures to resolve melody contours. Std mapping generally yields significantly more failures at octave 3 with 1 semitone intervals compared to either Smt-MF of Smt-LF, which is consistent with the expected compression of partials in the lower frequencies. The failures become less frequent as the interval size is increased or at a higher octave. For Smt-LF, there is a significant increase in such resolution failures at octave 5 with 1 interval. This corresponds to the reduction in scores in Figure (4.4), and is probably due to the Smt-LF mapping filtering out overtones higher than 1502 Hz. Thereby the tones are reduced to only their fundamental component which makes tone resolution in higher octaves difficult.

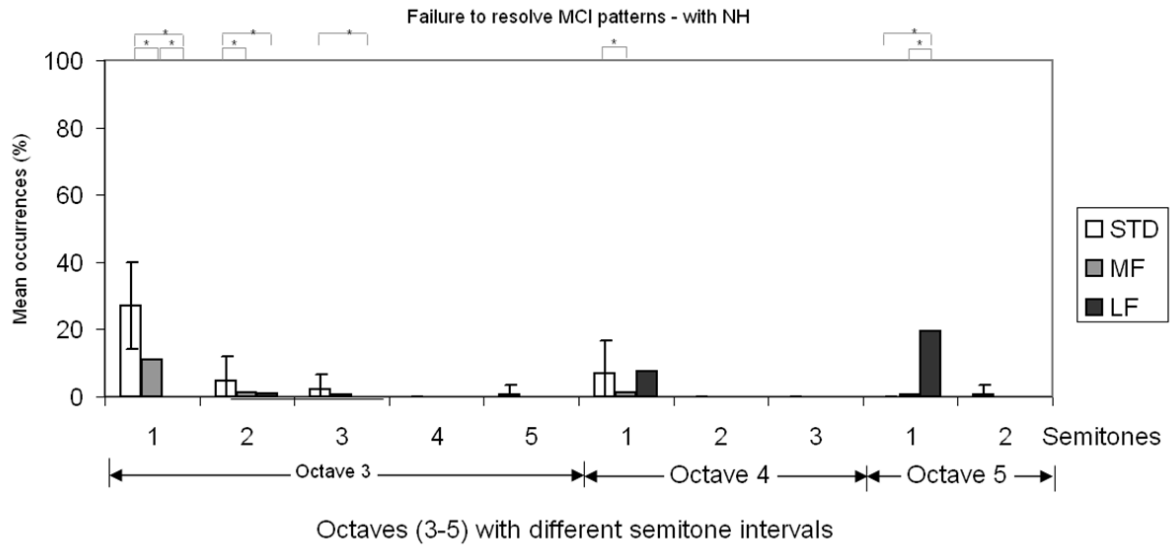


Figure (4.9): Mean frequency of occurrence of failures to resolve a contour pattern for NH subjects with AMO outputs for standard (white), semitone Smt-MF (gray) and Smt-LF (white) mappings. An asterisk between two columns indicates that the corresponding scores are significantly different (p = 0.05) from each other.

The results in Figure (4.7) and Figure (4.9) also show that there is generally little difference between the three mappings with large (4 and 5) intervals, and that these are therefore superfluous for this test. Also, MCI contours in the higher octaves (4 and 5), except at 1-semitone interval, are also largely redundant. Furthermore, Smt-MF mapping filters out too many of the partials from tones in octave 5, making it difficult to perform meaningful comparisons. Consequently, it was decided that the subsequent testing with CI subjects would concentrate on octaves 3 and 4, with 1, 2 and 3 semitone intervals.

The MCI test was repeated using a reduced number (5 instead of 9) of contour patterns (see Figure (4.6) with CI recipients. Eight CI recipients conducted the MCI test with twice the number of repetitions and the same mapping conditions.

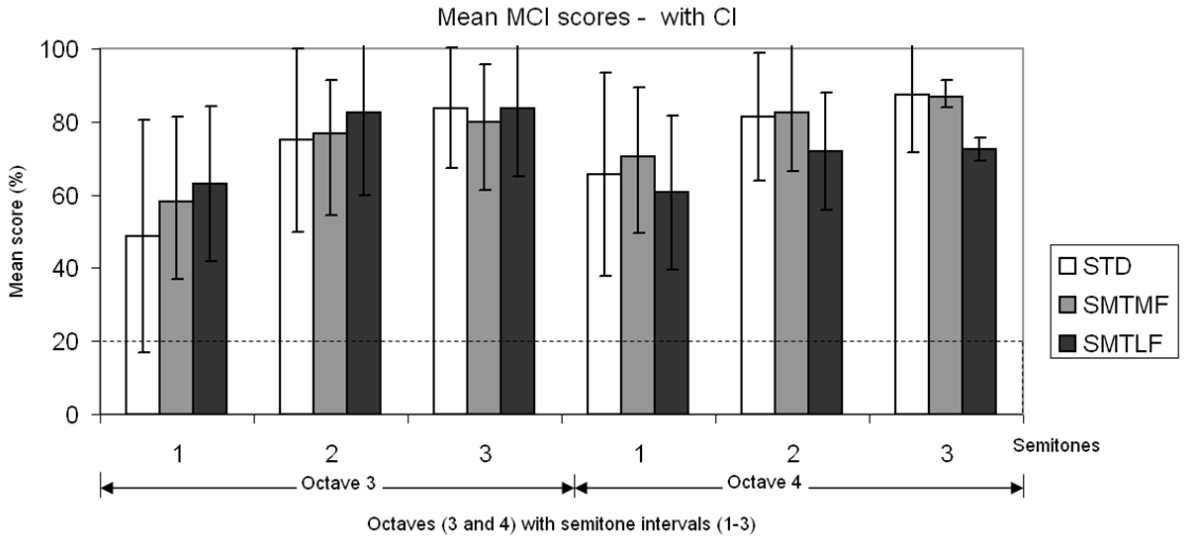


Figure (4.10): MCI test results with 8 CI recipients for standard (white), semitone Smt-MF (grey) and Smt-LF (black) mappings. Two octaves (3 and 4) were tested with semitone intervals from 1 to 3. Chance level is indicated by the dashed line.

Figure (4.10) shows the results for CI recipients with Std, Smt-MF and Smt-LF mappings. With all three mappings, the identification scores generally improved when the interval size was increased from 1 to 2 semitones, whereas the differences in scores were smaller when the interval size was increased from 2 to 3 semitones. No significant differences were found between all three mappings. In octave 4, the Smt-LF score was lower than in octave 3, and also lower than the scores compared with Std and Smt-MF mappings. This decrease may be due to filtering out of HF partials with Smt-LF. This is illustrated in the electrodograms in Figure (4.11) for the rise-fall pattern in octave 3 (Figure 4.11-Left) and octave 4 (Figure 4.11-right) with 2 semitone intervals. Figure (4.11) also shows that the Smt-LF pattern is transposed to channels with higher characteristic frequencies, and that HF overtones are filtered out from the 4th octave signal's pattern (right), leaving less cues in the resulting signal to perform the contour identification.

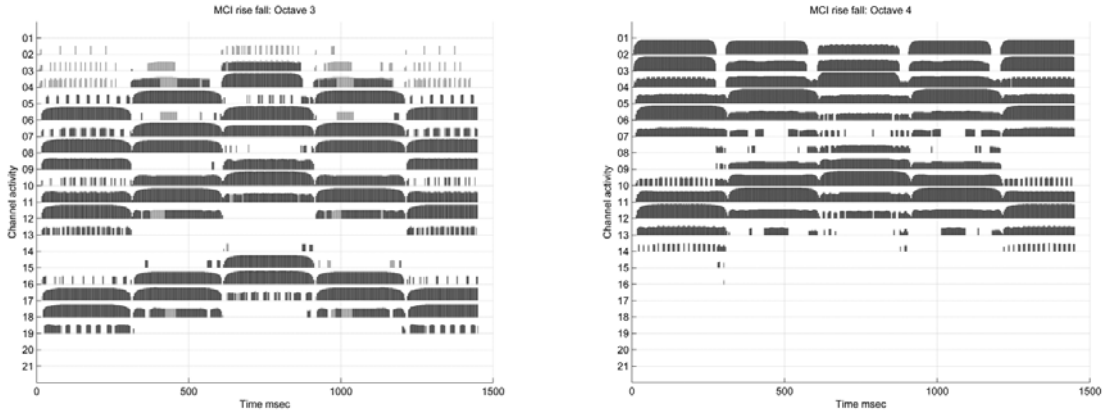


Figure (4.11): Electrodograms for the MCI rise-fall pattern in octave 3 (left) and octave 4 (right) with 2 semitone intervals, using Smt-LF mapping. Smt-LF, which has an upper cut-off frequency of 1502 Hz, has filtered out most of the octave 4 signal's higher partials. The two electrodograms also demonstrate how Smt-LF results in a transposition to higher frequencies.

The CI recipients' failure to resolve melody contours is shown in Figure (4.12). A significant decrease in the number of failures to resolve the contours with Smt-MF at octave 3 with 1 interval was found in comparison with Std mapping. Smt-LF mapping also showed decreased number of failures. The difficulties in resolving the contours with Std are most likely due to the poor representation at lower frequencies. In octave 3, with Smt-MF the lower frequency partials (the fundamental in particular) were filtered out in contrast to Smt-LF which preserved the LF partials (see Figure (4.13). Even with the semitone mapping, lower partials are generally better resolved than higher partials, due to the logarithmic frequency-to-channel assignment, resulting in a spatially denser representation of the higher partials. Together with effects like the SOE, this makes it more difficult to resolve contours when the lower partials are missing. The importance of the lower partials is supported by the observation that with Smt-LF in octave 4, where the higher frequency partials have been filtered out, the performance improved compared to octave 3.

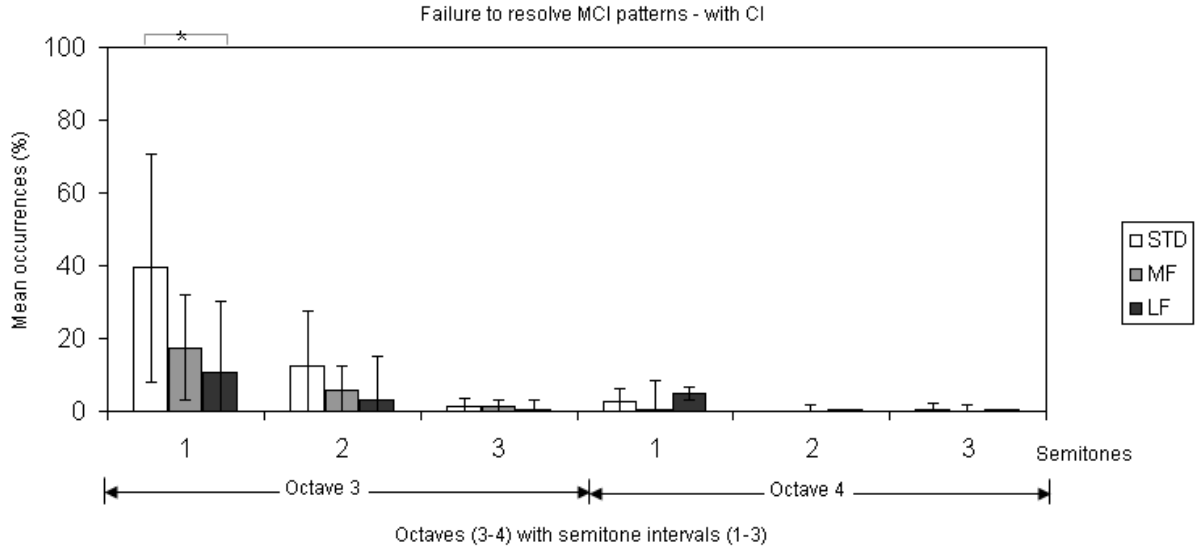


Figure (4.12): Mean frequency of occurrence of failures to resolve a contour pattern for 8 CI recipients for standard (white), semitone Smt-MF (gray) and Smt-LF (black) mappings. Two octaves (3 and 4) are plotted with different semitone intervals. An asterisk between two columns indicates that the corresponding scores are significantly different (p = 0.05) from one another.

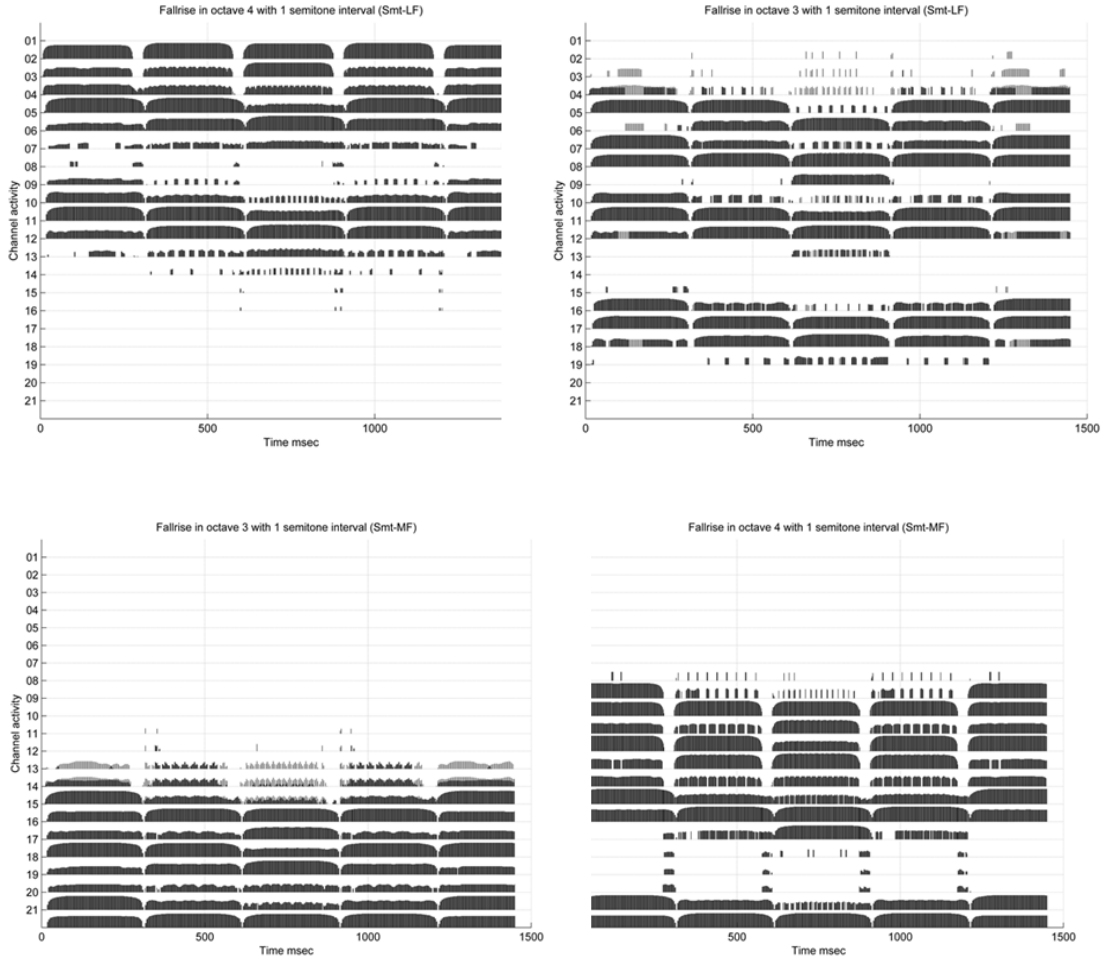


Figure (4.13): Results of Smt-LF mapping (above) for the fall-rise pattern in octave 3 (left) and octave 4 (right) using 1 semitone intervals. Results of Smt-MF mapping (below) for the same pattern in octave 3 (left) and octave 4 (right) with the same semitone intervals.

Overall, CI scores were lower than simulation scores. The significant benefits of semitone mappings which was found for NH listeners and AMO synthesized signals were not observed in CI users with MCI test. This may be due to lack of a long term familiarization. However, a significant reduction in failure to resolve tones was noticed with Smt-LF. More importantly, unlike NH subjects listening to simulations, CI users did not seem to have pitch reversals because their Smt-MF scores were not poorer than their Std scores in octave 3 for the 1-semitone interval condition.

4.4 Instrument Recognition

Pitch and melody are two aspects of music and timbre or tone color is another important one. Different instruments are characterized by their tone color (Helmholtz 1954). Timbre is defined physically by the temporal envelope (particularly the onset) and the spectral shape of the acoustic sound (Handel 1995). In this test patients have to identify sounds played by a musical instrument with different mappings.

4.4.1 Hypothesis

Improving frequency representation with Smt mapping may improve IR compared to the Std mapping.

4.4.2 Method

The first 8 bars from the music piece "Vem kan segla förutan wind?" (traditional Swedish folksong) played by professional musicians on eight different instruments (Trumpet, Trombone, Flute, Clarinet, Violin, Cello, Guitar and Piano) were recorded and used as the basis of the test material. Dividing each recording into sub melodies of 2 bars each then produced a total of 4 "pieces" per instrument. The instruments could be divided into four families, namely Brass, Woodwind, Bowed Strings and Struck Strings, each consisting of two instruments (see Figure (4.14)). In the IR test, the listener was required to listen and identify the instrument used to play the piece being presented.

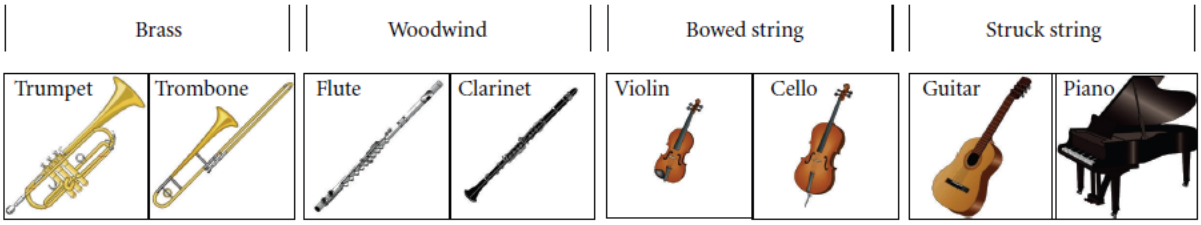


Figure (4.14): The eight different instruments from four instrument families (Brass, Woodwind, Bowed Strings and Struck Strings) used in the instrument recognition test.

As with the MCI test, the signals were presented via streaming to the CI recipients with the L34 speech processor. The signals were preprocessed with the Matlab utility "*Checker*" for all three mappings (Std, Smt-MF and Smt-LF), using patient specific settings of minimum and maximum current levels per electrode retrieved from a clinical database and the processed signals were first saved as XML files prior to the test being conducted. The input signals to the Checker were calibrated to correspond to an equivalent acoustic (loudspeaker) mean level of 60 dB SPL.

CI recipients were seated in front of a touch screen and an XML file was streamed to the L34 speech processor from the MACarena test environment in combination with NIC. The CI recipients had to select the instrument that corresponded to the perceived sound using eight response buttons corresponding to the eight instruments shown on the touch screen.

Before testing began, the CI recipients practiced with a limited set of signals in familiarization and training sessions. In a familiarization session, the CI recipient pressed a button on the screen to listen to the corresponding sound. In a training session, feedback was provided as to whether the response was correct or wrong. If a response was wrong, the correct response would be indicated on the screen and the same sounds could be repeatedly presented. The final test involved presenting each of the 8 instruments a total of 4 times (corresponding to a single presentation of each of the 4 sub melodies) without feedback. 8 adult post-lingual CI recipients performed the test. All subjects had at least 1 year's experience using a CI device. All of them used the Nucleus Cochlear Implant.

4.4.3 Results

The results were analyzed for the percentage correct scores for identifying the individual instrument (8 possibilities) and the instrument family (4 possibilities).

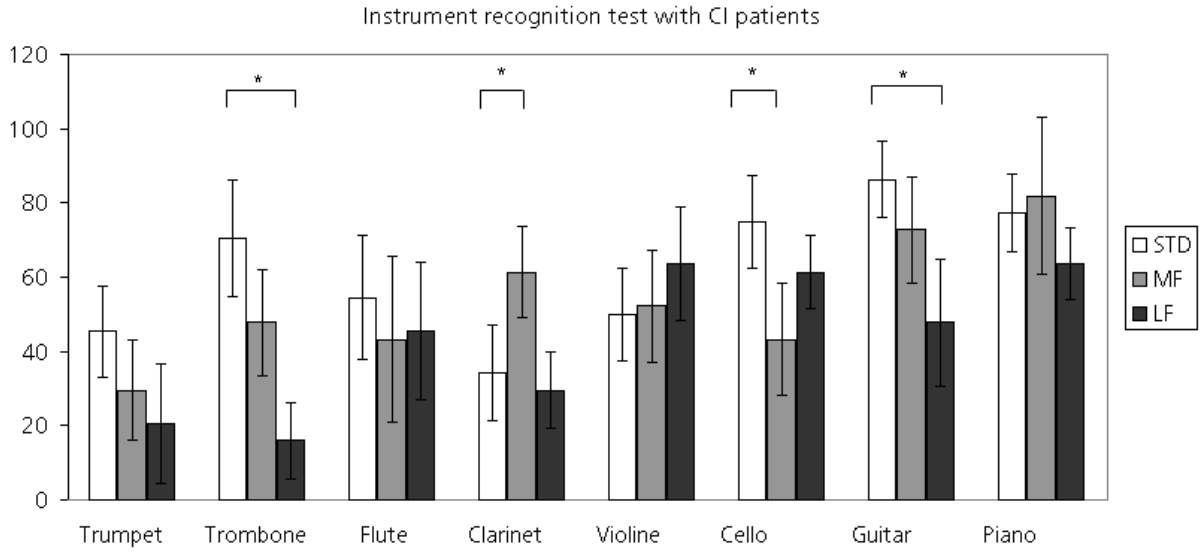


Figure (4.15): Instrument recognition scores with 8 CI patients for different instruments using Std (white), Smt-MF (gray) and Smt-LF (black) mappings. A significant improvement was detected with the Smt-MF for the Clarinet instrument.

Figure (4.15) shows the IR scores with CI patients with the three mappings (Std, Smt-LF and Smt-MF). In general it shows that the Std mapping yielded better identification scores. Piano and Violin tones were slightly better recognized using Smt-MF mapping. Smt-MF was significantly better than Std and Smt-LF for the Clarinet instrument. One reason may be that in general Clarinet partials are more harmonically related than other instruments (for example the Cello). Violin and Cello were better recognized with Smt-LF than Std mapping.

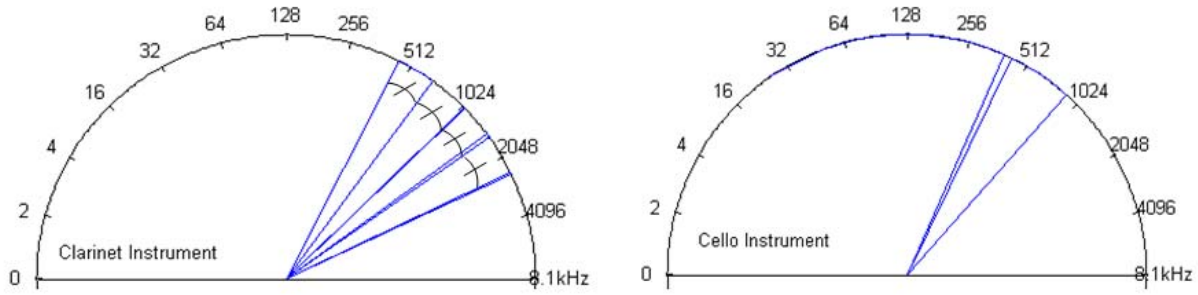


Figure (4.16): A polar representation of frequency components along an octave spacing binary spectrum for both Clarinet (left) and Cello (right) instruments. It illustrates that angular distance or in other words semitone spacing between different components in the Clarinet are almost. Partials amplitudes were extracted from logarithmic amplitude FFT with a threshold at -90 dB and then replaced with a constant value. Blue points on the circumference represent noise allocated at corresponding frequencies.

Figure (4.16) shows a comparison between unprocessed tones from Clarinet and Cello instruments. The figures show polar plots of frequency values of existing partials allocated on a binary spectrum to represent octave spacing. The figure shows that the angular differences between partials in the clarinet instrument are almost equal, which is not the case with Cello (see Figure 4.16 right). This equal spacing of harmonics in a natural instrument was probably well recognized with Smt-MF as shown in Figure (4.15).

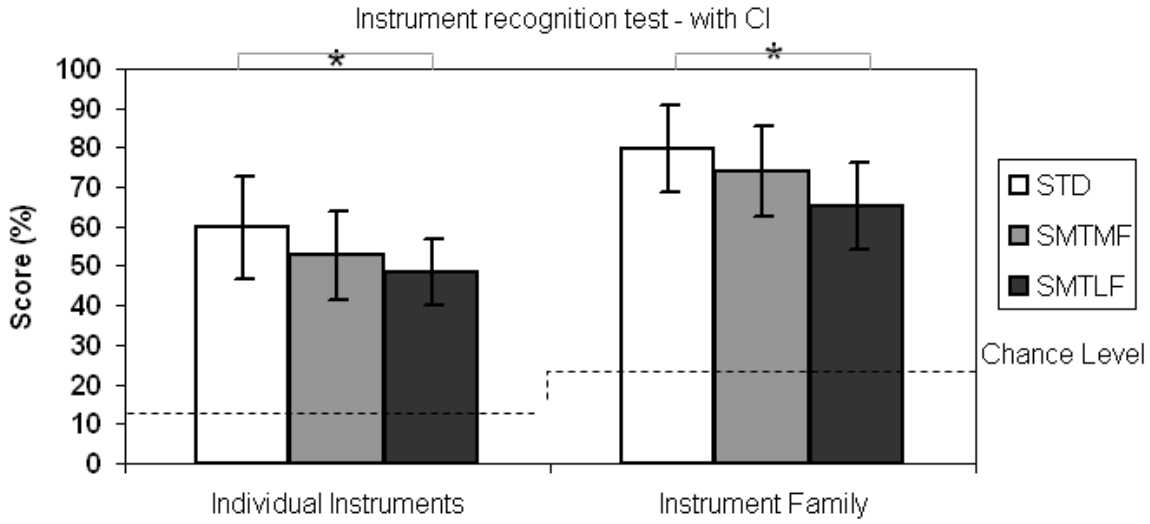


Figure (4.17): Results of individual musical instruments and instrument family recognition test with CI recipients using standard (Std) (bricked) and Smt-MF (gray) and Smt-LF (black) mappings. Dashed lines illustrate chance level. An asterisk between two columns indicates that the corresponding scores are significantly different ($p = 0.05$) from one another.

Figure (4.17) summarizes the average results with Std, Smt-MF and Smt-LF mappings. The average identification scores decreased significantly with Smt-LF mappings compared to Std mappings for individual instruments as well as instrument families. This may be because characteristic differences between instruments such as timbre are contained in the temporal fine structure rather than the tonotopic frequency allocation (Handel 1995). The three mappings Std, Smt-LF and Smt-LF use different window lengths of 128, 512 and 512 respectively for their processing algorithms. In addition, Smt-LF halves the sampling rate to increase the frequency resolution for frequencies below 1054 Hz, which account for the majority of its input frequency range. Consequently, the temporal resolution is expected to be

best with Std and poorest with Smt-LF. Additionally, as these strategies do not encode the temporal fine structure properly, patients can only rely on the spectrum to identify different instruments. Since the Std mapping is covering the widest frequency range (180-7800 Hz) compared to semitone mapping Smt-LF and Smt-MF ranges (130-1502 Hz) and (440-5009 Hz) respectively, the highest amount of spectral information is transmitted with Std mapping. Another possible reason could be that the subjects were more familiar with the Std mapping, which is very similar to the mapping in their daily used speech processor.

Chapter 5 - Discussion and Conclusion

Chapter 5 – Discussion

5.1 Discussion

CIs and current coding strategies have limitations arising from the physical size of the electrode array and the insertion depth which contribute to CI recipients' difficulty to recognize melodic sounds. Furthermore, the available 22 electrodes of the Nucleus CI are probably not sufficient to provide a good representation of melodies. (Drennan and Rubinstein 2008) report that with monopolar stimulation, recipients perceive pitch based on a degraded spatial representation, suggesting that only 8-9 distinguishable channels of the available 22 may exist. To increase the number of channels, VCs (Busby and Plant 2005) were suggested. However, VCs may not necessarily increase the number of independent channels because stimulating two electrodes simultaneously increases the SOE due to interaction of electrical fields.

Preserving the harmonic structure of musical notes is expected to ameliorate melody representation of musical sounds with CIs. Melody preservation is investigated in this thesis by employing semitone mapping using two ranges together with the use of VCs. The frequency distribution of a semitone mapping is determined by the number of semitones that is to be mapped over the number of available channels. When incorporating VCs, the CF of each channel is one semitone apart from the previous one. If VCs were not used (i.e. only the original 22 physical channels (electrodes) are used instead), the CF of each channel would be two semitones apart, in order to cover the same input frequency range.

The Smt-LF mapping [130 – 1502 Hz] needs a resolution of ~8 Hz at the low frequency end (130 Hz) to resolve the two lowest semitones. Although the resolution can be increased by increasing the number of points in the FFT analysis, this also causes the signal to be smeared in the time domain. A compromise is achieved by using frequency sub bands, in which different frequency bands can have different resolutions. With Smt-LF, frequencies below 1054 Hz are analyzed with a resolution of ~8 Hz while higher frequencies are analyzed with a 31.25 Hz resolution. Note that the frequency allocation for each channel will not be exact for each semitone due to the discrete nature of the FFT to bin allocation. There is a limitation with the Smt-LF mapping in that frequencies higher than 1502 Hz as well as frequencies lower than 130 Hz are filtered out. This might lead to filtering out some fundamental frequency that could then cause pitch reversals of tones or degradations in the perceived sound quality. In the Smt-LF mapping, frequencies are transposed into higher ranges and this would make tones sound higher pitched. It is expected that CI patients with Smt-LF mapping may distinguish low frequencies better but music may not sound natural.

In Smt-MF mapping, the characteristic frequencies of channels approach their tonotopical frequencies according to the Greenwood function. Consequently, Smt-MF mapping may sound more natural than Smt-LF mapping. Smt-MF does not require sub band decomposition when a FFT size of 512 points is used with a resolution of 31.25 Hz (except for the lowest two tones). There is however a limitation with the Smt-MF mapping at low frequencies, where components below 440 Hz are filtered out. This filtering may remove the fundamental frequency of some tones which could then result in pitch reversals compared to tones that have all frequency components intact.

An exact matching of the electrode location to the tonotopical CF location is not possible, nor is it desirable (Dorman et al. 2007). Thus, the phenomenon of missing fundamentals is not expected to occur with CI patients (Oxenham 2008). Equation (2.6) in Chapter 2 yields a

range of 0.19 mm to 0.38 mm for the spacing between successive semitones for the Smt-LF mapping. For the Smt-MF mapping, the range is from 0.3 mm to 0.4 mm. Assuming that VCs are located exactly midway between two adjacent physical electrodes, the channel spacing for the Nucleus Contour Advance electrode array is about 0.2 mm to 0.4 mm. This is the same order of magnitude as the spacing required for the Smt-LF and Smt-MF mappings. Although the match is not exact, the Semitone mapping minimizes the spatial distortion of the harmonic structure representation on the BM.

Smt mapping leads to the stimulation of different groups of electrodes for each tone on the musical scale. Each tone has a harmonic structure of overtones which will in turn be regularly spaced along the BM. The analysis results showed that Semitone mapping preserved the ratio of partials of the processed tones. This may lead to a better preservation of the harmonic structure when presented to a CI recipient. The mapping approach outlined in this thesis has been tested analytically with DDM and harmonicity index in Chapter 3 and psychoacoustically with NH listeners as well as CI recipients in Chapter 4.

The Smt mapping algorithm illustrated in this thesis follows the regular signal processing methods as in the Std mapping algorithm, but the STFT processing and the subbands blocks may be optimized using the S transform (Stockwell 2007) for limited memory speech processors. It may be also worth investigating the effect of other transforms such as constant Q transform (Brown 1991) where the frequency domain is decomposed into logarithmically spaced bands with a constant number of semitones per octave. Derivatives of this transform may be considered as well such as the modified constant Q fast filter bank transform (Diniz et al. 2007).

Smt mapping may be one possible step on a path to improve music representation with CI devices by preserving the harmonic structure of overtones to enhance melody representation, but other aspects of music such as timbre have also to be considered in the future as well. Timbre depends partially on the representation of temporal fine structure (FS) (Handel 1995). Including FS analysis in new strategies may improve music representation (Nie et al. 2005; Chen and Zhang 2008; Drennan and Rubinstein 2008) but this is beyond the scope of the present study.

In an elementary analysis both Smt and Std mapping are used to process synthetic musical tones with 22 and 43 channels using the AMO. In a further step, the resulting electrodograms are processed employing DDM, a newly proposed matrix for diagonals detection. It has been used to estimate a relative amount of harmonic overtones preservation with Std and Smt mappings. The results show that incorporating 43 channels provides a higher amount of preservation than 22 channels with Smt-LF mapping. It showed also that the Smt-LF highly ameliorates harmonic structure preservation of overtones in musical notes with respect to Smt-MF mapping. Both Smt mappings were superior to Std mapping.

Music in general does not have a measure that quantifies its different aspects in a simple value. However, harmony, which is one aspect of music, was selected to be estimated with a newly proposed index that calculates approximately the probability that overtones in an acoustic signal are harmonically related. This index showed differences between Std and Smt mappings. Thus, the harmonicity index proposed in Chapter 3 was used to quantify musical tones from eight musical instruments after being processed and resynthesized with the Std, Smt-LF and Smt-MF mappings using the AMO. The results of the unprocessed and the processed tones were compared.

Three different harmonicity index approaches (V1, V2 and V3) were studied; the first two approaches mainly depend on a correct estimation of the fundamental frequency. Then the algorithm searches for expected harmonic overtones with a tolerance of ±½ Bark. Algorithm V2 is a modified version that calculated energy distribution per critical bands to refine existing peaks. The result showed that algorithm V1 was frequency dependent while algorithm V2 failed in many cases in proper estimation of the fundamental frequency estimation. These results motivated a search for another approach that does not depend on the fundamental frequency. The third approach (V3) fulfils the required condition; it does not depend on estimating the fundamental frequency. An approximation is employed to determine the cumulative error that describes the relative shift of two partials from expected harmonic overtones with a tolerance of ± ½ Bark from each partial. V3 yielded more stable results than the previous two approaches and thus this approach was extended to investigate the effect of processing mappings on the harmonic structure of musical tones.

The V3 method was validated by computing the harmonicity index for chords composed of two tones and 5 harmonic overtones; the first tone had a fundamental frequency 250 Hz while the fundamental frequency of the second was shifting stepwise from 250 Hz to one octave above. The harmonicity index was calculated for each shift and the output showed in general similar results than a consonance and dissonance study by (Pierce 1983). Slight differences were observed such as shifts in lobes' positions, but the percentage of the relative error was less than 5 %. The shift in lobe positions may be due to the approximation involved in the method which grows as the fundamental frequency decreases.

Another evaluation test was conduced in more dimensions using additional chords. Each chord was again composed of two tones and each tone had 5 harmonic overtones similar to the previous test but with a shift in the harmonic structure of the overtones. The fundamental frequency of the first tone and the second were fixed while the harmonic overtones of the second were shifted with a percentage from the fundamental frequency. The shift varied from 0 upto 110 percent and the fundamental frequency varied from 32 to 2217 Hz in one semitone interval step. The results showed that some zones at different frequency shifts exist; the highest index was found when the shift was from 0 to 9 % and from 100 to 109 % for frequencies above 55 Hz. The zones were relatively wide due to the tolerance width of ± ½ Bark. For shifts between 10 and 99 % the index value varied. It increased when the shift approached a zone and vice versa. The existence of zones confirms the results from the previous experiment and of the consonance and dissonance study by (Pierce 1983). One drawback in the algorithm may be the approximation used.

Processing real musical tones from eight instruments by the three mappings (Std, Smt-LF and Smt-MF) produced harmonicity indices which were significantly lower than those of the unprocessed versions. The harmonicity index with Smt-LF was significantly different from the unprocessed with SOE = 0.5 mm. It was lower than for the Std mapping when using sinusoidal vocoders and SOE = 1mm. One reason may be the frequency transposition caused by Smt-LF mapping. Such shifts in fundamental and partial frequencies decrease the number of expected overtones below the maximum frequency (8 kHz); e.g. for a 4 kHz frequency component to be harmonic, at least a partial at 2 kHz has to exist because 8 kHz is rejected. Thus transposing a signal into higher frequencies below 8 kHz, decreases its likelihood to be harmonic. There was a significant enhancement in the harmonicity index using Smt-MF with sinusoidal and noise vocoders with all SOE values over Std mapping and this coincides with results obtained using DDM.

Although implant recipients perceive basic rhythm patterns similarly to NH subjects (Gfeller et al. 1997a), perception for pitch, pitch sequences and melody recognition is significantly

poorer than that of NH subjects (Dorman et al. 1990; Dorman et al. 1991; Gfeller and Lansing 1991; Pijl and Schwarz 1995b; Gfeller et al. 1997a; Pijl 1997; Fujita and Ito 1999a).

Pitch discrimination was tested with NH subjects using sounds processed by the AMO. The AMO was based on a noise band vocoder. One of the parameters needed for the AMO was the SOE (width of stimulation). (Laneau and Wouters 2004; Laneau et al. 2006a) found that a width of stimulation of around 1 mm produced electrode discrimination similar to that of average Nucleus CI24 recipients. Prior to using the AMO for testing with NH subjects for the present study, a pilot test was conducted to examine the effect of the width of stimulation. The Oldenburg sentence recognition test (Wagener et al. 1999b) in quiet was chosen for this purpose with the Std mapping using different widths of simulation (1, 3.3 and 10 mm). The results shown in Figure (5.1) indicate that widths of 1 and 3.3 mm gave very similar results (90% and 87% respectively). With 10 mm, the results were very poor and were considered to be not representative of CI recipients performances (Mueller-Deile 2009). A 1 mm width of stimulation was selected for further tests with the AMO as this matches well with the recommendation by (Laneau and Wouters 2004; Laneau et al. 2006a).

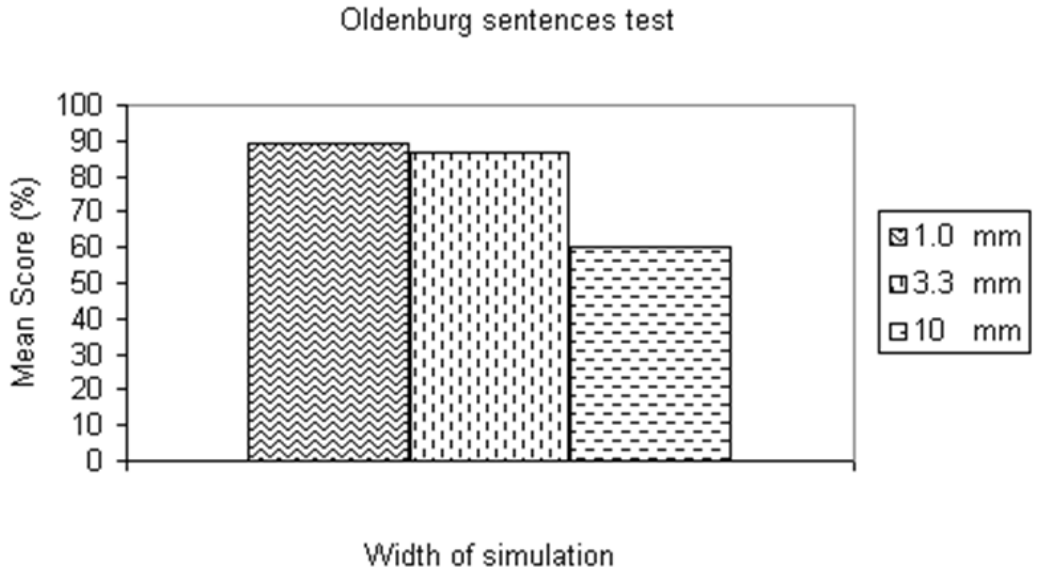


Figure (5.1): Average correct scores for the Oldenburg sentences test performed with different widths of simulation (1, 3.3 and 10 mm) for the AMO using two lists with 10 sentences each. Only native German speaking subjects were tested.

Two pitch discrimination pilot tests were carried out with Std mapping employing 43 and 22 channels, one using clarinet tones and the other using pure tones. The results obtained from the pure tone test showed the expected increase of semitone difference limens. They were different from the results with complex tones. No significant enhancement was found when VCs were used. To examine clues that may provide extra details on psychoacoustic effects governing these results, the spectral-contrast model was utilized with the test signals. Figures (5.2 and 5.3) show the results of the spectral-contrast index calculated for pure and complex clarinet tones processed with Std mapping using 22 and 43 channels. The spectral-contrast results with pure tones coincided with the pitch discrimination results using 22 and 43 channels in which tones in higher octaves were better discriminable than in lower octaves and wider semitones intervals were advantageous to narrower intervals. However the spectral-contrast model could not provide clues for the complex tones except showing a rising trend in octave 5. This may be due to psychoacoustic processing of the brain which was not included in the spectral-contrast model (e.g. masking).

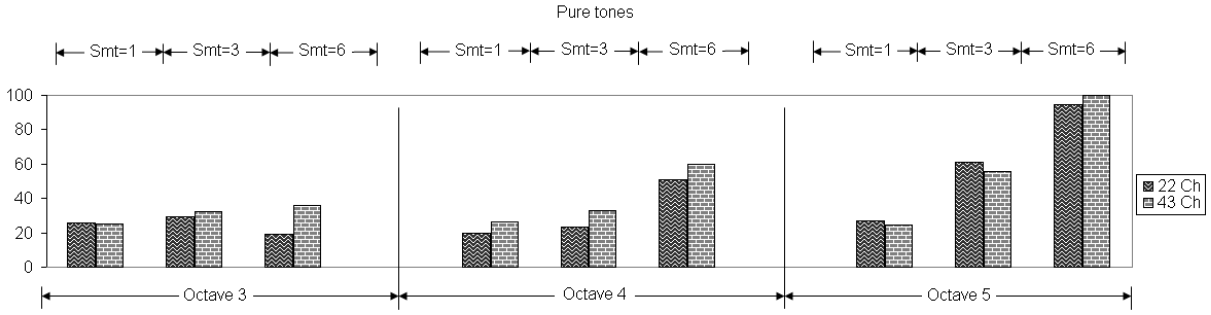


Figure (5.2): Results of spectral-contrast index calculated for pure tones processed with Std mapping using 22 (black dots) and 43 (gray bricks) channels in octaves 3, 4 and 5 with 1, 3 and 6 semitone intervals between sounds.

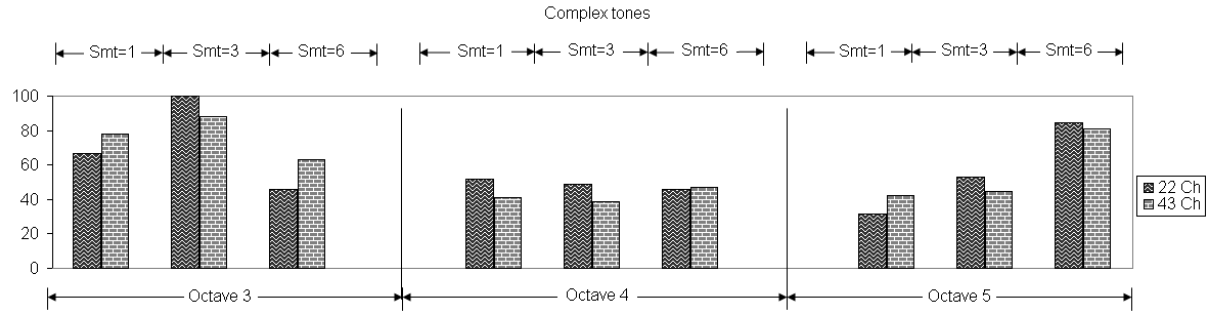


Figure (5.3): Results of spectral-contrast index calculated for clarinet tones processed with Std mapping using 22 (black dots) and 43 (gray bricks) channels in octaves 3, 4 and 5 with 1, 3 and 6 semitone intervals between sounds

Another pitch discrimination test was conducted with two reference tones (D and G#) with different semitones intervals for the three mappings (Std, Smt-MF and Smt-LF) using the AMO with NH subjects (see section 4.2.2). This test with NH subjects was intended to examine whether the Semitone mappings would indeed produce better representation of complex tones over Std mapping. Results with unprocessed synthetic complex tones confirmed that (a) the test material was suitable for such a task, and (b) the subjects were able to perform the task. Results tended to be poorer with smaller intervals between the probe and reference, and also poorer in a lower octave range. This is consistent with the reduction in critical band sizes at the frequencies in concern (i.e. below 500 Hz) (Pierce 1983).

The pitch discrimination results with the AMO model (see Figure (4.4)) showed that Std mapping was significantly poorer than either of the Smt mappings for the tone pair D-D# (1-semitone interval) in all three octave ranges. With 3-semitone intervals, Std mapping was significantly poorer than Smt-LF mapping at the lowest octave (D3-F3) only. With a higher pitched reference (G#), these difficulties with the Std mapping were not observed. This is consistent with the fact that Std mapping compresses the representation of lower frequency partials (as described in section 2.4), thereby making it difficult to distinguish between tones that are close to each other. Smt mapping in general improves the representation of the partials. Pitch reversals were seen with the Smt-MF mapping in Octave 3 with the D reference, and in Octave 4 with the G# reference. A closer examination of the power spectrum estimates for the AMO generated tones, for instance G4# and D5 (with fundamental frequencies of 392 Hz and 554 Hz respectively), shows that the loss of partials below 440 Hz filtered out by Smt-MF shifts the lowest remaining partial of G4# to a frequency higher than that of D5 (see Figure (5.4)). Thus, the loss of lower frequency partials due to the cutoff frequency of Smt-MF is a likely cause of the observed pitch reversals.

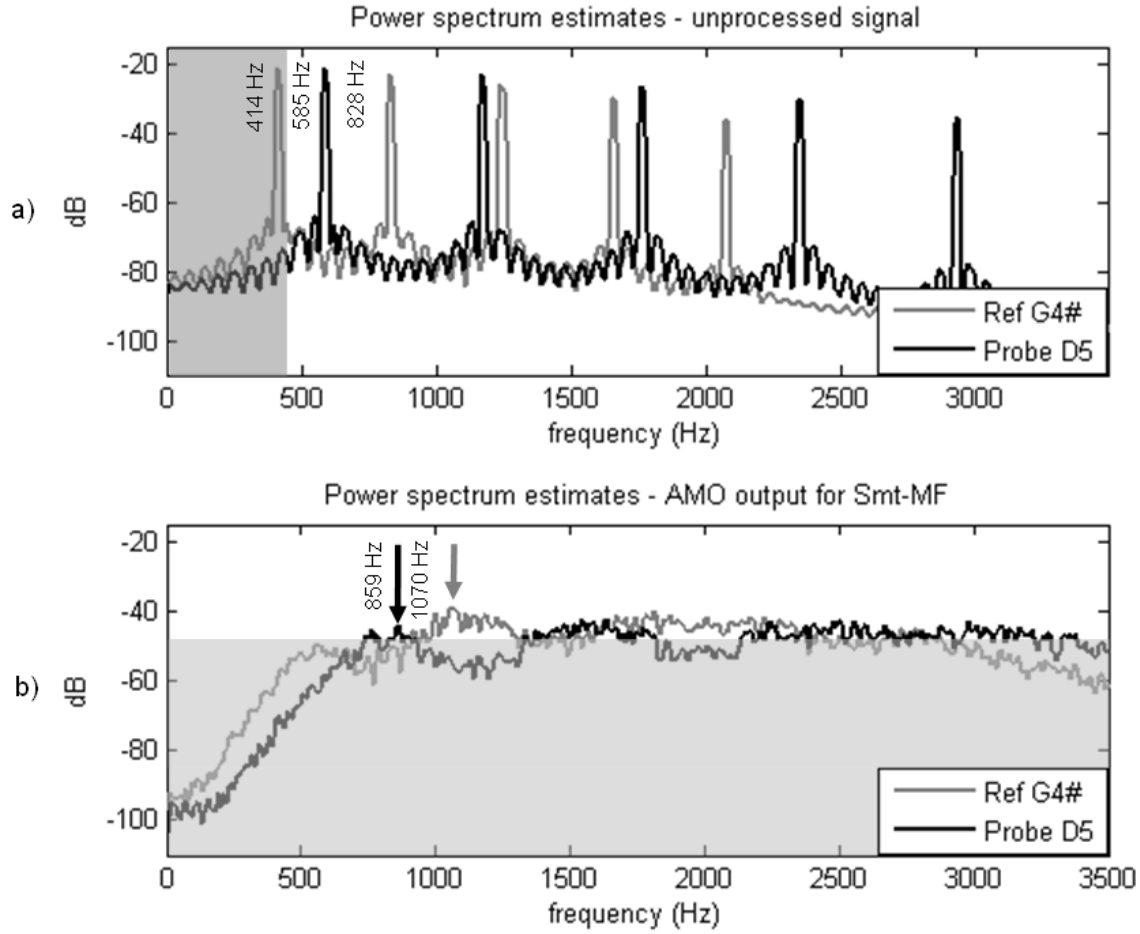


Figure (5.4): Power spectrum estimates for both the G4# Reference (black) and D5 Probe (6-semitones interval) (gray) signals for the unprocessed tones (a) and using Smt-MF mapping after AMO processing (b). The fundamental frequency of the reference tone (414Hz) is filtered out because it is below 440 Hz (shaded area in figure 14-a), while that of the Probe (585 Hz) is retained. After Smt-MF processing with the AMO, comparisons of the unfiltered first overtone's peak above the theoretical noise floor of the power spectral density (–48.2dB) which is double the theoretical noise floor of FFT in dB (Kester 2003) (shaded area in Figure 5.2-b) in the Reference (1070 Hz) with the fundamental of the Probe (~ 859 Hz) would then result in a pitch reversal.

These results may not be transferred directly to CI recipients, as the AMO only produces an approximation of the CI perceptions (Dorman et al. 2007). However, the results did show that in principle, Semitone mapping has the potential to produce better pitch discrimination of complex tones that possess a harmonic structure. Pitch reversals arising due to filtering out of the lower frequency partials had a negative effect on the identification scores. Smt-LF mapping filters out partials below 130 Hz and above 1502 Hz, while Smt-MF mapping filters out partials below 440 Hz and above 5009 Hz. For the range of tones tested here, Smt-LF caused no loss of lower frequency partials due to filtering. Smt-MF, on the other hand, is more likely to result in the lower partials of the lower pitched tones being affected by its band-pass filter cutoff. Smt-MF mapping was designed to be as close as possible to the characteristic tonotopic frequencies of the electrode array according to the Greenwood function (Greenwood 1990), assuming an average cochlea length of 33 mm and an insertion depth of 22 mm. It is not possible to map Smt-LF to characteristic tonotopical locations because the lowest input frequency of 130 Hz is far too distant from the characteristic frequency of the most apical electrode (whose location should correspond to a characteristic

frequency of around 400 to 600 Hz) according to Greenwood. Instead, Smt-LF mapping effectively transposes frequencies from 130 to 1502 Hz into a higher tonal range. This may simplify pitch discrimination of tones in the lower frequency octaves (e.g. octave 3) but may cause them to be perceived as unnaturally high pitched. The results nevertheless demonstrated that Semitone mapping may improve pitch discrimination due to improving the frequency representation.

The pitch discrimination results also showed that this test has different sensitivity in different tonal ranges. With the G# reference, it was not sensitive enough to detect differences between the various mappings being investigated. Pitch discrimination is possibly a too simple task and further studies should involve a more complex test that is sensitive enough to show differences between mappings. Such a test will be first assessed with NH subjects and then subsequently with CI recipients. This study suggested investigating melody which is another aspect of music.

Melody can be described as a group of tones that are perceived as a single entity (Terhardt 1998). Different melody tests exist, such as simple melody recognition with lyrics (Gfeller et al. 2000b) or a sequence of familiar notes (Gfeller et al. 2002a; Olszewski et al. 2005; Looi et al. 2008), complex song recognition and complex song appraisal (Gfeller et al. 2000b). Although it would have been more appropriate to perform a melody test, the existing tests do not involve the perception of melody alone, but also involve other perceptual mechanisms such as pattern recognition as well as memory (familiarity). For instance, familiar melody recognition has been used to directly asses CI listeners' music perception abilities (e.g. (Pijl and Schwarz 1995b; Pijl and Schwarz 1995a; Fujita and Ito 1999b; Kong et al. 2004a)) but general results showed that CI recipients are much worse than NH subjects (Galvin et al. 2007). In addition, Lynch et al. found that musical experience or education and cultural background greatly influenced melody recognition performance when notes within a melody were mistuned (Lynch et al. 1991). (Gfeller and Lansing 1991) measured CI recipients' melody and rhythm discrimination using the primary measures of musical audition (PMMA) test. Their results showed that CI recipients are more sensitive to changes in rhythm than in melody. Familiar melody identification may involve paying attention to the pitch contour of the melody (Dowling 1994). As such, deviations from the expected intervals for a familiar melody may strongly affect identification performance. CI recipients may depend less on the exact intervals and more on the general contour in pitch, rhythm and timbre. Because CI recipients' perception of these musical aspects will be limited by the amount of information transmitted by the devices, we chose to investigate the CI recipients' ability to identify melody contours. Galvin et al. introduced the MCI test which assesses the listener's ability to detect and identify interval changes between successive tones in a short sequence (Galvin et al. 2007). Among the advantages of this test is that confounding factors such as rhythm can be eliminated, and the contour patterns do not need any previous familiarity for the listener to perform the task.

The results of the MCI test with NH subjects (see Figure (4.7)) showed similarities with the results from the pitch discrimination test (see Figure (4.4)) in that significant improvements over Std mapping were obtained for Smt-LF mapping, particularly in Octave 3 with 1 and 3 semitone intervals, as well as in Octave 4 with 1 semitone interval. However, the pitch discrimination improvements were found with the D reference but not the G# reference, whereas the MCI patterns had a root note of A, and the tone intervals are more similar to the pitch discrimination intervals with the G# reference. Thus, the pitch discrimination and MCI results cannot be directly inferred from one another. The MCI test is probably a more difficult task as the listener had to concentrate on the contrasts between upto 5 tones, whereas the pitch discrimination task only involved a single contrast. A given tonal range which was relatively

easy for pitch discrimination may thus be expected to be more difficult when multiple contrasts are involved. The observation that the MCI test results showed the same trend at a higher "reference" or "root" tone suggests that MCI is not merely a more complex form of pitch discrimination involving sequential tones, but is also a more difficult form.

With Smt-MF, the poor MCI results in Octave 3 with 1 semitone intervals is most probably caused by pitch reversals in specific tones as a result of the lower partials being filtered out. Note that pitch reversals in specific tones are probably more crucial for contour patterns with smaller intervals, which are inherently more difficult to resolve. When larger intervals are involved, the subjects may still be able to use the other segments of the contour to perform the identification. With Smt-LF there was a significant decrease in contour identification at octave 5 with 1-semitone intervals most likely because of filtering out HF partials, resulting thereby in some patterns being identified as flat when they were not. The results with these particular patterns were further analyzed, and the inability or failure to resolve melody contours (see Figure (4.9)) in this manner was found to correspond to the observed reduction in identification score. The inability to resolve partials also accounts for the significantly higher number of errors with the Std mapping at Octave 3 with 1 semitone intervals, since frequency components in the octave 3 range tend to be mapped to a very small number of channels with the Std mapping.

The number of patterns was reduced to 5 contours (rise, rise-fall, flat, fall-rise and fall) for testing with CI recipients. Based on the results from the pitch discrimination and MCI tests with NH subjects, it was also decided to restrict the MCI test in octaves 3 and 4, using the more difficult interval-size conditions, namely with semitone intervals 1 to 3 in each contour.

Incorporating VCs, thereby increasing the number of available channels from 22 to 43, for testing with CI recipients is expected to produce improvements in performance. This study, however, was aimed primarily at comparing Semitone mapping against the Std mapping, and as the CI subjects did not use VCs in their regular daily routine, it was decided that the number of varying parameters should be minimized for the comparisons. With 22 channels, the resolution of the frequency to channel mapping was also reduced by a factor of two, meaning that always two semitones will be mapped to a single channel. The MCI test was therefore carried out with both NH and CI subjects using 22 channel mode in order to be able to compare the results directly. Patients did not have a short-term or long-term familiarization to Smt mapping due to technical constraints. Since Smt mapping uses a slightly different processing technique (Sub-bands and mapping matrices) which requires building a new firmware for the implant processor in order to set patients familiarization on a daily basis and the experiments could only be done in the laboratory environments. Performance may gradually improve with short-term and long-term experience with Smt mapping.

The MCI test results with CI recipients showed a general improvement in identification scores with increasing interval size. The enhancements found of Smt-LF and Smt-MF in the average scores were not significant for a given octave and interval size condition. The Smt-LF scores appeared to be lower for the octave 4 compared to the octave 3 conditions, most probably due to filtering out of higher partials resulting in less cues to distinguish between the contour patterns. Both Smt-MF and Smt-LF mappings were better than Std mapping in terms of resolving contours, especially in lower octaves (octave 3) with small (1-semitone) interval sizes. This emphasizes that semitone mapping may be advantageous to the Std mapping. Again, this is consistent with the expectation that Std mapping is unable to resolve tones well in these frequency ranges.

In general, Smt mapping showed some improvements over Std mapping with the MCI test. However melody contour is only one aspect of music perception. Timbre is another aspect that is involved in characterizing different instruments (Helmholtz 1954). Timbre depends on the frequency spectrum as well as the temporal fine structure of the perceived sounds. To investigate whether music with semitone mapping would be perceived as musical, a music IR test was carried out. In this test, the timbre was coded more in the temporal patterns rather than the frequency spectrum. Results with CI recipients in the IR test showed that there was a statistically significant enhancement of Smt-MF over the Std mapping with Clarinet. However, in general, there was a decrease in the average of individual instrument and instrument family recognition score with semitone mappings. The decrease was found to be significant with Smt-LF mapping. Semitone mapping is based upon modifying the frequency allocation compared to the Std mapping of the ACE strategy and uses different numbers of points in the FFT frames and the overlap (Omran et al. 2011). Because there were no changes to the specific coding of temporal information for all three mappings, the Semitone mappings effectively changed the spectral density representations compared to the Std mapping. Thus, the CI recipients may have been strongly relying on the PSD of signals as suggested by (Drennan and Rubinstein 2008) for identifying the instruments. Although the difference in the window size (number of points) used in Smt-MF and Smt-LF compared to Std, (512 versus 128) and the additional subband decomposition of Smt-LF, improves the frequency resolution with Semitone mapping. It also decreases the temporal resolution. Furthermore, the Std mapping covers a range from 188-7980 Hz, while Smt-LF and Smt-MF cover the frequency ranges from 130-1502 Hz and 440-5009 Hz respectively. Since the Std mapping has a wider input frequency range than the Semitone mappings, the average encoded spectrum will be greater than with either Semitone mappings. Thus, the larger spectral representation as well as the CI recipients' familiarity with the Std mapping are other likely reasons for its superior performance in the IR test. This also highlights the importance of training as well as the need to encode appropriate cues for specific purposes (temporal fine structure in this case for timbre perception). An additional reason may be the harmonic relationship of frequency components in an instrument sound. The more harmonic structure an instrument has, the better recognition with semitone mapping especially Smt-MF is expected to occur. Instrument recognition may be dependent on the energy per octave. Furthermore, the observation that Smt-MF performed better than Smt-LF could have been due to the effective transposition to a higher pitch range that occurs with Smt-LF mapping. The resultant sounds were commented by CI recipients as being unnaturally high pitched and unpleasant, making it more difficult for them to distinguish and identify the instruments.

5.2 Conclusion

Semitone mapping preserves the representation of the harmonic structure of overtones. This is expected to improve melody recognition with CI recipients. This mapping may introduce pitch reversals with tones when high and low frequency partials are filtered out. DDM was used successfully to determine diagonal activities in matrices. It was employed to evaluate the amount of harmonic structure preservation with different mappings. The results showed that 43 channels is expected to preserve the harmonic structure better than 22 channels with Smt-LF mapping. It showed also that the Smt-LF mapping highly ameliorates harmonic structure preservation of overtones in musical notes with respect to Smt-MF mapping which is higher inturn than the Std mapping. Pitch discrimination tests with NH subjects using Std mapping incorporating VCs showed no significant benefit of VCs. A harmonicity index was proposed and utilized in measuring the amount of harmonic overtones preservation in processed musical tones with Std, Smt-LF and Smt-MF mapping. The results showed that at least one Semitone mapping range produces higher preservation than Std mapping with many musical tones from different instruments.

Pitch discrimination and MCI (Galvin et al. 2007) tests showed that there was an improvement with Semitone mapping over Std mapping. The pitch discrimination results support the hypothesis that better preservation of the harmonic structure through semitone mapping will improve the discriminability of complex tones. Similarly, the hypothesis that this improvement in discrimination can be applied to a more complex task such as MCI appears to be also justified. However, the frequency limits of both Smt-LF and Smt-MF can produce difficulties when not all partials of complex tones are present. This is more likely to occur when the tones have partials close to the frequency limits of either semitone mappings. The improvement differed between Smt-MF and Smt-LF. Although Smt-LF mapping provided better pitch discrimination and melody identification results, the perceived sounds were much higher in pitch and some CI recipients did not like it. Smt-MF maps the tones closer to their natural characteristic frequencies and probably sounded more natural than Smt-LF. The Instrument recognition test showed a significant enhancement with Clarinet using Smt-MF but in general revealed a significant decrease in average scores with Semitone mapping. The results illustrate that Semitone mapping alone is not sufficient to improve instrument recognition of all instruments. Temporal fine structure information, which is also important to discriminate timbre (and hence identify instruments), is not explicitly coded in Semitone mapping, and needs to get incorporated in future developments of coding strategies intended to present music. The benefits of semitone mappings were significant in simulations but were not significant in CI recipients with the MCI test. Long term familiarization with the new mappings and use of VCs may be necessary before significant benefits in CI users can be observed.

Chapter 6 – Future Perspectives

Chapter 6 – Future Perspectives

6.1 Electric and Normal Hearing

CIs were originally aimed at restoring speech perception for patients with profound hearing loss. Although CI recipients report around 80% of speech comprehension (Mueller-Deile 2009) with Oldenburg sentence tests (Wagener et al. 1999a; Wagener et al. 1999b; Wagener et al. 1999c), many postlingually CI patients report that music is not well perceived while others enjoy it (Gfeller and Lansing 1991; Gfeller and Knutson 2003). This may depend on the listening habits prior implantation and on training after implantation (Gfeller et al. 1997a; Gfeller et al. 2000a; Gfeller et al. 2000b; Gfeller et al. 2002a; Gfeller et al. 2002b). The harmonic structure of music is not well preserved by the current Std ACE mapping strategies due to the limited number of frequency bands (Drennan and Rubinstein 2008). Additionally, information found in the temporal FS is largely ignored (Wilson et al. 2005).

The Freedom Nucleus Cochlear Implant provides 22 electrodes. The electrode spacing x between n successive electrodes in the Nucleus implant for the contour electrode is non-uniform and ranges between 0.57 and 0.81mm (Eisen and Franck 2005). Stimulating an electrode expended electric field which excites related large groups of ganglion cells. SOE is the effective width in mm of the electric field along the BM. Several studies estimated the average width of stimulation in CI recipients psychoacoustically (Laneau and Wouters 2004; Laneau et al. 2006a). Laneau et al. (2006) found that the average width of stimulation is around 1.3 mm in CI recipients. Other researches used finite element modeling to calculate the SOE (Tognola et al. ; Frijns et al. 1995; Hanekom 2001).

To improve the representation of spectral components, VCs were previously examined which are obtained by stimulating two electrodes either simultaneously (Busby and Plant 2005) or sequentially (McDermott and McKay 1994) with same or different current levels (Firszt et al. 2007). Although VCs may generate an intermediate pitch perception they do not reduce the effective field (SOE).

Simulation experiments with AMO using NH subjects showed that for 80% correctness of melody perception depending on the spectrum, more than 40 independent channels are required (Smith et al. 2002).

6.2 Hypothesis

- SOE width may have to decrease in future devices (i.e fewer nerve fibers should be stimulated with each electrode).
- Sounds perceived through electric hearing with “many” electrodes will be similar to those perceived by NH subjects.

6.3. Methods and Procedures

This section illustrates an analytical approach to compare between acoustic and electric hearing from a transfer functions’ perspective. In fact, there is much evidence that CI recipients do not perceive sound as well as NH subjects do (Kong et al. 2004b). Normal ears provide much higher frequency and temporal resolutions of perceived sounds than with hearing with CIs. NH subjects hear due to motion of hair cells inside the cochlea. Whereas CI recipients hear due to stimulating group of auditory nerve fibers with electric fields (Miller

and Spelman 1989). The electric fields stimulate the auditory nerve with an exponential decay distribution along the BM (Black and Clark 1980). Helmholz suggested that the ear resolves sounds using filter-banks (Helmholtz 1954). Later in 1972, Johannesma suggested a better representation of auditory filters; GTF which were later implemented by de Boer (1975) in the following section (Johannesma 1972; de Boer 1975). A relationship between Gammatone and exponential decay filter-banks found in NH and CI recipients respectively will be derived.

6.3.1 Gammatone Filter-Banks

The Gammatone function is a result of collecting data through physiological impulse-response measurements from primary auditory fibers in cat (Johannesma 1972). The GTF was first suggested by de Boer (1975) (de Boer 1975). The envelope of the impulse response is approximated by a statistical Gamma distribution. FS of the impulse response is a sinusoid at the center frequency of the filter bank. The impulse response is described as follows:

$$g(t) = At^{N-1}\exp(-2\pi bt)\cos(2\pi f_c t + \phi) \quad t\geq 0,\ N\geq 1 \quad \text{(Slaney 1993)} \tag{6.1}$$

where
A: is the filter's amplitude used for normalization
b: Bandwidth of the Gammatone filters in the filter-bank, as described by Equation (6.2).
t: Spatial tonotopical representation in mm from the cochlear apex.

$$b(fc) = 1.019 * 24.7 * (1 + 4.37\frac{f_c}{1000}) \tag{6.2}$$

where,
f_c: Center frequency of the filter

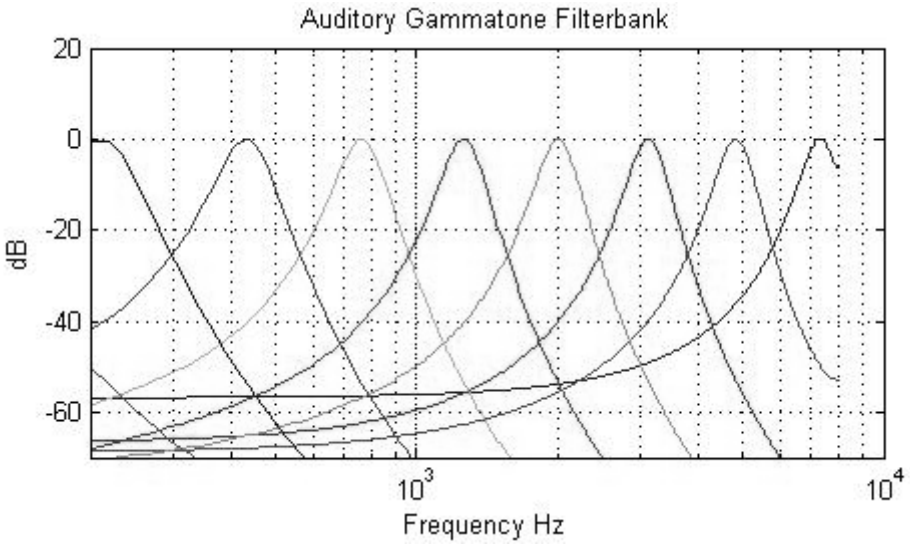


Figure (6.1): Gammatone auditory filter bank plotted for nine selected channels upto 8 kHz.

6.3.2 Exponential Decay Filter-Banks

The ear delivers auditory information through the 8th cranial (auditory) nerve to the brain. For profound hearing subjects, CIs are used to stimulate the auditory nerve by means of an electric field. This electric field is produced due to electric current flowing from an active electrode to a ground through conductive tissue to a reference electrode. For bipolar stimulation, both electrodes are placed within the cochlea whereas for monopolar stimulation the reference is

placed mostly in or below the muscle. Bipolar stimulation is the most common one (Miller and Spelman 1989). The electric narrow band field stimulates different tonotopical positions along the BM producing the percept of a noise signal.

The exponential decay filter *h(x)* is described by Equation (6.3)

$$\ln(h(x)) = -b * \left|(x - x_0 - cochlength\ )\right| \qquad \text{(Laneau et al. 2006a)} \qquad (6.3)$$
$$\Rightarrow h(x) = e^{-\left|(x - x_0 - cochlength\ )\right| / \lambda}$$

where,
cochlength: length of cochlea (~ 33 mm)
x: insertion depth of the implant (~ 22 mm)
x_0: electrodes positions in the cochlea at which filters are calculated in mm from the basal end
$\lambda = 1/b$: with of stimulation of electric field (typically in the order of 1.3 mm)

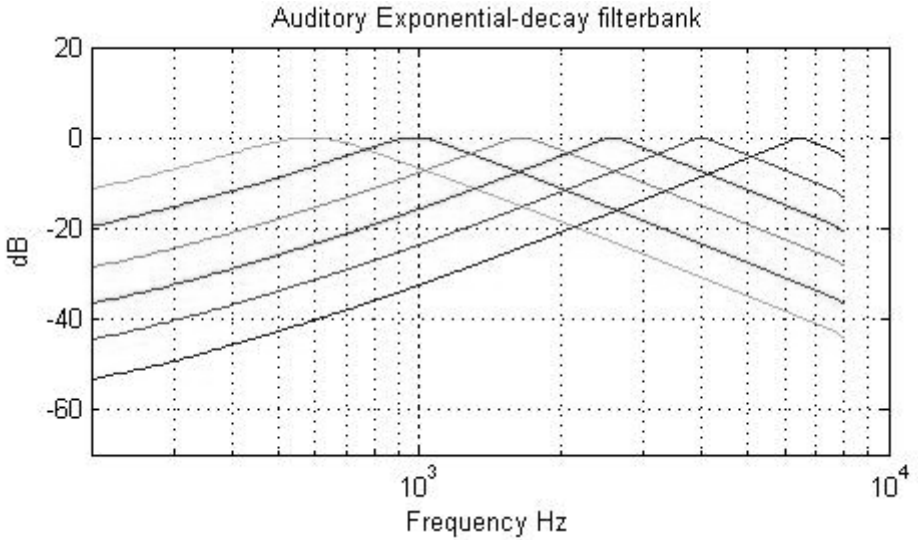


Figure (6.2): Exponential-decay filter-bank plotted for selected channels with Greenwood center frequency similar to center frequencies of plotted Gammatone channels.

Calculating the Fourier transform *H(f)* of the exponential decay spatial filter *h(x)* produces:

$$H(f) = \int_0^\infty h(x).e^{2\pi fx}dx = \int_0^\infty e^{-b(x+x0-cochlength)}e^{-j2\pi fx}dx \qquad (6.4)$$

6.3.3 Analysis

Assuming that sounds perceived by electric hearing are equivalent to those perceived by NH, an unknown missing filter bank M(f) that can be introduced represents required frequency shaping for electric hearing pathway.

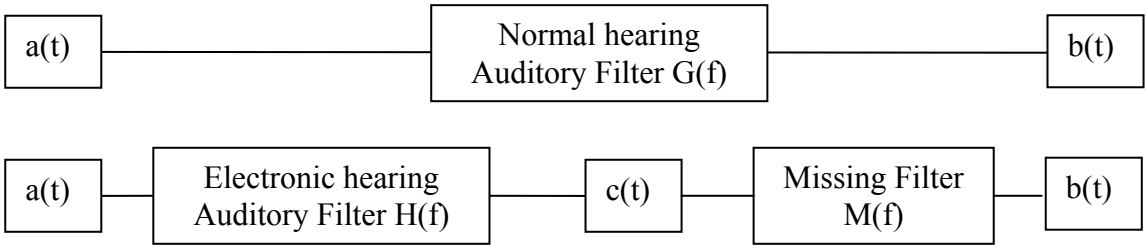


Figure (6.3): Block diagram of normal and electric hearings.

An acoustic sound a(t) causes perception of a signal b(t) in NH subjects, and perception of a signal c(t) in CI recipients. Assuming that for CI there are still some missing processing functions to be optimized M(f) filter is introduced. If a M(f) filter were to exist, CI recipients would be able to perceive a normal sound b(t) as described by Figure (6.3).

In this section, the missing function M(f) will be calculated based on the relationship between a Gammtone and exponential-decay filter-bank.

B(f)=A(f).G(f)
C(f)=A(f).H(f)
B(f)=C(f).M(f)

From Equation (6.4), the missing filter M(f) can be derived as

$$M(f)=\frac{B(f)}{C(f)}=\frac{A(f).G(f)}{A(f).H(f)}=\frac{G(f)}{H(f)} \tag{6.5}$$

From Equation (6.5) the M(f) is the ratio between the NH auditory filter-bank and the electric hearing filter-bank.

In order to find the missing filter M(f), the frequency transform of the GTF g(t) described by Equation (6.1) is calculated using the Laplace transform instead of direct integration.

$$G(s)=\ell(g(t))=\ell\left\{At^{N-1}\exp(-2\pi bt)\cos(2\pi f_c t+\phi)\right\}$$
$$G(s)=A*\ell\left\{t^{N-1}\exp(-2\pi bt)\cos(2\pi f_c t+\phi)\right\}$$

Using the Laplace property $\ell\left\{t^n f(t)\right\}=(-1)^n F^{(n)}(s)$

$$G(s)=(-1)^{n-1}.\left[\ell\left\{e^{(-2\pi bt)}.\cos(2\pi f_c t+\phi)\right\}\right]^{n-1} \tag{6.6}$$

Using the shift Laplace property $\ell\left\{e^{at}f(t)\right\}=F(s-a)$

Let $f(t)=\cos(2\pi f_c t+\phi)$, a=-2πb,

Since $\ell\left\{\cos(wt)\right\}=\frac{s}{s^2+w^2}$, and with a phase φ shift, it is described as follows:-

$$\ell(\cos(wt+\phi))=\int_0^\infty e^{-st}\cos(wt+\phi)dt$$

$$\ell(\cos(wt+\phi))=\frac{1}{2}\int_0^\infty e^{-st}\left\{e^{j(wt+\phi)}+e^{-j(wt+\phi)}\right\}dt$$

$$\ell(\cos(wt+\phi))=\frac{1}{2}\left[\frac{e^{-st+jwt+j\phi}}{-s+jw}+\frac{e^{-st-jwt-j\phi}}{-s-jw}\right]_0^\infty$$

$$\ell(\cos(wt+\phi))=\frac{1}{2}\left[\frac{e^{-st+jwt+j\phi}}{jw-s}-\frac{e^{-st-jwt-j\phi}}{jw+s}\right]_0^\infty$$

$$\ell(\cos(wt+\phi)) = \frac{1}{2}\left[\frac{e^{j\phi}(e^{-t(s-jw)})}{jw-s} - \frac{e^{-j\phi}(e^{-t(jw+s)})}{s+jw}\right]_0^{\infty}$$

$$\ell(\cos(wt+\phi)) = \frac{1}{2}\left[\frac{-e^{j\phi}(s+jw)+e^{-j\phi}(jw-s)}{-w^2-s^2}\right]$$

$$\ell(\cos(wt+\phi)) = \frac{-1}{2}\left[\frac{-s(e^{j\phi}+e^{-j\phi})-jw(e^{j\phi}-e^{-j\phi})}{w^2+s^2}\right]$$

$$\ell(\cos(wt+\phi)) = \frac{-1}{2}\left[\frac{-2s\cos\phi - w(2\sin\phi)}{w^2+s^2}\right]$$

$$\ell(\cos(wt+\phi)) = \frac{s\cos\phi - w\sin\phi}{w^2+s^2} \tag{6.7}$$

Substituting Equation (6.7) in to Equation (6.6)

$$\ell\{\cos(2\pi f_c t+\phi)\} = \frac{s\cos\phi - 2\pi f_c \sin\phi}{s^2+(2\pi f_c)^2}$$

$$\ell\left\{e^{at}\cos(2\pi f_c t+\phi)\right\} = \frac{(s-a)\cos\phi - 2\pi f_c \sin\phi}{(s-a)^2+(2\pi f_c)^2}$$

Using Equation (6.1) the Laplace transform of a Gammatone filter-bank is:-

$$\mathrm{G(s)} = (-1)^{n-1}.\left[\frac{(s+2\pi b)\cos\phi - 2\pi f_c \sin\phi}{(s+2\pi b)^2+(2\pi f_c)^2}\right]^{n-1} \tag{6.8}$$

Equation (6.8) represents the Laplace transform of GTF. Replace each s with j2πf to get the Fourier transform. Since auditory filters could be illustrated with fourth order Gammatone filter bank (Slaney 1993), therefore substitute n=4 in Equation (6.8).

$$\mathrm{G(f)} = -1.\left[\frac{(j2\pi f+2\pi b)\cos\phi - 2\pi f_c \sin\phi}{(j2\pi f+2\pi b)^2+(2\pi f_c)^2}\right]^{3} \tag{6.9}$$

Finding the transfer function H(s) of the electric hearing filter bank is as follows:-

From Equation (6.3) $h(t) = e^{-b_1*(t+x0-cochlength)}$

$$H(s) = \ell\left\{e^{-b_1 t}.e^{-b(x0-cochlength)}\right\} = \frac{e^{-b_1(x0-cochlength)}}{s+b_1} \tag{6.10}$$

The missing filter M(s)=H(s)/G(s). From Equations (6.8 and 6.9)

$$M(s) = \frac{H(s)}{G(s)} = (-1)^{n-1}.\left[\frac{(s+2\pi b)\cos\phi - 2\pi f_c \sin\phi}{(s+2\pi b)^2+(2\pi f_c)^2}\right]^{n-1}.\left[\frac{s+b_1}{e^{-b_1(x0-cochlength)}}\right]$$

Transforming the filter M(s) into the time domain:-

$$M(s) = \frac{(-1.\cos\phi)^{n-1}}{e^{-b_1(x0-cochlength)}}.\left[\frac{(s+2\pi b)-2\pi f_c \tan\phi}{(s+2\pi b)^2+(2\pi f_c)^2}\right]^{n-1}.[s+b_1]$$

Let $Amp = \frac{(-1.\cos\phi)^{n-1}}{e^{-b_1(x0-cochlength)}}$ where n=4 for a four pole Gammatone filter (Slaney 1993)

$$M(s) = Amp.\left[\frac{(s+2\pi b) - 2\pi f_c \tan\phi}{(s+2\pi b)^2 + (2\pi f_c)^2}\right]^3 .[s+b_1]$$

Let X= $2\pi b - 2\pi f_c \tan\phi$, Y= $2\pi b$ and Z= $2\pi f_c$

$$M(s) = A.\frac{(s+X)^3.(s+b_1)}{[(s+Y)^2 + (Z)^2]^3}$$

$$M(s) = A.\left[\frac{(s+X)^2}{(s+Y)^2 + (Z)^2}\right].\left[\frac{(s+X)}{(s+Y)^2 + (Z)^2}\right].\left[\frac{(s+b_1)}{(s+Y)^2 + (Z)^2}\right]$$

Using symbolic reduction with Matlab symbolic toolbox, the inverse Laplace transforms of the missing filter *M(s)* in the time domain would be as follows:-

$$m(t) = A.e^{-2\pi b.t}\left[B*\sin(2\pi f_c t) + t\left(C*\sin(2\pi f_c t) + E*\cos(2\pi f_c t)\right) + t^2\left(D*\sin(2\pi f_c t) + F*\cos(2\pi f_c t)\right)\right]$$

$$m(t) = A.e^{-2\pi b.t}\left[B*\sin(2\pi f_c t) + t.\sin(2\pi f_c t + \tan^{-1}\frac{E}{C}) + t^2\sin(2\pi f_c t + \tan^{-1}\frac{F}{D})\right] \quad (6.11)$$

where,

A= $\frac{-1}{16}\frac{\cos^3\phi}{(\pi f_c)^2}e^{b_1(x0-l)}$

B= $(6 - 3b_1 + 6b\pi)$

C= $(-12b\pi^2 - 20\pi^2 f_c^{\,2} + 6b_1\pi f_c)$

D= $-4b_1\pi^2 f_c^{\,2} + 8\pi^3 f_c^{\,3} + 8\pi^3 b f_c^{\,2}$

E= $4\pi^2 f_c^{\,2} + 6b_1\pi f_c - 12b\pi^2 f_c$

F= $-8\pi^3 f_c^{\,3} + 8\pi^3 b f_c^{\,2} - 4b_1\pi^2 f_c^{\,2}$

$b(fc) = 1.019*24.7*(1 + 4.37\frac{f_c}{1000})$ is from NH Gammatone parameter

fc is the center frequency of used filters

x0: position of the electric stimulation filter on the cochlea starting from the apex in mm.
l: total cochlea length in mm (~32 mm).
fc: center frequency of the filter and is calculated from the greenwood function at x=x0.
b1: reciprocal of electric hearing stimulation width (SOE) (λ) (~0.3 mm^{-1})

6.4 Results

In the previous section, the missing filter was determined as expressed in Equation (6.10). This filter bank represents missing extra processing in the electric hearing. Each filter in this filter bank has different band widths at different frequencies.

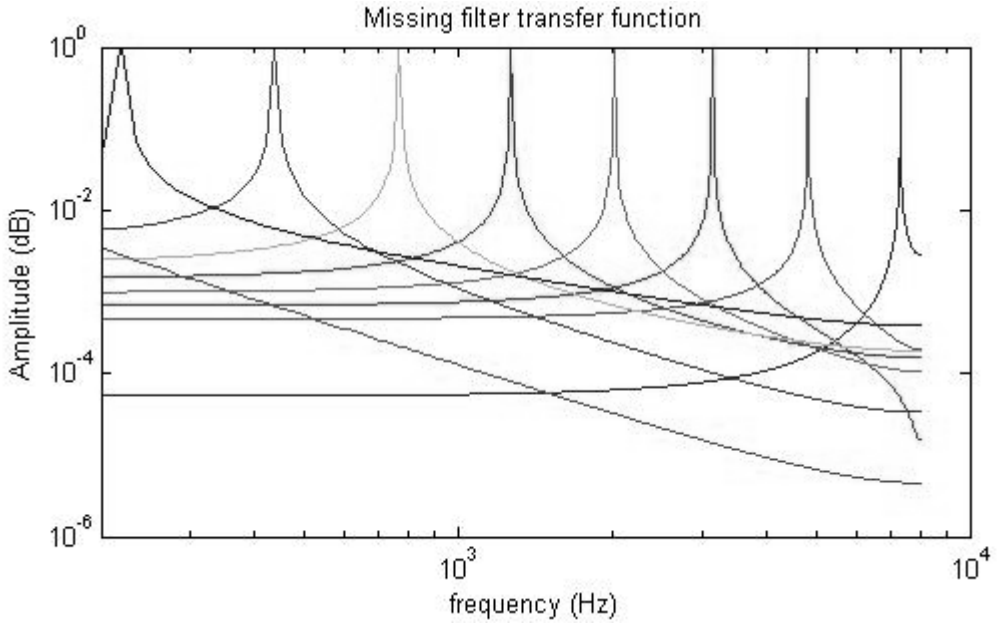


Figure (6.4): Missing filter-bank M(f) plotted for selected channels having Greenwood center frequencies identical to center frequencies of plotted Gammatone channels.

Figure (6.4) shows the missing filter-bank M(f) defined in Equation (6.10). It illustrates a very sharp selective filter-bank processing that would be required in order to generate normal auditory filter selectivity stimulation with CI devices. SOE of each filter is calculated assuming that the ganglion cells are located at points where there is a -3 dB reduction from the peak amplitude of each filter. The resulting calculated bandwidth is shown in Figure (6.5).

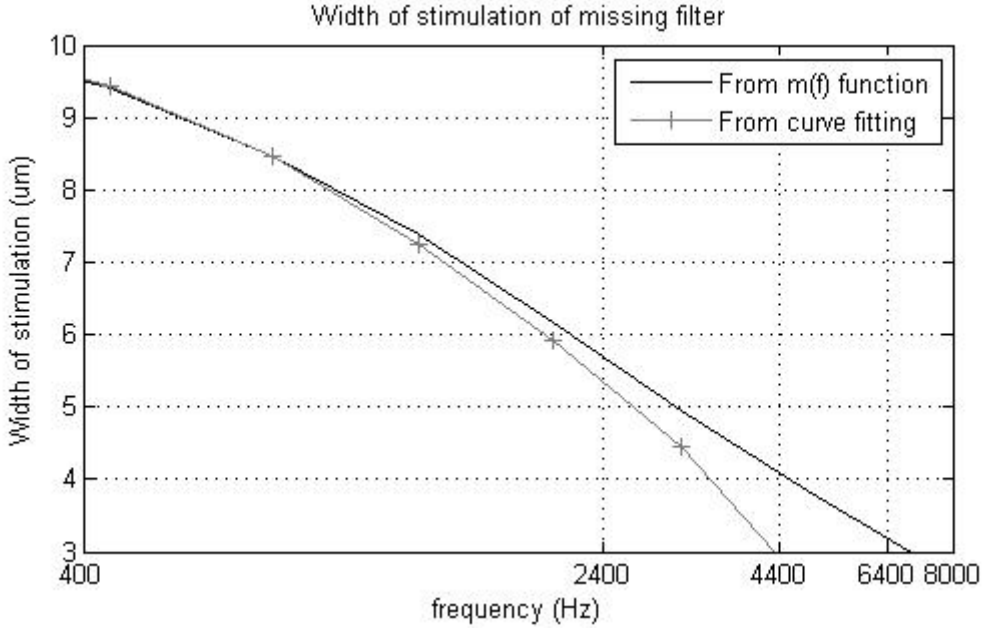


Figure (6.5): Width of stimulation of ideal electric hearing filter-banks to provide a perception almost similar to normal hearing Gammatone filter-banks.

Figure (6.5) shows the SOE assuming it is equal to the bandwidth of the missing filters, calculated with respect to the center frequency of each channel as shown in Figure (6.4). This missing filter-bank has wider width of stimulation at low frequencies (~9.5 μm) than at high frequencies (~3 μm). The curve was fitted with a cubic polynomial in Equation (6.12)

$$y = (8.1*10^{-18}).f^4 + (-1.5*10^{-13}).f^3 + (10^{-9}).f^2 + (-4*10^{-6}).f + 0.011 \quad (6.12)$$

Currently, CI recipients have an average given width of stimulation of about 1.3 mm. Figure (6.5) shows that the width of stimulation may have to be improved in future devices more by a than a factor of 100 and decrease to as little as9 μm and around 3 μm at low frequencies (400 Hz) and high frequencies (6400 Hz) respectively.

6.5 Conclusion

This theoretical analysis compared between acoustic and electric hearing to estimate necessary improvements in frequency selectivity for future devices. It illustrated that the SOE of CI devices may have to decrease from an average of 1.3 mm as measured psychoacoustically, by more than two orders of magnitude. The results here are calculated from an analytic perspective only. SOE is dependent on technical and biological parameters. This study shows that there may be a benefit in improving the physical filter and shape of electrodes and their proximity to neural elements.

Appendix – A (Electrode Schematics)

A.1 Electrodes schematics

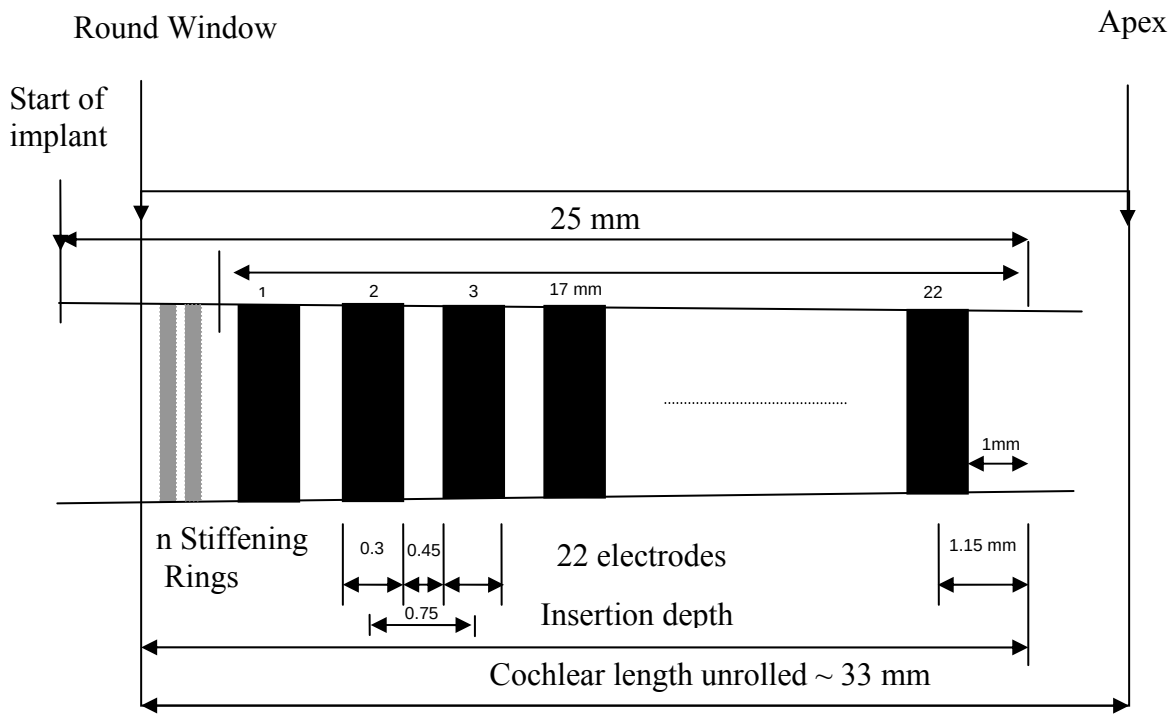


Figure (A.1): Illustrating distances on the Nucleus 24 straight array implant.

To calculate the Greenwood tonotopical place to frequency pitch for normal hearing use the following function: Greenwood_x2cf.m

```
function cf = Greenwood_x2cf(x)
%Calculates the center frequencies corresponding to position
%along the Basilar Membrane (BM). Based on Greenwoods formula :-
%
% cf = Greenwood_x2cf(x);
%
% x is the distance from the apical end of the BM in mm
% cf is the resulting frequency in Hertz
% define constants of Greenwoods formula
A=165.4;
a=0.06;
k=1;

cf = A * (10.^(a*x) - k);
return;
```

Figure (A.2): Matlab function to convert positions along the BM into its tonotopical frequency representation according to Greenwood (Greenwood 1990; Greenwood 1991).

Ex 1: Insertion depth = 25 and cochlear length = 33
First basal electrode at = 33-25 +1+16= 25 mm
Greenwood frequency of first basal electrode = f(33-25+1.15+16)= 5.17 k
Greenwood frequency of first apical electrode =f(33-25+1.15 mm) = 420 Hz

Ex 2: Insertion depth = 20, Cochlear length = 33 mm
Highest frequency = f(33-20+1.15+16) = 10.4 kHz
Lowest frequency = f(33-20+1.15) = 1000.2 Hz

Appendix – B (Bark GUI)

B.1 BarkGUI

BarkGUI is a Matlab utility used to calculate bin-to-bands distribution for virtual channels. The virtual channel mode uses 256 points in the FFT algorithm. This produces 128 bins of 62.5 Hz. The first bin is a direct current bin that is not used in the processing. This leaves a total of 127 bins.

In some Cochlear Implant recipients, one or more channels are disabled. A technical consequence of this is having fewer virtual channels available. The bin-to-band algorithm covers the frequency band (188-7980 Hz) in the standard ACE mapping. The bins are distributed differently when some channels are disabled. The number of available channels has to be specified according to a patient's map.

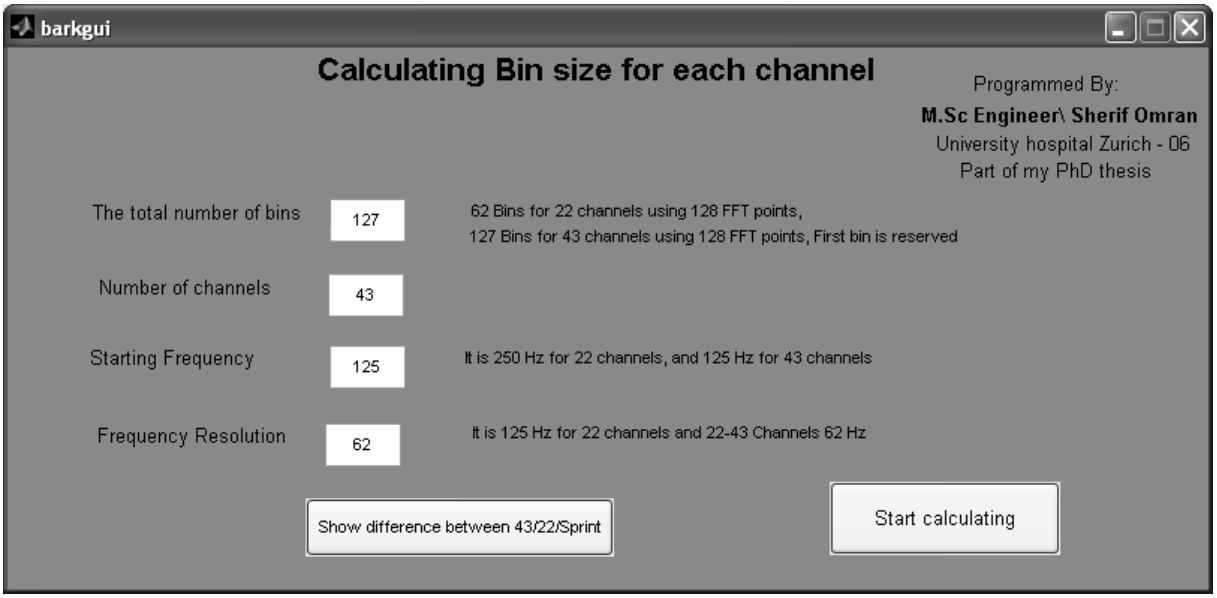


Figure (B.1): BarkGUI utility graphical user interface.

B.1 Implementation of 43 channels

The NMT as described before is a toolbox written by Cochlear Ltd., to be used with Matlab. The toolbox supports 22 channels only but with some adjustments it can be extended to 43 channels.

The bin size is calculated for each of the 43 channels as described in section 2.1.1 using the BarkGUI to produce all related graphs. The corresponding bin sizes for channels from 22-43 are shown in the Table (B.1) in the syntax form of Matlab matrices. Notice that for channel number 22, there are two values (22 and 22+), the first is used in the setting of the Sprint device provided by Cochlear and the second is the optimum value to maximize the number of used bins (see section 2.1.1 for more details).

Channels	Bin sizes
43	[1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,4,4,4,4,5,5,5,6,6,6,7,7,8,8];
42	[1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,4,4,4,4,5,5,5,6,6,6,7,7,8,9];
41	[1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,4,4,4,4,5,5,5,6,6,7,7,8,8,9];
40	[1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,4,4,4,4,5,5,5,6,6,7,7,8,8,9];
39	[1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,4,4,4,4,5,5,6,6,6,7,7,8,8,9];
38	[1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,4,4,4,4,5,5,6,6,6,7,7,8,9,9];
37	[1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,3,4,4,4,4,5,5,6,6,7,7,8,8,9,10];
36	[1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,3,4,4,4,5,5,5,6,6,7,7,8,8,9,10];
35	[1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,4,4,4,5,5,5,6,6,7,7,8,9,9,10];
34	[1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,4,4,4,5,5,5,6,6,7,7,8,9,10,10];
33	[1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,4,4,4,5,5,5,6,6,7,8,8,9,10,11];
32	[1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,4,4,4,5,5,5,6,6,7,8,8,9,10,11];
31	[1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,4,4,4,5,5,5,6,7,7,8,9,9,10,11];
30	[1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,4,4,5,5,5,6,7,7,8,9,10,10,11];
29	[1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,4,4,5,5,6,6,7,7,8,9,10,11,12];
28	[1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,4,4,5,5,6,6,7,7,8,9,10,11,12];
27	[1,2,2,2,2,2,2,2,2,2,2,3,3,3,3,4,4,5,5,6,6,7,8,8,9,10,11,13];
26	[1,2,2,2,2,2,2,2,2,2,3,3,3,3,4,4,5,5,6,6,7,8,8,9,10,12,13];
25	[1,2,2,2,2,2,2,2,2,3,3,3,3,4,4,5,5,6,6,7,8,9,10,11,12,13];
24	[1,2,2,2,2,2,2,2,3,3,3,3,4,4,5,5,6,6,7,8,9,10,11,12,14];
23	[1,2,2,2,2,2,2,3,3,3,3,4,4,4,5,6,6,7,8,9,10,11,13,14];
22+	[1,1,1,1,1,1,1,1,2,2,2,2,2,3,3,4,4,5,5,6,7,7];
22	[1,1,1,1,1,1,1,1,1,2,2,2,2,3,3,4,4,5,5,6,7,8];
Cochlear	From Cochlear Ltd., builtin Sprint device

Table (B.1): Bin sizes for channels from 22-43.

Example using 43 channels withing NMT as follows:-

```
p.block_length=256;
p.num_bins=256;
p.num_bands=43;
p.electrodes=[43:-1:1]';
p.threshold_levels=ones(43,1)*80;       %T-Levels set at 80
p.comfort_levels=ones(43,1)*120;        %C-Levels set at 120
p=ACE_map (p);
q=Process(p,'asa');                      %Output is the pulse sequences.
```

Appendix – C (AMO and Checker)

C.1 Acoustic Model

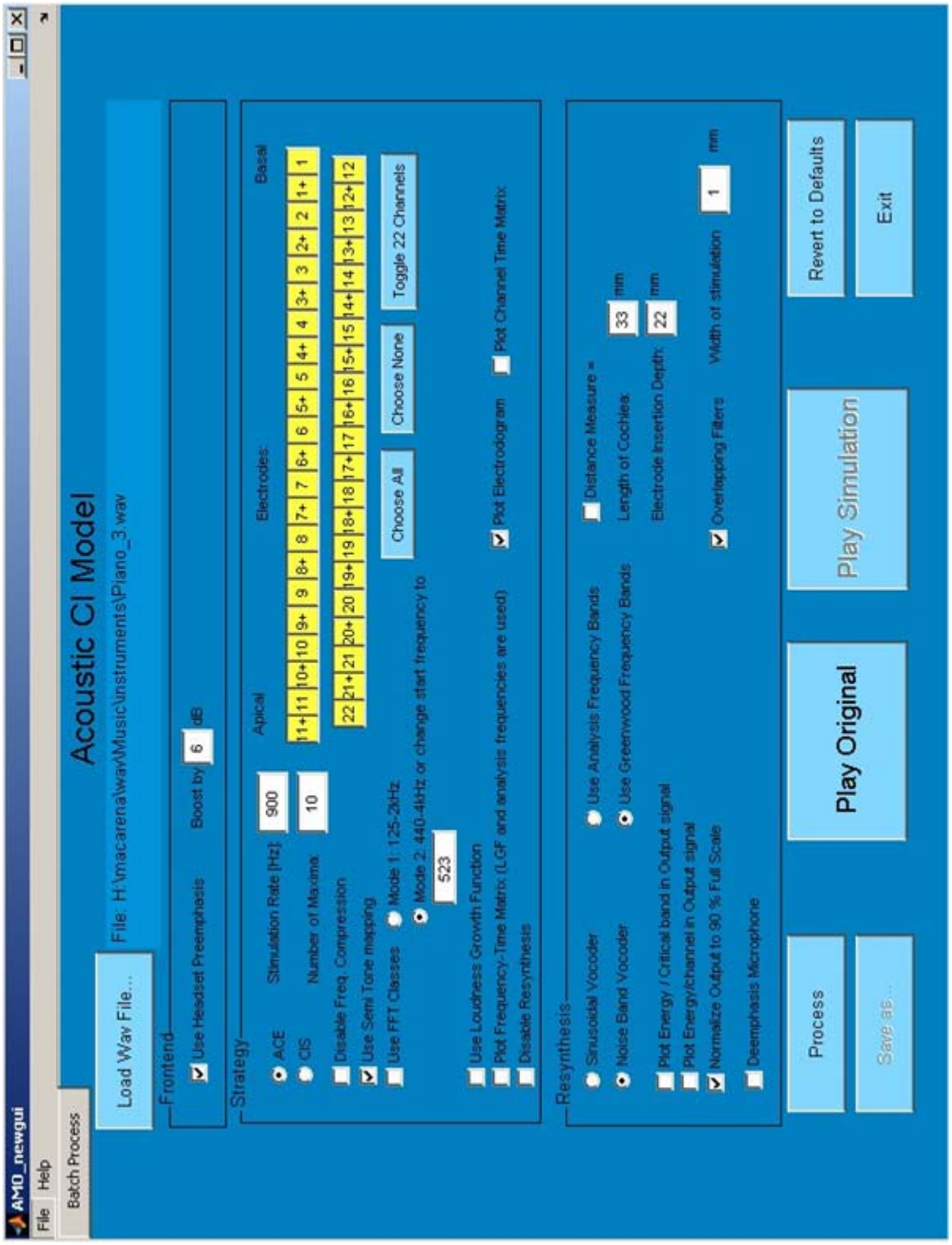


Figure (C.1): The acoustic model; a cochlear implant simulator

Features:

- Process wave files similarly to a Cochlear Implant process
- Resynthesizes channel activities with either sinusoidal or noise vocoder
- Support 43 channels mode
- Support CIS strategies with 22 channels
- Implemented mapping algorithms are the standard ACE, Greenwood and Semitone mapping
- Support varying start frequency of Smt-MF (440 Hz is the default value)
- Produces many plots (electrodogram, channel time matrix, frequency time matrix, energy of resynthesized signals per critical band or per channel)
- Supports loudness growth function
- Supports headset pre-emphasis filter
- Supports Spread of excitation (width of stimulation) variations
- Supports Batch processing of files; useful for automated overnight processing

C.2 Checker

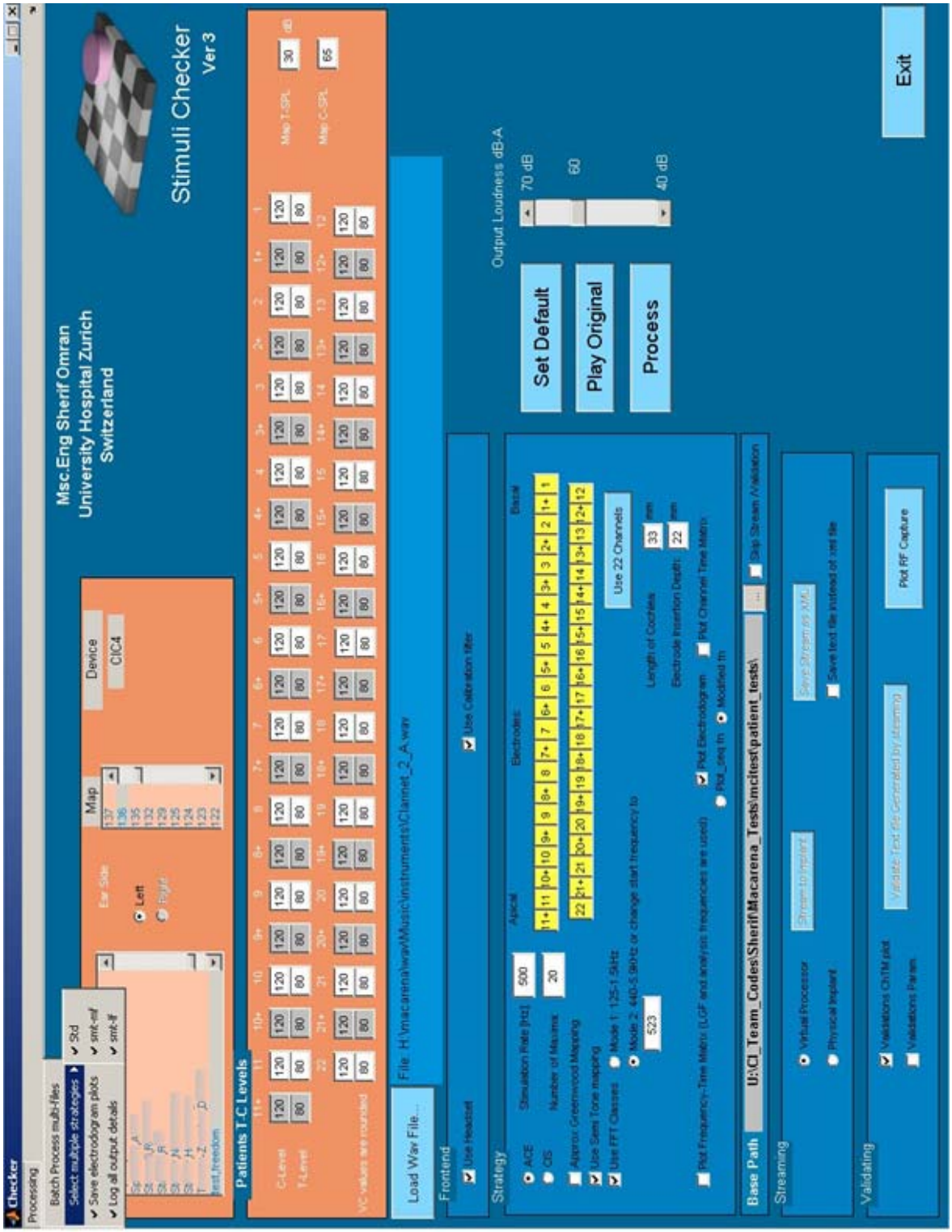


Figure (C.2): The graphical user interface of the Checker utility.

Features

- loads patient data from a clinical database
- loads implanted ear side for the selected patient
- loads different map numbers and device type
- loads map settings, stimulation rate, T-Levels and C-Levels
- Supports 43 channels mode
- Interpolates T-Levels and C-Levels of 22 channels to get values for virtual channels
- Process selected wave file(s) similarly as a Cochlear Implant processor
- Support CIS strategies with 22 channels
- Implemented mapping algorithms are standard ACE, Greenwood and Semitone mappings
- Supports varying start frequency of Smt-MF (440 Hz is the default value)
- Produces many plots (electrodogram, channel time matrix, frequency time matrix, energy of resynthesized signals per critical band or per channel)
- Supports enable/disable of loudness growth function
- Supports headset pre-emphasis filter
- Supports batch processing of files, useful for automated overnight processing, includes saving an output report and/or electrodograms outputs
- Electrodograms are calibrated to dBA values.
- Streams data to physical implant or virtual processor (for offline files production)
- Supports production of 43 channels XML files
- Can plot streamed txt files only; does not support XML files electrodograms plotting
- Validates different types of errors

C.2.1 Rules

Pulse sequence and other parameters are validated. The Checker follows a systematic validation process according to some rules as follows:-

The intensity of each channel must be the range [1-254] current levels. Current level equal to 0 is a power up pulse. Each channel activity must lie between the T-Level and C-Level according to the patient's map. The T-Levels and C-Levels for each electrode are the minimum and maximum current intensities perceived as soft and loud respectively.

The number of channels picked from each temporal window is the number of maxima. This number is calculated from the sequence file; since channels in each temporal frame are sorted by the channel number. Channel sorting can be changed to be either increasing or decreasing. Increasing means that channels are activated from the basal to apical direction and decreasing is from the apical to basal direction. The Checker does not support random (unsorted) channel activity. The number of maxima calculated from the sequence file must match the number of maxima used in processing the signal.

Stimulation rate is calculated from the stimuli file using Equation (3.2) and is compared to the stimulation rate used in processing.

$$\text{Stimulation rate} = \text{Number of Maxima} / \text{average pulse period} \qquad (0.1)$$

The CTM is reconstructed from the stimuli file and is compared to a previously stored copy before being converted into a stimuli file.

The Freedom Implant is called a "*CIC4 type*" implant while previous versions were called "*CIC 1-3*" implants. Both types have three mono-polar stimulation modes MP1, MP2 and MP1+2 represented in Matlab as 101, 102 and 103 respectively. The stimuli are stored first in a text file called "Stimuli file". CIC3 implants decode values 24, 25 and 30 as MP1, MP2 and MP1+2 respectively, while CIC4 implants decode numbers 24, 25 and 28 as MP1, MP2 and MP1+2 respectively. CIC4 implants decode the value 30 as a double electrode mode, in which adjacent electrode from the apical direction is excited simultaneously. Due to a patient protection scheme, using CIC3 implants is disabled in the Checker. These values are described in Table (3.1).

Implant Type	Stimulation mode	Matlab	Stream
CIC3	MP1	101	24
	MP2	102	25
	MP1+2	103	30
CIC4	MP1	101	24
	MP2	102	25
	MP1+2	103	28
	Double Electrodes	104	30

Table (C.1): Implant types with their modes as represented stimuli files.

Additional to examining different parameters, the Checker plots a reconstructed version of the CTM side by side to another plot of the CTM before streaming. This visual comparison is for double checking. Figures (C.3 and C.4) show a flow chart of the Checker.

C.2.2 Virtual Channels Work-Around

NIC is an interface layer between programming languages and the implant device as described previously. Unfortunately, it doesn't support direct stimulation of double electrodes. However, there is a work around by changing the mode and the implant type (refer to Table (C.1)).

The double electrode mode for CIC4 types and the MP1+2 mode for CIC3 both have the same value in the stimuli file. The NIC streaming function stops if any active electrode has a value greater than 22. In case of 43 channels, all the active electrodes where changed in to their physical electrode values before calling the streaming function. Flags of VCs are stored in another matrix. The Streaming function in NIC generates a stimuli text file which is used to generate a streaming xml file that can be direct upload to the speech processor. The work-around Matlab functions are included in Appendix C.

The generated stimuli file in case of VCs **must not** be used for stimulation or any further processing in case of using VCs because the modes of virtual channels are currently not properly set. It is used internally to generate the streaming XML file which the MACarena application uploads to the speech processor for testing with CI recipients.

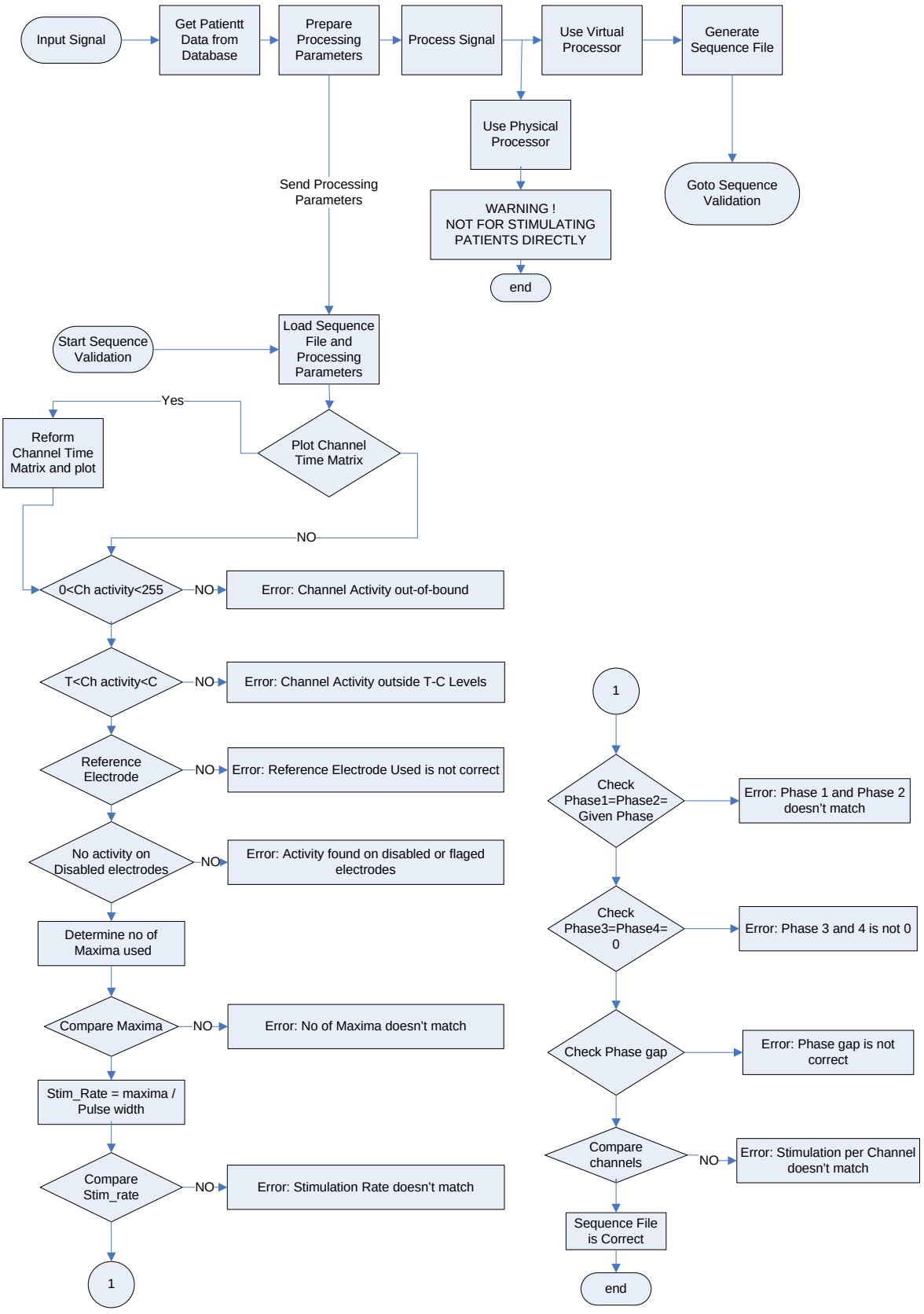


Figure (C.3): Checking and validation steps flow chart.

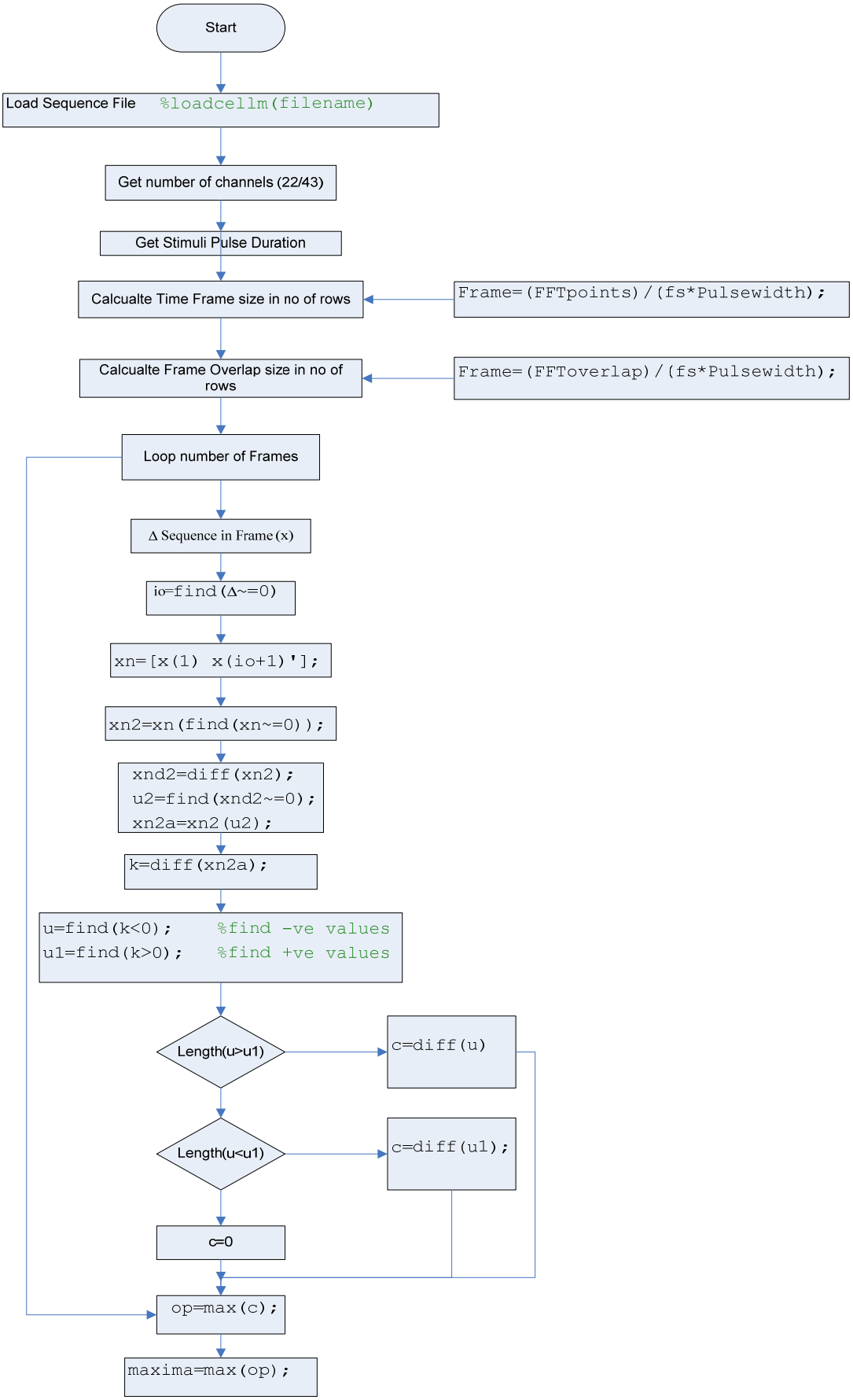


Figure (C.4): Get Number of Maxima using Sequence File.

C.2.3 Sequence File Structure

As shown in the flow chart in Figures (C.3 and C.4) many parameters are calculated from the sequence file and are checked to coincide with the given parameters.

The sequence file has the following structure:-

```
standard  5cpc    0    0   20  25.0  7.0  25.0   0.0   0.0  666.6
standard  5cpc    0    0   20  25.0  7.0  25.0   0.0   0.0  666.6
standard  5cpc    0    0   20  25.0  7.0  25.0   0.0   0.0  666.6
standard  5cpc    9   28  193  25.0  8.0  25.0   0.0   0.0  666.6
standard  5cpc   12   28  206  25.0  8.0  25.0   0.0   0.0  666.6
standard  5cpc    0    0   20  25.0  7.0  25.0   0.0   0.0  666.6
standard  5cpc   20   28  205  25.0  8.0  25.0   0.0   0.0  666.6
standard  5cpc   21   28  195  25.0  8.0  25.0   0.0   0.0  666.6
standard  5cpc   22   28  189  25.0  8.0  25.0   0.0   0.0  666.6
standard  5cpc    0    0   20  25.0  7.0  25.0   0.0   0.0  666.6
standard  5cpc    0    0   20  25.0  7.0  25.0   0.0   0.0  666.6
standard  5cpc    0    0   20  25.0  7.0  25.0   0.0   0.0  666.6
standard  5cpc    0    0   20  25.0  7.0  25.0   0.0   0.0  666.6
```

Figure (C.5): Structure of sequence files.

Where,

Column 3: Active Electrode
Column 4: Reference Electrode Mode
Column 5: Current Level
Column 6: Phase Width
Column 7: Phase Gap
Column 8: Phase 2 Width
Column 9: Phase 3 Width (not used, no phase 3)
Column 10: Phase 4 Width (not used too)
Column 11: Duration in micro sec between 2 pulses

The generated text files in conjunction with MACarena program could be streamed to CI recipients to conduct MCI test.

Parameter Validation Output

The following is an example of a parameter validation output:-

```
Pass: No intenstity >255 or <0
pass: Channels activities lies between T-C level
Pass: No activity found on the disabled electrodes
Pass: Number of maxima Correct = 8
Pass: No of channels=22 and mode is 28 Correct
pass: Stimulation rate Correct
Pass: Channel Time Matrixes are correct comparison is done pulse by pulse
```

Proposed Channel Sequence checking method

A third validation scheme could be using a sequence pulse of 100 ms signal, followed by a pause of 50 ms, frequency increased stepwise after every pause on a discrete frequency scale

that matches the center frequencies of the channels. The pause in between the signal would show the delay limitation of the implant's processor to switch on and off and the increasing frequency would show the use of each channel separately for both 22 and 43 channel modes.

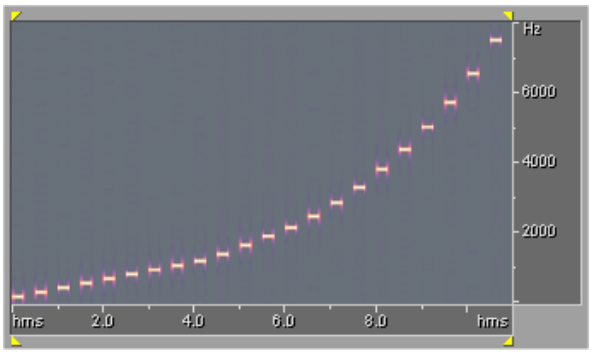


Figure (C.6): Spectrogram of the proposed checking signal.

```
-----------------------------------------------------
Parameters Validation Report
---------------------------
Pass: No intenstity >255 or <0
pass: Channels activities lies between T-C level
Pass: No activity found on the disabled electrodes
Pass: Number of maxima Correct = 6 < or = 10
Pass: No of channels=22 and mode is 28 Correct
pass: Stimulation rate Correct
Pass: Channel Time Matrixes are correct comparison is done pulse by pulse
```

Validation Plots

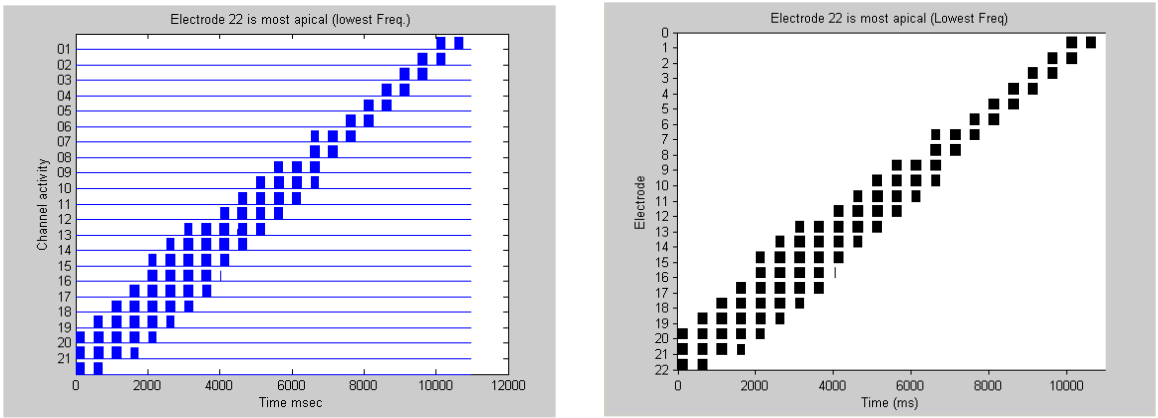


Figure (C.7): Channel time matrix from the generator (left) and the Checker (right).

Folder and File nomenclature

By clicking the save xml button in the Checker, the file and folder structure will be used. The base path could be changed using the basepath button, other wise it is set at line 4499 by default to

`U:\CI_Team_Codes\Sherif\Macarena_Tests\mcitest\patient_tests\`

```
4498    % Prepare directory Structure
4499    BasePath=['U:\CI Team Codes\Sherif\Macarena Tests\mcitest\patient tests\'];
4500    %FolderO=[ACEorCIS(handles) '_' num2str(u)];                %ACE_22
```

The main folder will be the strategy and the number of channels used ex: STD_22, SMTLF_22, or SMTMF_43. The subfolder is named by patient family name, first name and the ID number separated by underscores respectively. Next subfolder is the ear side, either left or right, followed by the map number. The structure is shown in Figure (C.8).

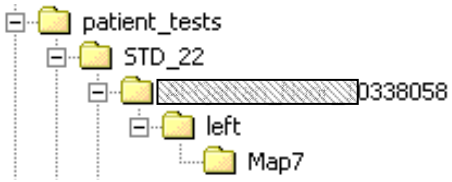


Figure (C.8): Folder structure.

The XML file name structure has 10 parts: ID number, Family name, First name, Map number, type of device, map type, number of channels, strategy and filename. All parts are separated with underscore. An example would be as follows

e.g: 10338058_Freedom_Test_left_Map126_CIC4_ACE_22_STD_Fall4a3d.xml

Checking MACarena's Output

MACarena is a playback program used to stream xml files generated by the Checker to the implant processor. Thus it was required to check that it correctly streams the given data and no packet is changed or lost due to an internal error. MACarena has an option which instead of streaming to the implant sends the output to a virtual implant whereby the electrode activity is then written to a text file. The same text file could be generated by the Checker. The output from MACarena and Checker was checked to be identical. The text files corresponding five different sound patterns (MCI contours rise, fall, risefall, fallrise and flat) were compared. The result of the comparison (DOS command screen output) is shown in the Figure (C.9).

```
H:\macarena\NIC\Test\STD_22\test_freedom_1234567\left\Map136>fc checker_risefall.txt mac_risefall.txt
Vergleichen der Dateien checker_risefall.txt und MAC_RISEFALL.TXT
FC: Keine Unterschiede gefunden

H:\macarena\NIC\Test\STD_22\test_freedom_1234567\left\Map136>fc mac_rise.txt checker_rise.txt
Vergleichen der Dateien mac_rise.txt und CHECKER_RISE.TXT
FC: Keine Unterschiede gefunden

H:\macarena\NIC\Test\STD_22\test_freedom_1234567\left\Map136>fc mac_flat.txt checker_flat.txt
Vergleichen der Dateien mac_flat.txt und CHECKER_FLAT.TXT
FC: Keine Unterschiede gefunden

H:\macarena\NIC\Test\STD_22\test_freedom_1234567\left\Map136>fc mac_fallrise.txt checker_fallrise.txt
Vergleichen der Dateien mac_fallrise.txt und CHECKER_FALLRISE.TXT
FC: Keine Unterschiede gefunden

H:\macarena\NIC\Test\STD_22\test_freedom_1234567\left\Map136>fc mac_checker_fall.txt checker_fall.txt
FC: Kann MAC_CHECKER_FALL.TXT nicht öffnen - Datei oder Ordner nicht vorhanden
```

Figure (C.9): Comparing MACarena and Checker outputs.

Testing the Checker algorithms

Verifying the Checker involves testing its algorithms against self produced errors. The following data was used among with this test.

Wave file: Singer_1.wav
Path: U:\CI_Team_Codes\Sherif\Matlab_Codes\AMO New release 43 Ch\Sounds_Original\
Patient Name: Test Freedom
Map number:136

The Checker processed the given wave file and streamed it to a virtual channel where the stimuli were written in a text file. This text file was then edited by hand to introduce errors that the Checker hat to detect.

1- Activity above 255

Three pulses were edited by to have a stimuli level greater than 255 as shown in gray in Figure (C.10).

6453	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	138.8
6454	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	138.8
6455	standard	5cpc	21	28	256	25.0	8.0	25.0	0.0	0.0	138.8
6456	standard	5cpc	22	28	256	25.0	8.0	25.0	0.0	0.0	138.8
6457	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	138.8
6458	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	138.8
6459	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	138.8

Figure (C.10): Part of the sequence file where stimuli were edited.

When validating the sequence file, the following results were printed on the Matlab command window where it shows the pulse number that has the error as shown in the next figure.

```
-------------------------------------------------------
Parameters Validation Report
---------------------------
Error: Stimulation level is >255 or <0 at pulses number

I =

        6455
        6456
        6473

Fix the error and repeat  validation to continue
>>
```

Figure (C.11): Screenshot from Matlab result window showing the detected error.

2- Activity out of Threshold-Comfort levels bound

Each electrode in a patient's map has two values, threshold and comfort levels, these values were measured by clinical audiologists using clinical software. A stimulus at a given electrode must not exceed the comfort level or be below the threshold level. Figure (C.12) shows the T-C levels of each electrode.

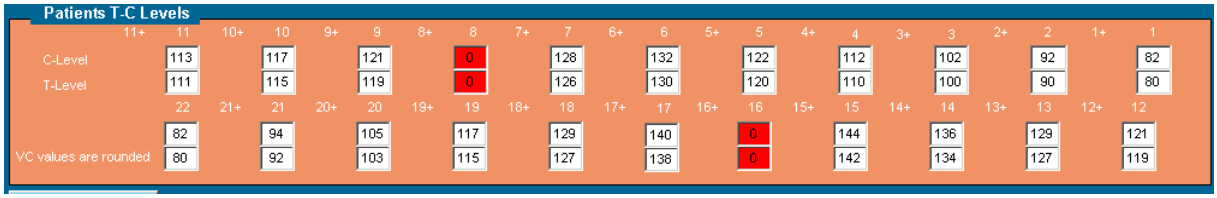


Figure (C.12): T-Levels and C-levels for each electrode using Test Freedom dummy patient with map number 138.

An error is introduced at lines 6509 and 6510 where the stimuli level exceeds Comfort level at electrode 11 and is below Threshold level at electrode 21.

```
6506  standard  5cpc   7   28  128  25.0  8.0  25.0  0.0  0.0  138.8
6507  standard  5cpc   9   28  120  25.0  8.0  25.0  0.0  0.0  138.8
6508  standard  5cpc   0    0   20  25.0  7.0  25.0  0.0  0.0  138.8
6509  standard  5cpc  11   28  117  25.0  8.0  25.0  0.0  0.0  138.8
6510  standard  5cpc  21   28   80  25.0  8.0  25.0  0.0  0.0  138.8
6511  standard  5cpc  20   28  104  25.0  8.0  25.0  0.0  0.0  138.8
6512  standard  5cpc  22   28   82  25.0  8.0  25.0  0.0  0.0  138.8
```

Figure (C.13): Out-of-bound errors were introduced at electrodes 11 and 21.

Figure (C.14) shows the results obtained from the Checker as it shows both electrodes 11 and 21 to have out-of-bound activity and their values.

```
-----------------------------------------------------
Parameters Validation Report
---------------------------
Pass: No Stimulation level >255 or <0
Error: Electrode 11 has out of T-C bound  T=111 C=113 intensity value =117
Error: Electrode 21 has out of T-C bound  T=92 C=94 intensity value =94
Pass: Number of maxima Correct = 8 < or = 8
Pass: No of channels=22 and mode is 28 Correct
>>
```

Figure (C.14): Out-of-Bound error report from the Checker.

3- Disabled channels

The patient's map is customized subjectively by audiologists, where some electrodes are switched off for different reasons; in some cases stimulating a certain electrode, may activate trigeminal of facial nerve. Accordingly, there must not be any activity on a disabled electrode. Figure (C.15) shows the disabled electrode for the loaded map.

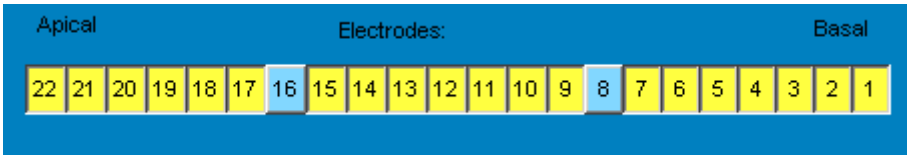


Figure (C.15): Disabled electrodes for Test Freedom patient with map 138.

Producing the error invokes adding an activity on both electrodes as shown in Figure (C.16).

6412	standard	5cpc	11	28	113	25.0	8.0	25.0	0.0	0.0	138.8
6413	standard	5cpc	12	28	121	25.0	8.0	25.0	0.0	0.0	138.8
6414	standard	5cpc	16	28	105	25.0	8.0	25.0	0.0	0.0	138.8
6415	standard	5cpc	11	28	113	25.0	8.0	25.0	0.0	0.0	138.8
6416	standard	5cpc	16	28	82	25.0	8.0	25.0	0.0	0.0	138.8
6417	standard	5cpc	8	28	132	25.0	8.0	25.0	0.0	0.0	138.8
6418	standard	5cpc	9	28	121	25.0	8.0	25.0	0.0	0.0	138.8
6419	standard	5cpc	10	28	117	25.0	8.0	25.0	0.0	0.0	138.8

Figure (C.16): Activity on disabled channels number 6 and 8 are added as shown in gray.

Figure (C.17) shows the validating report and it shows that there is out-of-bound activity at electrodes 8 and 16 because they are disabled and their T and C levels should be 0. Additionally it shows an activity of disabled channels and its pulse position.

```
-------------------------------------------------------
Parameters Validation Report
---------------------------
Pass: No Stimulation level >255 or <0
Error: Electrode 8 has out of T-C bound  T=0 C=0 intensity value =132
Error: Electrode 16 has out of T-C bound  T=0 C=0 intensity value =105
Error: Electrode Activity found on a disabled channel number 16
 The error in sequence is at the following pulse numbers

Io =

        6414
        6416

Error: Electrode Activity found on a disabled channel number 8
 The error in sequence is at the following pulse numbers

Io =

        6417

Pass: Number of maxima Correct = 8 < or = 8
```

Figure (C.17): Screenshot from Matlab showing the disabled channels activity.

4- Number of Maxima

It is a value that describes the number of peaks could be extracted from each FFT frame that is analyzed and electrode activities are prepared. The channel number is normally sorted either from basal-apical or apical-basal for each frame.

Number of maxima = Total Stimulation rate / Stimulation rate

Thus changing the active electrode value from 22 to 5 at line 6320 should invoke the algorithm to find the number of maxima as 9 and not 8.

```
6318  standard  5cpc   20   28  105  25.0  8.0  25.0  0.0  0.0  138.8
6319  standard  5cpc   21   28   94  25.0  8.0  25.0  0.0  0.0  138.8
6320  standard  5cpc    5   28   82  25.0  8.0  25.0  0.0  0.0  138.8
6321  standard  5cpc    6   28  132  25.0  8.0  25.0  0.0  0.0  138.8
6322  standard  5cpc    9   28  121  25.0  8.0  25.0  0.0  0.0  138.8
6323  standard  5cpc   10   28  117  25.0  8.0  25.0  0.0  0.0  138.8
6324  standard  5cpc   11   28  113  25.0  8.0  25.0  0.0  0.0  138.8
6325  standard  5cpc   12   28  121  25.0  8.0  25.0  0.0  0.0  138.8
6326  standard  5cpc   20   28  105  25.0  8.0  25.0  0.0  0.0  138.8
6327  standard  5cpc   21   28   94  25.0  8.0  25.0  0.0  0.0  138.8
6328  standard  5cpc   22   28   82  25.0  8.0  25.0  0.0  0.0  138.8
6329  standard  5cpc    6   28  132  25.0  8.0  25.0  0.0  0.0  138.8
```

Figure (C.18): Changing the active electrode at such location should invoke an error.

Results showed that Checker could detect such faults where it showed an out-of-bound at electrode 5 because its stimuli is outside the T-C levels and also showed that the detected maxima is 9 and not 8 as defined by this map.

```
-----------------------------------------------------
Parameters Validation Report
---------------------------
Pass: No Stimulation level >255 or <0
Error: Electrode 5 has out of T-C bound  T=120 C=122 intensity value =122
Pass: No activity found on the disabled electrodes
Error: Number of Maxima is wrong or maybe the sequence has a random order. Detected maxima = 9 and not 8 or less
```

Figure (C.19): Error invoked in the number of maxima algorithm.

5- Mode checking

As described previously in Table (3.1) that 22 channel stimulation strategy should have a mode of number 28. This is also check and the error could be invoked by changing the mode of any pulse (see Figure (C.20)). Checker could detect this mistake as shown in the result in Figure (C.21).

```
6319  standard  5cpc   21   28   94  25.0  8.0  25.0  0.0  0.0  138.8
6320  standard  5cpc    5   30   82  25.0  8.0  25.0  0.0  0.0  138.8
6321  standard  5cpc    6   28  132  25.0  8.0  25.0  0.0  0.0  138.8
```

Figure (C.20): Changing mode of stimulation from 28 to 30 in line 6320 to validate the Checker.

```
-----------------------------------------------------
Parameters Validation Report
----------------------------
Pass: No Stimulation level >255 or <0
Error: Electrode 5 has out of T-C bound  T=120 C=122 intensity value =122
Pass: No activity found on the disabled electrodes
Error: Number of Maxima is wrong or maybe the sequence has a random order. Detected maxima = 9 and not 8 or less
Error: Number of channels is 22 and mode not 28
```

Figure (C.21): result screen shot invoking an error in mode.

6- Rate checking

Stimulation rate is an important parameter that has a limit of 32 kHz for all electrodes overall in the Freedom implant and this rate is related to the pulse width for each stimulus and the interpulse interval and also to the number of maxima selected. This was invoked by changing the interpulse interval from 138.8 to 140.8 as shown in Figure (C.22). Due to approximation errors in calculations, a tolerance of 0.005% is included on the estimated value.

2864	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	140.8
2865	standard	5cpc	5	28	122	25.0	8.0	25.0	0.0	0.0	140.8
2866	standard	5cpc	6	28	132	25.0	8.0	25.0	0.0	0.0	140.8
2867	standard	5cpc	7	28	128	25.0	8.0	25.0	0.0	0.0	140.8
2868	standard	5cpc	9	28	121	25.0	8.0	25.0	0.0	0.0	140.8
2869	standard	5cpc	11	28	113	25.0	8.0	25.0	0.0	0.0	140.8
2870	standard	5cpc	12	28	121	25.0	8.0	25.0	0.0	0.0	140.8

Figure (C.22): Changing interpulse interval with to invoke rate error.

Figure (C.23) shows that this error could be found with the Checker.

```
-----------------------------------------------------
Parameters Validation Report
----------------------------
Pass: No Stimulation level >255 or <0
pass: Channels activities lies between T-C level
Pass: No activity found on the disabled electrodes
Pass: Number of maxima Correct = 8 < or = 8
Error: Number of channels is 22 and mode not 28
Error: Total Pulse rate does not match the pulse duration with respect to the detected number of maxima
```

Figure (C.23): Error invoked due to difference in rate.

6- Checking Phase 1 and Phase 2

Both phase 1 and phase 2 are checked to be similar and equal to the given processing value. This error is invoked as shown in Figure (C.24). The Checkers pops up an error as shown in the Figure (C.25).

2868	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0
2869	standard	5cpc	20	28	111	26.0	8.0	25.0	0.0	0.0	200.0
2870	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0
2871	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0

Figure (C.24): Simulating phase width error.

```
pass: Stimulation rate Correct
Error: Phase 1 and Phase 2 are not equal at pulses number

i =

        2869

Error: Phase 1 and Phase 2 values are not Correct
pass: Phase 3 is zero. No phase 3 used.
```

Figure (C.25): Error invoked due to difference in phase 1 and phase 2.

7- Checking Phase 3 and Phase 4

Phase 3 and phase 4 are not used and there must be no activity on them. This is checked in this step to be correct otherwise an error is produced.

2867	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0
2868	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0
2869	standard	5cpc	20	28	111	25.0	8.0	25.0	1.0	0.0	200.0
2870	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0
2871	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0

Figure (C.26): Simulating phase 3 and/or phase 4 width error.

```
Parameters Validation Report
----------------------------
Pass: No Stimulation level >255 or <0
pass: Channels activities lies between T-C level
Pass: No activity found on the disabled electrodes
Pass: Number of maxima Correct = 10 < or = 10
Pass: No of channels=22 and mode is 28 Correct
pass: Stimulation rate Correct
pass: Phase 1 and Phase 2 are equal
Pass: Phase 1 and Phase 2 values Correct
Error: Phase 3 is not zero at pulses

i =

        2869

pass: Phase 4 is zero. No phase 4 used.
pass: Phase gap is correct
Pass: Channel Time Matrixes are correct comparison is done pulse by pulse
>>
```

Figure (C.27): Error invoked due to activity in phase 3.

8- Checking Phase gap

The phase gap is checked to be the same as the processing values, taking into consideration that the phase gap is different for powering pulses. This was invoked and checked as shown in Figure (C.28) and Figure (C.29) respectively.

2867	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0
2868	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0
2869	standard	5cpc	20	28	111	25.0	6.0	25.0	0.0	0.0	200.0
2870	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0
2871	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	200.0

Figure (C.28): Simulating a phase gap error.

```
-----------------------------------------------------
Parameters Validation Report
----------------------------
Pass: No Stimulation level >255 or <0
pass: Channels activities lies between T-C level
Pass: No activity found on the disabled electrodes
Pass: Number of maxima Correct = 10 < or = 10
Pass: No of channels=22 and mode is 28 Correct
pass: Stimulation rate Correct
pass: Phase 1 and Phase 2 are equal
Pass: Phase 1 and Phase 2 values Correct
pass: Phase 3 is zero. No phase 3 used.
pass: Phase 4 is zero. No phase 4 used.
Error: Phase gap is not correct
Pass: Channel Time Matrixes are correct comparison is done pulse by pulse
>>
```

Figure (C.29): Error invoked due to change in phase gap.

9- Sequence matrix

This section checks the channels to be stimulated and their current level with those were generated and their current levels. By changing any electrode number of its current level, this test will not be passed. Active electrode in line 2877 was changed from 11 to 12 as shown in Figure (C.30).

2876	standard	5cpc	9	28	121	25.0	8.0	25.0	0.0	0.0	138.8
2877	standard	5cpc	12	28	112	25.0	8.0	25.0	0.0	0.0	138.8
2878	standard	5cpc	0	0	20	25.0	7.0	25.0	0.0	0.0	138.8

Figure (C.30): Active electrode being changed to invoke sequence matrix mismatch error.

Checker could detect this error and present it in the report as shown in Figure (C.31).

```
-----------------------------------------------------
Parameters Validation Report
----------------------------
Pass: No Stimulation level >255 or <0
Error: Electrode 12 has out of T-C bound  T=119 C=121 intensity value =121
Pass: No activity found on the disabled electrodes
Pass: Number of maxima Correct = 8 < or = 8
Pass: No of channels=22 and mode is 28 Correct
pass: Stimulation rate Correct
Error: Channel Time Matrices dont Match.. Pulse by pulse comparison mismatch found
>>
```

Figure (C.31): Result from Checker showing an error invoked in the sequence matrix test.

The Checker was calibrated to produce electrodograms activities with almost the same loudness as if they were acoustically stimulated. Accordingly, a new slider in the Checker was introduced to logarithmically amplify the input sound. By amplifying the sound around 15 dB, the electrodogram is Figure (C.30-left) seems to have lost the fine modulations. An additional inspection to the software and NMT indicated that that the Q value in the loudness growth function "LGF_proc.m" should be changed. It was varied from 20 to 40 and the result is shown in Figure (C.30-right). In order to keep the old programs running smoothly, an additional field was added to the p structure "p.newbaselevel" as shown in Figure (C.33). For more details about the calibration process (see Appendix D).

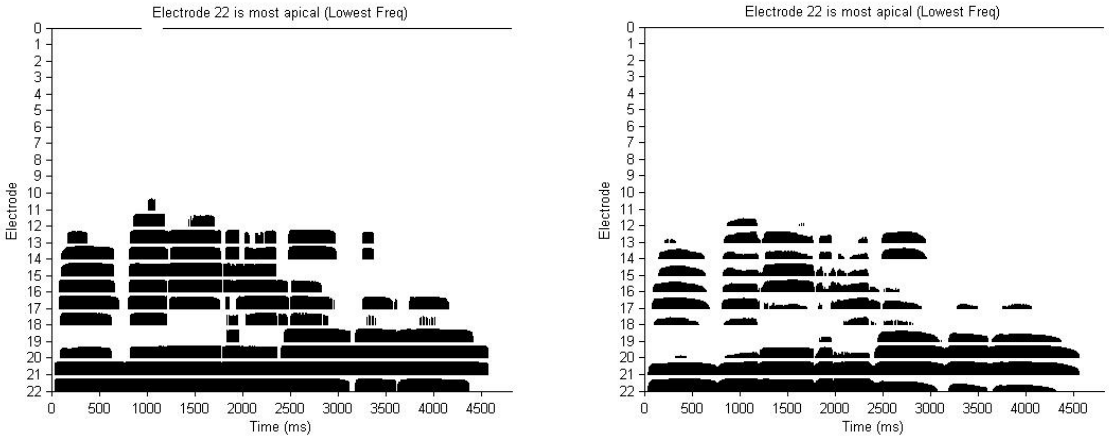


Figure (C.32): Electrodograms of a clarinet sound amplified 15dB with Q=20 (left) and Q=40 (right).

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
case 1  % Parameter calculations
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % Defaults:

    % Newbaselevel is a field in the P structure to be added,if we want
    % to set the processing according to the T-SPL level of a map

    if isfield(p,'newbaselevel')==0,     % If newbaselevel is not passed from the AMO,checker %Sherif
        p = Ensure_field(p,'base_level',  4/256);
        p = Ensure_field(p,'Q',          20);
    elseif isfield(p,'newbaselevel')==1, % if it is passed.
        Tspl=150/(256*From_dB(65 - p.newbaselevel));
        p = Ensure_field(p,'base_level',Tspl);
        p = Ensure_field(p,'Q',          40);   %Changed by Sherif: It was 20
    end

```

Figure (C.33): Changes to "LFG_proc.m" file; when the "p.newbaselevel" field is passed to the LGF in the NMT, the base level will be adjusted according to the patient TSPL value in his map and the Q will be changed to 40, otherwise no change is used.

Extended SQL function for NIC

```
function data=LoadPatientNames()
%% Please load the patient names and their USZ numbers
% get all Patient IDs and Names which are NEITHER archived NOR explanted
% default looks CIC4 implants only
querytext = sprintf(['SELECT DISTINCT Name_First, Name_Last,
CrossReference'...
    ' FROM Recipient, Implant, Part where T_RecordStatus=\''0\'''...    %
RecordStatus = 0: recipient NOT archived
    ' AND FK_GUID_Recipient=GUID_Recipient'...
    ' AND FK_Part=UID_Part'...
    ' AND Implant.T_Status<>1000000130'...  % LRD_ImplantStatus_Explanted
    ' AND National_ID<>\''Intraop\'''...
    ' AND (Part_Desc=\''COCHLEAR IMPLANT,CI24RE (CA)\'''...
    ' OR Part_Desc=\''COCHLEAR IMPLANT,CI24RE (CS)\'''...
    ' OR Part_Desc=\''COCHLEAR IMPLANT,CI24RE (ST)\'''...
    ' OR Part_Desc=\''COCHLEAR IMPLANT, CI24RE (ST)\'')'...
    ' order by Name_Last, Name_First' ]);
data = Get_data(querytext);
return
```

```
function op=LoadPatientImplantedEar(USZnumber)
%% Please tell me if the patient has Left/Right/Both ears implanted
% get the side implanted for patient no USZnumber, if left output 'L', if
right output 'R', if both output 'B'
%
querytext = sprintf(['SELECT DISTINCT  T_Locus'...
    ' FROM Recipient, Implant, Part where T_RecordStatus=\''0\'''...    %
RecordStatus = 0: recipient NOT archived
    ' AND FK_GUID_Recipient=GUID_Recipient'...
    ' AND FK_Part=UID_Part'...
    ' AND Implant.T_Status<>1000000130'...  % LRD_ImplantStatus_Explanted
    ' AND CrossReference=' char(39) num2str(USZnumber) char(39)  ...
    ' AND (Part_Desc=\''COCHLEAR IMPLANT,CI24RE (CA)\'''...
    ' OR Part_Desc=\''COCHLEAR IMPLANT,CI24RE (CS)\'''...
    ' OR Part_Desc=\''COCHLEAR IMPLANT,CI24RE (ST)\'''...
    ' OR Part_Desc=\''COCHLEAR IMPLANT, CI24RE (ST)\'')'...
    ' order by T_Locus' ]);
op = Get_data(querytext);
return
```

```
function op=LoadPatientMAPS(USZnumber, EarSideNO)
%% Please tell me What Maps does the patient have given an ear side and his
%% USZnumber give the usznumber of the patient and the ear side as a number
% Left Ear=1000000126, Right Ear = 1000000127 and output is the map numbers
querytext = sprintf(['SELECT DISTINCT MAPNumber '...
    ' FROM Recipient, MAP, Implant, Part WHERE '...    % RecordStatus = 0:
recipient NOT archived
    ' Recipient.GUID_Recipient=MAP.FK_GUID_Recipient'...
    ' AND FK_Part=UID_Part'...
    ' AND Implant.T_Status<>1000000130'...  % LRD_ImplantStatus_Explanted
    ' AND Recipient.CrossReference=' char(39) num2str(USZnumber) char(39)
...
    ' AND Implant.T_Locus=' char(39) num2str(EarSideNO) char(39) ...
    ' AND Implant.GUID_Implant=MAP.FK_GUID_Implant' ...
    ' AND (Part_Desc=\''COCHLEAR IMPLANT,CI24RE (CA)\'''...
    ' OR Part_Desc=\''COCHLEAR IMPLANT,CI24RE (CS)\'''...
    ' OR Part_Desc=\''COCHLEAR IMPLANT,CI24RE (ST)\'''...
    ' OR Part_Desc=\''COCHLEAR IMPLANT, CI24RE (ST)\'')'...
    ' AND Recipient.T_RecordStatus=' char(39) '0' char(39) ...
    ' ORDER BY MAPNUMBER;' ]);
op = Get_data(querytext);
return
```

Determining number of maxima

Finding the number of maxima goes on two steps; loading the file and calling findnumberofmaxima function.

Example

[Cell1,Cell2,Cell3,Cell4,Cell5,Cell6,Cell7,Cell8,Cell9,Cell10,Cell11]=loadcellm('Nic_test.txt');
maxima=findnumberofmaxima(Cell3,Cell11,22) % for 22 Channels

Function 1: Loadcellm – Load a sequence text file

```
function
[Cell1,Cell2,Cell3,Cell4,Cell5,Cell6,Cell7,Cell8,Cell9,Cell10,Cell11]=loadc
ellm(fname);
%function [lc,dflag,numdata]=loadcell(fname,delim,exclusions);
%
%  loadcellm loads a RS capture file into cells, it is the fastest optimum
%  way done, loading 1.5MB needs 83 Sec only, other methods took 30 min.
%  File is not tab delimited and no other way to loadit v.fast and properly
%  with matlab except this.
%
%  fname: Filename according the RS-Capture Structure
%  Written by: Sherif Omran
%  University Hospital Zurich: 14-Feb-08
%Col3 : Active Electrode
%Col4 : Reference Electrode Mode
%Col5 : Current Level
%Col6 : Phase Width
%Col7 : Phase Gap
%Col8 : Phase 2 Width
%Col9 : Phase 3 Width (not used, no phase 3)
%Col10 : Phase 4 Width (not used too)
%Col11 : Duration in microsec between 2 pulses
```

```
%Example:
[Cell1,Cell2,Cell3,Cell4,Cell5,Cell6,Cell7,Cell8,Cell9,Cell10,Cell11]=loadc
ellm('Nic_test.txt');

%Open file
fid=fopen(fname,'rt');
%Cannot open: return -1
if (fid<0)
  lc=-1;
else
  fullfile=fread(fid,'uchar=>char')';

  %Find all eol
  endpos=find(fullfile==char(10));
  st=1;
 % tic
  for row=1:length(endpos)
    line=fullfile(st:endpos(row));
    line=[line ' '];
    if length(line)>2, %entries in the line and not the last line
         [a,b,c,d,e,f,g,h,j,k,l]=GetCells(line);

         Cell1(row,:)=a;           Cell2(row,:)=b;
         Cell3(row,:)=str2num(c);  Cell4(row,:)=str2num(d);
         Cell5(row,:)=str2num(e);  Cell6(row,:)=str2num(f);
         Cell7(row,:)=str2num(g);  Cell8(row,:)=str2num(h);
         Cell9(row,:)=str2num(j);  Cell10(row,:)=str2num(k);
         Cell11(row,:)=str2num(l);
    end
    st=endpos(row)+1;
  end
%toc
end
return;



function
[Cellx1,Cellx2,Cellx3,Cellx4,Cellx5,Cellx6,Cellx7,Cellx8,Cellx9,Cellx10,Cel
lx11]=GetCells(line)
%This function take a line in the following strucutre
% line='standard  5cpc    0    0   20  25.0  7.0  25.0   0.0   0.0  166.6'
% the separation is depending on the spaces between cells
% Output should be
% Cell1:Standard
% Cell2: 5cpc
% and so on

line=deblank(line);

x=findstr(line,' ');
a=diff(x);
b=find(a>1);
st=(x(b)+1);
ed=(x(b)+1+a(b)-2);

row=1;

Cellx1(row,:)=line(1:x(1));   %1st cell
for i=1:9
    sl=['Cellx' num2str(i+1) '(row,:)=line(' num2str(st(i)) ':'
num2str(ed(i)) ');']; % ex: Cell2(row,:)=line(9:13);  %2st cell
```

```
    eval(sl);
end

Cellx11(row,:)=line(x(end)+1+a(end)-2:end); %11th cell

return;
```

Function 2: findnumberofmaxima – Find number of maxima

```
% Function: findnumberofmaxima
%
% It takes Column 3 from the sequence text file and gives the number of
% maxima used in it.
%
% Procedure:
% The column is divided into overlapping frames with the same duration as
% those were used for FFT analysis, each frame undergoes the maxima search
% algorithm.
% The algorithm can handle both Apical-Base and Base-Apical sequences
% while Random Sequence is not supported
%
% Example
%[Cell1,Cell2,Cell3,Cell4,Cell5,Cell6,Cell7,Cell8,Cell9,Cell10,Cell11]=load
cellm('Nic_test.txt');
% maxima=findnumberofmaxima(Cell3,Cell11,22) % for 22 Channels
% maxima=findnumberofmaxima(Cell3,Cell11,43) % for 43 Channels
%
% Written By:
% Sherif Omran
% University Hospital Zurich
% 28. April 08

function  output=findnumberofmaxima(x,Cell11,numberofchannels)
%x is Cell3 from the sequence File

%Window frame size in number of rows from the sequence file
fs=16000;
if numberofchannels==22,
    nofpoints=128;              %used in FFT analysis
elseif numberofchannels==43,
    nofpoints=256;              %used in FFT analysis
end
overlap=32;                 %used in FFT analysis
Pulsewidth=mean(Cell11)*10^-6;

WindowSizePoints=(nofpoints)/(fs*Pulsewidth);
OverlapSizePoints=(overlap)/(fs*Pulsewidth);

u=(buffer(x,WindowSizePoints,OverlapSizePoints))';

for i=1:size(u,1)
    op=CountMaxima(u(i,:)');
    if isempty(op)==0 %not empty
        ou(i)=op;
    end
end

output=max(ou);

return
```

```
function op=CountMaxima(x)

 % no repetition
 d=diff(x);
 io=find(d~=0);
 xn=[x(1) x(io+1)']; %remove repetitions

 % Exclude 0 channels Indices-> powers
 u=find(xn~=0);
 xn2=xn(u);
 if isempty(xn2)==0,
     xn2(end)=[];
 end

 % Remove duplications after removing the power signals indices
 xnd2=diff(xn2);
 u2=find(xnd2~=0);
 xn2a=xn2(u2);

 % Find -ve points represent end point of rising sequence
 k=diff(xn2a);

 u=find(k<0);    %find -ve values
 u1=find(k>0);   %find +ve values

 t=length(u);
 t1=length(u1);

 if t<t1,       %Basal-Apical
  c=diff(u);    %find number of point between -ve values (Frames)
 elseif t1<t,   %Apical-Basal
  c=diff(u1);
 elseif t==t1,  % could be Random Sequence or very few stimuli
     c=0;
 end

 op=max(c);

return
```

Changing sequence text file from 43 to 22 “seq_43to22”

```
% Function seq_43to22
%
% This function takes a sequence structure used in generating the text
sequence file
% Due to limitation in NIC, it doesn't accept 43 channels, this work around
is done.
% Channel numbering in the sequence matrix are changed from 1-43 into 1-22
% so that the streaming function can accept it.
% However, to know which channels were the virtual channels I added another
% field in the structure called.
% sequence.virtualchannel
%                         = 1   for virtual channels
%                         = 0   for physical electrodes
%
% sequence.electrodes     = holds the new electrode sequence
%
% sequence.oldsequence    = holds the old sequence (1-43)
%
%
% Syntax:       op=seq_43to22(sequence)
%
% Author: Sherif Omran
%         University Hospital Zurich
%         22 April 08
%

function op=seq_43to22(sequence)

%Channel sequence in case of 22 ch is from 1:22 and not from 0:21

sequence.virtualchannel=zeros(length(sequence.electrodes),1);
newsequence=zeros(length(sequence.electrodes),1);

for i=1:length(sequence.electrodes)
    current_channel=sequence.electrodes(i);
    ab=physical_channel(current_channel);
    newsequence(i)=ab;
    a=IsPhysical(current_channel);
    if current_channel~=0, %0 means power pulse
        if a==1 , %yes it is physical channel
            sequence.virtualchannel(i)=0;
        else     %no it is virtual channel
            sequence.virtualchannel(i)=1;
        end
    end
end

sequence.oldsequence=sequence.electrodes;
sequence.electrodes=newsequence;

op=sequence;

return;

function opp=IsPhysical(electrodeno)
% This function tells if the electrodeno represent a physical electrode
% (op=1) or not (op=0)
a=fix(electrodeno/2)-electrodeno/2;
if a~=0, %physical electrode
    opp=1;
else
```

```
    opp=0;
end
return

function op=physical_channel(channel)
% This function tells what the physical electrode is given any channel
% number
% Example: inputchannel = 5, output=3
%          inputchannel = 6, output=3
a=IsEven(channel);
if a==0, % it is odd channel
    op=(channel+1)/2;
else
    op=channel/2;
end

return

function op=IsEven(n)
% is this number even? yes=1, no=0
a=fix(n/2)-n/2;
if a==0,
    op=1; %yes even
else
    op=0; %no odd
end
return
```

Changing xml file from 22 to 43 channels "seq_xml_22to43"

```
%function seq_xml_22to43(filename,sequence)
%
%  seq_xml_22to43 is a work around for stimulating 43 channels using NIC
%  The idea depends on that the sequence were edited to have the 22 channel
%  format, there is a field sequence.virtualchannels that points to the VC
%  The XML is generated using the 22 channel format, and using this field
%  sequence.virtualchannels, this function will edit the xml file to change
%  the mode of the virtual channels.
%
%  Written by: Sherif Omran
%  University Hospital Zurich: 21-April-08
%
% Ex:
% a=seq_xml_22to43('10338058_Al-
Qattan_Noor_left_Map7_CIC4_ACE_22_STD_Fall4a3d.xml',sequence);
%
% Notes: The new algorithm in the upper part is ~120 times faster than the
% lower one. It makes the changes on the fly using pointers.
% I counted the number of virtual electrodes should be changed in the file
% before calling the function using matlab and after calling it using
% notepad++ and is correct.
function op=seq_xml_22to43(fname,sequence);

% Load the xml file in a string

maximumlinelength=250; %maximum length of anyline in the xml file

fid=fopen(fname,'r');
if (fid<0)
    op=-1;
    return;
else
        fullfilebin=fread(fid)'; %Read it as binary file
        refelectrodestart=strfind(fullfilebin,char('<re>'));
        %start position of the referenc electrode
        refelectrodeend=strfind(fullfilebin,char('</re>'));
        %end position of the reference electrode
        VCStmuliNumber=find(sequence.virtualchannel==1);
        xs=refelectrodestart(VCStmuliNumber);
        %start positions of the reference electrodes for only VC
        xe=refelectrodeend(VCStmuliNumber);
        %end positions of the reference electrodes for only VC
        %fullfile(xs:xe)='<re>-4<'; %Edit the virtual channels on the fly
        for i=1:length(xs)
      fullfilebin(xs(i):xe(i))=abs(strrep(char(fullfilebin(xs(i):xe(i))),'-
3','-4'));
        end
end
fclose(fid);
% Now write the new output xml file
fid2 = fopen(fname,'w');
fwrite(fid2,fullfilebin);
fclose(fid2);
op=1;
```

Appendix – D (Calibration)

D.1 Calibration

Semitone mapping may provide a better perception for Cochlear Implant recipients. There are two ways to test with CI patients, either through an acoustic stimulation where a signal is transmitted over loudspeakers, or through a direct stimulation of electrodes, where signals are processed, analyzed and transmitted to the implant (see Figure (D.1)). To calibrate, both outputs are compared to have similar electrode activity per time (electrodogram) using similar patient map and processing strategy. Figure (D.1) shows the two pathways.

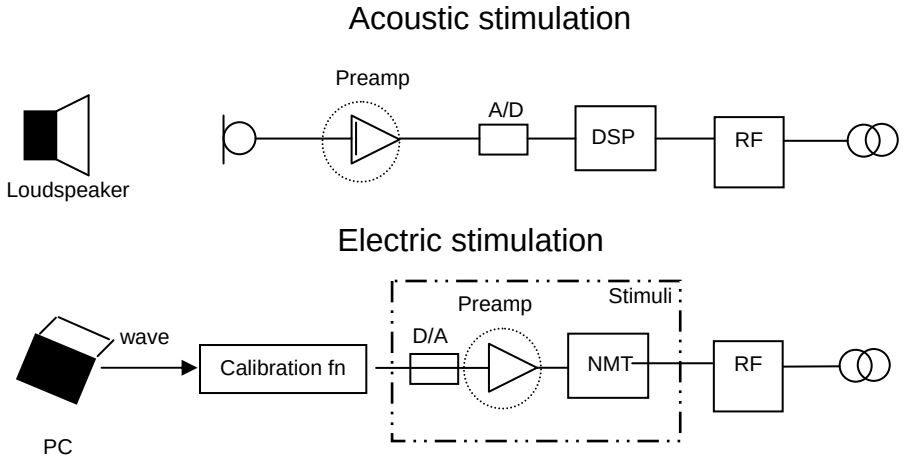


Figure (D.1): Acoustic and electric stimulation pathways

Figure (D.1) shows the two path ways (electric and acoustic stimulation) to be calibrated and their underlying components. Assuming that in the electric stimulation there is an unknown calibration filter that could match the output of the electrodograms from both paths.

D.1.1 Approach

Psychoacoustic tests were performed through a loudspeaker adjusted at a level of 65 dB(A) and 1.5 m away from the patient and compared with other tests conducted by streaming the transfer function of a calibrating filtering to be determined. In order to find the transfer function of the calibrating filter, a CI Freedom processor was first programmed with a test patient map that has a constant threshold and comfort levels (T-Level=80 and C-Level=120). The processor was placed inside a calibrated hearing and test box where its coil was connected to a radio frequency capture box that was connected to a computer in a way that electrode activities could be captured and saved in a text file using a python program. The calibration was started by playing different single harmonic tones through an audiometer and capturing the output electrodograms. The loudness of the audiometer tones was varied from 40-70 dB(SPL) and the frequency was varied from 125Hz-6kHz.

For each frequency the output text files at different presentation levels were inspected to find the loudness level (Phon) at which comfort level was just hit. The corresponding electrode

with such a maximum activity was noted. Then the audiometer presentation value was varied and the captured activity at the same electrode was recorded as shown in Table (D.1).

Audiometer											
Current Levels											
Electrodes	**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**
loudness dB(SPL)	125 Hz	200 Hz	250 Hz	400 Hz	500 Hz	630 Hz	800 Hz	1000 Hz	2000 Hz	4000 Hz	6000 Hz
40	89	95	96	99	100	105	103	103	106	107	97
45	93	99	100	104	107	109	108	107	111	111	102
50	98	103	105	108	111	113	112	112	115	115	107
55	101	105	108	112	115	117	116	116	118	119	111
60	102	110	113	116	118	120	120	120	120	120	115
65	107	114	117	120	120	120	120	120	120	120	119
70	110	118	120	120	120	120	120	120	120	120	120

Table (D.1): Audiometer sinusoidal tones sound pressure level and frequency values were varied and the electrode with maximum activity at c-level was recorded.

From Table (D.1) different current level values were plotted for different SPL value along the frequency domain.

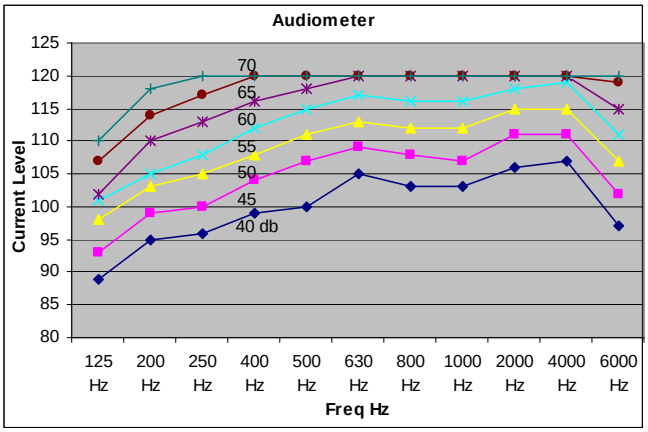


Figure (D.2): Audiometer current levels for different Phons per frequency.

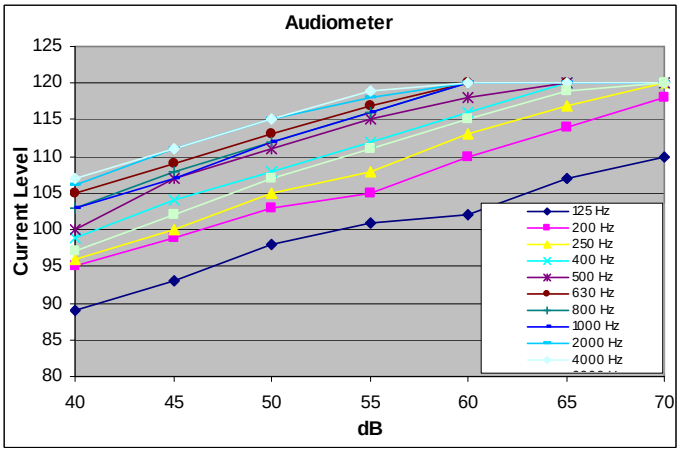


Figure (D.3): Audiometer current levels for different Phons value (dB-SPL) per frequency.

Figures (D.2 and D.3) show the current level variation for each sound pressure level value at discrete frequencies. They show that the current level increases for an additional 5 dB increase. These graphs represent an output from an acoustic stimulation path that will be compared to outputs from an electric stimulation path to find the calibration filter function.

D.1.2 NIC Analysis

A sinusoidal single harmonic tone was generated using Matlab at different frequencies where the amplitude of the tone was varied in 5 dB steps between 70 and 25 dB. Signals with 70 dB had an absolute maximum equal to 1 in the wave file.

A programmed Matlab utility called "Checker" (see Appendix C) was used to process the generated wave files and produce a stimuli file. Checker uses NMT.

NMT block diagram according to Cochlear's documentation is shown in Figure (D.4), where there are two paths for a signal to flow, either the signal is picked up through a headset microphone and through a preamplifier or it flows directly from an external input cable to a pre-emphasis and then to the analogue digital converter (ADC). Previous psychoacoustic tests were presented acoustically from a loudspeaker to the processor's microphone and accordingly the first path is used in the calibration.

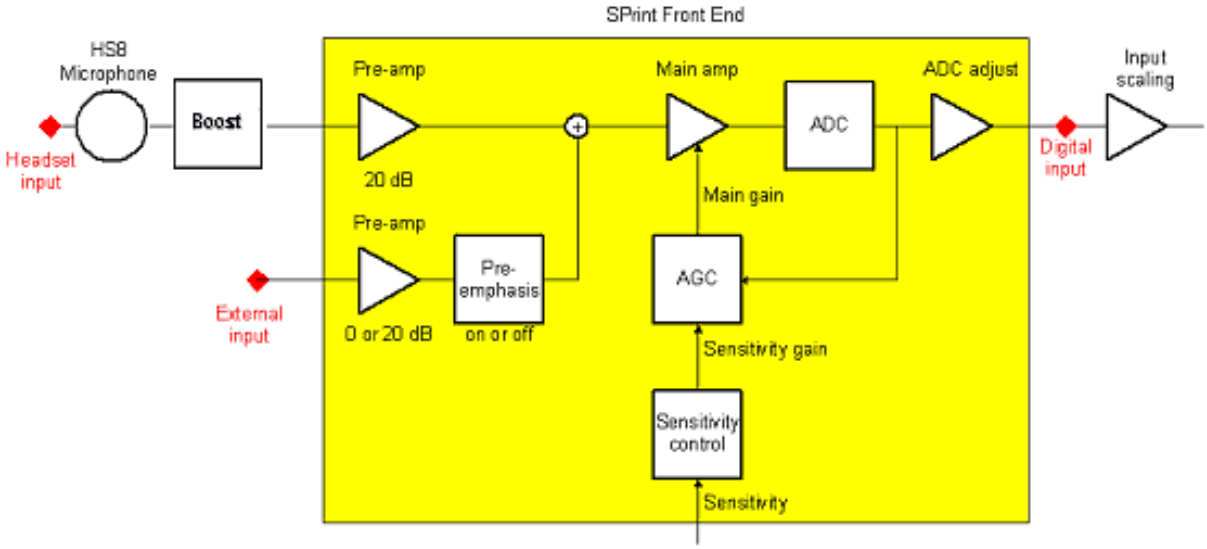


Figure (D.4): Front-end block diagram for NMT.

This block diagram is implemented in the "p.processes" structure in NMT. The structure has the following entries:

Wav_proc	Read wave file and ensure sampling rate
HS8_microphone_proc	Add microphone gain and transfer function
SPrint_front_end_proc	Add sensitivity and preamplifier gain
Input_scaling_proc	Additional scaling factor
FFT_filterbank_proc	FFT filter bank analysis
Power_sum_envelope_proc	Power spectral density calculation
Reject_smallest_proc	Reject peaks below T-level
LGF_proc	Add a dynamic range gain

D.1.3 Microphone Transfer Function

According to the documents from Cochlear PTY, the microphone transfer function is shown in Figure (D.5) and has a pre-amplifier gain of 20 dB. The microphone has a sensitivity (Default = 26.2500 dB SPL). Accordingly the front end gain is the sum of the preamplifier and the sensitivity = 46.5 dB SPL, which is implemented in Matlab as a 1st order Butterworth BPF with pass band [1750 4800] Hz as shown in Figure (D.5).

Figure (D.5): Microphone transfer function according to Cochlear documents.

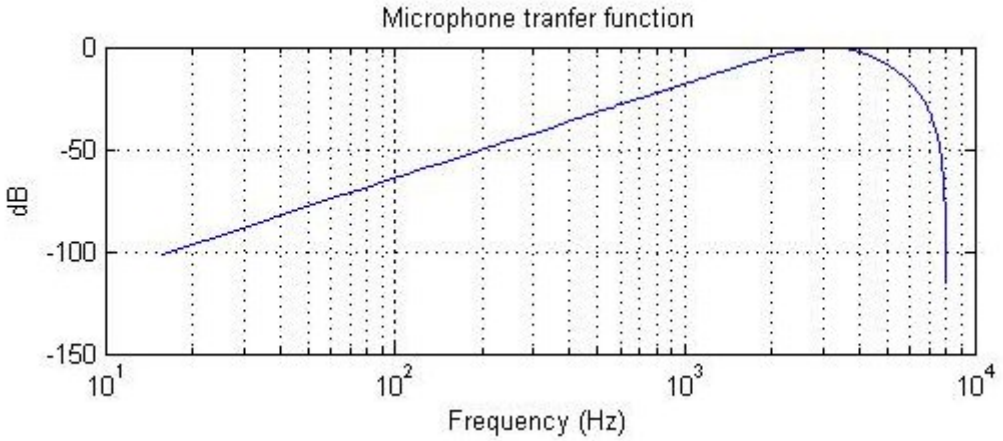


Figure (D.6): Implemented microphone transfer function.

It was noticed that the microphone needs a linear boost and this was implemented in the AMO and as well as in Checker as shown in

Figure (D.7).

```
3153 -      else  % Don't include calibration filter, thus check if headset is required
3154 -          p.reference_SPL_dB = 56.6 + str2double(get(handles.BoostNum,'String')); % boost up if HS used
3155 -      end
```

Figure (D.7): implantation of a linear boost.

The boost was set to 6 dB. Different sinusoidal wave files at different frequencies and loudness levels were processed (see Figure (D.8) and compared to Figure (D.2)).

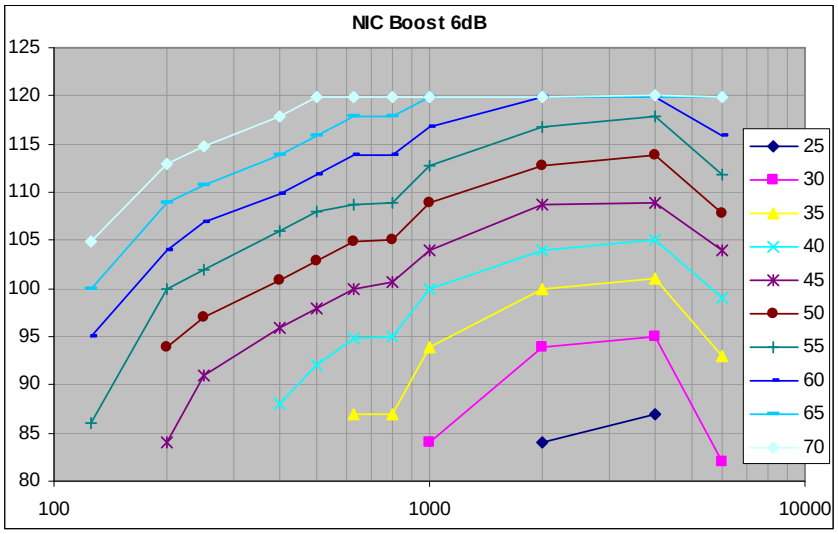


Figure (D.8): Current level at different frequencies with different loudness levels with 6 dB boost.

For different loudness values from 25-70 dB, the boost was changed at 1 kHz and the output current level was plotted as shown in Figure (D.9) and it was found that the gain is linear.

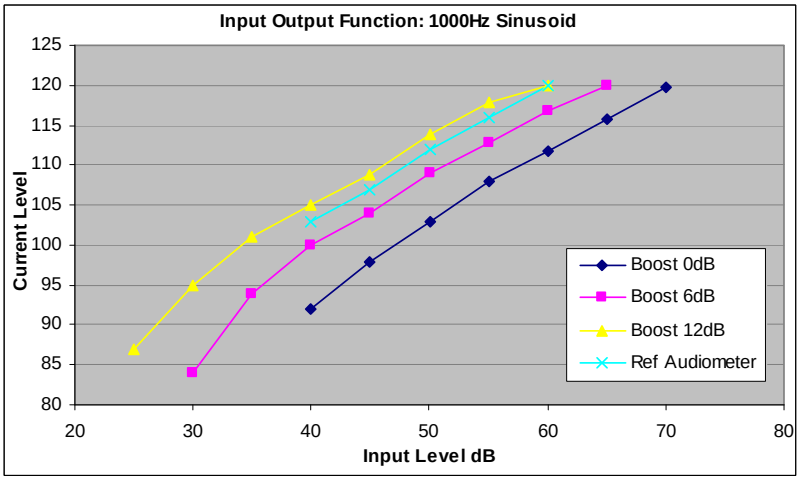


Figure (D.9): Gain increase in current level with boost increase.

Because the boost was the only available parameter to change, it was varied between 6 and 19 with 1 dB step for each frequency and each Phons value. An example at 1 kHz is shown in Figure (D.9), where current level is plotted at different loudness levels. From Figure (D.9) one can see that the reference audiometer curve lies between boost 6-12 for 1 kHz accordingly other boost values were tested for all frequencies. The boost value at each frequency that aligns the gain with the audiometer is noted as shown in Table (D.2) and Figure (D.10).

Frequency	125	250	500	800	1000	2000	4000	6000	8000
dB	14	14	14	14	14	9	7	6	6

Table (D.2): Calibration filter transfer function values.

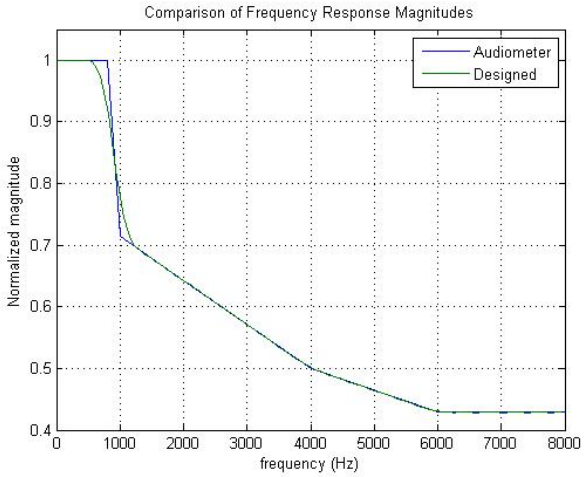


Figure (D.10): Determined calibration filter transfer function (left) and implemented (right).

The nearest filter characteristics were determined using FIR2 in Matlab and the filter parameters are described by Equation (D.1) and are calculated by the code below:

$$H(e^{j2\pi.f}) = \frac{B(e^{j2\pi.f})}{A(e^{j2\pi.f})} = \frac{b(1)+b(2).e^{-j2\pi.f} + ... + b(m+1).e^{-j2\pi.f.m}}{a(1)+a(2).e^{-j2\pi.f} + ... + a(m+1).e^{-j2\pi.f.m}} \qquad \text{Equation (D.1)}$$

```
f=[0 125 250 500 800 1000 2000 4000 6000 8000];
fn=f/8000;
hcalib=[14 14 14 14 14 10 9 7 6 6];
hcalibn = hcalib/max(hcalib);
B = 3.981*fir2 (80, fn, hcalibn); A=1;
```

D.2 Verification

Since the calibration function was determined using single harmonic tones at different loudness levels and different frequencies, an elementary verification test is to process the tones with the Checker application and using the calibration filter replacing the boost block (refer to Figure (D.4). Figure (D.11) shows a comparison between the outputs using the calibration filter (right) and the filter shape determined from the audiometer (left) (more details see verification of calibration function in the appendix).

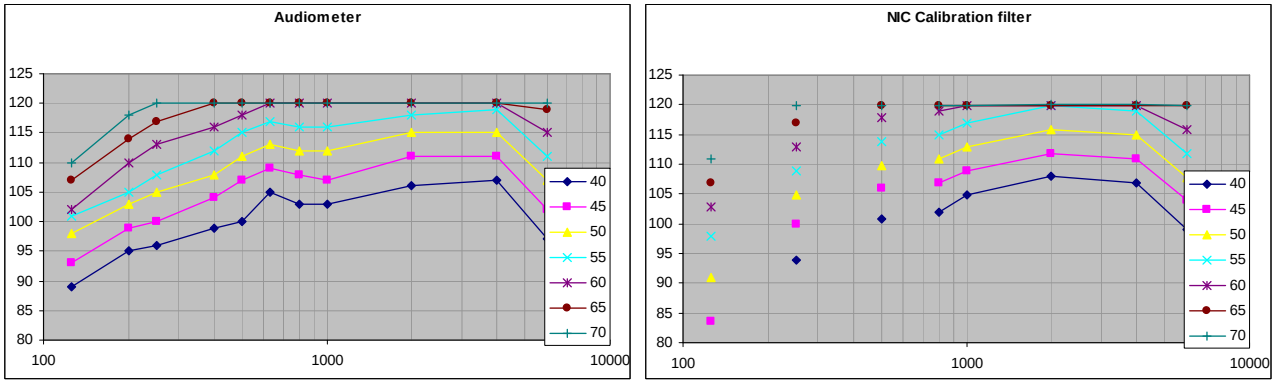


Figure (D.11): Filter shape determined from the audiometer (left) and calibration filter (right).

The calibration's targets finding the nearest electrodograms to each other when patients are acoustically or electrically stimulated and because the previous graphs were plotted using sinusoidal tones, accordingly a relation between average root mean square (RMS) and loudness for the sinusoidal tones is plotted in Figure (D.12).

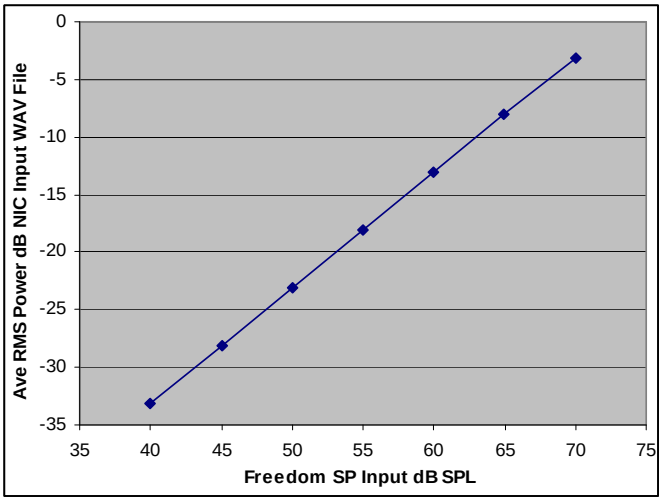


Figure (D.12): Relation between average RMS and loudness dB SPL

For a complex signal, an A-weighted filter (see Figure (D.13) is used to filter the audio signal and calculate the average RMS of the output signal (*R2*). On the other hand, the user has to decide at which dB(A) level the signal should be calibrated to and from Figure (D.13) the average RMS (*R1*) is calculated for the required dB(A) value.

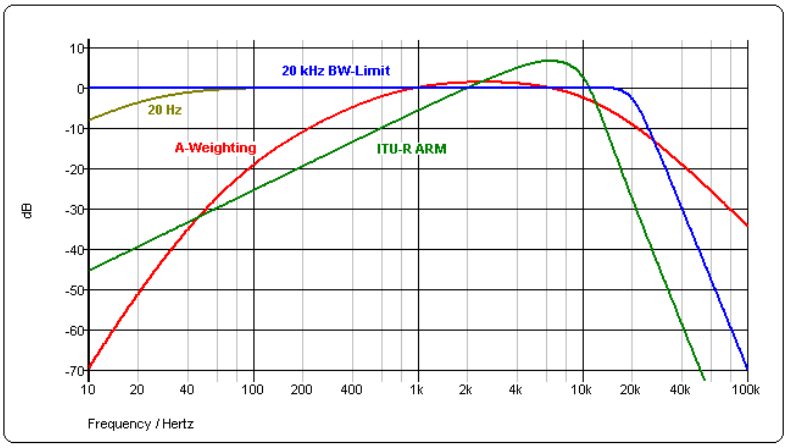


Figure (D.13): A-Weighted filter.

The value *R2:R1* is calculated and the input signal is then amplified with such a value. Now that the calibration parameters are adjusted, the RF captured output from a CI processor can be compared with the output from Checker as shown in Figure (D.14). A random map no 107 for patient (M.Be) is used in the verification. Figure (D.14) shows that the captured signal and the generated signal look generally similar with minor differences in channels 7 to10 this is due to free field acoustic external noise.

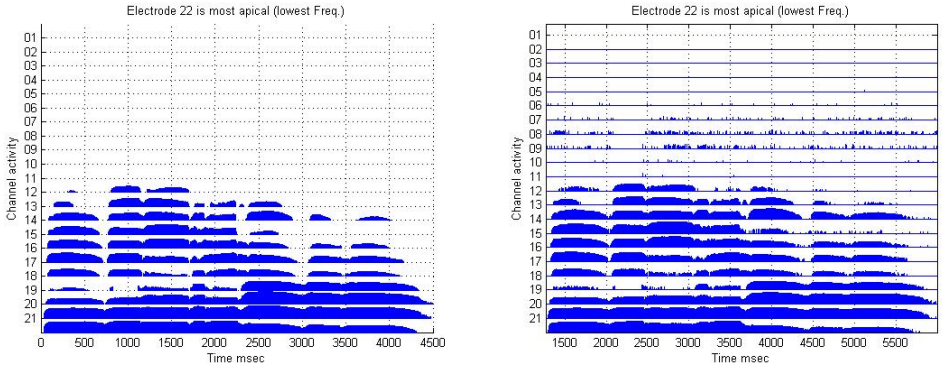


Figure (D.14): Checker generated signal at 63 dBA (left) and RF captured signal at 60 dBA (right) with a clarinet signal.

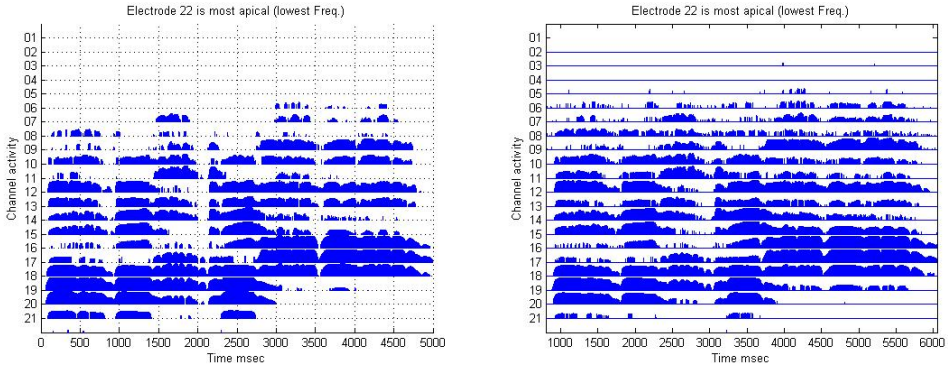


Figure (D.15): Checker generated signal at 63 dBA (left) and RF captured signal at 60 dBA (right) with a flute signal.

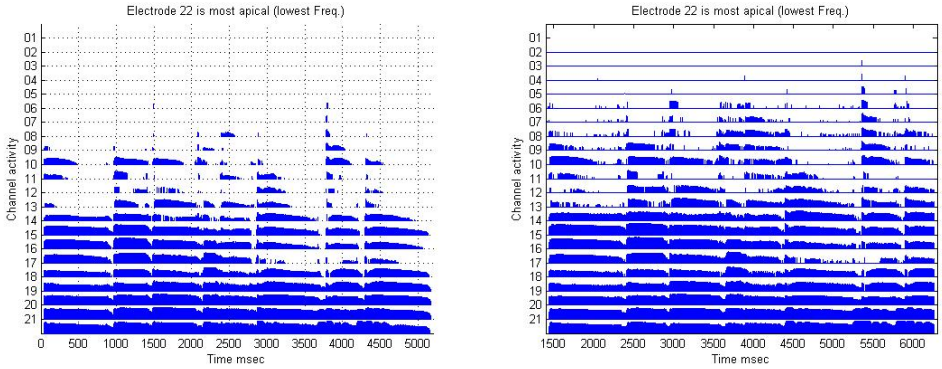


Figure (D.16): Checker generated signal at 63 dBA (left) and RF captured signal at 60 dBA (right) with a guitar signal.

D.2.1 Semitone Mapping Calibration

Semitone mappings (Smt-MF and Smt-LF) were calibrated with respect to Std mapping at a loudness level where all electrodograms are almost the same when single harmonic tones are applied. Figures (D.17 and D.18) show a cello signal being processed at 55dBA with the three maps and in the header the maximum current level is written to be 114, 110 and 113 in Std, Smt-MF and Smt-LF respectively.

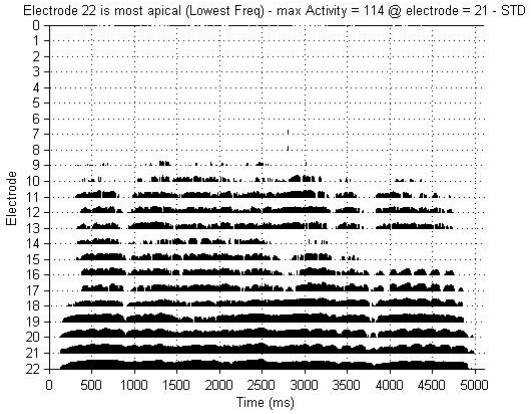


Figure (D.17): Cello signal processing at 55 dBA with Std mapping.

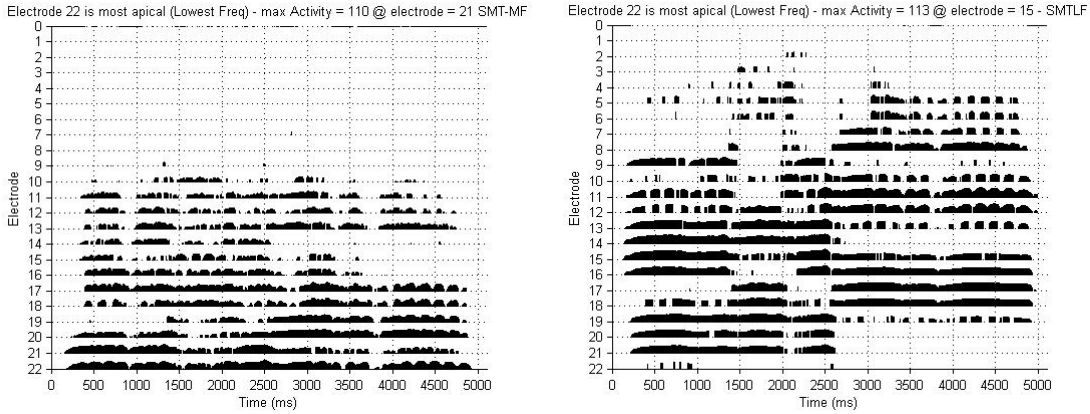
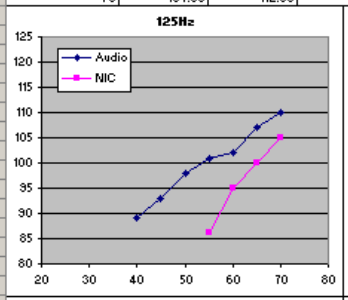
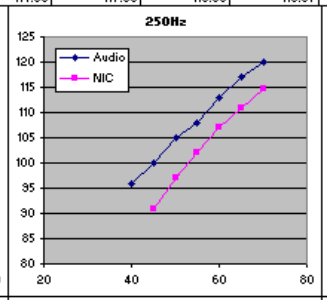


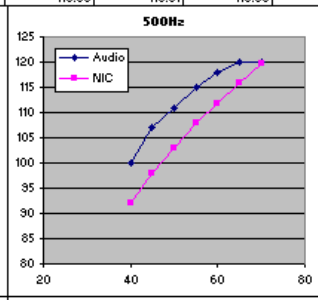
Figure (D.18): Cello signal processing at 55 dBA with Smt-MF (left) map and Smt-Lf (right) map.

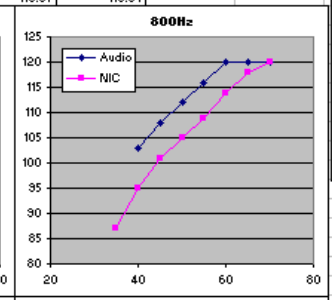
Boost 6 dB

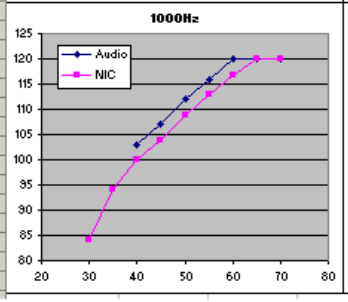
	A	B	C	D	E	F	G	H	I	J	K	L
1	**Audiometer**											
2		**frequency**										
3	**loudness dB**	**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**
4		125	200	250	400	500	630	800	1000	2000	4000	6000
5	40	89	95	96	99	100	105	103	103	106	107	97
6	45	93	99	100	104	107	109	108	107	111	111	102
7	50	98	103	105	108	111	113	112	112	115	115	107
8	55	101	105	108	112	115	117	116	116	118	119	111
9	60	102	110	113	116	118	120	120	120	120	120	115
10	65	107	114	117	120	120	120	120	120	120	120	119
11	70	110	118	120	120	120	120	120	120	120	120	120
12												
13	**NIC Boost=6**											
14		**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**
15	**Ref+Loudness**	**frequency**										
16		125	200	250	400	500	630	800	1000	2000	4000	6000
17	20											
18	25									83.98	86.98	
19	30								83.973	93.92	94.94	82
20	35						86.96	86.987	93.973	99.947	100.95	92.933
21	40				87.973	91.987	94.9	94.947	99.933	103.97	104.97	98.953
22	45		83.98	90.973	95.913	97.933	99.953	100.75	103.95	108.79	108.86	103.96
23	50		93.92	96.927	100.94	102.95	104.95	104.97	108.95	112.83	113.85	107.87
24	55	86.047	99.947	101.95	105.95	107.95	108.79	108.83	112.8	116.85	117.86	111.88
25	60	95.02	103.97	106.95	109.79	111.81	113.81	113.84	116.83	119.87	119.9	115.89
26	65	100.03	108.82	110.82	113.83	115.83	117.82	117.85	119.86	119.93	119.94	119.89
27	70	104.93	112.85	114.83	117.83	119.83	119.87	119.89	119.91	119.95	119.97	119.94

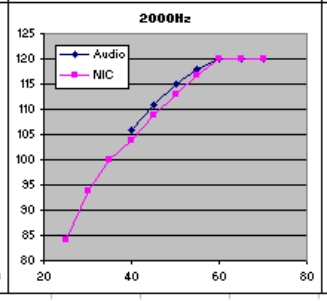


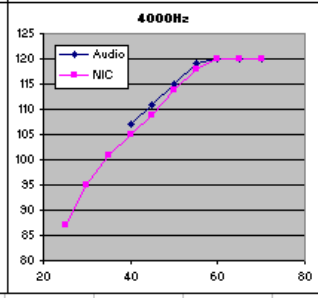


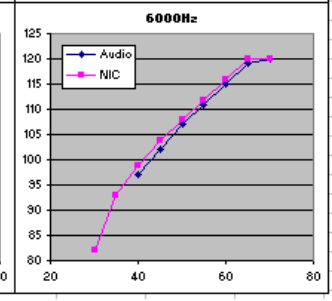


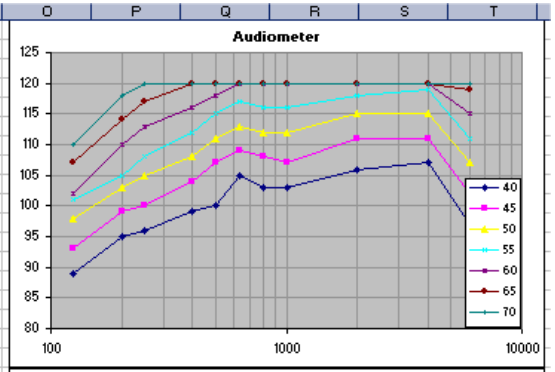


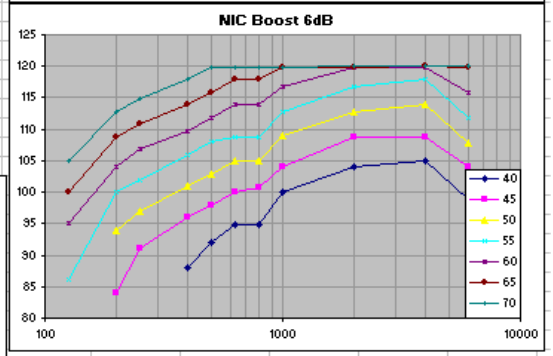






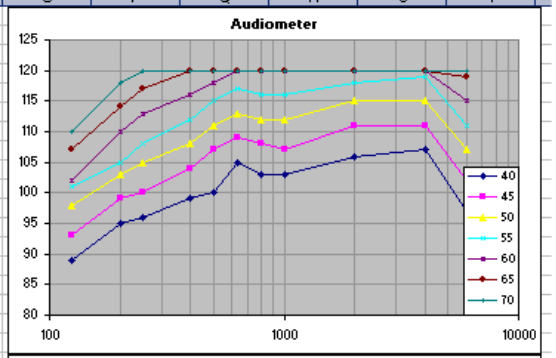


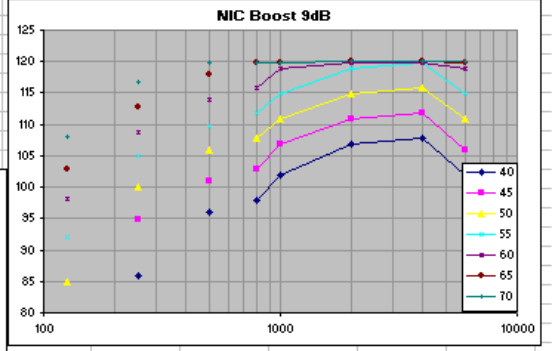


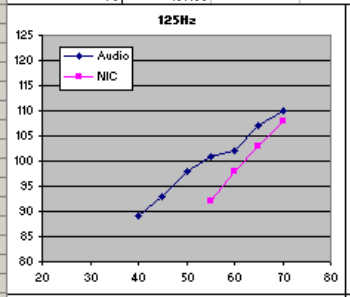


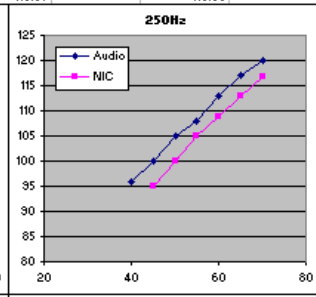
Boost 9 dB

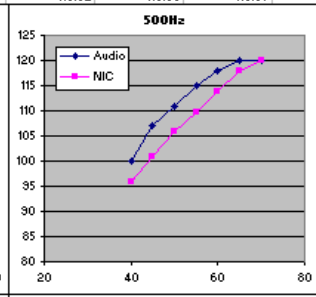
	Audiometer											
		frequency										
loudness dB	**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
	125	200	250	400	500	630	800	1000	2000	4000	6000	
40	89	95	96	99	100	105	103	103	106	107	97	
45	93	99	100	104	107	109	108	107	111	111	102	
50	98	103	105	108	111	113	112	112	115	115	107	
55	101	105	108	112	115	117	116	116	118	119	111	
60	102	110	113	116	118	120	120	120	120	120	115	
65	107	114	117	120	120	120	120	120	120	120	119	
70	110	118	120	120	120	120	120	120	120	120	120	
	NIC Boost=9											
	22	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
Ref+Loudness						**frequency**						
	125	200	250	400	500	630	800	1000	2000	4000	6000	
20												
25									90.987	91.933		
30								90.973	97.94	97.967	89.987	
35					88.973		91.927	96.927	102.95	102.97	96.947	
40			85.973		95.92		97.967	101.96	106.97	107.83	101.97	
45			94.907		100.96		102.97	106.95	110.84	111.86	105.85	
50	85		99.94		105.95		107.81	110.8	114.85	115.87	110.85	
55	92.027		104.95		109.8		111.85	114.83	118.87	119.88	114.86	
60	98.027		108.82		113.84		115.86	118.83	119.91	119.93	118.87	
65	102.93		112.85		117.85		119.87	119.89	119.94	119.96	119.92	
70	107.93		116.87		119.88		119.92	119.93	119.97	119.99	119.96	

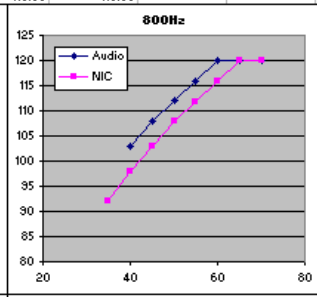


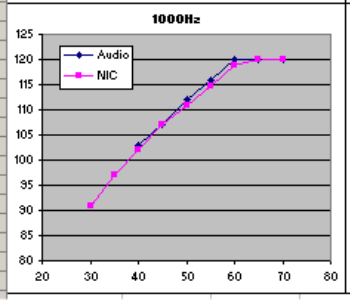


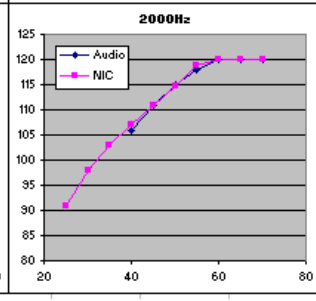


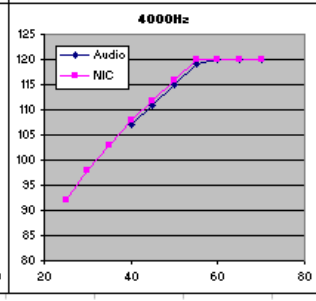


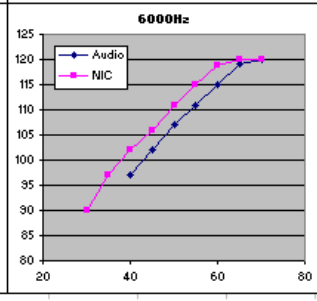






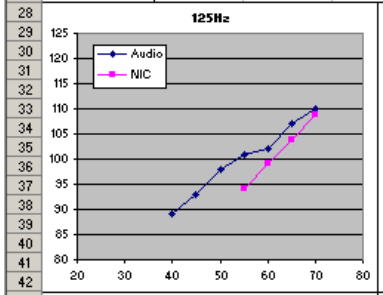


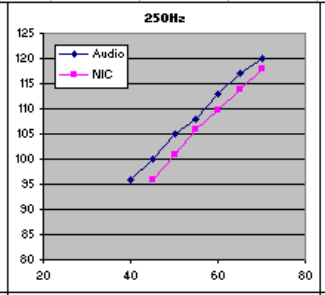


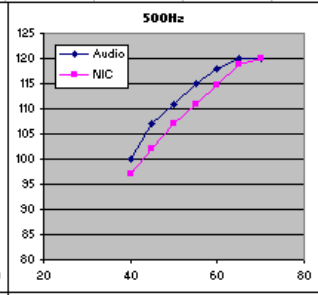


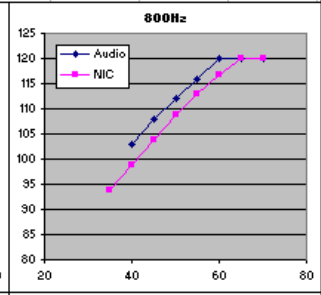
Boost 10 dB

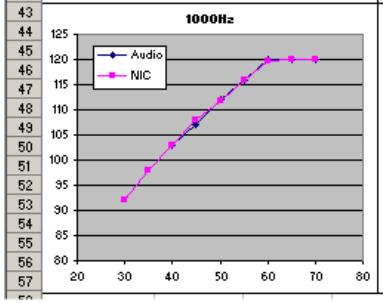
	A	B	C	D	E	F	G	H	I	J	K	L
1	**Audiometer**											
2		**frequency**										
3	**loudness dB**	**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**
4		125	200	250	400	500	630	800	1000	2000	4000	6000
5	40	89	95	96	99	100	105	103	103	106	107	97
6	45	93	99	100	104	107	109	108	107	111	111	102
7	50	98	103	105	108	111	113	112	112	115	115	107
8	55	101	105	108	112	115	117	116	116	118	119	111
9	60	102	110	113	116	118	120	120	120	120	120	115
10	65	107	114	117	120	120	120	120	120	120	120	119
11	70	110	118	120	120	120	120	120	120	120	120	120
12												
13	**NIC Boost=10**											
14		**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**
15	**Ref+Loudness**	**frequency**										
16		125	200	250	400	500	630	800	1000	2000	4000	6000
17	20											
18	25									92.987	93.933	
19	30							83.987	91.987	98.94	99.953	90.933
20	35					90.973		93.927	97.94	103.95	103.97	97.953
21	40			88.966		96.927		98.967	102.96	107.95	108.83	102.96
22	45			95.92		101.96		103.97	107.95	111.84	112.85	106.86
23	50	83.053		100.95		106.95		108.81	111.81	115.86	116.86	110.88
24	55	94.013		105.95		110.8		112.85	115.83	119.86	119.89	115.86
25	60	99.027		109.81		114.83		116.86	119.83	119.92	119.94	118.89
26	65	103.93		113.84		118.83		119.88	119.9	119.95	119.97	119.93
27	70	108.93		117.85		119.89		119.93	119.94	119.97	119.99	119.96

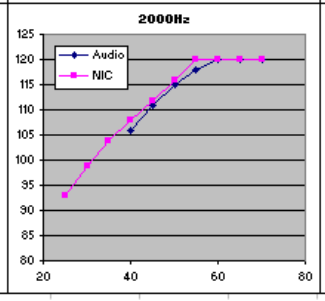


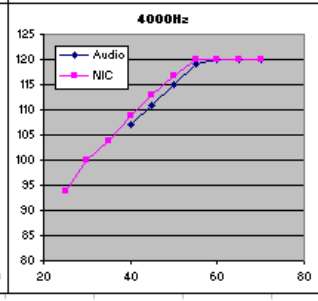


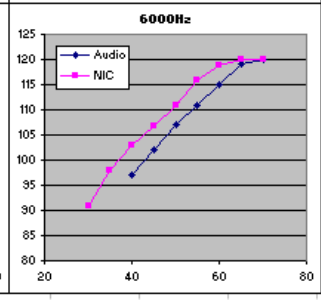


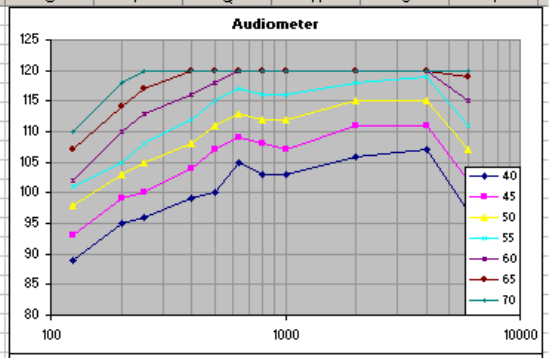


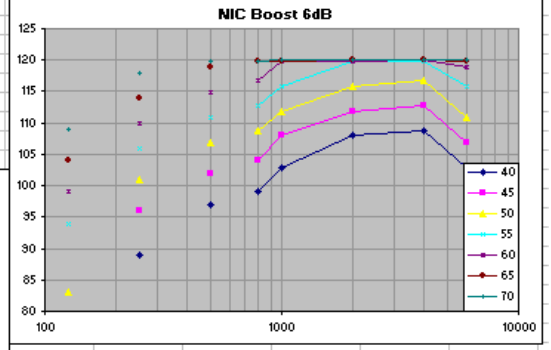






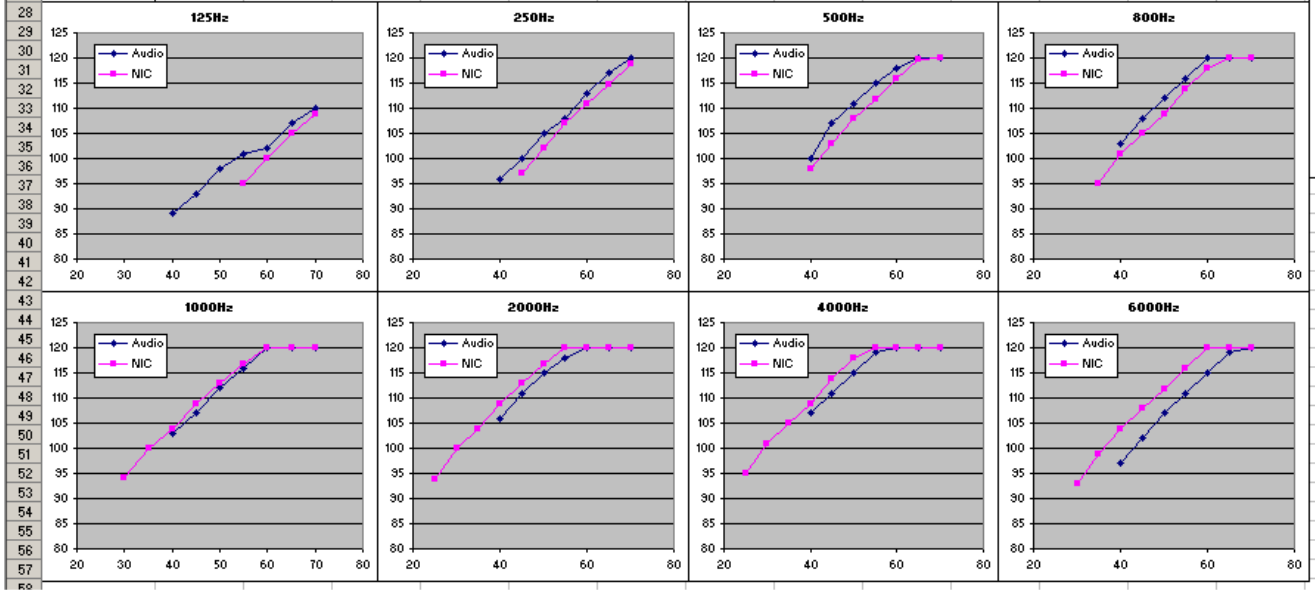


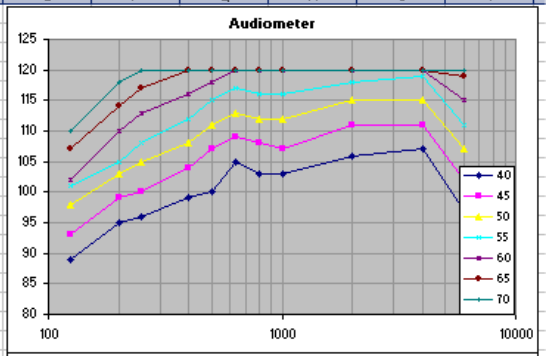


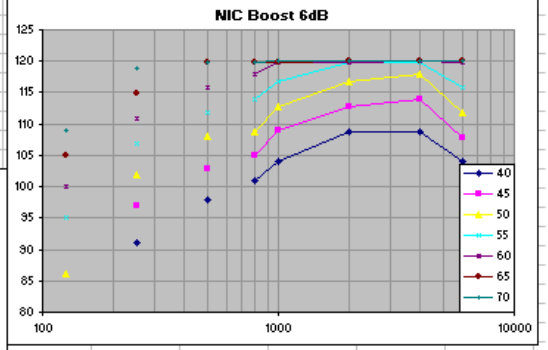


Boost 11 dB

	Audiometer											
		frequency										
loudness dB	**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
	125	200	250	400	500	630	800	1000	2000	4000	6000	
40	89	95	96	99	100	105	103	103	106	107	97	
45	93	99	100	104	107	109	108	107	111	111	102	
50	98	103	105	108	111	113	112	112	115	115	107	
55	101	105	108	112	115	117	116	116	118	119	111	
60	102	110	113	116	118	120	120	120	120	120	115	
65	107	114	117	120	120	120	120	120	120	120	119	
70	110	118	120	120	120	120	120	120	120	120	120	
	NIC Boost=11											
	22	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
Ref+Loudness	**frequency**											
	125	200	250	400	500	630	800	1000	2000	4000	6000	
20												
25								83.973	93.92	94.94	82	
30							86.987	93.973	99.947	100.95	92.933	
35					91.987		94.947	99.933	103.97	104.97	98.953	
40			90.973		97.933		100.95	103.95	108.79	108.86	103.96	
45			96.927		102.95		104.97	108.95	112.83	113.85	107.87	
50	86.047		101.95		107.95		108.83	112.8	116.85	117.86	111.88	
55	95.02		106.95		111.81		113.84	116.83	119.87	119.9	115.89	
60	100.03		110.82		115.83		117.85	119.86	119.93	119.94	119.89	
65	104.93		114.83		119.83		119.89	119.91	119.95	119.97	119.94	
70	108.95		118.85		119.91		119.93	119.94	119.98	119.99	119.97	

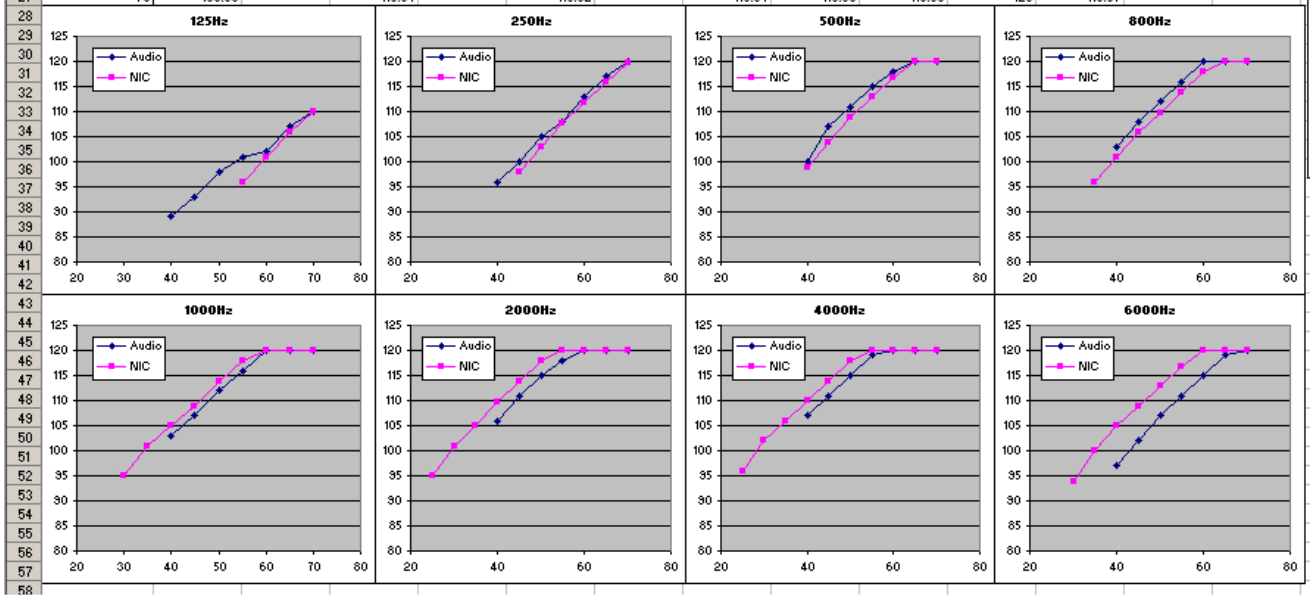


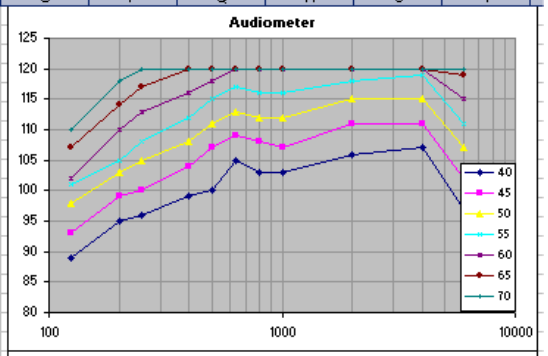


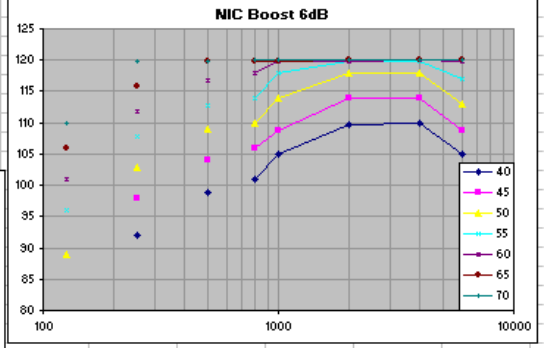


Boost 12 dB

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Audiometer**											
2		**frequency**										
3	**loudness dB**	**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**
4		125	200	250	400	500	630	800	1000	2000	4000	6000
5	40	89	95	96	99	100	105	103	103	106	107	97
6	45	93	99	100	104	107	109	108	107	111	111	102
7	50	98	103	105	108	111	113	112	112	115	115	107
8	55	101	105	108	112	115	117	116	116	118	119	111
9	60	102	110	113	116	118	120	120	120	120	120	115
10	65	107	114	117	120	120	120	120	120	120	120	119
11	70	110	118	120	120	120	120	120	120	120	120	120
12												
13	**NIC Boost=12**											
14		**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**
15	**Ref+Loudness**	**frequency**										
16		125	200	250	400	500	630	800	1000	2000	4000	6000
17	20											
18	25								86.96	94.933	95.96	84.987
19	30					83.973		88.987	94.907	100.95	101.95	93.953
20	35					93.973		95.953	100.93	104.97	105.81	99.967
21	40			91.987		98.947		100.97	104.95	109.8	109.87	104.96
22	45			97.933		103.95		105.97	108.79	113.83	113.88	108.86
23	50	89.02		102.95		108.95		109.84	113.81	117.85	117.89	112.88
24	55	96.027		107.76		112.81		113.87	117.82	119.89	119.91	116.89
25	60	100.91		111.82		116.83		117.87	119.87	119.93	119.95	119.91
26	65	105.93		115.84		119.86		119.91	119.92	119.96	119.97	119.95
27	70	109.95		119.84		119.92		119.94	119.95	119.99	120	119.97

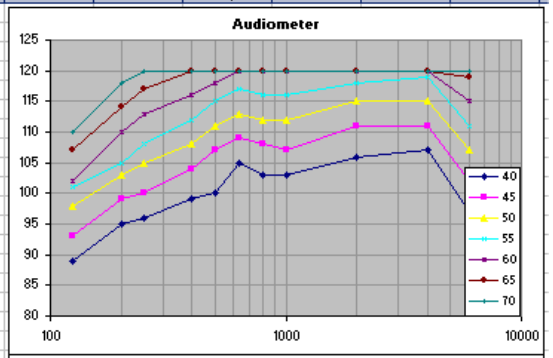


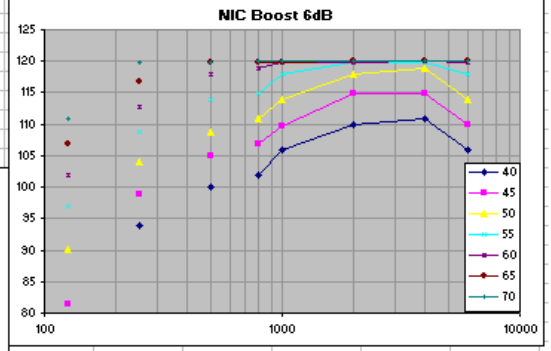


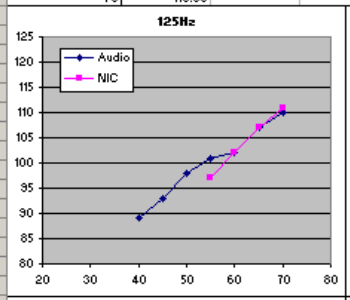


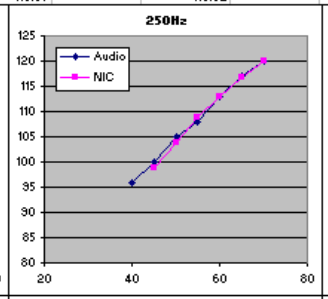
Boost 13 dB

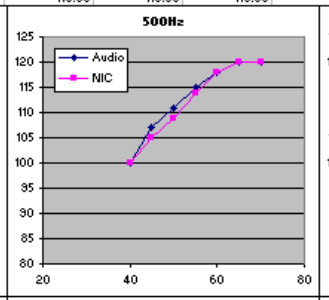
	A	B	C	D	E	F	G	H	I	J	K	L
1	Audiometer											
2		frequency										
3	loudness dB	22	22	22	21	20	19	17	16	11	6	3
4		125	200	250	400	500	630	800	1000	2000	4000	6000
5	40	89	95	96	99	100	105	103	103	106	107	97
6	45	93	99	100	104	107	109	108	107	111	111	102
7	50	98	103	105	108	111	113	112	112	115	115	107
8	55	101	105	108	112	115	117	116	116	118	119	111
9	60	102	110	113	116	118	120	120	120	120	120	115
10	65	107	114	117	120	120	120	120	120	120	120	119
11	70	110	118	120	120	120	120	120	120	120	120	120
12												
13	NIC Boost=13											
14		22	22	22	21	20	19	17	16	11	6	3
15	Ref+Loudness	frequency										
16		125	200	250	400	500	630	800	1000	2000	4000	6000
17	20											
18	25								88.973	95.953	96.967	87.987
19	30					85.973		90.987	95.92	101.95	101.97	95.94
20	35			83		94.907		96.96	100.96	105.97	106.83	100.97
21	40			93.973		99.953		101.97	105.95	109.83	110.86	105.81
22	45	81.5		98.94		104.95		106.97	109.8	114.83	114.88	109.85
23	50	90.04		103.95		108.79		110.85	113.83	117.87	118.89	113.88
24	55	97.027		108.78		113.81		114.86	117.85	119.89	119.93	117.89
25	60	101.93		112.83		117.82		118.87	119.88	119.93	119.95	119.92
26	65	106.93		116.84		119.87		119.91	119.93	119.96	119.98	119.95
27	70	110.95		119.87		119.92		119.95	119.95	119.99	120	119.98

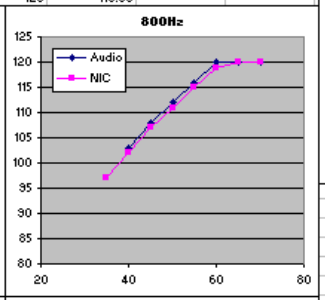


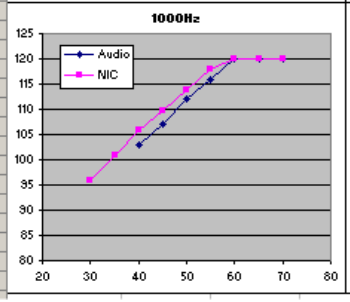


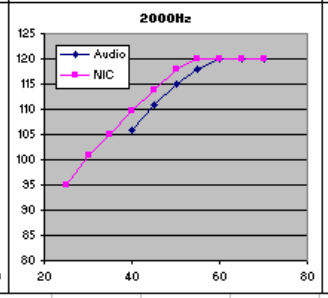


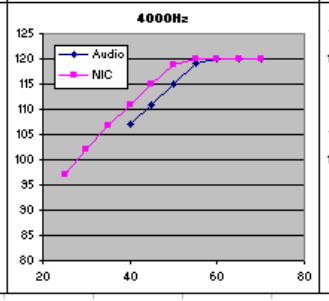


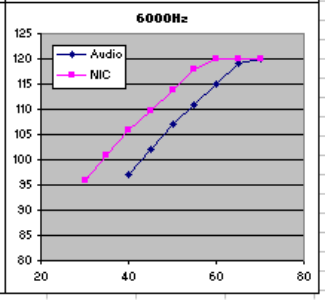






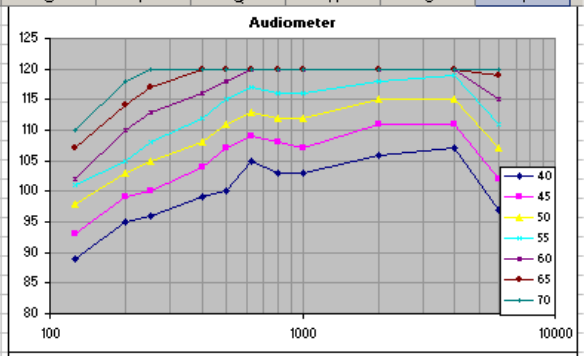


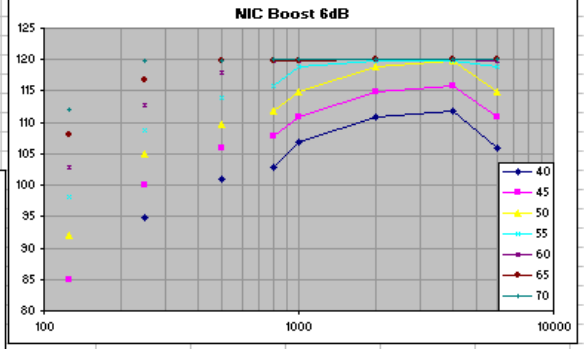


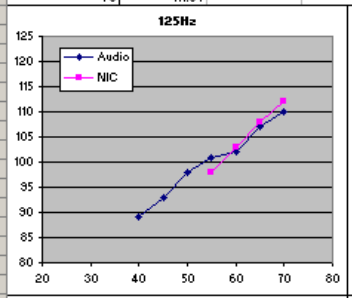


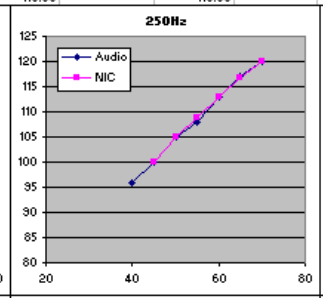
Boost 14 dB

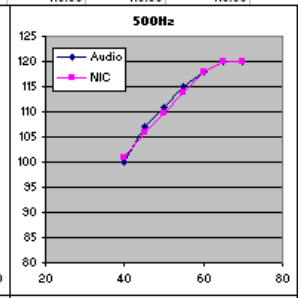
	Audiometer											
	frequency											
loudness dB	22	22	22	21	20	19	17	16	11	6	3	
	125	200	250	400	500	630	800	1000	2000	4000	6000	
40	89	95	96	99	100	105	103	103	106	107	97	
45	93	99	100	104	107	109	108	107	111	111	102	
50	98	103	105	108	111	113	112	112	115	115	107	
55	101	105	108	112	115	117	116	116	118	119	111	
60	102	110	113	116	118	120	120	120	120	120	115	
65	107	114	117	120	120	120	120	120	120	120	119	
70	110	118	120	120	120	120	120	120	120	120	120	
	NIC Boost=14											
	22	22	22	21	20	19	17	16	11	6	3	
Ref+Loudness	frequency											
	125	200	250	400	500	630	800	1000	2000	4000	6000	
20												
25								90.973	97.94	97.967	89.987	
30					88.973		91.927	96.927	102.95	102.97	96.947	
35			85.973		95.92		97.967	101.96	106.97	107.83	101.97	
40			94.907		100.96		102.97	106.95	110.84	111.86	105.85	
45	85		99.94		105.95		107.81	110.8	114.85	115.87	110.85	
50	92.027		104.95		109.8		111.85	114.83	118.87	119.88	114.86	
55	98.027		108.82		113.84		115.86	118.83	119.91	119.93	118.87	
60	102.93		112.85		117.85		119.87	119.89	119.94	119.96	119.92	
65	107.93		116.87		119.88		119.92	119.93	119.97	119.99	119.96	
70	111.94		119.88		119.93		119.95	119.96	119.99	120	119.99	

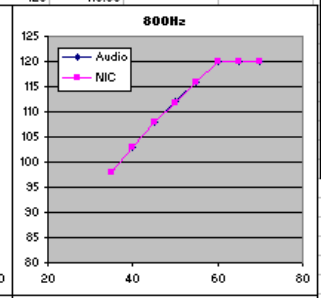


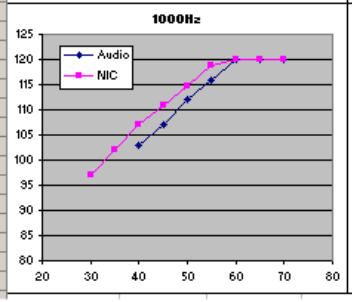


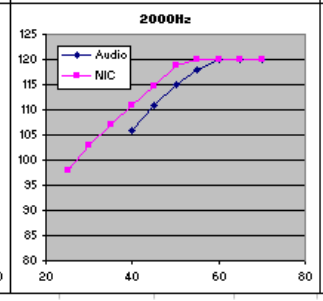


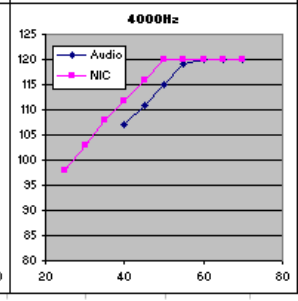


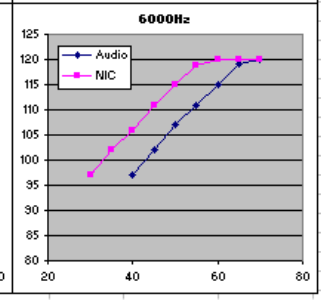






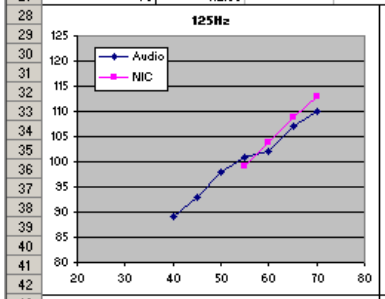


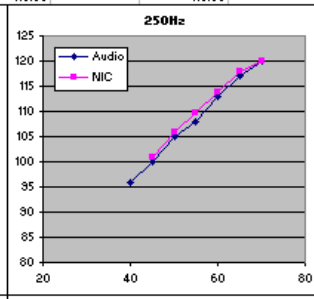


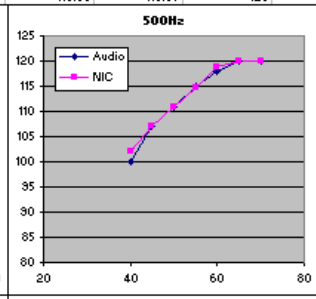


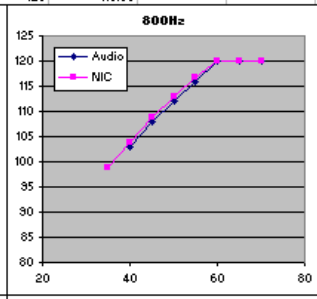
Boost 15 dB

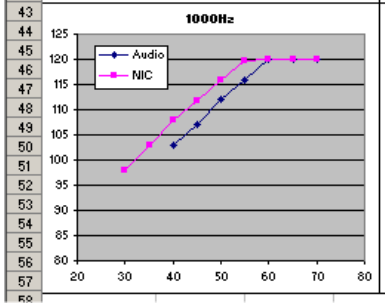
	Audiometer											
		frequency										
loudness dB	**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
	125	200	250	400	500	630	800	1000	2000	4000	6000	
40	89	95	96	99	100	105	103	103	106	107	97	
45	93	99	100	104	107	109	108	107	111	111	102	
50	98	103	105	108	111	113	112	112	115	115	107	
55	101	105	108	112	115	117	116	116	118	119	111	
60	102	110	113	116	118	120	120	120	120	120	115	
65	107	114	117	120	120	120	120	120	120	120	119	
70	110	118	120	120	120	120	120	120	120	120	120	
	NIC Boost=15											
	22	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
Ref+Loudness	**frequency**											
	125	200	250	400	500	630	800	1000	2000	4000	6000	
20												
25							83.987	91.987	98.94	99.947	90.933	
30					90.973		93.927	97.94	103.95	103.97	97.953	
35			88.966		96.927		98.967	102.96	107.95	108.83	102.96	
40			95.92		101.96		103.97	107.95	111.84	112.85	106.86	
45	83.053		100.95		106.95		108.81	111.81	115.86	116.86	110.88	
50	94.013		105.95		110.8		112.85	115.83	119.86	119.89	115.86	
55	99.027		109.81		114.83		116.86	119.83	119.92	119.94	118.89	
60	103.93		113.84		118.83		119.88	119.9	119.95	119.97	119.93	
65	108.93		117.85		119.89		119.93	119.94	119.97	119.99	119.96	
70	112.93		119.89		119.93		119.95	119.97	120	120	119.99	

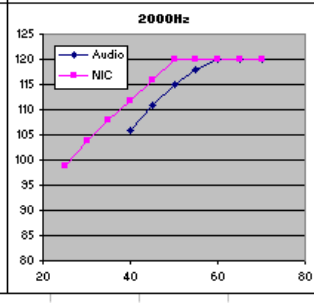


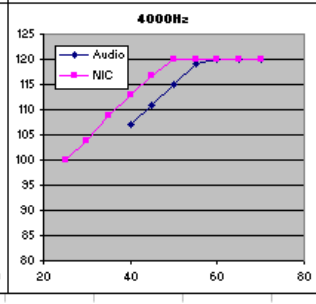


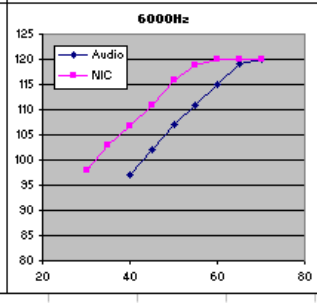


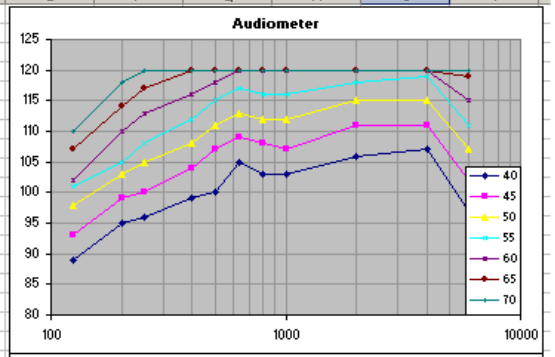


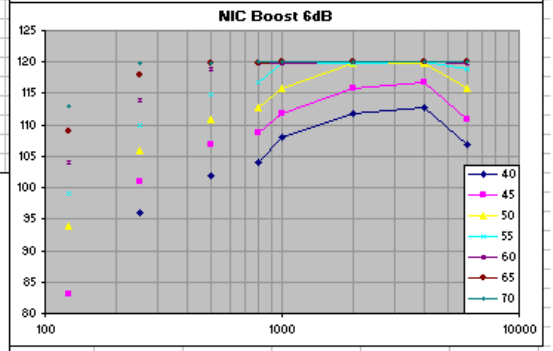






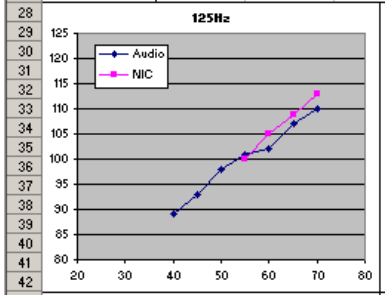


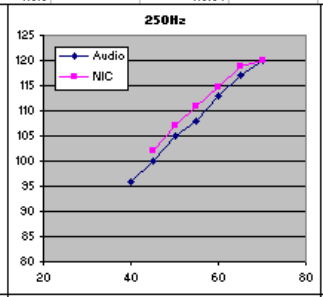


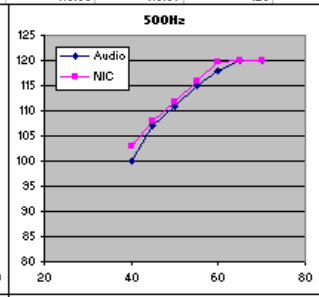


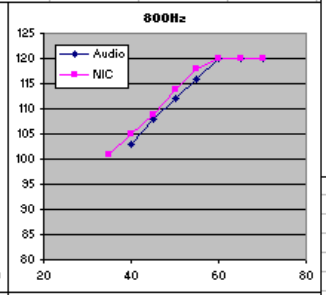
Boost 16 dB

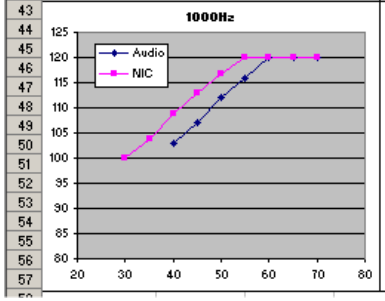
	Audiometer											
		frequency										
loudness dB	**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
	125	200	250	400	500	630	800	1000	2000	4000	6000	
40	89	95	96	99	100	105	103	103	106	107	97	
45	93	99	100	104	107	109	108	107	111	111	102	
50	98	103	105	108	111	113	112	112	115	115	107	
55	101	105	108	112	115	117	116	116	118	119	111	
60	102	110	113	116	118	120	120	120	120	120	115	
65	107	114	117	120	120	120	120	120	120	120	119	
70	110	118	120	120	120	120	120	120	120	120	120	
	NIC Boost=16											
	22	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
Ref+Loudness						**frequency**						
	125	200	250	400	500	630	800	1000	2000	4000	6000	
20												
25							86.987	93.98	99.947	100.95	92.933	
30					91.987		94.947	99.933	103.97	104.97	98.953	
35			90.973		97.933		100.95	103.95	108.79	108.86	103.96	
40			96.927		102.95		104.97	108.95	112.83	113.85	107.87	
45	86.047		101.95		107.95		108.83	112.8	116.85	117.86	111.88	
50	95.02		106.95		111.81		113.84	116.83	119.87	119.9	115.89	
55	100.03		110.82		115.83		117.85	119.86	119.93	119.94	119.89	
60	104.93		114.83		119.83		119.89	119.91	119.95	119.97	119.94	
65	108.95		118.85		119.91		119.93	119.94	119.98	119.99	119.97	
70	112.96		119.9		119.94		119.96	119.97	120	120	119.99	

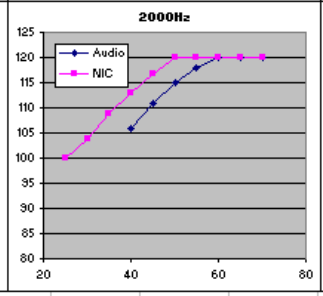


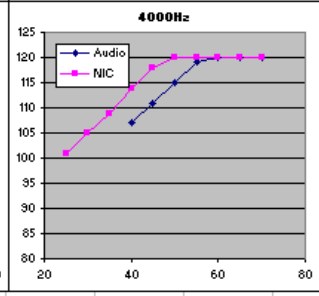


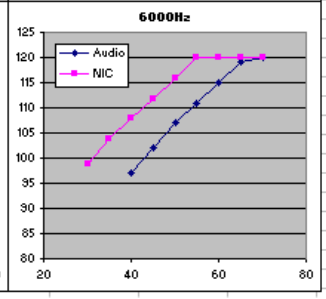


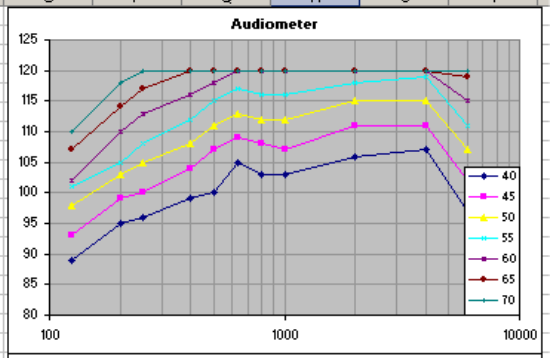


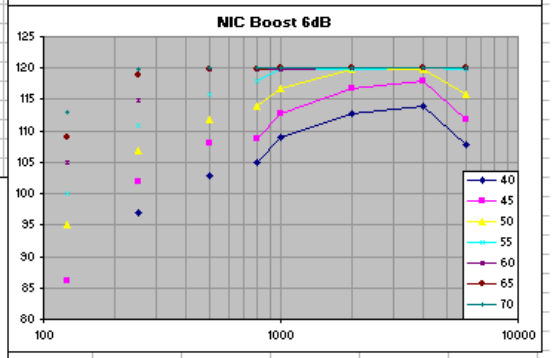






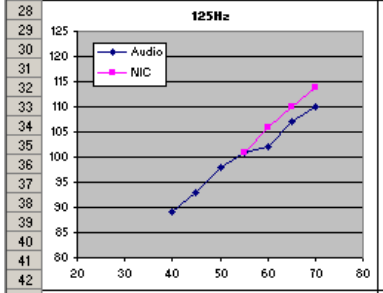


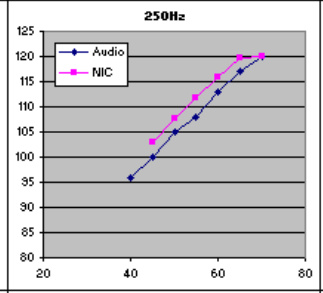


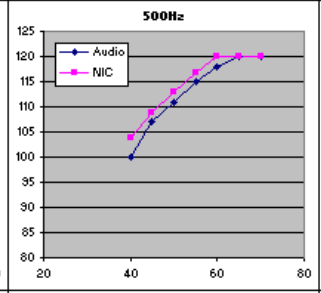


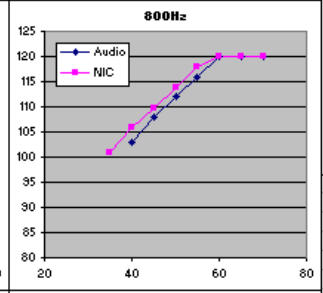
Boost 17 dB

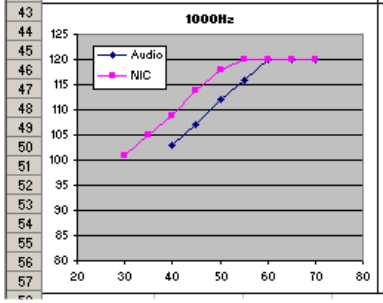
	Audiometer											
		frequency										
loudness dB	**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
	125	200	250	400	500	630	800	1000	2000	4000	6000	
40	89	95	96	99	100	105	103	103	106	107	97	
45	93	99	100	104	107	109	108	107	111	111	102	
50	98	103	105	108	111	113	112	112	115	115	107	
55	101	105	108	112	115	117	116	116	118	119	111	
60	102	110	113	116	118	120	120	120	120	120	115	
65	107	114	117	120	120	120	120	120	120	120	119	
70	110	118	120	120	120	120	120	120	120	120	120	
	NIC Boost=17											
	22	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
Ref+Loudness	**frequency**											
	125	200	250	400	500	630	800	1000	2000	4000	6000	
20												
25					83.973		88.987	94.907	100.95	100.97	93.953	
30					93.973		95.953	100.93	104.97	105.81	99.96	
35			91.987		98.947		100.97	104.95	109.8	109.87	104.96	
40			97.933		103.95		105.97	108.79	113.83	113.88	108.86	
45	89.02		102.95		108.95		109.84	113.81	117.85	117.89	112.88	
50	96.027		107.76		112.81		113.87	117.82	119.89	119.91	116.89	
55	100.91		111.82		116.83		117.87	119.87	119.93	119.95	119.91	
60	105.93		115.84		119.86		119.91	119.92	119.96	119.97	119.95	
65	109.95		119.84		119.92		119.94	119.95	119.99	120	119.97	
70	113.95		119.91		119.95		119.97	119.97	120	120	120	

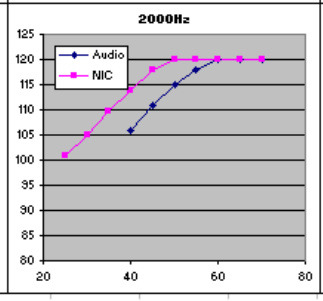


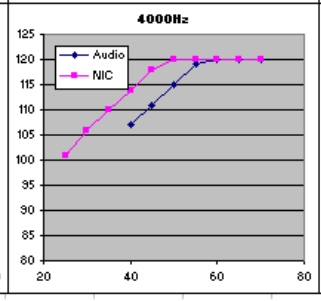


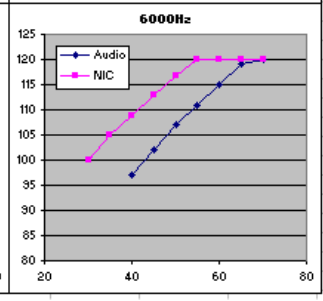


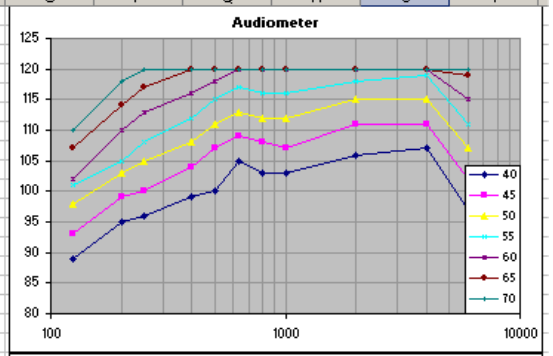


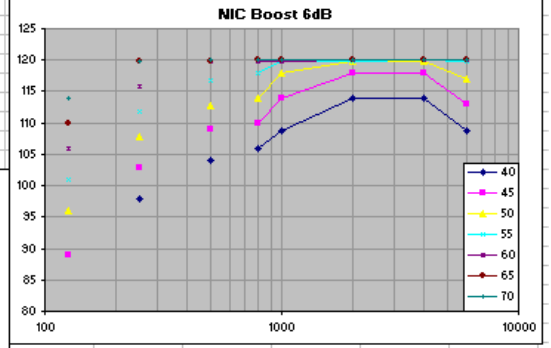






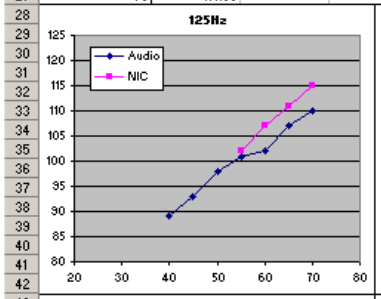


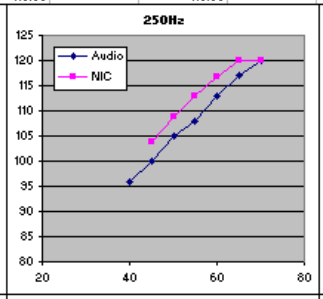


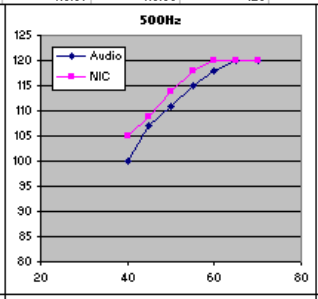


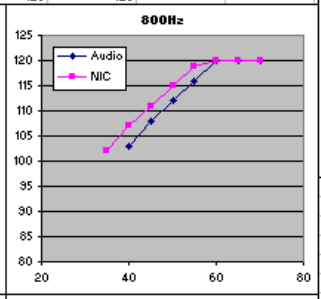
Boost 18 dB

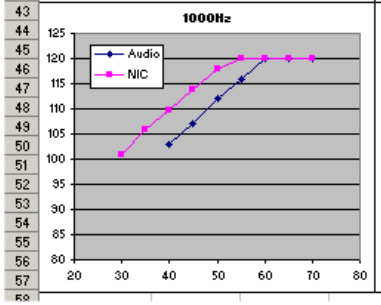
	Audiometer											
		frequency										
loudness dB	**22**	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
	125	200	250	400	500	630	800	1000	2000	4000	6000	
40	89	95	96	99	100	105	103	103	106	107	97	
45	93	99	100	104	107	109	108	107	111	111	102	
50	98	103	105	108	111	113	112	112	115	115	107	
55	101	105	108	112	115	117	116	116	118	119	111	
60	102	110	113	116	118	120	120	120	120	120	115	
65	107	114	117	120	120	120	120	120	120	120	119	
70	110	118	120	120	120	120	120	120	120	120	120	
	NIC Boost=11											
	22	**22**	**22**	**21**	**20**	**19**	**17**	**16**	**11**	**6**	**3**	
Ref+Loudness	**frequency**											
	125	200	250	400	500	630	800	1000	2000	4000	6000	
20												
25					86.96		90.987	95.92	101.95	101.97	95.94	
30			83		94.907		96.96	100.96	105.97	106.83	100.97	
35			93.973		99.953		101.97	105.95	109.83	110.86	105.81	
40	81.5		98.94		104.95		106.97	109.8	114.83	114.88	109.85	
45	90.04		103.95		108.79		110.85	113.83	117.87	118.89	113.88	
50	97.027		108.78		113.81		114.86	117.85	119.89	119.93	117.89	
55	101.93		112.83		117.82		118.87	119.88	119.93	119.95	119.92	
60	106.93		116.84		119.87		119.91	119.93	119.96	119.98	119.95	
65	110.95		119.87		119.92		119.95	119.95	119.99	120	119.98	
70	114.95		119.93		119.95		119.97	119.98	120	120	120	

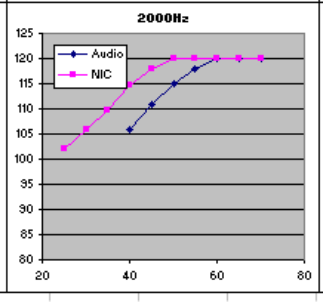


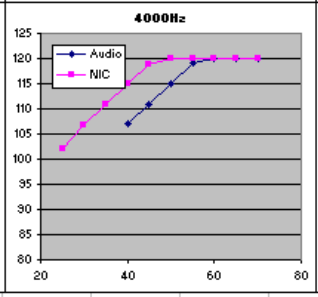


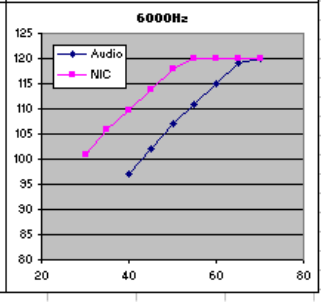


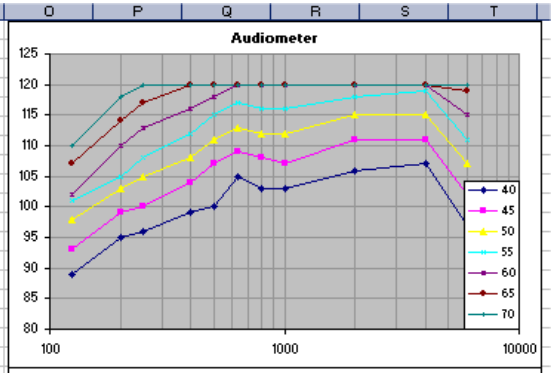


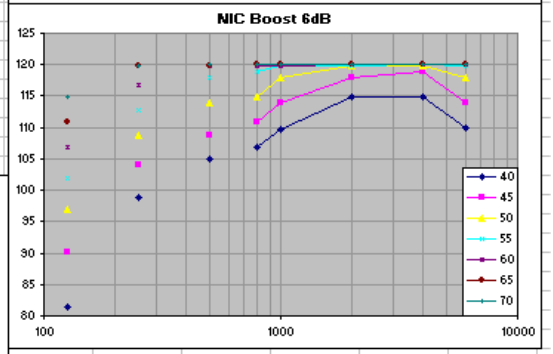






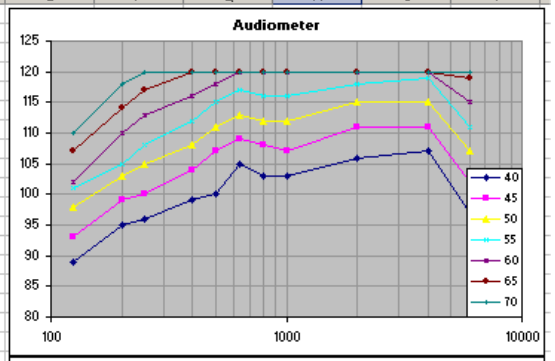


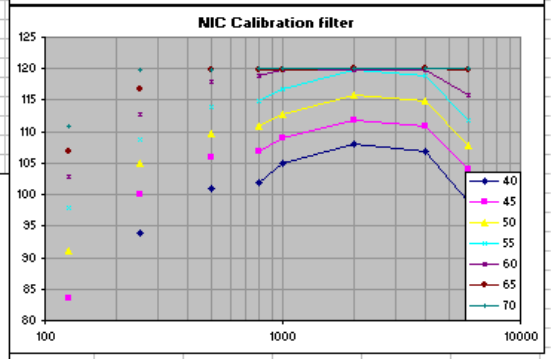


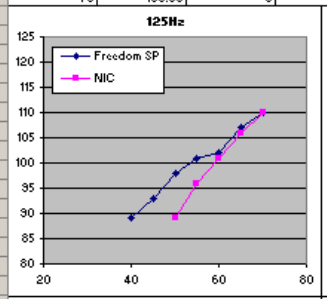


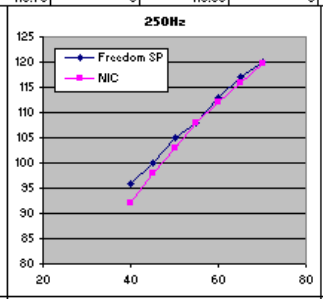
D.3 Verification of the calibration function

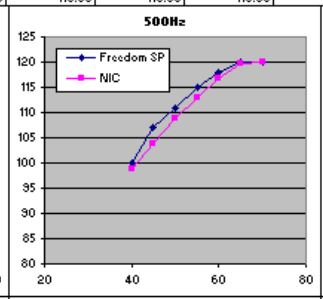
					Audiometer							
					frequency							
loudness dl	22	22	22	21	20	19	17	16	11	6	3	
	125	200	250	400	500	630	800	1000	2000	4000	6000	
40	89	95	96	99	100	105	103	103	106	107	97	
45	93	99	100	104	107	109	108	107	111	111	102	
50	98	103	105	108	111	113	112	112	115	115	107	
55	101	105	108	112	115	117	116	116	118	119	111	
60	102	110	113	116	118	120	120	120	120	120	115	
65	107	114	117	120	120	120	120	120	120	120	119	
70	110	118	120	120	120	120	120	120	120	120	120	
					Calibration Filter							
	22	22	22	21	20	19	17	16	11	6	3	
Ref+Loudn						frequency						
	125	200	250	400	500	630	800	1000	2000	4000	6000	
20												
25												
30												
35												
40		0	91.966	0	98.913	0	100.93	102.93	105.95	104.95	96.94	
45		0	97.893	0	103.93	0	104.95	107.92	110.94	108.97	102.94	
50	88.98	0	102.93	0	108.93	0	109.94	111.93	114.73	113.77	106.96	
55	95.993	0	107.93	0	112.93	0	113.75	115.71	118.78	117.8	110.79	
60	101.01	0	111.93	0	116.75	0	117.79	119.75	119.84	119.85	114.83	
65	106	0	115.77	0	119.79	0	119.83	119.84	119.89	119.91	118.84	
70	109.88	0	119.79	0	119.86	0	119.89	119.89	119.93	119.93	119.89	

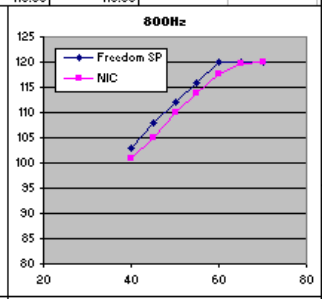


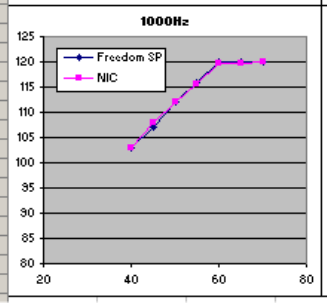


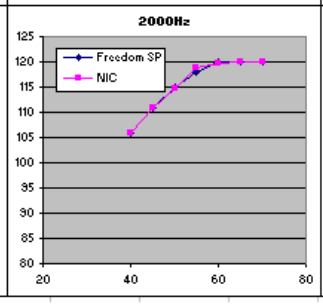


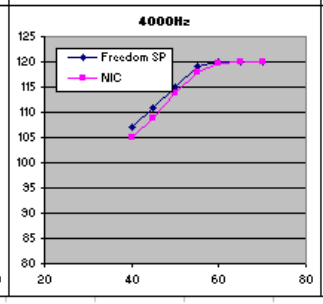


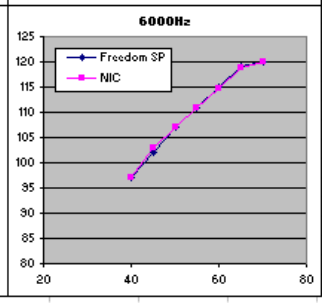












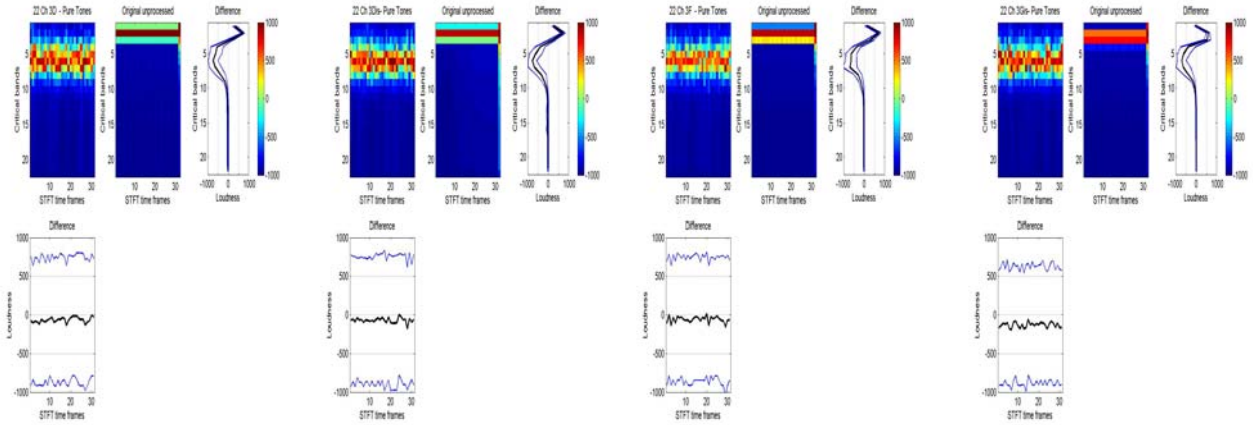
Appendix – E (Spectral-Contrast analysis)

E.1 Processed and Unprocessed Pure Tones

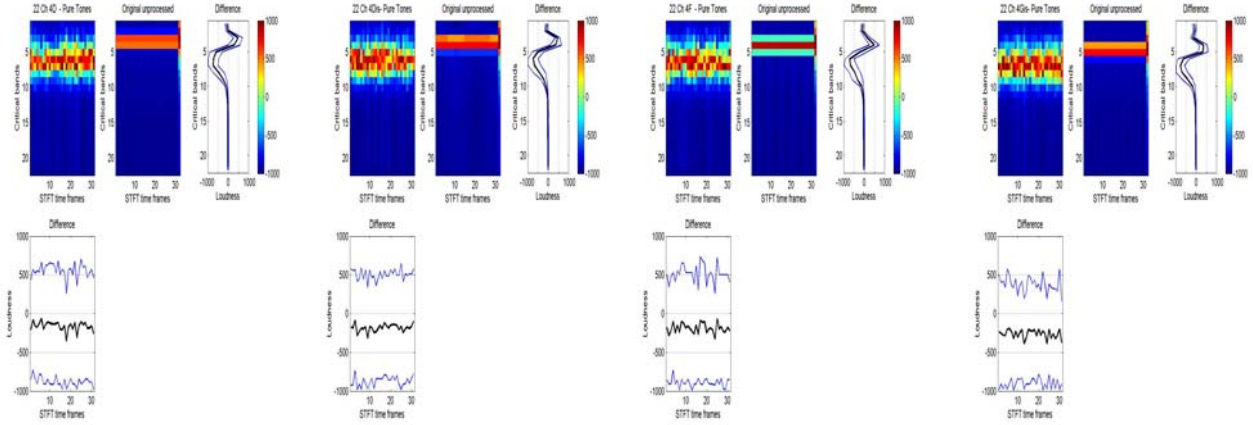
This section shows comparisons between processed (left) and unprocessed (right) pure tones using 22 channels. Signals were processed with the AMO and resynthesizing used noise band vocoder. In each block, the contrast in the critical bands and time dimension are on the right side and beneath the main graph with a blue background respectively. The main graph shows the processed signal (top-left) and the pure tone (top-middle). This was done for a reference note (D), 1, 3 and 6 semitones shift in octaves 3, 4 and 5.

Reference 1 Semitone 3 Semitone 6 Semitone

Octave 3



Octave 4



Octave 5

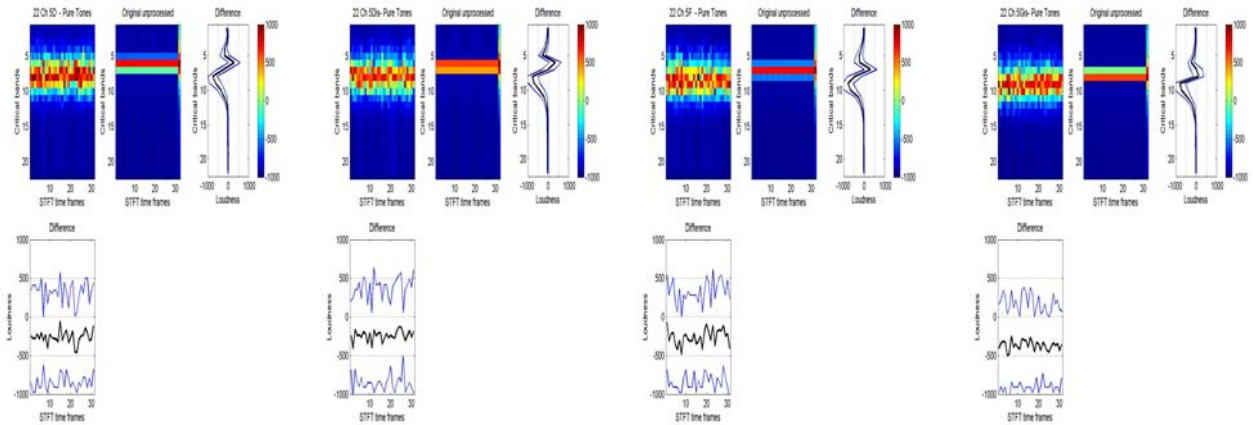


Figure (E.1): Comparison between processed for 22 channels and unprocessed pure tones.

This section shows comparisons between processed (left) and unprocessed (right) pure tones using 43 channels. Signals were processed with the AMO and resynthesizing used noise band vocoder. In each block, the contrast in the critical bands and time dimension are on the right side and beneath the main graph with a blue background respectively. The main graph shows the processed signal (top-left) and the pure tone (top-middle). This was done for a reference note (D), 1, 3 and 6 semitones shift in octaves 3, 4 and 5.

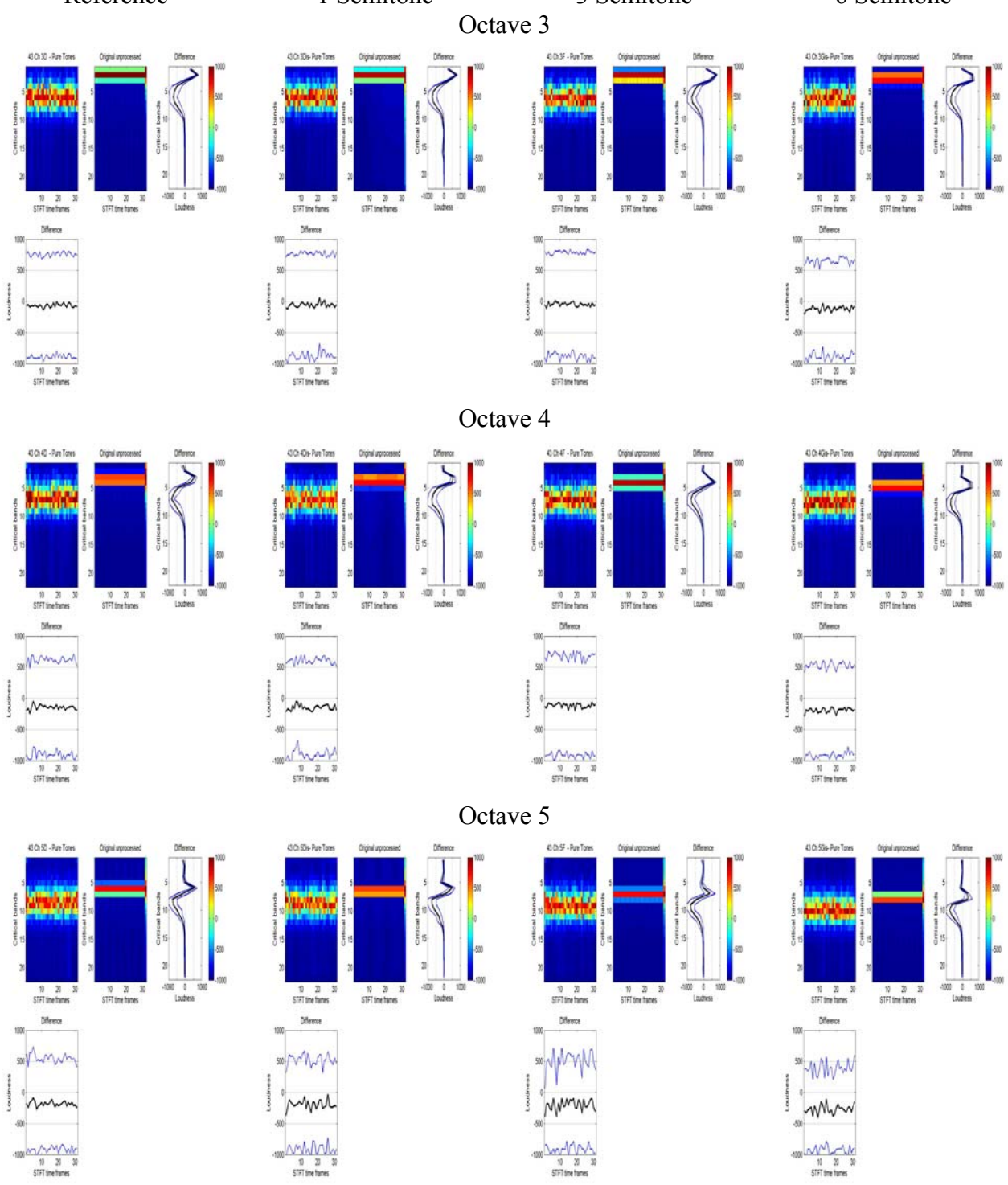


Figure (E.2): Comparison between processed for 43 channels and unprocessed pure tones.

E.2 Processed and Reference Pure Tones

Comparison between pure and reference tones, both were processed and resynthesized with the AMO using noise vocoder with Std mapping and 22 channels. The distance between the pure tone and the reference was varied among 1, 3 and 6 semitones in octaves 3, 4 and 5.

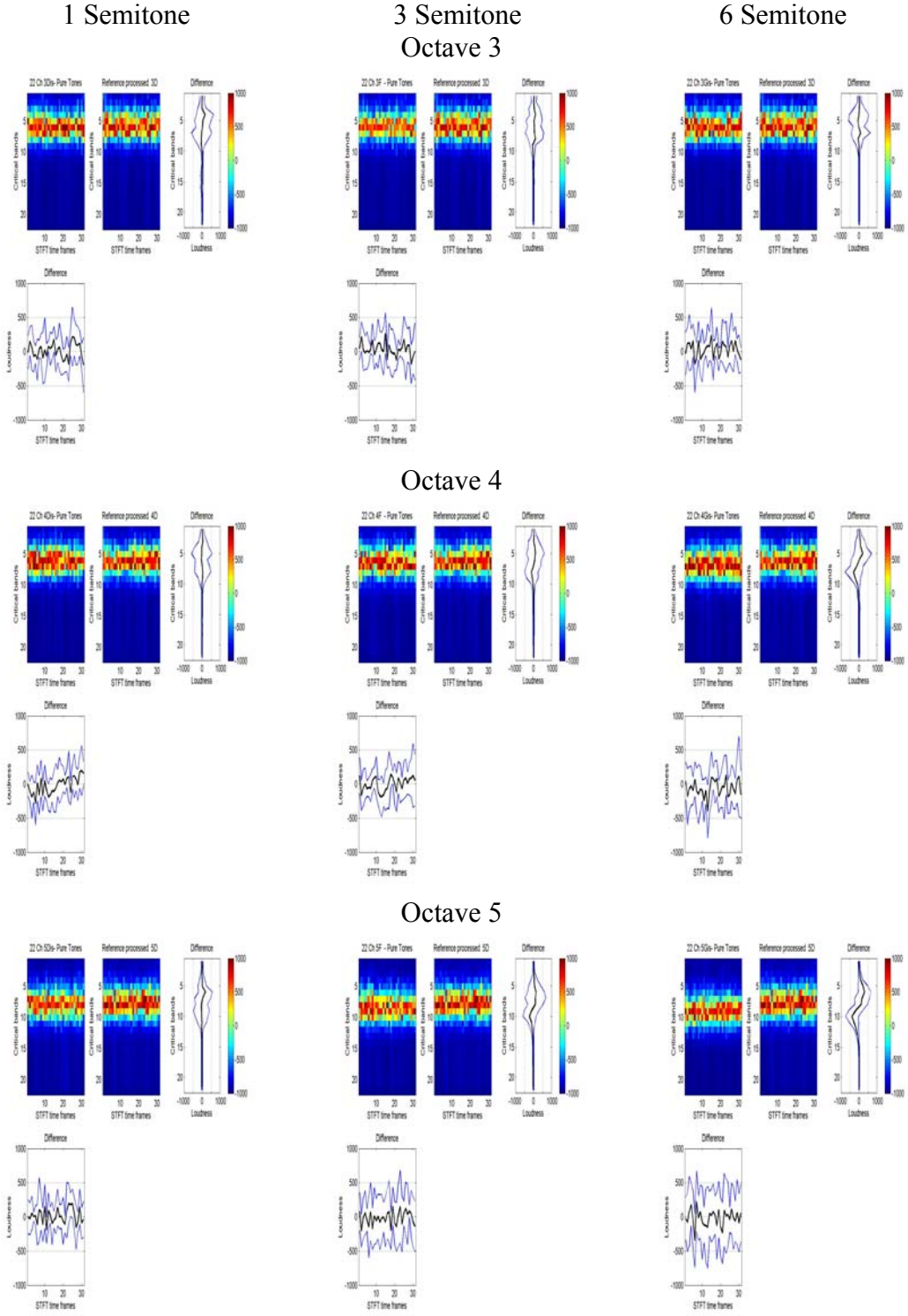


Figure (E.3): Comparison between pure and reference tones processed and resynthesized with the AMO using 22 channels and noise band vocoder for 1, 3 and 6 semitones differences.

Comparison between pure and reference tones, both were processed and resynthesized with the AMO using noise vocoder using Std mapping with 43 channels. The distance between the pure tone and the reference was varied among 1, 3 and 6 semitones in octaves 3, 4 and 5.

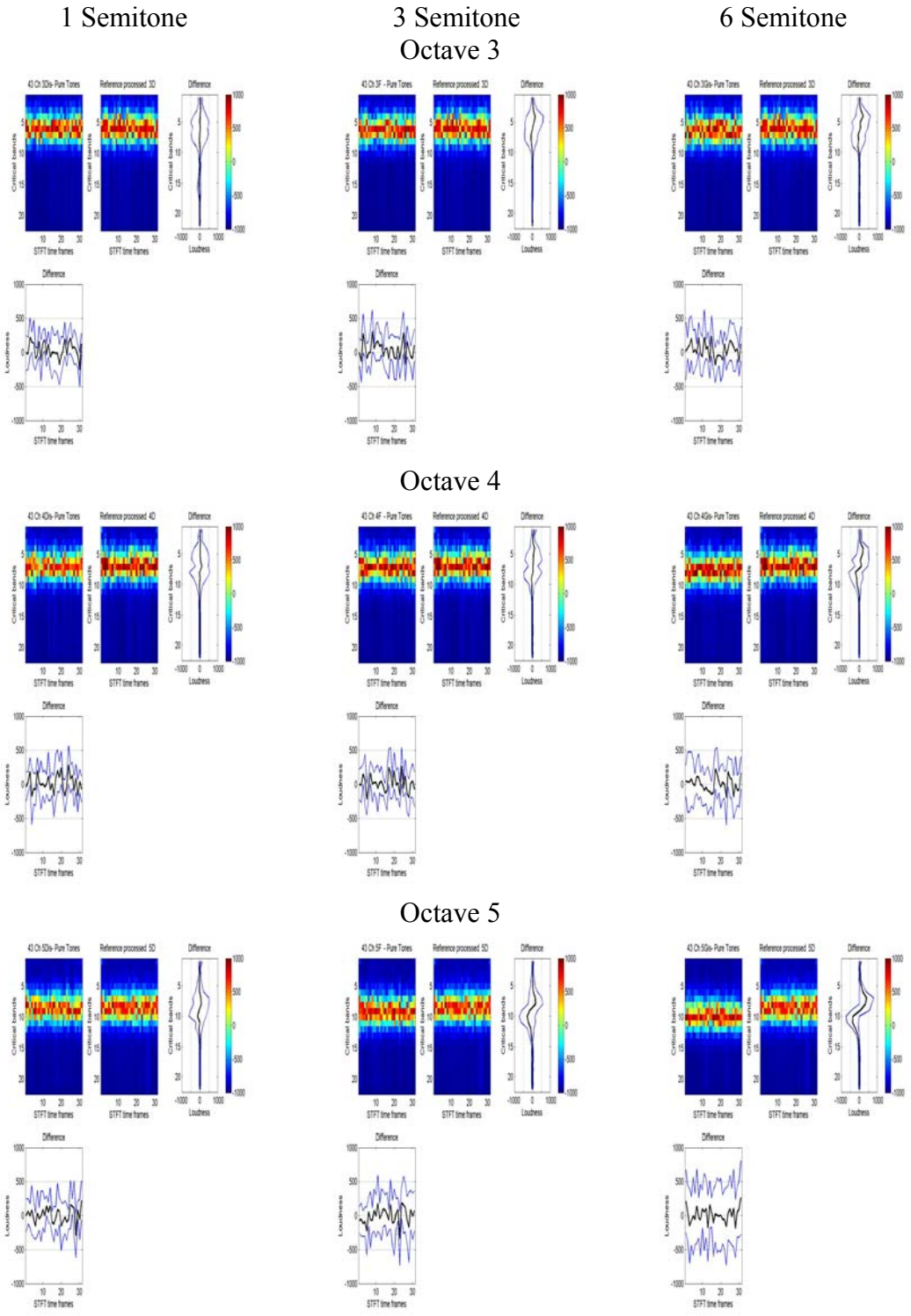


Figure (E.4): Comparison between pure and reference tones processed and resynthesized with the AMO using 43 channels and noise band vocoder for 1, 3 and 6 semitones differences.

E.3 Processed and Unprocessed Complex Tones

Comparison between processed complex tones and unprocessed versions; the processed tones were resynthesized with the AMO using noise vocoder using Std mapping with 22 channels. The distance between the pure tone and the reference was varied among 1, 3 and 6 semitones in octaves 3, 4 and 5.

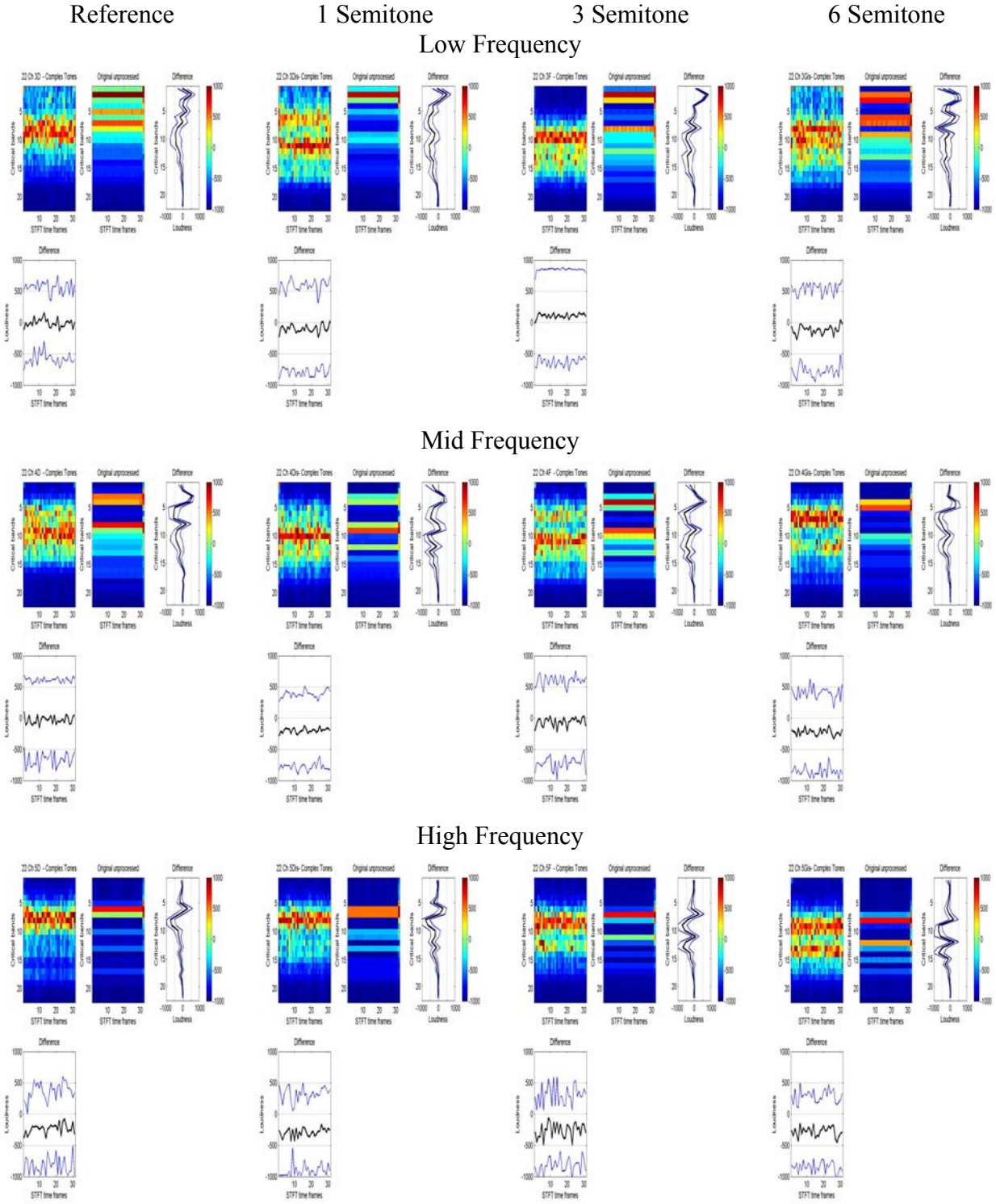


Figure (E.5): Comparison between processed complex tones and unprocessed versions. The processed tones were resynthesized with the AMO using 22 channels and noise vocoders.

Comparison between processed complex tones and unprocessed versions; the processed tones were resynthesized with the AMO using noise vocoder using Std mapping with 43 channels. The distance between the pure tone and the reference was varied among 1, 3 and 6 semitones in octaves 3, 4 and 5.

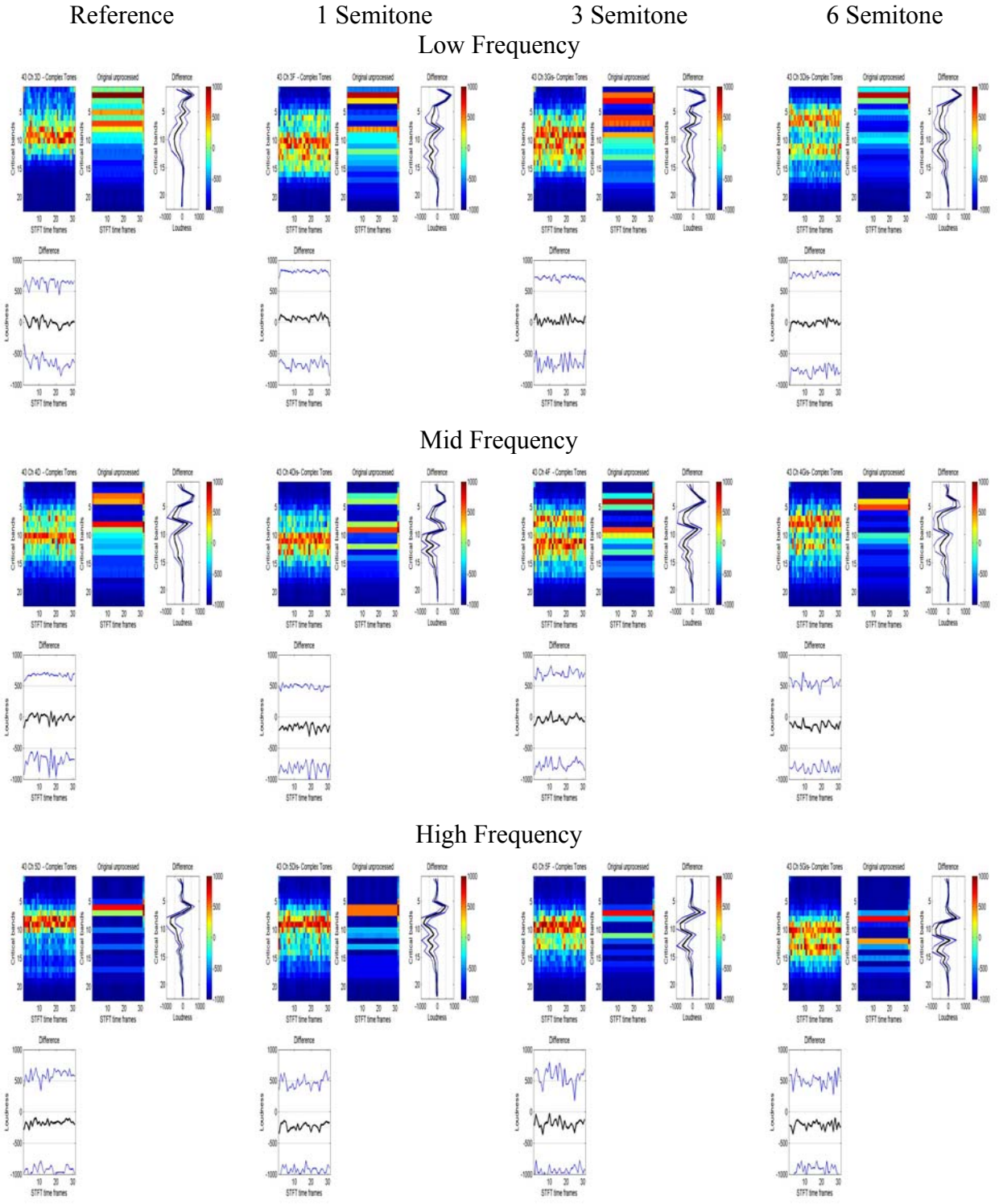


Figure (E.6): Comparison between processed complex tones and unprocessed versions. The processed tones were resynthesized with the AMO using 22 channels and noise vocoders.

E.4 Processed and Reference Complex Tones

Comparison between complex and reference tones, both were processed and resynthesized with the AMO using noise vocoder with Std mapping and 22 channels. The distance between the complex tone and the reference was varied among 1, 3 and 6 semitones in octaves 3, 4 and 5.

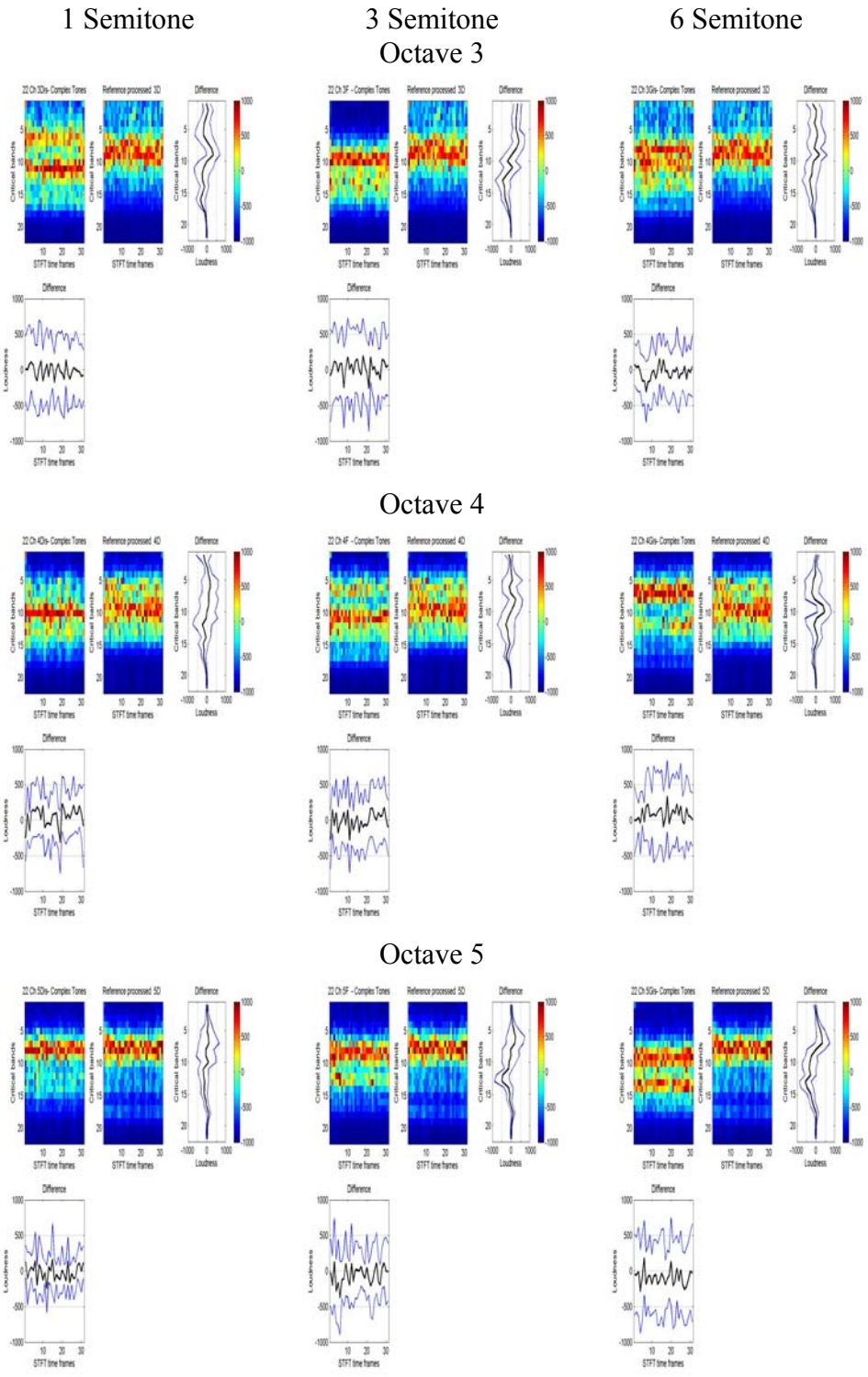


Figure (E.7): Comparison between complex and reference tones processed and resynthesized with the AMO using 22 channels and noise band vocoder for 1, 3 and 6 semitones differences.

Comparison between complex and reference tones, both were processed and resynthesized with the AMO using noise vocoder with Std mapping and 43 channels. The distance between the complex tone and the reference was varied among 1, 3 and 6 semitones in octaves 3, 4 and 5.

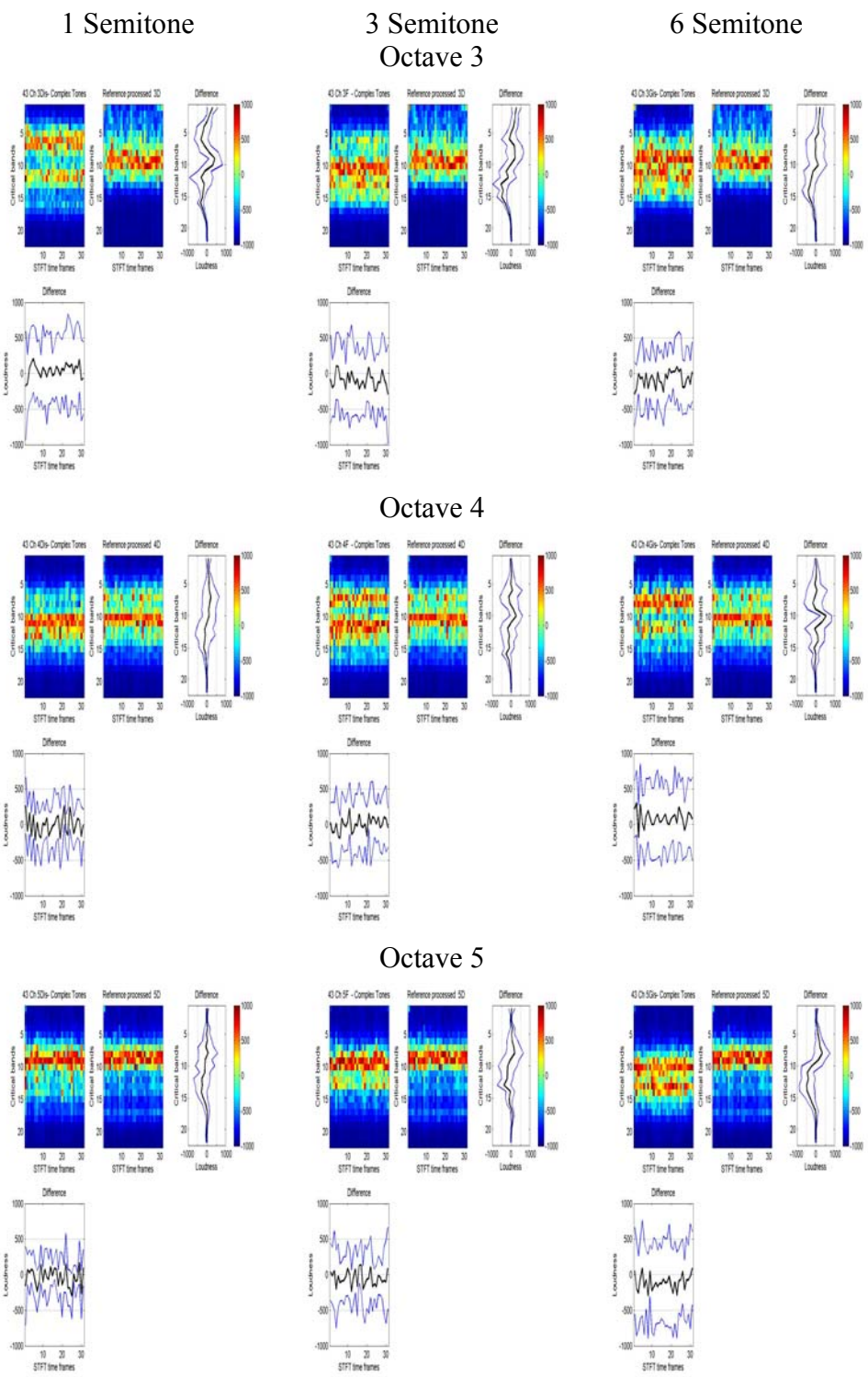


Figure (E.8): Comparison between complex and reference tones processed and resynthesized with the AMO using 22 channels and noise band vocoder for 1, 3 and 6 semitones differences.

E.5 Summary - Processed against Unprocessed (Pure and Complex) Tones

Comparison between 22 against 43 channels using pure tones and complex tones with respect to the original unprocessed signal.

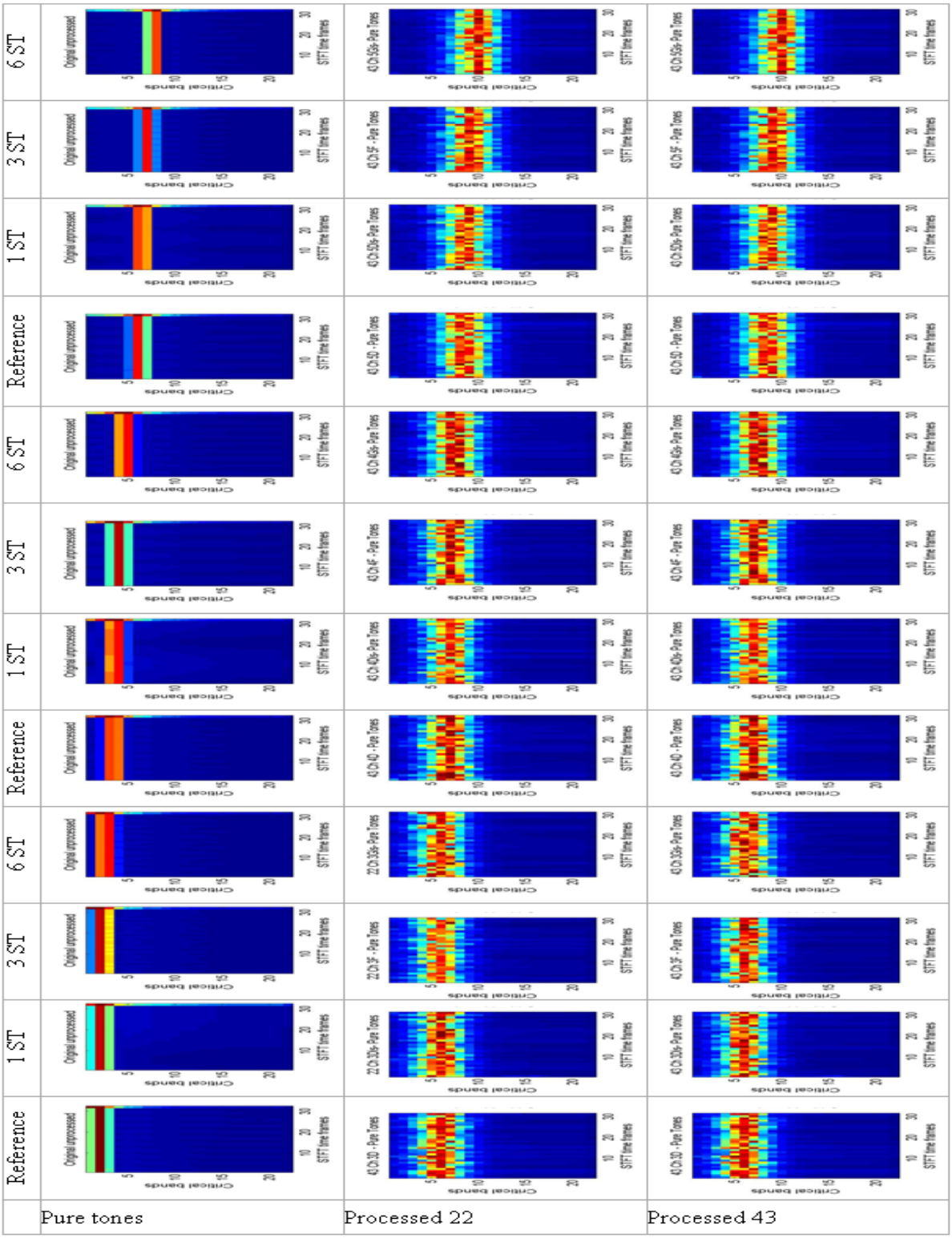


Figure (E.9): 22 Channels against 43 channels using pure tones and Complex tones with respect to the original unprocessed signal

E.6 Processed against Reference (Pure and Complex) Tones

Comparison between the 22 against 43 channels using pure tones and complex tones with respect to the reference processed signal.

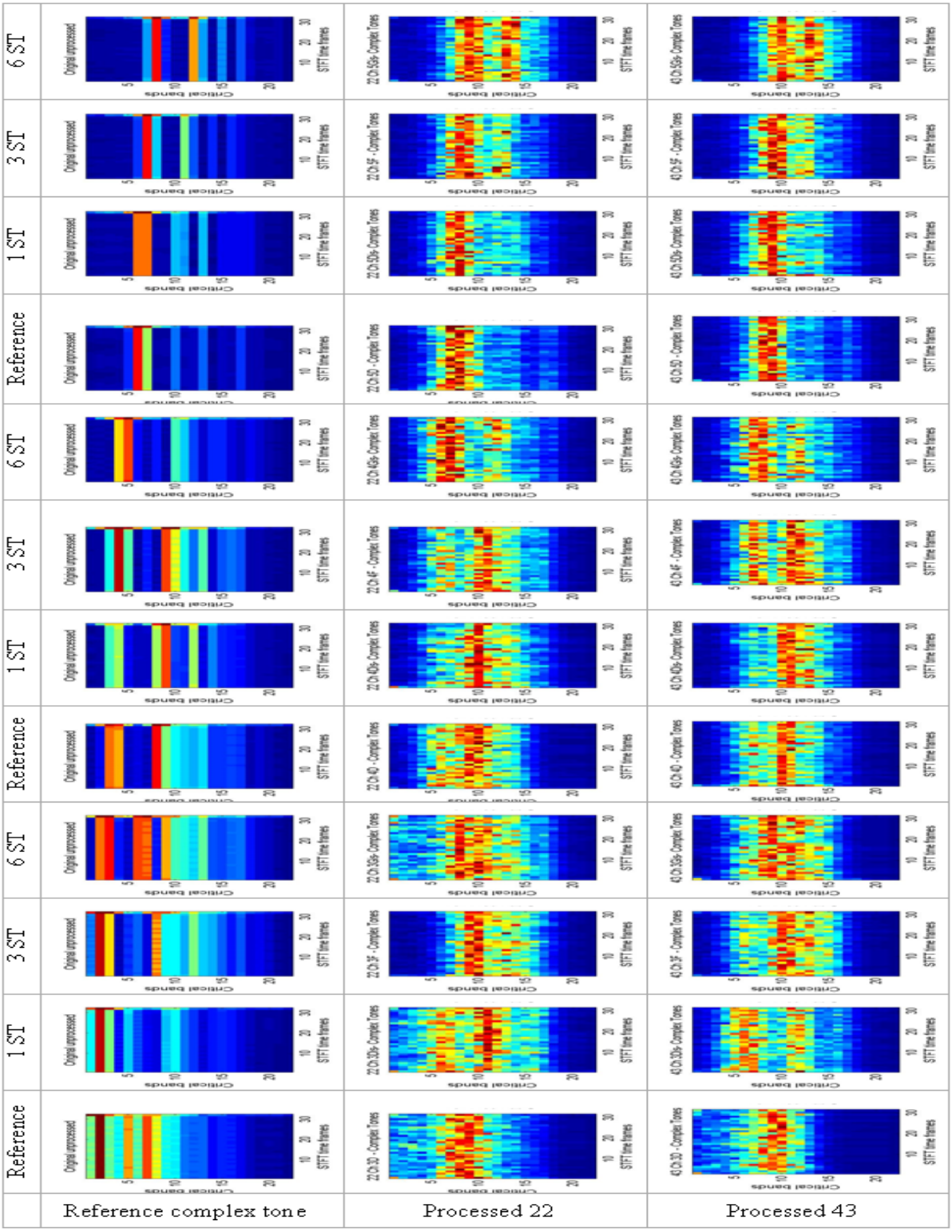


Figure (E.10): 22 Channels against 43 channels using pure tones and Complex tones with respect to the reference processed signal

Appendix - F (Conferences and Posters)

G.1 Conferences and Symposia

G.1.1 German Society of Audiology (DGA 2009)

Melody contour identification and instrument recognition using semitone mapping in Nucleus Cochlea Implant recipients

Sherif A. Omran[1,2], Waikong Lai[1], Michael Büchler[1], Norbert Dillier[1]

[1] University Hospital Zurich, ENT Dept, CH-8091 Zurich, Switzerland
[2] Institute of Neuroinformatics, CH-8057 Zurich, Switzerland

Abstract

Background
Cochlear Implants (CI) were originally developed for speech perception. The current frequency mapping compresses low frequency ranges and distorts the harmonic structure of musical sounds. Two new semitone (Smt) mapping strategies (Smt-MF and Smt-LF) for CIs were tested in a pilot study with normal hearing (NH) subjects [1]. The first map covers low and mid frequencies (Smt-LF) that are common in most musical instruments, and the second covers mid and high frequencies (Smt-MF) that are common between music and speech.

Hypothesis
Given its improved frequency representation, Smt mapping is expected to ameliorate melody contour identification (MCI) [2] compared with the standard (Std) mapping since Smt-MF and Smt-LF mappings cover the ranges from [500 Hz - 6kHz] and [125 Hz - 1.5kHz] respectively while Std mapping covers the range from [125 Hz – 8 kHz]. Better preservation of the harmonic structure in Smt-MF could improve musical instrument recognition (IR).

Methods
Two psychoacoustic music tests were conducted; MCI (which is dependent on harmonic frequency representation) and IR (which is based on timbre and the temporal structure). Acoustic and direct electric stimulation were matched [1]. Different synthetically musical tones were generated for the MCI test [2]. Each tone has 5 different harmonic overtones with a successive decrease in the amplitude by 20%. The fundamental frequency with 100% amplitude matches those of musical tones. The same music piece was played with eight different instruments and presented for the IR test. Eight patients are planned to be tested.

Discussion
The study is currently ongoing; the data will be presented and discussed during the conference.

References

1. Omran, S., et al. *Pitch perception using a Semitone mapping to improve music representation in Nucleus Cochlear Implants*. in *DGA Congress*. 2008. Kiel, Germany.
2. Galvin, J.J.I., Q.-J. Fu, and G. Nogaki, *Melodic Contour Identification by Cochlear Implant Listeners*. Ear and Hearing, 2007. **28**(3): p. 302-319.

This work was supported by research grant Nr. 32-110043/1 from the Swiss National Science Foundation.

G.1.2 German Society of Audiology (DGA 2008)

Pitch perception using a Semitone mapping to improve music representation in Nucleus Cochlear Implants.

Sherif A. Omran[1,2], Michael Büchler[1], Waikong Lai[1], Norbert Dillier[1]

[1] University Hospital Zurich, ENT Dept, CH-8091 Zurich, Switzerland
[2] Institute of Neuroinformatics, CH-8057 Zurich, Switzerland

Abstract

Introduction
Two new mapping strategies are presented that are believed to improve music representation with Cochlear Implants (CI). The CI was originally developed for speech perception and existing coding strategies are not optimal for music representation. Music consists of complex sounds that have a harmonic structure. The current frequency mappings lead to a compression of the lower frequency ranges that does not preserve the harmonic structure and produces distortions. We present two new strategies based on Semitone Mapping (Smt) for two different frequency ranges. The first mapping covers low and mid frequencies (Smt-LF) that are common in most musical instruments, and the second one covers mid and high frequencies (Smt-MF).

Methods
Different musical tones were synthetically generated using the fundamental frequency of music notes and 5 different harmonic overtones, where the amplitude of successive overtones decreased by 20%. The tones were processed with an Acoustic Simulation Model of the CI implemented in MATLAB, using a standard mapping strategy and the two experimental strategies. 8 normal Hearing (NH) subjects performed a 2AFC pitch discrimination test by listening to pairs of processed tones. [1].

Discussion
Preliminary results with NH subjects showed a consistently better or least similar discrimination of complex tone pairs using Smt-LF in comparison with either Smt-MF or standard mapping. They showed also that Smt-MF produces some pitch reversals most probably due to missing signal components of notes having lower frequencies than the mapping range. ANOVA tests and different descriptive statistics will be presented in detail at the conference.

References

1. Omran, S.A., et al. *Pitch Ranking of Complex Tones using a Model of the Virtual Channels in the Nucleus Freedom System.* in *EFAS 2007*. 2007. Heidelberg - Germany: German Society for Audiology (Deutsche Gesellschaft für Audiologie).

This work was supported by research grant Nr. 32-110043/1 from the Swiss National Science Foundation.

G.1.3 ZNZ Symposium (2008)

Comparison between Pitch Ranking of Complex Tones using an Acoustic Model of the Virtual Channels in Cochlear implants and an acoustic distance measure

Sherif A. Omran[1,2], Michael Büchler[1], Waikong Lai[1], Rodney Douglas[2], Norbert Dillier[1]

[1] University Hospital Zurich, ENT Dept, CH-8091 Zurich, Switzerland
[2] Institute of Neuroinformatics, CH-8057 Zurich, Switzerland

Abstract

Introduction

Nucleus Cochlear Implant (CI) devices provide 22 intracochlear stimulation electrodes which can be used as 22 stimulation channels. There are physical limitations preventing a substantial increase in the number of stimulation electrodes although theoretically a higher number of stimulation channels are expected to improve music notes discrimination. One way to do this in spite of the physical constraints of currently available electrode arrays would be to use Virtual Channels (VC). These are channels formed by stimulating two adjacent electrodes simultaneously. It was shown in [1] that VC stimulation can result in the perception of an intermediate frequency between adjacent electrodes and it was shown in [2] that there is no statistically significance difference in pitch between 22 and 43 channel mode; hence it was required to find a mathematic measure to show how much two signals are apart.

Methods

VCs increase the number of available stimulation channels from 22 to 43. In this study, frequency discrimination for 43 and 22 channels using an Acoustic Model (AM) was measured. The degree of adjacent music notes discrimination is measured through a pitch ranking test using complex tones. The test involves a set of music notes separated by 1, 3 or 6 semitones respectively. Adjacent pairs of notes are presented to the subject who then has to indicate which note sounded higher in pitch. In addition to the processed sounds undergo a distance measure algorithm to show how far the signals from the reference are that theoretically speaking could be correlated to the results obtained from the psychoacoustic tests.

Discussion

Testing is currently in progress and the final results will be presented at the symposium. It is expected that the distance measure might give some results that could have some similarity to the results from psychoacoustic tests. These tests with NH subjects are carried out to obtain a relation between potential improvements in complex tone discrimination using VC with different strategies and the expected mathematical results using the distance measure.

References

1. Busby, P.A. and K.L. Plant, *Dual electrode stimulation using the nucleus CI24RE cochlear implant: electrode impedance and pitch ranking studies.* Ear Hear, 2005. **26**(5): p. 504-11.
2. Omran, S.A., et al. *Pitch Ranking of Complex Tones using a Model of the Virtual Channels in the Nucleus Freedom System.* in *EFAS 2007*. 2007. Heidelberg - Germany: German Society for Audiology (Deutsche Gesellschaft für Audiologie).

This work was supported by grant Nr. 32-110043/1 from the Swiss National Science Foundation

G.1.4 German Society of Audiology (EFAS 2007)

Pitch Ranking of Complex Tones using a Model of the Virtual Channels in the Nucleus Freedom System.

Sherif A. Omran[1,2], Michael Büchler[1], Waikong Lai[1], Norbert Dillier[1]

[1] University Hospital Zurich, ENT Dept, CH-8091 Zurich, Switzerland
[2] Institute of Neuroinformatics, CH-8057 Zurich, Switzerland

Abstract

Introduction
Nucleus Cochlear Implant (CI) devices provide 22 intracochlear stimulation electrodes which can be used as 22 stimulation channels. There are physical limitations preventing a substantial increase in the number of stimulation electrodes although theoretically a higher number of stimulation channels are expected to improve music notes discrimination. One way to do this in spite of the physical constraints of currently available electrode arrays would be to use Virtual Channels (VC). These are channels formed by stimulating two adjacent electrodes simultaneously. It was shown in [1] that VC stimulation can result in the perception of an intermediate frequency between adjacent electrodes.

Methods
VCs increase the number of available stimulation channels from 22 to 43. In this study, frequency discrimination for 43 and 22 channels using an Acoustic Model (AM) was measured. The degree of adjacent music notes discrimination is measured through a pitch ranking test using complex tones. The test involves a set of music notes separated by 6, 4 or 2 semitones respectively. Adjacent pairs of notes are presented to the subject who then has to indicate which note sounded higher in pitch. This test compares the subjects' ability to discriminate notes on a musical scale using 43 and 22 channels. The sounds were presented to Normal Hearing (NH) subjects after pre-processing them with an AM.

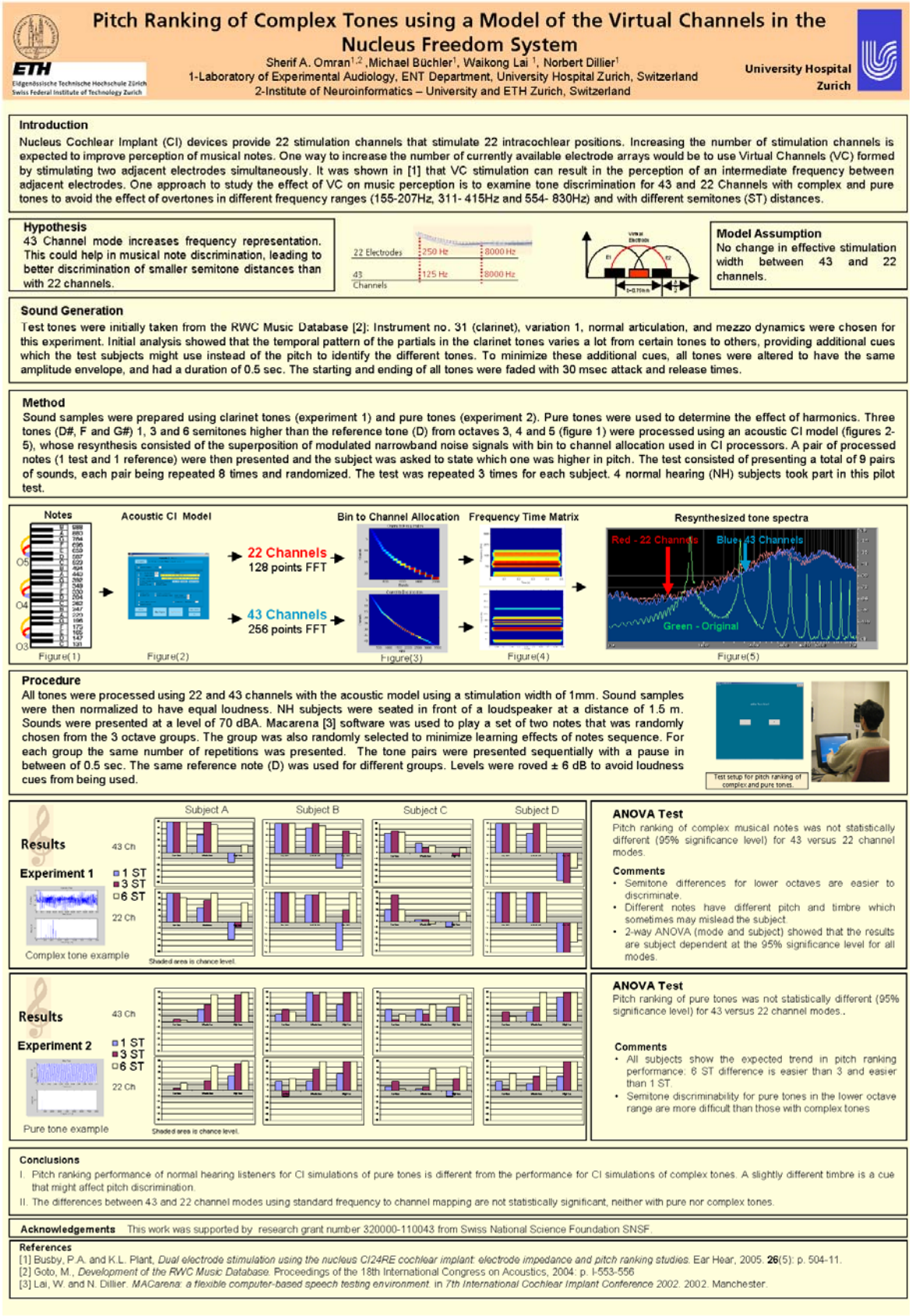
Discussion
Testing is currently in progress and the final results will be presented at the conference. It is expected that if the ability to discriminate music notes increases using 43 channels, this would be due to an increase in frequency representation that is beneficial for music appreciation. These tests with NH subjects are carried out to obtain an estimate of potential improvements in complex tone discrimination using VC instead of standard stimulation channels. The same tests will be applied to CI users subsequently.

References
[1] Busby, P. A., and Plant, K. L. (2005) "Dual electrode stimulation using the nucleus CI24RE cochlear implant: electrode impedance and pitch ranking studies," Ear and Hearing 26(5), 504-511

This work was supported by research grant Nr. 32-110043/1 from the Swiss National Science Foundation.

G.2 Posters

G.2.1 EFAS 2007



Pitch Ranking of Complex Tones using a Model of the Virtual Channels in the Nucleus Freedom System

Sherif A. Omran[1,2], Michael Büchler[1], Waikong Lai [1], Norbert Dillier[1]
1-Laboratory of Experimental Audiology, ENT Department, University Hospital Zurich, Switzerland
2-Institute of Neuroinformatics – University and ETH Zurich, Switzerland
ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich
University Hospital Zurich

Introduction
Nucleus Cochlear Implant (CI) devices provide 22 stimulation channels that stimulate 22 intracochlear positions. Increasing the number of stimulation channels is expected to improve perception of musical notes. One way to increase the number of currently available electrode arrays would be to use Virtual Channels (VC) formed by stimulating two adjacent electrodes simultaneously. It was shown in [1] that VC stimulation can result in the perception of an intermediate frequency between adjacent electrodes. One approach to study the effect of VC on music perception is to examine tone discrimination for 43 and 22 Channels with complex and pure tones to avoid the effect of overtones in different frequency ranges (155-207Hz, 311- 415Hz and 554- 830Hz) and with different semitones (ST) distances.

Hypothesis
43 Channel mode increases frequency representation. This could help in musical note discrimination, leading to better discrimination of smaller semitone distances than with 22 channels.

Model Assumption
No change in effective stimulation width between 43 and 22 channels.

Sound Generation
Test tones were initially taken from the RWC Music Database [2]: Instrument no. 31 (clarinet), variation 1, normal articulation, and mezzo dynamics were chosen for this experiment. Initial analysis showed that the temporal pattern of the partials in the clarinet tones varies a lot from certain tones to others, providing additional cues which the test subjects might use instead of the pitch to identify the different tones. To minimize these additional cues, all tones were altered to have the same amplitude envelope, and had a duration of 0.5 sec. The starting and ending of all tones were faded with 30 msec attack and release times.

Method
Sound samples were prepared using clarinet tones (experiment 1) and pure tones (experiment 2). Pure tones were used to determine the effect of harmonics. Three tones (D#, F and G#) 1, 3 and 6 semitones higher than the reference tone (D) from octaves 3, 4 and 5 (figure 1) were processed using an acoustic CI model (figures 2-5), whose resynthesis consisted of the superposition of modulated narrowband noise signals with bin to channel allocation used in CI processors. A pair of processed notes (1 test and 1 reference) were then presented and the subject was asked to state which one was higher in pitch. The test consisted of presenting a total of 9 pairs of sounds, each pair being repeated 8 times and randomized. The test was repeated 3 times for each subject. 4 normal hearing (NH) subjects took part in this pilot test.

Procedure
All tones were processed using 22 and 43 channels with the acoustic model using a stimulation width of 1mm. Sound samples were then normalized to have equal loudness. NH subjects were seated in front of a loudspeaker at a distance of 1.5 m. Sounds were presented at a level of 70 dBA. Macarena [3] software was used to play a set of two notes that was randomly chosen from the 3 octave groups. The group was also randomly selected to minimize learning effects of notes sequence. For each group the same number of repetitions was presented. The tone pairs were presented sequentially with a pause in between of 0.5 sec. The same reference note (D) was used for different groups. Levels were roved ± 6 dB to avoid loudness cues from being used.

ANOVA Test
Pitch ranking of complex musical notes was not statistically different (95% significance level) for 43 versus 22 channel modes.

Comments
- Semitone differences for lower octaves are easier to discriminate.
- Different notes have different pitch and timbre which sometimes may mislead the subject.
- 2-way ANOVA (mode and subject) showed that the results are subject dependent at the 95% significance level for all modes.

ANOVA Test
Pitch ranking of pure tones was not statistically different (95% significance level) for 43 versus 22 channel modes.

Comments
- All subjects show the expected trend in pitch ranking performance: 6 ST difference is easier than 3 and easier than 1 ST.
- Semitone discriminability for pure tones in the lower octave range are more difficult than those with complex tones

Conclusions
I. Pitch ranking performance of normal hearing listeners for CI simulations of pure tones is different from the performance for CI simulations of complex tones. A slightly different timbre is a cue that might affect pitch discrimination.
II. The differences between 43 and 22 channel modes using standard frequency to channel mapping are not statistically significant, neither with pure nor complex tones.

Acknowledgements This work was supported by research grant number 320000-110043 from Swiss National Science Foundation SNSF.

References
[1] Busby, P.A. and K.L. Plant, *Dual electrode stimulation using the nucleus CI24RE cochlear implant: electrode impedance and pitch ranking studies.* Ear Hear, 2005. **26**(5): p. 504-11.
[2] Goto, M., *Development of the RWC Music Database.* Proceedings of the 18th International Congress on Acoustics, 2004: p. I-553-556
[3] Lai, W. and N. Dillier. *MACarena: a flexible computer-based speech testing environment.* in *7th International Cochlear Implant Conference 2002.* 2002. Manchester.

G.2.2 ZNZ Symposium (Zurich University - 2008)

Comparison between Pitch Ranking of Complex Tones using an Acoustic Model of the Virtual Channels in Cochlear implants and an acoustic distance measure

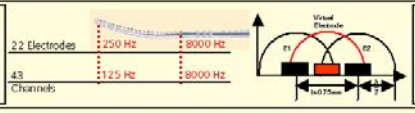
Sherif A. Omran[1,2], Michael Buechler[1], Waikong Lai[1], Rodney Douglas[2], Norbert Dillier[1]
1-Laboratory of Experimental Audiology, ENT Department, University Hospital Zurich, Switzerland
2-Institute of Neuroinformatics - University and ETH Zurich, Switzerland

Introduction

Individuals with severe hearing loss are treated with a Cochlear Implant (CI), bypassing the outer and middle ear to present auditory signals directly to the inner ear. Nucleus CI devices provide 22 intracochlear stimulation electrodes (channels). There are physical limitations preventing a substantial increase in the number of stimulation electrodes although theoretically a higher number of stimulation channels are expected to improve music notes discrimination. One way to achieve this without increasing the number of physical electrodes is to use Virtual Channels (VC). These are channels formed by stimulating two adjacent electrodes simultaneously and hence increases the number of available channels from 22 to 43. Busby & Plant [1] showed that VC stimulation can result in the perception of an intermediate frequency between adjacent electrodes, but Omran et al [2] found no statistically significant pitch differences between 22 and 43 channel mode. It is necessary to understand the factors influencing the psychoacoustic results. A distance measure (DM) model is presented here which might explain psychoacoustic results and in turn could help to optimize parameter settings.

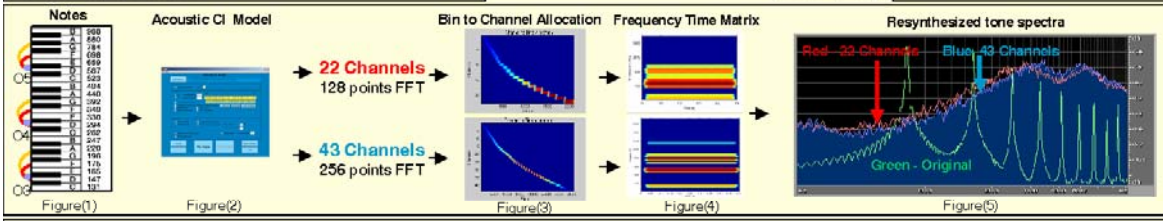
Hypothesis

The perceptual difference between two sounds could be described by the loudness difference between similar Critical Band (CB) numbers of the two sounds. This is taken into account by calculating the loudness in Sones for each CB in the DM.



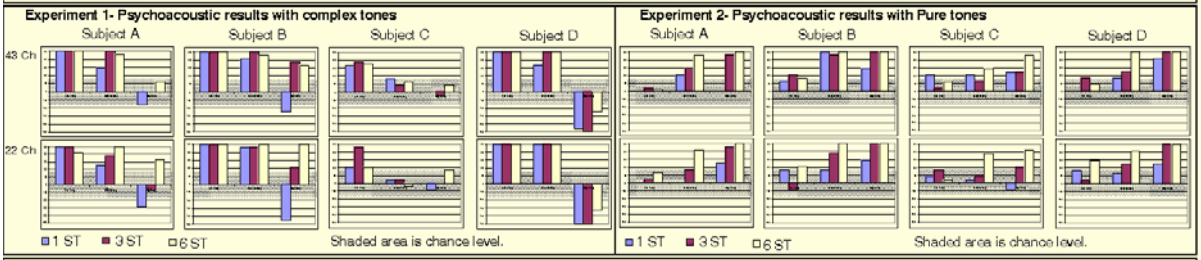
Model Limitation

This model summates the absolute differences between the give pair of signals; resulting in an absolute value. Pitch reversals are not accounted for in this model.



Psychoacoustic tests procedure

Recorded clarinet tones (Complex Tones) differing by 1, 3 and 6 Semitones (namely D#, F, G#) from octaves 3, 4 and 5 were processed using the acoustic model for 22 and 43 channels, assuming a stimulation width (excitation area) of 1mm. The resulting sound samples were normalized to produce equal loudness perception and presented to Normal-hearing subjects seated 1.5 m in front of a loudspeaker at a sound level of 70 dBA. MACarena [3] software was used to play a pair of two notes; in each presentation, a reference tone (always D) and a probe tone (either D#, F or G#) from the same octave were played back in a random order, separated by 0.5 seconds. The subject has to determine which tone sounded higher in pitch. To avoid loudness cues from affecting the task, the playback level was roved by ± 6 dB. Each pair of tones were presented a total of 24 times. The above was also repeated using pure tones at the fundamental frequency of the corresponding complex tones.



Acoustic Distance Measure Model

The model depends on calculating the differences in loudness between frequency components of two audible sounds. The loudness is calculated in Sones for different frequencies' intensity with respect to their CB location. Short Time Fast Fourier Transform (STFFT) first processes an audio signal with a narrow window. The frequency domain is divided into CBs [4]. All the intensities in the same CB are added and converted to dB SPL and then converted to their equivalence in Phones using the equal loudness curve. Phones-Sones curve is used to find the corresponding value in Sones. The means and variances at each CB and each time window are subtracted from each other to form the DM values and then all the values are summed to give an absolute value.

Calculation Steps

Calculate STFFT of an input sound of length L with number of points N, dividing the Spectrum into k bins with Bin Size B.

$$N = 2^{\frac{\log(L)}{\log(2)}} \qquad B = \frac{fs}{N}$$

The bark scale relates the frequencies to their corresponding CB.

$$Bark(f) = 7\ln\left(\frac{f}{650} + \sqrt{1+\left(\frac{f}{650}\right)^2}\right)$$

Intensities in the same CB are added together and converted to dB SPL.

dB SPL of each CB are converted to phones using equal loudness contour as shown in figure (6) and then converted to loudness in Sones using phones to Sones relation in figure (7)

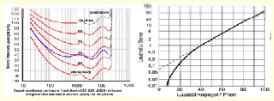
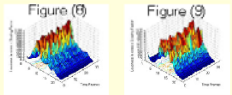


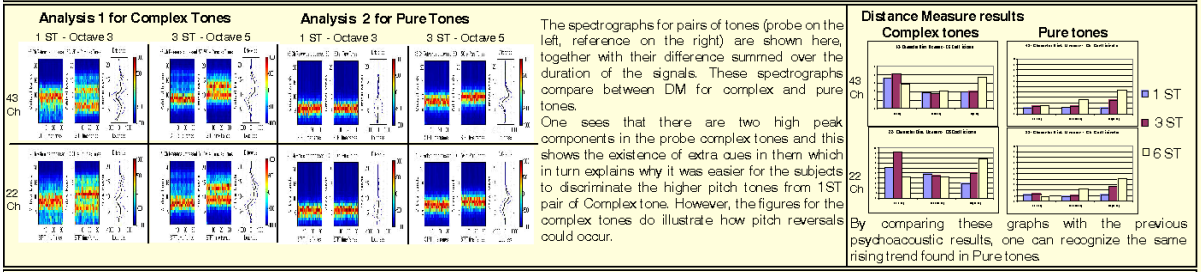
Figure (6) Figure (7)

Critical Band activity

Figures (8,9) show the processed probe and the processed reference complex tones respectively.



The means at each CB and each time window are subtracted from each other to form the DM.



The spectrographs for pairs of tones (probe on the left, reference on the right) are shown here, together with their difference summed over the duration of the signals. These spectrographs compare between DM for complex and pure tones.

One sees that there are two high peak components in the probe complex tones and this shows the existence of extra cues in them which in turn explains why it was easier for the subjects to discriminate the higher pitch tones from 1ST pair of Complex tone. However, the figures for the complex tones do illustrate how pitch reversals could occur.

By comparing these graphs with the previous psychoacoustic results, one can recognize the same rising trend found in Pure tones.

Conclusion

1. Semitone discrimination for lower octaves is easier than higher octaves. In the complex tone there are pitch reversals and this is due to other cues misleading the subject. A 2-way ANOVA (mode and subject) showed that the results are subject dependent with 95% significance level. DM and subjects showed the same rising pitch ranking trend with pure tones.
2. DM could be used to explain some psychoacoustic results.

Acknowledgements This work was supported by research grant number 320000-110043 from Swiss National Science Foundation SNSF.

References

[1] Busby, P.A. and K.L. Plant, *Dual electrode stimulation using the nucleus CI24RE cochlear implant: electrode impedance and pitch ranking studies.* Ear Hear, 2005. **26**(5): p. 504-11.
[2] Omran, S.A., et al. *Pitch Ranking of Complex Tones using a Model of the Virtual Channels in the Nucleus Freedom System.* EFAS 2007. Heidelberg - Germany: German Society for Audiology.
[3] Lai, W. and N. Dillier. *MACarena: a flexible computer-based speech testing environment.* 7th International Cochlear Implant Conference 2002. Manchester.
[4] Zwicker., E., *Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen).* The Journal of the Acoustical Society of America, 1961.

G.2.3 Retreat (ZNZ Zurich – 2008)

Mechanism of Music Perception Through Cochlea Implant

Sherif A. Omran[1,2], Michael Buechler[1], Waikong Lai[1], Rodney Douglas[2], Norbert Dillier[1]
1-Laboratory of Experimental Audiology, ENT Department, University Hospital Zurich, Switzerland
2-Institute of Neuroinformatics - University and ETH Zurich, Switzerland

ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich

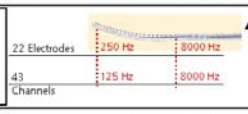
University Hospital Zurich

Introduction

Individuals with severe hearing loss are treated with a Cochlear Implant (CI), bypassing the outer and middle ear to present auditory signals directly to the inner ear. Nucleus CI devices provide 22 intracochlear stimulation electrodes (channels). There are physical limitations preventing a substantial increase in the number of stimulation electrodes although theoretically a higher number of stimulation channels are expected to improve music notes discrimination. One way to achieve this without increasing the number of physical electrodes is to use Virtual Channels (VC). These are channels formed by stimulating two adjacent electrodes simultaneously and hence increases the number of available channels from 22 to 43. Busby & Plant [1] showed that VC stimulation can result in the perception of an intermediate frequency between adjacent electrodes, but Omran et al [2] found no statistically significant pitch differences between 22 and 43 channel mode with standard map. It is necessary to understand the factors influencing the psychoacoustic results. A distance measure (DM) model is presented here which might explain psychoacoustic results and in turn could help to optimize parameter settings and a semitone map that is believed to improve the harmonic structure for a better music representation.

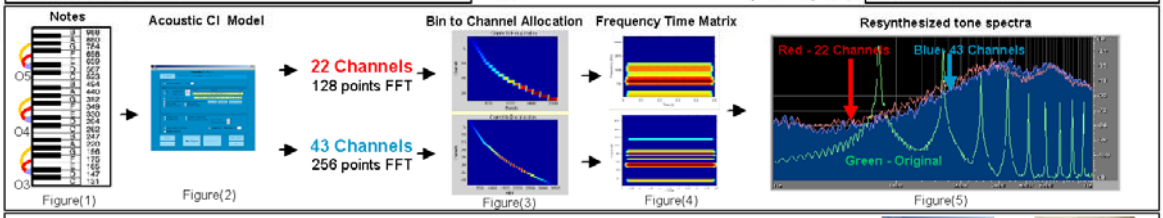
Hypothesis

The perceptual difference between two sounds could be described by the loudness difference between similar Critical Band (CB) numbers of the two sounds. This is taken into account by calculating the loudness in Sones for each CB in the DM.



Model Limitation

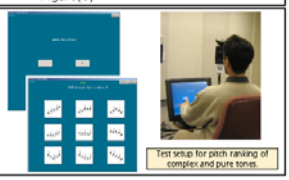
This model summates the absolute differences between the give pair of signals, resulting in an absolute value. Pitch reversals are not accounted for in this model.



Figure(1) Figure(2) Figure(3) Figure(4) Figure(5)

Psychoacoustic tests

Recorded clarinet tones (Complex Tones) differing by 1, 3 and 6 Semitones (namely D#, F, G#) from octaves 3, 4 and 5 were processed using the acoustic model for 22 and 43 channels, assuming a stimulation width (excitation area) of 1mm. The resulting sound samples were normalized to produce equal loudness perception and presented to Normal-hearing subjects seated 1.5 m in front of a loudspeaker at a sound level of 70 dBA. MACarena [3] software was used to play a pair of two notes; in each presentation, a reference tone (always D) and a probe tone (either D#, F or G#) from the same octave were played back in a random order, separated by 0.5 seconds. The subject has to determine which tone sounded higher in pitch. To avoid loudness cues from affecting the task, the playback level was roved by ± 6 dB. Each pair of tones were presented a total of 24 times. The above was also repeated using pure tones at the fundamental frequency of the corresponding complex tones.



Test setup for pitch ranking of complex and pure tones.

Acoustic Distance Measure Model

The model depends on calculating the differences in loudness between frequency components of two audible sounds. The loudness is calculated in Sones for different frequencies' intensity with respect to their CB location. Short Time Fast Fourier Transform (STFFT) first processes an audio signal with a narrow window. The frequency domain is divided into CBs [4]. All the intensities in the same CB are added and converted to dB SPL and then converted to their equivalence in Phones using the equal loudness curve. Phones-Sones curve is used to find the corresponding value in Sones. The means and variances at each CB and each time window are subtracted from each other to form the DM values and then all the values are summed to give an absolute value.

Calculation Steps

Calculate STFFT of an input sound of length L with number of points N, dividing the Spectrum into k bins with Bin Size B.

$$N = 2^{\frac{\log(L)}{\log(2)}} \qquad B = \frac{fs}{N}$$

The bark scale relates the frequencies to their corresponding CB.

$$Bark(f) = 7\ln\left(\frac{f}{650} + \sqrt{1+\left(\frac{f}{650}\right)^2}\right)$$

Intensities in the same CB are added together and converted to dB SPL.

dB SPL of each CB are converted to phones using equal loudness contour as shown in figure (6) and then converted to loudness in Sones using phones to Sones relation in figure (7)

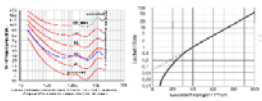


Figure (6) Figure (7)

Critical Band activity

Figures (8,9) show the processed probe and the processed reference complex tones respectively.

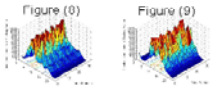


Figure (8) Figure (9)

The means at each CB and each time window are subtracted from each other to form the DM.

Semitone Map and Harmonic Structure

Semitone map is a logarithmic spacing method where each electrode gets activated to the fundamental frequency of a note. Two ranges are selected Low frequency (red arrows) and Mid Frequency range (green arrows).

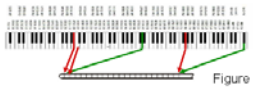


Figure (10)

Figure (11) shows a block diagram of the acoustic model. A signal is inputted and is splitter in to overlapping time frames, where frequencies are analyzed. A loudness growth function LGF is used to include loudness model effect and diminish fine frequencies. Accordingly a Frequency Time Matrix is produced. This in turn used the map function to match frequency bands to their corresponding channels that should be activated . The state-of-the-art standard map, semitone map LF and MF are compared.

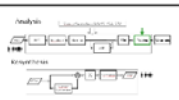


Figure (11)

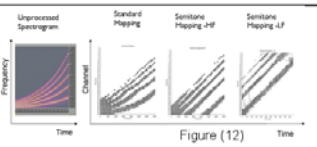


Figure (12)

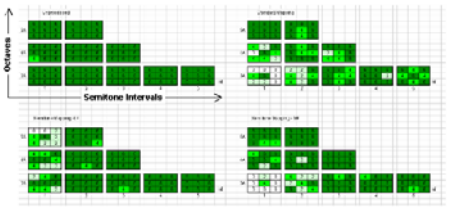
A group of tones with 4 harmonic overtones are synthesized and been processed with different maps shown in figure(12). It could be shown that the standard map causes a distortion in the harmonic structure while Semitone not. However, the MF model causes pitch reversals at low frequencies due to low frequency filtration while the LF model filtrates high frequency components. This was tested wit NH subjects.

Checker and Stimuli Validation

Testing with Patients is strongly restricted and accordingly acute checking must be done before it. Thus we created a tool by which one can select a patient from the hospital database and get his map according to last testing done with by the audiologist. One can select a file, process it with the different proposed strategies and different channel numbers to generate a sequence file. This defines which channel should be stimulated at which time. This application reads the sequence file and extract the patient parameters from it such that they are compared to the database. Once every thing is correct, the file will be reported as validated for testing other wise an error will pop defining the pulse number where the mistake it. This application was tested by manually introducing mistakes and it identified them correctly.

Melody Contour Identification (MCI)

Music has many aspects among them the melody contour. This is a test where 5 processed contours are presented with different maps and the subject has to select the correct one.



Figure(2): Results with different mappings and the unprocessed notes.

Conclusion

1. Semitone discrimination for lower octaves is easier than higher octaves. In the complex tone there are pitch reversals and this is due to other cues misleading the subject. A 2-way ANOVA (mode and subject) showed that the results are subject dependent with 95% significance level. DM and subjects showed the same rising pitch ranking trend with pure tones.
2. DM could be used to explain some psychoacoustic results.

Acknowledgements This work was supported by research grant number 320000-110043 from Swiss National Science Foundation SNSF.

References

[1] Busby, P.A. and K.L. Plant, *Dual electrode stimulation using the nucleus CI24RE cochlear implant: electrode impedance and pitch ranking studies.* Ear Hear, 2005. **26**(5): p. 504-11.
[2] Omran, S.A., et al. *Pitch Ranking of Complex Tones using a Model of the Virtual Channels in the Nucleus Freedom System.* EFAS 2007. Heidelberg - Germany: German Society for Audiology.
[3] Lai, W. and N. Dillier. *MACarena: a flexible computer-based speech testing environment.* 7th International Cochlear Implant Conference 2002. Manchester.
[4] Zwicker., E., *Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen).* The Journal of the Acoustical Society of America, 1961.

List of Figures

List of Tables

References

Arapova AA, Gersuni GV, Volokhov AA (1936) A Further Analysis of The Action of Alternating Currents on The Auditory Apparatur. J. Physiol. U.S.S.R. 89: 122-131

Arnoldner C, Riss D, Brunner M, Durisin M, Baumgartner WD, Hamzavi JS (2007) Speech and music perception with the new fine structure speech coding strategy: preliminary results. Acta Otolaryngol 127: 1298-1303

Barkhausen H (1954) Lehrbuch der Elektronenröhren, Elektronenröhren und ihre technischen Anwendungen. Hirzel-Verlag Leipzig

Baskent D, Shannon RV (2003) Speech recognition under conditions of frequency-place compression and expansion. J Acoust Soc Am 113: 2064-2076

Baskent D, Shannon RV (2004) Frequency-place compression and expansion in cochlear implant listeners. J Acoust Soc Am 116: 3130-3140

Baumann U, Nobbe A (2004) Pitch ranking with deeply inserted electrode arrays. Ear Hear 25: 275-283

Bekesy GV (1960) Travelling wave profiles in Experiments in Hearing. McGraw-Hill, New York

Besouw Rv (2008) The Feasibility of Encoding Harmonic Pitch Using the Nucleus Matlab Toolbox. In: Cochlear Implant Music Workshop 2008, University hospital zurich

Bhattacharya A, Zeng FG (2007) Companding to improve cochlear-implant speech recognition in speech-shaped noise. J Acoust Soc Am 122: 1079-1089

Black RC, Clark GM (1980) Differential electrical excitation of the auditory nerve. J. Acoust. Soc. Am. 67: 868-874

Blamey P, Dowell R, Clark G (1987) Acoustic parameters measured by a formant- estimating speech processor for a multiple-channel cochlear implant. J Acoust Soc Am 82: 38-47

Blume F (1961) Die Musik in Geschichte und Gegenwart. Finscher, Ludwig, Kassel

Bosco E, D'Agosta L, Mancini P, Traisci G, D'Elia C, Filipo R (2005) Speech perception results in children implanted with Clarion devices: Hi-Resolution and Standard Resolution modes. Acta Otolaryngol 125: 148-158

Bosi M, Goldberg RE (2003) Introduction to Digital Audio Coding and Standards. Springer

Boulogne D (1855) De l'Électrisation localisée et de son application à la physiologie, à la pathologie et à la thérapeutique In:

Bregman AS (1990) Auditory Scene Analysis: The Perceptual Organization of sound. The MIT Press, Cambridge, Massachusetts

Brockmeier SJ, Grasmeder M, Passow S, Mawmann D, Vischer M, Jappel A, Baumgartner W, Stark T, Muller J, Brill S, Steffens T, Strutz J, Kiefer J, Baumann U, Arnold W (2007) Comparison of musical activities of cochlear implant users with different speech-coding strategies. Ear Hear 28: 49S-51S

Brown J (1991) Calculation of a constant Q spectral transform. J. Acoust. Soc. Am. 89: 425–434

Büchner A, Frohne-Buechner C, Stoever T, Gaertner L, Battmer R-D, Lenarz T (2005) Comparison of a Paired or Sequential Stimulation Paradigm with Advanced Bionics' High-Resolution Mode. Otol Neurotol 26: pp 941-947

Busby PA, Plant KL (2005) Dual electrode stimulation using the nucleus CI24RE cochlear implant: electrode impedance and pitch ranking studies. Ear Hear 26: 504-511

Carney LH, Geiseler CD (1986) A temporal analysis of auditory nerve fiber responses to spoken stop consonant vowel syllables. J. Acoust. Soc. Am. 79: 1896-1914

Carroll J, Zeng FG (2007) Fundamental frequency discrimination and speech perception in noise in cochlear implant simulations. Hear Res 231: 42-53

Chen F, Zhang Y-T (2008) A novel temporal fine structure-based speech synthesis model for cochlear implant. J. Sigpro 88 2693– 2699

Chen H, and Zeng, F. (2003) Improving pitch perception in cochlear implant users. In: 2003 Conference on Implantable Auditory Prosthesis, Asilomar, California
Clark G (1987) The University of Melbourne-Nucleus multi-electrode cochlear implant. Advances in Oto-Rhino-Laryngology 38: 1-189
Clark G (2003) Cochlear implants: fundamentals and applications. Springer-Verlag, Heidelberg
Cohen NL, Roland JT, Fishman JA (2002) Surgical Technique For The Nucleus Contour Cochlear Implant. Ear & Hearing. 23: 59S-66S
Crochiere RE, Webber SA, Flanagan. JL (1976) Digital Coding of Speech in Sub-bands. . Bell System Technical Journal 55: 1069-1085
de Boer E (1975) Synthetic whole-nerve action potentials for the cat. J. Acoust. Soc. Am. 58: 1030–1045
Delgutte B, Kiang NY (1984) Speech coding in the auditory nerve: I. Vowel-like sounds. J Acoust Soc Am 75: 866-878
Deutsch S (1977) Two-Stage lateral inhibition for auditory selectivity. Bull. Math. Biol. 39: 259-266
Dillier N, Battmer RD, Doring WH, Muller-Deile J (1995) Multicentric field evaluation of a new speech coding strategy for cochlear implants. Audiology 34: 145-159
Diniz FCCB, Kothe I, Biscainho LWP, Netto SL (2006) A Bounded-Q Fast Filter Bank for Audio Signal Analysis. In: 8th International Conference on Intelligent Tutoring Systems, Taiwan, pp 1015-1019
Diniz FCCB, Kothe I, Netto SL, Biscainho LP (2007) High-Selectivity Filter Banks for Spectral Analysis of Music Signals. EURASIP Journal on Advances in Signal Processing 2007: 12 pages
Divenyi PL, Hirsh IJ (1974) Identification of temporal order in three-tone sequences. J Acoust Soc Am 56: 144-151
Dorman M, Basham K, McCandles G, Dove H (1991) Speech understanding and music appreciation with the Ineraid cochlear implant. The Hearing Journal 44: 32-37
Dorman M, Smith L, McCandles G, Dunnavant G, Parking J, Dankowski K (1990) Pitch scaling and speech understanding by patients who use the Ineraid cochlear implant. Ear Hearing II: 310-315
Dorman M, Spahr T, Gifford R, Loiselle L, McKarns S, Holden T, Skinner M, Finley C (2007) An Electric Frequency-to-place Map for a Cochlear Implant Patient with Hearing in the Nonimplanted Ear. J. Assoc. Res. Otolaryngol. Volume 8,: 234-240
Dowell R, Dawson P, Dettman S, Shepherd R, Whitford L, Seligman P, Clark G (1991) Multichannel cochlear implantation in children: A summary of current work at the University of Melbourne. American Journal of Otology 12: 137-143
Dowell R, Mecklenburg D, Clark G (1986) Speech recognition for 40 patients receiving multichannel cochlear implants. Arch Otolaryngol 112: 1054-1059
Dowell R, Seligman P, Blamey P, Clark G (1987) Evaluation of a two-formant speech processing strategy for a multichannel cochlear prosthesis. ANN OTO RHINOL LARYN 96 132-134
Dowling WJ (1994) Melodic contour in hearing and remembering melodies. Oxford University Press, New York, NY
Doyle J, Doyle JJ, Turnbull FJ (1963) Electrical Stimulation of Eighth Cranial Nerve. Arch Otolaryngol 80: 388-391
Drennan WR, Rubinstein JT (2008) Music perception in cochlear implant users and its relationship with psychophysical capabilities. J Rehabil Res Dev 45: 779-789
Eisen MD (2003) Djourno, Eyries, and the First Implanted Electrical Neural Stimulator to Restore Hearing. Otol Neurotol 24: 500-506
Eisen MD, Franck KH (2005) Electrode Interaction in Pediatric Cochlear Implant Subjects. J Assoc Res Otolaryngol. 6: 160–170

Fearn R. (2001). Music and pitch perception of cochlea implant recipients. *PhD thesis, School of physics, The University of new south wales*

Fearn R, Wolfe J (2000) Relative importance of rate and place: experiments using pitch scaling techniques with cochlear implants recipients. Ann. Otol. Rhinol. Laryngol. Suppl. 185: 51-53

Feth LL, O'Malley H, Ramsey JJ (1982) Pitch of unresolved, two-component complex tones. J. Acoust. Soc. Am. 72: 1403-1412

Firszt JB, Koch DB, Downing M, Litvak L (2007) Current steering creates additional pitch percepts in adult cochlear implant recipients. Otol Neurotol 28: 629-636

Flanagan JL (1980) Parametric coding of speech spectra. J. Acoust. Soc. Am. 68: 412-419

Flanagan JL, Golden RM (1966) Phase Vocoder. Bell System Technical Journal 45: 1493-1509

Flottorp G (1952) Effect of Different Types of Electrodes in Electrophonic Hearing (A). J Acoust Soc Am 24: 447-448

Fourakis MS, Hawks JW, Holden LK, Skinner MW, Holden TA (2004) Effect of frequency boundary assignment on vowel recognition with the Nucleus 24 ACE speech coding strategy. J Am Acad Audiol 15: 281-299

Fourakis MS, Hawks JW, Holden LK, Skinner MW, Holden TA (2007) Effect of frequency boundary assignment on speech recognition with the Nucleus 24 ACE speech coding strategy. J Am Acad Audiol 18: 700-717

Frijns JHM, Snoo SLd, Schoonhoven R (1995) Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea Hear Res 87: pp 170-186

Fujita S, Ito J (1999a) Ability of nucleus cochlear implantees to recognize music. Annals of Otology, Rhinology, and Laryngology 108: 634-640

Fujita S, Ito J (1999b) Ability of Nucleus cochlear implantees to recognize music. ANN OTO RHINOL LARYN 108: 634-640

Galvin JJI, Fu Q-J, Nogaki G (2007) Melodic Contour Identification by Cochlear Implant Listeners. Ear Hearing 28: 302-319

Gantmacher F (1990) The Theory of Matrices. American Mathematical Society, Rhode Island

Gantz BJ, Turner C, Gfeller K (2004) Expanding cochlear implant technology: Combined electrical and acoustical speech processing. . Cochlear Implants International 5: 8-14

Gfeller K, Christ A, Knutson JF, Witt S, Murray KT, Tyler RS (2000a) Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. J Am Acad Audiol 11: 390-406

Gfeller K, Knutson JF (2003) Music to the Impaired or Implanted Ear; Psychosocial Implications for Aural Rehabilitation.

Gfeller K, Lansing CR (1991) Melodic, rhythmic, and timbral perception of adult cochlear implant users. J Speech Hear Res 34: 916-920

Gfeller K, Olszewski C, Rychener M, Sena K, Knutson JF, Witt S, Macpherson B (2005) Recognition of "real-world" musical excerpts by cochlear implant recipients and normal-hearing adults. Ear Hear 26: 237-250

Gfeller K, Turner C, Woodworth G, Mehr M, Fearn R, Knutson J, Witt S, Stordahl J (2002a) Recognition of familiar melodies by adult cochlear implant recipients and normal hearing adults. Cochlear Implants International, 3: 29-53

Gfeller K, Witt S, Adamek M, Mehr M, Rogers J, Stordahl J (2002b) Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. J Am Acad Audiol 13

Gfeller K, Witt S, Stordahl J, Mehr M, Woodworth G (2000b) The effect of training on melody recognition and appraisal by adult cochlear implant recipients. Journal of the Academy of Rehabilitative Audiology 23: 115-138

Gfeller K, Woodworth G, Robin DA, Witt S, Knutson JF (1997a) Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. Ear Hear 18: 252-260
Gfeller K, Woodworth G, Robin DA, Witt S, Knutson JF (1997b) Perception of Rhythmic and Sequential Pitch Patterns by Normally Hearing Adults and Adult Cochlear Implant Users. Ear & Hearing 18: 252-260
Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. Hearing Research 47: 103-108
Goldstein JL (1973) An optimum processor theory for the central formation of the pitch of complex tones. J Acoust Soc Am 54: 1496-1516
Goto M (2004) Development of the RWC Music Database. Proceedings of the 18th International Congress on Acoustics I-553-556
Goto M, Hashiguchi H, Nishimura T, Oka R (2001) RWC Music Database: Popular music database and royalty-free music database (in Japanese). Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002): 287-288
Grayden DB, Burkitt AN, Kenny OP, Clarey JC, Paolini AG, Clark GM (2004) A Cochlear Implant Speech Processing Strategy Based On An Auditory Model. In: International Conference on Intelligent Sensors, Sensor Networks and Information Processing. IEEE, Melbourne, Australia, pp 491-496
Greenwood D (1961) Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane. J Acoust Soc Am 33: 12
Greenwood D (1990) A cochlear frequency-position function for several species--29 years later. J Acoust Soc Am 87: 2592-2605
Greenwood D (1991) Critical bandwidth and consonance in relation to cochlear frequency-position coordinates. Hear Res 54: 164-208
Guerts L, Wouters J (1999) Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants. J Acoust Soc Am 105
Hall DE (2001) Musical Acoustics. Brooks Cole, California
Hallpike CS, Rawdon-Smith AF (1934) The origin of the Wever and Bray phenomenon. jphysiol 83: 243-254
Handel S (1995) Timbre perception and auditory object formation. In: BC M (ed) Hearing. Academic Press, San Diego (CA), pp 425-461
Hanekom T (2001) Three-Dimensional Spiraling Finite Element Model of the Electrically Stimulated Cochlea. Ear & Hearing. 22: pp 330-315
Hans B (1980) Einführung in die Akustik. Wissenschafts Verlag, Mannhein, Wien, Zürich
Helmholtz H (1863) Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. . Vieweg, Braunschweig
Helmholtz HLF (1954) On the Sensations of Tone as a Physiological Basis for the Theory of Music. Dover, New York
Hofman-Jablan J (1995) Antisymmetry and colored symmetry of musical work. Symmetry: Culture and Science 6: 249-251
Hong RS, Rubinstein JT (2003) High-rate conditioning pulse trains in cochlear implants: dynamic range measures with sinusoidal stimuli. J Acoust Soc Am 114: 3327-3342
Houle G (1987) Meter in Music, 1600-1800: Performance, Perception, and Notation. Indiana University Press
Idson WL, Massaro DW (1978) A bidimensional model of pitch in the recognition of melodies. Percept Psychophys 24: 551-565
ISO-226 (:2003 revision) Acoustics - Normal equal-loudness-level contours. In:
ISO-13818-3 (1994) Information Technology - Generic Coding of Moving Pictures and Associated Audio: Part 3 - Audio. In: 13818-3, vol NO803
James E, Barros AK, Yoshinori T (2004) Speech Enhancement by lateral inhibition and binaural masking. In: Machine learning for signal processing. IEEE

Ji-Jon S, M. SA, J. OA, A. FM, Rahul S (2007) A low-power asynchronous interleaved sampling algorithm for cochlear implants that encodes envelope and phase information. IEEE transactions on biomedical engineering 54: 138-149

Johannesma PIM (1972) The pre-response stimulus ensemble of neurons in the cochlear nucleus,". In: Symposium on Hearing Theory IPO, Eindhoven, The Netherlands, pp 58–69

Jones RC, Stevesn SS, Lurie MH (1940) Three Mechanisms of Hearing by Electrical Stimulation. J Acoust Soc Am 12: 281-290

Kasturi K, Loizou P (2007) Effect of filter spacing on melody recognition: acoustic and electric hearing. J Acoust Soc Am 122: EL29-34

Katz J, Burkard RF, Medwetsky L (2002) Handbook of Clinical Audiology. Lippincott Williams & Wilkins

Kester W (2003) Mixed signal and DSP design techniques. Analog Devices and Newnes

Kiefer J, Hohl S, Sturzebecher E, Pfennigdorff T, Gstoettner W (2001) Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system. Audiology 40: 32-42

Kiefer J, Pok M, Adunka O, Stürzebecher E, Baumgartner W, Schmidt M, Tillein J, Ye Q, Gstoettner W (2005) Combined electric and acoustic stimulation (EAS) of the auditory system - Results of a clinical study. Audiology & Neuro-Otology 10: 134-144

Kong YY, Cruz R, Jones JA, Zeng F-G (2004a) Music perception with temporal cues in acoustic and electric hearing. Ear and Hearing 25: 173-185

Kong YY, Cruz R, Jones JA, Zeng FG (2004b) Music perception with temporal cues in acoustic and electric hearing. Ear Hear 25: 173-185

Lai W. (1990). Psychophysical studies investigating a place/rate coding strategy for a multiple-channel cochlear implant. *PhD thesis, The University Melbourne*

Lai W, Dillier N (2002) MACarena: a flexible computer-based speech testing environment. In: 7th International Cochlear Implant Conference 2002, Manchester

Lai W, Dillier N (2008) Investigating the MP3000 coding strategy for music perception. In: 11. Jahrestagung der Deutschen Gesellschaft fuer Audiologie, Kiel - Germany

Landsberger DM, McKay CM (2004) Perceptual differences between low and high rates of stimulation on single electrodes for cochlear implantees. J Acoust Soc Am 117: 319-327

Laneau J, Moonen M, Wouters J (2006a) Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. J Acoust Soc Am 119: 491-506

Laneau J, Wouters J (2004) Multi-channel place pitch sensitivity in cochlear implant recipients. J. Assoc. Res. Otolaryngol. 5: 285-294

Laneau J, Wouters J, Moonen M (2006b) Improved Music Perception with Explicit Pitch Coding in Cochlear Implants. Audiol Neurotol 11: 38-52

Langner G (1992) Periodicity coding in the auditory system. Hear Res 60: 115-142

Licklider JC (1951) A duplex theory of pitch perception. Experientia 7: 128-134

Loeb GE, White MW, Merzenich MM (1983) Spatial cross-correlation. A proposed mechanism for acoustic pitch perception. Biol Cybern 47: 149-163

Loizou P (1998) Introduction to Cochlear Implants. IEEE signal processing magazine: 101-130

Loizou P, Stickney G, Mishra L, Assmann P (2003a) Comparison of speech processing strategies used in the Clarion implant processor. Ear and Hearing 24: 12-19

Loizou PC, Dorman M, Tu Z (1999) On the number of channels needed to understand speech. J Acoust Soc Am 106: 2097-2103

Loizou PC, Stickney G, Mishra L, Assmann P (2003b) Comparison of speech processing strategies used in the Clarion implant processor. Ear Hear 24: 12-19

Looi V, McDermott H, McKay C, Hickson L (2007) Comparisons of quality ratings for music by cochlear implant and hearing aid users. Ear Hear 28: 59S-61S
Looi V, McDermott H, McKay C, Hickson L (2008) Music perception of cochlear implant users compared with that of hearing aid users. Ear Hear 29: 421-434
Lynch MP, Eilers RE, Oller KD, Urbano RC, Wilson P (1991) Influences of acculturation and musical sophistication on perception of musical interval patterns. . J Exp Psychol Hum Percept Perform 17: 467-475
Maragos P, Kaiser JF, Quatieri TF (1993) On amplitude and frequency demodulations using energy operators. IEEE Trans. Signal Processing 41: 1532
McDermott HJ, al. E (1993) A Portable Programmable Digital Sound Processor for Cochlear Implant Research. IEEE TRANSACTIONS ON REHABILITATION ENGINEERING 1
McDermott HJ, McKay CM (1994) Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea. J Acoust Soc Am 96: 155-162
McDermott HJ, McKay CM, Richardson LM, Henshall KR (2003a) Application of loudness models to sound processing for cochlear implants. J Acoust Soc Am 114: 2190-2197
McDermott HJ, McKay CM, Vandali AE (1992) A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. J Acoust Soc Am 91: 3367-3371
McDermott HJ, r CMSuche, McKay CM (2003b) Speech Intelligibility with a Cochlear Implant Sound-Processing Scheme Based on Loudness Models. In: 25th Annual Intemational Conference of the IEEE EMBS, Cancun, Mexico
McDermott HJ, Sucher CM, McKay CM (2005) Speech perception with a cochlear implant sound processor incorporating loudness models. Acoustics Research Letters Online 6: 7-13
McKay CM, Henshall KR, Farrell RJ, McDermott HJ (2003) A practical method of predicting the loudness of complex electrical stimuli. J Acoust Soc Am 113: 2054-2063
McKay CM, McDermott HJ (1998) Loudness perception with pulsatile electrical stimulation: the effect of interpulse intervals. J Acoust Soc Am 104: 1061-1074
McKay CM, McDermott HJ, Vandali AE, Clark GM (1992) A comparison of speech perception of cochlear implantees using the Spectral Maxima Sound Processor (SMSP) and the MSP (MULTIPEAK) processor. Acta Otolaryngol 112: 752-761
McKay CM, Remine MD, McDermott HJ (2001) Loudness summation for pulsatile electrical stimulation of the cochlea: effects of rate, electrode separation, level, and mode of stimulation. J Acoust Soc Am 110: 1514-1524
Milczynski M, Wieringen Av, Wouters J (2008) F0 modulation strategy. In: Cochlea Implant Music Workshop 2008. experimental orl - Katholieke university of Leuven - belgium, Leuven
Miller JM, Spelman FA (1989) Cochlear Implants - Models of the stimulated ear. Springer Verlag New York
Mishra LN. (2000). Analysis of speech processing strategies for the Clarion implant processor. *M.Sc thesis, The University of Texas*
Möller S (2005) Quality of Telephone-Based Spoken Dialogue Systems. Springer US
Moore B (2003) An introduction to the psychology of hearing. Academic Press
Mueller-Deile J (2009) Sprachverständlichkeitsuntersuchungen bei Kochleaimplantatpatienten HNO 57: 11
Neubauer A (2003) Irreguläre Abtastung: Signaltheorie und Signalverarbeitung (Irregular sampling: signal theory and signal processing). Springer, Berlin
Neukom M (2005) Signale, Systeme und Klangsynthesse: Grundlagen der Computermusik Peter Lang, Bern
Nie K, Stickney G, Zeng FG (2005) Encoding frequency modulation to improve cochlear implant performance in noise. IEEE Trans Biomed Eng 52: 64-73

Nie K, Zeng FG (2004) A perception-based processing strategy for cochlear implants and speech coding. Conf Proc IEEE Eng Med Biol Soc 6: 4205-4208
Niet K, Lant N, Gao S (1998) Implementation of CIS Speech Processing Strategy for Cochlea Implants by Using Wavelet Transform. Signal Processing Proceedings, ICSP '98 2: 1395-1398
Nobbe A. (2004). Pitch Perception and Signal Processing in Electrical Hearing. *Dr thesis, HNO, The LMU*
Nobbe A, Schleich P, Zierhofer C, Nopp P (2007) Frequency discrimination with sequential or simultaneous stimulation in MED-EL cochlear implants. Acta Otolaryngol 127: 1266-1272
Nogueira W, Büchner A, Lenarz T, Edler B (2005) A Psychoacoustic N-of-M Type Speech Coding Strategy for Cochlear Implants. EURASIP Journal on Applied Signal Processing 18: 3044-3059
Nogueira W, Giese A, Edler B, Büchner A (2006) Wavelet Packet Filterbank for Speech Processing Strategies in Cochlear Implants. In: 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, France
Olszewski C, Kate Gfeller, Rebecca Froman, Julie Stordahl, Bruce Tomblin (2005) Familiar melody recognition by children and adults using cochlear implants and normal hearing children. Cochlear Implants International 6: 123-140
Omran S (2011) Detecting diagonal activity to quantify harmonic structure preservation with cochlear implant mappings. International Journal of Robotics and Automation 1(5): 100-112
Omran S, Büchler M, Lai W, Dillier N (2008) Pitch perception using a Semitone mapping to improve music representation in Nucleus Cochlear Implants. In: DGA Congress, Kiel, Germany
Omran S, Lai W, Büchler M, Dillier N (2011) Semitone Frequency Maps to Improve Music representation for Nucleus Cochlear Implants. Audio speech and music processing journal Submitted
Omran SA, Büchler M, Lai W, Douglas R, Dillier N (2007) Comparison between Pitch Ranking of Complex Tones using an Acoustic Model of the Virtual Channels in Cochlear implants and an acoustic distance measure. In: ZNZ-Symposium, Zurich - Switzerland
Omran SA, Lai WK, Dillier N (2010) Pitch ranking, Melody contour and instrument recognition tests using two semitone frequency maps for Nucleus Cochlear Implants. EURASIP Journal on Audio, Speech, and Music Processing 2010
Ostroff JM, David EA, Shipp DB, Chen JM, Nedzelski JM (2003) Evaluation of the high-resolution speech coding strategy for the Clarion CII cochlear implant system. J Otolaryngol 32: 81-86
Oxenham AJ (2008) Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. Trends Amplif 12: 316-331
Pancaldi G (2003) Volta: Science and culture in the age of enlightenment. Princeton University Press
Paolini A, Clarey J, Needham K, Clark G (2004) Fast Inhibition Alters First Spike Timing in Auditory Brainstem Neurons. In: Journal of neurophysiology, vol 92, pp 2615-2621
Pasanisi E, Bacciu A, Vincenti V, Guida M, Berghenti MT, Barbot A, Panu F, Bacciu S (2002) Comparison of speech perception benefits with SPEAK and ACE coding strategies in pediatric Nucleus CI24M cochlear implant recipients. Int J Pediatr Otorhinolaryngol 64: 159-163
Patrick J, Clark G (1991) The Nucleus 22-channel cochlear implant system. Ear Hear 12: 3-9
Petitcolas F (1998) MATLAB implementation of the MPEG psychoacoustic model I [online]. In. www.petitcolas.net/fabien/software/index.html [Accessed June 1st 2009]
Pierce J (1983) The science of musical sound. Scientific American Books, New York

Pierce JR (1991) Periodicity and pitch perception. J Acoust Soc Am 90: 1889-1893
Pijl S (1997) Labeling of musical interval size by cochlear implant patients and normally hearing subjects. Ear Hear 18: 364-372
Pijl S, Schwarz DWF (1995a) Intonation of musical intervals by deaf subjects stimulated with single bipolar cochlear implant electrodes. Hear. Res. 89: 203-211
Pijl S, Schwarz DWF (1995b) Melody recognition and musical interval perception by deaf subjects with electrical pulse trains through single cochlear implant electrodes. J. Acoust. Soc. Am. 98: 886-894
Plomp R (1964) The ear as a frequency analyzer. J Acoust Soc Am 36
Rhode WS (1974) Evidence from Mässbauer experiments for nonlinear vibrations in the cochlea. J. Acoust. Soc. Am 55: 588–597
Rhode WS, Geisler CD, Kennedy DT (1978) Auditory nerve fiber responses to wide-band noise and tone combinations. J. Neurophysiol. 41: 692–704
Riemann B (2000) Musik Lexikon. Directmedia Publishing GmbH, Berlin
Robinson DW, Dadson RS (1956) A re-determination of the equal-loudness relations for pure tones. Br. J. Appl. Phys. 7: 166-181
Ruggero M, Robles L, Rich N (1992) Two-tone Suppression in the Basilar Membrane of the cochlea: Mechanical Basis of Auditor-Nerve Rate suppression. J Neurophysiol 68: 1087-1099
Sachs MB, Kiang NY (1968) Two-tone inhibition in auditorynerve fibers. J. Acoust. Soc. Am. 43: 1120-1128
Sachs MB, Voigt HF, Young ED (1983) Auditory nerve representation of vowels in background noise. J. Neurophysiol. 50: 27–45
Sadie S, Grove G (1995) The New Grove Dictionary of Music and Musicians. In, vol 17. Grove, London
Salzer F (1982) Structural hearing: tonal coherence in music. Dover Publications New York
Schouten JKNA (1938) Wetenschappen Proc. 41: 1086–1093
Scruton R (1997) The Aesthetics of Music. Oxford: Clarendon Press
Seligman P, Patrick J, Tong Y, Clark G, Dowell R, Crosby P (1984) A signal processor for a multiple-electrode hearing prosthesis. Acta Otolaryngologica: 135-139
Seligman PM (1987) Speech-processing strategies and their implementation. Annals of Otology, Rhinology & Laryngology 96: 71-73.
Shamma S, Klein D (2000) The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. J Acoust Soc Am 107: 2631-2644
Shamma SA (1985a) Speech processing in the auditory system. I: The representation of speech sounds in the responses of the auditory nerve. J Acoust Soc Am 78: 1612-1621
Shamma SA (1985b) Speech processing in the auditory system. II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. J Acoust Soc Am 78: 1622-1632
Simmons FB (1966) Electrical stimulation of the auditory nerve in man. . Arch Otolaryngol 84 24-76
Sit JJ, Simonson AM, Oxenham AJ, Faltys MA, Sarpeshkar R (2007) A low-power asynchronous interleaved sampling algorithm for cochlear implants that encodes envelope and phase information. IEEE Trans Biomed Eng 54: 138-149
Skinner M, Holden L, Holden T, Dowell R, Seligman P, Brimacombe J, Beiter A (1991) Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucleus multi-electrode cochlear implant. Ear Hear 12: 3-22
Skinner MW, Clark GM, Whitford LA, Seligman PM, Staller SJ, Shipp DB, Shallop JK, Everingham C, Menapace CM, Arndt PL, et al. (1994) Evaluation of a new spectral peak coding strategy for the Nucleus 22 Channel Cochlear Implant System. Am J Otol 15 Suppl 2: 15-27

Slaney M (1993) An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank. In. Apple Computer Technical Report
Slaney M (1998) Auditory toolbox ver 2 [online]. In. http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/ [Accessed October 1st 2009]
Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416: 87-90
Souza SD, D'souzaa N, D'souza DJF (2003) Recent advances in cochlear implant devices and techniques in India International Federation of Otorhinolaryngological Societies (IFOS) 1240: 365– 368
Sridhar D, Stakhovskaya O, Leake PA (2006) A Frequency-Position Function for the Human Cochlear Spiral Ganglion. Audiol Neurootol. 11: 16-20
Stevens SS, Jones RC (1939) The Mechanism of Hearing by Electrical Stimulation. J Acoust Soc Am 10: 8
Stockwell RG (2007) A basis for efficient representation of the S-transform. In: Digital Signal Processing, vol 17, pp Pages 371-393
Stoop R, Kern A (2004) Essential auditory contrast-sharpening is preneuronal Proc. Natl. Acad. Sci. 101: 9179–9181
Swanson B (2001) Nucleus® Technical Reference manual,Cochlear Ltd. Australia.
Swanson B (2005) Nucleus Matlab Toolbox 3.02. In. Cochlear Company
Swanson B. (2008). Pitch Perception with Cochlear Implants. *PhD thesis, Faculty of Medicine, Dentistry & Health Sciences, Otolaryngology Eye and Ear Hospital, The university of Melbourne*
Swanson B, Van Baelen E, Janssens M, Goorevich M, Nygard T, Van Herck K (2007) Cochlear Implant Signal Processing ICs. In: IEEE 2007 Custom Intergrated Circuits Conference (CICC), San Jose, CA, pp 437-442
Terhardt E (1998) Akustische Kommunikation. Springer-Verlag, Berlin
Thornton AR, Raffin MJM (1978) Speech - discrimination scores modeled as a binomial variable. J Speech Hear Res 21: 507-518
Tognola G, Pesatori A, Norgia M, Parazzini M, Rienzo LD, Ravazzani P, Burdo S, Grandori F, Svelto C Numerical Modeling and Experimental Measurements of the Electric Potential Generated by Cochlear Implants in Physiological Tissues. IEEE Transactions on Instrumentation and Measurement 56: pp 187-193
Townshend B, Cotter N, Van Compernolle D, White RL (1987) Pitch perception by cochlear implant subjects. J Acoust Soc Am 82: 106-115
Turicchia L, Sarpeshkar R (2005) A Bio-Inspired Companding Strategy for Spectral Enhancement. IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING 13: 243-253
Turner CW, Gantz BJ, Vidal C, Behrens A, Henry BA (2004) Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. J Acoust Soc Am 115: 1729-1735
Tye-Murray N, Lowder M, Tyler R (1990) Comparison of the F0/F2 and F0/F1/F2 processing strategies for the Cochlear Corporation cochlear implant. Ear Hear 11: 195-200
Van-Hoesel R (2001) Peak-Derived timing stimulation strategy for a multichannel cochlear implant. In: J Acoust Soc Am., Vol. 123, No. 6, Australia, p 4036
Vandali AE, Sucher C, Tsang DJ, McKay CM, Chew JW, McDermott HJ (2005) Pitch ranking ability of cochlear implant recipients: a comparison of sound-processing strategies. J Acoust Soc Am 117: 3126-3138
Vary P, Heute U, Weiss W (1998) Digitale Sprachsignalverarbeitung. B.G. Teubner, Stuttgart
Wagener K, Brand T, Kollmeier B (1999a) Development and evaluation of a German sentence test II: Optimization of the Oldenburg sentence test. . Audiologie 38 44-56

Wagener K, Brand T, Kollmeier B (1999b) Development and evaluation of a German sentence test III: Evaluation of the Oldenburg sentence test. Audiologie 38: 86-95
Wagener K, Kühnel V, Kollmeier B (1999c) Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test. (1), 4-15. Audiologie 38: 4-15
Wallenberger E, Battmer R (1991) Comparative speech recognition results in eight subjects using two different coding strategies with the Nucleus 22 channel cochlear implant. British Journal of Audiology 25: 371-380
Walser H (1993) Der Goldene Schnitt. Teubner, Stuttgart
Warren RM (1999) Auditory Perception: A New Analysis and Synthesis. Cambridge University Press, New York
Whitford LA, Seligman PM, Blamey PJ, McDermott HJ, Patrick JF (1993) Comparison of current speech coding strategies. Adv Otorhinolaryngol 48: 85-90
Wightman FL (1981) Pitch perception: an example of auditory pattern recognition. D.J. Getty and J.H. Howard, New Jersey
Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991) Better speech recognition with cochlear implants. Nature 352: 236-238
Wilson BS, Schatzer R, Lopez-Poveda EA, Sun X, Lawson DT, Wolford RD (2005) Two new directions in speech processor design for cochlear implants. Ear Hear 26: 73S-81S
Wilson BS, Wolford RD, Lawson DT, Schatzer R (2002) Speech processors for auditory prostheses: additional perspectives on speech reception with combined electric and acoustic stimulation. Third Quarterly Progress Report, Bethesda, MD. In: NIH project NO1-DC-2-1002, Neural Prosthesis Program, National Institutes of Health
Wyatt K, Schroeder C (1998) Harmony and Theory: a comprehensive source for all musicians. Musicians Institute Press
Wyrsch S. (2000). Adaptive Subband Signal Processing for Hearing Instruments. *Diss. ETH No. 13577. PhD thesis, The ETH Zurich*
Xu J, Stevenson AW, Gao D, Tykocinski M, Lawrence D, Wilkins SW, Clark GM, Elaine Saunders, Cowan RS (2001) The Role of Radiographic Phase-Contrast Imaging in the Development of Intracochlear Electrode Arrays. Otol Neurotol 22: 6
Xu L, Wei W (2007) Melody Perception Using a Novel Harmonic Spacing Filtering Scheme in Cochlear Implant Sound Processing. In: ARO 2007
Yan Z, Jie Z, Zhitong H (1998) Speech recognition based on lateral inhibition network auditory model. In: IEEE International Conference on Systems, Man, and Cybernetics, vol 5. IEEE, San Diego, CA, USA, pp 4154-4159
Zeng FG, Shannon R (1999) Psychophysical laws revealed by electric hearing. NeuroReport 10: 193l-1935
Zwicker E (1961) Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen). J Acoust Soc Am VOLUME 33: 248
Zwicker E, Fastl H (1990) Psychoacoustics: Facts and Models. . Springer-Verlag, 1990, Berlin, Germany

Kept blank intentionally

Acknowledgments

Many thanks to all the colleagues who supported me heavily during working in the audiometric laboratory; unfortunately I had hard times in the first 3 months when I joined the research group, due to pass-away of my beloved father. I like thank Belia, Erica and Cornelia who supported me when I got the news. Cornelia could luckily find me an airplane in a difficult season time.

I would like to thank my colleague Michael Büchler who cooperated with me much in the first 2½ years of my PhD, he is a nice guy and I was glad to work with, as I learned from him many technical details. I enjoyed working with Michi, because I used to come out with new signal processing ideas and he could follow me instantaneously without describing much of the details. Also, I was glad to work together with Waikong Lai. Waikong is a wonderful guy who was patient enough to guide me during my learning process; I learned from him many wonderful stuff, mainly in visual illustrations, concepts and text organizations. He has an interesting talent in using color to stress visual effects especially with Excel, which one may think that it is a humble application, but in fact Waikong showed me many excel tricks to produce difficult plots. We had many discussions and cooperation together. I was glad working with all the cochlear implant team; I had many interesting technical discussions with Markus who is an electronics Guru, and many supportive friends as Ann-Lena Hinze, Andrea Kegel, Steven Schimmel and Martin Mueller.

Many thanks to Dr. Martin Neukom and John Flury from the music institute for their keen support with musical stuff and ideas.

Many thanks to Professor Douglas who supported me in joining the institute of neural science, he guided me as well through my lab meeting presentations and gave me recommendations on how to hold interesting seminars.

Thank you professor Dillier for the strong support and weekly guidance; we had a weekly meeting in which I used to present my new results. Norbert encouraged me strongly when I started studying Neural-science courses and I had difficulties understanding many parts in it. With his moral encouragements, I could progress and get good results in it. Thank you, Norbert.

Many thanks to Andrea Kegel for her hard work translating my summary into german.

This thesis was supported in the first 3 years and 3 months with a grant number 320000-110043 from Swiss national foundation. In the last one and half year, it was also supported by 3 scholarships covering only institution fees from the University of Zurich and the rest was financed by my self.

Curriculum Vitae

M.Sc. Engineer (Dipl.ing)\ Sherif Omran

Post Address	**Nationality**	Egyptian	
Sherif A. Omran	**Born on**	1st August 1977	
Vogelsangst 10,	**Gender**	Male	
8006, Zurich	**Status**	Single	
Switzerland	**Email**	sherif.omran@gmx.de	
Cell Phone	+41-79-394-0965	**Home**	+ 41-44-337-9872
SIP Phone	770800@voxalot.com	**Home** (Enum & VOIP)	+ 49-89-721-0108-7914

EDUCATION

Feb 2006-Sep 2009	**Institute of Neuro-informatics (University of Zurich & ETH)** PhD program (Dr.Sc = German Dr.rer.nat) started. Expected to finish by June 2010. Thesis title "*Mechanism of music perception with Cochlear implants*".Developing new processing algorithms to improve music perception.
Feb 2007-Feb 2008	**University of Zurich –Switzerland** Studied neural science and graduated from "Centre of neural science Zurich -ZNZ"
Feb 2005	**University of Karlsruhe – Germany** Recognition of Egyptian master of science in engineering (German Dipl.ing)
Sep 2000 - Sep 2004	**University of Alexandria – Egypt** Master degree of science (M.Sc) degree in electrical engineering. Field: digital signal processing (DSP). Graduating thesis title "*New Approaches for Cardiac Sound Processing*". Developed new sound analyzing algorithms.
Sep1995 - Jun 2000	**University of Alexandria – Egypt** Bachelor degree of science BSc in telecommunications and electrical engineering. Graduating project (93%). Graduating thesis title "*Computer interfaced monitoring system*".
	Ennasr Boys' School – Alexandria, Egypt
Jun 1995	The Egyptian general secondary school certificate „Thanaweya". Final degree 93%.

SKILLS

Development Tools	Applications	Expert level using MS Office programs including VBA macro development.
	Operating system	Windows, Linux (Suse / Ubuntu) and MAC
	Programming languages and development tools	Expert in MATLAB, VB and VBA Very good knowledge in web development (ASP, PHP, MySQL, HTML and XML) and content management systems (Drupal, Joomla & Magento). Good knowledge in C and C++

Spoken Languages	Arabic	English	German	French	Greek
	Native	Very Good	Very Good	Good	Fair

WORK EXPERIENCE

Feb 2006-Apr 2009 (Part time – 50 %)	Private research - telecommunication engineer Working remotely to building voice over IP (VOIP) networks and setting them up (Asterisk server, call routing and building new firmware for VOIP boxes)
Feb 2006-Apr 2009 (Part time - 50 %)	Audiology department (ORL) - university hospital of Zurich –Switzerland. Signal Processing research scientist in a PhD program.
May 2005-Oct 2005	Neurology clinic of Grosshadern - university of Munich (LMU) – Germany Signal processing research including Matlab programming for a project titled "Correlation between eye movements and brain spikes".
Oct 2004-Feb 2005	University of Karlsruhe – Germany Guest scientific research in dynamic wireless Adhoc networks.
Feb 2004-Sep 2004	Ericsson telecom – Cairo – Egypt Ericsson support team; network administrator for Alexandria landline ATM networks. Client support and follow up.
Jun 2001-Feb 2004	Telecom Egypt – Alexandria - Egypt Installation and configuration of the AXD301 routers and AXE exchanges. Transformation of circuit switching to packet switching technology project. Set up & maintenance of 24 Zones. Duties included analyzing SS7 routing protocol and synchronising E1 telephone channels.
Jun 2004-Jul 2004	Ericsson training centre – Dublin – Ireland Installation, configuration and maintenance of bridging routers between circuit and packet switching exchanges.
2001 – 2003 (Part time job)	Different software companies Part time working in software and web production departments. Programming languages used ASP, PHP, SQL, C, Visual Basic, Assembly and MATLAB.
Sep 2001-Aug 2001	ERICSSON TRAINING CENTER – Dublin – Ireland - Data communication networking and configuration - Ericsson data communication solutions overview - ATM network operation & maintenance - Engine Introduction - Engine bridgehead network configuration and verification - AXD301 Installation, configuration and maintenance
Jan 2001-Jun 2001	FACULTY OF ENGINEERING– Alexandria University – Egypt Assistant teacher in the Electrical Engineering department Teaching of electrical measuring techniques, fibre optics telecommunication and digital logic circuit.

ACTIVITES AND HOBBIES

Networking	Member of the networks society of the IEEE organisation
Signal Processing	Member of the digital signal processing society of the IEEE organisation
Neuralscience	Member in the Swiss neural science
Hobbies	Running, Ping pong, Chess playing, programming, and learning languages

PUBLICATIONS AND CONFERENCES

Omran, S., Omran matrix: detecting diagonal activity to quantify harmonic structure preservation with cochlear implant mappings. in submitting, 2009
Omran, S., How far cochlear implant strategies have been developed: a comprehensive review. Under review, 2009.
Omran, S. et al. *Pitch ranking, Melody contour and instrument recognition tests using two semitone frequency maps for Nucleus Cochlear Implants*, in submitting process
Omran, S. et al. *Semitone Frequency Maps to Improve Music representation for Nucleus Cochlear Implants*, in submitting process
Omran, S., et al. *Pitch perception using a Semitone mapping to improve music representation in Nucleus Cochlear Implants.* in *DGA Congress*. 2008. Kiel, Germany.
Omran, S.A., et al. *Pitch Ranking of Complex Tones using a Model of the Virtual Channels in the Nucleus Freedom System.* in *EFAS 2007*. 2007. Heidelberg - Germany: German Society for Audiology (Deutsche Gesellschaft für Audiologie).
Omran, S.A., et al. *Comparison between Pitch Ranking of Complex Tones using an Acoustic Model of the Virtual Channels in Cochlear implants and an acoustic distance measure* in *ZNZ-Symposium*. 2007. Zurich - Switzerland.
Omran, S and Mazhar Tayerl. *Heart Sound Segmentation and Feature Extraction Algorithm.* IEEE Midwest conference 2003. Cairo - Egypt

AWARDS

Feb 06 – Apr 09	PhD working support from Swiss national science foundation.
May 05-Oct 05	Fellowship from Bernstein institute – Munich – Germany.

TRAININGS

EIO – Alexandria – Egypt (3 month)	Testing and fault tracing of Biomedical electronic boards.
NCR - Alexandria – Egypt (1 month)	Network configurations and adjustments.
RAGAB (Ford) – Alexandria - Egypt (2 Month)	Fault Tracing course to fix electronic printed circuit boards (PCB).
FOX – Alexandria - Egypt (3 Month)	Hardware development and PCB production.

REFERENCES

Dr. Michael Buechler	michael.buechler@usz.ch
Prof. Dr. Norbert Dillier	norbert.dillier@usz.ch
Dr. Von Bülow, Isolde	ivbuelow@nefo.med.uni-muenchen.de
Dr. Waikong Lai	waikong.lai@usz.ch
Dr. Ing Glassauer, Stefan	sglasauer@nefo.med.uni-muenchen.de

The End

تم بحمد الله