Speech recognition tested at fixed, positive signal-to-noise ratios using time compression: Methods and applications

Von der Carl von Ossietzky Universität Oldenburg
- Fakultät für Mathematik und Naturwissenschaften -
zur Erlangung des Grades und Titels eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

angenommene Dissertation

von

Anne Schlüter

geboren am 14.08.1978
in Münster

Gutachter:	Prof. Dr. Dr. Birger Kollmeier
Zweitgutachter(in):	Prof. Dr. Inga Holube
	Prof. dr. ir. Wouter Dreschler
	Prof. Dr. Tim Jürgens
Tag der Disputation:	21. Mai 2015

Abstract

Positive signal-to-noise ratios (SNRs) characterize listening situations that are most relevant for hearing-impaired listeners in daily life and need to be considered when evaluating hearing aid algorithms. Since this is difficult to assess using perception measurements such as the Acceptable Noise Level test or an adjustment method, a speech-in-noise test was developed in which the background noise is presented at fixed positive SNRs and the speech rate is adjusted to alter speech recognition. In an initial study, both a uniform and a non-uniform algorithm were used to compress the sentences of the German Oldenburg Sentence Test (OLSA) at different speech rates. For normal-hearing participants, measurements of time-compressed sentences in background noise at different SNRs confirmed decreasing recognition with increasing speech rate. For the application in speech-in-noise tests, subjective and objective measures indicated a clear advantage of the uniform algorithm in comparison to the non-uniform algorithm. Therefore, the uniform algorithm was selected for further research. In a second study, learning effects were analyzed using original and time-compressed speech in noise. Normal-hearing and hearing-impaired participants completed repeated measurements of the OLSA during successive sessions. The largest improvements in speech recognition thresholds were observed within the first measurements of the first session, indicating a rapid initial adaptation phase. Additional inter-session improvements indicated a longer phase of ongoing learning. Generally, observed effects were larger for time-compressed than for original speech, but the taking into account of learning effects is recommended for all OLSA versions. Based on these results, a procedure for adaptive adjustment of the speech rate (i.e. the time compression of the speech material) was developed in a third study. Two methods, which differed in adaptive procedure and step size, were compared for younger normal-hearing and older hearing-impaired participants. Analysis of the measurements with regard to list length and estimation strategy for thresholds resulted in a practical method for measuring the time compression for 50% recognition in background noise. In a fourth study, the procedure was used to shift the SNR for the evaluation of hearing aid algorithms towards fixed positive values. For hearing-impaired participants, recognition scores were measured with and without two different single-microphone noise reduction algorithms and using speech individually compressed in time at positive SNRs. Results showed a high potential of the method for discriminating across algorithms at positive SNRs. Even though objective measurements showed an improvement for both algorithms, a benefit for the participants was only obtained with the algorithm that used a priori knowledge of the noise. Taken together, the methods developed and validated here were shown to extend the available repertoire for testing speech in noise and for hearing aid benefit in humans to a range of ecologically relevant SNR values.

Zusammenfassung

Positive Signal-Rausch-Verhältnisse (SNR) charakterisieren Hörsituationen, die im täglichen Leben von Schwerhörigen oft vorkommen. Deshalb ist es wichtig, bei der Evaluation von Hörgeräte-Algorithmen diese Gegebenheiten besonders zu berücksichtigen. Perzeptive Messungen wie der Acceptable Noise Level Test oder eine Einregelungsmethode lösen dies Problem nicht. So wurde ein Sprachverständlichkeitstest entwickelt, bei dem das Hintergrundgeräusch bei festen positiven SNRs präsentiert und die Sprechrate eingestellt wird, um das Sprachverstehen zu variieren. In einer ersten Studie wurden ein uniformer und ein nicht-uniformer Algorithmus verwendet, um die Sätze des deutschen Oldenburger Satztests (OLSA) zu komprimieren und sie dann in verschiedenen Sprechraten darbieten zu können. Messungen zeitkomprimierter Sprache präsentiert mit Hintergrundgeräusch bestätigen, dass bei normalhörenden Probanden mit zunehmender Sprechrate die Verständlichkeit abnimmt. Subjektive und objektive Messungen zeigen für die Anwendung in einem Sprachverständlichkeitstest einen deutlichen Vorteil des uniformen Algorithmus im Vergleich zum nicht-uniformen Algorithmus. Für die weiteren Untersuchungen wurde daher der uniforme Algorithmus ausgewählt. In einer zweiten Studie wurde untersucht, ob bei originaler und bei zeitkomprimierter Sprache im Störgeräusch Lerneffekte festzustellen sind. Dazu führten normalhörende und schwerhörige Probanden Wiederholungsmessungen des OLSA in mehreren Sitzungen durch. Die größten Verbesserungen der Sprachverständlichkeitsschwellen wurden in den ersten Messungen der ersten Sitzung beobachtet. Sie weisen auf eine schnelle frühe Anpassungsphase hin. Zusätzliche zwischen den Sitzungen auftretende Verbesserungen sind ein Zeichen für längeres anhaltendes Lernen. Generell sind die beobachteten Effekte für zeitkomprimierte Sprache größer als für originale Sprache. Die Berücksichtigung von Lerneffekten wird jedoch für alle OLSA-Versionen empfohlen. Basierend auf diesen Ergebnissen wurde in einer dritten Studie ein Verfahren für die adaptive Einregelung der Sprechrate (d.h. der Zeitkompression des Sprachmaterials) entwickelt. Zwei Methoden, die sich im adaptiven Verfahren und der Schrittgröße unterschieden, wurden mit jungen normalhörenden und älteren schwerhörigen Probanden durchgeführt und verglichen. Analysen der Messungen berücksichtigten die Listenlänge und die Schwellenschätzung und ergaben eine praktikable Methode für die Messung der Zeitkompression, die für das Verstehen von 50% der Sprache notwendig ist. In einer vierten Studie wurden diese Methoden verwendet, um für die Evaluation von Hörgeräte-Algorithmen den SNR zu festen positiven Werten zu verschieben. Mit schwerhörigen Probanden wurde die Verständlichkeit ohne und mit zwei unterschiedlichen einkanaligen Störgeräuschreduktionsalgorithmen gemessen unter Verwendung von individuell zeitkomprimierter Sprache bei festen positiven SNR. Die Ergebnisse zeigen ein hohes Potential dieser Methode, zwischen den Algorithmen bei positiven SNR zu unterscheiden. Obwohl objektive Untersuchungen eine Verbesserung durch beide Algorithmen zeigten, wurde ein Nutzen für die Probanden nur bei dem Algorithmus erreicht, der a-priori-Wissen des Hintergrundgeräusches nutzte. Alle Ergebnisse zeigen, dass die hier entwickelte und validierte Methode das Repertoire der Prüfmethoden von Sprache im Hintergrundgeräusch und des Hörgerätenutzens für den Menschen auf einen ökologisch relevanten SNR-Bereich erweitert.

Publications associated with this thesis

Peer-reviewed articles

Fredelake, S., Holube, I., Schlueter, A., Hansen, M. (2012) "Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms", International Journal of Audiology, 51(4), 299-308.

Schlueter, A., Lemke, U., Kollmeier, B., Holube, I. (2014) "Intelligibility of time-compressed speech: The effect of uniform versus non-uniform time-compression algorithms", Journal of the Acoustical Society of America, 135, 1541-1555.

Schlüter, A., Aderhold, J., Koifman, S., Krüger, M., Nüsse, T., Lemke, U., and Holube, I. (2014) "Evaluation eines Einregelungsverfahrens zur Bestimmung des Nutzens einkanaliger Algorithmen zur Störgeräuschreduktion - Evaluation of an adjustment method to determine the benefit of single-microphone noise reduction algorithms", Zeitschrift für Audiologie, 53(2), 50-58.

Schlueter, A., Lemke, U., Kollmeier, B., Holube, I. (2014) "Normal and time-compressed speech: How does learning affect speech recognition thresholds in noise?", International Journal of Audiology, submitted.

Schlueter, A., Brand, T., Lemke, U., Nitzschner, S., Kollmeier, B., Holube, I. (2014) "Speech perception at positive signal-to-noise ratios using adaptive adjustment of time compression", Journal of the Acoustical Society of America, submitted.

Non-peer-reviewed articles and conference presentations

Kallinger, M., Ochsenfeld, H., Schlüter, A. (2009) "A Novel Listening Test-Based Measure of Intelligibility Enhancement", 127th Convention of the Audio Engineering Society, New York, USA.

Schlüter, A., Holube, I. und Lemke, U. (2010) "Untersuchung eines subjektiven SNR-Vergleichs zur Bestimmung des Nutzens einkanaliger Störgeräuschreduktionen", 13. Jahrestagung der Deutschen Gesellschaft für Audiologie (DGA), Frankfurt, Germany.

Schlüter, A., Holube, I. (2010) "Perzeptive Maße zur Evaluation von Hörgeräteversorgungen bei Sprache im Störgeräusch", Zeitschrift für Audiologie, 49(3), 103-111.

Schlüter, A., Holube, I., Lemke, U. (2011) "Sprachverstehen beschleunigter Sprache im Störgeräusch", 14. Jahrestagung der Deutschen Gesellschaft für Audiologie (DGA), Jena, Germany.

Schlueter, A., Holube, I., Lemke, U. (2011) “Speech intelligibility as a function of time compression, age, word position, and signal-to-noise ratio”, Speech Perception and Auditory Disorders - 3rd International Symposium on Auditory and Audiological Research (ISAAR), Nyborg, Denmark, pages 191-198.

Schlüter, A., Holube, I., Lemke, U. (2012) “Trainingseffekte bei normaler und schneller Sprache”, 15. Jahrestagung der Deutschen Gesellschaft für Audiologie (DGA), Erlangen, Germany.

Lemke, U., Schlüter, A. und Holube, I. (2012) “Sprachverständlichkeit und Hörqualität in positivem SNR-Bereich: Neue Ansätze zur Evaluation von Hörtechnologien”, Abstractband der 43. Jahrestagung der Deutschen Gesellschaft für Medizinische Physik, Jena, S. 146.

Schlueter, A., Holube, I. and Lemke, U. (2012) “Training effects in a German speech-in-noise test with original and fast speech”, IHCON 2012, International Hearing Aid Research Conference, Tahoe City, California, USA.

Schlüter, A., Holube, I., Lemke, U. (2013) “Verfahren zur Bestimmung der Zeitkompressionsschwelle von Sprache im Störgeräusch”, 16. Jahrestagung der Deutschen Gesellschaft für Audiologie (DGA), Rostock, Germany.

Schlueter, A., Holube, I. and Lemke, U. (2013) “A speech intelligibility test assessing time-compression thresholds for speech in noise”, Second International Conference on Cognitive Hearing Science for Communication, CHSCOM, Linköping, Sweden.

Schlüter, A., Holube, I., Lemke, U., Herzog, D. (2014) “Verständlichkeitsschwellen im Göttinger und Oldenburger Satztest bei Variation der Sprachgeschwindigkeit”, 17. Jahrestagung der Deutschen Gesellschaft für Audiologie (DGA), Oldenburg, Germany.

Schlueter, A., Holube, I. and Lemke, U. (2014) “Intelligibility improvement of noise reduction algorithms: Does the presentation of positive signal-to-noise ratios help?”, IHCON 2014, International Hearing Aid Research Conference, Tahoe City, California, USA.

Contents

List of Figures

List of Tables

Abbreviations

ANL	Acceptable Noise Level
BNL	Background Noise Level
GÖSA	Göttingen sentence test
HI	Hearing-impaired participants
HL	Hearing Level
MCL	Most Comfortable Level
NH	Normal-hearing participants
OHI	Older hearing-impaired participants
OLSA	Oldenburg sentence test
ONH	Older normal-hearing participants
PSOLA	Pitch Synchronous Overlap-Add technique
RHT	Reversed Hierarchy Theory
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
SRT	Speech Recognition Threshold
STI	Speech Transmission Index
TC	Time-compressed
TCT	Time Compression Threshold
YNH	Younger normal-hearing participants

1

Introduction – Objective and outline

Everyday verbal communication often occurs in noisy environments. Among other physical characteristics such as, e.g., the spectrum or overall sound level, listening situations can be described by the signal-to-noise ratio (SNR). The SNR is the difference between the level of the speech signal and that of the background noise. Analysis of listening situations containing speech frequently show positive SNRs (Olsen, 1998; Smeds et al., 2015). That is, in daily communication, the speech signal is often at a higher level than the background noise.

Speech perception has an important role in daily communication and is frequently affected by background noise and by a variety of other factors, for instance the acoustics of the hearing situation and the hearing ability of the listener. For studies of speech perception in various hearing situations, different aspects are usually analyzed, such as quality, comprehension and recognition of speech. In this context, Kondo (2012) describes speech quality as being based on the subjective evaluation of speech samples. Listeners were, for example, asked to rate the overall quality on a one-dimensional scale or to use a multidimensional scale and rate aspects related to the quality such as roughness or dullness. In contrast, comprehension measurements test the ability to understand the message of utterances or words, and participants have to answer questions related to the content of the samples presented. For example, the participants could perform a lexical decision task in which they have to decide whether the signal presented was a real word or a semantically correct sentence (e.g. Banai and Lavner, 2012). Further examples ask for details in the samples presented such as "who was performing an action?" (e.g., Wingfield et al., 2006; Carroll and Ruigendijk, 2013). In comprehension tasks, reaction times or the number of correct answers are usually compared between different hearing situations. Moreover, speech recognition is often analyzed to better understand speech perception. To test speech recognition, especially in background noise, speech-in-noise tests are a common approach in many studies. Examples of those tests are digit triplet tests (Smits et al., 2004; Zokoll et al., 2012), sentence tests presenting short meaningful sentences (Kollmeier and Wesselkamp, 1997; Nilsson et al., 1994; Plomp and Mimpen, 1979), or the matrix test using limited speech material of the same structure (e.g., Wagener et al., 1999c; Ozimek et al., 2010; Hochmuth et al., 2012). During these tests, participants listen to the speech and have to repeat what they perceived. To compare hearing situations, the number of correct repetitions is

counted. However, depending on the correctly repeated words, the speech level (or the background noise level) is often adaptively adjusted towards a certain threshold, e.g., the SNR for which the participant can understand 50% (or, e.g., 80%) of the words. This threshold is called speech recognition threshold (SRT). For German speech-in-noise tests, e.g., the Oldenburg sentence test, the mean SRT is at about -7 dB SNR for normal-hearing participants (Wagener et al., 1999a). Even hearing-impaired participants with hearing thresholds ranging between 10 and 100 dB HL reach a mean SRT of about –3 dB SNR (Wagener and Brand, 2005).

When compared to the above-mentioned positive SNR of everyday verbal communication, SRT results of speech-in-noise tests do not present ecologically relevant hearing situations. However, these tests have advantages for the analysis of speech perception, since they usually show a steep discrimination function (e.g., Wagener et al., 1999a) and yield a high reliability (e.g., Wagener et al., 1999a; Wagener and Brand, 2005). Furthermore, Naylor (2010) described the tests' results as convenient for statistical analysis and carrying out the test is relatively fast. As a result, speech-in-noise tests are applied in clinical studies for diagnostical purposes and also in scientific studies of speech recognition. In scientific studies, these tests are used, for example, to analyze effects of hearing situations (such as with different background noises, with different speech material or with spatial separation of sound sources). Speech-in-noise tests are also of interest for the application in hearing aid development regarding the analysis of speech recognition as depending on hearing aids and their algorithms.

Hearing aid developers are aware of the importance of speech communication and the difficulties occurring for communication in noise. These difficulties arise because hearing impairment is characterized by a variety of deficits such as decreased audibility, decreased dynamic range, decreased frequency resolution, decreased temporal resolution and their diverse interactions (Dillon, 2012). Apart from providing, e.g., frequency-dependent gain and dynamic range compression, hearing aid developers implement different algorithms into the hearing aids to improve recognition, ease of communication and listening comfort. They make use of algorithms such as beamforming to extract sounds (e.g., a single speaker in a noisy background) but also use knowledge about binaural processing to facilitate communication of hearing-impaired listeners (for an overview of digital signal processing in hearing aids see Holube et al., 2014a). Optimally, hearing aids and their algorithms are able to adapt to the everyday life of the user, which is full of varying hearing situations. In addition, many algorithms show processing dependent on the SNR of the current hearing situation. Prominent examples are single-microphone noise reduction algorithms, which are also implemented to improve noisy communication situations. In the context of this thesis, they were used to illustrate the SNR's role and the limitations of some "typical" hearing aid algorithms.

Single-microphone noise reduction algorithms receive only the mixed speech and noise signal as their input and have to estimate the background noise out of the mixed signal. Using a gain rule based on certain statistical assumptions, they suppress those parts of the input signal that contain most of the noise. As a result, even though the time- and frequency-resolved SNR might not be changed by this pure manipulation in local intensity, an improvement in the overall SNR will result in the ideal case of a perfect SNR estimate. However, the separation of speech and background noise is difficult, especially if the background noise has a higher level

than the speech. Therefore, for some single-microphone noise reduction algorithms, overall SNR improvements will mainly result at positive SNRs and the algorithms introduce increasing distortions with decreasing SNR (as described by, e.g., Marzinzik, 2000; Fredelake et al., 2012; Marzinzik and Kollmeier, 2002; Neher et al., 2014a; Brons et al., 2014). Frequently, speech recognition tests do not show improvement after single-microphone noise reduction processing (e.g., Brons et al., 2013, 2014; Hu and Loizou, 2007; Neher et al., 2014b). In general, the application of speech-in-noise tests for the analysis of hearing aids and their algorithms creates problems. For the evaluation of hearing aids and their algorithms, Naylor (2010) warned of the difficulties that may be produced by variable SNR resulting from SRT measurements. In addition, low or even negative SRTs (meaning a low or even negative SNR at the input of an algorithm) are challenging for some single-microphone noise reduction algorithms (e.g., Luts et al., 2010; Fredelake et al., 2012; Brons et al., 2014). Finally, the comparison of normal-hearing and hearing-impaired listeners' results are difficult if they are obtained at individual SRT with hearing aid algorithms that are usually SNR-dependent in their processing.

The combination of all three factors – the need for considering everyday communication situations, speech-in-noise tests, and SNR-dependent processing of some hearing aid algorithms – points out the importance of ecologically relevant positive SNRs in studies using speech in background noise. Generally, this makes sense not only for diagnostic purposes but also for clinical and scientific research with objectives such as analyzing speech recognition, evaluating hearing aids and hearing aid algorithms, as well as for documenting the rehabilitative success of hearing aids.

The problems described above imply the following requirement for a speech test in background noise: Speech and background noise need to be presented at fixed positive SNRs. The fixed SNR can be used to consider the SNR dependent processing of the hearing aid algorithms and furthermore allows for the comparison of the hearing ability of normal-hearing and hearing-impaired listeners. In addition, the positive SNR permits evaluating everyday communication situations and, e.g., maximal SNR improvement of single-microphone noise reductions.

The acceptable noise level test is a procedure that partly meets these requirements (for an overview see Schlüter and Holube, 2010; Olsen and Brännström, 2014). This procedure answers the question as to how much background noise listeners tolerate while following speech (Nabelek et al., 1991). In this test, participants have to adjust maximum ('too loud'), minimum ('too soft') and comfortable levels (Most Comfortable Level, MCL) of a single speech signal. Thereafter, noise is presented, together with speech at MCL, and the subjects modify the noise in the same way as the speech to a maximum ('loud, until you cannot understand the speech'), minimum ('soft, until the speech is very clear and you can follow the story easily') and acceptable level ('to the most noise that you would be willing to "put-up-with" and still follow the sentences for a long period of time without becoming tense or tired'). The acceptable level is noted as Background Noise Level (BNL). The MCL and BNL are used to calculate a SNR, called Acceptable Noise Level (ANL). This procedure was characterized by Nabelek (2005) as a simple method that tends to achieve positive ANL values. Positive ANL values are reached if the listeners adjust the noise to a smaller level compared to the level of speech and implies little acceptance of background noise. Measurements of the ANL showed large differences for

individual listeners. Nabelek (2005) described ANL values ranging from -2 to 29 dB SNR for hearing-impaired listeners and from -2 to 38 dB SNR for normal-hearing listeners. The mean interval for both groups was at positive SNRs of 10 to 11 dB. The large interindividual variance of the results shows that the subjectively adjusted levels for the speech and noise signals are prone to subjective, individual criteria (Olsen and Brännström, 2014), which are presumably not transferable between listeners. Furthermore, the accuracy and precision of the test was controversially discussed (Olsen and Brännström, 2014). Differences of the ANL between participants were also measured by Schlüter (2007), who investigated single-microphone noise reduction algorithms. For realistic algorithms, only normal-hearing participants obtained an improvement of the ANL compared to situations without processing. For the same situations, hearing-impaired participants showed no improvement. Schlüter (2007) explained this result with the high interindividual variability in the ANL and the related varying SNR improvement.

Wittkop (2001) and Wittkop et al. (1997) developed a subjective comparison method to be performed at fixed positive SNRs. The listeners' task is to adjust the speech recognition of a test signal (mixed speech and noise) comparable to a reference signal that is processed, e.g., with a noise reduction algorithm. Therefore, the listener modifies the SNR of the test signal, and the SNR improvement due to processing can be measured. In contrast to the ANL test, this procedure presents an anchor (reference signal) and thus reduces the variability of the participants' responses. However, participants perform the comparison using individual criteria to account for the processing differences perceived between the reference and the test signal, resulting in interindividual differences. Schlüter et al. (2014a, see Appendix A) conducted this procedure at three different SNRs (individual SRT, 5 dB and mean of both values). Additionally, they used the shadow-filtering method, which was proposed by Kallinger et al. (2009) in combination with a speech-in-noise test. For shadow filtering, speech and noise signals were mixed at 5 dB SNR and the resulting signal was processed by the noise reduction algorithm. Coefficients calculated by the algorithm for the mixed signal were also used to filter the two signal components of the mixed signal separately. Then, the filtered signal components were mixed at the SNR of the individual SRT and presented to the participant. As a result, participants listened to reference signals at negative SNRs, although the noise reduction was applied at positive SNRs. Results showed that realistic noise reduction did not achieve any improvement at different SNRs without shadow filtering. Only the shadow-filtering method obtained significant improvements of the SNR due to the noise reduction algorithm. However, this subjective comparison method using shadow filtering cannot be applied to hearing aids, due to the essential a priori knowledge about the unmixed parts of the signal. Schlüter et al. (2014a) also analyzed differences of the variances to show the difficulty of the adjustment method at different SNRs. The variance increased with SNR presented in the reference signal. This means that participants performed the adjustment more easily for reference signals presented at an SNR of the individual SRT, as compared to presentations at a positive SNR of 5 dB. This indicates that in the adjustment method, participants were uncertain about the application of their criteria, especially for highly intelligible speech (at positive SNRs).

The procedures described so far lack either the possibility of presenting fixed positive SNRs and/or participants show difficulties in building reliable, comparable, subjective criteria for

their assessment, due to the procedures' tasks. An investigation of a speech test with fast speech in quiet (Versfeld and Dreschler, 2002) demonstrated a way to overcome the problems and led to the idea of this thesis: to modify a speech-in-noise test and adaptively adjust the rate of speech presented in background noise at fixed SNRs. It is expected that this approach offers participants a simple task (repetition of the speech they perceived) and will also permit variation of the recognition rate. Hence, a threshold measurement becomes feasible, just like the adjustment of the SNR in original speech-in-noise tests. To develop this procedure, different aspects were investigated.

In order to increase the speech rate of existing recordings, the signals are compressed in time while the pitch of the speech is preserved. In that process, time-compression algorithms mimicking natural fast speech either vary the compression rate according to the phonetical content of the speech (non-uniform compression) or compress the signal uniformly. To select an algorithm for a speech-in-noise test with time-compressed speech, these time-compression algorithms have to be compared and the comparison should consider the alterations of perceptually relevant cues for understanding speech in noise. Selection criteria therefore include, e.g., deviations from the targeted compression across time, changes of phoneme durations, long-term spectra, and modulation spectra. Additionally, recognition of speech compressed in time with different algorithms should be compared. Hence, the speech has to be presented in background noise and the relation of time compression and SNR to recognition has to be investigated. Also, studies of recognition scores are expected to show the time compression necessary to reach 50% recognition at positive SNRs. Furthermore, results can indicate factors influencing the reliability of a speech-in-noise test with time-compressed speech. All these aspects are considered in **Chapter 2** and include the objective comparison, as well as the measurement of speech recognition scores of speech compressed in time with a uniform and a non-uniform algorithm at different SNRs.

Besides the selection of a time-compression algorithm and the relation of time compression and recognition, learning effects for time-compressed speech have to be considered for studying a speech-in-noise test. Recognition for time-compressed speech improves, the longer the participant listens to this kind of speech processing (e.g., Gordon-Salant and Friedman, 2011). This adaptation or learning effect comes in addition to the well-known learning effect of speech-in-noise tests that use a matrix structure of the sentences (Wagener et al., 1999a). Hence, learning effects of time-compressed matrix sentences are studied and permit comparison to original sentences. **Chapter 3** describes these studies, discusses recommendations for training procedures in speech audiometry and relates the results to a theoretical model for learning.

For the investigation of a speech-in-noise test presenting fixed positive SNRs, two different procedures for the adaptive presentation of time compression in a speech-in-noise test are compared in **Chapter 4**. These procedures are based on adaptive procedures introduced by Versfeld and Dreschler (2002) as well as Brand and Kollmeier (2002). Results are discussed in relation to two factors: the stimulus placement procedure and the strategy for estimating the resulting threshold of the test. This leads to the selection of an adaptive procedure. The reliability of this approach is studied for groups of participants that differ in age and in hearing loss. In consideration of learning effects, training protocols are recommended.

Then, the speech-in-noise test developed is applied to the study of hearing aid algorithms that show SNR-dependent processing. The procedure is used to individually adjust the time-compressed speech and allows for the individual adaptation of the test difficulty for recognition measurements with and without noise reduction algorithms. **Chapter 5** shows the results and clarifies whether ceiling effects of recognition scores are avoided with time-compressed speech even though speech is presented at positive fixed SNRs. Furthermore, the SNRs presented allow for beneficial processing, which is verified by objective SNR improvement. Recognition results for two different single-microphone noise reduction algorithms are measured and discussed.

In **Chapter 6**, general results of studies are summarized with regard to their discussions and conclusions. Furthermore, findings of previous studies are related to perspectives for future work.

2

Intelligibility of time-compressed speech: The effect of uniform versus non-uniform time-compression algorithms

For assessing hearing aid algorithms, a method is sought to shift the threshold of a speech-in-noise test to (mostly positive) signal-to-noise ratios (SNRs) that allow discrimination across algorithmic settings and are most relevant for hearing-impaired listeners in daily life. Hence, time-compressed speech with higher speech rates was evaluated to parametrically increase the difficulty of the test while preserving most of the relevant acoustical speech cues. A uniform and a non-uniform algorithm were used to compress the sentences of the German Oldenburg Sentence Test at different speech rates. In comparison, the non-uniform algorithm exhibited greater deviations from the targeted time compression, as well as greater changes of the phoneme duration, spectra, and modulation spectra. Speech intelligibility for fast Oldenburg sentences in background noise at different SNRs was determined with 48 normal-hearing listeners. The results confirmed decreasing intelligibility with increasing speech rate. Speech had to be compressed to more than 30% of its original length to reach 50% intelligibility at positive SNRs. Characteristics influencing the discrimination ability of the test for assessing effective SNR changes were investigated. Subjective and objective measures indicated a clear advantage of the uniform algorithm in comparison to the non-uniform algorithm for the application in speech-in-noise tests.

Adapted from:
Schlueter, A., Lemke, U., Kollmeier, B., Holube, I. (2014) "Intelligibility of time-compressed speech: The effect of uniform versus non-uniform time-compression algorithms", J. Acoust. Soc. Am., 135, 1541-1555.

2.1 Introduction

Speech-in-noise tests like the Hearing in Noise Test (Nilsson et al., 1994) and the German Oldenburg or Göttingen Sentence Tests (Wagener et al., 1999a; Kollmeier and Wesselkamp, 1997) present quite simple hearing situations with speech in a stationary background noise. They can be used to assess the influence of speech level or speech-to-noise ratio on intelligibility. The signal-to-noise ratio (SNR), where the listener understands 50% of the speech, is measured and documented as the speech recognition threshold (SRT). Those speech-in-noise tests typically yield negative SRTs even for hearing-impaired listeners and exhibit a steep slope of the discrimination function (between approximately 10%/dB and 20%/dB) such that the SNR range where the test is sensitive to small changes of the speech level is limited. Wagener and Brand (2005), for example, found a mean SRT of about -3 dB SNR for a group of listeners with pure-tone hearing thresholds ranging from 10 dB hearing level (HL) to more than 100 dB HL. In combination with the slope of 18%/dB for the discrimination function, this limits the operational test range (i.e., the SNR range where such a test discriminates best between two conditions with different effective SNRs) between about -9 dB (i.e., approximately 20% intelligibility for normal-hearing listeners with a mean SRT of about -7 dB; Wagener et al., 1999a) and about -1 dB (i.e., approximately 80% intelligibility for hearing-impaired listeners).

For the evaluation of hearing aid algorithms, however, speech-in-noise tests are required that are able to discriminate across settings of the respective algorithms when operating at positive SNRs for several reasons: First, several hearing aid algorithms are most effective at positive SNRs such as single-microphone noise reduction algorithms (Fredelake et al., 2012). Second, most conversations in everyday life take place in noisy environments at positive SNRs (Olsen, 1998; Smeds et al., 2012), i.e., speech levels are generally higher than the noise level. In addition, hearing-impaired listeners require a higher SNR for the same intelligibility than normal-hearing listeners so that testing hearing aid algorithms with normal-hearing listeners should be done at higher SNR values than represented by their individual SRTs. For these reasons, methods are sought to shift the operational range of speech-in-noise tests towards positive SNRs and to allow for the presentation of more realistic SNR situations during the evaluation of hearing aid algorithms.

There are different approaches to shift the operational SNR range of a speech-in-noise test towards positive SNRs. One class of approaches modifies the acoustic characteristics of the speech signal (e.g., by introducing reverberation, time compression of speech, or more complex background noise situations). Another class modifies the overall cognitive demand of the comprehension task, e.g., by varying sentence predictability or syntactic complexity. The current study raises the difficulty of a speech-in-noise test by increasing the speech rate. This approach aims at preserving the acoustical characteristics and sensory cues of the speech as far as possible while challenging the cognitive processing of the listener with the effect that the overall SRT in noise is increased.

In general, the ability to comprehend time-compressed speech decreases as speech rate increases (e.g., Dupoux and Green, 1997; Versfeld and Dreschler, 2002; Humes et al., 2007). This is also

true for time-compressed speech in background noise (e.g., Tun, 1998; Schneider et al., 2005; Liu and Zeng, 2006; Gordon-Salant and Friedman, 2011; Adams et al., 2012).

For the development of a speech-in-noise test at positive SNRs utilizing time-compressed speech, the following specifications have to be met: First, the algorithms employed for time-compression have to yield a high signal quality with as few artifacts as possible in order not to deteriorate intelligibility. Second, these algorithms should be applicable to a wide range of speech signals which may be processed by hearing aid algorithms and presented with and without interfering noise in order to be able to compare relevant results across, e.g., hearing aid processing or speech enhancement schemes. Third, algorithms for time compression should deteriorate intelligibility across all items of the speech material in a very similar way in order to preserve the homogeneity of the test items typically encountered in speech-in-noise tests. This homogeneity leads to a steep discrimination function of the test (Kollmeier, 1990; Kollmeier and Wesselkamp, 1997) which in turn is required to discriminate in a sensitive way across different "effective" SNRs produced, e.g., by a speech enhancement algorithm. A specific concern therefore is to keep any position effects in intelligibility across words in an utterance or sentence as small as possible.

Different strategies can be used for increasing the speech rate of existing recordings, which maintain the pitch of the recorded voice. Within the time-compression techniques, uniform and non-uniform algorithms are differentiated. In simple terms, uniform time-compression algorithms eliminate small segments of the speech at defined regular intervals with overlap-add techniques. Dorran (2005) gives an overview of the different uniform methods. One processing strategy that provides a very good signal quality is the pitch synchronous overlap-add technique (PSOLA). It deletes pitch periods to compress the speech signal. This strategy is the basis for the time-compression capabilities found in several signal processing software packages (e.g., Adobe AUDITION and Windows EDW (Speech Research Lab, A.I. DuPont Hospital for Children and the University of Delaware, 2012)). The PSOLA, according to Moulines and Charpentier (1990), is implemented in the free phonetics software Praat (Boersma and Weenink, 2009). This method was previously used for the investigation of time-compressed speech (e.g., Janse, 2003; Adank and Janse, 2009; Shibuya et al., 2012).

In contrast, non-uniform time-compression algorithms consider speech characteristics with the aim of preserving as much of the time-dependent speech cues as possible. Therefore, the non-uniform time-compression algorithms require a further stage of analysis compared to the uniform processing. They determine special features to describe the structure of a speech signal and use this information to calculate a signal- and a time-dependent compression. Different strategies for the feature extraction are described in the literature and only a selection will be introduced here. Some algorithms make use of acoustic measures. Chu and Lashkari (2003), for example, calculate the short-term energy of a speech signal. They assume that high energy segments of the speech are more important than low energy segments and adjust the time-compression accordingly by compressing high energy segments less than those with low energy. Lee et al. (1997) and Kapilow et al. (1999) detect transient information and non-stationarity, respectively. Both algorithms only change steady segments of the speech signal. The transient parts are left unprocessed because they are important for comprehension and a good signal

quality. Demol et al. (2005) adopt and complement the idea and classify the speech into plosive-, vowel-, and consonant-like segments as well as phone transitions and pauses. Different time-compression factors are applied to these classes. As a result, pauses are compressed more than other segments and plosive-like segments remain unprocessed. Thomas et al. (2008) used a similar approach in order to distinguish voice, unvoiced, and silent intervals from unvoiced and transient segments of on-going speech with the goal to not process the unvoiced and transient components. Höpfner (2006, 2007, 2008) goes beyond this and subdivided detected phonemes themselves into stationary parts and non-stationary transitions to neighboring phonemes. The algorithm compresses silence and stationary parts of voiced and unvoiced phonemes differently. Again, plosives and transient parts of phonemes remain unprocessed. Despite the complexity of these algorithms, all have several disadvantages when applied to speech material of speech-in-noise tests. Either the algorithms consider only few aspects of natural fast speech or they show compression for only parts of the signals. Both may generate unnatural speech timing at high speech rates, which are necessary to reach positive SNRs. In worst cases, high speech rates cannot be processed and speech signals with missing and unprocessed parts are generated. Additionally, at high speech rates the signal quality is often influenced by artifacts. A promising and available non-uniform approach to overcome these problems was described by Covell et al. (1998). Their Mach1 algorithm preserves the "natural timing of [natural] fast speech" (Covell et al., 1998, p. 349) without the use of a classification and with compression of the entire signal. It applies the most compression to pauses and silence; intermediate compression is used for unstressed vowels; consonants are compressed based on their adjacent vowels and in total consonants are compressed to a higher amount than vowels. Previously, this algorithm was used by Covell et al. (1998), He and Gupta (2001), and Tucker and Whittacker (2006) for their investigations.

Covell et al. (1998) compared the Mach1 algorithm with a uniform synchronous overlap-add algorithm at the same global compression. They compressed short and long dialogs as well as monologs with both algorithms to 24%-39% of their original length (calculated from the time-compression rates given by the authors, unprocessed speech: 155-302 syllables per minute (syll/min); processed speech: 546-942 syll/min[1] and stated no restrictions for the selection of the time-compression factor. Listeners then were asked to answer questions regarding the content of the presented speech and performed paired comparisons of signals processed with Mach1 and the uniform algorithm. Covell et al. (1998) found that compared with uniform processing, comprehension improved by 17% and participants preferred listening to the Mach1 algorithm. Furthermore, the advantages of Mach1 increased as the speech rate increased. These results were confirmed by He and Gupta (2001) who investigated a synchronous overlap-add method and a modified version of Mach1 and compressed speech down to 40% of their original length (unprocessed speech: 249-286 syll/min, processed speech: up to 623-714 syll/min[1]). Adank and Janse (2009) measured the comprehension of normal, time-compressed and natural fast speech (presented at 46% of the original length and 612 syll/min, calculated from the speech rate

[1] Speech rate in syllables per minute was calculated based on the average number of syllables per word of about 1.4 for English (Lamel et al., 1989) and the speech rate in words per minute stated by the authors.

given by the authors). In contrast to the advantage of the mimicked natural fast speech by Mach1 observed by Covell et al. (1998) and He and Gupta (2001), Adank and Janse (2009) found that listeners exhibited greater difficulties when processing natural-fast speech than time-compressed speech. According to Adank and Janse (2009) these discrepant findings may stem from the greater spectrotemporal variations of the natural fast speech compared to the original speech. Such variations are not only caused by differences of uniform and natural speech production but are also expected for non-uniform processing strategies. They all result in alterations of perceptually relevant cues for speech intelligibility and signal quality. These results necessitate a comparison of the uniform PSOLA and the non-uniform Mach1 for the application in a speech-in-noise test at positive SNRs. The above cited results do not explain the change of intelligibility (as measured by a speech intelligibility test) according to the processing strategy of a specific time-compression algorithm. Furthermore, the influence of an additional background noise is unknown and, therefore, no hint is offered for the time compression needed to reach positive SNRs. For a speech-in-noise test with time-compressed speech, it is necessary to consider these factors when selecting a time-compression algorithm in order to present speech material with equal intelligibility and therefore to reach a high sensitivity of a speech-in-noise test. In addition to subjective comparisons of the different algorithms it is necessary to compare objective performance. The alteration of perceptually relevant cues for understanding speech in noise, such as, e.g., overall time-compression, long-term spectra, and the modulation spectra of the compressed speech should be comparable between the two algorithms. This ensures that only those differences across time-compression strategies which are related to the modification of temporal cues, such as, e.g., the phoneme durations result in intelligibility differences.

As mentioned before, the material for a sensitive speech-in-noise test is normally designed to reach a homogeneous intelligibility across words and comparable lists of words or sentences. For example, Wagener et al. (1999b) achieved this by adapting the presentation level of single words. As a result, Wagener and Brand (2005) reported only a small serial word position effect for the Oldenburg sentence test. The serial position effect occurs in free recall and consists of higher recall accuracy for words presented at the beginning and the end of a series compared to words presented at intermediate positions (e.g., Murdock Jr., 1962; Glanzer and Cunitz, 1966). Wagener and Brand (2005), for example, observed a more favorable SRT for names (at the beginning of their sentences) and subsequent words with a difference amounting to -0.6 and -1 dB for normal-hearing and hearing-impaired participants, respectively. In comparison to a sequence of normal speech, Aaronson et al. (1971) found less order errors for the time-compressed sequence of digits which were presented with enlarged pauses between the elements and at the same point in time as the elements of the normal speech. Therefore, Aaronson et al. (1971) suggested a positive effect of the enlarged pause time. The missing of pauses between words of a sentence will probably increase the difficulty and enlarge the serial position effect.

For the development of a speech-in-noise test with time-compressed speech it has to be tested if the small serial position effect observed by Wagener and Brand (2005) is kept within reasonable limits for time-compressed speech in order not to reduce the sensitivity of a speech-in-noise test against effective SNR changes too much.

Summarizing, the goal of this study was to find the most appropriate method for constructing a speech intelligibility test that yields best discrimination across effective SNR changes produced e.g., by speech enhancement algorithms at positive SNRs. This goal was followed by comparing the uniform PSOLA time-compression algorithm available in Praat software (Boersma and Weenink, 2009) and the non-uniform Mach1 time-compression algorithm mimicking natural fast speech even at high speech rates. Comparison of the two algorithms should not only be performed in terms of speech intelligibility across the speech test items employed, but should also consider the alterations of perceptually relevant cues for understanding speech in noise such as deviations from the targeted compression across time, changes of the phoneme durations, long-term spectra, and modulation spectra.

2.2 Methods

2.2.1 Signals

Evaluation of the two algorithms was conducted using sentences of the German matrix test (Oldenburg sentence test, OLSA, Wagener et al., 1999c). The sentences consist of a fixed syntactic structure (name, verb, numeral, adjective, and object). Each position of the sentence contains one of ten possible words, which were randomly selected to construct the sentences (examples: Peter bekommt vier grüne Messer. (Peter gets four green knifes.); Thomas kauft neun schwere Tassen. (Thomas buys nine heavy cups.)). Fixed syntactic structure and random selection induce unpredictable semantics of the sentences. As a result the sentences cannot be easily memorized and there is no benefit from sentence context (Wagener and Brand, 2005). In total, 100 sentences were available in 20 test lists of 30 sentences each with equal intelligibility. Co-articulation between the words was taken into account by recoding single words with all ten possible subsequent word alternatives. For the construction of sentences recordings of single words were selected, where co-articulation matched to the subsequent word, and the ending of the co-articulation was faded into the beginning of the subsequent word. This synthesis of sentences produced natural sound (Wagener and Brand, 2005). Sentences were spoken with a normal to moderate speech rate of on average 233 (+/-27) syll/min (Wagener et al., 1999c). The sentences were presented together with the noise stimulus of the OLSA which has the same long-term spectrum as the speech and is composed by superposition of all sentences (Wagener et al., 1999c).

2.2.2 Methods of time compression

Current time-compression algorithms maintain the pitch of the recorded voice and are distinguished into uniform and non-uniform algorithms. There are different descriptions characterizing the amount of time compression. This paper defines the time-compression factor ρ as the duration of the compressed signal compared to the original duration in percent. As an example, $\rho = 25\%$ corresponds to a speech rate which is four times faster than the original and increases from 233 to 932 syll/min for OLSA material. Hence, smaller time-compression factors result in higher time compression and faster synthesized speech.

2.2.2.1 Uniform time-compression algorithm

The uniform time-compression algorithm PSOLA (Moulines and Charpentier, 1990) analyzes the pitch of a speech signal, sets pitch marks, and segments the original signal into windowed frames. At unvoiced parts of the speech the pitch period is set constant. For the synthesis of time-compressed speech, a new set of pitch marks depending on the time compression is calculated, and the analyzed frames are rearranged to the synthesized compressed signal. In simple terms, for a given compression factor, some segments are deleted and the remaining segments are concatenated to the time-compressed signal. Position and number of deleted segments are dependent on the time-compression factor. This method was used as implemented in the Praat software (Boersma and Weenink, 2009).

2.2.2.2 Non-uniform time-compression algorithm

In contrast to uniform time-compression algorithms, non-uniform algorithms take the structure of the speech into account. The Mach1 algorithm (Covell et al., 1998) estimates two continuous parameters termed "local emphasis" and "relative speaking rate". These values are combined for an estimate of the "audio tension" which is defined as "the degree to which the local speech segments resist against changes in rate" (Covell et al., 1998, p. 349). With more audio tension, Mach1 applies less compression for a given segment of the speech signal. Depending on the estimated audio tension, a local time-compression factor is derived and used to control a uniform time-compression technique like the synchronous overlap-add method. As a result, the Mach1 algorithm mimics natural fast speech. The experiments were run with a Mach1 implementation by Tucker and Whittacker (2006).

The PSOLA and Mach1 algorithm were applied to all sentences of the OLSA and compressed the speech material to $\rho = 25\%$, 30%, 45%, and 40% (representing speech rates of 932, 777, 666, 583 syll/min). Covell et al. (1998) describe that the Mach1 algorithm uses time-compression factors close to the targeted factor. To achieve the targeted time-compression factor, they recommended a "slow response feedback loop" (Covell et al., 1998, p. 351) to correct for the time-compression factor. This is consistent with the observations of the current study whereby an increase of the time compression with time was observed. To consider this adaptation of the Mach1 algorithm in the current study, each sentence was presented five times in succession (separated by a break of 0.15 s duration). Subsequently, the third presentation of the sentence was selected and cut out for further usage. All of these processed sentences were arranged in the same way as the original list structure.

2.2.3 Participants

In total, 48 normal-hearing listeners (mean age: 23 yr, range: 20-33 yr; 38 female, 10 male) participated in the experiments. They were recruited through an online advertisement found on the webpages of the Universities of Oldenburg. Based on pure-tone audiometric testing, all participants exhibited auditory thresholds of 20 dB HL or better at all tested frequencies between 0.125 and 8 kHz. All listeners spoke German as their native language and had no or little experience with the OLSA. They were paid 10 Euro (about 14 US$) for each hour of their participation.

2.2.4 Experiments

2.2.4.1 Objective analysis of time-compression algorithms

For the objective description and comparison of the uniform and non-uniform algorithms, all 100 sentences of the OLSA were processed with the uniform PSOLA method and the non-uniform Mach1 algorithm. Two different time-compression factors of $\rho = 25\%$ (932 syll/min) and $\rho = 50\%$ (466 syll/min) were applied. These compression factors were chosen as the boundaries of the range of interest, where the listeners need a high or even positive SNR for understanding 50% of the speech. Speech compressed to a time compression factor below $\rho = 25\%$ (932 syll/min) was observed to be too distorted for all algorithms under investigation. Four different measures were conducted to describe the processing of the algorithms.

a) Deviations from the targeted time compression: Prior to the processing by the time-compression algorithms, a targeted time-compression factor has to be specified. For the calculation of deviance from this targeted time compression, the duration of the time-compressed signal was compared with the duration of the original signal and both were used to calculate the time-compression factor after processing. The differences between the time-compression factor after the processing and the targeted time-compression factor describe the deviations.
b) Alterations of the phoneme durations: To analyze the alteration of the phoneme duration, the first ten sentences of the OLSA were time compressed with a factor of $\rho = 50\%$ (466 syll/min). These sentences include all words of the test. None of these words was presented twice. The time-compressed sentences were segmented into phonemes and labeled in Praat. Then, the duration of the phonemes in the original speech and after time compression was analyzed. The results were analyzed per phoneme classes consisting of plosives, nasals, fricatives, approximants, vowels, and affricates.
c) Long-term spectra: The long-term spectra were calculated in 1/3-octave bands for all sentences.
d) Modulation spectra: The temporal envelope fluctuations of the time-compressed speech were characterized by the modulation spectra in 1/3-octave modulation frequency bands. A fast Fourier transformation of the Hilbert envelope of all sentences was applied for the calculation of the modulation spectra in order to compare the original and time-compressed speech material.

2.2.4.2 Perceptual analysis of time-compression algorithms

Testing occurred in a sound-treated booth and the presentation of stimuli was controlled by a PC using MATLAB (MathWorks, Natick, MA) based programming. Signals were routed through a sound card (AD/DA-Interface ADI 2, RME, Audio AG, Haimhausen, Germany) and a head phone amplifier (HB 7 Headphone Driver, Tucker Davis Technologies, Alachua, FL) to headphones (HDA 200, Sennheiser, Wedemark-Wennebostel, Germany). The headphones were free-field equalized according to international standard (IEC 60645-2, 2010; ISO 389-8, 2004). The noise was presented at 65 dB sound pressure level. The presented SNRs were dependent on the time-compression factor and algorithm (see Table 2.1). They were selected based on a pre-study to reach intelligibility between 20% and 80%, which is necessary for a

reliable fit of the discrimination function (described below). Participation was limited to one session of two hours duration in order to avoid training effects from session to session. Participants were divided randomly into two groups listening to signals processed with PSOLA or Mach1, respectively. Measurements of the original OLSA occurred at the end of the experiments, and resulted in nearly equal median SRT values for both participating groups. Participants who listened to PSOLA or Mach1 showed median SRTs for original speech of -8.2 dB SNR and -8.3 dB SNR, respectively.

In a pre-study, 14 normal hearing young adult participants listened to original and time-compressed speech (ρ = 25%, 30%, 45%, 40% representing speech rates of 932, 777, 666, 583 syll/min) in background noise at different SNRs. They were asked to repeat each of the sentences that were presented and intelligibility was determined using a word scoring procedure. The sentence started after a countdown, which counted from three backwards and was displayed on a monitor. Participants were familiarized with the OLSA procedure by listening to one training list of original (i.e., unprocessed) sentences. During training, after each response the correct sentence was displayed on the screen as feedback to the listener. The SNR was changed adaptively to reach a threshold of 50% intelligibility. The training of the original speech was followed by a training list of the time-compressed speech processed with the first factor presented in the subsequent test lists. After the training, all participants completed the intelligibility task at fixed SNRs for two out of four possible time-compression factors. Therefore, after the initial training list, four to five test lists were presented at different SNRs at the time-compression factor for which training was received. A training list was then presented for a second time-compression factor, followed by another four to five test lists and different SNRs. In total, each participant listened to 11 or 12 lists. The presented time-compression factors and the order of the presented SNRs were randomized across all participants using a Latin square design. Based on the results from these pre-study measurements, different SNRs for the main investigation were adapted to reach intelligibilities between 20% and 80%. Subsequently, participants of the main study conducted the same training and speech intelligibility tests at the determined fixed SNRs. The SNRs presented in the main investigation are summarized in Table 2.1. The group of participants was divided randomly into two subgroups of 24 participants each, who listened to sentences processed with PSOLA or Mach1, respectively.

Table 2.1: Presented SNRs for different time-compression factors and algorithms.

ρ [%]	**PSOLA** SNR [dB]					**Mach1** SNR [dB]				
25	-3	1	5	9	-	-1	1	5	9	-
30	-5	-3	-1	1	3	-5	-3	-1	1	3
35	-5	-3	-1	1	-	-5	-3	-1	1	-
40	-7	-5	-3	-1	-	-7	-5	-3	-1	-

For the presentation of the results, performance on the intelligibility task for each time-compression factor at different SNRs was used to model discrimination functions described by

Equation 2.1. This discrimination function was suggested by Wagener and Brand (2005). Intelligibility is defined as the mean probability (p_{in}) of correctly repeated words if the sentences are presented at a certain SNR,

$$p_{in}(\mathrm{SNR}, \mathrm{SRT}, \mathrm{slope}) = \frac{1}{1+e^{4*\mathrm{slope}*(\mathrm{SRT}-\mathrm{SNR})}} . \qquad (2.1)$$

This function was applied to estimate the SRT and slope of the function at SRT for the measured intelligibility of each participant using a maximum likelihood fit.

2.3 Results

The presentation of the results is organized as follows. First, the results from the objective analysis focusing on the deviation from the targeted time compression, alterations of phoneme durations, long-term spectra, and modulation spectra are presented. Second, results of the intelligibility measurements are presented and analyzed as a function of the two time-compression algorithms explored in this study and the serial word position within the sentences.

2.3.1 Objective analysis of time-compression algorithms

2.3.1.1 Deviation from targeted time compression

To analyze the effect of a time dependent time-compression factor, Figure 2.1 shows the deviations from the targeted time-compression factor ρ = 25% (932 syll/min) and ρ = 50% (466 syll/min). Only the Mach1-processed materials differ from the targeted factor and the difference is higher for the lower time-compression factor. The median deviation amounts to 0.9% for a time-compression factor of ρ = 25% (932 syll/min). In other words, a median time-compression factor of about 26% (896 syll/min) was reached instead of 25% (932 syll/min). The deviation for a time-compression factor of ρ = 50% (466 syll/min) is smaller and shows a median of 0.2%.

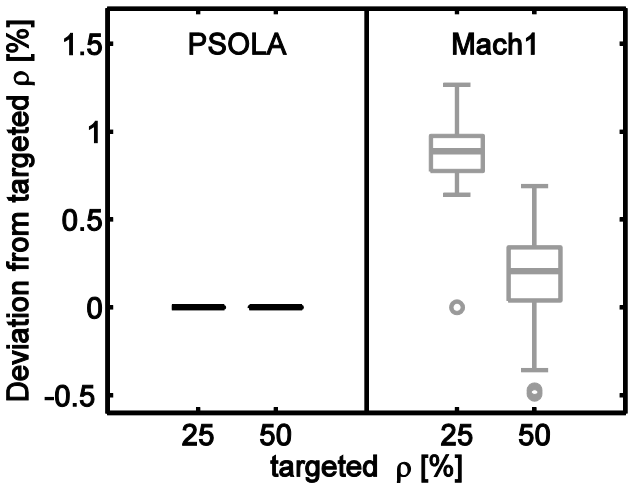

Figure 2.1: Deviations from the targeted time-compression factors ρ = 25% (932 syll/min) and ρ = 50% (466 syll/min) of speech time compressed using the PSOLA or Mach1 algorithm.

2.3.1.2 Alterations of phoneme durations

Given the differences between the two algorithms, it was expected that Mach1 would produce a non-linear reduction in phoneme duration. The scatter plots in Figure 2.2 show the phoneme durations after time compression in relation to their durations before compression for the different phoneme classes. A linear regression analysis is depicted by the solid line and shows the changes of the phoneme durations due to time compression. The dashed lines represent the linear reduction of the phonemes to half of their duration. The PSOLA algorithm compressed the phonemes nearly equally to this linear reduction. In contrast, Mach1 reduced the durations of long phonemes to a higher amount than a linear compression. The compression of long consonants (e.g., nasals) was higher than the compression of vowels. Short consonants were compressed linearly to about half of their duration. Additionally, short and intermediate vowels showed a higher variability of compression. The analysis of their duration varies between nearly no and linear compression. Furthermore, some phonemes like plosives and vowels are missing (see the number of detected phonemes "n"). Also, as described in Section 2.2.2.2, the time-compression increased within a sentence for Mach1. Therefore the initial names were longer for the time-compressed sentences compared to the names processed with PSOLA.

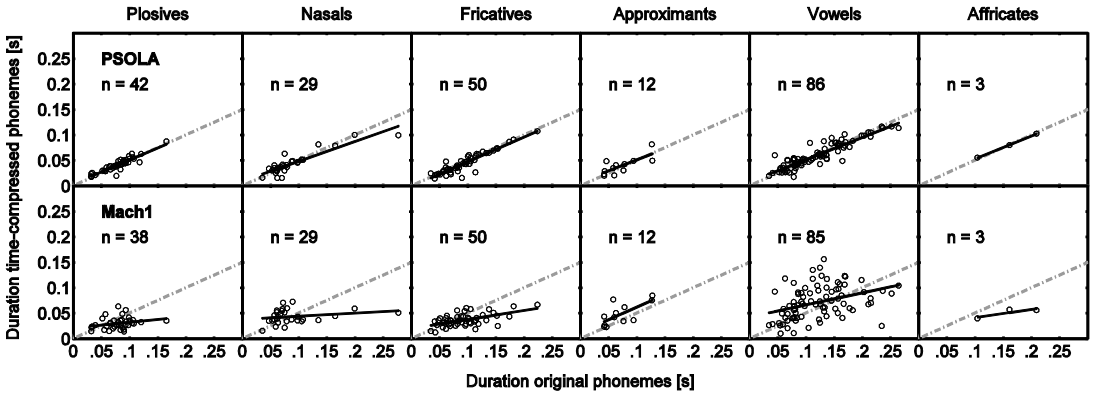

Figure 2.2: Scatter plot of the phoneme durations for the different phoneme classes after time compression as a function of their durations before compression. A linear regression analysis is depicted by the solid line and shows the change of the phoneme durations due to time compression. The dashed line represents the linear reduction of the phonemes to the half of their duration with a factor of ρ = 50% (466 syll/min).

2.3.1.3 Long-term spectra

Figure 2.3 shows the 1/3-octave band long-term spectra of the original speech with the time-compressed speech for all OLSA sentences (upper panel) and the differences between the spectra of the original and time-compressed speech (lower panel). The signals were either processed with PSOLA or Mach1 algorithm at time-compression factors ρ = 25% (932 syll/min) and ρ = 50% (466 syll/min). Compared to the original speech, the main difference is how PSOLA processes low-frequency information and how Mach1 processes high-frequency information, with larger differences being observed with the low time-compression factor. The long-term RMS and peak level of PSOLA and Mach1, respectively, were comparable to the levels of the original speech.

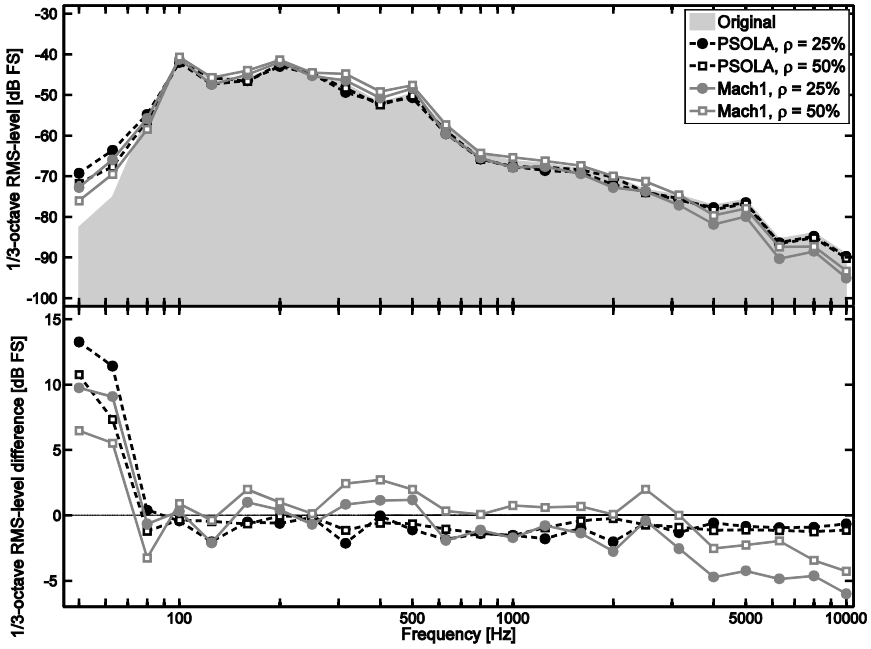

Figure 2.3: 1/3-octave band long-term spectra of original and time-compressed speech (upper panel) and difference of the 1/3-octave band long-term spectra of original and time-compressed speech (lower panel). The spectrum of the original speech is shown as the gray area. The spectra of signals processed with the time-compression factors ρ = 25% (932 syll/min, filled circles) and ρ = 50% (466 syll/min, open squares) are displayed with black dashed and gray solid lines for PSOLA and Mach1, respectively.

2.3.1.4 Modulation spectra

Time compression of the speech was expected to produce changes in its temporal envelope fluctuations described by the modulation frequencies. Figure 2.4 depicts the power within 1/3-octave bands in the modulation frequency domain against the modulation frequency. Compared to the original speech with a peak of the modulation spectrum at about 2 Hz, time-compressed speech shows a shift of the peak to higher modulation frequencies. This peak is dependent on the time-compression factor and is shifted to higher frequencies for lower factors (higher speech rates). For example, the modulation spectrum of speech compressed with PSOLA at a factor of ρ = 25% (932 syll/min) is shifted to four times higher modulation frequencies than the spectrum of the original speech. The processing of Mach1 also shifted the modulation spectra to higher modulation frequencies, but with slightly different shapes and slightly broader peaks compared to original speech.

2.3.2 Perceptual analysis of time-compression algorithms

Figure 2.5 depicts boxplots that document the performance on the intelligibility task determined with the two time-compression algorithms at different SNRs and time-compression factors. Each boxplot summarizes results of twelve participants. As expected, intelligibility increases with increasing SNR and increasing time-compression factor. Discrimination functions were estimated for the results from each individual listener. The median SRT and median slope were determined across twelve listeners and the corresponding median discrimination function is depicted in Figure 2.5 as well.

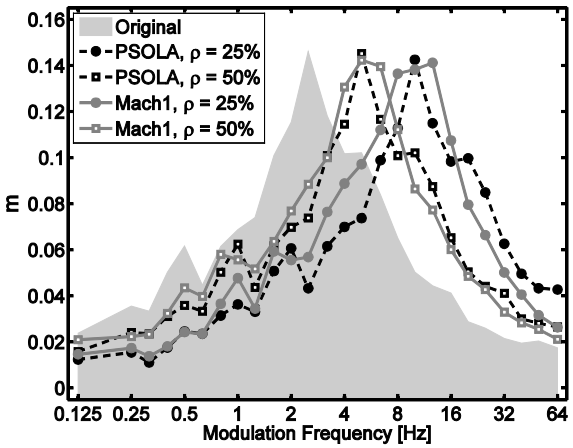

Figure 2.4: Modulation spectra of the original speech and speech time compressed with PSOLA and Mach1 and factors of $\rho = 25\%$ (932 syll/min) and $\rho = 50\%$ (466 syll/min), respectively. The modulation spectrum of the original speech is depicted by the gray area. The signals processed with the time-compression factors $\rho = 25\%$ (filled circles) and $\rho = 50\%$ (open squares) and the algorithms PSOLA and Mach1 are displayed with black dashed and gray solid lines, respectively.

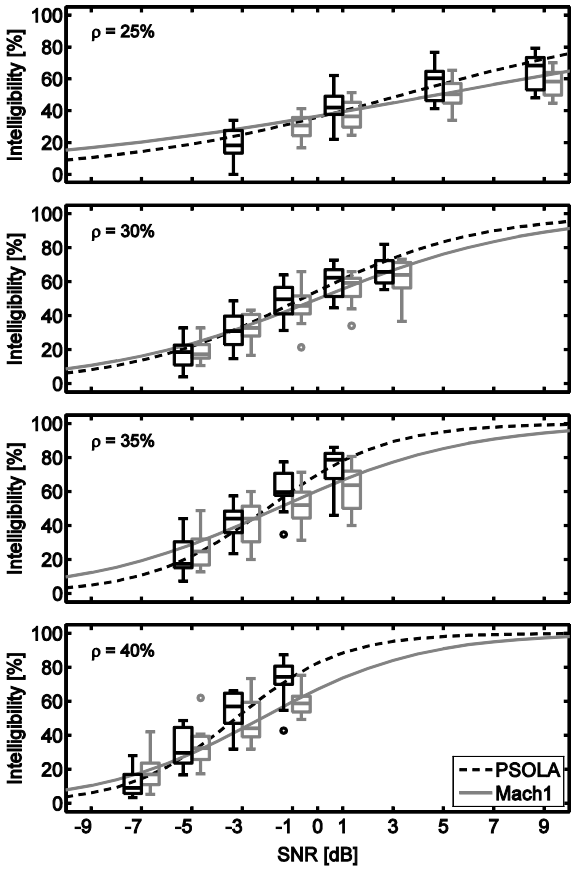

Figure 2.5: Speech intelligibility at different SNRs and corresponding discrimination functions for the two time-compression algorithms and for time-compression factors between $\rho = 25\%$ (932 syll/min) and $\rho = 40\%$ (583 syll/min). The black and gray boxplots indicate the results measured with PSOLA and Mach1, respectively. The median discrimination functions across twelve listeners for PSOLA are represented with dashed black lines and the functions for Mach1 are displayed with solid gray lines.

In addition, median SRT and slope of speech that was time-compressed with PSOLA and Mach1 are documented in Table 2.2. The SRT increases and slope decreases as the time-compression factor decreases. The speech compressed with PSOLA leads to lower SRTs and steeper slopes than the speech compressed with Mach1. The distribution of the SRT results was investigated. A Kolmogorow-Smirnow test with a Lillifors correction confirmed the normal distribution of the data. Although the results of a Levene statistic showed significantly different variances of the results, significant differences of the calculated SRTs in dependence on the time-compression algorithm and time-compression factor were analyzed with a two-way analysis of variance (ANOVA), because the ANOVA is robust against the violation of the variance homogeneity. It showed a significant main effect in the SRT data of the factors algorithm ($F(1,88) = 4.63$, $p = 0.034$) and time-compression factor ($F(3,88) = 73.92$, $p < 0.001$). Their interaction was not significant. Post hoc testing with a paired t-test and Bonferroni correction showed for all combinations of time-compression factor significance except for $\rho = 35\%$ (666 syll/min) and $\rho = 40\%$ (583 syll/min).

Table 2.2: Median SRTs, slopes, and interquartile ranges (in brackets) for the median fitted discrimination functions.

ρ [%]		**PSOLA**	**Mach1**
25	SRT [dB SNR]	3.4 (4.3)	4.8 (5.6)
	Slope [%/dB]	4.3 (1.2)	2.9 (0.8)
30	SRT [dB SNR]	-0.6 (2.3)	0.0 (2.3)
	Slope [%/dB]	7.2 (1.2)	5.9 (0.7)
35	SRT [dB SNR]	-2.0 (1.5)	-1.6 (3.0)
	Slope [%/dB]	10.6 (2.8)	6.7 (1.1)
40	SRT [dB SNR]	-3.3 (1.0)	-2.3 (1.9)
	Slope [%/dB]	12.0 (3.0)	8.0 (1.6)

2.3.3 The effect of serial word position on speech intelligibility

Further analysis of the results investigated the dependence of speech intelligibility on the word position within the sentences. Figure 2.6 depicts intelligibility as a function of SNR for different word positions. Results are shown for the two time-compression algorithms and time-compression factors. Discrimination functions were calculated for each of the twelve participants. Be aware that these fits may be less reliable than the fits to the speech material depicted in Figure 2.5 because not the full range of performance on the intelligibility task was between 20% and 80% in Figure 2.6. For example, intelligibility of Mach1 was found to be greater than 60% at the time-compression factor $\rho = 25\%$ (932 syll/min) for names. Additionally, the respective median discrimination function is plotted in Figure 2.6. The figures within each row exhibit the same time-compression factor while figures within each column show the results of words with the same position within the sentences. Table 2.3 summarizes the median SRTs, slopes, and their interquartile ranges for the plotted discrimination functions as a function of the time-compression algorithm, factor, and word position. The initial names show smaller median SRTs and a higher intelligibility than verbs or objects. Higher median

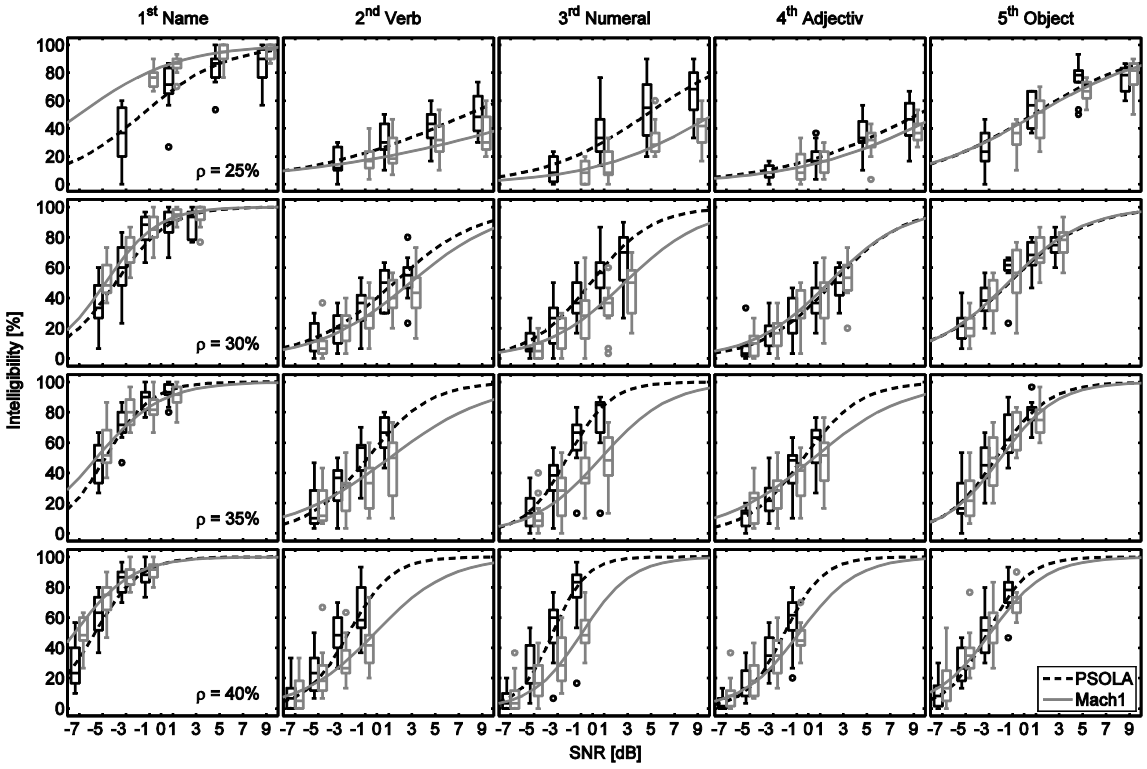

Figure 2.6: Intelligibility measured with different time-compression algorithms at different SNRs and time-compression factors ranging between $\rho = 25\%$ (932 syll/min) and $\rho = 40\%$ (583 syll/min) and the associated discrimination functions. Each column shows the results for each word position within the sentences. The black and gray boxplots depict performance for PSOLA and Mach1, respectively. The median discrimination functions across twelve listeners for PSOLA is represented with a dashed black line and the function for Mach1 is displayed with solid gray lines.

SRTs (poorer intelligibility) were observed for adjectives (fourth position in each sentence). In general, the results measured with PSOLA describe lower SRTs and a higher intelligibility than the results measured with Mach1. One exception is that lower SRTs and higher intelligibility performance was observed for names with Mach1 than with PSOLA. Statistical analyses of the SRT data were conducted using a three-way ANOVA of the factors word position, time-compression algorithm, and time-compression factor. For all factors and their interactions a significant effect was observed (word position: $F(4,440) = 177.31$, $p < 0.001$; algorithm: $F(1,440) = 15.52$, $p < 0.001$; time-compression factor: $F(3,440) = 195.37$, $p < 0.001$, word position * time-compression factor: $F(12,440) = 14.25$, $p < 0.001$, word position * algorithm: $F(4,440) = 18.9$, p < 0.001, word position * time-compression factor * algorithm: $F(12,440) = 1.93$, $p = 0.029$). For the interaction of time-compression factor and algorithm no significant effect was found ($F(3,440) = 1.14$, $p = 0.333$). A two-way ANOVA was used to analyze the significant differences of the factors word position, time-compression factor and their interactions for PSOLA and Mach1 separately. For all factors and their interactions a significant effect was observed (see Table 2.4). Post hoc testing of the factor word position using paired t-tests and a Bonferroni correction was conducted. Verbs showed no significant differences to numerals and adjectives if the signals were processed with PSOLA. All paired comparisons of the word positions were significant different for Mach1 with exception of

Table 2.3: Median SRTs, slopes, and interquartile ranges for the median fitted discrimination functions as a function of word position within the sentences.

ρ [%]		PSOLA					Mach1				
		Name	Verb	Numeral	Adjective	Object	Name	Verb	Numeral	Adjective	Object
25	SRT [dB]	-1.6 (4.6)	7.8 (4.8)	4.6 (5.3)	9.4 (6.1)	0.8 (4.3)	-7.0 (4.6)	15.0 (10.5)	10.4 (2.6)	11.4 (2.0)	1.0 (3.6)
	Slope [%/dB]	7.0 (3.0)	3.6 (1.6)	5.7 (3.2)	4.3 (1.2)	5.1 (2.3)	6.0 (7.8)	2.5 (1.3)	4.8 (2.7)	4.0 (3.3)	4.9 (1.6)
30	SRT [dB]	-3.9 (1.9)	1.5 (1.9)	-0.1 (2.5)	2.1 (2.4)	-1.3 (1.3)	-4.8 (1.3)	2.9 (3.3)	2.8 (3.1)	1.8 (2.9)	-1.4 (1.6)
	Slope [%/dB]	11.1 (5.7)	6.9 (2.9)	9.6 (3.2)	8.2 (2.4)	7.5 (2.4)	11.2 (4.0)	6.4 (2.3)	7.4 (2.9)	7.6 (5.3)	7.7 (2.7)
35	SRT [dB]	-4.8 (1.8)	-0.8 (2.1)	-2.1 (1.9)	-0.4 (2.3)	-2.4 (1.8)	-5.5 (4.5)	1.1 (6.8)	0.8 (2.9)	0.5 (3.0)	-2.0 (2.8)
	Slope [%/dB]	13.0 (6.1)	9.6 (4.7)	13.8 (3.6)	10.3 (4.1)	11.3 (3.4)	9.0 (3.4)	5.8 (4.0)	8.6 (3.7)	6.4 (3.5)	10.4 (3.1)
40	SRT [dB]	-5.6 (1.6)	-2.4 (2.1)	-3.4 (1.4)	-2.0 (1.3)	-3.1 (0.4)	-7.4 (1.1)	-0.1 (2.8)	-0.9 (2.1)	-0.8 (1.6)	-2.9 (1.8)
	Slope [%/dB]	13.2 (6.5)	13.7 (3.9)	17.4 (8.0)	15.3 (4.9)	14.5 (4.3)	10.2 (6.1)	7.9 (4.2)	11.7 (10.0)	10.3 (2.6)	10.5 (3.3)

Table 2.4: Results of a two-way ANOVA for SRTs measured with PSOLA and Mach1.

	PSOLA	Mach1
Word position	$F(4,220) = 64.89$, $p < 0.001$	$F(4, 220) = 116.69$, $p < 0.001$
Time-compression factor	$F(3, 220) = 121.42$, $p < 0.001$	$F(3, 220) = 85.29$, $p < 0.001$
Word position * time-compression factor	$F(12, 220) = 6.04$, $p < 0.001$	$F(12, 220) = 9.26$, $p < 0.001$

adjectives versus verbs and adjectives versus numerals. This supports the general trend already seen in Figure 2.6 that intelligibility performance was best when listening to names. The poorest performance was determined for verbs and adjectives. All paired comparisons for Mach1 of the time-compression factor were significantly different except for $\rho = 30\%$ (777 syll/min) and $\rho = 35\%$ (666 syll/min) where no significant differences were observed. The significant main effect of the time-compression factor with PSOLA was also observed to be significant in a paired post hoc t-test with Bonferroni correction for all time-compression factors.

2.4 Discussion

2.4.1 Objective analysis of time-compression algorithms

The objective measures computed in this study verified the intended changes of the targeted time-compression factor, the phoneme duration, as well as the differences of the long-term spectra and modulation spectra produced by the uniform PSOLA and the non-uniform Mach1 algorithms. Larger changes of all objective measures were observed for signals processed by Mach1 than by PSOLA.

In detail, PSOLA showed no deviation from the targeted time-compression factor, while Mach1 compressed the signals with very small variations of the time-compression factor. The variations increased slightly with decreasing time-compression factor. Time-dependent processing of Mach1 caused these deviations. For the investigation of comparable signals with Mach1, it was necessary to process a series of five sentences with only short pauses and to select the third sentence. For the first sentences of the series, processing of the stimuli by Mach1 resulted in an increase of the time-compression factor. The deviations from the targeted time-compression factor might have occurred because the processing of Mach1 was still not completely adapted. Furthermore, Covell et al. (1998) described that the Mach1 algorithm processes time-compression factors near the targeted factor with no claim of actually achieving this factor. As a result, they recommended a "slow-response feedback loop [...] to correct the long term errors" (Covell et al., 1998, p. 351) in the time compression. The small deviation from the targeted time-compression factor measured with Mach1 generated negligible modification of the intelligibility. Analysis of the discrimination functions displaying intelligibility and time compression suggest an improvement of about 2% in intelligibility. However, although Mach1 showed less compression at low time-compression factors speech was less intelligible compared to PSOLA. Thus, the observed perceptive results are not fully explainable by deviations from the targeted time-compression factor for Mach1 compared to PSOLA.

Investigation of the phoneme duration of the PSOLA algorithms showed the expected linear reduction of all phoneme durations independent of their classification or original duration. In contrast, Mach1 used a higher time compression for long phonemes than for short phonemes. This result reflects the processing goals of the algorithm as previously described. For stressed and unstressed vowels, the algorithm attempts to use little and intermediate compression, respectively (Covell et al., 1998). In total, the Mach1 algorithms should compress consonants to a higher amount than vowels (Covell et al., 1998). The results of this study support this

expectation and showed partly higher compression for long consonants than for long vowels. In general, short consonants, rather than vowels were compressed to the expected linear compression. Finally, vowels showed a higher variety of the duration after compression, although some were not compressed at all.

The time-compression algorithms produced several changes to the frequency characteristics of the signals. Both algorithms showed a variation at low frequencies. These variations and the long-term RMS-level, which is equal to the original speech, induced a difference of the long-term spectra of time-compressed and original speech, which was not equal to zero for PSOLA at medium and high frequencies. With Mach1, the long-term spectra produced a variation at high frequencies compared to the spectrum of the original speech. This implies that high-frequency segments of the speech were primarily changed. These segments likely consisted of missing phonemes, which were mainly plosives and sibilant parts of consonants because consonants were compressed to a greater extent than vowels and vowels consist of less high-frequency spectra compared to consonants (Kent and Read, 2002). The noise used for the OLSA has the same long-term spectrum as the speech signal (Wagener et al., 1999c) in order to mask the speech in an optimal way. Large differences of the spectra would affect the spectral masking and shift the SRT to lower values. Because the aim of this study was to provide a speech-in-noise test with positive SRT results, only negligible deviations of the long-term spectrum of the time-compressed speech would be appropriate. Since the measured long-term spectra showed only small differences, no deterioration of spectral masking was expected.

Both algorithms produced a time-compression dependent shift of the modulation spectra to higher frequencies, which is an expected result. The modulation spectra of PSOLA maintained the shape of the original speech, while the modulation spectra of Mach1 showed a slightly broader peak and other small variations. The shape of the modulation spectra can be viewed in terms of Houtgast and Steeneken (1985). They investigated speech intelligibility as a function of modulation depth which is represented by the height of the maximum in the modulation spectrum. They found that intelligibility decreases if the modulation depth is reduced due to background noise or reverberation. Because modulation depth was not reduced in the current study, there is evidence for the assumption that changes in intelligibility were mainly caused by the shift of the spectra according to the time compression and additional noise, and likely not by the change of the shape of the modulation spectra generated by Mach1.

Generally, the objective analysis showed that differences in speech intelligibility for the two time-compression algorithms were not caused by deviation from the targeted time-compression factor, variations of the long-term spectra, changes of the modulation depth, or shape of the modulation spectra. Therefore, the difference in intelligibility observed for the two time-compression algorithms may have aroused primarily due to variations of time-dependent speech cues, such as, e.g., the phoneme duration.

The objective analysis of the signals highlights the change to the speech signal after time compression. These changes may limit the application of a speech-in-noise test with time-compressed speech for the investigation of hearing aid processing. These processing strategies use assumptions and statistics of normal speech, which may no longer apply for the case of

time-compressed speech material. Consequently, the output signal of a hearing aid might reflect unexpected processing patterns.

2.4.2 Perceptual analysis of time-compression algorithms

For the perceptual analysis of the time-compression algorithms, the intelligibility of time-compressed speech at different SNRs was assessed. Both background noise and time compression deteriorated speech intelligibility. The overall results showed that speech was less intelligible at higher speech rates. The calculated SRTs of the modeled discrimination functions increase with increasing speech rate. They rise from -7.1 dB for the original speech (Wagener et al., 1999a) to 3.4 dB with PSOLA and to 4.8 dB with Mach1 for speech compressed to 25% (932 syll/min) of its original length. These findings are consistent with previous studies by Tun (1998) and Liu and Zeng (2006). In order to compare the results of this paper with their results, the slopes and intercepts of a linear regression analysis described by Tun (1998) were used to calculate the respective SRT values of younger normal-hearing listeners as follows. Tun (1998) measured a SRT with babble noise of -5.4 dB for original speech, -4.3 dB, and -2.5 dB for speech time compressed with a factor of $\rho = 80\%$ (301 syll/min[1]) and $\rho = 60\%$ (399 syll/min[1]), respectively. Liu and Zeng (2006) presented speech spectrum shaped noise together with speech at normal rate and time compressed to 75% of its original length (calculated from the time compression given by the authors; no information about the speech rate is given by the authors). For the original speech, the SRT was -8.8 dB and for time-compressed speech, the SRT was -4.1 dB. Because the aim of the study was to provide a speech-in-noise test with positive SRT results, lower time-compression factors (corresponding to an expected higher speech rate) were employed as compared to Tun (1998) and Liu and Zeng (2006). The measured SRTs for the PSOLA and Mach1 algorithm with a time-compression factor of $\rho = 30\%$ (777 syll/min) amounted to -0.63 and 0 dB, respectively. As a result, positive SRTs were only obtained when speech was presented at a time-compression factor below $\rho = 30\%$ (777 syll/min).

In addition to the increasing SRT with decreasing time-compression factor, the current results also showed a decline of the slope of the discrimination function after reduction of the time-compression factor. The slope decreases from 17.1%/dB (Wagener et al., 1999a) for the original speech material to 4.3%/dB and 2.9%/dB for speech compressed to 25% of its original length with PSOLA and Mach1, respectively. Tun (1998) and Liu and Zeng (2006) documented different alterations of the slope. Whereas Tun (1998) observed smaller reductions of the slope (i.e., 7.1, 6.7%/dB to 6.3%/dB for original speech and speech time-compressed with a factor of $\rho = 80\%$ and 60%, respectively), Liu and Zeng (2006) measured an increase of the slope (i.e., 10.2%/dB to 12.8%/dB for the original and the time compressed speech ($\rho = 75\%$), respectively). The differences of the slopes between this study and those of Tun (1998) and Liu and Zeng (2006) likely resulted from use of smaller time-compression factors and speech material with low redundancy and low predictability in this study.

The shallow slope of the discrimination function is at least partly due to a high variability in intelligibility across the time-compressed speech items. This clearly limits the accuracy of the speech-in-noise test and its ability to discriminate across different effective SNR conditions

(Wagener et al., 1999b). Since these variations across speech items may have aroused from a serial word position effect which depends on the choice of the time-compression algorithm, the effect of both factors on the accuracy of a speech-in-noise test with time-compressed speech are discussed in the following Sections 2.4.3 and 2.4.4.

2.4.3 The effect of serial word position on speech intelligibility

First, it should be mentioned that the presented SNRs were selected for the measurement of discrimination functions of the complete sentences and not for the investigation of word groups. To fit functions successfully, intelligibility accuracy should vary between 20% and 80%. Missing data for some word groups at low or high intelligibilities might have generated unreliable fits of the discrimination function, SRTs, and slopes. Nevertheless, analysis of the word position in the sentences employed here with the fixed structure (i.e., name-verb-numeral-adjective-object) led to different patterns of performance between the positions in the sentences with the exception of verbs and adjectives. Names showed the highest intelligibility and verbs and adjectives were less intelligible. This pattern of results could be described as a serial position effect. Serial position effects refer to the U-shaped recall accuracy pattern that arises as a result of an item's position within a sequence such that recall is better for items presented at the beginning (primacy effect) and end (recency effect) of a list than for intermediate items (e.g., Murdock Jr., 1962; Glanzer and Cunitz, 1966). However, Wagener and Brand (2005) detected only a small serial position effect depending on the free recall of the sentences for the original OLSA. Names showed only a minimally higher intelligibility than the other words. Hence, the much larger word position effect observed in the current data using the time-compressed OLSA can not only be explained by the classical word position effect due to, e.g., attention or serial masking which is supposedly similar for the original and the time-compressed OLSA and is largely independent of semantic structure. Instead, the large word position effect observed here might at least partly be a consequence of the fixed semantic structure of the test.

Aaronson et al. (1971) observed serial position effects for unprocessed and time-compressed speech with a sequence of seven digits. Pauses were included between the time-compressed digits, thus, the signals of the unprocessed and time-compressed speech were presented at the same point in time. Aaronson et al. (1971) determined fewer errors in the time-compressed sequences compared to the unprocessed speech and suggested a positive effect of the enlarged pause duration. The advantage of the enlarged pauses between the words was intentionally eliminated for the current investigation. As a result, the test was more difficult so as to yield performance at positive SNRs. Notably, the observed serial position effect in this study was likely enhanced in light of the presentation method which guided the participant's attention to the names and objects. A countdown was presented before each sentence that indicated the beginning of the sentence. The countdown was added because the signals were very short at high speech rates. It may also be hypothesized that listeners could potentially miss the beginning and therewith the whole sentences without a countdown. Sukowski et al. (2008) investigated the effect of an introductory sentence on speech intelligibility, and did not rule out completely the benefit arising from the introductory sentence. Additionally, the participants' knowledge of the fixed structure of the sentences increased the possibility to understand the

initial name and the last object because they have a precise position within the presentation. The temporal position of verb, numeral and adjective is less determined, because the time-compression smeared the perception of the word boundaries within the sentence, distorted the information about the position of a word within the sentence, and as a result, may have decreased their intelligibility. Taken together, the current study cannot provide an explicit reason for the comparatively large serial word position effect which might be due to a combination of several factors. The possible primacy and recency effect cannot be distinguished from the effects of the fixed semantic structure of the sentences or the effects of the measurement procedure.

The serial word position effect was more prominent for the Mach1 algorithm compared to PSOLA. This was caused by non-uniform processing and the adaptation of the algorithm. The adaptation generated longer names compared to PSOLA and therefore a higher intelligibility was observed (see Section 2.3.3, Figure 2.6). Then, the non-uniform processing is expected to reduce subsequent sentence parts following names more than PSOLA to obtain the global compression. Based on the idea of Mach1, the beginning of a sentence consisted of parts with high audio tension while the parts within the sentences showed less audio tension. Such time-dependent compression additionally affected the serial word position effect.

Note that an alternative way of administering the test is to use sentence-based scoring rather than word-based scoring. This would mean a shift of the discrimination function towards higher SNR values or less severe time compression to achieve 50% intelligibility, respectively. Even though the disadvantages of time compression (including the high serial word position effect) would be alleviated and the discrimination function would be steeper, the measurement procedure would be considerably less efficient because sentence-based scoring would mean that the intelligibility would be determined by the least intelligible word. In contrast, word-based scoring sums up the intelligible words of a sentence and results rely on the intelligibility of a larger number of speech items.

2.4.4 A comparison of uniform and non-uniform time-compression

Even though the results cannot be generalized for all uniform and non-uniform algorithms, the objective and perceptual data favors the use of the uniform PSOLA algorithm for the application of time-compressed speech in speech-in-noise tests at positive SNRs. The use of PSOLA led to fewer deviations of the objective measures from their respective expected values (see Section 2.4.1). The comparison of the perceptive measurements showed that PSOLA produces speech signals with a higher intelligibility than the non-uniform processing by Mach1. The SRTs obtained with PSOLA exhibited lower values than the SRTs obtained with Mach1 and as a result PSOLA approached positive SNRs only for more extreme time compression (very low time-compression factors). This would argue in favor of Mach1 as positive SNRs could be reached easier compared to PSOLA. However, sentences modified by the PSOLA algorithm yielded discrimination functions with steeper slopes indicating less variable intelligibility of the speech items. Interestingly, Mach1 showed a higher intelligibility of names than PSOLA although the overall intelligibility of Mach1 was lower and therefore the intelligibility of certain word classes (e.g., verbs) was lower than the intelligibility using PSOLA. According to the

signal processing of Mach1, the names had a longer duration than with PSOLA leading to a higher intelligibility above 60% at ρ = 25% (932 syll/min). In contrast, PSOLA resulted in a steeper slope and a more similar intelligibility of the words. The results measured with PSOLA are more consistent compared to the results of Mach1. Therefore, PSOLA is expected to produce a higher accuracy in a speech-in-noise test with time-compressed speech. An informal subjective rating of three listeners confirmed a higher quality of signals processed with PSOLA than with Mach1. These results are in contrast to Covell et al. (1998) and He and Gupta (2001), who found an advantage for Mach1. Covell et al. (1998) presented signals processed with time-compression factors ranging between ρ = 24% and 39% (calculated from the time-compression rates given by the authors, 546-942 syll/min[1]) and therefore used the same range of time-compression factors and comparable speech rates as in the current study (ρ = 25%-40% and 583-932 syll/min[1]). He and Gupta (2001) used time-compression factors of either ρ = 40% or 67% (speech rate: 623-714 syll/min[1] or 374-428 syll/min[1]), which are partly higher than the presented factors. Therefore, all studies presented speech faster than the natural fast speech recorded by Janse (2003, Chapter 5, Section 3), which reached a time-compression factor of 72% (510 syll/min, calculated from the stated speech rate). Listeners of investigations according to Covell et al. (1998) and He and Gupta (2001) showed better comprehension and preference for Mach1 compared to a uniform synchronous overlap-add algorithm. These differences might be caused by the performance of the synchronous overlap-add algorithm used by those authors which is possibly poorer than the PSOLA processing. Other reasons might be the speech material, which consisted of short German sentences instead of English dialogues and monologues with pauses used by Covell et al. (1998) and He and Gupta (2001). The advantage of pauses for the Mach1 algorithm was previously discussed by Janse (2003). These pauses were highly compressed and the remaining speech did not have to be shortened as much as the signals processed with the uniform algorithm in order to achieve the same time compression. The current speech material consists of single sentences with no pauses. For the processing of the speech material five sentences were concatenated with very short pauses between (see Section 2.2.2.2). It is suggested that the algorithm was not able to use the advantage of these short pauses. Mach1 had to compress the speech to the desired time-compression factor like the uniform PSOLA algorithm. In addition, the listeners conducted different tasks. Covell et al. (1998) investigated comprehension by asking questions about the content of the presented speech. Participants in the current study listened to sentences with low redundancy, some of which did not convey meaning. They repeated every understood word and intelligibility was assessed. In contrast to the results of Covell et al. (1998) and He and Gupta (2001), the documented advantage of PSOLA is supported by the findings of Adank and Janse (2009). They described that natural-fast speech is more difficult to perceive than artificially time-compressed speech obtained with PSOLA. Natural-fast speech shows greater spectrotemporal variation than artificially time-compressed speech, when compared to speech presented at a normal rate. Such variation arises from articulatory limitations imposed when speaking at fast rates.

Summarizing, time compression deteriorates speech intelligibility. The non-uniform Mach1 algorithm induces greater changes and spectrotemporal variations of the time-compressed speech signal than the uniform PSOLA algorithm, especially if the signals are compared to the original

speech at normal rate. These deviations from the original speech increase the difficulty for listeners to recognize and process time-compressed speech, and therefore result in poorer intelligibility.

The current investigation was motivated by the need to develop speech-in-noise tests that yield performance thresholds at positive SNRs relevant for hearing aid processing schemes. It was observed that positive SNRs are achievable with high time compression only. This means that more realistic SNRs can be reached at a cost of presenting speech at an artificially high and less realistic speech rate. It is expected that in future work, two factors will lead to higher (i.e., less obtrusive) time-compression factors. First and as mentioned before, the use of sentence scoring will increase the time-compression factor because the intelligibility of an entire sentence is dependent on the least intelligible word. Second, measurements with older participants will shift the time-compression factor to higher values. Apart from less realistic time-compression factors, a speech-in-noise test does not represent a realistic hearing situation, because sentences were created from a matrix with low context and the background noise has the same spectrum as the speech. But it is a very sensitive test procedure, well defined, and therefore easy to reproduce in scientific settings and in standard hearing evaluation procedures. Therefore, the possible disadvantage of low time-compression factors is counterbalanced with several advantages of such a test.

In addition to previously mentioned factors, future investigations of time-compressed speech-in-noise tests should consider the following. First, if time-compressed material is used for the evaluation of hearing instruments, it is necessary to also investigate the effects of hearing instrument processing of time-compressed speech. Furthermore, the intelligibility of time-compressed speech differs for younger and older normal hearing listeners, as well as hearing-impaired participants (e.g., Schneider et al., 2005; Gordon-Salant and Friedman, 2011; Adams et al., 2012). Additionally, previous investigations have suggested that training can influence speech intelligibility performance when listening to time-compressed speech (e.g., Adank and Janse, 2009; Gordon-Salant and Friedman, 2011).

2.5 Conclusions

This study compared two time-compression algorithms, namely, the uniform PSOLA (as implemented in Praat) and the non-uniform Mach1. The goal was to investigate their potential application in a speech-in-noise test in order to evaluate speech perception at positive SNRs. On both the objective and perceptual measures, an advantage was observed for PSOLA compared to Mach1. The uniform PSOLA exhibited less variation of the objectively determined deviations of the targeted time compression, changes of the phoneme duration, spectra and modulation spectra. Thus, these signals are believed to be more similar to original speech at a normal rate and showed a higher intelligibility than Mach1. The more similar intelligibilities for words measured with PSOLA will likely result in more reliable performance when used in a speech-in-noise test. However, very high speech rates with a time-compression factor below 30% were necessary to reach 50% intelligibility at positive SNRs. Additionally, discrimination

functions measured with time-compressed speech showed a shallower slope compared to original speech material indicating limitations of the speech-in-noise test and its ability to discriminate across different effective SNRs.

Acknowledgements

We would like to thank Lüder Bentz, Micha Lundbeck, Maxi Susanne Moritz, and Stefan Nitzschner for their support on data collection and Simon Tucker (Department of Computer Science, University of Sheffield) for providing a MATLAB script for the Mach1 algorithm. The first author was supported by Phonak AG to perform her work on this project.

Parts of this work were presented at the International Symposium on Auditory and Audiological Research (2011) in Nyborg, Denmark.

3

Normal and time-compressed speech: How does learning affect speech recognition thresholds in noise?

Objective: Learning effects were investigated for the German Oldenburg sentence test (OLSA) with original and time-compressed fast speech in noise. Intra- and inter-session as well as transfer effects from time-compressed to original speech were analyzed. *Design:* Normal-hearing and hearing-impaired participants completed six lists of the OLSA during five sessions. *Study Sample*: Two groups of normal-hearing listeners (24 and 12 listeners) and two groups of hearing-impaired listeners (9 listeners each) performed the test with original or time-compressed speech. *Results:* In general, original speech resulted in better speech recognition thresholds than time-compressed speech. Thresholds decreased with repetition for both material sets. The largest improvements were observed within the first measurements of the first session, indicating a rapid initial adaptation phase, and were larger for time-compressed than for original speech. Additional inter-session improvements were present, indicating a longer phase of ongoing learning, especially for time-compressed speech. However, no transfer of learning benefits from time-compressed to original speech was observed for normal-hearing participants. *Conclusions*: Results are consistent with the Reverse Hierarchy Theory. It is recommended that learning effects are considered during the initial adaptation phase when administering the OLSA. Furthermore, prolonged inter-session learning effects should be taken into account, especially for time-compressed speech.

Adapted from:
Schlueter, A., Lemke, U., Kollmeier, B., Holube, I. (2014) "Normal and time-compressed speech: How does learning affect speech recognition thresholds in noise?", International Journal of Audiology, submitted.

3.1 Introduction

Perceptual learning while performing speech audiometry is of considerable importance in audiology: When hearing abilities are assessed with a speech test, optimal performance requires learning of the speech recognition task prior to data collection. Ideally, the training required for a stable level of performance is achieved quickly, so that the test is completed in a time efficient fashion. With matrix sentence tests, for example, it is typically recommended that listeners complete at least one or two training lists prior to the actual measurement (e.g., Wagener et al., 1999a; Hochmuth et al., 2012, see Section 3.1.1).

Learning effects have also been reported for speech recognition tests using time-compressed speech (e.g., Dupoux and Green, 1997, see Section 3.1.2). Time-compressed speech consists of original speech that has been processed with a time-compression algorithm in order to produce speech with a faster speaking rate. These algorithms retain the pitch of the speech and can achieve high compression of the signals. Furthermore, time-compressed speech used in a matrix test is currently studied to measure speech recognition thresholds (SRTs) at higher signal-to-noise ratios (SNRs, Schlueter et al., 2014b). The SRT describes the SNR at which 50% recognition is reached. With time-compressed speech, the SNRs presented are closer to conversations in everyday life that take place in noisy environments at positive SNRs (Olsen, 1998; Smeds et al., 2012). For speech tests with time-compressed speech, it is unclear whether the learning process is comparable to original speech. Hence, time-compressed speech appears to be a good tool to study the influence of different parameters on learning effects in speech audiology. This study therefore considered learning effects in time-compressed German matrix sentences. Furthermore, results were discussed from the perspective of a theory concerning perceptual learning.

3.1.1 Learning effects for original speech presented in a matrix test format

Learning effects were observed in speech audiometry. The German Oldenburg sentence test (OLSA, Wagener et al., 1999c) as well as its counterparts in different languages, such as the Swedish Hagerman test (Hagerman and Kinnefors, 1995), the Danish Dantale II (Wagener et al., 2003), Spanish or Polish matrix tests (Hochmuth et al., 2012; Ozimek et al., 2010) use sentences that were created from a matrix of 50 words by random selection. For SRT measurements, the sentences are usually presented at a normal speech rate in background noise at a fixed level and the level of the speech is adaptively adjusted to reach SRT.

There are several advantages of speech tests with a matrix structure, including the high number of different sentences that can be created from the limited set of material. These sentences show relatively low redundancy for specific words and word combinations. Also, most sentences do not make sense and are therefore characterized by low predictability. Moreover, typically, sentences are not used twice. Consequently, the chance of learning entire sentences is minimized. On the other hand, matrix tests are expected to be prone to learning effects due to the limited speech material. Also, learning effects are probably more pronounced at the beginning of the measurements due to the participants' need for familiarization with the unaccustomed test situation and with the sentences of low-predictability.

For matrix tests with original speech, learning effects clearly follow a learning curve, with more pronounced performance improvements within the first repeated measurements and decreasing improvements for further repetitions. In the following, this learning effect due to repeated measurements within a session is called *intra-session learning effect*. Hagerman and Kinnefors (1995) documented a mean decrease of SRT by 0.13 dB for normal-hearing participants (NH) and 0.07 dB to 0.5 dB for hearing-impaired participants (HI) for each repetition within a single session. Wagener et al. (1999a, 2003), Hernvig and Olsen (2005), and Hochmuth et al. (2012) measured similar or even higher learning effects and recommended the routine use of training lists before measuring speech recognition. For the OLSA, and based on measurements with young NH Wagener et al. (1999a) suggested one or two lists of training (i.e. up to 60 sentences). After the presentation of two lists, the learning effect of about 2 dB (observed between the first and last measurement of six consecutive lists) was reduced and a test accuracy of about 0.5 dB (SRT difference between following measurements after two training lists) was achieved. The size of the learning effect seems to be small and irrelevant when compared to speech recognition improvements with hearing aids for speech in quiet, but is relevant with regard to SRTs for speech in noise, which cannot always be improved by hearing aids. According to the German governmental guidelines for medical aids (Bundesministerium der Justiz, 2012), SRTs are expected to improve by at least 1.5 dB SNR for a beneficial hearing aid fitting. The required minimum improvement is below the reported accuracy of the OLSA when conducted without training (Wagener et al., 1999a). As a result, when conducting the OLSA without training lists, an improved SRT might falsely be interpreted as a benefit due to a hearing aid fitting, while in fact the improvement is the result of learning.

In addition to intra-session effects, learning effects can also be observed when comparing data collected on different days (i.e. *inter-session learning effect*). This effect is of interest especially for scientific studies, where repeated measurements of different settings (e.g. hearing aid settings in different listening situations) are conducted in different sessions and small differences between settings need to be resolved. After the initial training, Hagerman (1982) as well as Hernvig and Olsen (2005) documented an additional small decrease of SRT values in two consecutive sessions. Wagener and Brand (2005) measured an inter-session learning effect (median test-retest difference of SRT across all settings, two sessions) of 0.67 dB and 0.2 dB for NH and HI, respectively.

The learning effects described seem to be dependent on the hearing ability of the participants. Hagerman (1984) and Hagerman and Kinnefors (1995) described a negligible learning effect for HI with SRT values less than 0 dB. However, they stated that a learning effect should be considered for participants with SRT values larger than 0 dB. In contrast, Wagener and Brand (2005) reported a learning effect for both NH and HI, although the effect was smaller for the HI.

Based on previous research observing intra-session learning effects, it is generally recommended that training lists be administered before measuring speech recognition. This recommendation is especially important for clinical purposes, because daily practice generally provides one session with a small number of repeated measurements e.g. for comparison of hearing aids. Until

now, intra-session learning was not investigated for repeated measures within different consecutive sessions on different days. Such a procedure is commonly used in scientific studies using complex test protocols, when e.g. different parameter settings of hearing aids are tested in acoustically different environments. Little is known as to how scientific investigations with several sessions of repeated SRT measures are affected by learning. One possibility is that a participant who completes training in an earlier session displays a different pattern of intra-session learning effect at following sessions. Moreover, the recommendation of administering training lists is based on studies of young NH rather than older HI. However, the majority of hearing aid users is older. Also, older HI often take part in scientific research and they may exhibit a different pattern of performance compared to young HI and NH. Inter-session results suggest an ongoing learning of the speech material and the test procedure. Importantly, the studies cited conducted measurements in a limited number of sessions. Therefore, they provide no indication whether and when learning discontinues and whether SRT values remain stable over longer study periods. This is especially of interest for scientific investigations in which small differences between e.g. parameter settings of hearing aid algorithms need to be resolved and therefore a highly accurate test is necessary. Finally, available data on the comparison of learning effects in NH and HI showed no consistent tendency. It therefore remains unclear whether differential learning effects due to hearing impairment (and age) have to be taken into account or whether different training protocols dependent on previous learning and hearing loss have to be considered. Moreover, the cited studies did not discuss the results with regard to different mechanisms underlying the learning process, such as perceptual learning that relies on different stages of auditory information processing, as has been suggested by the Reverse Hierarchy Theory (RHT, Nahum et al., 2008, 2010; Adank and Janse, 2009; Banai and Lavner, 2012; Ahissar et al., 2009).

3.1.2 Learning effects for time-compressed speech

Beside matrix tests using original speech, time-compressed speech becomes relevant for the investigation of speech at SNRs of everyday life (Schlueter et al., 2014b). Time-compressed speech material shows learning effects and the combination of speech rate and matrix material is expected to emphasize these effects. The literature concerning learning of time-compressed speech reports intra-session learning effects. Dupoux and Green (1997), Peelle and Wingfield (2005), Adank and Janse (2009) and Adank and Devin (2010) consistently documented increasing comprehension of speech within the first sentences. Dupoux and Green (1997), for instance, described improved recognition after the presentation of only five or ten time-compressed sentences, indicating e.g. “a short term adjustment to local speech rate parameters” (p. 926). In contrast, Golomb et al. (2007) observed inter-session learning effects and measured an increase of recall accuracy between the first and second session about one week later for both younger and older participants. Additionally, for time spent listening to time-compressed speech, Gordon-Salant and Friedman (2011) showed increasing speech recognition with increasing hours per week.

Different factors seem to affect the learning of time-compressed speech. Dupoux and Green (1997) described longer learning effects for speech that had higher compression. Although older

listeners achieved a smaller recall performance of time-compressed speech compared to younger participants, initial adaptation to time-compressed speech was independent of age (Peelle and Wingfield, 2005): rate and magnitude of initial learning effects were comparable for younger and older participants. Unlike young participants, older participants' learning was dependent on compression factors; they were impaired in transferring learning effects from one speech rate to another. Golomb et al. (2007) did not confirm this result, but found for older participants, beyond an initial plateau, learning effects comparable to younger listeners. The absence of age-related effects indicated learning processes that were resistant to cognitive decline, or, alternatively, compensatory mechanisms that were applied by older participants. In order to explore the level of representations used for perceptual learning, transfer of learning was also studied for different settings. Adank and Janse (2009) documented a transfer of learning from time-compressed to natural fast speech, but not for the reverse presentation of signals. Referring to the RHT, they argued "that participants relied more on lower level acoustic cues for conditions that require more attention [natural fast speech], i.e. those that were more difficult" (p. 2656). Pallier et al. (1998) and Sebastián-Gallés et al. (2000) showed a transfer for different languages and time-compressed speech. They indicated a prelexical processing level for learning of time-compressed speech.

Additionally, Banai and Lavner (2012) suggested that intra- and inter-session learning effects reflect two phases of the perceptual learning process. They based their findings on verification and lexical decision tasks that non-native listeners executed on time-compressed Hebrew sentences and words. Participants' performance for five different conditions was measured in pre and post training sessions. To analyze the generalization of learning, different conditions were presented using trained and untrained sentences or words of the same speaker as well as trained sentences of a different male or female speaker. Banai and Lavner (2012) documented that learning of time-compressed speech continued after adaptation to few sentences and throughout training sessions. Since participants only showed generalization of the learned to the trained/untrained sentences and different male speakers, Banai and Lavner (2012) related their findings to the RHT and concluded that during an initial phase of brief adaptation, perception relied on abstract acoustic representations at higher levels of the auditory pathway. Also, they interpreted the subsequent learning process as using spetro-temporal representations at lower levels with increasing precision.

With respect to speech recognition tasks in noise, the results of Banai and Lavner (2012) suggest a transfer of learning at higher levels during the initial phase of learning. However, it is further assumed that the transfer of learning does not take place during the second phase, during which specific lower-level representations are used. Based on this assumption, participants who have been trained with time-compressed speech should reach a better threshold for subsequently-presented original speech than naïve participants, because they have adapted their higher level-representations. Furthermore, participants are expected to exhibit worse thresholds than those participants who developed learning during prolonged practice of original speech, because they cannot rely on lower-level representations of original speech. Comparison of the thresholds should allow a distinction between the initial phase of brief adaptation and the subsequent learning with relevant spectro-temporal representations at lower levels.

3.1.3 Research questions of the current study

To investigate potential learning effects, the OLSA was conducted in repeated measures within consecutive sessions. NH and HI listened to the original speech material and to time-compressed sentences. It was expected that learning of the speech presented progresses throughout an initial general phase and a subsequent prolonged and more stimulus-specific phase. Consequently, SRT values should improve with repetition within and between sessions. These principle mechanisms of learning were expected to be the same for NH and HI as well as for original and time-compressed speech. Primarily, the consequences of learning mechanisms for speech audiometry and hearing instrument evaluation were analyzed. For the reliable usage of a matrix test with original and time-compressed speech, it is necessary to know for both NH and HI whether reliable results are achievable within and across sessions and, if so, how many training lists are necessary. Furthermore, observed results need to be related to a model of perceptual learning. This was facilitated by additional studies on the transfer of learning between original and time-compressed speech.

3.2 Methods

3.2.1 Participants

Participants were assigned to the four groups outlined in Table 3.1, i.e. NH and HI that were presented either with original or with time-compressed (TC) speech material. Based on pure-tone audiometry testing, NH showed hearing levels of 20 dB HL or better at all octave frequencies between 0.25 kHz and 8 kHz for both ears. Thresholds of the HI are depicted in Figure 3.1. All listeners had German as their native language and no prior experience with the OLSA. They were paid a small amount for their participation, to compensate for their expenses.

Table 3.1: Characteristics of participating groups and test conditions.

Group	**Age**	**Number of participants**	**Gender**	**Hearing ability**	**Test condition**
NH-Original	Mean: 22 years, 19-27 years	24	21 female, 3 male	Normal	Original speech
HI-Original	Mean: 71 years, 58-78 years	9	3 female, 6 male	Impaired	Original speech
NH-TC	Mean: 22 years, 19-27 years	12	6 female, 6 male	Normal	Time-compressed speech, compression: 30%
HI-TC	Mean: 66 years, 58-75 years	9	6 female, 3 male	Impaired	Time-compressed speech, compression: 50%

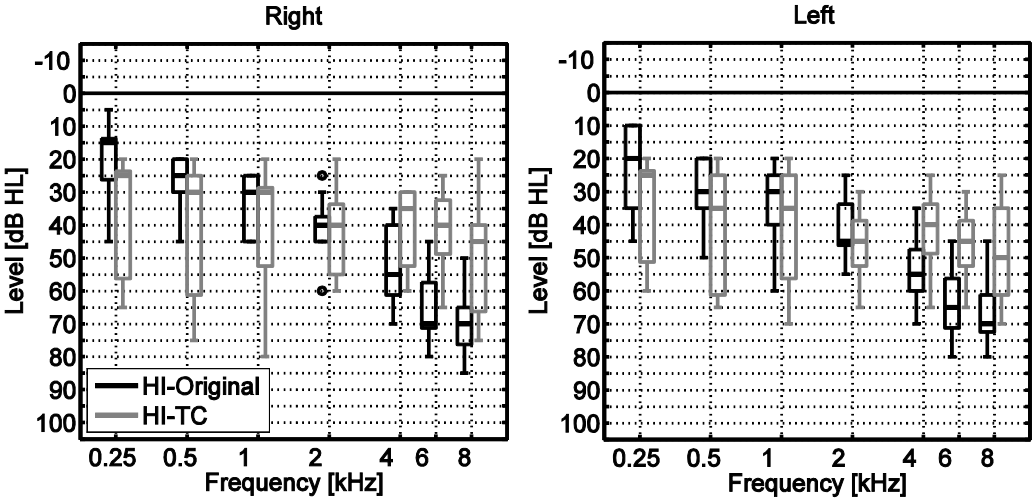

Figure 3.1: Results of pure-tone audiometry testing for the right and left ear of the hearing-impaired participants (HI), who listened to original (black) or time-compressed (TC, gray) speech material.

3.2.2 Signals

Speech recognition performance was determined using the German matrix test (OLSA, Wagener et al, 1999a). Its sentences have the same structure (name, verb, numeral, adjective, object). Sentences were generated from a random selection of one out of ten words for each structural element of the sentence. The selection process produced 100 sentences with low redundancy, for instance "Peter kauft zehn nasse Messer." ("Peter buys ten wet knifes."). The test consists of 40 lists with 30 sentences each with equal speech recognition and included a background noise stimulus. The noise resulted from a superposition of all sentences and has the same long-term spectrum as the speech (Wagener et al., 1999a).

For the presentation of time-compressed speech with original fundamental frequency, sentences of the OLSA were processed with a pitch synchronous overlap-add procedure implemented in the software Praat (Boersma and Weenink, 2009). This approach analyzes the pitch of a speech signal, sets pitch marks, and segments the original signal into windowed frames. Afterwards, segments at regular intervals are deleted and the remaining segments are concatenated to the time-compressed signal. Position and number of deleted segments are dependent only on the time-compression factor and are not influenced by speech characteristics. This compression is different to natural fast speech in which e.g. pauses and vowels are compressed the most as compared to other parts of the speech (e.g., Covell et al., 1998). Schlueter et al. (2014b) showed that only very small differences between long-term spectra of time-compressed speech processed with Praat and original speech. Among other reasons, they recommended Praat for speech tests because it showed signal quality comparable to the original speech material. For the speech tests in this study, sentences were compressed to 30% (for NH) and 50% (for HI) of their original length. Compression was selected after pretests, in which younger NH reached 50% speech recognition at a compression of about 30% and 1 dB SNR. HI showed equal recognition at a compression to 50% and SNRs of 1 dB or higher (Schlüter et al, 2013).

3.2.3 Measurements

After brief anamnesis, otoscopic examination, and pure-tone audiometric testing, the OLSA measurements were performed in an acoustically-insulated audiometric booth. Signal presentation was controlled by a laptop running the Oldenburg Measurement Application (Hörtech, Oldenburg, Germany). Signals were routed through a sound card (Fireface 400, RME, Audio AG, Haimhausen, Germany) and a headphone amplifier (HB 7 Headphone Driver, Tucker Davis Technologies, Alachua, FL) to headphones (HDA 200, Sennheiser, Wedemark-Wennebostel, Germany). The noise was presented at a fixed level of 65 dB SPL for NH. HI listened to a noise level dependent on their hearing threshold at both 0.5 and 1 kHz and the level correction was based on by the half-gain rule but without frequency shaping. The corrected overall presentation level l was calculated using Equation 3.1.

$$l = 65 + \frac{\text{hearing threshold}_{0.5\text{ kHz}} + \text{hearing threshold}_{1\text{ kHz}}}{4}\ \text{[dB SPL]} \qquad (3.1)$$

If HI complained about the loudness of the background noise, the presentation level was decreased in 5 dB steps until an acceptable level was achieved. The resulting median presentation level of the background noise was 77 dB SPL for HI (range: 70-85 dB SPL). The measurement of the SRT was conducted with an adaptive procedure, adjusting the level of sentences as a function of their recognition. Therefore, participants listened to one test list of OLSA sentences (30 sentences) in background noise and repeated orally what they understood without visual presentation of the speech matrix. The level of the speech was adaptively adjusted after each sentence and depended on the recognition of previous sentence, the target recognition of 50%, the slope of the assumed discrimination function and a rate of convergence (Brand and Kollmeier, 2002). The first sentence of each list was presented at 0 dB SNR or 10 dB SNR for normal and time-compressed speech, respectively. After each test list, the SRT was estimated using a maximum likelihood method (Brand and Kollmeier, 2002) and based on all SNRs and sentence-specific recognition scores within that test list.

All participants took part in five sessions, which were arranged at one-to-three day intervals. During each session, they performed six lists with randomly-selected list numbers of the OLSA. In the last session, participants of the two groups who listened in all previous measurements to the time-compressed speech condition conducted an additional seventh list of original speech material. The study was approved by the Ethical Committee of the Carl von Ossietzky University, Oldenburg.

3.3 Results

3.3.1 Normal-hearing participants

3.3.1.1 General results

Figure 3.2a provides an overview of the SRT results for six successively-measured lists at each of five sessions. These results are presented for groups of NH-Original and -TC. In general, a trend of decreasing SRTs within and between sessions was observed, clearly showing higher

SRTs for time-compressed than for original speech. Also, the overall improvement in SRT across measurements and sessions was larger for time-compressed than for original speech.

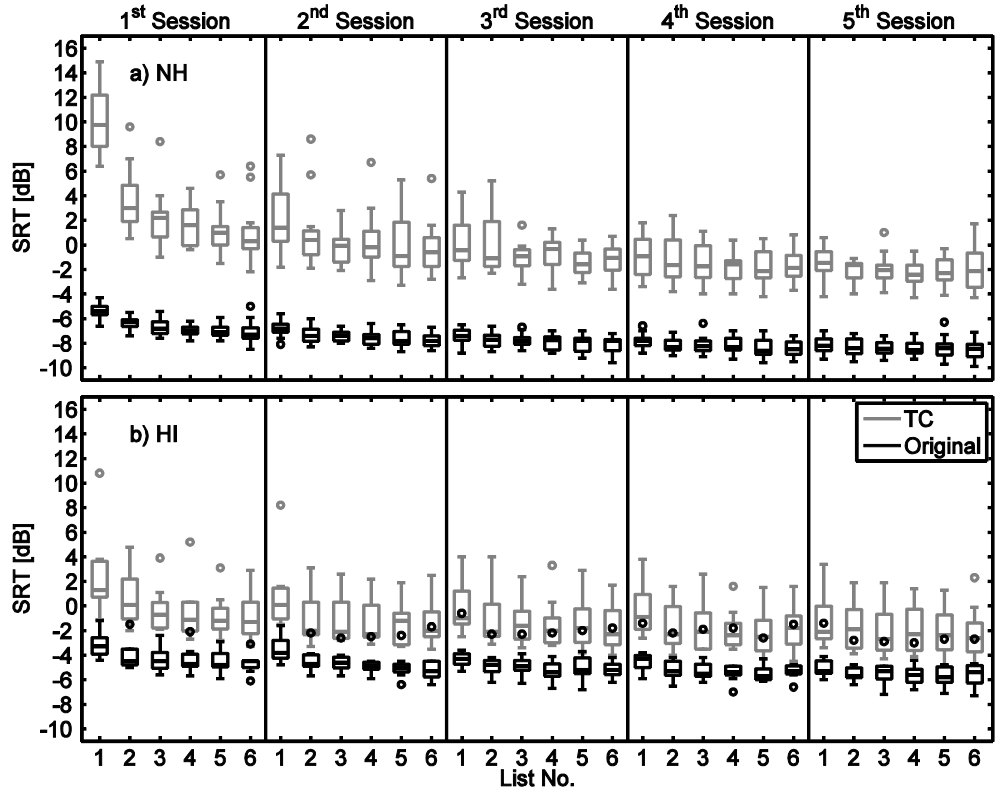

Figure 3.2: Boxplot of SRT values measured in five sessions with six successively-measured lists. Results are from groups of a) normal hearing (NH) and b) hearing impaired (HI) as well as original (black) and time-compressed (TC, gray) speech material.

3.3.1.2 Intra-session learning

For the first session, intra-session learning effects were studied using differences between the first and third lists. This difference is shown in Figure 3.3 and reflects the improvement that can be achieved by the performance of two recommended training lists (up to 60 sentences, Wagener et al., 1999a). The results of NH-Original (see Figure 3.3a) can serve as a reference for the typical improvements to be expected when applying the OLSA in the normal clinical setting of one single session. In Figure 3.3a, NH achieved smaller improvement for original (median difference of SRTs: NH-Original 1.2 dB) than for time-compressed speech (median difference of SRTs: NH-TC 7.8 dB) and these differences were significant (U-Test; NH: $p < 0.001$).

Since no statistically-significant differences were found between results from fifth and sixth list for all sessions and groups (Wilcoxon test for all sessions and groups: $p > 0.078$), the mean of the fifth and sixth list served as a representation of the SRT at the end of a session. For all five sessions, intra-session learning effects after the third list were calculated by individual differences between results of the third list and the end of each session (mean of fifth and sixth list). Median differences for this measure of intra-session learning effects are shown in Table 3.2. In general, for NH this learning effect was observed to be largest for the first session and

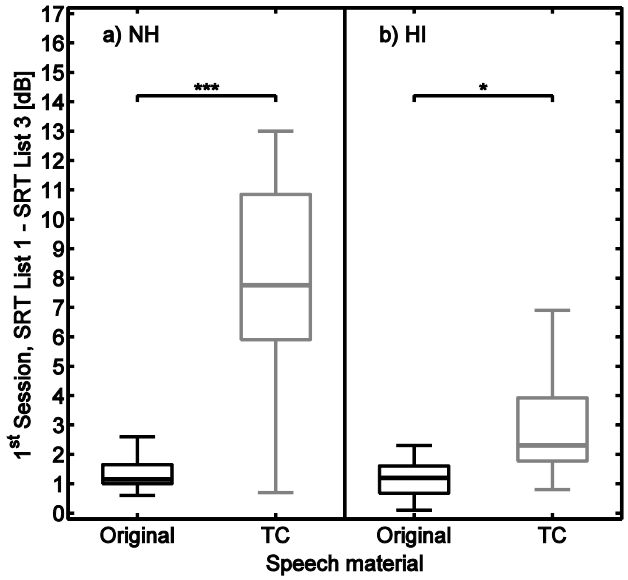

*Figure 3.3: Intra-session learning - SRT-differences between the first and third list in the first session for a) normal hearing (NH) and b) hearing impaired (HI) as well as original (black) and time-compressed (TC, gray) speech material. Significance was analyzed with a U-Test and significant p-values are displayed ($p < 0.05$: *; $p < 0.01$: **; $p < 0.001$: ***).*

Table 3.2: Intra and inter-session learning – Intra-session learning effects are described by the median SRT-difference between the third and the mean of fifth and sixth lists (and the interquartile range). For inter-session learning effects, mean SRTs of the fifth and sixth list were first computed for each participant and session. Subsequently, the median difference (and the interquartile range) between the first and fifth session were calculated. Results are displayed for normal hearing (NH) and hearing impaired (HI) who listened to original or time-compressed speech (Original, TC) as well as results of all participants and speech signals (All).

Learning	**Session**	**NH-Original**	**NH-TC**	**HI-Original**	**HI-TC**	**All**
Intra-session	1st	0.4 dB (0.6 dB)	1.3 dB (1.3 dB)	0.5 dB (1.0 dB)	0.7 dB (0.5 dB)	0.6 dB (1.0 dB)
	2nd	0.3 dB (0.3 dB)	0.2 dB (2.0 dB)	0.6 dB (0.7 dB)	0.3 dB (0.5 dB)	0.3 dB (0.6 dB)
	3rd	0.4 dB (0.7 dB)	0.3 dB (0.6 dB)	0.2 dB (0.6 dB)	0.1 dB (0.6 dB)	0.3 dB (0.7 dB)
	4th	0.3 dB (0.7 dB)	0.3 dB (1.4 dB)	0.2 dB (0.5 dB)	0.4 dB (0.6 dB)	0.2 dB (0.8 dB)
	5th	0.1 dB (0.9 dB)	0.2 dB (1.6 dB)	0.1 dB (0.5 dB)	0.3 dB (0.9 dB)	0.1 dB (0.9 dB)
Inter-session		1.4 dB (0.7dB)	3.0 dB (2.2 dB)	0.9 dB (0.7 dB)	1.2 dB (1.0 dB)	

further decreased for nearly all subsequent sessions. Additionally, for the first session, a larger improvement was detected for time-compressed than for original speech. However, the intra-session learning effects did not reach statistical significance between sessions when analyzed for each normal-hearing group separately (Friedman, NH-Original: $p = 0.058$, NH-TC: $p = 0.167$).

To determine the number of training lists after which no significant difference of the SRT occurred, Friedman tests were calculated. First, a statistical analysis was performed for the entire data of each session. Then SRTs of successive lists, beginning with the first list, were eliminated and the Friedman test was again performed. In doing so, the number of lists was identified after which no significant difference occurred compared to the subsequent lists. This analysis showed significant differences for all sessions and groups. In general, for NH-Original, significant differences were found for the first three lists of a session, while NH-TC showed significant differences for the first four lists. It has already been noted that results of the fifth and sixth list were not significantly different from each other.

3.3.1.3 Inter-session learning

Inter-session learning effects were assessed by analyzing the mean SRTs of the fifth and sixth list within the first to fifth session for each participant. These results are depicted in Figure 3.4a for NH and show decreasing SRTs for all sessions. In order to determine the session after which no significant difference occurred, Friedman tests were performed after eliminating successive sessions from the data beginning with the first session. NH-Original and -TC showed significant differences of the SRT for the first three sessions (Friedman, for all evaluations, $p < 0.019$). No significant difference in performance was observed to initially appear after the fourth session (Friedman, comparison of fourth and fifth session, NH-Original $p = 0.683$, NH-TC $p = 0.248$). In addition, Table 3.2 depicts the median difference between the first and fifth session for all groups. For time-compressed speech material, greater inter-session learning effects were found than for original speech. However, this effect was only significant for the comparison of NH-Original and NH-TC (U-test, NH: $p < 0.001$).

3.3.1.4 Transfer effects

Participants of the group NH-TC completed one additional, final list using the original speech material. Figure 3.5a shows the SRTs from this list, together with results of the NH-Original from their last measurement with original speech. NH-TC who trained on time-compressed speech during the experiment achieved higher SRTs for original speech than NH-Original, who trained on original speech during the experiment (Figure 3.5a). A U-test confirmed the results ($p < 0.001$).

3.3.2 Hearing-impaired participants

3.3.2.1 General results

Figure 3.2b shows also an overview of SRT results for the repeated measurements of HI. The general trend of decreasing SRTs within and between sessions was confirmed by HI as well as higher SRTs for time-compressed than for original speech.

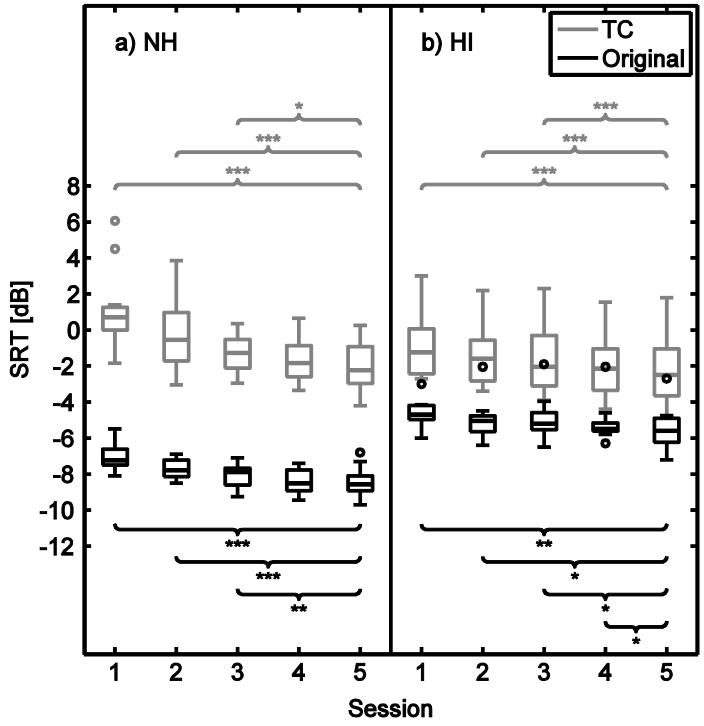

*Figure 3.4: Inter-session learning - Mean SRTs of the fifth and sixth list for five sessions performed by a) normal hearing (NH) and b) hearing impaired (HI) using original (black) and time-compressed (gray) speech material. Statistical testing consisted of a Friedman test and significant p-values are displayed ($p < 0.05$: *; $p < 0.01$: **; $p < 0.001$: ***). Brackets enclose all lists compared using the Friedman test.*

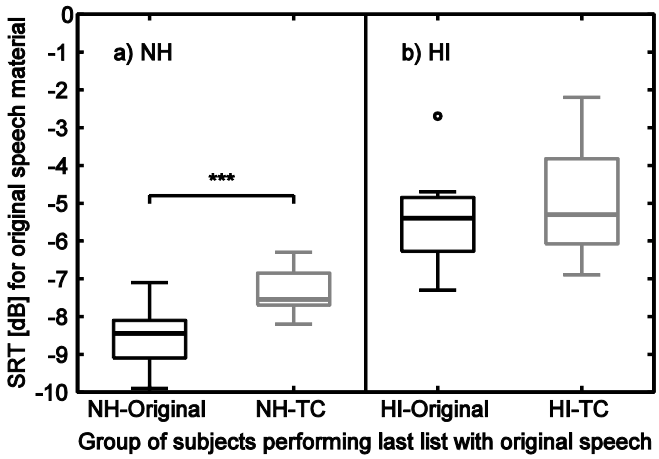

*Figure 3.5: SRTs for original speech. Shown are results of the sixth list within the fifth session of NH/HI-Original and results of a seventh list within the fifth session of NH/HI-TC. Significance was analyzed with a U-Test and significant p-values are displayed ($p < 0.05$: *; $p < 0.01$: **; $p < 0.001$: ***).*

3.3.2.2 Intra-session learning

Again for HI, differences between the first and third lists are shown in Figure 3.3b and reflect the improvement after two recommended training lists. As NH, HI obtained smaller improvement for original (median difference of SRTs: HI-Original 1.2 dB) than for time-compressed speech (median difference of SRTs: HI-TC 2.3 dB) and these differences were significant (U-Test; HI: $p = 0.012$).

As for NH, and because of no significant differences (Wilcoxon test for all sessions and groups: $p > 0.067$), the mean of the fifth and sixth list was used to represent the SRT at the end of a session. Again, intra-session learning effects (individual differences between results of the third list and mean of fifth and sixth list) were investigated and medians are listed in Table 3.2. As seen before for NH, learning effects for HI were largest for the first session and generally decrease for following sessions. When analyzed for each HI-groups separately, the intra-session learning effects did not reach statistical significance between sessions (Friedman, HI-Original: $p = 0.422$, HI-TC: $p = 0.082$).

As explained for the NH, Friedman tests were calculated to determine the number of training lists. For HI-Original and HI-TC, significant differences were only found for the first two lists of a session, not for any of the following lists three through six.

3.3.2.3 Inter-session learning

Inter-session learning effects (mean SRTs of the fifth and sixth list) of HI are depicted in Figure 3.4b and showed decreasing SRTs for all sessions. Again, Friedman tests were performed to determine the session after which no significant difference occurred. For HI-Original, results of all sessions were significantly different (Friedman, for all evaluations: $p < 0.044$). HI-TC results showed comparable significance to NH data. They also were significantly different (Friedman, for all evaluations: $p < 0.001$), with the exception of fourth and fifth session (Friedman, $p = 0.257$). In addition, Table 3.2 depicts the median difference between the first and fifth session for all HI-groups. The trend of greater inter-session learning effects for time-compressed than for original speech was not confirmed by the U-test ($p = 0.222$).

3.3.2.4 Transfer effects

As NH-TC, participants of the group HI-TC completed one additional, final list using the original speech material. Figure 3.5b shows these SRTs together with results from the last measurement with original speech of the group HI-Original. HI showed comparable SRTs, whether they listened to original or time-compressed speech during the experiment (U-test, $p = 0.566$).

3.4 Discussion

The presentation of time-compressed speech affected the recognition of the signals to a higher degree than original speech in NH and HI. Therefore, higher SRT values were measured with time-compressed speech than with original speech. Thus, SRT values of time-compressed

speech attained values that are closer to SNR values of conversations in everyday life (Olsen, 1998; Smeds et al., 2012).

Generally, intra- and inter-session learning effects were observed for repeated measures with original and with time-compressed speech presented in a German matrix test for both NH and HI. The SRTs improved over 30 different lists in five sessions indicating an initial phase of brief adaptation and a subsequent prolonged learning phase.

Additionally, during the first three presented lists, NH and HI achieved larger improvements for time-compressed speech than for original speech. These results indicate a different initial learning pattern for time-compressed speech material compared to results for original speech. The learning was observed in addition to the general familiarization with the measurement procedure and the formal structure of the sentences. Furthermore, the inter-session learning effect was more pronounced, showing larger improvements for time-compressed than for original speech, but this was only statistically significant for NH. NH-TC were able to transfer learning of the initial brief adaptation phase from time-compressed speech to their performance in original speech.

Methodological considerations with respect to speech audiometry applied repeatedly in scientific investigations are discussed here separately for speech tests with original and with time-compressed speech. Finally, the discussion will compare the present results to previous work and to the RHT of perceptual learning in general.

3.4.1 Consequences for speech audiometry

The observed learning effects have an impact on speech audiometry that involves a matrix test presenting speech in noise. In clinical applications of speech-in-noise tests, the number of repetitions and of sessions is commonly smaller than in scientific investigations and as a result, intra- and inter-session learning effects can affect these applications differently. In the following, results are compared to previous studies using speech tests for original and time-compressed speech separately.

3.4.1.1 Original speech

As expected, different SRT values for NH and HI were determined for original speech (Wagener et al., 1999a; Wagener and Brand, 2005). Nevertheless, the intra-session learning effect of NH and HI showed a similar median improvement of the SRT values of about 1 dB over the first three lists presented. Our results confirm Wagener et al. (1999a) and support their recommendation for the use of one or two training lists with at least 20 sentences each before a reliable measurement can be conducted. Commonly, the OLSA with original speech is conducted with lists of 20 sentences each. To increase accuracy, it is also possible to apply lists with a length of 30 sentences, as in the current study. Wagener et al. (1999a) did not differentiate between the length of lists and recommended up to 60 sentences (two lists with 30 sentences each) for the purpose of training. In doing so, they posited that such training facilitates familiarization with the measurement procedure and adaptation to the structure of the OLSA sentences. In the current study, although NH showed significance for the last four lists in each session (after

two training lists), differences were smaller than 0.4 dB. This result is in agreement with Wagener et al. (1999a), who observed an accuracy of about 0.5 dB for SRT measurements after initial training. HI showed a difference of 0.6 dB or smaller, but often only achieved significance for the first two lists within the third to fifth session. Therefore, the question remains whether one training list is sufficient for measurements at consecutive sessions. Still, since the current results were investigated with a high number of measurements within one session, learning effects for the first list within the third to fifth session were partly due to the large number of lists presented in the previous sessions and could be more pronounced for a smaller number of measurements per session. Hence, two training lists are recommended for each session, which also serves the common request to use identical protocols for both NH and HI.

Regarding inter-session learning effects, statistically significant improvements of SRT values were observed between nearly all sessions measured with NH, except for the fourth and fifth session. For HI, all sessions were significantly different, thus providing support for continuing perceptual learning. Statistical significance was probably due to uniform behavior and small inter-individual differences across participants. Hence, for practical purposes, this effect can likely be disregarded, because NH showed a larger median improvement of the SRT from the first to the fifth session of 1.4 dB (interquartile range: 0.7 dB) compared to HI of 0.9 dB (interquartile range: 0.7 dB). These results are in contrast to Hagerman (1984) and Hagerman and Kinnefors (1995). They observed a negligible learning effect for HI with SRT values smaller than 0 dB, while suggesting that learning effects have to be considered if participants have SRT values larger than 0 dB. According to Hagerman (1984) and Hagerman and Kinnefors (1995), the effect of learning was more important with poorer hearing ability. However, the results of the current study showed larger inter-session learning effects for NH than for HI. This inter-session learning effect confirms results by Wagener and Brand (2005), who found a larger difference of SRT values for NH than for HI. However, the improvement due to inter-session learning observed over five sessions was larger in the current study compared to Wagener and Brand (2005), who compared results of only two subsequent sessions. These results confirm the recommendation of two training lists as given above. The same training protocol is recommended, although the groups studied differed in age and hearing ability. This makes it clear that both age and hearing ability can be neglected for recommendation of training lists for matrix tests with original speech in both clinical and scientific practice. Nevertheless, for scientific practice it is important to account for training in each subsequent session, particularly when small differences in speech recognition and therefore a high accuracy of the speech-in-noise test are of interest. Otherwise, the improvement of e.g. 1.5 dB in SRT for speech in noise, which is supposed to underlie a beneficial hearing aid fitting (see Section 3.1, Bundesministerium der Justiz, 2012), will be generated due to learning.

As described, two initial training lists in each session are recommended. Nevertheless, the comparison of results between different sessions, especially in scientific studies with a high number of measurement repetitions, can cause difficulties because significant differences can occur due to learning effects. Therefore, it may be advisable to conduct studies with experienced listeners who have performed the test before and are well trained. Also, the number of

test conditions should be limited and/or the number of participants increased. These recommendations are in line with Hernvig and Olsen (2005) for the performance of a Danish matrix test. In addition to these suggestions, the test conditions should be randomized across participants and sessions.

3.4.1.2 Time-compressed speech

NH showed poorer SRT values than HI only for the first measurements with time-compressed speech and subsequent measurements were similar. This was achieved by the use of different rates of time compression for the two groups. NH listened to speech with higher compression (speech compressed to 30% of original length) than HI (speech compressed to 50% of original length). The different time compression was selected to reach comparable perception levels for both groups with regard to speech recognition and SRT values after the initial learning process.

Differences in intra-session learning effects were observed for NH and HI. The main learning effect occurred within the first two measurements and improvement during the first three lists was larger for NH than for HI. The learning effects of NH are comparable to those observed by Dupoux and Green (1997), who described a longer-lasting initial learning effect for faster speech. This was confirmed by the statistical comparison of all measurements within the sessions, where NH showed significant differences up to the fourth list and HI until the second list. Also, for time-compressed speech the number of training lists necessary to establish reliable results was investigated. As for original speech, when listening to time-compressed speech it is recommended that efficient test administration is best achieved for both NH and HI by completing two training lists. However, the accuracy of measurements with time-compressed speech is smaller than for original speech. The improvement after the first two training lists was about 1 dB for the first session and smaller than 0.4 dB for following sessions.

Also, the results on inter-session learning effects add to the observation of intra-session learning. These results were determined during five sessions, which were arranged at one-to-three day intervals. Significant differences of mean SRT for each session were calculated for NH and HI until the fourth session. The fourth and fifth session showed equal results. Effect of learning over all sessions was larger for NH than for HI. The results of intra- and inter-session learning showed an explicit adaptation to time-compressed speech and confirm the observations of e.g. Dupoux and Green (1997) and Golomb et al. (2007).

The differences found for time-compressed speech presented to NH and HI cannot only be explained by the hearing ability of participants. Previous studies documented a deterioration of recognition for fast speech associated with age. Older NH performed poorer in speech recognition tasks with time-compressed speech than younger NH (Golomb et al., 2007). However, the current study used an incomplete design; it was conducted with younger NH and older HI, whereas data of older NH are not available. Thus, no conclusion on the differential effects of hearing ability and age can be drawn. As for original speech, this does not affect the consequences for speech audiometry with time-compressed speech because the same training protocol is recommended for young NH and old HI. Likewise, smaller learning effects were observed for older HI than for young NH. Therefore, it is expected that the recommended protocol also is suitable for older NH and younger HI.

A similar set of recommendations is given when completing a matrix test with either time-compressed or original speech. Specifically, two training lists should be performed before the measurement of the SRT with time-compressed speech in each session. Since the results of different sessions were significantly different for time-compressed speech, SRT values measured in different sessions should not be compared. It is therefore advisable to design the study with measurements conducted within one single session, with a sufficient high number of participants and also to randomize test conditions across participants. Participants with previous experience listening to time-compressed speech do not necessarily have to be excluded, as long as differences between settings are investigated within one session. The degree of experience with time-compressed speech for the investigation of absolute SRT values becomes important, however, as recognition correlates positively with the time of exposure to time-compressed speech (Gordon-Salant and Friedman, 2011).

3.4.2 Results of original and time-compressed speech in relation to perceptual learning theory

In general, the results observed support the existence of perceptual learning, described as the improvement of perceptual task performance that occurs after practice (Adank and Janse, 2009). The RHT (Nahum et al., 2008, 2010; Adank and Janse, 2009; Banai and Lavner, 2012; Ahissar et al., 2009) provides an excellent framework to explain auditory perception and can be applied to the perception of time-compressed speech as well. It suggests that in a perceptual learning process, the initial performance relies on immediate access to abstract acoustic representations at high levels of the auditory pathway. These abstract representations allow for relatively good initial performance in everyday life. However, under privileged conditions (e.g., repeated listening), performance can be improved as learning progresses and specific lower-level representations that are beneficial for the current task become accessible. Based on this idea, the RHT predicts that learning that is based on high–level representations can be generalized and transferred to different tasks, whereas learning that applies lower–level representations is task-specific.

In detail, Banai and Lavner (2012) reported initial adaptive learning effects during early sessions and additional ongoing learning effects in later sessions. This pattern is comparable to the results of the current study, where larger learning effects were observed during early measurements and ongoing learning effects were observed within and between sessions. Banai and Lavner (2012) considered perceptual learning of time-compressed speech using the RHT framework. On the basis of this theoretical framework, the current findings on initial intra-session learning effects can be explained by brief adaptation processes. These include learning of the test procedure, instructions, but also basic characteristics of the speech material (e.g. speaker's voice, sentence structure) making use of high-level abstract acoustical representations. Also, the present results on later intra- and inter-session learning effects can be related to the application of increasing knowledge about detailed low-level spectro-temporal representations of the presented speech. These principle mechanisms of learning seem to remain the same for NH and HI as well as for original and time-compressed speech. Moreover, the results of the current

study showed a general trend of decreasing SRT-values after practice, although in detail participants started within sessions at higher SRT-values than the SRT-values obtained at the end of previous sessions. This may indicate that part of the improvement due to learning was lost between sessions. Participants had to re-adapt to reach SRT-values as in previous sessions because they had forgotten learned parts.

The RHT can also explain the larger improvement achieved by the time-compressed, compared to original speech: Normal speech rates used in the original speech are expected to be more often applied in ecologically-likely hearing situations than time-compressed speech rates. This results in higher-level representations of auditory processing, which rely primarily on normal speech rates. Since time-compressed speech material was rather unfamiliar to the participants, auditory processing showed larger and longer learning effects.

In addition, transfer of learning from time-compressed to original speech was further explored and can be discussed using RHT. All participants who listened to the time-compressed speech condition (NH-TC, HI-TC) performed one final measurement with original speech. Their SRT results for original speech were compared to results of NH and HI who participated in the original speech condition (NH-Original, HI-Original). Since transfer effects were only tested from time-compressed speech to original speech, observations cannot be supported by transfer effects from original to time-compressed speech. Interestingly NH-TC showed lower SRT values than naïve NH within their first measurement of original speech. The median threshold of original speech for NH-TC is comparable to the median SRT values of -7.1 dB measured with the original OLSA for younger NH after two initial training lists (Wagener et al., 1999a). Therefore, participants were able to apply the learning of higher levels, which RHT and Banai and Lavner (2012) expected to be transferable. According to Wagener et al. (1999a) this first phase included customization to the measurement procedure and to the formal structure of the sentences, both information independent of the speech rate. The comparison to the thresholds observed by Wagener et al. (1999a) indicates that the initial phase of brief adaptation was completed after about two initial training lists and the second phase of detailed low-level learning dominated the effects subsequently observed. This is supported by the observation that NH-TC failed to completely transfer learning. They showed nearly 1 dB poorer SRT values for original speech than NH-Original who trained using original speech. The absence of transfer indicates that no additive interaction of the learning processes for original and time-compressed speech occurred. In agreement with Banai and Lavner (2012) and the RHT, participants of the current study failed to apply the trained spectro-temporal representations of the time-compressed speech at lower levels of the auditory pathway to the original signals. Therefore, NH-TC learned a different content than NH-Original during the second phase and increased their knowledge about different detailed low-level spectro-temporal representations of the presented speech. The absence of transfer of subsequent detailed low-level learning might also be valid for further alterations of the speech signal or matrix tests with different languages. This is in line with observations made by Pallier et al. (1998) and Sebastián-Gallés et al. (2000). They showed that the adaptation to time-compressed native speech was supported by previous exposure to time-compressed non-native speech that was similar to native speech. For

this adaptation to take place, however, the listener did not necessarily need to know the foreign language.

The non-significant difference of the SRT values for original speech between TC-trained and Original-trained HI was unexpected and different reasons might explain these results. First, the variation of measured SRT values was larger for HI than for NH. This might have concealed the differences in SRT due to different learning mechanisms. Second, the two hearing-impaired groups that were trained with original or time-compressed speech showed different hearing levels in pure-tone audiometric testing. At high frequencies between 4 and 8 kHz, HI-TC exhibited a lower average hearing loss than HI-Original. Even though the effect of high-frequency hearing loss on SRT in noise is limited, this still might have contributed to a slightly better (i.e., lower) SRT of the HI-TC group than expected from the result of NH and the HI-Original.

Conclusions and consequences for speech audiometry

The results presented for original and time-compressed speech provide further evidence that learning of speech in a matrix test progresses through an initial general phase (1-2 lists) to a subsequent prolonged and more stimulus-specific phase (at least up to 6 lists and 5 sessions). The appearance of these two phases was discussed with regard to the RHT, which refers the initial phase of brief adaptation to perception that relies on nonspecific high level auditory representations, and the subsequent prolonged learning phase to perception of stimulus specific low level representations. These general mechanisms of perceptual learning are similar for NH and HI as well as for original and time-compressed speech. Only the extent of the learning effects differs for the two groups and speech materials. The application of time-compressed speech has been shown to be useful to differentiate between the two learning phases with (initial brief learning phase) and without transferable learning (subsequent prolonged learning phase).

The observed results allow for specific recommendations for clinical as well as scientific applications of matrix tests in speech audiometry with repeated measurements. If matrix tests are used with original or time-compressed speech, two training lists should be administered in each session before measuring SRTs. In scientific applications that use original speech and aim to compare small differences or results of different sessions, the potential effect of intra- and inter-session learning requires a careful randomization of test situations across sessions. Also, the recruiting of experienced listeners who are well trained on the test materials may be beneficial. If tests are conducted with time-compressed speech, comparison of results of different sessions is not recommended, but instead measurements should be performed within one session. By using this approach, the known effects of inter-session learning can be avoided.

Acknowledgements

We thank Kristina Anton, Lüder Bentz, Nina Blase, Karin Brand, Maximilian Busse, Fehime Cigir, Shiran Koifman, Theresa Nüsse, Patrycia Piktel, Martin Seidel, Johanna Weigel, Mareike Wemheuer and Nathalie Zimmermann for their support with data collection, Gurjit Singh

for helping to prepare the manuscript and G.A. Manley (www.stels-ol.de) for advising on language issues. This project was supported by Phonak AG; the Ministry for Culture and Science of Lower Saxony, Germany; the European Regional Development Fund (ERDF, project HURDIG) as well as Niedersächsisches Vorab (project AKOSIA).

Parts of the results were already presented on the 15th annual conference of the Deutsche Gesellschaft für Audiologie (2012) in Erlangen, Germany and the International Hearing Aid Conference (2012) in Tahoe City, USA.

4

Speech perception at positive signal-to-noise ratios using adaptive adjustment of time compression

Positive signal-to-noise ratios (SNRs) characterize listening situations most relevant for hearing-impaired listeners in daily life and should therefore be considered when evaluating hearing aid algorithms. For this, a speech-in-noise test was developed and evaluated, in which the background noise is presented at fixed positive SNRs and the speech rate (i.e. the time compression of the speech material) is adaptively adjusted. In total, 29 younger and 12 older normal-hearing, as well as 24 older hearing-impaired listeners took part in repeated measurements. Younger normal-hearing and older hearing-impaired listeners conducted one of two adaptive methods which differed in adaptive procedure and step size. Analysis of the measurements with regard to list length and estimation strategy for thresholds resulted in a practical method measuring the time compression for 50% recognition. This method uses time-compression adjustment and step sizes according to Versfeld and Dreschler (2002), with sentence scoring, lists of 30 sentences, and a maximum likelihood method for threshold estimation. Evaluation of the procedure showed that older participants obtained higher test-retest reliability compared to younger participants. Depending on the group of listeners, one or two lists are required for training prior to data collection.

Adapted from:
Schlueter, A., Brand, T., Lemke, U., Nitzschner, S., Kollmeier, B., Holube, I. (2014) "Speech perception at positive signal-to-noise ratios using adaptive adjustment of time compression", J. Acoust. Soc. Am., submitted.

4.1 Introduction

Speech-in-noise tests reflect the typical situation of speech embedded in background noise. Typical examples of these tests are digits triplet tests (e.g., Zokoll et al., 2012), sentence tests employing short meaningful sentences (e.g., Plomp and Mimpen, 1979; Kollmeier and Wesselkamp, 1997) or the matrix test structure that uses limited speech material always presented in the same arrangement (e.g., Wagener et al., 1999c; Hochmuth et al., 2012). All these tests are administered by adaptively adjusting the signal-to-noise ratio (SNR) to determine the speech recognition threshold (SRT), which commonly marks the SNR for 50% recognition. These usually show a steep discrimination function (e.g., Wagener et al., 1999a) and yield SRT values that are usually at negative SNRs even for hearing-impaired listeners (Wagener et al., 1999a; Wagener and Brand, 2005).

The negative SNR range has unfavorable effects when employing the tests for the assessment of a hearing loss or for the evaluation of hearing aids (see also Schlueter et al., 2014b). First, negative SNRs do not represent realistic hearing situations. Everyday conversations in noisy environments typically take place at positive SNRs (Olsen, 1998; Smeds et al., 2012). Second, several hearing aid algorithms showed SNR-dependent processing. For example, the gain in SNR from single-microphone noise reduction algorithms is highly dependent on the SNR in the input signal (e.g., Brons et al., 2013). Fredelake et al. (2012), for example, found the largest SNR improvement of single microphone noise reductions at positive SNRs. Conversely, low or even negative SNRs are challenging for these algorithms (Luts et al., 2010). Hence, employing positive and fixed SNRs for the assessment and parameter optimization of hearing aid algorithms would make it possible to test the hearing aid at its normal point of operation (Naylor, 2010). Third, normal-hearing and hearing-impaired listeners gain comparable recognition results at different SNRs. This complicates the comparison of results for normal-hearing and hearing-impaired listeners when evaluating hearing aid algorithms. Thus, there is a need for a speech-in-noise test that provides fixed and comparable positive SNRs for normal-hearing and hearing-impaired listeners.

To reach positive SNRs, Schlueter et al. (2014b) changed the difficulty of a speech-in-noise test by increasing the speech rate of German matrix-type sentences using time-compression algorithms. For time compression, they recommended a pitch synchronous overlap-add procedure. It preserves the pitch of the speech and deletes regularly-spaced parts of the signal to increase the speech rate (Moulines and Charpentier, 1990). Schlueter et al. (2014b) conducted recognition measurements with younger, normal-hearing listeners and determined positive SRTs for sentences compressed from 100% down to 30% or less of their original length.

In order to use fixed SNRs, a modification of the adaptive procedure in a speech-in-noise test is necessary. Instead of the SNR, Versfeld and Dreschler (2002) adaptively adjusted the speech rate of everyday sentences in quiet and measured the time-compression threshold (TCT). This is the speech rate that leads to 50% recognition. The current study explored the idea of an adaptive procedure according to Versfeld and Dreschler (2002) and an alternative method for adjusting the time compression (i.e. speech rate) for speech in background noise, while keeping fixed positive SNRs.

For adaptive methods, Levitt (1971) described two important principles: a) placing of observations and b) estimation of the resulting data. The placing of observations describes in a speech-in-noise test the presented SNR, which is adaptively adjusted. Step size, starting level and homogeneity of the presented material mainly determines placing (Leek, 2001; Smits and Houtgast, 2006). The estimation of the resulting data, e.g., denotes the SRT in a speech-in-noise test and is mainly defined by the number of presentations and the calculation strategy (Leek, 2001; Smits and Houtgast, 2006). Speech-in-noise tests such as the Oldenburg sentence test (OLSA, Wagener et al., 1999c), as well as the Plomp and Mimpen sentence test (Plomp and Mimpen, 1979), have considered these principles differently. The OLSA uses matrix-type sentences of the same structure (name, verb, numeral, adjective, object), which show relatively low redundancy and low predictability. OLSA lists consist of 20 or 30 sentences. During the measurement, the first sentence is easily understood and a 1up-1down staircase procedure adjusts the level of the speech (or background noise). The step size and direction is chosen depending on the number of reversals and recognition scores of the previous sentence. Finally, SRTs are estimated with the maximum likelihood method, applying recognition and SNR of all presented sentences in one list (Brand and Kollmeier, 2002). In contrast, the Plomp and Mimpen test applies a limited number of everyday sentences with high redundancy and predictability. During the test, participants first listen to sentences at unintelligible SNRs. Then the procedure decreases the background noise level until the first sentence is intelligible. Afterwards, a 1up-1down staircase procedure with fixed step sizes adjusts the level in the twelve following sentences. The SRT is estimated by the mean SNR of the last nine sentences and the SNR for a 14^{th} sentence. Versfeld and Dreschler (2002) modified this test procedure. They presented the signals in quiet, adaptively adjusted the speech rate and used the geometric mean of the last ten sentences for estimating TCT values. Based on these investigations, the current study combined the German matrix test OLSA with the idea explored by Versfeld and Dreschler (2002), to find a practical method for the adaptive adjustment of time-compression in a speech-in-noise test at positive SNRs. Differences of the underlying procedures required the analysis and comparison of parameters such as adaptive procedure and step size, list length, and estimation strategy of the threshold.

As noted above, the OLSA presents speech in stationary background noise, whereas the method developed by Versfeld and Dreschler (2002) applies time-compressed speech in quiet. Background noise, as well as fast speech, are known to decrease speech recognition for older and for hearing-impaired listeners (e.g., Gordon-Salant and Friedman, 2011; Wagener and Brand, 2005). These effects were attributed mainly to a loss of spectral and temporal resolution caused by a hearing impairment, and to central processing changes related to age (e.g., Adams et al., 2012; Gordon-Salant and Fitzgibbons, 2001; Tun, 1998). The new procedure incorporates both background noise and time-compressed speech and is expected to resolve the differences that were described between older and hearing-impaired listeners. Therefore, it is necessary to explore the adaptive procedure developed using participants of different age and different hearing ability, and to explain the results in connection with knowledge of previous studies about cognitive and perceptive declines in the discussion (e.g., Adams et al., 2012; Gordon-Salant and Fitzgibbons, 2001; Gordon-Salant and Friedman, 2011; Janse, 2009; Schneider et al., 2005; Tun, 1998; Wingfield et al., 2006).

Everyday sentences of the Plomp and Mimpen test mainly show learning for repetition of the same test list, while speech-in-noise tests with matrix type sentences show a different pattern of learning (e.g., Hochmuth et al., 2012; Wagener et al., 1999a). SRTs measured with the OLSA decrease with repetition because listeners become familiar with the test procedure, sentence structure and word material (Schlueter et al., 2014c; Wagener et al., 1999a). Wagener et al. (1999a) recommended using up to two training lists to overcome the learning effect. Learning effects are also dependent on the hearing ability, because hearing-impaired listeners showed less improvement of the SRTs with repetition (Wagener and Brand, 2005). For time-compressed matrix sentences presented in repeated measurements, the learning effect was even more pronounced and also generated an SRT improvement between sessions on different days (Schlueter et al., 2014c). The observed learning effects require evaluating the reliability of the adaptive procedure to adjust time-compressed speech. It is hypothesized that learning can be observed within sessions of repeated measures and between sessions on different days. However, it is expected that this training effect saturates and that it is possible to identify the extent of training that is required for obtaining stable results.

The current study consists of two parts. The first part examined different parameter settings to find a practical method for adaptive adjustment of time compression in a speech-in-noise test. The parameters tested were adaptive procedure, step size, list length, and estimation strategy of the threshold. The second part evaluated the selected method. TCTs in noise were explored for participants of different age and hearing status. Additionally, the expected training effect was investigated and the number of required training lists was determined.

4.2 Method

4.2.1 Signals

Speech recognition was determined with the Oldenburg sentence test (OLSA, Wagener et al., 1999a). The sentences always have the same structure (name, verb, numeral, adjective, object). They were composed from a random selection of one out of ten possible words for each element of the sentence structure. After a selection process, the test includes 100 possible syntactically fixed sentences with low predictability, for instance “Peter kauft zehn nasse Messer.” (“Peter buys ten wet knifes.”). From these sentences, equally intelligible lists with 30 sentences each were composed. Sentences were spoken by a male speaker with a normal to moderate speech rate of, on average, 233 (±27) syllables/minute (Wagener et al., 1999c). A repeated superposition of all sentences resulted in a stationary noise with the same long-term spectrum as the speech (Wagener et al., 1999c).

For time compression, OLSA sentences were processed with the software Praat (Boersma and Weenink, 2009). This software uses a pitch synchronous overlap-add procedure, that preserves the original fundamental frequency and applies different time-compression factors ρ (see Section 4.2.2.1). This paper defines the time-compression factor ρ as the duration of the compressed signal compared to the original duration in percent. For example, $\rho = 25\%$ corresponds to a speech rate which is four times faster than original speech. Therefore, smaller time-compression factors result in higher time compression and faster synthesized speech.

4.2.2 Measurements

4.2.2.1 Adaptive procedures

Method A: Method A was implemented according to Versfeld and Dreschler (2002). Sentences of the OLSA were compressed in time to different factors ρ calculated using Equation 4.1 where N is a natural number between 0 and 10. The calculated time-compression factors ρ were rounded to the next integer.

$$\rho = 0.85^N * 100\ [\%] \tag{4.1}$$

Speech was thus presented at the original speech rate (N = 0) but also compressed up to 20% of its original length (N = 10). A 1up-1down procedure decreased or increased the time-compression factor adaptively whether a sentence was understood correctly or not (sentence scoring). The presented time-compression factor was ρ = 100% (meaning original speech) for the first sentence within one list.

Method B: This method was based on the adaptive procedure used in the OLSA (Brand and Kollmeier, 2002) and was applied with sentence scoring. The steps of the time-compression factor presented in the adaptive procedure were also calculated using Equation 4.1, with N varying in 0.5 steps between 0 and 12. The calculated time-compression factors ρ were rounded to one decimal place. Thus, method B realized more steps with a smaller step size compared to method A. The speech could be presented at its original length (N = 0) but also compressed down to 14.2% of its original length (N = 12). The step size varied during the adaptive procedure and included a change of N by +4 until the second reversal. After each second reversal, the step size was divided by two until N = 0.5. Again, the presented time-compression factor was ρ = 100% for the first sentence within one list.

4.2.2.2 Estimation of TCT

After presenting one OLSA list, recognition values (i.e. 0 for at least one mistake in the repetition of the sentence or 1 for correct repetition of all words within a sentence) were available for each sentence, together with the respective time-compression factor used for this sentence. The estimation of the TCT for each list of sentences was performed in the linear domain marked by N in Equation 4.1, rather than by averaging or extrapolating across the time-compression factors directly. For the presentation of results, N as well as ρ values are given. Respective TCT values are specified by TCT_N or TCT_ρ. Three different estimation strategies were used to calculate the TCT_N: a) mean of N, b) mean of reversals, and c) maximum likelihood method (Brand and Kollmeier, 2002). Within the maximum likelihood method, the discrimination function (see Equation 4.2)

$$\mathrm{p(N)} = 1 - \frac{1}{1+\mathrm{e}^{4*\mathrm{slope}*(\mathrm{TCT_N}-\mathrm{N})}} \tag{4.2}$$

was fitted to the data. In Equation 4.2, p is defined as the mean probability that sentences are repeated correctly. This probability is dependent on the time compression described by N. TCT_N denotes the time-compression threshold specified by N, which refers to 50% probability of correct responses. The parameter slope describes the slope of the discrimination function at

TCT_N. The result is the parameter setting of TCT_N and slope that produces the observed data with the maximum likelihood. The resulting TCT_N estimate was used for further data analysis.

Although lists of 30 sentences were presented, these three strategies (mean of N, mean of reversals, and maximum likelihood method) were calculated for different list lengths of 20 or 30 sentences. For a list of 30 sentences, the strategy "mean of N" applied the N values for sentences 11 to 30. The strategy "mean of reversals" used N values obtained for reversals within the same range. The maximum likelihood method used recognition scores and values of N for sentences 1 to 30. For a list of 20 sentences, it was assumed that only the first 20 sentences were presented and the last ten sentences were omitted. Therefore, the strategy "mean of N" used the values of N for the sentences 11 to 20 and the maximum likelihood method applied the recognition scores and values of N for the sentences 1 to 20. Again, the strategy "mean of reversals" used the N values obtained for reversals within the sentences 11 to 20. In total, six different strategies for the estimation of the TCT_N were compared.

4.2.3 Participants

Listeners took part in five different subgroups. Table 4.1 characterizes the participants belonging to the groups and summarizes the executed adaptive method, age, number of participants, sex, and hearing loss. Based on pure-tone audiometry testing, younger normal-hearing listeners (YNH) exhibited hearing levels of 20 dB HL or better at all octave frequencies between 0.25 and 8 kHz. Older normal-hearing listeners (ONH) showed hearing levels of 20 dB HL or better between 0.25 and 4 kHz, 30 dB HL or better at 6 kHz and 40 dB HL or better at 8 kHz. Older hearing-impaired listeners (OHI) exhibited a mean hearing threshold at the frequencies 0.5, 1, 2 and 4 kHz of 25 to 60 dB HL. Their hearing impairment was symmetrical, because differences of the thresholds were 10 dB HL or less between the ears at frequencies between 0.125 and 8 kHz. Figure 4.1 depicts the thresholds of all participating groups. All listeners had German as their native language and no prior experience with the Oldenburg sentence test. In addition, participants conducted the Trial Making Test and the verbal Digit Span forward and backward. These tests explored cognitive abilities of psychomotor information processing speed, as well as short-term and working memory. Those results in the cognitive tests that were 1.5 standard deviations poorer than the mean age related standard (Härting et al., 2000; Tombaugh, 2004), excluded participants from the investigation. Therefore, all participants showed cognitive abilities appropriate to their age or better. All participants were paid a small sum for their participation, to compensate their expenses.

4.2.4 Setup and schedule of measurements

The experiments were conducted in a sound-insulated booth. A PC with Matlab-based (MathWorks, Natick, MA) programming controlled the presentation of the signals. Signals were routed through a sound card (Fireface 400, RME, Audio AG, Haimhausen, Germany) and a headphone amplifier (HB 7, Tucker Davis Technologies, Alachua, FL) to headphones (HDA 200, Sennheiser, Wedemark-Wennebostel, Germany). The headphones were free-field equalized according to international standards (IEC 60645-2, 2010; ISO 389-8, 2004) and presented signals diotically.

Table 4.1: Characterization of the subgroups and their abbreviation used in the text.

Participating group	**Age [years]**	**Hearing**	**Method**	**Number**	**Sex**	**Speech level [dB SPL]**
YNH-A	Mean: 23, range: 20-28	normal	A	15	5 male, 10 female	Mean: 59.3, range: 55-69
YNH-B	Mean: 23, range: 20-26	normal	B	14	2 male, 11 female	Mean: 60.2, range: 55-75
OHI-A	Mean: 70, range:65-74	impaired	A	12	8 male, 4 female	Mean: 65.3, range: 60-76
OHI-B	Mean: 69, range: 62-74	impaired	B	12	9 male, 3 female	Mean: 66.6, range: 60-78
ONH-A	Mean: 68, range:61-78	normal	A	12	5 male, 7 female	Mean: 57.3, range: 55-64

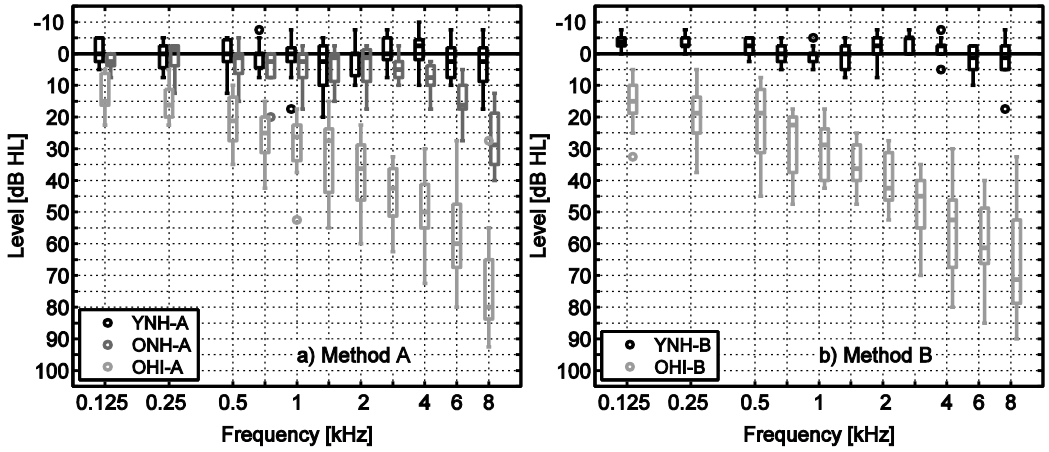

Figure 4.1: Pure tone audiograms for groups conducting measurements with methods A and B. Thresholds of both participants' ears were averaged.

During the first session, participants conducted a questionnaire about their education, hearing ability, and anamnesis, as well as audiometric measurements. Afterwards, their cognitive abilities were investigated with the Trial Making Test and Digit Span forward and backward. During the second session, participants listened to the original OLSA sentences and adjusted their level louder or softer than comfortable and finally to a comfortable level. Listeners repeated these adjustments three times. The median comfortable level (MCL) of the three adjustments was used as speech level in the following sessions. This speech level was limited to 55-75 dB SPL for YNH-A/B and ONH-B as well as 60-80 dB SPL for OHI-A/B. Table 4.1 gives the mean speech levels of the groups. For initial practice of the OLSA, participants performed two lists with the original speech material, while a screen visually displayed correct sentences after the response. Afterwards, a third list was presented without visual confirmation. Listeners conducted these three measurements with adaptive adjustment of the noise level. A speech recognition threshold of 80% recognition (SRT_{80}) with sentence scoring was measured.

The noise was continuously presented and a visual countdown marked the beginning of the sentence presentation. For the following measurements, an SNR of at least 80% recognition was required. Therefore, the SRT_{80} for the third list was verified by using the SRT_{80} as fixed SNR and determining the recognition score. Listeners with an SRT_{80} smaller than or equal to 1 dB SNR conducted the control measurement at 1 dB SNR. Participants with an SRT_{80} larger than 1 dB SNR listened to the signals at their individual SRT_{80}. If listeners reached recognition scores below 80%, they performed the control measurement again with an SNR increased by 2 dB. This step was repeated until a recognition score of 80% or above was reached. Afterwards, listeners executed six lists of the TCT measurement. The noise was presented at 1 dB SNR or a higher SNR at which participants understood at least 80% of original speech. YNH-A/B, ONH-A and most of OHI-A/B listened to signals presented at 1 dB SNR – except two participants of OHI-A that listened to 8 and 4 dB SNR, respectively and three participants of OHI-B that listened to 2 dB SNR. During the third session, participants repeated the six lists of the TCT measurement. The time interval between the second and third session was three to ten days. YNH performed measurements of the first two appointments within the first session.

4.3 Results and discussion

4.3.1 Effect of different adaptive methods and estimation strategies on the TCT

4.3.1.1 Results

Adaptive methods A and B were used for adjusting time compression during the test lists. Figure 4.2 shows the time compression as a function of sentence number within a presented list and therefore displays the average progress of all lists for methods A and B. All panels show rapidly increasing N values within the first sentences until N is close to threshold. In contrast to method A, which showed a continuous growth of N until an asymptotic value was reached, the average N values in method B showed an overshoot effect: The initial increase was faster than with method A, and a small oscillation was detected (i.e., a maximum followed by a minimum) until the asymptotic value was reached from below (see Figure 4.2c and d). Then N values decreased with method B and afterwards increased slowly until they remained static. The interquartile ranges of presented N values at the end of all lists were larger for method A than for method B.

Figure 4.3 shows the TCT values for the threshold estimation strategies “mean of N” and “maximum likelihood method” in comparison to the time-compression values gained at the end of a list (20^{th} or 30^{th} sentence). The analysis did not include an estimation strategy for thresholds that average across reversals, because a sufficient number of reversals for calculation of the mean was not available for several lists of 20 sentences. Fewer than two reversals were found within a list measured with one listener of group YNH-A and with six listeners of group YNH-B.

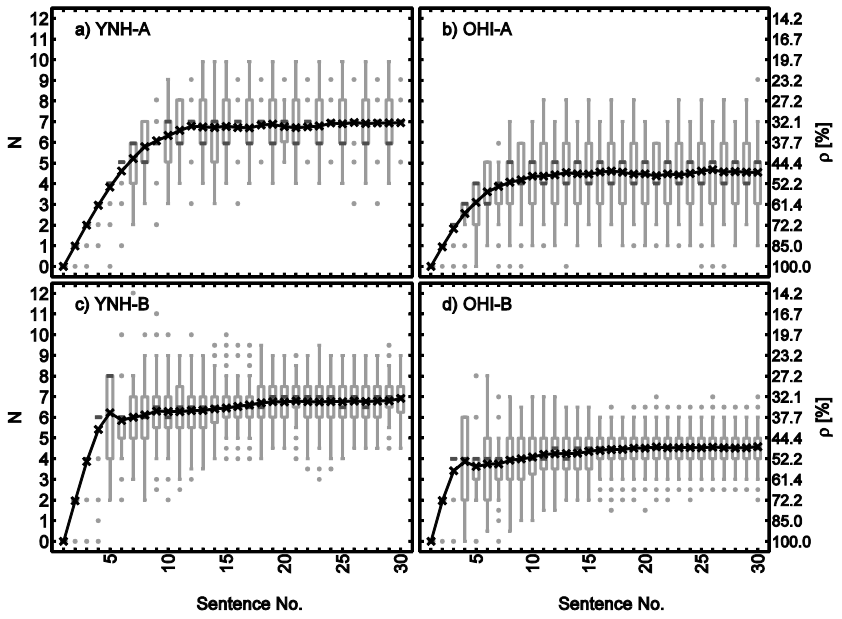

Figure 4.2: Time compression as a function of sentence number within a presented list. Results are displayed for the groups of a) young normal-hearing participants listening to adaptive method A (YNH-A), b) older hearing-impaired participants listening to adaptive method A (OHI-A), c) young normal-hearing participants listening to adaptive method B (YNH-B), and d) older hearing-impaired participants listening to adaptive method B (OHI-B). Colors black and gray display means and boxplots, respectively.

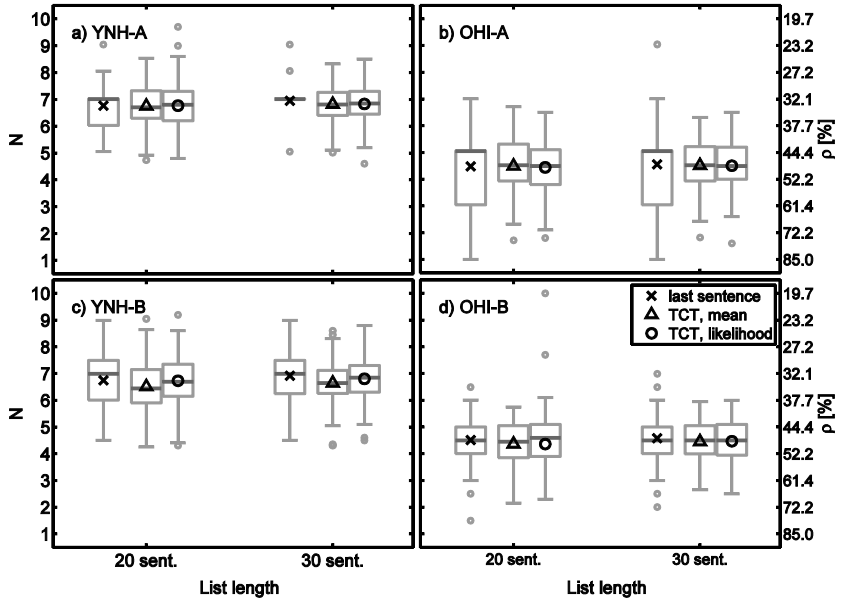

Figure 4.3: Time compression for list length of 20 and 30 sentences (20 sent. and 30 sent.) presented to a) young normal-hearing participants listening to adaptive method A (YNH-A), b) older hearing-impaired participants listening to adaptive method A (OHI-A), c) young normal-hearing participants listening to adaptive method B (YNH-B), and d) older hearing-impaired participants listening to adaptive method B (OHI-B). Boxplots and mean values marked by a cross represent the time-compression values adjusted in sentences of number 20 and 30, respectively. Boxplots and mean values displayed with a triangle or a circle show the TCT values, which were estimated with the mean and maximum likelihood method, respectively.

In general, the estimated results do not differ between adaptive method and estimation strategy for thresholds within each group of listeners. For YNH-B in Figure 4.3c, the maximum likelihood method resulted in slightly higher TCT_N values than the mean. A comparison of these results with the progress within the lists (see Figure 4.2) and the values for the last sentences (see Figure 4.3) suggests an underestimation of TCT_N based on the mean, due to the increase of N after the eleventh sentence. In contrast, the result of the maximum likelihood method is consistent with the values for late sentences at the end of a list. Similar effects as in Figure 4.3c are shown in Figure 4.3d for OHI-B. In addition, Figure 4.3d shows explicit outliers for the TCT calculated with the maximum likelihood method and a list of 20 sentences, outliers that are not present for 30 sentences. This pattern of results can also be seen in Figure 4.3a.

Statistical comparison of the four different strategies for TCT_N estimation shown in Figure 4.3 included pooled data of all twelve lists in both sessions for each adaptive method and each group and used a significance level of $\alpha = 0.05$. The data showed normal distribution and homogeneity of variances (Shapiro-Wilk and Levene test), except for three TCT_N values estimated for OHI-B where no normal distribution was obtained. Two mixed two-way analyses of variances (ANOVAs) were conducted for each group. This analysis obtained no significant differences of TCT_N values for both groups resulting from the method (YNH: $F(1,346) = 2.26$, $p = 0.133$; OHI: $F(1,286) = 0.17$, $p = 0.682$). TCT_N values of YNH showed significant effects of the estimation strategy ($F(1.992,689.139) = 24.56$, $p < 0.001$) and the interaction estimation strategy*method ($F(1.992,689.139) = 11.23$, $p < 0.001$). OHIs' results yielded only the significant interaction estimation strategy*method ($F(2.078,594.450) = 5.42$, $p = 0.004$). Since for both groups interaction of estimation strategy and method occurred, post-hoc t-tests were run with Bonferroni correction of the calculated probability values. Analyses considered every combination of possible estimation strategies for the method A and B and participating groups separately. Table 4.2 summarizes the results and mainly shows significant differences of the TCT estimation strategy mean for method B compared to other strategies.

Table 4.2: Probability values calculated with paired t-tests and Bonferroni corrections. Analysis compared TCT_N values that were measured with younger normal-hearing (YNH) and older hearing-impaired (OHI) listeners and were calculated for lists of 20 or 30 sentences length (20, 30). Mean or maximum likelihood method (mean, likelihood) were used for estimating the TCT_N values. Probability values in the upper right and lower left triangular part were determined for the adaptive methods A and B, respectively. Asterisks mark significance (: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).*

	YNH				**OHI**				
	mean, 20	likelihood, 20	mean, 30	likelihood, 30	mean, 20	likelihood, 20	mean, 30	likelihood, 30	
mean, 20	-	1.000	0.205	0.233	-	0.236	1.000	1.000	method A
likelihood, 20	0.000***	-	0.777	0.530	0.029*	-	0.219	0.257	
mean, 30	0.000***	0.038*	-	1.000	0.001***	1.000	-	1.000	
likelihood, 30	0.000***	0.075	0.000***	-	0.002**	1.000	0.726	-	
	method B				method B				

4.3.1.2 Discussion

The applied combination of German, low-predictability, matrix-type sentences and time-compressed speech showed small differences between the investigated adaptive methods A and B as well as the estimation strategies for TCT. In order to choose an adaptive method and estimation strategy, results have to be discussed with respect to the adaptive stimulus placement procedure and the subsequent estimation strategy for thresholds. The methods investigated in the current study are based on established speech tests developed by Plomp and Mimpen (1979), Versfeld and Dreschler (2002) as well as Wagener et al. (1999c) and Brand and Kollmeier (2002) applying staircase methods with adaptive levels or speech rates.

Adaptive stimulus placement method

In detail, the adaptive stimulus placement procedures controlled the step size differently: Method A provided a fixed, comparatively small step size whereas method B started with a larger step size with an adaptive decrease depending on the reversals until half of the value in method A was reached. This affected the average progress during the lists. After general adaptation close to the threshold, method A used speech at a similar time compression on average over a large number of sentences. This observation implies that the step size was large enough to exceed the threshold within a single step once the time compression presented was in the targeted area. On the other hand, the step size could be too large. A too large step size would result in a pattern of constant "jumping" between two consecutive time compressions, beginning at early sentences within a list and would allow for no convergence closer to threshold. This phenomenon could result in the missing variance of the steps presented for single sentences observed in Figure 4.2a (i.e. no interquartile range because more than 50% of the presented N values were the same for single sentences). Therefore, all single lists measured with method A were reviewed. Only few lists for YNH-A showed "jumping" between two consecutive time compressions at the end of the lists, indicating that the step size was efficient in method A.

In contrast, method B showed decreasing N values after general adaptation. This observation might be caused by the learning effect for time-compressed speech (see Section 4.3.2.2). Participants might have given incorrect responses for early sentences due to unfinished learning, which caused reversals. As a result, the step size reduced too fast and a gradual increase of time compression was observed for most of the sentences. As a consequence, the small step size of method B required more sentences to reach the threshold than the constant step size of method A. Also, the larger step size in method A simplified the perception of changes in time compression and possibly increased motivation for the participants as compared to method B. Incidentally, the missing of variance for single sentences noted in Figure 4.2a had no impact on the estimated TCTs, because methods A and B showed similar distributions of the TCTs (see Figure 4.3a and c). Although statistical analysis failed to show differences of the adaptive methods A and B, there is also a practical reason to support method A. Presenting different time compressions needs preprocessing of the speech material. Since method A applies fewer steps of time compression than method B, less preprocessing is necessary and fewer sentences have to be stored.

According to Smits and Houtgast (2006), an efficient up-down procedure provides the starting level close to threshold. Method A and B did not take this recommendation into consideration and provided original speech within the first sentence. This approach might have been less efficient and required a higher number of sentences. However, the methods followed the original procedure for OLSA, which starts with an SNR for speech recognition of 100% (Brand and Kollmeier, 2002). In addition, this approach took pronounced learning effects for time-compressed speech (Schlueter et al., 2014c) into account. In contrast to the approach recommended by Smits and Houtgast (2006), the applied procedure allowed a gentle adaptation to time-compressed speech. Starting with original speech at the beginning of a list and adapting to the threshold by increasing the time compression forced and limited learning effects mainly to the beginning of a list.

Homogeneity of the speech material also affects placing observations (Leek, 2001; Smits and Houtgast, 2006). Wagener et al. (1999b) controlled for the homogeneity of the OLSA sentences, selecting words with similar recognition and adjusting word levels to increase homogeneity. Thus, the OLSA shows a steep discrimination function (Wagener et al., 1999a). Unfortunately, time compression influences the homogeneity of the material, depending on the word's position within the sentences (Schlueter et al., 2014b). Names and objects at the beginning and end of sentences showed higher recognition than words in between. Therefore, assuming that time compression does not impact the homogeneity between sentences, Schlueter et al. (2014b) recommended the application of sentence scoring according to Versfeld and Dreschler (2002) instead of word scoring. Although estimation of the TCT values based on sentence scoring uses less information and items than word scoring (Brand and Kollmeier, 2002), recognition differences of words within sentences no longer influence the placing of observations, because the least intelligible word defines the recognition of an entire sentence.

Estimation strategy

In addition to placing of observations, the outcome of an adaptive procedure is specified by the estimation of the resulting data (Levitt, 1971). More precisely, the TCT is dependent on the estimation strategy and the number of presentations used for the estimation (Leek, 2001; Smits and Houtgast, 2006). In the current study, estimation of the TCT used the strategies of Versfeld and Dreschler (2002) as well as Brand and Kollmeier (2002). Versfeld and Dreschler (2002) calculated the TCT for a single measurement by the geometric mean of the speaking rate of the last ten sentences of one list (length:13 sentences) which is the same as the arithmetic mean of N. However, they also used a maximum likelihood method to estimate the discrimination function for grouped data. The original OLSA estimates SRT over 20 or 30 sentences with a maximum likelihood method by fitting the SRT and slope of a discrimination function (Brand and Kollmeier, 2002). Therefore, the current study estimated the TCT with the mean of N values or the mean of reversals, as well as with the maximum likelihood method and explored lists with a length of 20 or 30 sentences. These strategies were compared to find a TCT estimate for determining suitable results for single lists. Analysis disclosed different problems.

First, a sufficient number of reversals was not established for all listeners and lists to allow for a confident estimation of TCT. Therefore, further analysis excluded TCT estimation on basis of reversals.

Second, the analysis of the estimation strategies showed small differences. Generally, the adaptive procedures started at original speed and therefore required several sentences to adapt close to threshold. Therefore, N values for early sentences within a list were not as close to the TCT_N as N values for later sentences. Kollmeier et al. (1988) modeled the temporal evolution of an adaptive measurement track as a Markov chain and showed that the distribution of initial values steadily converges toward the limiting distribution, which is independent of the starting distribution. Hence, any bias of the estimated result decreases with increasing length of the adaptive track and with the number of initial trails discarded for estimating the results. Especially with method B (e.g. see Figure 4.3c), an increase beyond the eleventh sentence could still be observed because of the slow convergence towards the limiting distribution. Even though the TCT_N estimation strategy based on the mean discarded the first ten sentences, it resulted in smaller TCT_N values compared to the N values of the last sentence. Only the maximum likelihood method assessed all N values, considered the history of individual lists and reflected the variety of the individual progress within a list. As a result, TCT values were closer to the N values for late sentences.

Unfortunately, the maximum likelihood method failed for single lists (see Figure 4.2d) if a list length of 20 sentences was used. This occurred in cases when the adaptation showed only one direction for a large number of sentences e.g. after early mistakes of the participant or after an incidental large number of right answers. Therefore, the lists should be long enough, i.e. 30 sentences for following explorations. The recommendation of 30 sentences and maximum likelihood method as estimation strategy is limited to the investigated adaptive procedures and might be different for a deviating starting level. In principle, it is also possible to apply the maximum likelihood method with values determined after the second reversal. However, the second reversal would occur at late sentences in adaptation processes, which show only one direction for a large number of sentences. Consequently, the maximum likelihood method would be calculated on the results for a reduced number of sentences after the second reversal. Again this supports a sufficient list length of e.g. 30 sentences because a higher number of results after the second reversal will be available for a reliable estimation of the thresholds than for list lengths of 20 sentences.

4.3.2 Evaluation of the selected method

4.3.2.1 Results

The selected adaptive method and estimation strategy (method A, maximum likelihood method, 30 sentences) was evaluated for group differences and learning effects. Figure 4.4a-c depicts TCT_N and TCT_ρ values for all lists and groups. TCT_N values are high for YNH-A, low for OHI-A and in-between for ONH-A. Statistical analysis explored differences between groups on the pooled data for all lists and sessions. Since the Shapiro-Wilk and the Levene test showed normal distribution but no variance homogeneity, nonparametric tests were used to analyze

the comparisons between groups. The Kruskal-Wallis test yielded a significant result ($\chi^2(2) = 319.15$, $p < 0.001$) and post hoc Mann-Whitney U-tests resulted in significant differences of all groups (YNH-OHI: $U = 330.50$, $p < 0.001$; YNH-ONH: $U = 1508.00$, $p < 0.001$; OHI-ONH: $U = 4397.50$, $p < 0.001$).

Figure 4.4 shows learning effects because TCT_N increases with increasing number of lists and sessions, at least for YNH-A. For the statistical analysis of these observations within the three groups, a Shapiro-Wilk test confirmed normal distribution for all groups, sessions and lists with only minor exception. A two-way repeated measures ANOVA explored significant differences of lists and sessions for each group. TCT_N values increased with increasing number of lists (YNH-A: $F(3.25,45.43) = 26.76$, $p < 0.001$; ONH-A: $F(5,55) = 2.36$, $p < 0.001$, OHI-A: $F(5,55) = 2.04$, $p < 0.001$). In addition, results of the first session showed only for YNH-A smaller TCT_N values than in the second session ($F(1,14) = 46.82$, $p < 0.001$). In contrast, for ONH-A and OHI-A the results of the two sessions did not differ significantly (ONH-A: $F(1,11) = 0.87$, $p = 0.370$; OHI-A: $F(1,11) = 0.17$, $p = 0.686$).

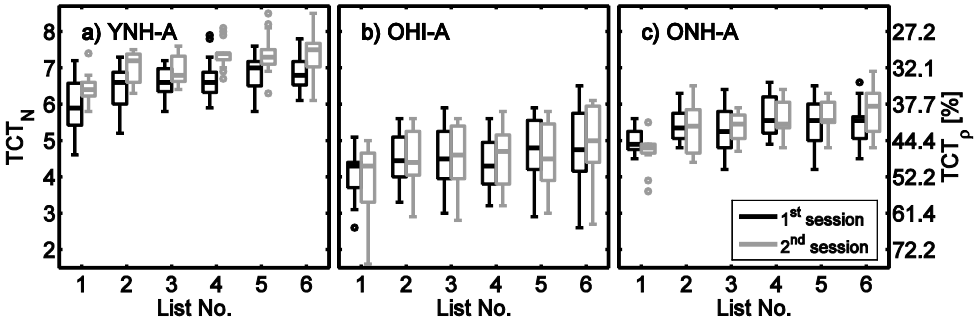

Figure 4.4: TCT values measured with adaptive method A and estimated with the likelihood method for lists of 30 sentences. a) YNH-A, b) OHI-A, and c) ONH-A conducted six lists within two sessions each. Black and grey display TCT values of the first and second session, respectively.

In order to determine the number of training lists after which no significant difference of the TCT_N occurred, *t*-tests were used to calculate probability values with Bonferroni correction. Within the first session, results of each list were compared statistically with results of subsequent lists. YNH-A showed significantly lower TCT_N values for the first two lists than for the following lists ($p \leq 0.028$). ONH-A and OHI-A reached only for the first list significantly lower TCT_N values compared to the subsequent lists ($p \leq 0.027$).

For analysis of test-retest reliability, scatterplots in Figure 4.5a-c show the results of the first and second session and respective correlation coefficients. A linear function with a slope of 1 was fitted to the data for estimating the bias, which is represented by the distance between the bisecting line and the fitted function. The bisecting line represents ideal reliability, when participants reach equal results in the first and second session. The percentage of TCT values is given for which the TCT values in the first session are larger, equal or smaller than in the second session. Figure 4.5 omits the learning effects examined in the previous paragraph, because only results of the third to sixth lists are presented. Overall, a majority of TCT_N values reached lower values in the first than in the second session. The correlation observed for OHI-A was higher than for YNH-A and ONH-A. In addition, YNH-A and OHI-A showed the largest

and smallest bias of 0.52 and 0.07, respectively. This is also represented by the groups' percentage of TCT values, which are larger in the second session than in the first session. Therefore, the YNH-A and OHI-A reached also the largest and the smallest percentage value, respectively.

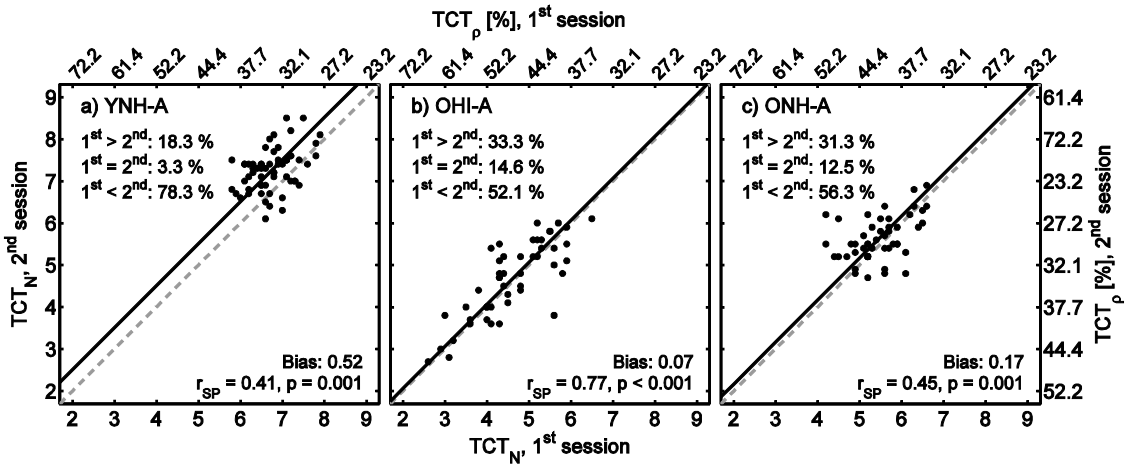

Figure 4.5: Scatterplots of TCT values reached for single lists in the first and second sessions for a) YNH-A, b) OHI-A and c) ONH-A. Number of TCT values in % is documented, where TCT values in the first session are larger, equal or smaller than in the second session. In addition, Spearman´s correlation coefficient r_{sp} is quoted. The dashed bisecting line represents equal results in the first and second session. The solid line with a fixed slope of 1 was fitted to the data. The distance between dashed and solid line indicates the bias.

4.3.2.2 Discussion

Groups of different age and hearing ability showed different TCT_N values for time-compressed speech in noise, i.e. YNH-A showed the highest thresholds, while thresholds of OHI-A and ONH-A were low and intermediate. These observations confirm studies by e.g. Adams et al. (2012), Gordon-Salant and Friedman (2011), and Tun (1998) who discussed several influencing factors. First, Adams et al. (2012) specified the effect of time compression. It was assumed to result in a "loss of subphonemic cues, such as place of articulation and vowel duration" (p. 29) and in the shortening of naturally occurring gaps, especially with added background noise (Adams et al., 2012). Second, hearing loss was a dominant factor for explaining results of older hearing-impaired listeners (Janse, 2009). Already for original OLSA, a difference in SRT was observed for hearing-impaired listeners compared to normal-hearing listeners (Wagener and Brand, 2005). According to Wingfield et al. (2006), the effect of hearing loss increased as speech rate increased. Limited temporal and spectral resolutions caused by hearing loss might impact the detection of time-compressed speech with shortened cues in background noise (Adams et al., 2012; Gordon-Salant and Fitzgibbons, 2001; Schneider et al., 2005; Wingfield et al., 2006). Third, larger effects of age were observed for time-compressed speech compared to original speech (Holube et al., 2009; Meister et al., 2011). For time-compressed speech, age-related cognitive factors or changes in central auditory processing were discussed to impact recognition in older listeners (e.g., Adams et al., 2012; Gordon-Salant and Fitzgibbons, 2004; Janse, 2009;

Wingfield et al., 2006). For example slowed information processing (Gordon-Salant and Fitzgibbons, 2004; Janse, 2009), reduced processing resources (Adams et al., 2012; Wingfield et al., 2006) and other age-related processing difficulties (e.g., Adams et al., 2012) served as explanations for these observations. Fourth, age-dependent cognitive factors and hearing impairment interact with each other and might explain results of hearing-impaired older participants (e.g., Adams et al., 2012; Wingfield et al., 2006). Pichora-Fuller and Singh (2006) described that a hearing loss may lead to inaccurate representations of hearing situations. These misrepresentations most probably continue along the auditory pathway and lead to inaccurate cognitive processing especially in challenging listening situations. Hearing-impaired listeners may also try to compensate reduced acoustic information with increased processing resources, which in turn might be missing at higher processing levels for e.g. comprehension (Gordon-Salant and Fitzgibbons, 2001; Wingfield et al., 2006). However, there is evidence that age-related decline of recognition for time-compressed speech is not simply dependent on an irreversible process of aging. Gordon-Salant and Friedman (2011) showed increasing recognition as a function of hours of listening to time-compressed speech for older blind participants. They concluded that blind adults' "greater attention to auditory information … may reduce the expected age related decline in auditory temporal processing" (p. 629). These relations and effects show that age is just a representative for different kinds of age-related changes, and investigations are needed to search for clear explanations. The current study permits no further analysis, but offers a new test procedure for more detailed investigations of aging processes.

In addition to age and hearing ability, learning affects the recognition in the introduced test procedure with time-compressed matrix sentences. Matrix tests like the OLSA with original speech also showed learning effects (Hochmuth et al., 2012; Wagener et al., 1999a; Wagener and Brand, 2005). Speech recognition improved during consecutive presented lists. Thus, different authors recommended one or two training lists before data collection (e.g., Schlueter et al., 2014b; Wagener et al., 1999a), to reduce the effect of learning within the test. These learning effects increased with time-compressed speech (Schlueter et al., 2014b). The current study supports these findings and showed learning effects within and between sessions on different days. Although listeners were trained using the original speech material before, the TCT improved significantly over the first one or two lists within the first session, depending on the group. Therefore, at least one list of training for ONH-A and OHI-A and two lists of training for YNH-A are recommended. Results of two different sessions determined with YNH-A should not be compared, because significant differences between sessions were observed and the bias was large compared to ONH-A and OHI-A. ONH-A and OHI-A can conduct measurements within different sessions because their bias was relatively small.

In contrast to the learning effect shown in Figure 4.4, Versfeld and Dreschler (2002) did not measure any learning effect for Plomp-type sentences, which are typical for everyday conversations. Schlüter et al. (2014b) confirmed this result and found no learning effect with the Goettingen sentence (Kollmeier and Wesselkamp, 1997) test, which also uses everyday language.

In addition, Wagener and Brand (2005) and Schlueter et al. (2014b) established a connection between training and hearing ability. They measured smaller learning effects with hearing

impairment for the original OLSA. In contrast, Schlueter et al. (2014b) recommended two training lists for presenting time-compressed speech to normal-hearing and hearing-impaired participants. The recommendation of longer training for hearing impaired, as compared to the current study, might have resulted from the deviating experimental setup: Schlueter et al. (2014b) did not train the original OLSA first and used a different adaptive procedure controlling the SNR.

Banai and Lavner (2012), Adank and Janse (2009) as well as Schlueter et al. (2014b) explained the learning effects with the Reverse Hierarchy Theory. During learning, specific lower-level representations in the auditory pathway become accessible if they are useful for performance. As a result, learning effects were observed for time-compressed speech in addition to earlier learning of the original OLSA material, because the extra adaptation to the fast speech cues was helpful for better recognition.

4.4 General discussion

The test procedure explored permitted the presentation of positive fixed SNR for normal-hearing and hearing-impaired listeners in order to measure a TCT. The SNR was selected to reach high (more than 80%) recognition for the original speech material. Therefore, and independent of their hearing ability, participants listened to hearing situations similar in SNR and recognition for original speech. Adjustment of the time compression (speech rate) modified the difficulty of the test and showed group-specific results.

The positive SNRs provided represent realistic listening situations. Everyday conversations often take place in noisy environments, in which speech is louder than the background noise (Olsen, 1998; Smeds et al., 2012). Moreover, applying defined fixed positive SNRs is useful for evaluating hearing aid algorithms. For example, single-microphone noise reduction algorithms are mainly beneficial when the input of the algorithm is at positive SNRs (Fredelake et al., 2012) while low and negative SNRs are challenging for these algorithms (Luts et al., 2010). However, time compression of the speech material leads to changes of speech statistics compared to original speech. If hearing aid processing relies on statistical information about the speech signal, which is altered by time compression, the algorithm may produce unexpected output.

Apart from positive SNRs, more complex hearing situations are more realistic. The current study applied background noise and time-compressed speech to challenge recognition, while in traditional speech-in-noise tests, including the original OLSA, only the background noise is used. Thus, the presented hearing situations of the current study were more complex. Unfortunately, the speech rates applied in this study especially for younger normal-hearing listeners represented more or less non-realistic hearing situations. Fast speech produced during a conversation or reading of a text reaches about 510 syllables/minute (Janse, 2003, Chapter 5, Section 3) and is therefore slower than the speech presented to YNH-A in this study (about 709 syllables/minute for median TCT). However, for older participants, the speech rate presented (484 syllables/minute and 560 syllables/minute for median TCT of OHI-A and ONH-A, respectively) is in the range of Janse (2003).

Since the method introduced provided a more complex hearing situation with background noise and time-compressed speech, the results cannot be directly compared to the original OLSA. In the original OLSA, hearing loss is considered the main determinant of participants' ability to separate speech from background noise. In contrast, the introduced test applied extra time compression, and therefore additional individual cognitive abilities might have been important and affected recognition in background noise (e.g., Wingfield et al., 2006, see also Section 4.3.2.2).

Conclusions

The current studies of a speech-in-noise test with time-compressed speech led to the following conclusions:

- A practicable method for the adaptive adjustment of time compression in a speech-in-noise test with OLSA sentences should use the following: time compression adjustment and step sizes according to Versfeld and Dreschler (2002) with sentence scoring, lists of 30 sentences, and a maximum likelihood method for threshold estimation.
- The TCT values measured deteriorated with age and hearing loss because of cognitive and perceptive decline.
- At least one or two lists should be used for training before data collection for older normal-hearing and hearing-impaired listeners and for younger normal-hearing listeners, respectively.
- Older normal-hearing and hearing-impaired participants showed higher test-retest reliability compared with younger normal-hearing participants. Therefore, older listeners can conduct measurements across different sessions, while data collection from younger listeners should be limited to one session.

Acknowledgements

We would like to thank Diana Herzog for her support on data collection and G.A. Manley (www.stels-ol.de) for advising on language issues. This project was funded by Phonak AG and the Niedersächsischen Vorab (project: AKOSIA).

Parts of this work were presented on the 16th annual conference of the Deutsche Gesellschaft für Audiologie (2013) in Rostock, Germany and the International Conference on Cognitive Hearing Science for Communication (2013) in Linköping, Sweden.

5

Evaluation of single-microphone noise reduction algorithms at fixed positive signal-to-noise ratios using individually time-compressed speech

To evaluate and compare hearing aid algorithms, it is often necessary to perform speech recognition tests at positive signal-to-noise ratios (SNRs) that correspond to a given level of performance (e.g. 50% recognition scores). For this purpose, this study examined the feasibility of a hybrid approach. This approach used results of a recently developed procedure that adaptively increases the speech rate to shift the SNR range towards fixed positive values for the evaluation of noise reduction algorithms. Eleven hearing-impaired listeners participated in the experiments. Their individual speech rate for reaching the speech recognition threshold was determined for SNRs of 1 or 5 dB. Then, recognition scores for the individually time-compressed sentences were measured with and without two different single-microphone noise reduction algorithms. As a result, no ceiling effects of recognition scores were observed, although measurements were conducted with participants having different hearing ability, with different set of speech signals, with different noise reduction algorithms and at different positive SNRs. Participants achieved a significant improvement of the recognition score only with one of the algorithms (an algorithm with a priori knowledge of the noise and Wiener filter). The second noise reduction algorithm (a realistic algorithm with minimum controlled recursive averaging and spectral subtraction) did not show a significant improvement of the recognition score even when positive SNRs were presented that, from objective measurements, may have been expected to bring such improvements.

5.1 Introduction

The objective of single-microphone noise reduction algorithms is to reduce background noise under speech-in-noise conditions. Frequently, the processing of single-microphone noise reductions is dependent on the SNR (e.g., Brons et al., 2013; e.g., Fredelake et al., 2012; Hoetink et al., 2009) and the algorithms often introduce increasing distortions with decreasing SNR (as described by ,e.g., Marzinzik, 2000; Fredelake et al., 2012; Marzinzik and Kollmeier, 2002; Neher et al., 2014a; Brons et al., 2014). Therefore, the effect of those algorithms should be evaluated using speech-in-noise tests that can be conducted at fixed signal-to-noise ratios (SNRs). Furthermore, the SNR should be high, i.e. positive, for an analysis of the algorithms in ecologically relevant hearing situations (Olsen, 1998; Smeds et al., 2014) and, presumably, with less distortion.

Up to now, these requirements are difficult to meet in the evaluation of speech following noise reduction processing. Although presentation of fixed positive SNRs is possible using adjustment methods such as the procedure proposed by Wittkop et al. (1997), or in subjective evaluation methods of, e.g., overall signal quality (e.g., Brons et al., 2013, 2014; Peissig, 1993, Chapters 5 and 6), participants perform the tasks presented with an individual subjective criterion that is difficult to compare between participants and leads to large interindividual variation. Therefore, frequently, speech-in-noise tests are applied to determine the speech recognition with and without noise reduction processing. Within these procedures, the participant's task is to repeat the perceived signals and the number of correct answers is counted. Positive SNRs, however, commonly lead to high recognition scores near 100% and are thus insensitive to possible improvements due to the algorithms and possible differentiation between the algorithms. Therefore, for example, Brons et al. (2013) and Neher et al. (2014a) also presented negative SNRs for the evaluation of noise reduction algorithms, and reached recognition scores below 100% and thus avoided ceiling effects. In addition, e.g., Bentler et al. (2008) or Wittkop (2001) analyzed noise reduction processing using speech-in-noise tests with adaptive procedures to adjust the SNR to the speech recognition threshold (SRT), i.e. a speech recognition score of 50%. This approach avoids ceiling effects but presents variable SNRs that are difficult to use for the evaluation of hearing aids and their algorithms (Naylor, 2010). Furthermore, SRTs are dependent on the individual hearing ability and vary between participants. As a result of the SNR variation, algorithm processing is frequently different between participants. Besides, this approach results in negative SNRs that occur especially in German speech-in-noise tests, even for hearing-impaired participants (e.g., Luts et al., 2010; Wagener and Brand, 2005). These negative SNRs in turn are not ecologically relevant and presumably lead to more distortions of the algorithms than at positive SNRs.

The current study applied a new speech-in-noise test in a hybrid approach to fulfill the requirement of presenting positive, fixed SNRs. Schlueter et al. (2014a) developed a speech-in-noise test with the objective of adapting the speech rate to a recognition score of 50% (time-compression threshold, TCT). The sentences presented were compressed in time, whereby the pitch of the signals was preserved. In order to allow for the adaptive measurement of the TCT, it is necessary to verify whether the recognition score at the intended fixed SNR is at least

75% when using the original speech rate. It is difficult to use this procedure directly for evaluation of the algorithms, because varying time compression during the measurement is expected to result in varying processing of the hearing aid algorithms, as stated above for varying SNR values. Individual TCTs can rather be used to individually adjust the difficulty of recognition measurements at fixed SNRs. These SNR values can be positive, in order to be ecologically relevant, and presumably lead to a processing of the algorithm with higher benefit than at negative SNRs. When selecting the SNR for the evaluation, it has also to be remembered that the higher the selected SNR, the faster the presented speech rates. After the selection of the individual hearing situation defined by SNR and time compression, recognition scores can be measured with and without the activation of noise reduction algorithms. As a result of that process, the speech recognition scores for unprocessed hearing situations are at about 50% for speech compressed to TCT. Compared to this situation, maximum possible changes of recognition due to the algorithms can be detected. If the noise reduction algorithms induce recognition changes in a speech-in-noise test that are less than 50%, floor or ceiling effects are not to be expected.

For reducing the background noise, realistic single-microphone noise reduction algorithms estimate the background noise on the basis of the mixed speech-in-noise signal. The estimate is often based on speech statistics and is applied in gain rules to filter this mixed signal. Since time-compressed speech has different speech statistics than original speech, and noise reduction is SNR-dependent (see above), it is necessary to study the processing and to objectively determine the amount of SNR improvement (as shown by, e.g., Neher et al., 2014a). Fredelake et al. (2012) used shadow filtering to investigate the objective overall SNR improvement as a function of the SNR at the input of three noise reduction algorithms. Based on the results of Fredelake et al. (2012), an objective analysis is expected to show an overall SNR improvement of the noise reduction algorithms for the listening situations of time-compressed speech presented at fixed positive SNRs. This observation leads to the assumption that no objective SNR improvement after noise reduction is not the reason for failure of recognition improvement.

Nevertheless, various studies did not obtain recognition improvement with single-microphone noise reduction algorithms (e.g., Brons et al., 2013, 2014; Hu and Loizou, 2007; Neher et al., 2014b). This observation is often attributed to a degradation of the speech signal, because reduction of noise also causes a partial reduction of speech and possibly introduces artifacts (e.g. Neher et al., 2014b). These findings resulted in consequences for the current study. Although the presented SNR was at maximal improvement of the SNR for a realistic noise reduction algorithm, recognition measurements might not show improvements in speech recognition compared to settings without noise reduction. Therefore, to test the measurement procedure described, an a priori knowledge-driven noise reduction algorithm was applied as a reference in the current study. It used a priori knowledge of the background noise and generated fewer artifacts when compared to realistic noise reductions. This algorithm was expected to improve speech recognition scores.

In addition to the evaluation of two noise reduction algorithms, two sets of speech signals were applied. Oldenburg and Göttingen sentences are available for the measure of the TCT. Previous studies showed that different sentence sets result in different TCTs (Schlueter et al., 2014a;

Schlüter et al., 2014b) and are therefore expected to yield different individual adjustments of time compression and SNR for recognition measurements with noise reduction algorithms. Besides the application of individually time-compressed speech for speech recognition scores of about 50% to evaluate noise reduction algorithms, an additional objective of the current study was to recommend one of the two sets of sentences for this approach.

In summary, the objective of the current study was to examine the feasibility of the hybrid approach using the procedure of adaptive adjustment of time-compression to find a hearing situation defined individually by time-compression and fixed positive SNR, which was presented in recognition measurements for the evaluation of noise reduction algorithms. For these evaluation measurements, the following results were expected: First, without noise reduction, speech recognition scores are in the range of 50% when speech is compressed at about TCT. Second, speech recognition after noise reduction does not show any ceiling effects if the change after processing is not too large. Third, the objective evaluation of the noise reduction algorithms shows an overall SNR improvement. Therefore, an improvement in recognition was expected, at least with the a priori knowledge-driven noise reduction algorithm. Furthermore, the current study applied different sets of sentences and may thus be in a position to recommend one set for the assessment of hearing aid algorithms with time-compressed speech.

5.2 Methods

5.2.1 Participants

Eleven hearing-impaired listeners (8 male, 3 female) participated in the measurements. Their mean age was 73 years (range: 68-76 years). All participants had German as their native language. Figure 5.1 depicts hearing thresholds of both of the listeners' ears as measured using air conduction pure tone audiometry. In addition, participants conducted the Trial Making Test and the verbal Digit Span forwards and backwards. These tests explored the cognitive abilities of psychomotor information processing speed and both short-term and working memory. Participants whose results in the cognitive tests were more than 1.5 standard deviations poorer than the mean age-related standard (Härting et al., 2000; Tombaugh, 2004) were excluded from the experiments. Therefore, all participants showed cognitive abilities appropriate to their age or better. All listeners were paid a small amount for their participation to compensate for their expenses.

5.2.2 Materials and measurements

5.2.2.1 OLSA

The Oldenburg sentence test (OLSA, Wagener et al., 1999) is a German matrix test. All sentences have the same structure (name, verb, numeral, adjective, object). Sentences were generated from a random selection of one out of ten words for each structural element of the sentences. After a selection process, the test included 100 sentences with low redundancy, for instance "Peter kauft zehn nasse Messer." ("Peter buys ten wet knifes."). These sentences were

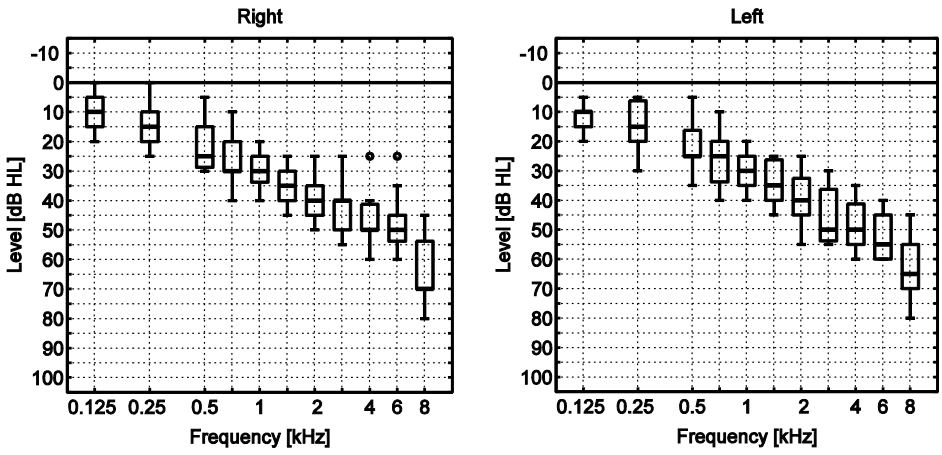

Figure 5.1: Boxplot of the results of pure tone audiometry using air conduction for the left and the right ears of all participants. Plots show the median (bold line within the box), the lower and upper quartile (lower and upper boundary of the box), lowest and highest results within 1.5 times the quartile range relative to the lower and upper quartile (whiskers), and outliers (circles).

sorted to lists with 30 sentences each that had equal recognition. The test included a background noise stimulus. The noise resulted from a superposition of all sentences and has the same long-term spectrum as the speech (Wagener et al., 1999). The OLSA was used to measure recognition scores for sentences. During all measurements, the presentation of the noise was continuous and a visual countdown marked the beginning of the sentence presentation. Participants listened to the sentences and repeated orally as much as they recognized.

5.2.2.2 GÖSA

The Goettingen sentence test (GÖSA) includes sentences from everyday live (Kollmeier and Wesselkamp, 1997). It consists of 200 sentences with three to seven words each. These sentences are apportioned to lists with 20 sentences each. The additional background noise stimulus was developed in the same way as the OLSA noise, with superposition of the speech signals belonging to the Einsilber-Reimtest (Sotscheck, 1982). The speaker was the same for the Reimtest and the GÖSA. Therefore, the long-term spectrum of the noise is similar to the spectrum of the applied speech signals. The same instructions and procedures as in the OLSA were applied to measure recognition scores.

5.2.2.3 Most Comfortable Level (MCL)

The measurement of the MCL was based on the procedure for the acceptable noise level (ANL) test (Nabelek et al., 1991). Participants listened to OLSA sentences in quiet, which were repeated in random order. They adjusted the level of the speech according to the following instructions: At first, speech in quiet was presented at 45 dB SPL and was adjusted to a level louder than comfortable, then afterwards to a level softer than comfortable and finally to the individual MCL. For the first and second speech adjustment ("louder than comfortable", "softer than comfortable"), a larger step size of 5 dB was used, while the last adjustment (individual MCL) used a step size of 2 dB.

For the level adjustments, a graphical user interface with the instruction and control elements was shown to the subjects on a touch screen. Using this interface, the subjects were able to adjust the level of the signals with an up and down button. When the subjects finished the adjustment, the level was approved with an "OK"-button and afterwards the instruction for the next measurement step was shown. The procedure was repeated three times. The median MCL of the three adjustments was calculated, if necessary limited to minimum of 60 or maximum of 80 dB SPL, and then presented as the speech level in the following session.

5.2.2.4 FastOLSA/FastGÖSA

The following description is taken from Section 4.2.2 and supplemented with the method of the FastGÖSA. FastOLSA and FastGÖSA were developed to measure the TCT. These tests were implemented according to Versfeld and Dreschler (2002) and evaluated by Schlueter et al. (2014a) and Schlüter et al. (2014b). Sentences of the OLSA and GÖSA were compressed in time with a pitch-synchronous overlap-add procedure implemented in Praat (Boersma and Weenink, 2009) to different time-compression factors, ρ, as calculated from Equation 5.1, where N is a natural number between 0 and 10. The calculated factors ρ were rounded to the next integer.

$$\rho = 0.85^{N} * 100\ [\%] \qquad (5.1)$$

By varying N, speech was presented at original speech rate (N = 0), but also compressed up to 20% of its original length (N = 10). A 1up-1down procedure decreased or increased the time-compression factor adaptively depending on sentence recognition. For the first sentence within each list, the time-compression factor presented was ρ = 100% (corresponding to original speech). Again, the presentation of the noise was continuous and a visual countdown marked the beginning of sentence presentation. Participants listened to the sentences and repeated orally as much as they recognized.

After presenting one OLSA or one GÖSA list, recognition values (i.e., 0 for at least one mistake in the repetition of the sentence or 1 for correct repetition of all words within a sentence) were available in relation to the respective time-compression factors with which the sentence was processed. To estimate the TCT_N, a maximum likelihood method was applied to the linear steps marked by N, which were calculated from Equation 5.1. Within this method, the discrimination function (see Equation 5.2)

$$p(N) = 1 - \frac{1}{1+e^{4*slope*(TCT_N - N)}} \qquad (5.2)$$

was fit to the data. In Equation 5.2, p is defined as the mean probability that sentences are repeated correctly. This probability is dependent on the time compression described by N. TCT_N denotes the time-compression threshold specified by N, which refers to 50% probability of correct responses. The parameter slope describes the slope of the discrimination function at TCT_N. The result is the parameter setting of TCT_N and slope that produces the observed data with the maximum likelihood. The resulting TCT_N estimate was used for further data analysis. For the presentation of results, N as well as ρ values were given. Respective TCT values were specified by TCT_N or TCT_ρ.

5.2.3 Single-microphone noise reduction algorithms

Single-microphone noise reduction algorithms use speech signals that are distorted with background noise, estimate the noise and filter the distorted signal in order to improve the SNR. The a priori knowledge-driven algorithm (*Apriori*) achieved this with an application of a priori knowledge of the separate speech and background noise signals and the Wiener gain rule for filtering (Vary et al., 1998). This algorithm is not a real world application, because it relies on the separate availability of the two signal components, speech and background noise. The real world algorithm (*Real8dB*), which was also used, estimates the background noise from the distorted speech signal with minimum controlled recursive averaging (Cohen and Berdugo, 2002) and applies spectral subtraction for filtering (Vary et al., 1998). Maximum reduction of both algorithms was set to 6 and 8 dB for *Apriori* and *Real8dB*, respectively.

For the objective measurement of the SNR improvement, the implemented noise reduction algorithms simultaneously processed the distorted speech signal as well as the separated speech and noise signal with the parameters determined for the mixed signal. Mixtures of OLSA and GÖSA sentences with the respective noise were evaluated at different SNRs. Ten seconds of noise preceded the speech. The processed speech signals included sentences concatenated to a monologue of about 1 min length. Separate speech and noise signals at the input and output of the noise reduction were used to calculate the SNRs before and after processing (SNR_{In} and SNR_{Out}). Then, the SNR improvement $SNR_{Out\text{-}In}$ was calculated according to Neher et al. (2014a). Thus the $SNR_{Out\text{-}In}$ was estimated in one-third octave bands and the mean was taken of all bands.

5.2.4 Setup and schedule

The experiments were conducted in a sound-isolation booth. PCs with Matlab-based programming (MathWorks, Natick, MA) controlled the presentation of the signals. Signals were routed through a sound card (Fireface 400, RME, Audio AG, Haimhausen, Germany) and a headphone amplifier (HB 7, Tucker Davis Technologies, Alachua, FL) to headphones (HDA 200, Sennheiser, Wedemark-Wennebostel, Germany). The headphones were free-field equalized according to international standards (IEC 60645-2, 2010; ISO 389-8, 2004) and presented signals diotically. Speech signals were offered at the individual MCL. Taking the presentation levels of all participants together, the mean speech level was 64.5 dB SPL, ranging between 60 and 70 dB SPL for all participants. The background noise was added at 1 and/or 5 dB SNR. The selection of the SNR was dependent on the recognition score obtained in the OLSA or GÖSA with original speech (see following explanations).

In total, participants visited the lab for three sessions on three different days. During the first session, participants conducted a questionnaire about their education, hearing ability, and anamnesis, as well as undergoing audiometric measurements. Afterwards, their cognitive abilities were determined with the Trial Making Test and Digit Span forward and backward. Then participants listened to the original OLSA sentences and adjusted their MCL (see Section 5.2.2.3).

Participants were separated randomly into two groups. During the second session, they started tests with either OLSA or GÖSA sentences and in the third session performed measurements with the remaining speech set. Both sessions consisted of three measurement units with objectives that built on one another. First, the individual positive fixed SNR (1 and/or 5 dB) at which the participants obtained at least 75% recognition for original speech was measured. Second, time-compression for 50% recognition of the speech signals was measured at this selected individual SNR. Third, individual SNR and time compression were applied, to measure speech recognition scores after noise reduction.

For achieving the first objective using the OLSA sentences, participants started with practice runs. They performed recognition measurements with two lists of the original speech signals (one at 5 dB SNR and one at 1 dB SNR) as well as visual confirmation of their answers on a screen. Afterwards, a third list was presented at 1 dB SNR but without visual confirmation. If the participants obtained recognition scores above 75% within the third list, subsequent measurements were conducted at 1 dB SNR. If the recognition scores were equal or below 75%, participants performed a further recognition measurement at 5 dB SNR. In case recognition scores were still below 75%, the participants were excluded from the study. If their scores were equal or above 75%, they performed subsequent measurements at 5 dB SNR.

For achieving the second goal – measurement of the time compression – participants performed the FastOLSA at an SNR (1 or 5 dB) obtained in the first part of the study. Participants trained the FastOLSA with two lists and visual confirmation. Results for the TCT of a final third list without confirmation served as time compression for the final part of the measurements. For this purpose, the measured TCTs were rounded to the next possible time-compression step available in the FastOLSA.

In the third part of the measurements, SNR and time compression determined in the first two parts were used. Three recognition measurements were conducted with time-compressed OLSA sentences processed with *Apriori* and *Real8dB* or without noise reduction (*NoAlgo*).

The measurements with the GÖSA sentences were generally performed similarly to the measurements with the OLSA sentences. The following explanation covers only deviations to the procedure described above. The first part of the measurements with GÖSA sentences started with a recognition measurement at 1 dB SNR. If the resultant recognition score was above 75%, the following measurements were conducted at 1 and 5 dB SNR. In case the recognition scores were equal or below 75%, recognition measurements were repeated at 5 dB SNR. Otherwise, the procedure to determine the SNR (first part of the measurements) was the same as the procedure using OLSA sentences.

In contrast to the procedure with OLSA sentences, the measurement of the TCT and speech recognition scores was conducted with GÖSA sentences without training. If the measurements could be conducted at 1 and 5 dB using the GÖSA sentences, the order of the SNR was chosen randomly. Then, measurements of TCT and recognition scores were performed as a block with the first SNR and repeated with the second SNR. In the entire study, list numbers, session for the presented speech sets (OLSA or GÖSA in first or second session), and order of the noise reduction settings (*NoAlgo*, *Apriori*, *Real8dB*) were randomly selected in the entire study.

5.3 Results

Beside the graphical representation of the results, effects were statistically evaluated with $\alpha = 0.05$. At first, all samples were tested for normal distribution with a Shapiro-Wilk test, which in all cases confirmed that the data were normally distributed. Thus subsequent analyses were conducted using *t*-tests. Some results are displayed with boxplots. These plots show the median using a central bold mark, the edges of the boxes are the lower and upper quartile and the whiskers end at the lowest and highest values that are within 1.5 times the interquartile range. Outliers are displayed by circles and are values outside the range defined by box and whiskers.

5.3.1 TCT

As described above, recognition measurements with the original speech signals determined SNRs for subsequent studies. At 1 dB SNR, eight and seven participants obtained recognitions scores above 75% with the OLSA and the GÖSA, respectively. At 5 dB SNR, three participants reached recognition scores of 75% or more with the OLSA and all eleven participants achieved these scores when performing the GÖSA. As a result, eight/seven participants listened to the OLSA/GÖSA sentences at 1 dB SNR, and three/all eleven participants conducted the OLSA/GÖSA measurements at 5 dB SNR, respectively. For the measurement of the required time compression, the FastOLSA and the FastGÖSA were conducted. Figure 5.2 displays the TCT for the FastOLSA and FastGÖSA at the different SNRs. The TCT_N is between 2 and 6 and therefore speech was presented compressed between 38% and 72% of its original length. An independent samples *t*-test confirmed that the TCT_N is significantly larger for the OLSA sentences than for the GÖSA sentences at 1 dB SNR ($t(13) = 4.01$, $p = 0.001$), i.e. the speech was presented faster in the FastOLSA than in the FastGÖSA condition.

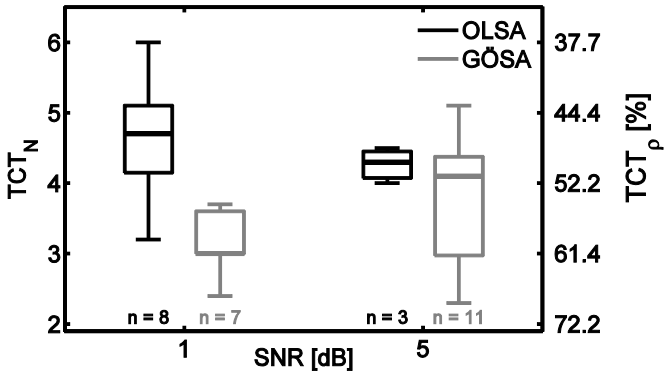

Figure 5.2: Boxplots of TCTs measured with FastOLSA and FastGÖSA at 1 or 5 dB SNR.

5.3.2 Objective improvement

Figure 5.3 shows the SNR improvement $SNR_{Out\text{-}In}$ as a function of SNR at the input of the applied noise reduction algorithms. $SNR_{Out\text{-}In}$ was determined with the original OLSA and GÖSA sentences. Both panels show a maximum in the range of 3-6 dB SNR_{In}. Larger values

for SNR_{Out-In} were observed for *Apriori* than for *Real8dB*. Note that the SNR improvement (SNR_{Out-In}) does not exceed 3 or 4 dB SNR, which is less than the maximum reduction of 8 or 6 dB for the *Real8dB* or *Apriori* algorithm, respectively. Additionally, values for SNR_{Out-In} obtained with GÖSA and OLSA show only small differences for *Real8dB* and *Apriori*.

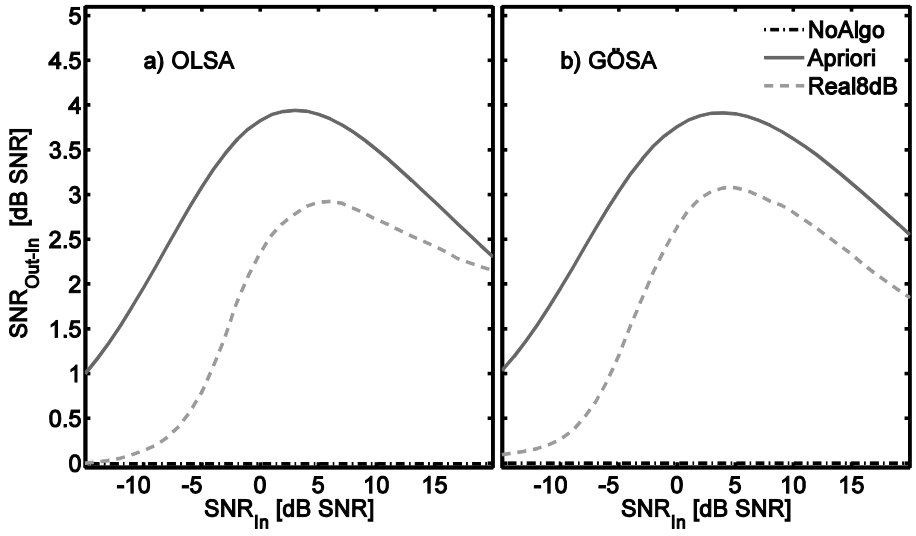

Figure 5.3: Mean objectively-determined SNR improvement SNR_{OUT-IN} of the a priori knowledge-driven (Apriori) and the realistic (Real8dB) noise reduction algorithm and without processing (NoAlgo), as a function of the SNR at the input of the algorithms SNR_{IN} and measured with original a) OLSA and b) GÖSA sentences without time compression.

The fundamental approach in the experiments was the individually-selected time compression of the speech presented to the noise reduction algorithm. Figure 5.4 shows the SNR improvement SNR_{Out-In} measured for the noise reduction algorithms *Apriori* and *Real8dB* when using time-compressed speech. Sentences of the OLSA and GÖSA were compressed in time to N values derived from the TCT measurement (see Sections 5.2.4 and 5.3.1) and presented at 1 and 5 dB SNR before they were processed with noise reduction algorithms. Most values for SNR_{Out-In} are between 2 and 4 dB SNR and therefore showed remaining SNR improvements. Again, larger SNR improvements were measured with *Apriori* than with *Real8dB,* and larger with time-compressed GÖSA sentences than with time-compressed OLSA sentences for *Real8dB*. Furthermore, *Real8dB* depended on the amount of time compression, as can be seen from the fact that the SNR_{Out-In} decreases with increasing N. In addition, the SNR_{Out-In} showed small differences for the SNR_{In}-values of 1 and 5 dB (see Figure 5.4a and b).

5.3.3 Recognition of time-compressed speech at fixed positive SNRs

Figure 5.5 shows recognition scores, which were measured with sentences individually compressed in time. Median recognition scores without any processing (*NoAlgo*) were between 40 and 63%. Although recognition was observed for different speech sets, for different noise reduction algorithms and at different fixed positive SNRs with participants having different hearing ability, all recognition scores were below 100% and therefore showed no ceiling effects. This was in accordance with the expected results for the selected procedure presented here.

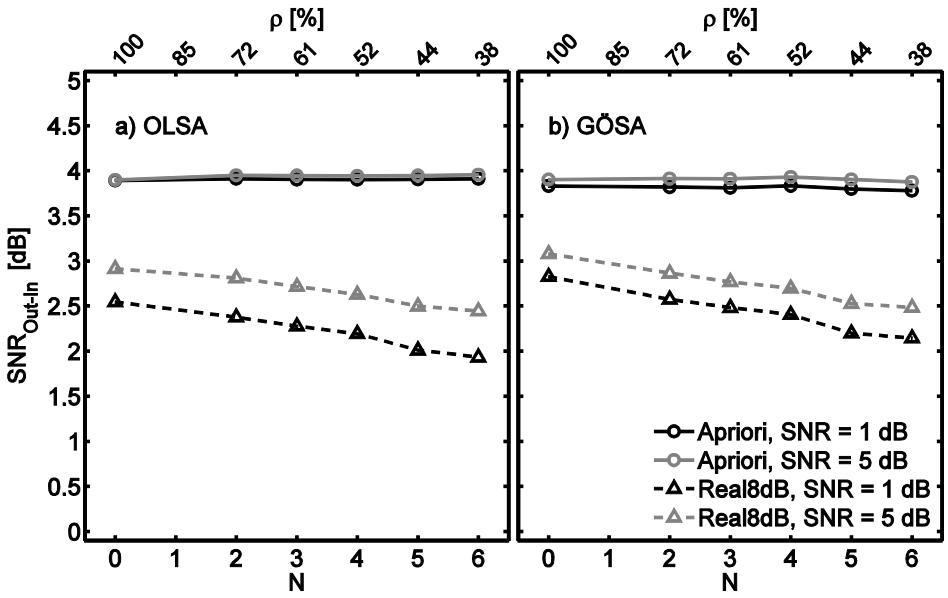

Figure 5.4: Mean objectively-determined SNR improvement $SNR_{Out\text{-}In}$ obtained for OLSA and GÖSA sentences with different time compression at 1 and 5 dB SNR. These signals were processed by the noise reduction algorithms Apriori and Real8dB.

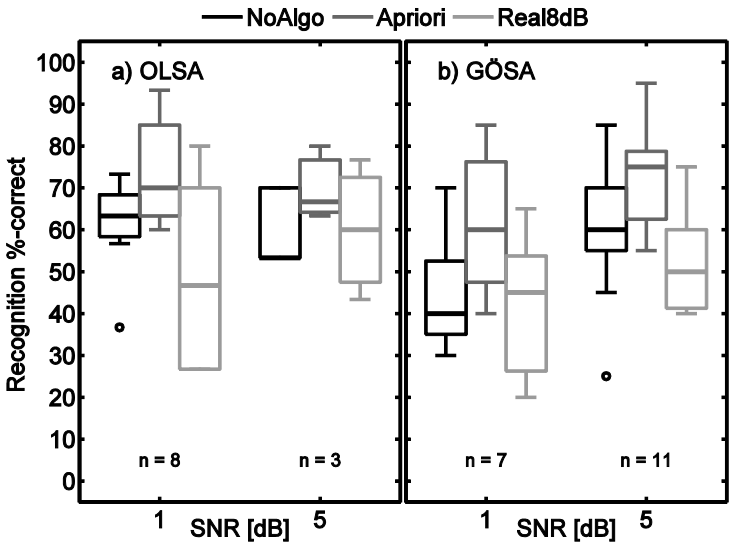

Figure 5.5: Boxplot of the recognition in %-correct for time-compressed a) OLSA and b) GÖSA at 1 or 5 dB SNR without noise reduction (NoAlgo) or processed with the noise reduction algorithms Apriori or Real8dB.

5.3.4 Improvement in recognition after noise reduction

To investigate the benefit of the noise reduction algorithms, improvements in recognition were analyzed by calculating the difference in recognition scores with and without noise reduction $\%\text{-correct}_{\text{Algo}}$ - $\%\text{-correct}_{\text{NoAlgo}}$. This difference is displayed in Figure 5.6. Results showed that only the algorithm *Apriori* improved recognition of time-compressed speech. No improvement, or even deterioration, was observed for the algorithm *Real8dB*. A one-sample *t*-test confirmed these results and showed significant deviation from 0 for *Apriori* using OLSA ($t(10) = 4.25$, $p = 0.002$) and GÖSA ($t(17) = 5.13$, $p < 0.001$) as well as for the combination of *Real8dB* and GÖSA ($t(17) = -2.12$, $p = 0.049$).

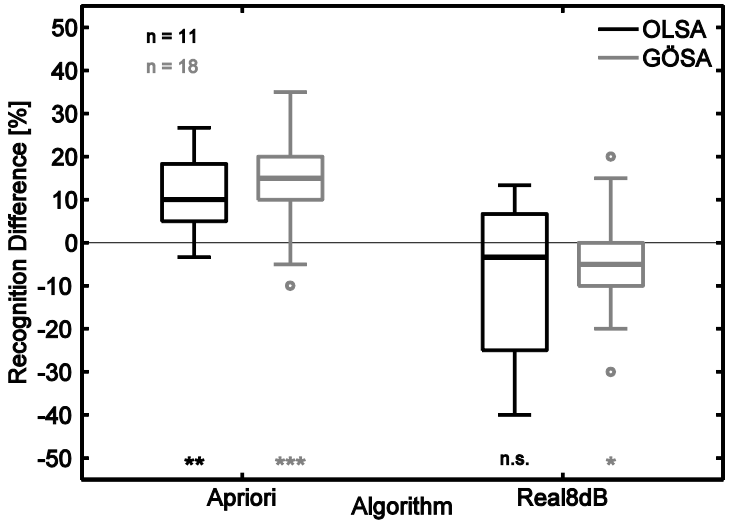

*Figure 5.6: Boxplot of the recognition improvements for the algorithms Apriori and Real8dB. Recognition scores were obtained with time-compressed OLSA or GÖSA sentences. Results of a t-test, which compares the values to 0, are displayed with asterisks (n.s.: not significant; *: p<0.05; **: p < 0.01; ***: p < 0.001).*

5.4 Discussion

The current study tested the feasibility of a new hybrid procedure (adaptive speech-in-noise test with time-compressed speech to select the presentation conditions) characterized by a fixed positive SNR and a fixed time-compression factor, which is, e.g., intended to test hearing aid algorithms for noise reduction processing and SNR improvement. The adaptive speech-in-noise test was used to measure individual TCTs, which are thresholds of time compression for 50% sentence recognition. TCTs were investigated at 1 and/or 5 dB SNR for two different sets of speech signals (OLSA and GÖSA sentences). In subsequent recognition measurements, time-compressed speech was presented at the fixed compression value, which was close to the individual TCT (see Section 5.2.4). This individual adjustment of the time-compressed speech permitted the individual adaptation of the test difficulty for recognition measurements without and with two different noise reduction algorithms.

The objective evaluation of the SNR improvements detected input-SNR-dependent performance in noise reduction as shown in previous studies (e.g., Brons et al., 2013; Fredelake et al., 2012; Hoetink et al., 2009). Both noise reduction algorithms showed the maximum SNR improvement for original speech at positive SNRs, and the maximum improvement was less than the maximum reduction of both algorithms. The calculation of overall SNR improvement was averaged for the entire signal and therefore presumably showed reduced improvement compared to the maximum reduction, which defines the maximum attenuation of local intensity. In addition, the algorithm *Real8dB* had to distinguish between speech and background noise to estimate the noise. Estimation errors may have contributed to the limited maximum improvement of *Real8dB* observed for original speech. Besides, time compression changes speech statistics and, e.g., shifts modulations to higher frequencies (see Schlueter et al., 2014b). Noise reduction algorithms such as *Real8dB* frequently rely on speech statistics to separate speech and noise (for an overview see Chung, 2004). This may also have contributed to the limited improvement of *Real8dB* observed for time-compressed speech in comparison to *Apriori*. Nevertheless, the decrease in SNR improvement was less than 1 dB from original speech as compared to speech with the highest time compression, and an improvement was still obtained for *Real8dB*. In contrast, SNR improvement of *Real8dB* showed higher variability, due to changes of the SNR at the input of a noise reduction algorithm as compared to the differences generated by the changes of the time compression. SNR improvement as a function of SNR at the input varied between no improvement and up to about 3 dB SNR for *Real8dB*. Hence, the presentation of fixed positive SNRs and individually time-compressed speech is presumed to have led to a higher comparability between test situations than in measurements with original OLSA or GÖSA at individual SRTs and therefore varying SNRs.

The objective assessment also revealed differences between OLSA and GÖSA. Those differences might be explained by differences between the stimuli: The OLSA applies sentences and noise that are exactly equal in long-term spectrum. The GÖSA sentences and noise only show similarity of the long-term spectrum. *Real8dB* used the differences of GÖSA's signal components for the estimate and the filtering. Therefore, this algorithm showed a greater SNR improvement for GÖSA than for OLSA. Since the *Apriori* algorithm applied a priori knowledge and has no need to estimate the signal components, it obtained similar SNR improvement for both speech sets.

To adapt the objective assessment for a comparison to the subjective evaluation, it is possible to calculate the overall SNR improvement with the band-importance function from the speech intelligibility index (as described by Neher et al., 2014a). Application of the frequency-dependent weighting, however, showed very similar SNR improvement in comparison to the calculation without weighting and therefore was not addressed further in this study.

A result of using individual time compression settings was that recognition scores obtained without noise reduction algorithm showed median values ranging between 40 and 63%. They were not exactly at 50% recognition, because the TCT measured was rounded to the next possible time-compression step presented in the adaptive procedure. Nevertheless, the recognition scores showed the expected accuracy of participants' results due, e.g., to individual fatigue or learning effects. A further result of the individual adjustment of the hearing situation

was that none of the recognition scores obtained with and without noise reduction algorithms showed any ceiling effects, although measurements were conducted at positive SNRs with different speech sets and different noise reduction processing. All results were included in the analysis, made possible by the careful selection of testing conditions (with a fixed positive SNR and fixed time-compression factor) performed by the adaptive pre-test of the hybrid procedure employed here. This is in contrast to studies by, e.g., Neher et al. (2014b), who had to exclude results because of ceiling effects that occurred in recognition measurements with noise reduction algorithms at positive SNRs.

In addition to preventing ceiling effects, application of time-compressed speech has further advantages. Brons et al. (2013) and Brons et al. (2014) studied the recognition of speech as processed with noise reduction algorithms using normal-hearing and hearing-impaired listeners and presented lower SNRs to normal-hearing than to hearing-impaired participants. Generally, in their studies, the individual presentation of time-compressed speech would allow for an equalization of the presented SNR for both groups of participants. In the current study, it also permitted the adjustment to the performance of the noise reduction algorithms, because both algorithms showed the largest SNR improvement at positive SNR_{In}. Further adjustments to other test situations are possible. In the study of Brons et al. (2013), participants also assessed, e.g., overall preference, noise annoyance and speech naturalness with complex results. In general, overall preference was dependent on the factors noise annoyance and speech naturalness. Individual participants, however, showed inconsistencies in weighting the two factors, depending on the SNR. Therefore, Brons et al. (2013) suggested that the individualization of noise reduction settings in hearing aids might be of benefit to users. In view of these results, in future research the comparison of noise reduction algorithms could be supported by the presentation of time-compressed speech. It also makes it possible to use listening situations with different SNRs but similar recognition. SNRs could be selected to present the highest or equal recognition of the signals processed with noise reductions. Thus, differences between noise reduction algorithms, i.e., their ability to improve SNRs and their generation of artifacts could be studied more precisely than with varying or inappropriate SNRs.

The application of time-compressed speech at positive SNRs and the equalization of SNRs, however, also have limitations. On the one hand, an SNR has to be selected at which participants also achieve high recognition scores for original speech. If the recognition of original speech is too low (e.g., 50 %), FastOLSA and FastGÖSA will not be able to adapt to the TCT. On the other hand, if the SNR is too high, a large amount of time compression (i.e., very fast speech) is necessary to obtain a recognition score of 50%, especially for normal-hearing participants. Due to these limitations of time-compressed speech, floor and ceiling effects can also occur, both in TCT measurements and in recognition measurements with time-compressed speech. Furthermore, the presentation of highly-compressed (i.e., very fast) speech can lead to unrealistic listening situations, because in real life, fast speech is limited by physiological factors (Adank and Janse, 2009). In addition, age effects have to be considered, because the recognition of time-compressed speech decreases with age (e.g. Schlueter et al., 2014a). Nevertheless, this effect is negligible if only relative changes in recognition scores, e.g., with and without noise reduction, are compared, as in the current study.

Analysis of improvements in recognition after noise reduction processing, as compared to unprocessed, only showed improvements for the *Apriori* algorithm. *Real8dB* did not show any improvement. These results are in line with previous research (e.g., Brons et al., 2013, 2014; Hu and Loizou, 2007; Neher et al., 2014b; Nordrum et al., 2006). Schlüter (2007) applied the same noise reduction algorithms for additional measurements, such as SRT measurements using the OLSA, Just Follow Conversation test, and ANL test. She observed a high variability between the results of different participants for the ANL test, and therefore a highly variable SNR improvement of the noise reduction algorithms. For normal-hearing participants, she found a significantly better ANL test results for *Apriori* and *Real8dB* relative to situations without noise reduction processing. For hearing-impaired participants, she found significantly better ANL test values only for *Apriori* as compared to without noise reduction. Furthermore, Schlüter (2007) measured negative SRTs without SNR improvement in the OLSA and the Just Follow Conversation test. For both test procedures and both groups of participants, she only observed significant recognition improvement for the *Apriori* algorithm as compared to situations without noise reduction. In contrast to previous research, positive fixed SNRs were presented in the current study with the objective of maximizing the effect of the noise reduction processing. Consequently, the lack of recognition improvement for *Real8dB* was possibly not due to disadvantageous SNRs for the algorithm.

According to Loizou and Kim (2011), an accurate noise estimate contributes to a better performance of noise reduction algorithms. *Apriori* used a priori knowledge of the background noise. In contrast, *Real8dB* applied an estimate of the background noise together with spectral subtraction. As it reduced the background noise from the distorted signal, it also degraded the speech more than *Apriori*. This was also confirmed by the objective measurements of the SNR improvement, in which *Real8dB* obtained less improvement compared to *Apriori*. Similar results were found by Hu and Loizou (2007), who measured improved recognition scores for a Wiener algorithm applied to car noise with an SNR of 5 dB. Other algorithms failed to show better recognition performance. According to Loizou and Kim (2011), distortions were introduced by algorithms applying, e.g., Wiener filter and spectral subtraction, and resulted from over- and underestimates of spectra by the noise reduction algorithm. Higher recognition performance resulted from better control of these estimates. Jørgensen and Dau (2011) explained decreased speech recognition of signals processed with spectral subtraction using an analysis of the envelope power-spectrum model. The model showed a decreased SNR at the output of a modulation-selective process for speech processed with noise reduction. More innovative and sophisticated noise reduction algorithms, as compared to the algorithms applied here might show improvements in recognition, especially at fixed positive SNRs.

Beside the aspect of time-compressed speech, two sets of speech signals were applied in the current study, OLSA and GÖSA sentences. The results of the TCT measurements showed higher TCT_N values for OLSA sentences than for GÖSA sentences. This means that, to reach equal recognition, the OLSA sentences had to be compressed by a greater amount than the GÖSA sentences. These results are in line with Schlüter et al. (2014b), who compared the FastOLSA to the Fast GÖSA. In addition, Schlüter et al. (2014b) found no training effects for the FastGÖSA, while FastOLSA had to be trained to obtain reliable results. Consequently,

GÖSA sentences and their modifications are more suitable for assessing noise reduction algorithms than are OLSA sentences. Disadvantages of the GÖSA sentences are the limited number of lists, and that repetitions of lists have to be avoided because participants are able to remember complete sentences.

Conclusions

The application of individually time-compressed speech in speech-in-noise tests for the assessment of single-microphone noise reduction algorithms led to the following conclusions:

- The feasibility of the hybrid approach introduced here was demonstrated with two different noise reduction algorithms. In the first phase of the approach, time-compressed speech was adaptively adjusted to the individual hearing ability, a predefined as well as fixed positive SNR and different sets of speech signals was used. In the second phase, the evaluation of the algorithms was performed with recognition measurements at these individually assigned hearing situations, using fixed positive SNR and time compression.
- The advantage of the approach is the ability to present speech and to test the algorithms under study at a predefined SNR, which may be selected to demonstrate the maximum effect of the noise reduction algorithm. Hence, the effective noise reduction processing and a correspondingly large SNR improvement was demonstrated for at least one of the two algorithms tested here.
- No ceiling effects of recognition scores were observed, because the effect of the algorithms considered here showed limited recognition changes close enough to the point of approx. 50% recognition that was achieved without the algorithm. This might not have been the case if algorithms with a larger effect on speech recognition were tested.
- Presentation of positive SNR resulted in improved recognition scores only for the *Apriori* algorithm. The *Real8dB* algorithm, however, did not show a convincing noise reduction effect in recognition measurements, although overall SNR improvement was objectively documented for the SNR values and time-compression factors presented.
- GÖSA sentences, rather than OLSA sentences, are more suitable for research, because less time compression and no training are necessary.

Acknowledgements

I would like to thank Vera Löw for her support on data collection and G.A. Manley (www.stels-ol.de) for advising on language issues.

Parts of this work were presented on the International Hearing Aid Conference (2014) in Tahoe City, USA.

6

General conclusions and future perspectives

The main objective of this thesis was to develop and evaluate a speech-in-noise test that presents fixed positive SNRs. Positive SNRs represent realistic communication situations (Olsen, 1998; Smeds et al., 2015) and are necessary for beneficial processing of some hearing aid algorithms (Naylor, 2010). Fixed SNRs permit taking the SNR-dependent processing of some hearing aid algorithms (e.g., single-microphone noise reduction algorithms) into account but also the comparison between normal-hearing and hearing-impaired participants. Therefore, a speech-in-noise test was developed and evaluated that adaptively adjusted the speech rate in order to measure a threshold of 50% recognition at fixed positive SNRs.

Different algorithms exist for the time compression of speech samples to increase the speech rate. In **Chapter 2** the uniform PSOLA (as implemented in Praat, Boersma and Weenink, 2009) and the non-uniform Mach1 (Covell et al., 1998) were compared. Both algorithms were used to compress sentences of the OLSA at different speech rates. In this comparison, the non-uniform algorithm exhibited greater deviations from the targeted time compression, as well as greater changes of the phoneme duration, spectra, and modulation spectra. Therefore, signals of the uniform algorithm are expected to be more similar to original speech. As a result, participants showed higher recognition scores for Praat than for Mach1. Subjective and objective measures indicated a clear advantage of the uniform algorithm in comparison to the non-uniform algorithm for the application in speech-in-noise tests. However, very high speech rates with a time-compression factor below 30% of the original sample length were necessary to reach 50% recognition at positive SNRs. Additionally, discrimination functions measured with time-compressed speech showed a shallower slope compared to original speech material, indicating limitations of the speech-in-noise test and of its ability to discriminate across different effective SNRs. This was caused by recognition that depends on the word's position within the sentences. As a result, sentence scoring was recommended and was applied in a speech-in-noise test presenting time-compressed speech. In addition, sentence scoring was expected to increase the time-compression factor for 50% recognition, because the recognition of an entire sentence depends on the least intelligible word.

Besides the relation of recognition to the time-compression algorithm, a dependency of recognition and learning of a certain type of speech material is known and was taken into account for the evaluation of a speech-in-noise test with time-compressed speech (see **Chapter 3**). Therefore, a series of measurements of speech recognition thresholds was performed for original and time-compressed speech, in five sessions each with six repeated lists. Generally, speech recognition thresholds improved with repeated measurements. The largest improvements were observed within the first measurements of the first session. These improvements were larger for time-compressed than for original speech. The observed perceptual learning process was explained on the basis of the Reverse Hierarchy Theory (e.g., Banai and Lavner, 2012). This suggests that learning of speech in a matrix test progresses through an initial general phase, with the access to abstract acoustic representations of high levels of the auditory pathway. This allows for relatively good initial performance in everyday life. However, this is followed by a prolonged learning phase of specific lower-level representations that are beneficial for the current task. Based on this idea, the RHT predicts that learning that is based on high–level representations can be generalized and transferred to different tasks (i.e., is accessible for perception of original and time-compressed speech in the same way). On the other hand, learning that applies lower–level representations is task-specific (i.e., is accessible only for perception of original or time-compressed speech). The observed partial transfer of what was learned between conditions of time-compressed and original speech confirms this theory. In general, learning progresses as long as it is beneficial for the current task. Therefore, the results presented might not resolve small learning effects after long learning phases, due to the accuracy of the method of SRT-measurements applied. Nevertheless, the observed learning effects permit and require consequences for speech audiometry in clinical, and especially in scientific, applications of matrix tests in speech audiometry that use repeated measurements. Studies using either original or time-compressed speech materials should include training lists in each session. In addition, a careful randomization of test situations across sessions, as well as recruiting experienced listeners, is appropriate for both speech materials, especially when small differences between hearing situations need to be analyzed. Improvements observed between sessions for time-compressed speech indicate large learning effects and therefore measurements should be carried out within one session. This separated observation is not necessary for the original speech material.

Results of Chapter 2 and 3 were applied to study and evaluate a speech-in-noise test using time-compressed speech (see **Chapter 4**). This test used an adaptive procedure to adjust the time compression of sentences presented at a fixed positive SNR. Two different adaptive methods were compared: the first one was based on Versfeld and Dreschler (2002) and the second on Brand and Kollmeier (2002). Analysis of the measurements regarding list lengths and estimation strategies for thresholds showed that a practical method for measuring the time compression for 50% recognition can be derived. This method applied time-compression adjustment and step sizes proposed by Versfeld and Dreschler (2002), with sentence scoring, lists of 30 sentences, and a maximum likelihood method for threshold estimation. In the evaluation measurements, older participants obtained higher test-retest reliability than younger participants. Additional analysis of learning effects led to requirements of including one or two lists in

training prior to data collection, depending on the group of listeners. Appendix B includes an estimate of the test accuracy and a comparison to the original OLSA.

In **Chapter 5,** the speech-in-noise test introduced was applied to study hearing aid algorithms. The test was used to increase the difficulty of the speech material individually for each participant, in order to present fixed positive SNRs and to decrease the possibility of ceiling effects in recognition measurements. Using this individual adaptation, recognition scores of time-compressed speech were measured with two different single-microphone noise reduction algorithms and without processing. Results confirmed that speech with an individually adjusted time-compression factor can avoid ceiling effects of recognition scores because of prior knowledge of hearing ability. Also, a given, fixed positive SNR makes it possible to select the appropriately compressed speech material for each individual participant. Hence, SNR values can be adjusted to yield large overall SNR improvements of noise reduction processing. The procedure makes it possible to exclude negative as well as variable SNR with missing or variable overall SNR improvement. Generally, everyday sentences of the Göttingen sentence test have to be compressed less than matrix-type sentences of the Oldenburg sentence test, and for Göttingen sentences, no training effects are expected. Therefore, Göttingen sentences are better suited for studies with time-compressed speech at fixed positive SNRs.

Overall, the development of a speech-in-noise test that offers fixed positive SNRs was successful. Besides the presentation of fixed positive SNRs due to the adaptive adjustment of time compression, its advantages are the presentation of a simple task for the participants (repeating the understood words) and the simple adaptive method. This method is based on a small but efficient number of time-compression steps and therefore requires little effort in signal processing in advance of the test. A freely available algorithm implemented in Praat is used for compressing the speech signals in time, which offers a good signal quality. In addition, the procedure allows for presentation of the signals online. This means that, for example, participants can conduct the measurements in a setup with loudspeakers, but also that differences with and without hearing aids can be analyzed. Hence, recording hearing aid processing for presentation to participants is not necessary. Besides these practical aspects, the procedure offers a more complex hearing situation using time-compressed speech in background noise and therefore a higher difficulty compared to speech in noise or in quiet. This probably makes it possible to study speech recognition that is not only affected by sensory, but also by cognitive abilities (Uslar, 2014; Wingfield et al., 2006). Furthermore, the results of the test can be used to adapt hearing situations. Thus hearing situations can be adjusted to measure recognition with a decreased possibility for ceiling effects.

Nevertheless, the procedure also has limitations. Since the recognition of the words depends on the words' positions within the sentences, sentence scoring, instead of word scoring, had to be applied for assessing sentence recognition. Sentence scoring denotes the technique of counting a response as correct only if every word of a sentence was correctly recognized, whereas word scoring describes counting the number of correctly recognized words in a sentence. Therefore, using word scoring offers a higher informational content for the strategy for estimating the resulting thresholds than sentence scoring.

Another constraint of the method is that the participants showed large learning effects during this procedure and this limits the reliability of the test. Therefore, combining training and measurements within one session is recommended.

Even though the test procedure offers a realistic situation in terms of positive SNRs, the application of very fast speech, especially for young normal-hearing listeners, is necessary. For further development of the test procedure and presentation of even more realistic hearing situations, additional interfering factors could be inserted. It is expected that these additional factors influence speech recognition and lead to lower speech rates. Work published by Rønne et al. (2013) and Simonsen et al. (2014) could support this idea because they systematically investigated the influence of spectral separation of the target speaker and of different interfering speakers on the speech recognition threshold. Moreover, signals could be processed with reverberation. Warzybok et al. (2013) studied recognition due to interaction between early reflections in rooms and to binaural processing. They observed a decay in recognition for frontal reflections having a delay of more than 25 ms. Additional deterioration of recognition was found for frontal reflections with a delay of 200 ms, while spatial separation of reflections decreased the effect of recognition deterioration. This study indicates that frontal reflections with a delay of 25 ms or more could probably increase the difficulty of the speech-in-noise test. Further studies of George et al. (2010), Rennies et al. (2014) and Holube et al. (2014b) analyzed sentence recognition after systematic changes of the reverberation and background noise in the context of the speech transmission index (STI). For mixtures of reverberation and noise, they observed increasing intelligibility with increasing STI. These results provide insights into the interacting effects of noise and reverberation with regard to the additional application of reverberation in more complex listening situations. Additionally, the studies of Holube et al. (2014b), Rennies et al. (2014), and George et al. (2010) could be models for the analysis of speech recognition that depends on systematic variation of time compression of the speech signal and SNR of the background noise. However, the STI is not expected to be the appropriate model for the systematic study of relations between time-compressed speech and background noise. After time compression, the statistics of speech signals is changed and, for example, the modulation spectrum is shifted to higher frequencies, which probably results in a miscalculation of the STI.

The above-mentioned variation of signals statistics after time compression can also affect processing of hearing aid algorithms, if they rely on parameters changed by the time compression. Therefore, the application of the procedure for the evaluation of hearing aid algorithms could be limited. Nevertheless, it is possible to account for this processing, depending on the time compression. As an example, Chapter 5 shows an application for single-microphone noise reduction algorithms. In contrast to the algorithms described and utilized in the current thesis, future work should focus on noise reduction algorithms that yield fewer artifacts and a larger improvement in recognition. However, future research need not focus exclusively on the evaluation of noise reduction algorithms. In general, hearing aids, their algorithms, and the interaction of these algorithms can be tested with time-compressed speech and with fixed ecologically-relevant SNRs. For example, this can be illustrated by noise reduction algorithms and dynamic compressions, both of which are implemented in modern hearing aids. Dynamic compression

and noise reduction reduce the gain if a high level or low SNR is present. In hearing aids, noise reduction processes the signals first and dynamic compression follows. This may lead to reduction of gain after the noise reduction, while subsequently the dynamic compression applies a higher gain due to the attenuated level of the signal (Holube et al., 2014a). Chung (2007) as well as Anderson et al. (2009) found, to some extent, effects of interactions between both algorithms on speech recognition and sound quality. When time-compressed speech is used to present positive SNRs in future research, interactions could be analyzed in hearing situations relevant for hearing aid users and with a decreased possibility for ceiling effects.

Generally, the test can be applied to adapt the test situation for each participant to equal SNR and similar recognition, even though the participants have different hearing ability. This advantage is not only applicable to recognition measurements as explained above. Brons et al. (2013) showed that subjective evaluation of naturalness and annoyance of signals processed with noise reduction algorithms depends on the SNR. Application of time-compressed speech could result in a compensation of the SNR effect. Furthermore, the presentation of time-compressed speech could be applied in the method of subjective adjustment developed by Wittkop (Wittkop et al., 1997; Wittkop, 2001) to increase the difficulty of the listening situation. As explained above, if the time-compressed speech is used to adapt the difficulty of the speech material, participants should have fewer uncertainties in adjusting "equal recognition", as they showed for original speech and background noise at SRT (Schlüter et al., 2014a). Furthermore, the signals could be presented at different SNRs in the reference signal. It is hypothesized that participants would obtain a reliable subjective criterion for the evaluation of the time-compressed reference and test signal. If they perceive an improvement of the subjective speech recognition in the reference signal, they probably will be able to adjust this in the test signal as well. Nevertheless, participants perform the comparison using their individual criteria to account for the perceived processing differences between the reference and the test signal.

Besides the application with hearing aids and their algorithms, the test procedure offers more extensive analysis of speech recognition compared to the test procedure with original speech. Speech recognition is dependent on sensory and cognitive abilities of listeners (e.g., Uslar, 2014; Wingfield et al., 2006) and effects of both aspects appear markedly when processing gets difficult (Wingfield et al., 2006). Again, time-compressed speech can be applied to vary the difficulty of a hearing situation and to reach listening situations in which both sensory and cognitive abilities are necessary for performance. Since sensory and cognitive abilities vary with age, the presentation of time-compressed speech permits a more detailed study of differences between younger and older participants than tests with speech delivered at normal rates. According to Wingfield et al. (2006), it is not clear whether this observation is caused by sensory deficits (e.g., in temporal and in spectral resolution) or cognitive abilities (loss of processing time at the linguistic level, see also Chapter 4). For the analysis of these effects, the studies by Uslar (2014) and Wingfield et al. (2006) could serve as a model. In both studies, the cognitive load was varied using sentences of different linguistic complexity and the relation to age and hearing ability was analyzed. Wingfield et al. (2006) showed that comprehension accuracy of time-compressed complex sentences in a 'who-did-it' paradigm was affected by age and time-compression factor (speech rate) and hearing ability. Uslar (2014) complemented this study

and presented sentences of different linguistic complexity in different listening conditions that were specified by different background noises to younger and to older, normal-hearing listeners, as well as to older hearing-impaired participants. In total, Uslar (2014) found that participants show fewer interactions between sentence complexity and background noise if they are older and have a declining hearing ability. This means that older participants use top-down and expectation-driven processing for speech recognition in difficult hearing situations (of particularly ambiguous sentences) and therefore show less variation of recognition in dependence on the background noise as compared to young participants. The approach of Uslar (2014) could be repeated with time-compressed speech, to vary the difficulty of the hearing situation. It is expected that increasing difficulty of the sentences would result in a higher application of top-down and experienced-driven processing, perhaps even for young normal-hearing participants. As a result, interactions between sentence complexity and hearing situations should not occur. Furthermore, it would be interesting to investigate whether older normal-hearing and older hearing-impaired participants apply cognitive abilities to the same extent for the compensation of their sensory deficits in recognition of time-compressed speech, as they did for measurements with background noise (Uslar, 2014).

All these future perspectives show that presentation of time-compressed speech and background noise can support the study and analysis of speech recognition and hearing aid development. At present, the main application of the procedure is probably in scientific studies. When sufficient experience with the procedures described in the current thesis has been gained, they will possibly also be applied in clinical studies, e.g., evaluating the rehabilitative success of hearing aids, hearing aid algorithms or adjustments at ecologically relevant SNRs. Furthermore, the test could perhaps even be applied to the detailed diagnostics of speech recognition problems in complex hearing situations caused by certain combinations of sensory and cognitive deficits.

Appendix A

Evaluation eines Einregelungsverfahrens zur Bestimmung des Nutzens einkanaliger Algorithmen zur Störgeräuschreduktion

Evaluation of an adjustment method to determine the benefit of single-microphone noise reduction algorithms

Zusammenfassung: Algorithmen zur Störgeräuschreduktion in Hörgeräten führen theoretisch vor allem bei positiven Signal-Rausch-Verhältnissen (SNR) zu einer Reduktion des Störgeräuschs. Diese Reduktion ist schwierig mit Satztestverfahren nachzuweisen, da diese Tests vor allem bei negativen SNR sensitiv sind. Deshalb wurde die Anwendbarkeit eines anderen subjektiven Testverfahrens zu Evaluation der Wirkung von Störgeräuschreduktionsalgorithmen untersucht. Bei dem von Wittkop et al. (1997) vorgeschlagenen Verfahren wird der SNR eines Testsignals, bestehend aus Sprache und Hintergrundgeräusch, so eingeregelt, dass die Sprachverständlichkeit derjenigen eines Referenzsignals entspricht. Das Referenzsignal wird dabei mit einem Störgeräuschreduktionsalgorithmus verarbeitet, so dass die Verbesserung des SNR durch den Algorithmus ermittelt werden kann. Das Einregelungsverfahren wurde für drei verschiedene SNRs durchgeführt - bei der individuellen Sprachverständlichkeitsschwelle (SRT), bei positivem SNR von 5 dB und beim Mittelwert aus diesen beiden Werten. Als Spezialfall wurde in dieser Studie außerdem das Shadow-Filtering-Verfahren (Kallinger et al., 2009) genutzt. Dafür arbeiten die Algorithmen bei einem SNR mit maximaler Störgeräuschreduktion und die dabei berechneten Koeffizienten werden auf die separat vorliegenden Signale (Sprache und Rauschen) angewendet.

Dadurch ist es möglich, eine Einregelung bei einem negativen SNR durchzuführen, obwohl die Störgeräuschreduktion bei einem positiven SNR angewandt wurde. Die Ergebnisse von zwölf normalhörenden und zwölf schwerhörigen Probanden (nach Ausschluss von vier Probanden mit Problemen bei der Testdurchführung) zeigen, dass die Einregelung beim SRT besser gelingt als bei einem positiven SNR von 5 dB. Signifikante Verbesserungen des SNR durch die Störgeräuschreduktion konnten nur mit dem Shadow-Filtering-Verfahren nachgewiesen werden, das jedoch aufgrund des notwendigen a-priori-Wissens über die Signalanteile bei realen Hörgeräten nicht angewendet werden kann.

Abstract: Noise reduction algorithms theoretically reduce noise mainly at positive signal-to-noise ratios (SNR). This improvement is difficult to verify with sentence-in-noise tests because these are most sensitive at negative SNRs. Therefore, the applicability of another subjective adjustment method for the evaluation of noise reduction algorithms was examined. This adjustment method was proposed by Wittkop et al. (1997). For this method, the SNR of a test signal, mixed speech and noise, is adjusted to the same speech intelligibility as those of a reference signal. The reference signal is processed with a noise reduction algorithm. Therefore, the improvement of the SNR by the algorithm can be determined. The adjustment method was applied at three different SNRs - at the individual speech reception threshold (SRT), at a positive SNR of 5 dB and at the mean of both values. Additionally, the shadow-filtering method proposed by Kallinger et al. (2009) was used for comparison. For this method, the algorithms are applied at a SNR with maximum noise reduction and the speech and noise signals are filtered separately with the respective coefficients. This procedure allows for an adjustment at negative SNRs even if the noise reduction is applied at positive SNRs. Results for twelve normal hearing and twelve hearing impaired listeners (after exclusion of four listeners with problems during test execution) showed that the adjustment is easier to perform at SRT compared to a positive SNR of 5 dB. Significant improvements of the SNR due to the noise reduction algorithm were only observed for the shadow-filtering method. Unfortunately, this method requires a priori knowledge about the parts of the signal and can therefore not be applied to real world hearing instruments.

Adapted from:
Schlüter, A., Aderhold, J., Koifman, S., Krüger, M., Nüsse, T., Lemke, U., and Holube, I. (2014) “Evaluation eines Einregelungsverfahrens zur Bestimmung des Nutzens einkanaliger Algorithmen zur Störgeräuschreduktion - Evaluation of an adjustment method to determine the benefit of single-microphone noise reduction algorithms”, Zeitschrift für Audiologie, 53(2), 50-58.

A.1 Einleitung

Schwierigkeiten bei der Kommunikation in Anwesenheit von Hintergrundgeräuschen sind allgemein bekannt und eines der häufigsten Probleme, von denen Hörgeräte-Träger berichten. Ansätze zur Kompensation dieses Problems in Hörgeräten sind Richtmikrofonsysteme und Algorithmen zur Störgeräuschreduktion. In der vorliegenden Untersuchung wird die Wirkung des zweiten Ansatzes, sogenannter einkanaliger Störgeräuschreduktionen[2] analysiert. Die Aufgabe der Störgeräuschreduktionsalgorithmen besteht darin, den Signal-Rausch-Abstand (engl. Signal-to-Noise Ratio, SNR) durch die Unterdrückung des Hintergrundgeräusches anzuheben und damit das Sprachverstehen zu verbessern. Dieses wird je nach eingesetzter Störgeräuschreduktion unterschiedlich realisiert. Einkanalige Störgeräuschreduktionen müssen in der Lage sein, aus dem gemischten Eingangssignal (bestehend aus Sprache und Hintergrundgeräusch) das Rauschen zu schätzen und herauszufiltern. Die Filterung kann anschließend z. B. über das spektrale Subtraktionsverfahren (Vary et al., 1998) realisiert werden und hat die größte Wirkung, wenn das Eingangssignal einen positiven SNR besitzt (d.h. die Sprache lauter ist als das Hintergrundgeräusch). Bei negativen SNR-Werten am Eingang zeigt sich nur eine geringe oder gar keine Verbesserung des SNR, da die Algorithmen in diesem Bereich Schwierigkeiten haben, Sprache und Rauschen voneinander zu trennen (Fredelake et al., 2012).

Zur Hörgeräteevaluation werden häufig Sprachverständlichkeitstests im Störgeräusch wie der Oldenburger Satztest (OLSA, Wagener et al., 1999c) eingesetzt. Dabei wird die Sprachverständlichkeitsschwelle (engl. Speech Reception Threshold, SRT) ermittelt, welche den SNR-Wert angibt, bei dem 50% des Sprachmaterials verstanden wurde. Bei Normalhörenden ist den Referenzwerten des OLSAs zufolge ein Ergebnis von -7,1 dB SNR mit einer Standardabweichung von 1,1 dB SNR zu erwarten (Wagener et al., 1999a). Erfahrungen zeigen, dass auch schwerhörige Probanden sehr häufig negative SNR-Werte erreichen (z. B. Wagener and Brand, 2005). Dies führt dazu, dass bei der Evaluation von Störgeräuschreduktionsalgorithmen ein Signal mit negativem SNR präsentiert wird, bei dem, wie zuvor beschrieben, die Störgeräuschreduktion nur einen geringen bzw. gar keinen Nutzen zeigt (Marzinzik and Kollmeier, 2002).

Für die Evaluation einer einkanaligen Störgeräuschreduktion muss nicht nur der Eingangs-SNR dem Algorithmus entsprechend gewählt werden. Als weitere Anforderung muss auch die Aufgabenstellung für die Probanden einfach genug sein, um zuverlässige und aussagekräftige Ergebnisse zu erhalten. Ein dafür in Frage kommendes Testverfahren ist der SNR-Vergleich nach Wittkop et al. (1997). Dabei werden ein Referenz- und Testsignal jeweils bestehend aus Sprache und Rauschen dargeboten. Das Referenzsignal wird mit einem Störgeräuschreduktionsalgorithmus verarbeitet und bei einem festen SNR präsentiert. Die Probanden haben die Aufgabe, den

[2] Störgeräuschreduktionen werden unterschieden nach der Anzahl der verwendeten Mikrofone in Ein-, Zwei- oder Mehrkanal-Ansätze. Das bedeutet, dass den Signalverarbeitungsalgorithmen zur Nutzsignalverbesserung bei einem einkanaligen Ansatz das vermischte Signal von Sprache und Rauschen zur Verfügung steht, während beim zweikanaligen Ansatz als Eingangssignale zwei vermischte Signale und bei einem mehrkanaligen Ansatz mit n Kanälen n Signale anliegen. Die Anzahl der Kanäle ist nicht zu verwechseln mit der Anzahl der Frequenzkanäle innerhalb der Signalverarbeitung des Hörgerätes.

Rauschpegel des unverarbeiteten Testsignals zu verändern und damit die Verständlichkeit von Referenz- und Testsignal anzugleichen. Wittkop et al. (1997) ermittelten mit diesem Verfahren eine SNR-Verbesserung zwischen 1 und 4 dB durch eine Störgeräuschreduktion. Dabei entsprach der SNR des Referenzsignals dem individuellen SRT-Wert der Probanden.

Schlüter et al. (2010) verwendeten das von Wittkop et al. (1997) entwickelte Verfahren bei vier verschiedenen Referenz-SNR-Werten (-5, 0, 3, 5 dB SNR). Dabei konnte allerdings nur ein geringer Effekt der Störgeräuschreduktionsalgorithmen nachgewiesen werden. Im Median lag die maximale Verbesserung nach der Störgeräuschreduktion, d.h. die Differenz zwischen dem von den Probanden eingestellten Testsignal-SNR und dem Referenz-SNR (ΔSNR), für normalhörende Probanden bei 1 dB SNR (bei einem Referenz-SNR von 5 dB). Für schwerhörige Probanden zeigte sich eine maximale Verbesserung von 2 dB SNR (bei einem Referenz-SNR von 3 dB). Diese eher geringen Verbesserungen wurden auf die hohen Referenz-SNR-Werte, bei denen die individuellen Sprachverständlichkeitsschwellen der Probanden nicht berücksichtigt wurden, zurückgeführt. Außerdem führte die Einregelung zu einer großen Streuung bei positivem Referenz-SNR, die vermutlich auf die Unsicherheit der Probanden beim Vergleich der Verständlichkeit von zwei deutlichen Signalen hindeutete.

Eine Möglichkeit, die im Vergleich zu positiven SNR bessere Differenzierbarkeit in der Sprachverständlichkeit nahe dem individuellen SRT und die bestmögliche Leistung der Algorithmen bei positivem SNR zu verbinden, zeigten Kallinger et al. (2009). Sie verarbeiteten das mit Rauschen unterlegte Sprachsignal mit Hilfe eines Störgeräuschreduktionsalgorithmus bei positivem und damit für den Algorithmus optimalem SNR und wendeten die dabei berechneten Koeffizienten auf die getrennt vorliegenden Signalanteile (Sprache und Rauschen) an. Danach wurden die getrennten verarbeiteten Signale bei einem neuen SNR kombiniert. Dieses Verfahren, das als „Shadow-Filtering“ bezeichnet wird, bietet die Möglichkeit, den Probanden ein Referenz-Signal bei ihrem individuellen SRT anzubieten. Bei diesem sollte ihnen die Einregelung gut gelingen, da auch kleine Veränderungen im SNR in schwierigen Hörsituationen gut wahrnehmbar sind. Gleichzeitig wird ein Signal dargeboten, das die Eigenschaften des jeweiligen Algorithmus adäquat abbildet, da dieser bei seinem optimalen Eingangs-SNR verwendet werden kann. Dabei ist allerdings zu beachten, dass sich diese Methode nur für Evaluationsmessungen eignet, die „offline“ (ohne Hörgerät) und mit Verarbeitung der Signale mittels PC-Software durchgeführt werden, da a-priori-Wissen über das Sprach- und Störsignal genutzt wird, um die Signale separat zu verarbeiten.

Das Ziel dieser Untersuchung war die Überprüfung der Eignung des Einregelungsverfahrens, um den Nutzen von Störgeräuschreduktionsalgorithmen subjektiv bestimmen zu können. Dazu wurden die Erfahrungen von Schlüter et al. (2010), Wittkop et al. (1997) und Kallinger et al. (2009) miteinander verknüpft. Im Folgenden wird zunächst das Konzept des Einregelungsverfahrens vorgestellt und dann dessen Eignung am Beispiel eines Algorithmus zur Störgeräuschreduktion analysiert. Da die Wirkung des Algorithmus wie bei Fredelake et al. (2012) beschrieben vom SNR am Eingang abhängt, wurde einerseits ein SNR im Maximum der Wirkungsweise, d. h. 5 dB, gewählt. Bei diesem Eingangs-SNR wird eine Verbesserung des SNR um ca. 3 dB erwartet, jedoch ist die Einregelung durch die Probanden aufgrund der sehr hohen

Sprachverständlichkeit vermutlich unsicher. Deshalb wurde andererseits der SNR auf den individuellen SRT eingestellt, da an der Schwelle vermutlich eine höhere Messgenauigkeit erzielt werden kann. Beim SRT zeigt die Störgeräuschreduktion jedoch voraussichtlich keine oder nur eine geringe Wirkung. Ein weiterer Eingangs-SNR wurde in der Mitte zwischen SRT und 5 dB gewählt. Um die Vorteile der beiden Messbedingungen SRT (verlässliche Einregelung) und 5 dB (maximale Wirkung der Störgeräuschreduktion) zu vereinen, wurde zusätzlich der Ansatz von Kallinger et al. (2009) mit Shadow-Filtering verfolgt. Bei diesem Verfahren wird sowohl eine Wirkung der Störgeräuschreduktion von ca. 3 dB als auch eine verlässliche Einregelung erwartet.

Die unterschiedlichen Testsituationen wurden sowohl von normalhörenden als auch von schwerhörigen Probanden durchlaufen. Im Vergleich der beiden Gruppen werden aufgrund der Hörschädigung und des höheren Alters der Schwerhörigen unterschiedliche Streubereiche bei den SNR-Verbesserungen durch den Störgeräuschreduktionsalgorithmus erwartet. Darüber hinaus wird vermutet, dass der gemessene Effekt in der Verbesserung der subjektiven Sprachverständlichkeit für die schwerhörigen Probanden stärker ausgeprägt ist als für die normalhörenden Probanden, da bei den schwerhörigen Probanden vermutlich eine größere Toleranz gegenüber Artefakten durch die Signalverarbeitung besteht. Somit könnte für die schwerhörigen Probanden das verbesserte Sprachverstehen und nicht die möglicherweise veränderte Signalqualität im Vordergrund stehen, während diese die Bewertung durch die normalhörenden Probanden beeinflussen könnte.

A.2 Methoden

A.2.1 Stimuli

Für alle Testverfahren wurde das Sprachmaterial des OLSA verwendet. Zusätzlich wurde das zugehörige stationäre Rauschen (OLnoise) verwendet, das das gleiche Langzeitspektrum wie die Sprache besitzt, da es aus mehrfachen Überlagerungen der OLSA-Sätze generiert wurde. Der verwendete Störgeräuschreduktionsalgorithmus nutzte zur Rauschschätzung das Verfahren des Minima Controlled Recursive Averaging (Cohen and Berdugo, 2002) und zur Filterung des Signals das Verfahren der Spektralen Subtraktion (Vary et al., 1998). Die maximale Reduktion des Hintergrundgeräusches wurde auf 8 dB beschränkt.

A.2.2 Testverfahren

A.2.2.1 Bestimmung des Most Comfortable Level

Zunächst wurde der persönlich angenehmste empfundene Sprachpegel (engl. Most Comfortable Level, MCL) bestimmt. Als dargebotenes Sprachmaterial wurden dafür die Sätze des OLSA in einer kontinuierlichen Darbietung ohne Störgeräusch verwendet. Analog zum Verfahren des Acceptable Noise Level (ANL, Nabelek et al., 1991) stellten die Probanden zunächst einen zu lauten Pegel ein. Anschließend stellten die Probanden einen zu leisen Sprachpegel ein. Zum Abschluss wurde der Pegel angenehmer Sprachlautstärke eingestellt. Die Schrittweite betrug 5 dB für die Einstellung „zu laut“ und „zu leise“. Für die Einstellung „angenehm“ betrug die

Schrittweite 2 dB. Der beschriebene Zyklus wurde drei Mal durchgeführt. Der MCL ergab sich als Median der drei eingestellten Messwerte jedes Zyklus. Die individuellen MCL-Werte wurden als Präsentationspegel für Sprache in den darauffolgenden Messungen verwendet.

A.2.2.2 Oldenburger Satztest

Der SRT wurde mit Listen von 30 Sätzen bestimmt. Im Unterschied zu dem üblichen Vorgehen beim OLSA wurde in dieser Untersuchung der Rauschpegel adaptiv gesteuert. Der Sprachpegel entsprach dem individuellen MCL. Das Rauschen wurde zwischen den einzelnen Satzdarbietungen unterbrochen.

A.2.2.3 Einregelungsverfahren

Den Probanden wurden zwei Signale, das Referenz- und das Testsignal, aufeinander folgend präsentiert. Sowohl das Referenz- als auch das Testsignal bestanden jeweils aus Sätzen des OLSA und OLnoise. Für eine paarweise Darbietung von Referenz- und Testsignal wurde jeweils der gleiche Satz verwendet. Für die nächste Kombination wurde entsprechend ein anderer Satz zufällig ausgewählt. Jeweils zuerst dargeboten wurde das Referenzsignal, das während eines Messdurchlaufs konstant gehalten wurde. Das Referenzsignal wurde mit dem Störgeräuschreduktionsalgorithmus verarbeitet und bei einem festen SNR präsentiert. Zeitlich nach dem Referenzsignal wurde jeweils das unverarbeitete Testsignal angeboten. Die Aufgabe der Probanden bestand darin, den Rauschpegel des unverarbeiteten Testsignals zu verändern und damit die Verständlichkeit von Referenz- und Testsignal anzugleichen. Der SNR des Testsignals zu Beginn jeder Einregelung wurde randomisiert und lag immer 3-7 dB über oder unter dem SNR des Referenzsignals. Dadurch unterschied sich in jedem Fall der Höreindruck von Referenz- und Testsignal zu Beginn jeder Einregelung. Die Probanden konnten durch entsprechende Auswahl auf einem Touchscreen den Rauschpegel um 1 oder um 3 dB erhöhen bzw. verringern. Nach jeder Pegelveränderung folgte automatisch eine Wiedergabe beider Signale mit der veränderten Einstellung des Testsignals. Bei Wahl der Option „Wiederholen“ wurden beide Signale mit der letzten Pegeleinstellung wiederholt. Wenn der Proband den Eindruck hatte, dass er die Aufgabe erfüllt hatte, so konnte mit der „OK“-Taste die Einregelung beendet werden.

Das Einregelungsverfahren wurde mit drei verschiedenen Verarbeitungsarten (*NoAlgo*, *Real8dB*, *Shadow-Filtering*) bei drei verschiedenen SNR-Bedingungen (*SRT*, *mean*, *5dB*) durchgeführt (siehe Tabelle A.1). Die unterschiedlichen Testkonditionen werden im Folgenden erläutert.

Tabelle A.1: Überblick über die verwendeten Testkonditionen

Table A.1: Overview of the test conditions

	SRT	*mean*	*5dB*
NoAlgo	x	x	x
Real8dB	x	x	x
Shadow-Filtering	x		

Die erste Kondition *NoAlgo* stellt die Kontrollsituation dar, in der sowohl das Testsignal als auch das Referenzsignal unverarbeitet, d.h. ohne Störgeräuschreduktion, dargeboten wurden. Die Aufgabe der Probanden bestand also in dieser Kondition darin, den vorgegebenen SNR im Referenzsignal mit dem Testsignal nachzubilden. Der vorgegebene SNR im Referenzsignal konnte dabei den mit dem OLSA bestimmten SRT (*SRT*), einen SNR von 5 dB (*5dB*) bzw. einen SNR, der dem Mittelwert aus beiden Werten entspricht (*mean*), betragen (siehe Abbildung A.1).

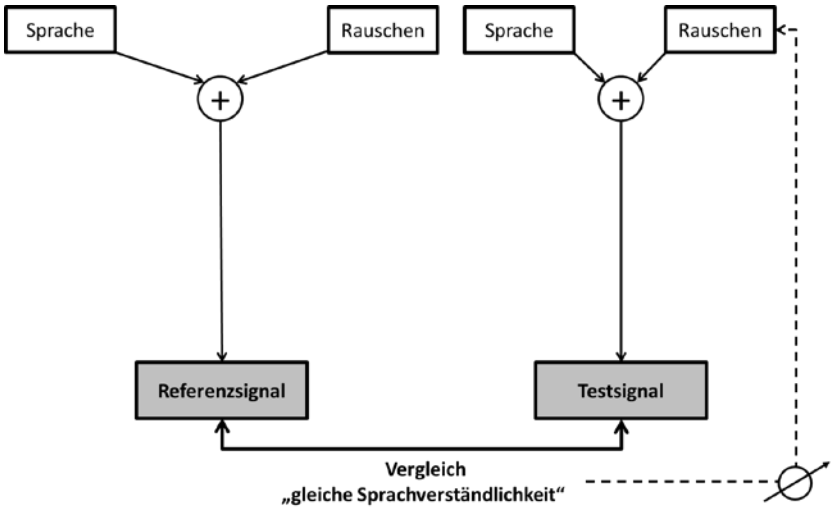

Abbildung A.1: Kondition NoAlgo. Kontrollsituation in der das Referenzsignal ohne Verarbeitung durch die Störgeräuschreduktion dargeboten wird. Der Rauschpegel des Testsignals wird verändert und damit die Verständlichkeit von Referenz- und Testsignal angeglichen.

Figure A.1: Condition NoAlgo. Control situation without noise reduction applied to the reference signal. The noise of the test signal was adjusted for similar speech intelligibility of reference signal and test signal.

Bei der zweiten Testkondition *Real8dB* wurde das von Wittkop et al. (1997) vorgeschlagene Verfahren verwendet (siehe Abbildung A.2). Dabei wurden Sprache und Störgeräusch im gewünschten SNR, d.h. wiederum bei *SRT*, *mean* bzw. *5dB* gemischt und der Störgeräuschreduktionsalgorithmus auf das resultierende Signal angewandt.

In der dritten Kondition *Shadow-Filtering* wurde das von Kallinger et al. (2009) vorgeschlagene Verfahren genutzt (siehe Abbildung A.3). Dabei wurden Sprache und Störgeräusch zunächst bei einem SNR von 5 dB gemischt und dem Störgeräuschreduktionsalgorithmus zugeleitet. Die Koeffizienten des Algorithmus wurden gespeichert und auf Sprache und Störgeräusch getrennt angewandt. Beide verarbeiteten Signale wurden dann bei dem SNR, der dem individuellen SRT entspricht, gemischt und als Referenzsignal dargeboten.

A.2.3 Messablauf

Die Messungen fanden jeweils an einem ca. zweistündigen Termin statt. Zunächst wurden ein kurzes Anamnesegespräch und eine Otoskopie durchgeführt, um sicherzustellen, dass die

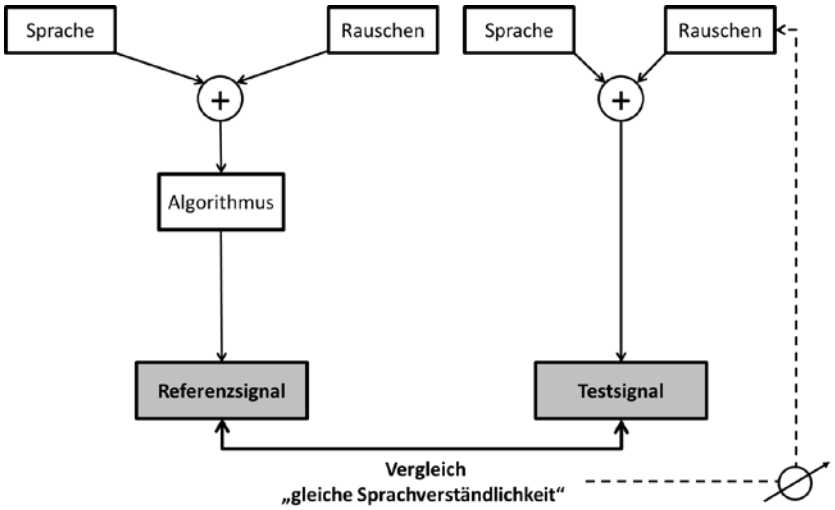

Abbildung A.2: Kondition Real8dB. Das Referenzsignal wird durch den Störgeräuschreduktionsalgorithmus verarbeitet. Der Rauschpegel des Testsignals wird verändert und damit die Verständlichkeit von Referenz- und Testsignal angeglichen.

Figure A.2: Condition Real8dB. The noise reduction algorithm was applied to the reference signal. The noise of the test signal was adjusted for similar speech intelligibility of reference signal and test signal.

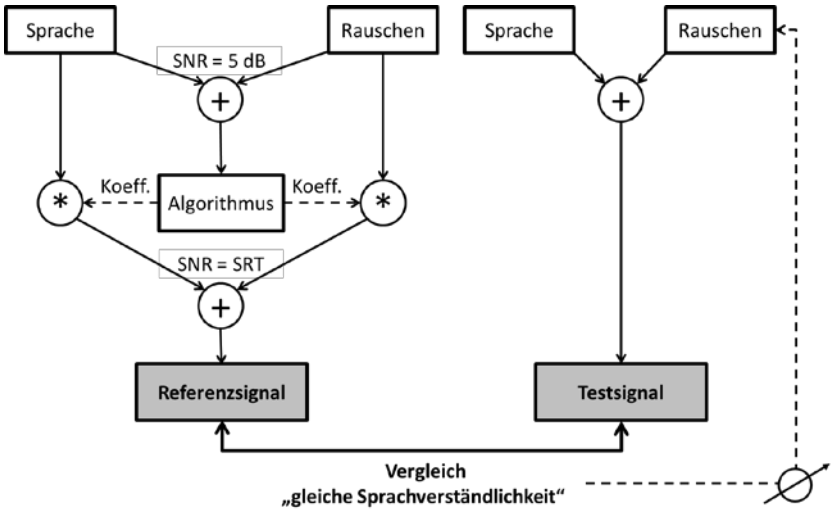

Abbildung A.3: Kondition Shadow-Filtering. Die Koeffizienten des Störgeräuschreduktionsalgorithmus werden für das gestörte Sprachsignal bei 5 dB SNR berechnet und anschließend auf das Sprach- und Rauschsignal getrennt angewendet. Als Referenzsignal werden beide verarbeiteten Signale dargeboten, die beim SNR, der dem individuellen SRT entspricht, gemischt wurden.

Figure A.3: Condition Shadow-Filtering. The coefficients of the noise reduction algorithm were calculated for a SNR of 5 dB and applied separately to speech and noise. The reference signal is derived by mixing of speech and noise with a SNR that corresponds to the individual SRT.

Probanden keine akuten Ohrerkrankungen hatten und der Gehörgang frei war. Nach dem Tonschwellen-Audiogramm wurde der MCL bestimmt, der als Grundlage für den Sprachpegel während der OLSA-Messung und dem Einregelungsverfahren verwendet wurde. Danach wurden zwei Trainingslisten und darauffolgend eine Testliste des OLSA durchgeführt.

Der gemessene SRT wurde für die gleichnamige Testsituation im Einregelungsverfahren verwendet. Im Anschluss wurde von jedem Probanden das Einregelungsverfahren durchlaufen. Dieses wurde zunächst dreimal geübt und dann mit je drei Messwiederholungen für die zuvor erläuterten Varianten durchgeführt. Die Abfolge der dargebotenen Referenz-SNR-Werte (*SRT*, *mean* oder *5dB*) und der Konditionen (*NoAlgo*, *Real8dB*, *Shadow-Filtering*) war dabei randomisiert. Jedoch wurden zunächst alle Einregelungen (sechs bzw. neun) für einen SNR-Wert durchgeführt bevor der nächste SNR-Wert dargeboten wurde.

A.2.4 Messaufbau

Für die im Weiteren vorgestellten Messungen wurde ein PC mit den Software-Paketen „Oldenburger Messprogramme“ (OMA, Hörtech gGmbH) und Matlab (Distribution R2010) verwendet. Die Darbietung der digital vorliegenden Sprach- und Rauschsignale erfolgte über eine externe Soundkarte mit zugehörigem Wandler (RME Fireface 400). Alle Signale wurden nach Verstärkung durch den externen Kopfhörerverstärker (Tucker Davis Technologies, TDT-HB7) diotisch über einen Kopfhörer des Typs HDA 200 (Sennheiser) dargeboten.

A.2.5 Probanden

An den Messungen nahmen insgesamt 28 Probanden teil, die in zwei Gruppen eingeteilt wurden. Die erste Gruppe bestand aus zwölf jungen Normalhörenden (7 weibl., 5 männl., Alter: 23-28 Jahre, MW: 25 Jahre) mit einem maximalen Hörverlust von 20 dB HL im Bereich von 125 Hz bis 8 kHz. Die zweite Gruppe bestand aus 16 älteren Probanden mit einem Hörverlust (11 weibl., 5 männl., Alter: 38-76 Jahre, MW: 66 Jahre), die aus der Datenbank des Hörzentrums Oldenburg rekrutiert wurden. Von dieser Gruppe wurden vier Probanden nachträglich aufgrund von Schwierigkeiten in der Umsetzung des Einregelungsverfahrens ausgeschlossen (vgl. Abschnitt A.3.2), so dass zwölf ältere Probanden mit einem Hörverlust (9 weibl., 3 männl., Alter: 38-76 Jahre, MW: 64 Jahre) verblieben. In Abbildung A.4 ist die Verteilung der Tonaudiogramme der gesamten schwerhörigen Probandengruppe dargestellt. Der Aufwand der schwerhörigen Probanden wurde mit 12 Euro pro Stunde vergütet.

A.3 Ergebnisse

A.3.1 MCL und OLSA

Abbildung A.5 zeigt die Ergebnisse der Bestimmung des MCL für die normalhörenden und die schwerhörigen Probanden. Im Median regelten die Normalhörenden den MCL für Sprache auf einen Pegel von 60 dB SPL (min: 43 dB SPL, max: 75 dB SPL) ein, während die schwerhörigen Probanden einen MCL von 69 dB SPL (min: 57 dB SPL, max: 80 dB SPL) erreichten.

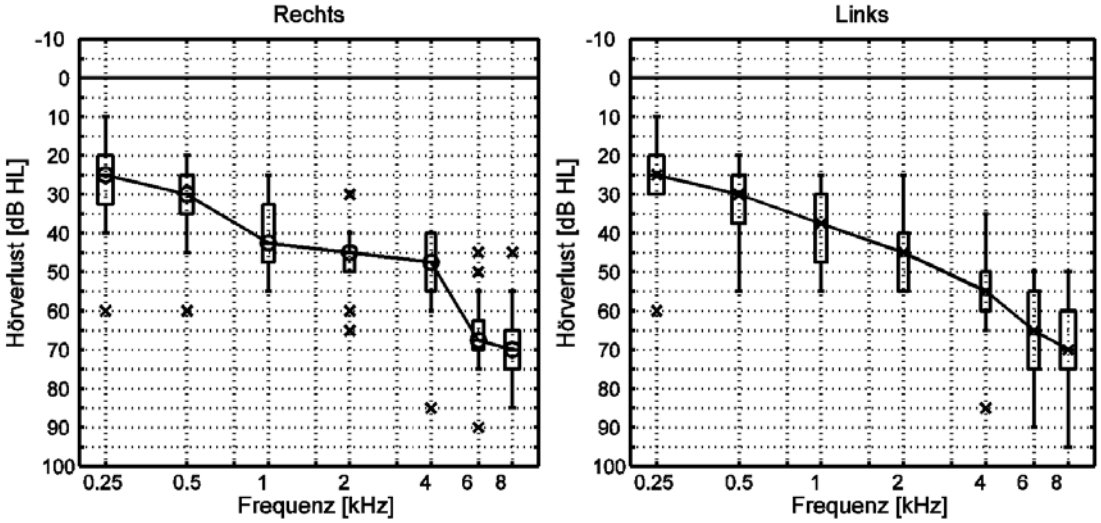

Abbildung A.4: Tonaudiogramme der schwerhörigen Probanden. Links sind die Ergebnisse der rechten Ohren und rechts die Ergebnisse der linken Ohren dargestellt.

Figure A.4: Results of a pure tone audiometric testing. The left and right plots show results of the right and left ears, respectively.

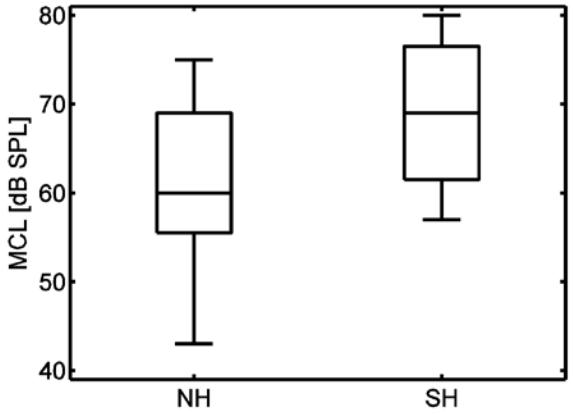

Abbildung A.5: MCL-Werte der normalhörenden (NH) und der schwerhörigen (SH) Probanden.

Figure A.5: Most comfortable level (MCL) values of the normal-hearing (NH) and the hearing-impaired (SH) participants.

Wie schon beschrieben (s. Abschnitt A.2.2), wurden die individuellen MCL-Werte als Präsentationspegel für Sprache in den darauffolgenden Messungen verwendet.

Der OLSA ergab für die Normalhörenden im Median einen SRT von -7,5 dB SNR (max: -5,7 dB SNR; min: -8,4 dB SNR) und für die Schwerhörigen von -4,8 dB SNR (max: -3,3 dB SNR; min: -6,1 dB SNR). Die Ergebnisse sind in Abbildung A.6 aufgetragen.

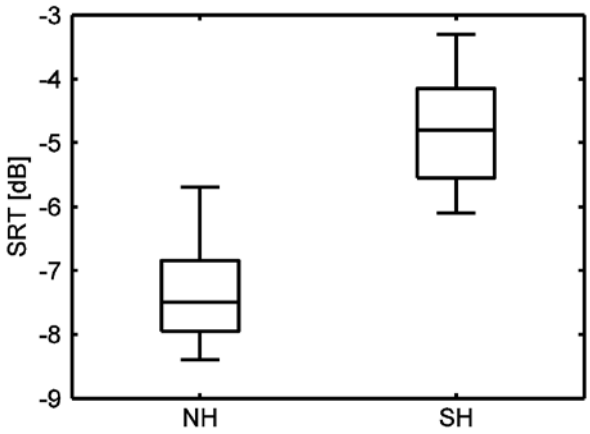

Abbildung A.6: SRT-Ergebnisse des OLSA für die normalhörenden (NH) und die schwerhörigen (SH) Probanden.

Figure A.6: Speech Reception Threshold (SRT) values of the normal-hearing (NH) and the hearing-impaired (SH) participants measured with OLSA.

A.3.2 Einregelungsverfahren

Zur Auswertung der Ergebnisse des Einregelungsverfahrens wurde der Median der eingeregelten SNR-Werte aus den drei Messwiederholungen jeder Testsituationen berechnet. Anschließend wurde die Differenz zwischen dem median eingestellten SNR des Testsignals und dem Eingangs-SNR des Referenzsignals ermittelt. Diese Größe wird im Folgenden als SNR-Verbesserung durch den Störgeräuschreduktionsalgorithmus bezeichnet (ΔSNR = eingestellter SNR - Referenz-SNR). Da fast alle Daten Normalverteilung zeigen (geprüft mit dem Shapiro-Wilk-Test), wurden die Ergebnisse mit parametrischen Tests und einem Signifikanzniveau von 5% statistisch geprüft.

Abbildung A.7 zeigt die Ergebnisse für die Normalhörenden in den drei SNR-Bedingungen. Für alle drei SNR-Bedingungen wurde die Einregelung für die Kondition *NoAlgo* durchgeführt (jeweils linker Boxplot). Für diese Kondition weichen die Mediane nur wenig vom Soll-Wert von ΔSNR = 0 dB ab. Die statistische Überprüfung mit einem t-Test gegenüber Null ergibt für die SNR-Bedingungen *5dB* und *mean* keinen signifikanten Unterschied (*5dB*: p = 0,339; *mean*: p = 1,0). Für den *SRT* findet sich zwar eine signifikante Abweichung von Null (p = 0,021), allerdings ist die Abweichung mit einem ΔSNR von im Median -1 dB (Mittelwert: -0,8 dB) gering. Die Probanden konnten also zum Großteil in der Kontrollsituation die Einstellung sicher durchführen.

In der Kondition Real8dB (rechter bzw. mittlerer Boxplot in Abbildung A.7 steigt der ΔSNR wie erwartet mit Erhöhung des Eingangs-SNR an. Die größte Verbesserung mit einem Median von ΔSNR = 1,5 dB ergibt sich für einen Eingangs-SNR von 5 dB. Nur mit *Shadow-Filtering* (rechter Boxplot für die SNR-Bedingung *SRT*) kann diese Verbesserung mit einem Median von ΔSNR = 2 dB noch übertroffen werden. Ein Vergleich zur Situation *NoAlgo* ergibt sogar im Median eine Verbesserung von ΔSNR = 2,5 dB (*5dB*) bzw. ΔSNR = 3 dB

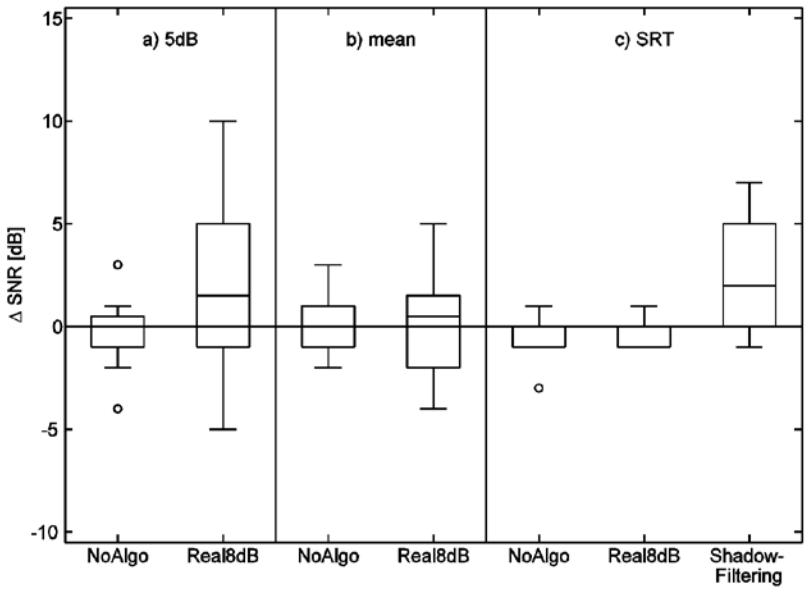

Abbildung A.7: SNR-Verbesserung (ΔSNR) der normalhörenden Probanden gemessen im Einregelungsverfahren. Der dargebotene SNR im Referenzsignal entsprach a) 5 dB, b) dem Mittelwert aus individuellem SRT und 5 dB SNR oder c) dem individuellen SRT gemessen im OLSA.

Figure A.7: SNR improvement (ΔSNR) for the normal-hearing listeners in the adjustment method. The presented SNR in the reference signal was a) 5 dB, b) the mean of individual SRT and 5 dB SNR or c) the individual SRT measured with OLSA. As expected, condition NoAlgo did not differ from zero. For Condition Real8dB, the difference between reference signal and test signal (ΔSNR) is increasing from SRT to 5dB. The largest ΔSNR is observed for Shadow-Filtering.

(*Shadow-Filtering*). Die Unterschiede zwischen den Konditionen mit Störgeräuschreduktion und der Kondition *NoAlgo* wurden mit einem *t*-Test statistisch überprüft. Bei der SNR-Bedingung *SRT* zeigt nur *Shadow-Filtering* ($p < 0{,}001$) einen signifikanten Unterschied zu *NoAlgo*. Bei allen anderen Paarvergleichen war keine signifikante Verbesserung durch die Störgeräuschreduktion nachweisbar (*5dB*: $p = 0{,}090$; *mean*: $p = 0{,}812$; $SRT_{Real8dB}$: $p = 0{,}053$).

Neben den Unterschieden in den Verbesserungen durch die Störgeräuschreduktion sind zusätzlich noch Unterschiede in der Varianz der Messergebnisse beobachtbar. Diese können einen Hinweis auf die Schwierigkeit des Einregelungsverfahrens bei den verschiedenen SNR-Bedingungen geben. In der Kondition *Real8dB* ist die Streubreite bei *5dB* deutlich größer als beim *SRT*. Die Interquartilsspannweite für *Real8dB* liegt in der *SRT*-Kondition bei 1,0 dB, in der *mean*-Kondition bei 3,5 dB und in der *5dB*-Kondition bei 6 dB. Um diese Beobachtungen statistisch zu überprüfen, wurden die absoluten Differenzen zwischen dem individuellen Median der ΔSNR-Werte der einzelnen Probanden und den Gruppen-Medianen berechnet sowie dann für die Testsituationen *Real8dB* die Ergebnisse verschiedener SNR-Präsentationen mit dem *t*-Test verglichen. Dabei zeigte sich eine signifikant abweichende Differenz, wenn *Real8dB* bei *SRT* mit den anderen SNR-Bedingungen verglichen wurde (*5dB*: $p = 0{,}006$; *mean*: $p = 0{,}014$). Dies bedeutet, dass sich die Streubreite der Ergebnisse gemessen in der Testsituation *Real8dB* signifikant unterscheidet und mit höherem SNR zunimmt.

Die Ergebnisse der schwerhörigen Probanden sind in Abbildung A.8 dargestellt. Für die Kondition *NoAlgo* sind keine signifikanten Abweichungen des ΔSNR von Null nachweisbar (*5dB*: $p = 0{,}741$; *mean*: $p = 0{,}059$; *SRT*: $p = 0{,}077$). Die größte Verbesserung durch die Störgeräuschreduktion ergibt mit ΔSNR = 4 dB die Kondition *Shadow-Filtering*. Die zweitgrößte Verbesserung mit ΔSNR = 2,5 dB wird für *Real8dB* in der *5dB*-Situation erreicht. Der Vergleich zur Situation *NoAlgo* ergibt im Median Verbesserungen von ΔSNR = 3,5 dB (*Shadow-Filtering*) und ΔSNR = 3 dB (*Real8dB*, *5dB*). Wird der *t*-Test auf die Fragestellung angewendet, ob sich eine Veränderung der ΔSNR-Werte gegenüber der *NoAlgo*-Situation bei Anwendung eines Algorithmus ergibt, so ist diese Veränderung für keine der getesteten SNR-Situationen signifikant (*5dB*: $p = 0{,}119$; *mean*: $p = 0{,}774$; *SRT*: $p = 0{,}252$; *Shadow-Filtering*: $p = 0{,}075$). Die im Vergleich zu den Normalhörenden (Abbildung A.8) größere Varianz der Daten deutet allerdings auf vermehrte Schwierigkeiten bei der Einregelung hin. Deshalb wurden in einer weiterführenden Betrachtung nur die Ergebnisse derjenigen schwerhörigen Probanden berücksichtigt, die in der Kondition *NoAlgo* bei *SRT* den ΔSNR relativ zuverlässig in einem Bereich zwischen -3 und 3 dB einregeln konnten. Diese Grenzen wurden aus pragmatischen Gründen so gewählt, dass einerseits Probanden mit extremen Unsicherheiten ausgeschlossen wurden und andererseits die Anzahl der ausgewerteten Ergebnisse und damit die Probandengruppe nicht zu klein wurden. Dadurch reduzierte sich die Probandengruppe von 16 auf zwölf Probanden. Die entsprechenden Ergebnisse sind in Abbildung A.9 dargestellt.

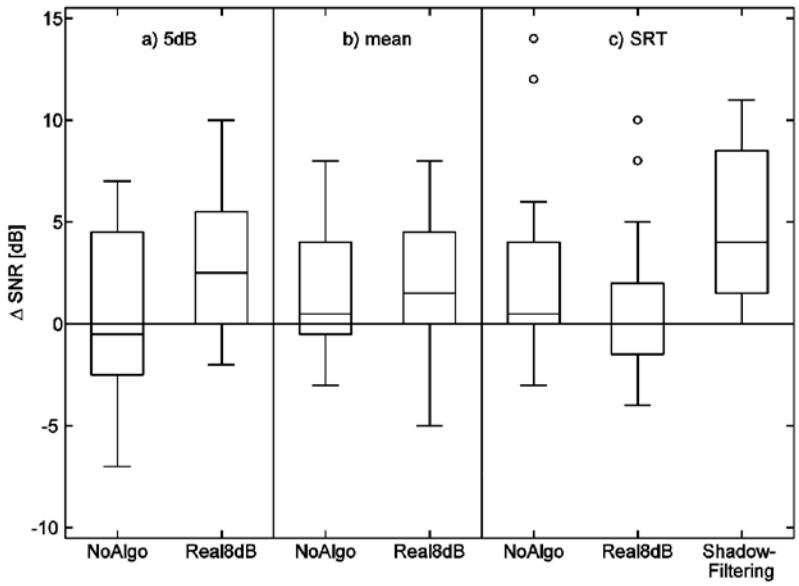

Abbildung A.8: SNR-Verbesserung (ΔSNR) der schwerhörigen Probanden gemessen im Einregelungsverfahren. Der dargebotene SNR im Referenzsignal entsprach a) 5 dB, b) dem Mittelwert aus individuellem SRT und 5 dB SNR oder c) dem individuellen SRT gemessen im OLSA.

Figure A.8: SNR improvement (ΔSNR) for the hearing-impaired listeners in the adjustment method. The presented SNR in the reference signal was a) 5 dB, b) the mean of individual SRT and 5 dB SNR or c) the individual SRT measured with OLSA. As expected, condition NoAlgo did not differ from zero. Due to the large variances, no other condition is significantly different from NoAlgo.

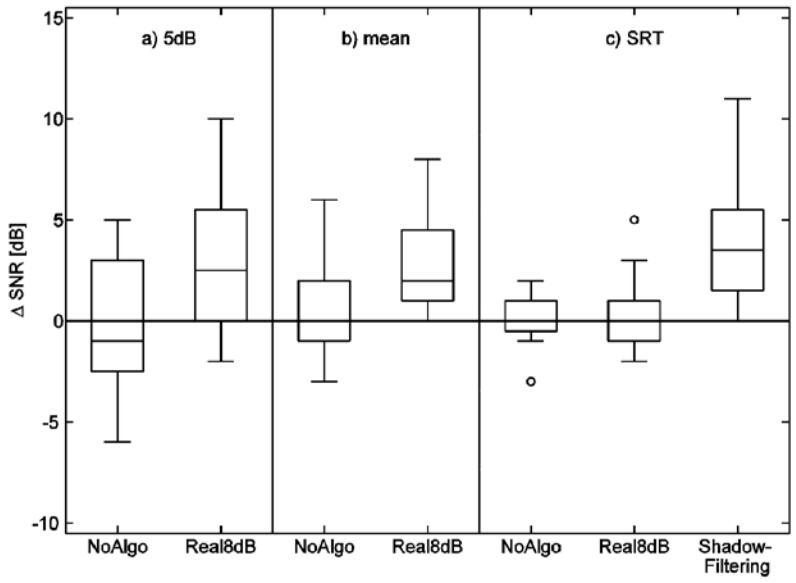

Abbildung A.9: SNR-Verbesserung (ΔSNR) der reduzierten Gruppe schwerhöriger Probanden gemessen im Einregelungsverfahren. Der dargebotene SNR im Referenzsignal entsprach a) 5 dB, b) dem Mittelwert aus individuellem SRT und 5 dB SNR oder c) dem individuellen SRT gemessen im OLSA.

Figure A.9: SNR improvement (ΔSNR) for the reduced hearing-impaired group in the adjustment method. The presented SNR in the reference signal was a) 5 dB, b) the mean of individual SRT and 5 dB SNR or c) the individual SRT measured with OLSA. Only the condition Shadow-Filtering shows a significantly different ΔSNR from the condition NoAlgo.

Die größte Verbesserung von $\Delta\text{SNR} = 3{,}5$ dB ergab sich auch bei dieser reduzierten Probandengruppe für die Kondition *Shadow-Filtering*. Die Verbesserung in der Kondition *Real8dB* beträgt im Median $\Delta\text{SNR} = 2{,}5$ dB für *5dB* und $\Delta\text{SNR} = 2$ dB für *mean*. Der Vergleich zur Situation *NoAlgo* zeigt im Median Verbesserungen von $\Delta\text{SNR} = 3{,}5$ dB (*Shadow-Filtering*), $\Delta\text{SNR} = 3{,}5$ dB (*Real8dB*, *5dB*) und $\Delta\text{SNR} = 2$ dB (*Real8dB*, *mean*). Auch für diese reduzierte Probandengruppe wurden die zuvor erläuterten statistischen Tests angewendet. Es zeigte sich, dass die Einregelung bei *NoAlgo* im Mittel gelang bzw. sich keine signifikante Abweichung der Ergebnisse von Null ergab (*5dB*: $p = 0{,}692$; *mean*: $p = 0{,}351$; *SRT*: $p = 0{,}862$).

Bei der Überprüfung des Effektes der Störgeräuschreduktion konnte jedoch nur für die Kondition *Shadow-Filtering* ein signifikanter Unterschied im Vergleich zu *NoAlgo* nachgewiesen werden ($p = 0{,}012$). Für *Real8dB* zeigte sich keine signifikante Verbesserung im Vergleich zu *NoAlgo* (*5dB*: $p = 0{,}053$; *mean*: $p = 0{,}098$; *SRT*: $p = 0{,}526$). Zusätzlich wurde auch für die reduzierte Probandengruppe auf Verteilungsunterschiede bei der Testsituation *Real8dB* mit einem *t*-Test geprüft, es wurde allerdings für keine der verglichenen Testkonfigurationen das Signifikanzniveau von 5% erreicht (*5dB* - *mean*: $p = 0{,}199$; *5dB* - *SRT*: $p = 0{,}074$; *mean* - *SRT*: $p = 0{,}484$).

Selbst bei der reduzierten Schwerhörigen-Probandengruppe sind die Streubreiten größer als bei der Normalhörenden-Probandengruppe (vgl. Abbildung A.9 und Abbildung A.7). Werden zum Beispiel die Interquartilsspannweiten für *5dB* in der Kondition *NoAlgo* betrachtet, so liegt der Wert für die Normalhörenden bei 1,5 dB, für die Schwerhörigen hingegen bei 5,5 dB. Ein *F*-

Test bestätigt die Beobachtungen der erhöhten Variabilität der Daten durch die Schwerhörigkeit der Probanden und zeigt eine signifikant unterschiedliche Verteilung bei dem Vergleich der Normalhörenden mit den Schwerhörigen für *NoAlgo* bei *5dB* ($p = 0{,}025$) sowie für *Real8dB* bei *SRT* ($p = 0{,}002$). Alle anderen Vergleiche zwischen den Probandengruppen waren nicht signifikant unterschiedlich (*5dB*, *Real8dB*: $p = 0{,}866$; *mean*, *NoAlgo*: $p = 0{,}062$; *mean*, *Real8dB*: $p = 0{,}904$; *SRT*, *NoAlgo*: $p = 0{,}100$; *SRT*, *Shadow-Filtering*: $p = 0{,}464$).

Zuletzt wurde noch mittels t-Test berechnet, ob es einen signifikanten Unterschied zwischen den Probandengruppen in der Ausprägung des Effektes gab. Da sich nur für das *Shadow-Filtering* ein messbarer Effekt bei beiden Probandengruppen herausgestellt hat, wurde dieser Test auch nur für das *Shadow-Filtering* durchgeführt. Dabei ergab sich kein signifikanter Unterschied zwischen den Probandengruppen ($p = 0{,}237$).

A.4 Diskussion

Die Ergebnisse des Einregelungsverfahrens zeigen sowohl durch ihre Werte als auch durch ihre Varianz, dass die normalhörenden Probanden in der Kondition *NoAlgo* zum Großteil in der Lage waren, das Testsignal dem Referenzsignal anzupassen und damit den ΔSNR reproduzierbar einzustellen. Lediglich beim *SRT* ergab sich eine statistisch signifikante Abweichung, die jedoch im Mittel kleiner als eine Schrittweite (1dB) ist und damit keine praktische Bedeutsamkeit besitzt. Bei Verwendung eines Störgeräuschreduktionsalgorithmus ist nur in der Kondition *Shadow-Filtering* eine signifikante Verbesserung im Vergleich zur Kondition *NoAlgo* beobachtbar. Die Kondition *Real8dB* weist zwar den erwarteten Trend, d.h. eine Erhöhung des ΔSNR von *SRT* zu *5dB*, auf, jedoch steigt ebenfalls die Streuung der Ergebnisse an. Damit verringert sich die statistische Aussagekraft. Die Vergrößerung der Streubreiten weist auf eine zunehmende Unsicherheit der Probanden bei der Einregelung hin, da ein eindeutiges und reproduzierbares Kriterium bei der Angleichung von zwei gut verständlichen Signalen bei einem SNR von *5dB* zu fehlen scheint. Dieses Kriterium liegt beim *SRT* vor, bei dem jedoch die Wirkung der Störgeräuschreduktion minimal ist.

Die sechzehn schwerhörigen Probanden weisen deutlich höhere Streubreiten auf als die Normalhörenden. Selbst in der Kondition *NoAlgo* konnten nicht alle schwerhörigen Probanden das Testsignal dem Referenzsignal zuverlässig angleichen. Diese Schwierigkeiten sind vermutlich einerseits auf die Komplexität der Aufgabe als auch andererseits auf die ungewohnte Bedienung des Touchscreens zurück zu führen. Obwohl alle Probanden angaben, die Aufgabe verstanden zu haben, tendierten manche Probanden aufgrund ihrer Vorerfahrungen aus anderen Untersuchungen dazu, das Referenzsignal auf die Sprachverständlichkeitsschwelle oder eine gute Sprachverständlichkeit und nicht auf die dem Referenzsignal entsprechende Sprachverständlichkeit einzuregeln. Deshalb wurden vier schwerhörige Probanden von der weiteren Analyse ausgeschlossen.

Die verbleibenden zwölf Probanden in der Gruppe der Schwerhörigen zeigten immer noch eine größere Streubreite, jedoch im Trend die gleichen Ergebnisse wie die Gruppe der Normalhörenden: Die Störgeräuschreduktion führt in der Kondition *Real8dB* zu einer zunehmenden Verbesserung des ΔSNR von *SRT* zu *5dB*. Jedoch konnte auch bei den schwerhörigen Probanden

nur ein signifikanter Unterschied zu *NoAlgo* in der Kondition *Shadow-Filtering* nachgewiesen werden. Der Effekt ist im Median größer als bei der Gruppe der Normalhörenden, jedoch nicht statistisch signifikant unterschiedlich. Dadurch lässt sich nur eine Tendenz erkennen, dass die schwerhörigen Probanden möglicherweise durch unerwünschte Effekte der Störgeräuschreduktion (Artefakte, teilweise Unterdrückung des gewünschten Sprachsignals) weniger beeinträchtigt werden als die normalhörenden Probanden und mehr von der Störgeräuschreduktion profitieren.

Mit diesen Ergebnissen wird die Untersuchung von Schlüter et al. (2010) bestätigt. Sie beschrieben bereits, dass nur eine geringe Verbesserung gemessen werden konnte und die Einregelung zu einer großen Streuung bei positivem Referenz-SNR führte. Allerdings wurden die Messungen bei festen SNR-Werten durchgeführt, die in der vorgestellten Studie durch individuelle Sprachverständlichkeitsschwellen ersetzt wurden. Diese Erweiterung der Methode lässt deshalb gezieltere Rückschlüsse auf die Anwendung der Einregelung in Abhängigkeit von der dargebotenen Sprachverständlichkeit zu. Im Gegensatz zu den gezeigten Ergebnissen wiesen Wittkop et al. (1997) den Nutzen von Störgeräuschreduktionsalgorithmen auch ohne *Shadow-Filtering* nach. Dies war ihnen möglich, da die untersuchten Algorithmen nicht einkanalig agierten und Modelle binauraler Interaktion nutzen. Diese Algorithmen erreichen im Gegensatz zu einkanaligen Störgeräuschreduktionen auch bei einem negativen SNR am Eingang eine Verbesserung. So war es auch möglich die Signale bei SNR-Werten anzubieten, die für die normalhörenden Probanden nahe ihrer Sprachverständlichkeitsschwelle lagen. Deshalb konnten die Probanden während der Einregelungen ein eindeutiges und reproduzierbares Kriterium verwenden, um eine Verbesserung des SNR durch die Störgeräuschreduktion zu beurteilen.

Insgesamt bleibt als Ergebnis der vorgestellten Studie festzuhalten, dass der Störgeräuschreduktionsalgorithmus zwar bei *5dB* den größten Effekt zeigt, dass das Einregelungsverfahren jedoch bei guter Verständlichkeit der Sprachsignale eine zu große Varianz aufweist, um Effekte von wenigen dB mit einer Anzahl von zwölf Probanden auflösen zu können. Dies gelingt nur mit dem Verfahren des *Shadow-Filtering*, das jedoch bei realen Hörgeräten nicht angewendet werden kann, da Sprache und Störgeräusch nicht separat verarbeitet und bei beliebigem SNR gemischt werden können.

A.5 Ausblick

Die Evaluation des hier vorgestellten Testverfahrens zur Bewertung einkanaliger Störgeräuschreduktion konnte in den Grundhypothesen für beide Probandengruppen bestätigt werden. Die Kondition *Shadow-Filtering* scheint für die Laborsituation eine gute Möglichkeit zu bieten, mit einem Einregelungsverfahren die Verbesserung des SNR durch einen Algorithmus zur Störgeräuschreduktion nachzuweisen. Jedoch ist a-priori-Wissen über die einzelnen Signalanteile notwendig, so dass dieses Verfahren keine Lösung für die Bewertung von Störgeräuschreduktionen in realen Hörgeräten darstellt. Deshalb besteht weiterhin die Notwendigkeit, ein adäquates Verfahren zur Evaluation einkanaliger Störgeräuschreduktionsalgorithmen realer Hörgeräte zu entwickeln, das kein a-priori-Wissen über die Signalanteile voraussetzt und bei positivem SNR mit hoher Validität und Reliabilität anwendbar ist.

Danksagung

Wir danken den Probanden für die Teilnahme an der Untersuchung und Martin Seidel für einen Teil der Datenaufnahme.

Appendix B

Comparison of accuracy between the speech-in-noise test using time-compressed speech and original speech material

For normal-hearing participants, the original OLSA shows a mean SRT of -7.1 dB SNR and a corresponding slope of 17.1 %/dB (Wagener et al., 1999a). These results are based on recognition scores measured with word scoring. Wagener et al. (1999a) state that the desired accuracy (standard deviation of the SRT) after training is about 0.5 dB SNR. This results in an 8.6% variation of recognition scores at the threshold.

The same model was applied to compare the accuracy of the speech-in-noise test with time-compressed speech to a test using original speech. TCT_N values obtained by young normal hearing participants measured at 1 dB SNR (see Chapter 4.3.2) were applied and the accuracy after training was estimated. Thereafter, the standard deviation of the TCT_N of all results measured in lists 3 to 6 was calculated and yielded a value of 0.5. The mean standard deviation of the repeated measurements for each listener is 0.3. Then the slope was estimated. This estimate was based on the recognition scores of young normal-hearing participants determined for sentences presented at 1 dB SNR and compressed to $\rho = 25$, 30 and 35% using Praat (see Chapter 2). The time compression presented was transformed using Equation 4.1. Then, a discrimination function was fitted, using a maximum likelihood method, to recognition scores and time-compression factors. This method was described in Chapter 4.2.2.2. The recognition scores, the fitted discrimination function and the estimated TCT and slope are shown in Figure B.1. The slope of the discrimination function was 15.9 %/N. Based on the standard deviation of all results, the test showed a variance in recognition scores of about 8% at the threshold. This variance is similar to that for original speech.

Although the observed results for time-compressed speech were measured at 1 dB SNR with young normal-hearing participants and speech was processed with Praat, this analysis can only serve as an estimate. The calculated slope of time-compressed speech relies on data determined with word scoring, while the accuracy (standard deviation of the repeated measurements after training) was determined with sentence scoring. This resulted in differences of the estimated

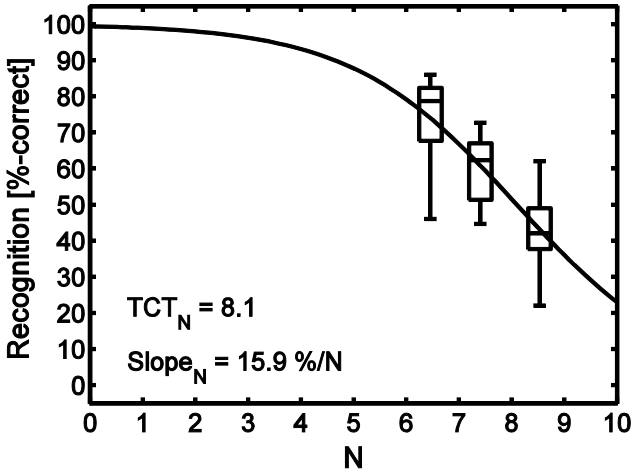

Figure B.1: Boxplots of recognition scores of time-compressed speech processed with Praat as obtained with twelve young normal-hearing participants at 1 dB SNR (for more details see Chapter 2). In addition, the discrimination function estimated with a maximum likelihood fit is displayed together with the corresponding TCT_N and slope.

TCT. While in Chapter 4 the mean TCT_N after training (i.e. using lists 3 to 6) was about 6.7, the TCT_N for recognition scores determined with word scoring (see Chapter 2) was about 8.1. It was expected that sentence scoring results in thresholds that correspond to less time compression (smaller speech rate) than word scoring (see Chapter 2). Furthermore, the application of sentence or word scoring in the different test procedures presenting time-compressed speech and original speech material makes the comparison difficult, since results are based on a different number of presented items (Brand and Wagener, 2005).

In addition, it is possible that the slope of the discrimination function for older or hearing-impaired participants is smaller as compared to young normal-hearing participants. This assumption is based on the comparison of slopes measured with normal-hearing and hearing-impaired participants using original speech. Wagener and Brand (2005) showed smaller slopes for hearing-impaired than for normal-hearing participants. The smaller slope could result in larger variations.

For a more appropriate comparison between the accuracy of a speech-in-noise test using original and time-compressed speech, it is necessary to measure discrimination functions described by threshold and slope using sentence scoring. Furthermore, normal-hearing as well as hearing-impaired participants of different ages should be studied for these measurements.

Bibliography

Aaronson, D., Markowitz, N., and Shapiro, H. (**1971**). "Perception and immediate recall of normal and 'compressed' auditory sequences," Atten. Percept. Psychophys., **9**, 338–344.

Adams, E. M., Gordon-Hickey, S., Morlas, H., and Moore, R. (**2012**). "Effect of rate-alteration on speech perception in noise in older adults with normal hearing and hearing impairment," Am. J. Audiol., **21**, 22–32.

Adank, P., and Devlin, J. T. (**2010**). "On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech," NeuroImage, **49**, 1124–1132.

Adank, P., and Janse, E. (**2009**). "Perceptual learning of time-compressed and natural fast speech," J. Acoust. Soc. Am., **126**, 2649–2659.

Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (**2009**). "Reverse hierarchies and sensory learning," Philos. Trans. R. Soc. B Biol. Sci., **364**, 285–299.

Anderson, M. C., Arehart, K. H., and Kates, J. M. (**2009**). "The acoustic and peceptual effects of series and parallel processing," EURASIP J. Adv. Signal Process., **2009**, Article ID 619805.

Banai, K., and Lavner, Y. (**2012**). "Perceptual learning of time-compressed speech: More than rapid adaptation," PLoS ONE, **7**, e47099.

Bentler, R., Wu, Y.-H., Kettel, J., and Hurtig, R. (**2008**). "Digital noise reduction: Outcomes from laboratory and field studies," Int. J. Audiol., **47**, 447–460.

Boersma, P., and Weenink, D. (**2009**). *Praat: doing phonetics by computer (Version 5.1.05) [Computer program]*, www. praat.org (Last viewed: 08/02/2011).

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am., **111**, 2801–2810.

Brand, T., and Wagener, K. (**2005**). "Wie lässt sich die maximale Verständlichkeit optimal bestimmen?," in 8. Jahrestagung der Deutschen Gesellschaft für Audiologie, Göttingen, Germany.

Brons, I., Houben, R., and Dreschler, W. A. (**2013**). "Perceptual effects of noise reduction with respect to personal preference, speech intelligibility, and listening effort," Ear Hear., **34**, 29–41.

Brons, I., Houben, R., and Dreschler, W. A. (**2014**). "Effects of noise reduction on speech intelligibility, perceived listening effort, and personal preference in hearing-impaired listeners," Trends Hear., **18**, 2331216514553924.

Bundesministerium der Justiz (**2012**). *Richtline des Gemeinsamen Bundesausschusses über die Verordnung von Hilfsmitteln in der vertragsärztlichen Versorgung (Hilfsmittel-Richtlinie/HilfsM-RL),* Bundesanzeiger, www.g-ba.de/downloads/39-261-1461/2012-03-15_HilfsM-RL_Neufassung-Hoerhilfen_BAnz.pdf (Last viewed: 22/10/2014).

Carroll, R., and Ruigendijk, E. (**2013**). "The Effects of Syntactic Complexity on Processing Sentences in Noise," J. Psycholinguist. Res., **42**, 139–159.

Chung, K. (**2004**). "Challenges and recent developments in hearing aids - Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms," Trends Amplif., **8**, 83–124.

Chung, K. (**2007**). "Effective compression and noise reduction configurations for hearing protectors," J. Acoust. Soc. Am., **121**, 1090–1101.

Chu, W. C., and Lashkari, K. (**2003**). "Energy-based nonuniform time-scale compression of audio signals," IEEE Trans. Consum. Electron., **49**, 183–187.

Cohen, I., and Berdugo, B. (**2002**). "Noise estimation by minima controlled recursive averaging for robust speech enhancement," IEEE Signal Process. Lett., **9**, 12–15.

Covell, M., Withgott, M., and Slaney, M. (**1998**). "MACH1: nonuniform time-scale modification of speech," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Seattle, WA, Vol. 1, pp. 349–352.

Demol, M., Verhelst, W., Struyve, K., and Verhoeve, P. (**2005**). "Efficient non-uniform time-scaling of speech with WSOLA," in Proceedings of the 10th International Conference on Speech and Computer (SPECOM), Patras, Greece, pp. 163–166.

Dillon, H. (**2012**). *Hearing Aids*, Boomerang Press, Turramurra, Australia, pp. 1-6.

Dorran, D. (**2005**). *Audio Time-Scale Modification,* Ph.D. thesis, School of Control Systems and Electrical Engineering, Dublin Institute of Technology, Ireland.

Dupoux, E., and Green, K. (**1997**). "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," J. Exp. Psychol. Hum. Percept. Perform., **23**, 914–927.

Fredelake, S., Holube, I., Schlueter, A., and Hansen, M. (**2012**). "Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms," Int. J. Audiol., **51**, 299–308.

George, E. L. J., Goverts, S. T., Festen, J. M., and Houtgast, T. (**2010**). "Measuring the effects of reverberation and noise on sentence intelligibility for hearing-impaired listeners," J. Speech Lang. Hear. Res., **53**, 1429–1439.

Glanzer, M., and Cunitz, A. R. (**1966**). "Two storage mechanisms in free recall," J. Verbal Learn. Verbal Behav., **5**, 351–360.

Golomb, J. D., Peelle, J. E., and Wingfield, A. (**2007**). "Effects of stimulus variability and adult aging on adaptation to time-compressed speech," J. Acoust. Soc. Am., **121**, 1701–1708.

Gordon-Salant, S., and Fitzgibbons, P. J. (**2001**). "Sources of age-related recognition difficulty for time-compressed speech," J Speech Lang Hear Res, **44**, 709–719.

Gordon-Salant, S., and Fitzgibbons, P. J. (**2004**). "Effects of stimulus and noise rate variability on speech perception by younger and older adults," J. Acoust. Soc. Am., **115**, 1808–1817.

Gordon-Salant, S., and Friedman, S. A. (**2011**). "Recognition of rapid speech by blind and sighted older adults," J. Speech Lang. Hear. Res., **54**, 622–631.

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

Hagerman, B. (**1984**). "Clinical measurements of speech reception threshold in noise," Scand. Audiol., **13**, 57–63.

Hagerman, B., and Kinnefors, C. (**1995**). "Efficient adaptive methods for measuring speech reception threshold in quiet and in noise," Scand. Audiol., **24**, 71–77.

Härting, C., Markowitsch, H. J., Neufeld, H., Calabrese, P., Deisinger, K., and Kessler, J. (**2000**). "Wechsler Gedächtnis Test–Revidierte Fassung (WMS-R)," Verlag Hans Huber Bern, Switzerland, p. 52.

He, L., and Gupta, A. (**2001**). "Exploring benefits of non-linear time compression," in Proceedings of the Ninth ACM International Conference on Multimedia, Ottawa, Canada, pp. 382–391.

Hernvig, L. H., and Olsen, S. Ø. (**2005**). "Learning effect when using the Danish Hagerman sentences (Dantale II) to determine speech reception threshold," Int. J. Audiol., **44**, 509–512.

Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., and Kollmeier, B. (**2012**). "A Spanish matrix sentence test for assessing speech reception thresholds in noise," Int. J. Audiol., **51**, 536–544.

Hoetink, A. E., Körössy, L., and Dreschler, W. A. (**2009**). "Classification of steady state gain reduction produced by amplitude modulation based noise reduction in digital hearing aids," Int. J. Audiol., **48**, 444–455.

Holube, I., Blab, S., Fürsen, K., Gürtler, S., Meisenbacher, K., Nguyen, D., and Taesler, S. (**2009**). "Einfluss eines Maskierers und der Testmethode auf die Sprachverständlichkeitsschwelle von jüngeren und älteren Normalhörenden - Influence of masker and measurement method on speech reception threshold of young and elderly normal hearing listeners," Z. Audiol., **48**, 120–127.

Holube, I., Puder, H., and Velde, T. M. (**2014a**). "Chapter 7: DSP Hearing Instruments," In M. J. Metz (Ed.), Sandlin's Textbook of Hearing Aid Amplification, Plural Publishing, Inc., San Diego, CA, USA, 3rd edition, pp. 221–293.

Holube, I., Schepker, H., Haeder, K., and Rennies, J. (**2014b**). "Listening effort and speech intelligibility in reverberation and noise," presented at the International Hearing Aid Research Conference (IHCON 2014), Tahoe City, California, USA.

Höpfner, D. (**2006**). "Echtzeitfähiger Algorithmus zur stufenlosen Geschwindigkeitserhöhung gespeicherter natürlicher Sprache," in Elektronische Sprachsignalverarbeitung, Tagungsband 17. Konferenz Freiberg, TUDpress Verlag der Wissenschaft, pp. 92–99.

Höpfner, D. (**2007**). "Untersuchungen zeitskalierter Sprachwiedergabe mit normal sehenden, sehbehinderten und blinden Probanden," in Elektronische Sprachverarbeitung, Cottbus, pp. 235–242.

Höpfner, D. (**2008**). "Nichtlinearer Zeitskalierungsalgorithmus für gespreicherte natürliche Sprache," in 8. ITG-Fachtagung, Sprachkommunikation 2008, Aachen, Germany.

Houtgast, T., and 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**, 1069–1077.

Humes, L. E., Burk, M. H., Coughlin, M. P., Busey, T. A., and Strauser, L. E. (**2007**). "Auditory speech recognition and visual text recognition in younger and older adults: Similarities and differences between modalities and the effects of presentation rate," J. Speech Lang. Hear. Res., **50**, 283–303.

Hu, Y., and Loizou, P. C. (**2007**). "A comparative intelligibility study of single-microphone noise reduction algorithms," J. Acoust. Soc. Am., **122**, 1777-1786.

IEC 60645-2 (**2010**). "Audiometric Equipment—Part 2: Equipment for Speech Audiometry" (International Electrotechnical Commission, IEC, Geneva, Switzerland), No. IEC 60645-2 29/714/CD.

ISO 389-8 (**2004**). "Acoustics—Reference zero for the calibration of audiometric equipment—Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones" (International Organization for Standardization, Geneva, Switzerland).

Janse, E. (**2003**). *Production and Perception of Fast Speech,* Ph.D. thesis, University of Utrecht, Netherlands.

Janse, E. (**2009**). "Processing of fast speech by elderly listeners," J. Acoust. Soc. Am., **125**, 2361–2373.

Jørgensen, S., and Dau, T. (**2011**). "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing," J. Acoust. Soc. Am., **130**, 1475–1487.

Kallinger, M., Ochsenfeld, H., and Schlüter, A. (**2009**). "A novel listening test-based measure of intelligibility enhancement," in Audio Engineering Society Convention 127, Audio Engineering Society, New York, NY, USA, Paper 7822

Kapilow, D., Stylianou, Y., and Schroeter, J. (**1999**). "Detection of non-stationarity in speech signals and its application to time-scaling," in Proceedings of Eurospeech, Budapest Hungary.

Kent, R. D., and Read, C. (**2002**). "The acoustic characteristics of consonants," In R. D. Kent (Ed.), *The Acoustic Analysis of Speech*, Singular Publishing Group, Delmar, New York, 2nd ed., pp. 139–188.

Kollmeier, B. (**1990**). *Messmethodik, Modellierung und Verbesserung der Verständlichkeit von Sprache,* Habilitation treatise, Fachbereich Physik, University of Göttingen, Germany.

Kollmeier, B., Gilkey, R. H., and Sieben, U. K. (**1988**). "Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model," J. Acoust. Soc. Am., **83**, 1852–1862.

Kollmeier, B., and Wesselkamp, M. (**1997**). "Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment," J. Acoust. Soc. Am., **102**, 2412–2421.

Kondo, K. (**2012**). "Chapter 2: Speech Quality," in *Subjective Quality Measures of Speech*, Springer, Heidelberg, New York, Dordrecht, London, pp. 7–20.

Lamel, L. F., Kassel, R. H., and Blab, S. (**1989**). "Speech database development: Design and analysis of the acoustic-phonetic corpus," in Proceedings of Speech Input/Output Assessment and Databases - 1989, Vol. 2, pp. 161–170.

Leek, M. R. (**2001**). "Adaptive procedures in psychophysical research," Percept. Psychophys., **63**, 1279–1292.

Lee, S., Kim, H. D., Kim, H. S., and Kim, H. (**1997**). "Variable time-scale modification of speech using transient information," in Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASP), Vol. 2, pp. 1319–1322.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am., **49**, 467–477.

Liu, S., and Zeng, F.-G. (**2006**). "Temporal properties in clear speech perception," J. Acoust. Soc. Am., **120**, 424–432.

Loizou, P. C., and Kim, G. (**2011**). "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions," IEEE Trans. Audio Speech Lang. Process., **19**, 47–56.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., and Spriet, A. (2010). "Multicenter evaluation of signal enhancement algorithms for hearing aids," J. Acoust. Soc. Am., 127, 1491–1505.

Marzinzik, M. (**2000**). *Noise Reduction Schemes for Digital Hearing Aids and their Use for the Hearing Impaired,* Ph.D. thesis, Carl von Ossietzky Universität, Oldenburg, Germany.

Marzinzik, M., and Kollmeier, B. (**2002**). "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics," Speech Audio Process. IEEE Trans. On, **10**, 109–118.

Meister, H., Schreitmüller, S., Grugel, L., Landwehr, M., von Wedel, H., Walger, M., and Meister, I. (**2011**). "Untersuchungen zum Sprachverstehen und zu kognitiven Fähigkeiten im Alter," HNO, **59**, 689-695.

Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun., **9**, 453–467.

Murdock Jr., B. B. (**1962**). "The serial position effect of free recall," J. Exp. Psychol., **64**, 482–488.

Nabelek, A. K. (**2005**). "Acceptance of background noise may be key to successful fittings," Hear. J., **58**, 10–15.

Nabelek, A. K., Tucker, F. M., and Letowski, T. R. (**1991**). "Toleration of background noises: relationship with patterns of hearing aid use by elderly persons," J Speech Hear Res, **34**, 679–685.

Nahum, M., Nelken, I., and Ahissar, M. (**2008**). "Low-level information and high-level perception: The case of speech in noise," PLoS Biol, **6**, e126.

Nahum, M., Nelken, I., and Ahissar, M. (**2010**). "Stimulus uncertainty and perceptual learning: Similar principles govern auditory and visual learning," Vision Res., **50**, 391–401.

Naylor, G. (**2010**). "Limitations of Speech Reception Threshold (SRT) as an outcome measure in hearing-aid research," presented at the International Hearing Aid Research Conference (IHCON 2010), Tahoe City, California, USA.

Neher, T., Grimm, G., and Hohmann, V. (**2014a**). "Perceptual consequences of different signal changes due to binaural noise reduction: Do hearing loss and working memory capacity play a role?," Ear Hear., **35**, e213–e227.

Neher, T., Grimm, G., Hohmann, V., and Kollmeier, B. (**2014b**). "Do hearing loss and cognitive function modulate benefit from different binaural noise-reduction settings?," Ear Hear., **35**, e52–e62.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am., **95**, 1085–1099.

Nordrum, S., Erler, S., Garstecki, D., and Dhar, S. (**2006**). "Comparison of performance on the hearing in noise test using directional microphones and digital noise reduction algorithms," Am. J. Audiol., **15**, 81–91.

Olsen, S. Ø., and Brännström, K. J. (**2014**). "Does the acceptable noise level (ANL) predict hearing-aid use?," Int. J. Audiol., **53**, 2–20.

Olsen, W. O. (**1998**). "Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report," Am. J. Audiol., **7**, 21–25.

Ozimek, E., Warzybok, A., and Kutzner, D. (**2010**). "Polish sentence matrix test for speech intelligibility measurement in noise," Int. J. Audiol., **49**, 444–454.

Pallier, C., Sebastián-Gallés, N., Dupoux, E., Christophe, A., and Mehler, J. (**1998**). "Perceptual adjustment to time-compressed speech: a cross-linguistic study," Mem. Cognit., **26**, 844–851.

Peelle, J. E., and Wingfield, A. (**2005**). "Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech," J. Exp. Psychol. Hum. Percept. Perform., **31**, 1315–1330.

Peissig, J. (**1993**). *Binaurale Hörgerätestrategien in komplexen Störschallsituationen,* Ph.D. thesis, Georg-August-Universität, Göttingen, Germany.

Pichora-Fuller, M. K., and Singh, G. (**2006**). "Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation," Trends Amplif., **10**, 29–59.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Int. J. Audiol., **18**, 43–52.

Rennies, J., Schepker, H., Holube, I., and Kollmeier, B. (**2014**). "Listening effort and speech intelligibility in listening situations affected by noise and reverberation," J. Acoust. Soc. Am., **136**, 2642–2653.

Rønne, F. M., Laugesen, S., Jensen, N. S., Hietkamp, R. K., and Pedersen, J. H. (**2013**). "Magnitude of speech-reception-threshold manipulators for a spatial speech-in-speech test that takes signal-to-noise ratio confounds and ecological validity into account," J. Acoust. Soc. Am., **133**, 3379–3379.

Schlueter, A., Brand, T., Lemke, U., Nitzschner, S., Kollmeier, B., and Holube, I. (**2014a**). "Speech perception at positive signal-to-noise ratios using adaptive adjustment of time compression," J. Acoust. Soc. Am., submitted.

Schlueter, A., Lemke, U., Kollmeier, B., and Holube, I. (**2014b**). "Intelligibility of time-compressed speech: The effect of uniform versus non-uniform time-compression algorithms," J. Acoust. Soc. Am., **135**, 1541–1555.

Schlueter, A., Lemke, U., Kollmeier, B., and Holube, I. (**2014c**). "Normal and time-compressed speech: How does learning affect speech recognition thresholds in noise?," Int. J. Audiol., submitted.

Schlüter, A. (**2007**). *Perzeptive Beurteilung von Sprache im Störgeräusch,* Master's thesis, Carl von Ossietzky Universität, Oldenburg, Germany.

Schlüter, A., Aderhold, J., Koifman, S., Krüger, M., Nüsse, T., Lemke, U., and Holube, I. (**2014a**). "Evaluation eines Einregelungsverfahrens zur Bestimmung des Nutzens einkanaliger Algorithmen zur Störgeräuschreduktion - Evaluation of an adjustment method to determine the benefit of single-microphone noise reduction algorithms," Z. Audiol., **53**, 50–58.

Schlüter, A., and Holube, I. (**2010**). "Perzeptive Maße zur Evaluation von Hörgeräteversorgungen bei Sprache im Störgeräusch - Perceptive measure to evaluate hearing aid fittings for speech in noise," Z. Audiol., **49**, 103–111.

Schlüter, A., Holube, I., and Lemke, U. (**2010**). "Untersuchung eines subjektiven SNR-Vergleichs zur Bestimmung des Nutzens einkanaliger Störgeräuschreduktionen," in 13. Jahrestagung der Deutschen Gesellschaft für Audiologie, Frankfurt, Germany.

Schlüter, A., Holube, I., and Lemke, U. (**2013**). "Verfahren zur Bestimmung der Zeitkompressionsschwelle von Sprache im Störgeräusch," in 16. Jahrestagung der Deutschen Gesellschaft für Audiologie, Rostock, Germany.

Schlüter, A., Holube, I., Lemke, U., and Herzog, D. (**2014b**). "Verständlichkeitsschwellen im Göttinger und Oldenburger Satztest bei Variation der Sprachgeschwindigkeit," in 17. Jahrestagung der Deutschen Gesellschaft für Audiologie, Oldenburg, Germany.

Schneider, B. A., Daneman, M., and Murphy, D. R. (**2005**). "Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing?," Psychol. Aging, **20**, 261–271.

Sebastián-Gallés, N., Dupoux, E., Costa, A., and Mehler, J. (**2000**). "Adaptation to time-compressed speech: phonological determinants," Percept. Psychophys., **62**, 834–842.

Shibuya, T., Kobayashi, Y., Watanabe, H., and Kondo, K. (**2012**). "Differences in the effect of time-expanded and time-contracted speech on intelligibility by phonetic feature," in Proceedings of the IEEE I International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, pp. 4489 –4492.

Simonsen, C. S., Kijne, J.-C. B., Hansen, L. B., Petersen, A. S., Laugesen, S., Rønne, F. M., and Jensen, N. S. (**2014**). "The Spatial Fixed-SNR (SFS) test: presentation and validation," presented at the International Hearing Aid Research Conference (IHCON 2014), Tahoe City, California, USA.

Smeds, K., Wolters, F., and Rung, M. (**2012**). "Estimation of realistic signal-to-noise ratios," presented at the International Hearing Aid Research Conference (IHCON 2012), Tahoe City, California, USA.

Smeds, K., Wolters, F., and Rung, M. (**2015**). "Estimation of signal-to-noise ratios in realistic sound scenarios," J. Am. Acad. Audiol., **26**, 183–196.

Smits, C., and Houtgast, T. (**2006**). "Measurements and calculations on the simple up-down adaptive procedure for speech-in-noise tests," J. Acoust. Soc. Am., **120**, 1608–1621.

Smits, C., Kapteyn, T. S., and Houtgast, T. (**2004**). "Development and validation of an automatic speech-in-noise screening test by telephone," Int. J. Audiol., **43**, 15–28.

Sotscheck, J. (**1982**). "Ein Reimtest für Verständlichkeitsmessungen mit deutscher Sprache als ein verbessertes Verfahren zur Bestimmung der Sprachübertragungsgüte," Fernmelde-Ingenieur, **36**, 1–84.

Speech Research Lab, A.I. DuPont Hospital for Children and the University of Delaware (**2012**). *WEDW,*. www.asel.udel.edu/speech/Spch_proc/software.htm (Last viewed: 08/10/2012)

Sukowski, H., Brand, T., Wagener, K. C., and Kollmeier, B. (**2008**). "Der Einfluss des Präsentations- und Antwortformates auf die Messung der Sprachverständlichkeit mit Einsilbertestverfahren," in 11. Jahrestagung der Deutschen Gesellschaft für Audiologie e.V. , Kiel, Germany.

Thomas, M. R. P., Gudnason, J., and Naylor, P. A. (**2008**). "Application of the DYPSA algorithm to segmented time-scale modification of speech," in Proceedings of the European Signal Processing Conference (EUSIPCO), Lausanne, Switzerland.

Tombaugh, T. N. (**2004**). "Trail making test A and B: normative data stratified by age and education," Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol., **19**, 203–214.

Tucker, S., and Whittaker, S. (**2006**). "Time is of the essence: an evaluation of temporal compression algorithms," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Québec, Canada, pp. 329–338.

Tun, P. A. (**1998**). "Fast noisy speech: age differences in processing rapid speech with background noise," Psychol. Aging, **13**, 424–434.

Uslar, V. N. (**2014**). *Speech perception, age, and hearing loss: Methods to assess the balance between bottom-up and top-down processing,* Ph.D. thesis, Carl von Ossietzky Universität, Oldenburg, Germany.

Vary, P., Heute, U., and Hess, W. (**1998**). *Digitale Sprachsignalverarbeitung*, B.G. Teubener, Stuttgart, Germany, pp. 387-379.

Versfeld, N. J., and Dreschler, W. A. (**2002**). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," J. Acoust. Soc. Am., **111**, 401–8.

Wagener, K., Brand, T., and Kollmeier, B. (**1999a**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests," Z. Audiol., **38**, 86–95.

Wagener, K., Brand, T., and Kollmeier, B. (**1999b**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests," Z. Audiol., **38**, 44–56.

Wagener, K. C., and Brand, T. (**2005**). "Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters," Int. J. Audiol., **44**, 144–156.

Wagener, K., Josvassen, J. L., and Ardenkjær, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol., **42**, 10–17.

Wagener, K., Kühnel, V., and Kollmeier, B. (**1999c**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests," Z. Audiol., **38**, 4–15.

Warzybok, A., Rennies, J., Brand, T., Doclo, S., and Kollmeier, B. (**2013**). "Effects of spatial and temporal integration of a single early reflection on speech intelligibility," J. Acoust. Soc. Am., **133**, 269–282.

Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., and Cox, C. L. (**2006**). "Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity," J. Am. Acad. Audiol., **17**, 487–497.

Wittkop, T. (**2001**). *Two-channel noise reduction algorithms motivated by models of binaural interaction*, Ph.D. thesis, Carl von Ossietzky Universität, Oldenburg, Germany.

Wittkop, T., Albani, S., Hohmann, V., Peissig, J., Woods, W. S., and Kollmeier, B. (**1997**). "Speech processing for hearing aids: Noise reduction motivated by models of binaural interaction," Acust. United Acta Acust., **83**, 684–699.

Zokoll, M. A., Wagener, K. C., Brand, T., Buschermöhle, M., and Kollmeier, B. (**2012**). "Internationally comparable screening tests for listening in noise in several European languages: The German digit triplet test as an optimization prototype," Int. J. Audiol., **51**, 697–707.

Acknowledgements

An dieser Stelle möchte ich mich bei allen bedanken, die mich während dieser Arbeit unterstützt haben und sie damit erst ermöglicht haben.

Prof. Dr. Inga Holube danke ich dafür, dass sie mir dieses interessante Promotionsthema übertragen hat und für die Möglichkeit, meine Arbeit in ihrer Arbeitsgruppe am Institut für Hörtechnik und Audiologie der Jade Hochschule Oldenburg durchführen zu können. Ihre Betreuung meiner Arbeit war für mich besonders wichtig und wertvoll. Sie hat mich immer maßgeblich gefördert und unermüdlich unterstützt.

Prof. Dr. Dr. Birger Kollmeier danke ich für die zielgerichtete Betreuung meiner Arbeit an der Carl von Ossietzky Universität. Nur durch die Zusammenarbeit mit ihm und seine Unterstützung war es mir möglich an der Jade Hochschule zu arbeiten und gleichzeitig an der Universität zu promovieren. Seine Anregungen und Ratschläge haben meine Arbeit wesentlich gefördert und ergänzt.

Bei Prof. dr. ir. Wouter Dreschler und Prof. Dr. Tim Jürgens bedanke ich mich für ihre Bereitschaft meine Arbeit zu begutachten und für ihre Spontanität, dies auch in einem verkürzten Zeitraum umzusetzen.

Dr. Ulrike Lemke und Dr. Thomas Brand danke ich für wertvolle und für mich lehrreiche Diskussionen und die produktive wissenschaftliche Zusammenarbeit.

Diana Herzog und Stefan Nitzschner danke ich für die Erfahrungen, die ich sammeln durfte bei der Betreuung ihrer Masterarbeiten. Außerdem danke ich ihnen, dass sie damit meine Arbeit unterstützt haben.

Meinen Dank richte ich auch an Lüder Bentz, Vera Löw, Micha Lundbeck, Maxi Susanne Moritz und an die Studenten der von mir betreuten Projektpaktika Kristina Anton, Nina Blase, Karin Brand, Maximilian Busse, Fehime Cigir, Shiran Koifman, Theresa Nüsse, Patrycia Piktel, Martin Seidel, Johanna Weigel, Mareike Wemheuer und Nathalie Zimmermann. Sie haben zur Datenaufnahme beigetragen.

Ich danke allen Probanden, die an den Untersuchungen teilgenommen haben und damit diese Arbeit unterstützt haben.

Allen Kollegen am Institut für Hörtechnik und Audiologie an der Jade Hochschule und den Mitgliedern der Medizinischen Physik an der Universität danke ich für das freundliche Arbeitsumfeld, sowie ihre technische und administrative Assistenz.

Danken möchte ich auch meinen Freunden, die mich in dieser Zeit begleitet und unterstützt haben. Besonders möchte ich mich aber bei Alex und Katharina für offene Ohren, Unterstützung, geteilte Freude, Ablenkung, Schokolade und so vieles mehr bedanken.

Meinen Eltern danke ich für ihren Beistand in so vielen Bereichen, besonders aber bei dieser spannenden, intellektuellen und aufregenden Arbeit.

Auch mein Mann Jan hat mich auf so vielen Ebenen unterstützt. Besonders sein Verständnis, seine Zuversicht, seine Motivation und seine Kraft haben mir während dieser Arbeit sehr geholfen. Danke!

Bei der Phonak AG und dem Niedersächsischen Ministerium für Kultur und Wissenschaft Niedersächsisches Vorab (Projekt AKOSIA) bedanke ich mich für die Finanzierung meiner Arbeitsstelle.

Eigenständigkeitserklärung

Ich, Anne Schlüter, erkläre hiermit, dass diese Dissertation mit dem Titel “Speech recognition tested at fixed, positive signal-to-noise ratios using time compression: Methods and applications” vollständig mein Eigentum ist. Ich bestätige, dass:

- ich die Dissertation selbständig verfasst habe, und dass die benutzten Hilfsmittel vollständig angegeben sind.
- die Dissertation in Teilen bereits veröffentlicht wurde (siehe Publikationsliste).
- die Dissertation weder in ihrer Gesamtheit noch in Teilen einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorliegt oder vorgelegen hat.
- die Leitlinien guter wissenschaftlicher Praxis an der Carl von Ossietzky Universität Oldenburg befolgt worden sind.
- im Zusammenhang mit dem Promotionsvorhaben keine kommerziellen Vermittlungs- oder Beratungsdienste (Promotionsberatung) in Anspruch genommen worden sind.
- dort wo diese Dissertation auf Arbeiten basiert, die gemeinsam mit anderen durchgeführt wurden, deutlich gemacht wurde, welche Teile von mir und welche von anderen stammen.

Datum: _____

Unterschrift: _____

Curriculum Vitae

Anne Schlüter

*14.08.1978

Education

since 04/2009	Carl von Ossietzky University, Oldenburg, Germany Member of Medical Physics section
08/2005 - 12/2007	Carl von Ossietzky University, Oldenburg, Germany M.Sc. Hearing Technology and Audiology
09/2001 - 08/2005	Jade University of Applied Sciences, Oldenburg, Germany B.Eng. Hearing Technology and Audiology
08/1998 - 08/2001	AQ Hörgeräte, Lüdinghausen, Germany Apprenticeship Hearing Aid Acoustician
1998	St. Antonius Gymnasium, Lüdinghausen, Germany A-Level

Work-Experience

since 01/2009	Jade University of Applied Sciences, Oldenburg, Germany Institute of Hearing Technology and Audiology Research associate
07/2008 - 12/2008	University of Cambridge, UK Department of Experimental Psychology Group: Auditory Perception Internship
01/2008 - 06/2008	Jade University of Applied Sciences, Oldenburg, Germany Institute of Hearing Technology and Audiology Research associate
04/2006 - 07/2006	Hörzentrum Oldenburg, Germany Freelancer
02/2006 - 03/2006	Leuphana University, Lüneburg, Germany Department of Business Psychology Internship
02/2005 - 07/2005	Phonak AG, Stäfa, Switzerland Research and Development Internship for developing diploma thesis
09/2004 - 01/2007	Jade University of Applied Sciences, Oldenburg, Germany Institute of Hearing Technology and Audiology Student assistant

09/2003 - 01/2004	Siemens Audiologische Technik, Erlangen, Germany Internship
09/2002 - 08/2003	Carl von Ossietzky University, Oldenburg, Germany Group: Medical Physics Student assistant

Awards

Publication award of the Deutsche Gesellschaft für Audiologie (2015)	The following paper was awarded: Schlüter, A., Aderhold, J., Koifman, S., Krüger, M., Nüsse, T., Lemke, U., and Holube, I. (2014) "Evaluation eines Einregelungsverfahrens zur Bestimmung des Nutzens einkanaliger Algorithmen zur Störgeräuschreduktion - Evaluation of an adjustment method to determine the benefit of single-microphone noise reduction algorithms", Zeitschrift für Audiologie, 53(2), 50-58.