Speech recognition

Birger Kollmeier

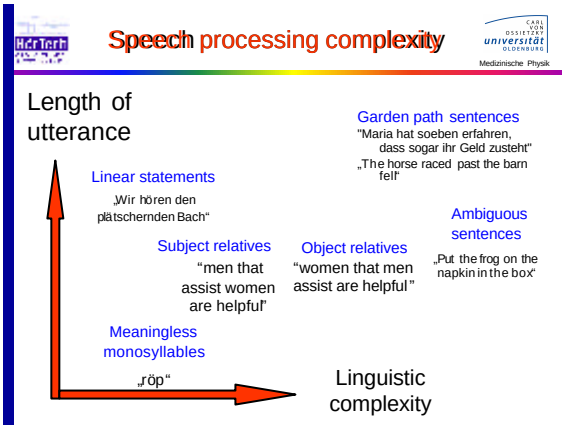
Medical Physics, Universität Oldenburg & Kompetenzzentrum HörTech, D-26111 Oldenburg, Germany

In order to better understand the effect of hearing impairment on speech perception in everyday listening situations as well as the limited effect of modern hearing instruments in improving the situation for hearing-impaired listeners, a thorough understanding of the mechanisms and factors influencing speech recognition in quiet and in noise is highly desirable. This contribution therefore reviews the theoretical background, the currently employed measurement methods, and the practical implications of measuring speech recognition in patients. A special emphasis is put on the comparability of speech intelligibility tests across different languages which is considered in the European HEARCOM project. Further on, the degree to which we understand how speech recognition "works" in normal and hearing-impaired listeners is discussed. Both bottom-up and top-down strategies have to be assumed when trying to understand speech reception in noise. Computer models that assume a near-to-perfect "world knowledge", i.e., an accurate anticipation of the speech unit to be recognized, can surprisingly well predict the performance of human listeners in noise and may provide a useful tool in hearing aid development. Finally, the cognitive abilities of human listeners when understanding speech are challenged by considering fluctuating background noise where hearing impaired listeners vary considerably in their respective ability to combine the information from "listening into the dips". In addition, the performance for syntactically "difficult" vs. "simple" sentence structures highlight the interaction between hearing impairment and cognitive processing structures, such as, e.g., working memory.

(Work supported by DFG, CEC-Project Hearcom, and the Audiologie-Initiative Niedersachsen).

MEASURING SPEECH RECOGNITION IN HUMANS

The perception of speech in normal and hearing-impaired listeners is mostly performed under non-ideal, i.e. "difficult" acoustical situations which has often been ignored in laboratory studies. An appropriate two-dimensional representation of a variety of such ecologically important communication situations can be given if one dimension represents the reverberation time (ranging from zero in free field situations to several seconds in large rooms with many acoustical reflections) and the other dimension represents the number of interfering noise sources (ranging from zero in "ideal" communication situations to large numbers in "cocktail-party-situations"). While the vast majority of communication situations are characterized by reverberation and interfering noise sources, in most standard audiometric listening situations only a very artificial situation with no reverberation and one or even zero interfering noise sources is used. A similar argument holds for the complexity of the speech test materials normally used for standard speech audiometric tests in clinical environments where another two-dimensional characterization of natural speech utterance is possible: Again, the length of the speech utterance to be recognized by the listener may represent one dimension (ranging from monosyllabic words (such as, e. g. logatomes) to multisyllabic words and up to complete sentences) and the linguistic complexity could represent the second dimension (ranging from simple linear word or sentence structures up to garden path sentences where the complete sentences can only be correctly understood if the last word has been processed and understood correctly: "The horse raced past the barn fell".) Again, in this two-dimensional representation most everyday communication will occur at a comparatively high utterance length with a moderate linguistic complexity while audiometric testing is usually done for short utterances at a very low linguistic complexity. These examples show that speech reception in every day communication is far more complex than usually considered in artificial situations in the laboratory.



Nevertheless, if we restrict ourselves to comparatively "simple" acoustical situations in the laboratory using comparatively simple speech materials it can be stated that substantial progress has been made within the last decades to understand speech reception and the specific influence of the various parameters involved. Typically, the speech reception threshold in noise is assessed, i. e., the speech-to-noise ratio required to achieve a certain percent correct (in most cases: 50%) of the speech material employed. This quantity measures indirectly the "maximum comfortable communication distance" in noisy situations, i. e. the spatial distance between listener and speaker in a real-life situation which is normally assumed by the listener as a compromise between maximizing the speech intelligibility and keeping a socially acceptable distance between talker and listener. In hearing-impaired listeners, this distance is typically significantly reduced and most of the complaints about hearing difficulties arise from hearing-impaired listeners not in quiet, but in noisy conditions. Hence, ways of reliably and efficiently quantifying the speech reception threshold are required as a measure of the hearing problem in every day listening situations. In addition, a better theoretical knowledge of the influence of a variety of parameters on the individuals speech recognition is highly desirable.

Existing German Speech tests

Medizinische Physik

Test items	Name of test	Test material per list	Reference
Logatomes (nonsense monosyllabic)	Kieler Logatomtest	CVC	Müller-Deile (pers. Comm.)
	OLLO – Oldenburg logatome corpus	150 VCV and CVC, 40 speaker, 6 variabilities	Meier et al., 2005
Monosyllabic meaningful	Freiburger Einsilbertest	20 common words	Hahlbrock, 1953
Test in noise	Dreinsilber-Test	3 repeated monosyllables	Döring& Hamacher, 1992
	Einsilber Reimtest	33+33+34 words per list, 6 rhyme alternatives	Sotscheck, 1982
Test in noise&quiet	Einsilber Reimtest (WAKO)	33+25+14 words per list, 5 rhyme minimum pair alternatives	v.Wallenberg & Kollmeier, 1989
	Verkürzter Reimtest	25 rhyme pairs	Brand &Wagener, 2005
Bisyllabic meaningful	Zweisilber-Reimtest	24+24+24 words, 4 rhyme minimum pair alternatives	Kliem & Kollmeier, 1994
	Oldenburger Kinder-Reimtest	12 words, 3 pictoral rhyme pairs	Kliem & Kollmeier,1995
	AAST-Test	6 spondees, pictoral response	Coninx, 2005
Multisyllabic	Freiburger Zahlentest	10 numerals, 4-5 syllables	Hahlbrock, 1953
	Zahlentripel-Test	10 3-digit strings	Wagener et al.,2005
Sentences	Marburger Satztest	10 short meaningful sentences	Niemeyer, 1967
	Basler Satztest	15 high predictable & 15 low predictable sentences	Tschopp & Ingold 1992
	Göttinger Satztest	10 short meaningful sentences	Wesselkamp & Kollmeier, 1994
	HSM-Satztest	20 short meaningful sentences	Hochmair et al., Schmidt et al., 1997
	Oldenburger Satztest	10 syntactically fixed, unpredictable sentences	Wagener et al., 1999

A classification of the various speech tests available for audiological and clinical usage in German is listed above. They vary with respect to the number of syllables per speech item, their respective construction principle and if the test has been designed to be performed in quiet (displayed in white background colour) or in noise (light yellow background colour). Several tests (like most of the tests developed in recent years in Oldenburg) have been evaluated both for being used in quiet and in noise (dark yellow background colour).

To extent the scope of speech tests to different European languages, the comparability of sentence tests was assessed within a multi-centre study of the HEARCOM project (Wagener et al., 2006, 2007). Five partner sites from four different European countries participated in the measurements. Netherlands: Academic Medical Center Amsterdam and VU University Medical Center Amsterdam, Sweden: Linköping University, Dept of Audiology, United Kingdom: University of Southampton Institute of Sound and Vibration Research, Germany: Hörzentrum Oldenburg. Sentence intelligibility was determined in different conditions: So-called Plomp type sentences (short meaningful sentences) were used to determine the binaural SRT in quiet, monaural SRT in non-modulated speech shaped icra1 noise (Dreschler et al, 2001) and in modulated speech shaped icra5-250 noise (modulations simulate one interfering talker, Wagener et al, 2006). The noise was either male or female frequency shaped regarding the speaker's gender of the applied sentence test. So-called Matrix sentences (syntactically fixed but semantically non predictable sentences, i.e., the Oldenburg Sentence test in German and its equivalent version in other languages) were used to determine binaural aspects of speech intelligibility like intelligibility level difference (ILD=benefit between SRTs of signal and noise presentation from same direction S_0N_0 and signal and noise presentation from different directions S_0N_{90}). Also, the binaural intelligibility level difference was determined (BILD= benefit between listening with only the contralateral ear to the noise source in S_0N_{90} and listening with both ears in this situation). All measurements were performed via free-field equalized Sennheiser HDA200 headphones. The binaural measurements were performed with virtual acoustics.

The sentence intelligibility measurements in noise were performed at a fixed noise presentation level of 65 dB SPL for normal-hearing listeners. For hearing-impaired listeners, an individual loudness level was chosen (according to a prior individual loudness scaling measurement included in the auditory profile: level yielding a loudness rating of 20 categorical units, i.e. between "soft" and "medium").

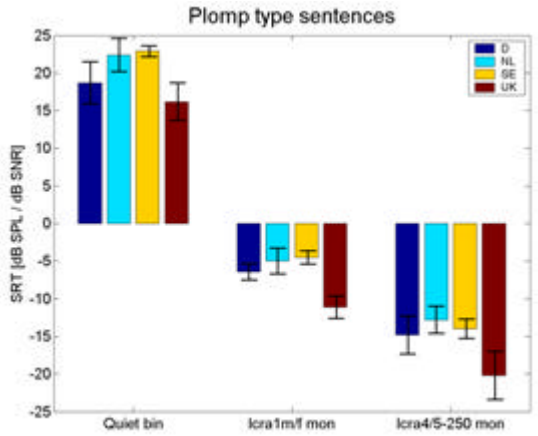


Fig. 1 shows the mean SRT results and the respective standard deviations of normal-hearing listeners who performed Plomp type sentence intelligibility tests. The binaural SRT data in quiet are shown in the left part of the figure (given in dB SPL), the monaural SRT data in non-modulated Icra noise are shown in the middle, and the monaural SRT data in modulated Icra noise are shown in the right part of the figure (both given in dB SNR). The country-specific data are indicated as follows: German: dark blue, Dutch: light blue, Swedish: yellow, British: red.

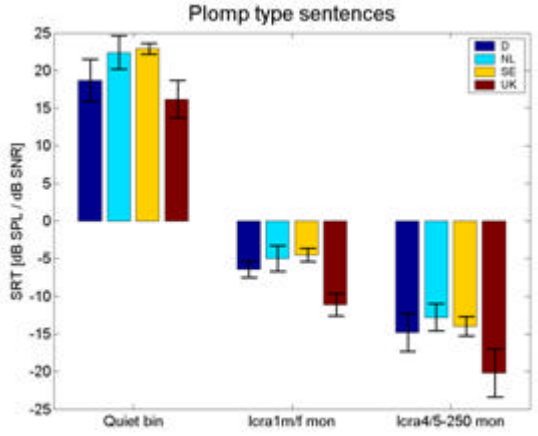


Fig. 1: HEARCOM project data. Mean country-specific normal-hearing SRT data and standard deviations of Plomp type sentences (German: dark blue, Dutch: light blue, Swedish: yellow, British: red). Three different conditions (binaural SRT in quiet, monaural SRT in non-modulated Icra noise, and monaural SRT in modulated Icra noise).

The different results across countries can partly be explained by the procedure differences across countries in applying Plomp type sentences. One difference is the scoring method: Both the Dutch and the Swedish test apply sentence scoring, the German test applies word scoring, and the British test applies key word scoring. Also the adaptive procedure of the Dutch test is different from the other tests: In the German, Swedish, and British tests, an adaptive procedure with decreasing step size was used that is described in Brand & Kollmeier 2002 by procedure A1. The Dutch test uses a 1up-1down adaptive procedure with fixed step size 2 dB. As a consequence of the different languages, the speakers differ across tests (Dutch and Swedish: female speaker, German and British: male speaker).

It seems that the scoring method mostly influences the results: When analyzing the German data according to sentence scoring (by applying the j factor concept by Boothroyd & Nittrouer 1988), the results are similar to the Dutch results.

Fig. **2** (left panel) shows the mean monaural SRT results and the respective standard deviations of normal-hearing listeners who performed Matrix sentence intelligibility tests. The monaural SRT data in quiet are shown in the left part of the figure (given in dB SPL), the monaural SRT data in non-modulated Icra noise are shown in the middle, and the monaural SRT data in modulated Icra noise are shown in the right part of the figure (both given in dB SNR). The country-specific data are indicated as follows: German: dark blue, Dutch: green, Swedish: red.

Fig. **2** (right panel) shows the mean binaural SRT results and the respective standard deviations of normal-hearing listeners who performed Matrix sentence intelligibility tests. The SRT data for S_0N_0 presentation are shown in the left part of the figure (given in dB SNR), the ILD data are shown in the middle, and the BILD data are shown in the right part of the figure. The country-specific data are indicated as follows: German: dark blue, Dutch: light blue, Swedish: yellow, British: red.

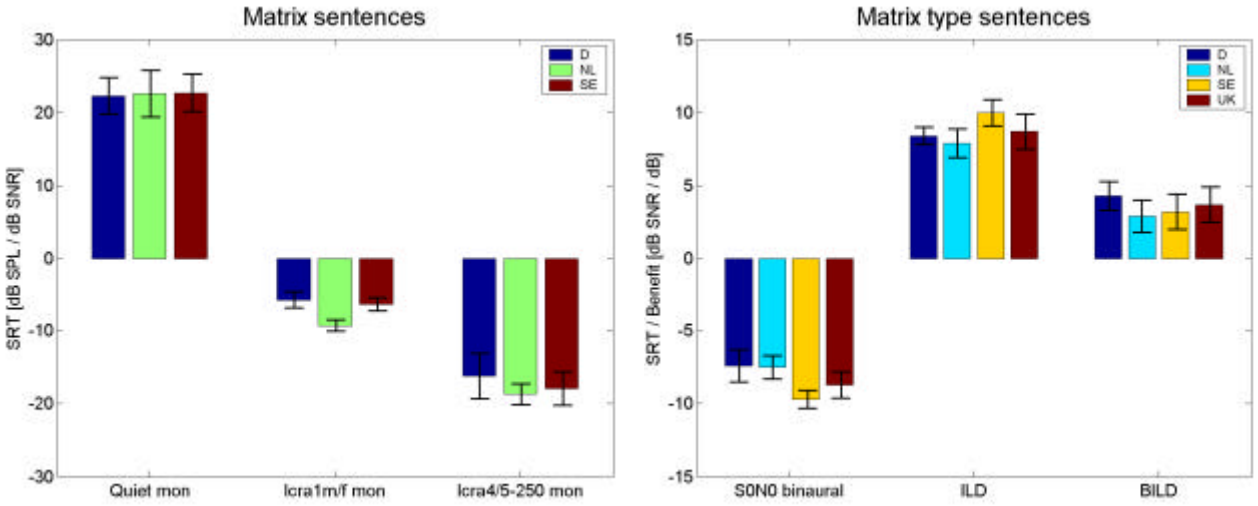


Fig. 2: HEARCOM project data. Left panel: Mean country-specific monaural normal-hearing SRT data ***and standard deviations of*** *Matrix sentences (German: dark blue, Dutch: green, Swedish: red). Three different conditions (SRT in quiet, SRT in non-modulated Icra noise, and SRT in modulated Icra noise).*
Right panel: Mean country-specific binaural normal-hearing SRT data and standard deviations of Matrix sentences (German: dark blue, Dutch: light blue, Swedish: yellow, British: red). Three different conditions (SRT in S_0N_0, ILD, and BILD).

As shown in the figures, the differences across countries are smaller compared to the Plomp type sentences data. This can be explained by the fact that for the Matrix sentence tests the same measurement procedure was used in all countries and the only difference apart from the language itself was the speaker of the test.

Taken together, the normal-hearing cross-validation data with two types of sentence intelligibility tests (Plomp type and Matrix sentences) indicates that some of the country-specific differences can be explained by procedure differences like word scoring versus sentence scoring. Since the procedure differences are less in the Matrix sentences, also the country-specific differences are smaller in these sentences. Hence, the aim of establishing compatible speech audiometric tests across Europe that produce the same test results for a given acoustical and audiological situation – irrespective of the subject´s language background - is getting closer.

MODELLING SPEECH RECOGNITION

The aim of modelling speech recognition in normal and hearing-impaired listeners is to obtain a comprehensive understanding on how hearing impairment affect the different processes involved in understand speech. A rough classification of these models can be given according to their intended level within the communication chain: acoustical layer – sensory layer and cognitive layer (Kollmeier et al., 2007).

The "classical" approach to model speech recognition under noise on the *acoustical layer* uses a spectral weighting of the long-term signal-to–noise-ratio and assumes that the total received information is the sum of the information transmitted in different frequency channels where the amount of information in each frequency channel is given by the respective signal-to-noise ratio (Fletcher and Galt, 1950). A vast literature exists on the articulation index, its further developments (Speech Transmission Index (STI, see Houtgast and Steeneken, 1985, and Speech Intelligibility Index (SII, ANSI, 1997) and its use for predicting speech intelligibility in hearing-impaired listeners. A modification of these procedures – adapted for the use with hearing-impaired listeners and the Oldenburg sentence intelligibility test – was tested by Brand & Kollmeier (2002). The time-independent SII reaches a reasonable well prediction accuracy for the SRT in quiet which, however, might not yield much additional information than the audiogram. The situation is different with suprathreshold speech tests in noise where the limits of the current SII models becomes clear. It is highly probable that the recruitment phenomenon and other suprathreshold processing deficits will be responsible for the observed deviations between empirical SRT data and audiogram- and external noise based SRT predictions. This finding calls for better modelling approaches (see Kollmeier et al., 2007). A more refined model for time-dependent speech recognition (that also models speech intelligibility in fluctuating noise is presented in the contribution by R. Meyer et al. (2007, this issue).

As a *clinical application* of the model evaluated above, the speech audiogram using the Freiburg monosyllabic speech intelligibility test in quiet can be predicted from the patient´s individual audiogram (see below, from Brand & Kollmeier, 2002). Any deviation between the predicted and the actually measured speech audiogram can either be attributed to inaccuracies of the tone audiogram or suprathreshold speech processing deficits that are not accounted for in the audiogram. Hence the difference between predicted and measured tone audiogram can reveal important information to the audiological clinician.

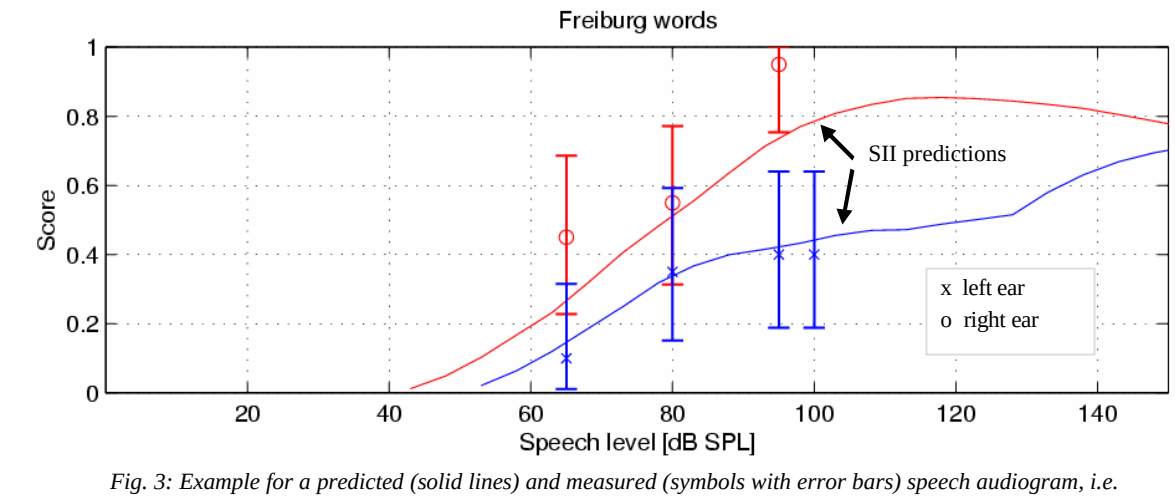


Fig. 3: Example for a predicted (solid lines) and measured (symbols with error bars) speech audiogram, i.e. Freiburg monosyllabic word test as a function of presentation level (from Brand&Kollmeier, 2002)

For the acoustical level in modelling speech reception, it is save to say that articulation index-based approaches (AI, STI, SII and modifications) appear to work well for threshold-dominated prediction tasks, i. e., for subjects with a mild to moderate hearing loss, for predictions in quiet and the average effect of continuous noise. However, the SRT in stationary noise is only partially predictable for hearing-impaired listeners since the variability among listeners seems to be highly influenced by non-acoustical factors (such as, e. g. sensory effects and cognitive effects). The short-term model extension evaluated by Meyer & Brand (this issue) for fluctuating noise seems to work well if spectro-temporal information of both signal and noise is accounted for within the approach.

A more refined approximation of modelling speech reception should take into account properties of the signal processing in the normal and hearing-impaired auditory system that reflect sensory processes in audition, i.e., the first steps of the physiological transformation of sound into neural activity and the neural representation of sound in the auditory system. Hence, the sensory layer can be thought of as in intermediate stage between the pure acoustical layer and a (perfectly operating) cognitive stage. This sensory layer can therefore be characterized by auditory models of "effective" signal processing that describe the neural transformation from the acoustical signal into some internal neural stage. We assume that the imperfections of the sensory processes involved in human auditory signal processing cause the main limitation in recognizing and discriminating speech sounds. These imperfections should be influenced by auditory signal processing properties that are relevant for human perception of sound, such as, e. g., bandwidth of the "effective" auditory critical bands, compression and adaptation in the auditory system, fine structure versus envelope cues and binaural interaction. The output of the sensory layer is fed into a cognitive layer that exploits the "internal representation" of speech signals in a perfect way by utilizing a-priori knowledge. This layer therefore can be modelled as an "optimal detector" which is assumed to include the whole "world knowledge" of the observer. In a more realistic approach, the cognitive layer can be approximated by a speech recognizer.

One approach to the sensory layer that aims at describing the binaural interaction and binaural noise

reduction in normal and hearing-impaired listeners during speech reception tasks was proposed by Beutelmann and Brand (2006). It is based on previous work by vom Hövel (1984) and a similar approach by Zurek (1990). A modification of the equalisation and cancellation (EC-) model of binaural interaction introduced by Durlach (1963) was used as a front end to an SII-type speech intelligibility prediction method.

A more direct way of addressing the sensory component in modelling speech reception was pursued by Holube and Kollmeier (1996) who used an "effective" signal processing model (Dau et al., 1996) of the normal and hearing-impaired listener as a front end to a standard Dynamic-Time Warp (DTW) speech recognizer. By determining the distances between a test utterance and training utterances "on a perceptual scale" (i.e., at the output of the "effective" signal processing model), the utterance with the least distance is taken as the recognized one. The "effective" auditory perception model emplyed (Dau *et al*,.1996) has been shown to model many different psychoacoustical experiments with different masking conditions as well as modulation detection tasks (Dau et al., 1997).

This approach of combining a perceptual signal processing model (representing the sensory layer) with a DTW speech recognizer (representing the cognitive layer) was further developed by Jürgens et al. (2007). They concluded that the prediction of speech reception appears to be quite successful if an "ideal detector" is assumed, i. e., a perfect world knowledge of the word to be expected. In such a configuration, an "effective model" of auditory signal processing seems to predict the availability of speech cues quite well. This is markedly different from the speech intelligibility index-based approach discussed above because speech discrimination is directly predicted from the speech signal without any prior normalisation of the intelligibility function for the respective speech material to be expected. On the other hand, the assumption of a perfect world knowledge (i. e., previous knowledge of the word to be expected as a kind of "Wizard of Oz" experiment) is only a very rough model of the cognitive system and does not take into account any individual differences in cognitive processing abilities. This calls for a better modelling of speech reception including the cognitive level.

Several approaches exist in the literature to examine the influence of inter-individual cognitive factors on obtained speech reception thresholds for normal and hearing-impaired listeners. The Linköping group (Larsby et al., 2005), for example, could demonstrate a high correlation between speech reception thresholds in noise and cognitive test outcomes, such as tests for assessing the individual working memory and maximum cognitive load by, e. g., performing a dual task memory span experiment. Based on their work, a cognitive test was included in the Hearcom auditory profile which is currently under consideration in a multicenter trial (Dreschler et al., 2007). However, in order to model the cognitive component in a more quantitative way and in order to connect this to models of the acoustical and sensory level (as given above), one will have to exchange the "ideal observer" concept outlined above with a "realistic observer" concept which includes a realistic pattern recognition model and various training procedures to account for priori knowledge in a scalable way. The best currently available pattern recognizers for speech stimuli are highly developed within the field of automatic speech recognition (ASR) so that a model of human speech recognition (HSR) based on elements of automatic speech recognition appears to be a meaningful approach. Since human listeners outperform ASR systems in almost all experiments (Lippmann, 1997), ASR may also profit from auditory feature extraction as proposed in (Kleinschmidt, 2003) or by using models of human word recognition (Scharenborg, 2005). In addition, a comparison between HSR and ASR should provide an appropriate basis for advancing such models of human speech recognition. Ideally, such a refined model should not only utilize bottom-up processes (such as, transforming the acoustical input signal into an internal representation which is recognized by a more or less ideal pattern recognizer), but should also incorporate aspects of top-down processing (such as, e. g. using learned patterns and a hypothesis-driven pattern recognition that may be influenced by the individual´s cognitive competence and working memory limitations) in order to model speech recognition in a more adequate way.

As a first step into this direction, a fair comparison of human and machine phoneme recognition was achieved by Meyer *et al*. (2007). They concluded that the total gap between human and automatic speech recognition in terms of SRT amounts to approx. 13 dB. This gap can be separated into a "sensory part", i. e. the gap between HSR for natural speech and for re-synthesized speech (i.e. a speech signal where only those speech cues are available for the human listeners that are provided to the computer in ASR) which amounts to 10 dB. This portion of the gap is due to non-ideal representation of the speech signal as the input pattern for the speech pattern recognition model. The remaining gap of 3 dB between HSR for resynthesized speech and ASR can be interpreted as the "cognitive" gap, i. e., the advantage of human

"top-down"-processing over the statistical-model-based pattern recognition in the ASR. Even though the HMM speech recognizer employed by Meyer et al. (2007) is only a poor model of the human cognitive system in recognising speech, this comparison still helps to quantitatively assess the effect of cognition for speech recognition in noise. Interestingly, the 3-dB gap is in the same order of magnitude as the difference between native and non-native listeners found in SRT measurements with sentences (for example Warzybok *et al*., 2007 this issue).

As a conclusion for the cognitive level, we can say that no promising "Ansatz" exists yet to adequately model the cognitive level in speech recognition. Hence, more work will have to be invested to achieve a satisfactory, complete model that will eventually also include individual differences in cognitive processing for the prediction of speech reception thresholds. However, the comparison between the perceptual, information-driven approach (bottom-up) and the world knowledge- and hypothesis-driven approach (top-down) pursued here appears to be a reasonable first step.

Acknowledgement

Supported by BMBF and research ministry of lower Saxony (Kompetenzzentrum HörTech) and the CEC (project Hearcom). We also thank all members of the Medical Physics group, the HörTech and the Hearcom consortium for their cooperation as well as the subjects for their patience.

References

ANSI (**1997**). "Methods for Calculation of the Speech Intelligibility Index", American National Standard S3.5-1997.

Beutelmann, R. and Brand, T. (**2006**). „Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners.", Journal of the Acoustical Society of America, 2006. **120**: p. 331-42.

Brand, T. and B. Kollmeier (**2002b**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests". Journal of the Acoustical Society of America, 2002. **111**(6): p. 2801-2810.

Brand, T. and Kollmeier, B. (**2002a**). „Vorhersage der Sprachverständlichkeit in Ruhe und im Störgeräusch aufgrund des Reintonaudiogramms", DGA 2002.

Dau, T. (**1997**). "Modeling auditory processing of amplitude modulation.", Journal of the Acoustical Society of America, 1997. **101**: p. 3061 (A).

Dau, T., D. Püschel, and A. Kohlrausch (**1996**). "A quantitative model of the "effective" signal processing in the auditory system: I. Model structure.", Journal of the Acoustical Society of America, 1996. **99**: p. 3615-3622.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation: I. Detection and masking with narrow band carrier," J. Acoust. Soc. Am., **102**, 2892-2905.

Demuynck, K., Garcia, O. and Dirk Van Compernolle (**2004**): "Synthesizing Speech from Speech Recognition Parameters", In Proc. ICSLP 2004, volume II, pages 945–948.

Dreschler, W.A. et al. (**2001**). "ICRA Noises: Artificial Noise Signals with Speech-like Spectral and Temporal Properties for Hearing Instrument Assessment", Audiology, 2001. **40**: p. 148–157.

Dreschler, W.A., van Esch, T.E.M. and Jeroen Sol, J. (**2007**). "Diagnosis of impaired speech perception by means of the Auditory Profile", this volume.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences", J. Acoust. Soc. Am. **35**(8), p. 1206–1218.

Fletcher, H., Galt (1950): "The Perception of Speech and Its Relation to Telephony" J Acoust Soc Am, 1950 **22**(2): p. 89-151.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a Gammatone filterbank". Acta acustica / Acustica, 2002. **88**(3): p. 433-442.

Holube, I. and B. Kollmeier (**1996**). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model", J Acoust Soc Am, 1996. **100**(3): p. 1703-16.

Houtgast, T., Steeneken, H.J.M. (1985) A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria" J. Acoust. Soc. Am. **77**(3), p. 1069-1077.

Jürgens, T., Brand, T. and Kollmeier, B. (**2007**). "Modelling the Human-Machine Gap in Speech Reception: Microscopic Speech Intelligibility Prediction for Normal-Hearing Subjects with an Auditory Model", In Proc. Interspeech 2007, Antwerpen.

Kleinschmidt, M. (**2003**). "Localized spectro-temporal features for automatic speech

recognition", Proc. Eurospeech/Interspeech, Geneva, 2003.

Kollmeier, B. (1990). "Meßmethodik, Modellierung und Verbesserung der Verständlichkeit von Sprache", Habilitationsschrift, Universität Göttingen

Kollmeier, B., Meyer, B., Jürgens, T., Beutelmann, R., Meyer R.M., Brand, T. (2007): Speech reception in noise: How much do we understand? In: ISAAR Symposium, DK-Helsingborg (in press)

Larsby, B., Hällgren, M., Lyxell, B., Arlinger, S. (2005) „Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects" Int. J. Audiol. 44(3), 131-143.

Lippmann, R.P. (**1997**). "Speech recognition by machines and humans", Speech Communication **22** (1) 1-15, 1997.

Meyer, B., Wächter, M., Brand, T. and Kollmeier, B. (**2007**): "Phoneme confusions in human and automatic speech recognition", In Proc. Interspeech 2007, Antwerpen, Belgium, 2007.

Meyer, R., Brand, T. (2007). " Prediction of speech intelligibility in fluctuating noise" in: EFAS/DGA 2007, Heidelberg (in press).

Pavlovic, C. V. (**1984**). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**(4), 1253–1258.

Payton, K.L., Braida, L.D. (1999):" A method to determine the speech transmission index from speech waveforms " J. Acoust. Soc. Am. **106**(6), 3637-3648.

Plomp, R. (1986). "A Signal-to-Noise Ratio Model for the Speech-Reception Threshold of the Hearing Impaired". J.Sp.Hear. Res. **29**, 146-154.

Rankovic, C. M. (1997). "Prediction of Speech Reception by Listeners With Sensorineural Hearing Loss" in: Jestaeadt, W. (ed.): Modeling Sensorineural Hearing Loss, Lawrence Earlbaum Associates, Mahwah, N.J.

Rhebergen, K. & Versfeld, N. (**2005**). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners." J Acoust Soc Am, **117**, 2181-92.

Sakoe, H. and S. Chiba (**1978**). "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Transactions on Acoustics, Speech, and Signal Processing, 1978. ASSP-26(1): p. 43-49.

Scharenborg, O. (**2005**). "Narrowing the gap between automatic and human word recognition", Ph.D. thesis, Radboud University Nijmegen, September 16th, 2005.

Sroka, J.J. and Braida, L.D. (**2005**). "Human and Machine Consonant Recognition", Speech Communication 45 (401-423), 2005.

Vom Hövel, H. (**1984**). "Zur Bedeutung der Übertragungseigenschaften des Außenohrs sowie des binauralen Hörsystems bei Gestörter Sprachübertragung," Dissertation, Fakultät für Elektrotechnik, RTWH Aachen.

Wagener K., Brand T., Kollmeier B. (**2006**). "The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners", Int J Audiol, **45**, 26-3.

Wagener K., Brand T., Kühnel V. und Kollmeier B (**1999**). „Entwicklung und Evaluation eines Satztestes für die deutsche Sprache I-III: Design, Optimierung und Evaluation des Oldenburger Satztestes", Zeitschrift für Audiologie **38**(1-3)

Wagener, K. and Brand, T. (**2005**). "Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: influence of measurement procedure and masking parameters", International Journal of Audiology, **44**(3), p. 144-156.

Wagener, K., Brand, T. and Kollmeier, B. (**2006**). „The role of silent intervals for sentence intelligibility in fluctuating noise in hearing–impaired listeners." International Journal of Audiology, **45**(1), p. 26-33.

Wagener, K.C., Brand, T. and Kollmeier, B (**2007**). "International cross-validation of sentence intelligibility tests", EFAS - Meeting 2007, Heidelberg (in press).

Warzybok, A., Wagener, K.C., Brand, T. (**2007**). "Intelligibility of German digit triplets for non-native German listeners", EFAS - Meeting 2007, Heidelberg (in press)

Wesker, T., Meyer, B., Wagener, K., Anemüller, J., Mertins, A. and Kollmeier, B. (**2005**). "Oldenburg Logatome Speech Corpus (OLLO) for Speech Recognition Experiments with Humans and Machines", In Proc. Interspeech 2005, Lisbon, Portugal, pp. 1273-1276.

Zurek, P. M. (**1990**). "Binaural advantages and directional effects in speech intelligibility," in Acoustical Factors Affecting Hearing Aid Performance, 2nd ed., edited by G. A. Studebaker and I. Hockberg (Allyn and Bacon, London), Chap. 15, pp. 255–276.