J.H.J. Allum, D.J. Allum-Mecklenburg, F.P. Harris and R. Probst
Progress in Brain Research, Vol. 97

CHAPTER 28

Speech encoding strategies for multielectrode cochlear implants: a digital signal processor approach

Norbert Dillier, Hans Bögli and Thomas Spillmann

Department of Otorhinolaryngology, University Hospital, Zurich, Institute for Biomedical Engineering, University of Zurich, and Swiss Federal Institute of Technology, Zurich, Switzerland

The following processing strategies have been implemented on an experimental laboratory system of a cochlear implant digital speech processor (CIDSP) for the Nucleus 22-channel cochlear prosthesis. The first approach (PES, Pitch Excited Sampler) is based on the maximum peak channel vocoder concept whereby the time-varying spectral energy of a number of frequency bands is transformed into electrical stimulation parameters for up to 22 electrodes. The pulse rate at any electrode is controlled by the voice pitch of the input speech signal. The second approach (CIS, Continuous Interleaved Sampler) uses a stimulation pulse rate which is independent of the input signal. The algorithm continuously scans all specified frequency bands (typically between four and 22) and samples their energy levels. As only one electrode can be stimulated at any instance of time, the maximally achievable rate of stimulation is limited by the required stimulus pulse widths (determined individually for each subject) and some additional constraints and parameters. A number of variations of the CIS approach have, therefore, been implemented which either maximize the number of quasi-simultaneous stimulation channels or the pulse rate on a reduced number of electrodes. Evaluation experiments with five experienced cochlear implant users showed significantly better performance in consonant identification tests with the new processing strategies than with the subjects' own wearable speech processors; improvements in vowel identification tasks were rarely observed. Modifications of the basic PES- and CIS strategies resulted in large variations of identification scores. Information transmission analysis of confusion matrices revealed a rather complex pattern across conditions and speech features. Optimization and fine-tuning of processing parameters for these coding strategies will require more data both from speech identification and discrimination evaluations and from psychophysical experiments.

Key words: Auditory prosthesis; Digital signal processing; Cochlear implants

Introduction

Today, cochlear implants can partially restore auditory sensations and limited speech recognition for profoundly deaf subjects despite severe technological and electrophysiological constraints imposed by the anatomical and physiological conditions of the human auditory system. Electrical stimulation via implanted electrodes still allows only an extremely limited approximation of normal neural excitation patterns in the auditory nerve. Signal processing for cochlear implants, therefore, is confronted with the problem of a severely restricted transmission channel capacity and the necessity to select and encode a subset of the information contained in complex sound signals.

The search for new signal processing schemes must consider the specific perceptual attributes of various electrical stimulation waveforms and patterns. In order to convey the contents of a particular transmitted message, its primary information elements (e.g., phonetic or acoustic speech features) should be transformed into those physical stimulation parameters which can generate distinctive perceptions for the listener. Practical experience with cochlear implants in the past indicate that some natural relationships (such as growth of loudness and voice pitch variations) should be maintained in

the encoding process.

Speech signals can be considered as sound intensity fluctuations in time and frequency. Due to the large redundancy implicit in speech signals, they can to a great extent be distorted and reduced, and still preserve their essential information content; a property which is most useful for communication systems such as vocoders. The consonant vowel transitions /ba/, /da/ and /ga/, for example, can be synthesized using very few parameters; mostly changes of first, second and third formant frequency (Fig. 1).

One method to decode such speech signals in a cochlear prosthesis is based on the tonotopic mapping of an array of intracochlear electrodes to frequency bands of the signal spectrum. Twenty bandpass filters (BPF) with logarithmically spaced center frequencies spanning the speech bandwidth of approximately 300 – 4000 Hz are indicated on the left of Fig. 1. A complicating constraint for all multielectrode coding schemes is the unnatural tonotopical relationship between the spectral components of the input signal and the place of stimulation along the basilar membrane as illustrated in Fig. 2. Electrodes are typically inserted between 15 and 20 mm into the scala tympani. Thus, the most apical electrode (transmitting the lowest frequency band of up to 300 Hz) will excite nerve fibers at place frequencies of approximately 700 – 1500 Hz (characteristic frequency). Although many CI-users have demonstrated their ability to adapt efficiently to these transposed and distorted frequency mappings, these aspects would certainly warrant additional systematic investigations. The instantaneous energy maxima of the speech spectra, which correspond roughly to the formants (F1, F2, F3 in Fig. 1), are determined and tracked over time. The fundamental voice frequency (about 100 Hz in the example of Fig. 1) triggers the stimulation pulses directed to the appropriate electrodes. Note the interleaved pulse sequence which is selected to avoid simultaneous stimulation of the same area by different electrodes. It is obvious that an encoding strategy based on F1 and F2 only (the strategy implemented in the wearable speech processor, WSP) will never allow a distinction between the synthetic syllable /da/ and /ga/ and also that the discrimination between /ba/ versus /da/ or /ga/ has to rely on only two or three low-amplitude stimulation pulses. It, therefore, seems logical to include information about higher formants in the stimulation patterns and to emphasize formant transitions by applying higher pulse rates.

Several processing strategies have already been designed and evaluated by varying the number of electrodes and the amount of specific speech feature extraction and mapping transformations used (Clark et al., 1990). Recently, Wilson et al. (1991)

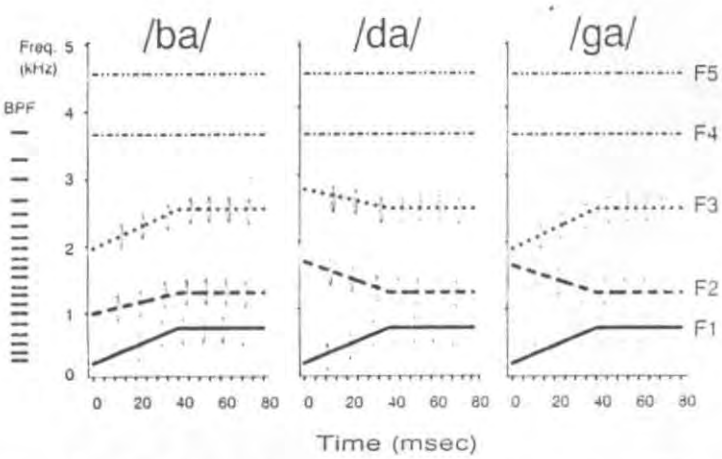

Fig. 1. Speech coding example: synthetic syllables /ba/, /da/, /ga/ (BPF: bandpass filters).

have reported astonishing improvements in speech test performance when they provided their subjects with high-rate interleaved pulsatile stimulation patterns (continuous interleaved sampler, CIS) on six rather than analog broad-band signals on four monopolar electrodes. They attributed this effect partly to the decreased current summation obtained by non-simultaneous stimulation of different electrodes (which might have stimulated the same nerve fibers and, thus, interacted in a non-linear fashion) and partly to a fundamentally different and, perhaps, more natural firing pattern due to an extremely high stimulation rate. Skinner et al. (1991) also found significantly higher scores on word and sentence tests in quiet and noise with a new multipeak speech coding strategy (MPEAK) implemented in the new miniature speech processor (Nucleus-MSP) as compared to the formerly used F0F1F2 of the Nucleus-WSP. Note that the MPEAK strategy does not correspond exactly to the formant extraction scheme denoted in Fig. 1; the presence of high-frequency spectral information (F3 and possibly F4) is signaled on two fixed electrodes within the range assigned to F2. Thus, the MPEAK strategy might generate identical patterns for the two different F3 trajectories of Fig. 1. In contrast, a formant or peak-picking channel vocoder strategy, as described below, would generate different patterns for the three syllables of Fig. 1., albeit with less mapping resolution for F2.

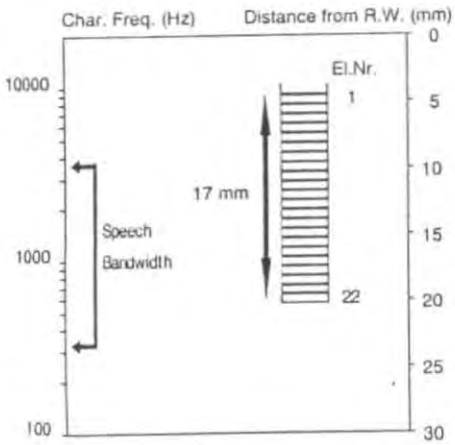

Fig. 2. Electrode positions along the basilar membrane and tonotopic relations (characteristic frequency) according to Greenwood (1990).

The present study was conducted in order to explore new ideas and concepts of multichannel pulsatile speech encoding for users of the Clark/Nucleus cochlear prosthesis. The optimization of signal processing schemes for existing implanted devices must consider a variety of technological and physiological aspects and is largely based on experiments with implanted subjects. With single-chip digital signal processors (DSP's) incorporated in personal computers, different speech coding strategies can be evaluated in relatively short laboratory experiments. In addition to the well known strategies realized with filters, amplifiers and logic circuits, a DSP approach allows the implementation of much more complex algorithms involving speech feature contrast enhancement, adaptive noise reduction and much more. Technological progress will most certainly allow further miniaturization and low-power operation of these processors in the near future.

Materials and methods

As previously described (Dillier et al., 1990), a cochlear implant digital speech processor (CIDSP) for the Nucleus 22-channel cochlear prosthesis was implemented using a single-chip digital signal processor (TMS320C25), Texas Instruments). For laboratory experiments, the CIDSP was incorporated in a general purpose computer (PDP11/73) which provided interactive parameter control, graphical display of input/output, intermediate buffers and off-line speech file processing facilities. For field studies and as a take-home device for patients, a wearable, battery-operated unit has been built. Advantages of a DSP implementation of speech encoding algorithms, as opposed to off-line prepared test lists, are increased flexibility; controlled, reproducible and fast modifications of processing parameters; and use of running speech for training and familiarization.

Speech signals were processed as follows: after analog low-pass filtering at 5 kHz and analog-to-digital-conversion (sampling rate 10 kHz), pre-

emphasis and Hanning windowing (12.8 msec, shifted by 6.4 msec or less per analysis cycle) was applied and the power spectrum calculated via fast Fourier transform (FFT); specified speech features such as formants and voice pitch were extracted and transformed according to the selected encoding strategy; and finally, the stimulus parameters (electrode position, stimulation mode, pulse amplitude and duration) were generated and transmitted via inductive coupling to the implanted receiver. In addition to the generation of stimulus parameters for the cochlear implant, an acoustic signal based on a perceptive model of auditory nerve stimulation was output simultaneously.

Two main processing strategies were implemented on this system. The first approach (PES, Pitch Excited Sampler) was based on the peak-picking channel vocoder concept (Flanagan, 1972) whereby the spectral energies of a number of frequency bands (typically third-octave bandwidths) are transformed into electrical stimulation parameters for up to 22 electrodes. Fig. 3 shows an example of a short-time power spectrum (vowel /a/) and its division into 22 frequency bands. The six largest peaks are marked by filled circles. Note that this coding scheme is not identical to a spectral maxima picker (as used, for example, by McKay et al., 1991) which, in this example, would pick more bands in the F1 – F2 region and disregard the high-frequency peaks. The pulse rate at any electrode is controlled by the voice pitch of the input speech signal. The pitch extractor algorithm calculates the autocorrelation function of a low-pass-filtered segment of the speech signal and searches for a peak within a specified window of time lags. A random pulse rate of about 150 – 250 Hz is used for unvoiced speech portions. Fig. 4 displays schematically the pitch-synchronous sequence of activated electrodes for the spectrum of Fig. 3.

The second approach (CIS, Continuous Interleaved Sampler) used a stimulation pulse rate which is independent of the input signal. The algorithm scans continuously all frequency bands and samples their energy levels as indicated in Fig. 3. As only one electrode can be stimulated at any instance of time, the rate of stimulation is limited mainly by the required stimulus pulse widths (determined individually for each subject) and the time to transmit additional stimulus parameters. As the information about the electrode number, the

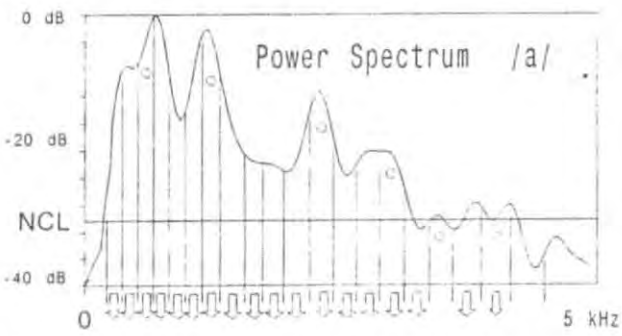

Fig. 3. Short time spectral analysis (128 point Hanning window, FFT, power spectrum) and mapping to *n* frequency bands corresponding to *n* electrode pairs (*n* may typically vary between 16 and 22) for a vowel segment. Frequency bands whose average power exceeds the noise cut level (NCL) are marked by arrows. Spectral peaks (locally maximal band energies relative to the preceding and following bands) are marked by circles.

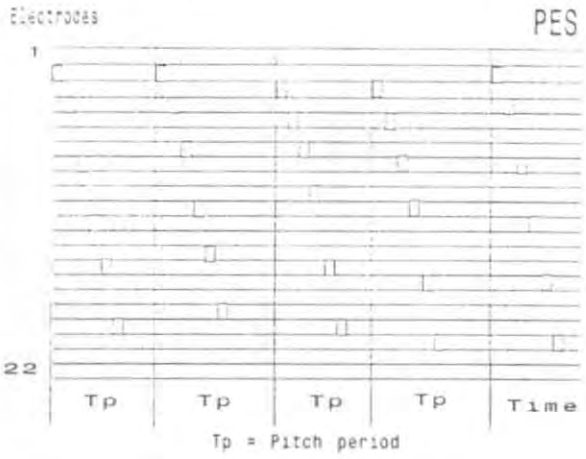

Fig. 4. Two-dimensional electrodogram (active electrodes vs. time) of the PES strategy. Electrodes corresponding to peaks in the short-time spectrum are stimulated pitch synchronously. The signal energy in the corresponding frequency band determines the stimulus amplitude. Five pitch periods are displayed. The third segment corresponds to the power spectrum of Fig. 3 where the six selected peaks are marked by circles. Electrode 22 is mapped to the lowest spectral band, electrode 1 to the highest frequency band.

stimulation mode, the pulse amplitude and width is encoded by high-frequency (2.5 MHz) bursts of different durations, the total transmission time for a specific stimulus depends on all of these parameters. This transmission time can be minimized by choosing the shortest possible pulse width combined with the maximal amplitude. For very short pulse durations, the overhead imposed by the transmission of the fixed stimulus parameters can become rather large. Consider, for example, the stimulation of electrode pair (21, 22) at 50 μsec. The maximally achievable rate varies from about 3600 Hz for high amplitudes to about 2700 Hz for low amplitudes whereas the theoretical limit would be close to 10000 Hz (biphasic pulses with minimal interpulse interval). In cases with higher pulse width requirements (which may be due to poor nerve survival or unfavorable electrode position or other unknown factors), the overhead gets smaller. Thus, typical maximal continuous rates for six electrodes range from 300 to 600 Hz.

In order to achieve maximally high stimulation rates for those portions of the speech input signals which are assumed to be most important for intelligibility, several modifications of the basic CIS strategy were designed. The analysis of the short time spectra was performed either for a large number of narrow frequency bands (approximately third-octave spacing, corresponding directly to the number of available electrodes) or for a small number (typically six) of wide frequency bands (approximately octave spacing) analogous to the approach suggested by Wilson et al. (1991). The mapping of spectral energy within any of these frequency bands to stimulus amplitude at a selected electrode was done according to several rules. The first scheme (CIS-NP) used a pre-selected number of peaks (typically again six) of the narrow-band analysis spectra relying on the same spectral feature extraction as the PES strategy. The second scheme (CIS-NA) was also based on the narrow-band spectral analysis but used all channels whose values exceeded a pre-set noise cut-out level (NCL), as shown in Fig. 5. Two variations of the wide-band analysis scheme were implemented which mapped the spectral energy of each band either to the same pre-selected electrode (fixed tonotopical mapping, CIS-WF) or to those electrodes within the wide bands which corresponded to local peaks (variable tonotopic mapping, CIS-WV). The CIS-WF scheme was supposed to minimize electrode interactions by establishing maximal distances between stimulated electrodes while the CIS-WV would make optimal use of tonotopical frequency selectivity. In both the PES- and the CIS-strategies, a high-frequency pre-emphasis was applied whenever a spectral gravity measure exceeded a pre-set threshold.

Subjects, fitting procedure, speech tests

To date, evaluation experiments have been conducted with seven cochlear implant users. Data of only five of these subjects will be presented in this paper as the remaining two subjects did not participate in all experiments. Table I lists some etiologic and technical information. All subjects were experienced users of their speech processors (time since implantation ranged from five months (KW) to nearly ten years (UT, implantation of single-channel device in 1980) with minimal open speech discrimination in monosyllabic word tests (scores

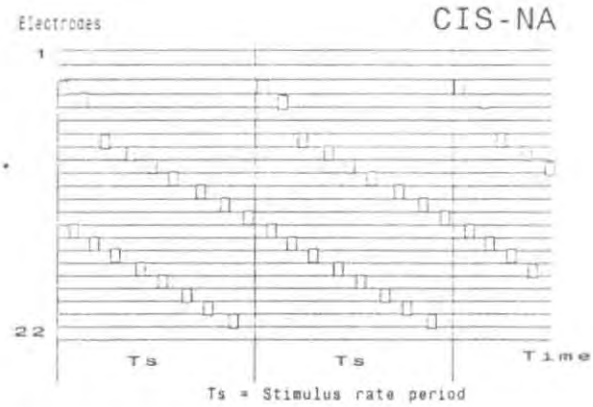

Fig. 5. Electrodogram of the CIS-NA strategy. All electrodes are stimulated sequentially at maximal rate. Spatial distances between sequentially activated electrodes are optimized as indicated in the graph. Values below a pre-set noise cut level (NCL) produce no output pulses. The second segment corresponds to the power spectrum of Fig. 3 where the frequency bands with energy above the NCL are marked by arrows.

TABLE I

Subjects information

Identification (Sex)	UT (f)	TH (m)	HS (m)	KW (m)	SA (f)
Date of birth	6/1941	2/1965	11/1944	3/1947	7/1962
Etiology (duration, years)	Sudden (15)	Trauma (3)	Sudden (14)	Meningitis (28)	Sudden (1)
Implementation date (side)	3/87 (L)	10/89 (R)	11/88 (R)	12/90 (L)	3/89 (L)
Speech processor (strategy)	WSP (F0F1F2)	MSP (F0F1F2)	MSP (MPEAK)	MSP (MPEAK)	MSP (MPEAK)
Electrode pairs	16	20	19	18	20
Stimulation mode	BP	BP + 1	BP	BP	BP
T-level (mean charge/phase, nC)	73	79	38	74	37
C-level (mean charge/phase, nC)	137	157	84	130	62
Pulse width (μsec)	150 – 204	204	100	204	100

Note: subject UT had pulse widths of 150 μsec on most bipolar electrode pairs except for electrodes 20, 17, 9 and 7 where the pulse width was set to 204 μsec.

from 5 to 25%) and limited use of the telephone. One subject (UT) still used the old wearable speech processor (WSP) which extracts only the first and second formant and stimulates only two electrodes per pitch period. Three subjects used the new miniature speech processor (MSP) with the so-called multipeak strategy whereby in addition to first and second formant information three fixed electrodes may be stimulated to convey information contained in three higher frequency bands. Subject TH used the MSP with the F0F1F2 strategy due to elevated stimulus levels.

The same measurement procedure was used to determine thresholds of hearing (T-levels) and comfortable listening (C-levels) for the CIDSP strategies as when fitting the WSP or MSP. Fig. 6 shows an example of these data for subject HS. Note that the stimulus levels are expressed as charge per phase in

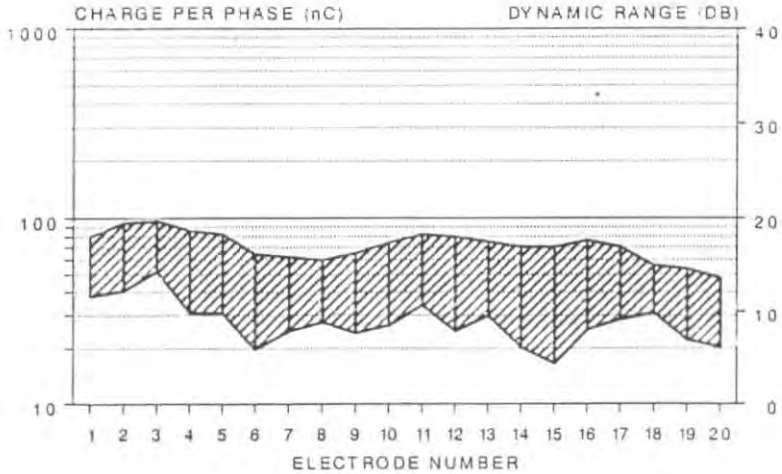

Fig. 6. Thresholds of hearing (T-levels) and comfortable loudness levels (C-levels) for subject HS, electrodes 1 to 20, bipolar stimulation (BP) with 100 μsec pulse width. The charge per phase was calculated as the product of amplitude times pulse width using a calibration table supplied by the electrode manufacturer (Cochlear Pty., Sydney).

nanoCoulomb (nC) instead of fitting software-specific current or stimulus levels. This was done to facilitate comparisons of different stimulus parameters and electrode configurations. As most subjects used fixed amplitudes and varying pulse widths (so-called stimulus levels) with their MSP's, and because the CIDSP algorithms required fixed pulse widths and varying amplitudes, all T- and C-levels were first remeasured prior to speech tests. Overall loudness of processed signals was adjusted by proportional factors (T- and C-modifiers) if necessary, following short listening sessions with ongoing speech and environmental sounds which were played from a tape recorder. T- and C-levels, averaged over all active electrodes, are shown in Table I.

Only minimal exposure to the new processing strategies was possible due to time restrictions. After about 5 – 10 min of listening to ongoing speech, one or two blocks of a 20-items, 2-digit numbers test with feedback of correct or wrong responses were administered. There was no feedback given during the actual test trials. Consonant and vowel identification, as well as multiple-choice, minimal-pair tests (rhyme tests, MAC battery) were performed. All test items were presented by a second computer via a D/A converter, low-pass filter (5 kHz) and attenuator. The subjects' responses were entered via a touch screen terminal (for multiple-choice tests) or keyboard (numbers tests and monosyllable word tests) and recorded automatically by the same computer. Speech signals were either presented via a loudspeaker in a sound-treated room (when patients were tested with their speech processors) or processed by the CIDSP in real time and fed directly to the transmitting coil at the subjects' head. Different speakers were used for the ongoing speech, the numbers tests and the actual speech tests, respectively. Table II shows the 12 consonants presented in /aCa/ context and the phonological and acoustic feature assignment matrices used in the following information transmission analyses (Miller and Nicely, 1955). The second set of features (Envelope, Lowfreq. and Duration) was obtained through an acoustic analysis of the digitized speech tokens used in the experiments and a correlation with the results of a multidimensional scaling analysis of listening experiments which included hearing aid users and cochlear implant users (Dillier and Spillmann, 1988).

Results

Fig. 7 summarizes the results for consonant identification tests for five subjects and six different processing conditions. The percentages of correctly identified tokens have been corrected for chance level. It can be seen that three subjects (UT, TH, SA) scored significantly lower with their own speech

TABLE II

Consonant phoneme features

Phoneme	p	t	k	b	d	g	m	n	l	r	f	s
Voicing	–	–	–	+	+	+	+	+	+	+	–	–
Nasality	–	–	–	–	–	–	+	+	–	–	–	–
Sonorance	–	–	–	+	+	+	+	+	+	+	–	–
Sibilance	–	–	–	–	–	–	–	–	–	–	–	+
Frication	–	–	–	–	–	–	–	–	–	–	+	+
Place	1	2	3	1	2	3	1	2	2	3	2	2
Manner	1	1	1	2	2	2	3	3	4	4	5	5
Envelope	3	3	3	2	2	3	1	1	1	1	2	2
Lowfreq.	+	+	+	–	–	+	–	–	–	+	+	+
Duration	3	3	2	2	3	2	1	1	2	1	3	3

processors than with any of the tested CIDSP strategies. Subject HS performed better with two of the four CIS variations, subject KW did not improve his scores with any of the digital processing strategies as compared to the performance with his MSP (note that KW had been using his speech processor for only five months. He had been deaf for nearly 28 years; all other subjects had lost their hearing less than 14 years before implantation). Most subjects reached more than 10% higher scores with the best CIS strategy compared to the PES coding. Among the four CIS variations, the NA and WF modes turned out to be superior. PES and CIS-NP resulted in about the same performance which could have been predicted by their algorithmic similarity. It is interesting to note that CIS-WV produced mostly lower scores than the similar CIS-WF (subject SA, unfortunately, was not tested with CIS-WV). This points to the importance of the mapping functions between spectral energy and stimulation electrodes. A variable mapping (i.e., a selection of the active electrode based on the local maximum within an octave frequency band) was apparently less effective than a fixed mapping which preserved constant maximal distances between activated electrode pairs. It is also interesting to note that activation of all possible electrodes (CIS-NA) within a stimulation cycle can lead to the same or even better performance as high-rate stimulation of only six electrodes (CIS-WF).

To gain more detailed insight into the performance differences with these processing conditions, an information transmission analysis of the pooled confusion matrices of all five subjects was performed. Fig. 8 shows the percentages of transmitted stimulus information for the phonological feature set of Table II. Again, the overall information transmitted is highest for CIS-NA and CIS-WF and lowest for the patient's own speech processor (WSP for UT and MSP for the other four subjects). Largest differences are seen for the sibilance and frication features (high-frequency a-periodic signals), whereas for sonorant consonants (and for the nasality distinction) the W/MSP seemed to perform as well as the peak-picker strategies (PES, CIS-NP, CIS-WV). It is well known that the phonological features are not orthogonal and may mislead interpretation due to their inherent redundancies. Therefore, a more detailed sequential analysis (Wang and Bilger, 1973) with the same phonological

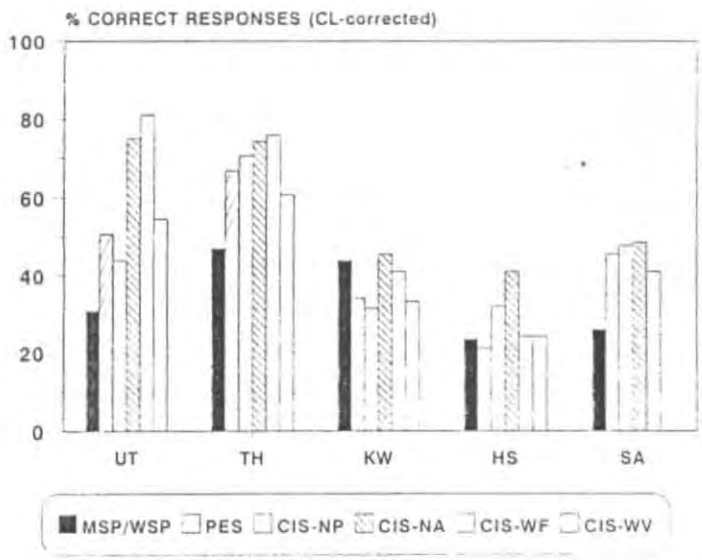

Fig. 7. Percent correct responses for five subjects tested with six different signal processing conditions (12 consonants test /aCa/, 144 trials). Chance level corrected scores: S = (R − CL)/(100 − CL), where R = raw score (%) and CL = chance level (%).

feature set and an additional set of acoustic features was carried out. Table III summarizes the main results of these analyses. The overall percentages are identical to the values displayed in Fig. 8. It can be seen that the phonological feature set (4 – 5 features selected out of seven) explains between 80 and 90% of the stimulus information ("Total" values) whereas the three acoustic features leave between 20

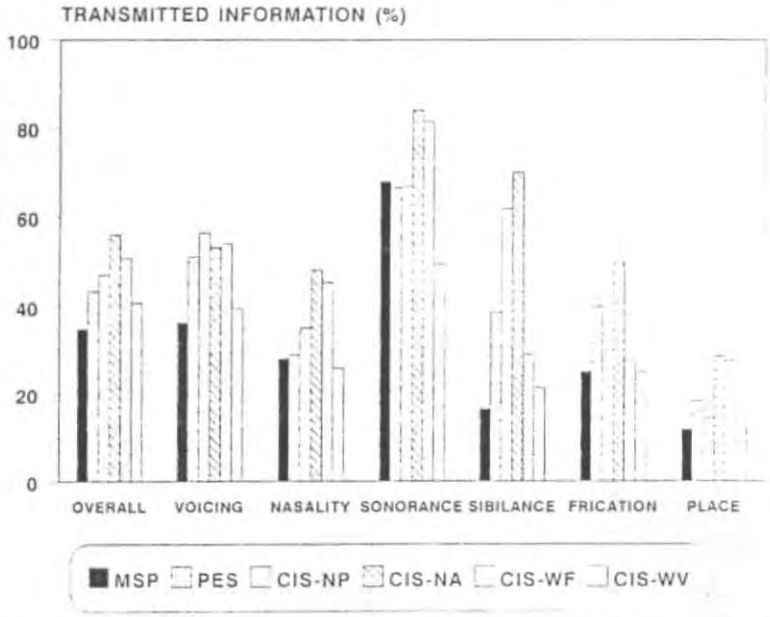

Fig. 8. Information transmission analysis of the consonant identification test data. Confusion matrices of the five subjects were pooled for the six processing conditions.

TABLE III

Sequential information transmission analysis (SINFA) of pooled consonant confusion matrices for five subjects with two different sets of speech features

	W/MSP	PES	CIS-NP	CIS-NA	CIS-WF	CIS-WV
Overall	34.8	45.3	47.0	56.0	50.6	40.7
Voicing	5.2	12.0	13.3	10.8	14.1	13.9
Nasality	3.9				5.9	
Sonorance	49.9	37.6	36.3	38.4	41.3	31.0
Sibilance		8.0	12.7	12.7	5.7	
Frication	7.6	2.3		3.6		8.5
Place	14.9	19.9	18.0	24.2	21.2	23.9
Manner			3.1			
Total	81.5	79.8	83.5	89.6	88.2	77.3
Envelope	57.1	50.5	46.4	50.3	52.6	42.9
Lowfreq.	13.8	16.2	13.1	18.8	14.4	11.7
Duration	6.1	6.9	9.2	8.6	13.3	11.0
Total	77.0	73.5	68.7	77.7	80.3	65.6

and 35% of the stimulus information unexplained. Sonorance information is highest for all conditions, followed by place of articulation and voicing (or sibilance, CIS-NA). As variations of articulatory positions are mainly correlated with changes in second and third formant frequencies, one might conclude that the CIS strategies, in this case, are better encoding information about formant transitions than the other approaches. Note, as well, that voicing information transmission is also improved although the CIS strategies do not explicitly encode the voice pitch as do the speech processor and the PES strategy;

Vowel identification scores, on the other hand, vere rarely improved by modifications of the signal processing strategy. Only one of the four subjects (HS) showed significant improvement in total scores, another subject (TH) had markedly deteriorated performance with either PES- and CIS strategies.

Discussion and conclusions

These speech test results should be considered as preliminary. The number of subjects is still very small and data collection for additional subjects is ongoing. It is hoped that other variations of the CIS strategy will lead to even larger improvements in recognition scores for these optimal cases. It should be kept in mind that these data were obtained in the laboratory without specific training and with only minimal exposure to the new processing schemes.

Inspection of the information transmission analysis seems to indicate a strong preference for a non-feature extraction approach such as maximal pulse rates on either all electrodes (narrow-band analysis with wide spatial separation of sequentially stimulated electrodes) or a limited fixed set of electrodes (wide-band analysis with preservation of fine temporal envelope information). The fact that the acoustic feature set used in the sequential analysis could not account for the same amount of stimulus information as the phonological feature set points to the importance of further investigations into the effects of transformation and encoding of speech signals for the specific case of multielectrode cochlear implant configurations (see Soli and Arabie, 1979, for further discussion). These findings need to be confirmed with more subjects. There are still many variables whose influence on the ability of implanted subjects to recognize and discriminate speech information should be explored further. Additional technical solutions may be found in the future which will allow the application of maximally high pulse rates even in subjects with elevated thresholds requiring stimulation with larger pulse widths. Whether the auditory nerve is reacting in a fundamentally different manner at continuously high excitation states and, thus, would generate a more natural auditory percept could be a very interesting basic research question. Psychophysical experiments using stochastic and jittered stimulus patterns (Dobie and Dillier, 1985) have demonstrated the ability of cochlear implant subjects to use pulse rate cues well above 300 Hz. Van den Honert and Stypulkowski (1987) have studied single-fiber response patterns for various pulsatile and analog stimulus waveforms and discussed possible consequences for speech encoding.

It is, however, very encouraging that new signal processing strategies can considerably improve speech discrimination. Consonant identification, apparently, may be enhanced by more detailed temporal information and specific speech feature transformations. Whether these improvements in phoneme discrimination can be transferred to a generally improved word and sentence recognition in quiet and noise environments also remains to be verified. Preferably further optimization of these processing strategies should include more specific data about loudness growth functions for individual electrodes or additional psychophysical measurements.

Although many aspects of speech encoding can be efficiently studied using a laboratory digital signal processor it would be desirable to allow subjects more time for adjustment to a new coding strategy. Several days or weeks of habituation are sometimes required until a new mapping can be fully exploited. Specific interactive training might shorten this

period in the future. Thus, for scientific as well as practical purposes, the miniaturization of wearable DSP's will be of great importance.

Acknowledgements

This work was supported by the Swiss National Research Foundation (grants no. 4018 – 10864 and 4018-10865). Implant surgery was performed by Prof. U. Fisch. Valuable help was also provided by Dr. E. von Wallenberg of Cochlear AG, Basel.

References

Clark, G.M., Tong, Y.T. and Patrick, J.F. (1990) *Cochlear Prostheses*, Churchill Livingstone, Edinburgh, London, Melbourne and New York, 264 pp.

Dillier, N. and Spillmann, T. (1988) Wahrnehmung akustischer Sprachmerkmale mit einkanaligen Cochlea-Implantaten und Hörgeräten. Eine vergleichende Analyse. *HNO*, 36: 335 – 341.

Dillier, N., Senn, C., Schlatter, T., Stöckli, M. and Utzinger, U. (1990) Wearable digital speech processor for cochlear implants using a TMS320C25. *Acta Otolaryngol. (Stockh.) (Suppl.)*, 469: 120 – 127.

Dobie, R.A. and Dillier, N. (1985) Some aspects of temporal coding for single-channel electrical stimulation of the cochlea. *Hearing Res.*, 18: 41 – 55.

Flanagan, J.L. (1972) *Speech Analysis, Synthesis and Perception*, Springer, Berlin, 2nd ed., pp. 321 – 405.

Greenwood, D.D. (1990) A cochlear frequency-position function for several species – 29 years later. *J. Acoust. Soc. Am.*, 87: 2592 – 2605.

McKay, C., McDermott, H., Vandali, A. and Clark, G. (1991) Preliminary results with a six spectral maxima sound processor for the University of Melbourne/Nucleus multiple-electrode cochlear implant. *J. Otolaryngol. Soc. Austr.*, *6:* 354 – 359.

Miller, G.A. and Nicely, P.E. (1955) An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.*, 27: 338 – 352.

Skinner, M.W., Holden, L.K., Holden, T.A., Dowell, R.C., et al. (1991) Performance of postlingually deaf adults with the wearable speech processor (WSP III) and mini speech processor (MSP) of the Nucleus Multi-Electrode Cochlear Implant. *Ear Hear.*, 12/1: 3 – 22.

Soli, S.D. and Arabie, P. (1979) Auditory versus phonetic accounts of observed confusions between consonant phonemes. *J. Acoust. Soc. Am.*, 66: 46 – 59.

Van den Honert, C. and Stypulkowski, P.H. (1987) Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli. *Hearing Res.*, 29: 207 – 222.

Wang, M.D. and Bilger, R.C. (1973) Consonant confusions in noise: a study of perceptual features. *J. Acoust. Soc. Am.*, 54: 1248 – 1266.

Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K. and Rabinowitz, W.M. (1991) Better speech recognition with cochlear implants. *Nature*, 352: 236 – 238.