Davis, M.H. and Johnsrude, I.S. (**2007**). “Hearing speech sounds, Top-down influences on the interface between audition and speech perception” Hear Res **229**, 132-147.

Foo, C., Rudner, M.., Ronnberg, J. and Lunner, T. (**2007**). “Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity” J Am Acad Audiol, **18**, 618-31.

Glasberg, B. and Moore, B. C. J. (**1989**). “Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech” Scand Audiol, **Suppl 32**, 1-25.

Hagerman, B. (**1982**). “Sentences for testing speech intelligibility in noise” Scand Audiol, **11**, 79-87.

Houtgast, T. and Festen, J.M. (**2008**). “On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise” Int J Audiol, **47**, 287-95.

Humes, L. (**2005**). “Do 'auditory processing' tests measure auditory processing in the elderly?” Ear and Hearing, **26**, 109-19.

Hällgren, M., Larsby, B. and Arlinger, S. (**2006**). “A Swedish version of the Hearing In Noise Test (HINT) for measurement of speech recognition” Int J Audiol, **45**, 227-237.

Hällgren, M., Larsby, B., Lyxell, B. and Arlinger, S. (**2001**). ”Evaluation of a cognitive test battery in young and elderly normal-hearing and hearing-impaired subjects” J Am Acad Audiol, **12**, 357-370.

Larsby, B., Hällgren, M., Lyxell, B. and 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**, 131-143.

Larsby, B., Hällgren, M. and Lyxell, B. (**2008**). “The interference of different background noises on speech processing” Int J Audiol, **47, Suppl 2**, S83-90.

Lunner, T. (**2003**). “Cognitive function in relation to hearing aid use” Int J Audiol, 42: S49-S58.

Lunner, T. and Sundewall-Thorén, E. (**2007**). “Interactions between cognition, compression, and listening conditions: effects on speech-in-noise performance in a two-channel hearing aid” J Am Acad Audiol, **18**, 539–552.

Lyxell, B., Andersson, U., Borg, E. and, Ohlsson, I.-S. (**2003**). “Working-memory capacity and phonological processing in deafened adults and individuals with severe hearing impairment” Int J Audiol, **42, Suppl 1**, S86-S89.

Pichora-Fuller, M.K. (**2003**). “Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension” Int J Audiol, **42, Suppl 1**, S59-67.

Plomp, R. (**1986**). “A signal-to-noise ratio model for the speech-perception threshold of the hearing impaired” J Speech Hear Res, **29**, 146-154.

Side-effects of binaural tone vocoding on recognising target speech presented against spatially separated speech maskers

MARTIN R. ANDERSEN, MICHAEL S. KRISTENSEN, TOBIAS NEHER AND THOMAS LUNNER

Eriksholm Research Centre, Oticon A/S, DK-3070 Snekkersten, Denmark

Previous experiments have indicated that monaural Temporal Fine Structure (mTFS) information aids Speech Reception. In these experiments mTFS was either kept or substituted using a tone-vocoder. Results showed that hearing-impaired subjects were not able to utilise mTFS information to the same degree as normal-hearing subjects. A first step towards a more ecological experiment would be to exploit the tone-vocoder paradigm in a simulated spatial setup, and measure binaural TFS (bTFS) benefit. However, by the introduction of a binaural tone-vocoder, a concern arose that artificial ITD cues pointing to a direction determined by the phase difference between the carriers of the two channels, would be introduced in addition to the intended removal of the original Interaural Time Difference (ITD) by vocoding. This experiment investigated this concern, by measuring speech reception for target speech presented against spatially separated speech maskers. 21 young normal hearing, 10 elderly normal hearing and 11 elderly hearing impaired subjects were tested in a fixed spatial condition with either the artificial ITD pointing forward (0º azimuth) or ±50º.

INTRODUCTION

Recent studies (Hopkins *et al*. 2008; Lunner *et al.*, in press) have shown that temporal fine structure (TFS) is important for speech intelligibility in complex situations with concurrent masking talkers. In these studies the benefit from TFS was measured in terms of Speech Reception threshold (SRT) changes using speech-on-speech tests. The speech signals were presented monaurally over headphones, using tone vocoding to remove the original monaural TFS (mTFS) cues.

A first step towards a more ecological experiment would be to exploit the tone-vocoder paradigm in a simulated spatial setup, and to measure binaural TFS (bTFS). However, by the introduction of a binaural tone-vocoder, a concern arose that artificial ITD (AITD) cues pointing to a direction determined by the phase difference between the carriers of the two channels, would be introduced in addition to the intended removal of the original Interaural Time Difference (ITD) by vocoding.

In an earlier pilot study, this concern was tested on 8 normal-hearing (NH) subjects (Andersen *et al*., 2010), where a condition with the AITD pointing forward (0°

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

azimuth) was compared to a condition where the direction for each AITD in each subband in the filter-bank was randomized on horizontal axis (−90° to 90° azimuth). The result is shown in Fig. 1 (the two data points to the right), and the choice of direction of AITD turned out not to be significantly different from each other, but with a tendency that the randomized condition was more difficult. However we were still concerned that a condition with AITD pointing in another fixed direction than the target speech would be detrimental for speech intelligibility. Therefore another experiment was conducted.

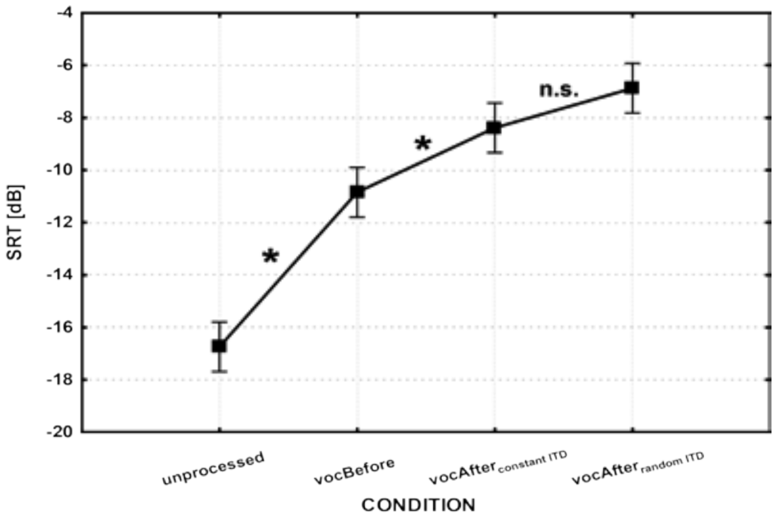

Fig. 1: Result from (Andersen and Kristensen, 2010). The two points to the right show the effect of the AITD direction. As seen, the difference is not significant (n.s.).

In this experiment a different approach was taken. Here the condition with the AITD pointing forward (0° azimuth) to the target speaker was compared to a condition with the AITD pointing towards one of the two competing talkers (±50° azimuth). The hypothesis was that if the AITD favored the source it is pointing towards, then the speech intelligibility should be worse for the condition pointing away from the target.

METHOD

Hypotheses

Two hypotheses were tested

- The AITD does affect speech perception, favoring the speaker it points towards.
- Normal-hearing subjects are affected more than hearing-impaired subjects.

Subjects

42 subjects were recruited, and divided into three groups.

- 21 young (age 21-40) normal hearing (NH) subjects (HL < 20 dB in the range from 125 Hz to 8 kHz): YNH
- 10 elderly (age 54-77) normal hearing (NH) subjects selected to match a given normal hearing range compared to their individual age (ISO, 2000): ENH
- 11 elderly (age 41-83) hearing impaired (HI) subjects with a symmetrical mild to moderated hearing loss. See the mean audiogram in Fig. 2.: EHI

Furthermore, the Cambridge formula hearing aid fitting prescription method (Moore and Glasberg, 1998) was applied to ensure that the speech signals were audible for the HI subjects.

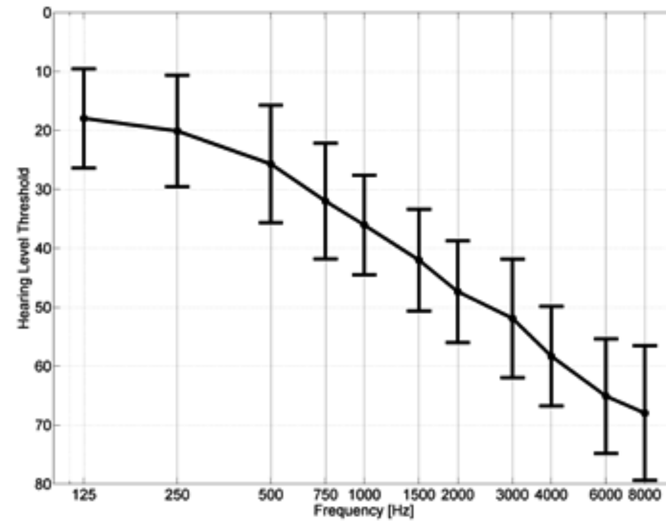

Fig. 2: Mean hearing Level of the elderly hearing impaired group, with the bars showing the standard deviation.

Speech Material

A special version of the Danish Dantale 2 corpus designed for spatial speech-on-speech testing (Behrens *et al.* 2007) was used for both target and masker talkers. The corpus was recorded with 5 different female speakers (only 3 of them were used). One of the speakers was consistently used as the target talker.

Simulated Spatial Setup

The simulated spatial condition is illustrated in Fig. 3. There is one target talker right in front of the test subject and two masker talkers at –50° and 50°, relative to the target talker. For the spatialisation, KEMAR HRTFs from the CIPIC database (Algazi *et al.* 2001) were used. The sources were only spatialised in the horizontal plane and not in the vertical plane.

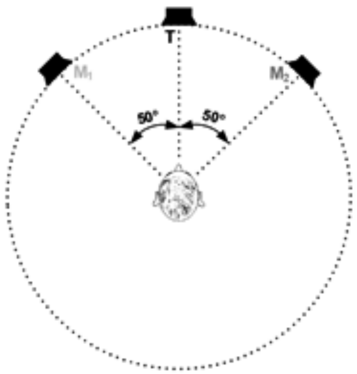

Fig. 3: Simulated spatial setup using HRTFs. Target is in front of the subject, and the two speech maskers are presented so one has an azimuth angle of +50 º and the other one of −50º.

Processing

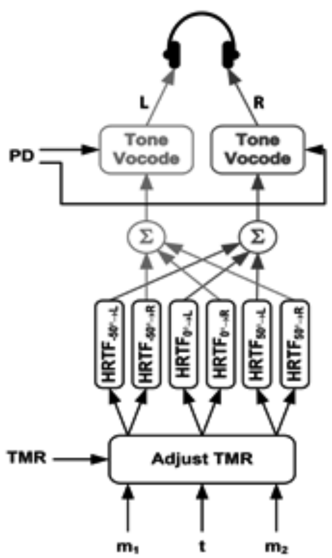

Fig. 4: Illustration of how sounds in the experiment are processed.

An overview of the processing is given in Fig. 4. First, the Target-to-Masker Ratio (TMR) was adjusted between the target *t* and the two competing talkers m_1 and m_2. Second, the individual sound sources were then convolved with the Head Related Impulse Reponses (HRIRs), which adds the binaural cues to each source individually. Third, the three right channels were added and the three left channels were added. Forth, the right and the left channels were vocoded individually, and a phase difference PD was inserted in between the two tone-carriers, to be able to control the overall ITD (Kuhn, 1977).

Condition

Two conditions was tested, see Table 1. In Fig. 5 an illustration showing how the ITD, Interaural Level Difference (ILD) and Interaural Envelope Difference (IED) are affected for the two conditions, is seen.

$AITD_{0°}$	*The instantaneous phase of the right and left side are all equal to 0 radians for all frequency bands; hence the ITDs give the perception that the sources are located in the median plane, while the perceived directions due to interaural level differences are determined by the HRTFs.*
$AITD_{50°}$	*Same as above but with the instantaneous phase difference between the right and left side are all equal to either 50° or − 50°.*

Table 1: Conditions in test

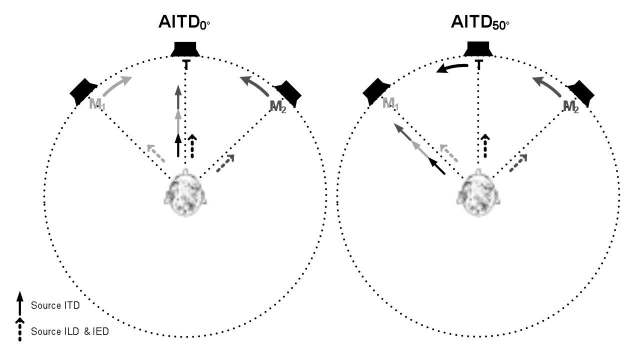

Fig. 5: An overview on how the ITD, and the ILD and IED are affected by the two processing conditions under test.

Tone-Vocoding

The tone vocoding processing paradigm was adopted from (Hopkins *et al.* 2008). The signals were filtered into 32 1-ERB_n wide frequency bands ranging from 100 to 10.000 Hz, and for each band the ENV was extracted and multiplied with a pure tone TFS with a frequency equal to the centre frequency of the band for the vocoded condition. Finally, the vocoded bands were filtered once again to limit the spreading in frequency and the unprocessed bands were delayed equivalently. In the vocoded conditions all bands were vocoded.

Training

Each test person received 45 min of training divided in two phases. First, each test person was trained in the speech-on-speech task procedure presented by headphones; tone-vocoding was not used here. Second, each test person was trained in the two research conditions of interest, given above.

RESULTS

In Fig. 6 the average SRT for both conditions and each group is shown. As seen, young NH subjects perform in a lower SNR than the elderly NH subjects, which again do perform in a lower SNR than the elderly HI subjects. Furthermore, there is a tendency that the $AITD_{0°}$ conditions giver lower SRT values than the $AITD_{50°}$ condition.

A repeated-measures ANOVA was performed, and it showed the effect of subject group to be significant [$F(2,2) = 26.23; p < 0.000$].

In order to test for significant differences between the two conditions, the $AITD_{50°}$ values was subtracted from the $AITD_{0°}$ values, as seen in Fig. 7. As seen, the mean for each group is similar and about −0.7 dB. A t-test for average means compared to 0 was conducted. The means for the three groups were collapsed into one single mean, and the result was $p = 0.012$, and therefore significantly different from *0*, meaning that $AITD_{0°}$ was significantly different from $AITD_{50°}$.

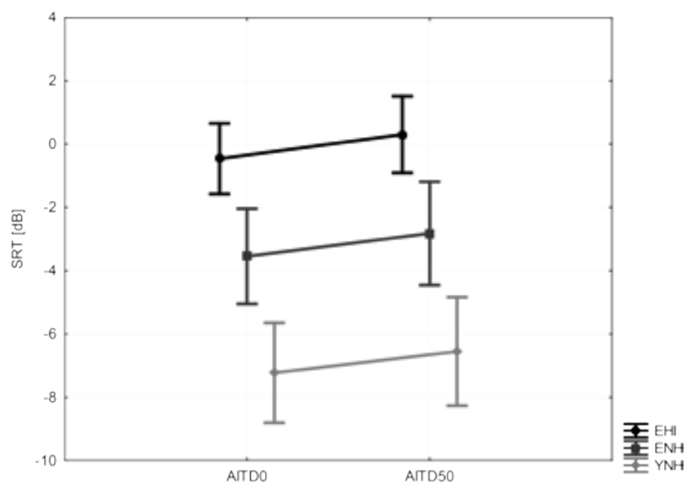

Fig. 6: Mean SRT for both conditions of each test subject group. Bars show the standard deviation

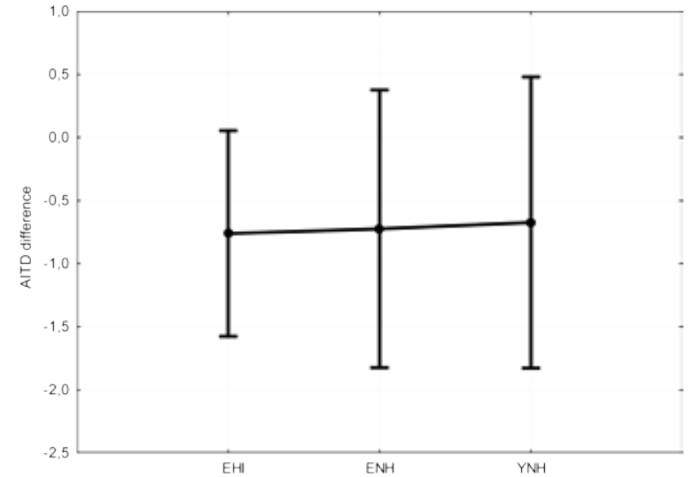

Fig. 7: AIDT Difference between the two conditions. Bars show the standard deviation

DISCUSSION

As expected a better SNR is needed, if the phase difference between the tone-carriers in the tone-vocoders point away from the target sound source. The overall difference measured in SRT was 0.7 dB between the two conditions where in one, the AITD was pointing in the direction of the target, and in the other it was pointing in the direction of one of the competing talkers. The difference was statistically significant. One could have thought that as the AITD only influences the ambiguous monaural TFS information, it would not affect speech perception, but the results show that it actually makes a difference.

What came as a surprise to us, was that the phase difference affected the speech intelligibility equally regardless of the group of test subjects; whether young or elderly, or whether normal-hearing or hearing-impaired. It was expected that if the phase difference favors one sound source over another, the effect would be larger for the YNH group, as they are more sensitive to, at least, monaural TFS information (Hopkins *et al.* 2008; Lunner *et al.* 2010).

So when measuring binaural tone-vocoding to assess binaural TFS benefit one has to consider the side-effects that the phase difference seems to favor the sound source it points towards regardless of test group, and at the same time it could potentially have a detrimental effect on the other sound sources. It can be discussed whether a correction value might be enough to compensate for the side-effects of tone-vocoding in a binaural setup. Further studies are needed for that.

CONCLUSION

Results show that the artificial ITD, which occurs when applying binaural tone Vocoding, affects speech perception. Furthermore, a better SNR is needed for the condition where the artificial ITD is pointing away from the target talker, compared to the condition where it points towards the target talker, for equal speech intelligibility.

Even though young normal-hearing subjects are more sensitive to monaural TFS than elderly hearing-impaired subjects [Hopkins *et al.* 2008], subjects with different age and different hearing status are equally affected by the artificial ITD.

Finally, using binaural tone-vocoding for measuring the benefit of binaural TFS produces side-effects that need to be considered.

REFERENCES

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF Database" IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pages 99-102.

Andersen, M. R., Kristensen, M. S., Neher, T. and Lunner T. (**2010**). "Effect of Binaural Tone Vocoding on Recognising Target Speech Presented Against Spatially Separated Speech Maskers" Poster at IHCON 2010.

Behrens, T., Neher, T., and Johannesson, R.B. (**2007**). "ERH-42-08-05 Evaluation of a Danish speech corpus for assessment of spatial unmasking" Poster at ISAAR.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech" J. Acoust. Soc. Am. **123** (2), 1140-1153.

ISO 7029:**2000**, "Acoustics – Statistical distribution of hearing thresholds as a function of age".

Kuhn, G.F. (**1977**). "Model for the interaural time differences in the azimuthal plane." J. Acoust. Soc. Am. **62** (1), 157-167.

Lunner, T., Hietkamp, R. K., Andersen, M. R., Hopkins, K., and Moore, B. C. J. (**in press**) "Effect of speech material on the benefit of temporal fine structure information in speech for normal-hearing and hearing-impaired subjects" Ear and Hearing.

Moore, B. C. J. and Glasberg, B. R. (**1998**). "Use of a loudness model for hearing aid fitting. I. Linear hearing aids" Br. J. Audioligy. **32**, 317-335.

Speech-inherent functional onomatopoeia as a basis for emotional analysis of phones

JENS BLAUERT

Institute of Communication Acoustics, Ruhr-Universität Bochum, D–44780 Bochum, Germany

Speech sounds (phones) originate in the context of a biological process where the articulators shape the vocal tract into a cascade of cavities and constrictions. This shaping requires muscle activity and these go along with feelings − which the speakers perceive as sitting inside their speech organs (phonetic feelings). The actual feelings depend on the specific form elements that are shaped. Formation of large open cavities associates with a feeling of emptiness, narrow, partly closed cavities are accompanied by the feeling of being pressed, constrictions feel stressed, short thick partitions depressed, long stretched-out partitions filled, satisfied. Each speech sounds designates (symbolically) a specific feeling that is potentially present in the talkers' perception while producing it. In other words, phones represent onomatopoeic acoustic descriptions of the talkers' phonetic feelings. It is considered whether this effect may be exploited, for instance, for word recognition, speaker-emotion recognition, sound design, speech synthesis and sound-quality assessment.

INTRODUCTION

The German philosopher and neurologist *Hans Lungwitz* (1881–1967) has, in 1933, published a monography on the *Psychobiology of Speech/Language* (Psychobiologie der Sprache) as part III of an eight-volume textbook of psychobiology (*Lehrbuch der Psychobiologie*). Due to the confusions of the second world war – *Lungwitz* was suspiciously observed by the Nazies and could hardly present his ideas in public – and since *Lungwitz*'s works are only available in (old fashioned) German, his *Psychobiology of Speech/Language* never really caught the attention of his peers. Recently, one of *Lungwitz*'s former students, *Reinhold Becker*, has newly edited the monography, added parts from other volumes of the textbook for better comprehensibility and "refurbished" the language for better readability (Lungwitz 1933, revised 2010). We take this opportunity to report on some fundamental ideas presented in the book, which – although almost eight decades old – may actually still have some relevance, for instance, for modern speech-and-language technology and for sound design. Further, we shall shortly discuss these ideas in the light of the current state of science and technology in the field.

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.