

Normal and Time-Compressed Speech: How Does Learning Affect Speech Recognition Thresholds in Noise?

Trends in Hearing
2016, Vol. 20: 1–13
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216516669889
tia.sagepub.com


Anne Schlueter¹, Ulrike Lemke², Birger Kollmeier^{3,4},
and Inga Holube^{1,4}

Abstract

Short-term and long-term learning effects were investigated for the German Oldenburg sentence test (OLSA) using original and time-compressed fast speech in noise. Normal-hearing and hearing-impaired participants completed six lists of the OLSA in five sessions. 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. In general, original speech resulted in better speech recognition thresholds than time-compressed speech. Thresholds decreased with repetition for both speech materials. Confirming earlier results, the largest improvements were observed within the first measurements of the first session, indicating a rapid initial adaptation phase. The improvements were larger for time-compressed than for original speech. The novel results on long-term learning effects when using the OLSA indicate a longer phase of ongoing learning, especially for time-compressed speech, which seems to be limited by a floor effect. In addition, for normal-hearing participants, no complete transfer of learning benefits from time-compressed to original speech was observed. These effects should be borne in mind when inviting listeners repeatedly, for example, in research settings.

Keywords

learning effect, speech in noise, matrix test, time-compressed speech

Date received: 18 January 2016; revised: 5 August 2016; accepted: 27 August 2016

Introduction

Perceptual learning can improve performance in listening tasks; this includes the familiarization with the task or the procedure as well as with the presented stimulus (Ortiz & Wright, 2009). The amount and type of practice affects the degree of learning and its generalization (i.e., transfer to unknown conditions; see Banai & Lavner, 2014). As a result, perceptual learning while performing speech tests is of considerable importance in audiology: When hearing abilities or hearing aid fittings are assessed with a speech test, performance also depends on learning of the speech recognition task and of the procedure. Ideally, this learning should be complete prior to data collection, and the amount of training required for a stable level of performance should be achieved quickly, so that the test can be completed in a time-efficient fashion.

Matrix sentence tests are frequently used to study speech recognition in background noise. The German Oldenburg sentence test (OLSA, Wagener, Kühnel, &

Kollmeier, 1999a) as well as its counterparts in different languages (such as the Swedish Hagerman test, Hagerman & Kinnefors, 1995; the Danish Dantale II, Wagener, Josvassen, & Ardenkjær, 2003; the Spanish, or the Polish matrix tests, Hochmuth et al., 2012; Ozimek, Warzybok, & Kutzner, 2010; for a review see Kollmeier et al., 2015) use sentences that were created by random selection from a matrix of 50 words. During the test, the sentences are usually presented at a normal speech rate in a fixed-level background noise, and the

¹Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, Oldenburg, Germany

²Phonak AG, Science & Technology, Cognitive & Ecological Audiology, Stäfa, Switzerland

³Medizinische Physik, Universität Oldenburg, Germany

⁴Cluster of Excellence "Hearing4All", Oldenburg, Germany

Corresponding author:

Anne Schlueter, Jade University of Applied Sciences, Ofener Str. 16, Oldenburg D-26121, Germany.

Email: anne.schlueter@jade-hs.de



level of the speech is adaptively adjusted depending on the response of the participant, who is asked to repeat every word. The goal of the adaptive adjustment procedure is to estimate the speech recognition threshold (SRT), which denotes the signal-to-noise ratio (SNR) for a recognition score of 50%.

Listeners with normal hearing (NH) and those with hearing impairment (HI) frequently obtain SRTs below 0 dB SNR (Wagener & Brand, 2005; Wagener, Brand, & Kollmeier, 1999b). However, conversations in noisy environments in everyday life normally take place at positive SNRs (Olsen, 1998; Smeds, Wolters, & Rung, 2015) and common noise reduction algorithms in hearing aids offer superior effects at similar positive SNRs (Fredelake, Holube, Schlueter, & Hansen, 2012). To measure SRTs at positive SNRs while using matrix tests, the recognition of the speech material has to be degraded. One possibility for degradation is speeding up the speech rate by time compression. The combination of positive SNRs and matrix tests in hearing aid evaluations by combining time-compressed (TC) speech with an adaptively adjusted speech rate is a topic of current research (Schlueter et al., 2015).

There are several advantages of using 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 through their meaning 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. In the following, the learning effects, training procedures, and resulting questions are introduced separately for original speech and TC speech.

Learning Effects for Original Speech Presented in a Matrix Test Format

For matrix tests using original speech, learning effects clearly follow a learning curve, with more pronounced improvements in performance 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 an *intrasession learning effect*. Within a single session, Hagerman and Kinnefors (1995) documented a mean decrease of SRT by 0.13 dB for NH and 0.07 dB to 0.5 dB for HI for each repetition. Wagener et al.

(1999b, 2003), Hernvig and Olsen (2005), and Hochmuth et al. (2012) measured similar or even greater learning effects and recommended the routine use of training lists before measuring speech recognition. For the OLSA, and based on measurements with young NH listeners, Wagener et al. (1999b) suggested one or two training lists or up to 60 sentences, dependent on the required measurement accuracy. After the presentation of two lists with 20 sentences each, the authors showed that the learning effect of about 2 dB (SRT difference observed between the first and last measurement of six consecutive lists) was reduced to a test–retest accuracy of about 0.5 dB (SRT difference between consecutive measurements after two training lists). Although 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, it is relevant with regard to SRTs for speech in noise. This is because a decrease of the SRT by 1 dB corresponds to an improvement in speech recognition of about 15% and, according to the German governmental guidelines for medical aids (Gemeinsamer Bundesausschuss, 2014), SRTs are expected to improve by at least 2 dB through a beneficial hearing aid fitting. The required minimum improvement is equal to the reported accuracy of the OLSA when conducted without training (Wagener et al., 1999b). As a result, when conducting the OLSA without training lists, an improved SRT might be falsely interpreted as a benefit from a hearing aid fitting, while in fact the improvement is the result of learning.

In addition to intrasession effects, learning effects can also be observed when comparing data collected on different days (i.e., *intersession 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) also measured an intersession 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 groups, respectively.

The learning effects 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 listeners with SRT values below 0 dB SNR. However, they stated that a learning effect should be considered for participants with SRT values above 0 dB SNR. In other words, learning effects are larger for participants with more severe HIs. In contrast, Wagener and Brand (2005) reported learning effects for both NH and HI groups, although the effect was smaller for the HI group. It remains unclear whether two

possible effects need to be considered: (a) Differential learning effects due to HI or age and (b) the use of different training protocols depending on previous learning or on hearing loss.

Based on the reported intrasession learning effects, for matrix tests, it is generally recommended that training lists are administered before measuring speech recognition, independent of the type of listener. This recommendation is especially important for clinical purposes, because daily practice generally provides one session with a small number of repeated measurements, such as when comparing hearing aids. Until now, intrasession learning was not examined for repeated measures within different consecutive sessions on different days. Such a procedure is commonly used in scientific studies using complex test protocols, for example, when different parameter settings of hearing aids are tested in acoustically different environments.

Little is known about how learning affects results in research studies when using several sessions of repeated SRT measures. One possibility is that a participant who completed training in an earlier session displays a different pattern of intrasession learning in following sessions. In addition, intersession differences would suggest an ongoing learning of the speech material and the test procedure. Moreover, the recommended use of particular training lists is based on studies of young NH rather than older HI listeners. However, the majority of hearing aid users are older, and older HI listeners often take part in scientific research during which they may exhibit a different pattern of performance from young HI and NH listeners. In addition, importantly, the studies cited conducted measurements in a limited number of sessions. Therefore, they provide no indication whether or when learning ceases, or whether SRT values remain stable over longer study periods. This is especially important for scientific research in which, for example, small differences between parameter settings of hearing aid algorithms need to be resolved and therefore a highly accurate test is necessary.

Learning Effects for TC Speech

The literature concerning learning of TC speech has reported both intrasession and intersession learning effects. Dupoux and Green (1997), Peelle and Wingfield (2005), Adank and Janse (2009), and Adank and Devin (2010) consistently documented an increase in speech comprehension during the first sentences of a speech presentation. Dupoux and Green, for instance, described improved recognition after the presentation of only 5 or 10 TC sentences, indicating “a short term adjustment to local speech rate parameters” (p. 926). In contrast, Golomb, Peelle, and Wingfield (2007) observed intersession learning effects and measured an increase of recall

accuracy between the first and second session about 1 week later for both younger and older participants. Additionally, for the time spent listening to TC speech, Gordon-Salant and Friedman (2011) showed an increasing speech recognition correlated with increasing hours spent per week.

Schlueter et al. (2015) is the only known study that applied matrix sentences of TC speech and analyzed both intrasession and intersession learning effects. When comparing the results of six measurements repeated in a second session, the results showed the same learning effects as documented for the matrix test with original speech. Therefore, Schlueter et al. (2015) recommended the use of at least one list for training if measurements were to be conducted with older NH and older HI listeners, whereas for younger NH listeners, the use of two lists was recommended. Older NH and HI listeners showed smaller effects of intersession learning and therefore a higher test-retest reliability compared with younger NH listeners. It thus seems reasonable to conduct complex measurements across several sessions for older listeners, whereas for younger listeners, data collection should be limited to one session. The combination of TC speech and matrix material appeared to emphasize learning effects. As the speech rate was changed adaptively in Schlueter et al., it was, however, not possible to analyze the learning effect of TC speech in the matrix test and compare it to the learning effect using the original speech material.

Different factors seem to affect the learning of TC speech. Dupoux and Green (1997) described longer learning effects for speech that had faster time compression. Although older listeners achieved poorer recall performance of TC speech compared with younger participants, initial adaptation to TC speech was independent of age (Peelle & Wingfield, 2005): The rate and magnitude of initial learning effects were comparable for younger and older participants. Unlike young participants, however, older participants' learning was dependent on compression ratios; they were impaired in transferring learning effects from one speech rate to another. Golomb et al. (2007) did not confirm this result, but found that, beyond an initial plateau, learning effects in older participants were comparable to those of younger listeners. The absence of age-related effects showed learning processes that were resistant to cognitive decline, or, alternatively, indicated compensatory mechanisms that were applied by older participants.

Transfer of learning from TC to natural fast speech, but not for the reversed presentation of signals, was observed by Adank and Janse (2009). In addition, Pallier, Sebastián-Gallés, Dupoux, Christophe, and Mehler (1998) and Sebastián-Gallés, Dupoux, Costa, and Mehler (2000) showed a transfer for different languages and TC speech. Banai and Lavner (2012)

Table 1. Characteristics of Participating Groups and Test Conditions.

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

Note. NH = normal hearing; HI = hearing impaired; TC = time compressed.

suggested that intrasession and intersession learning effects reflect two phases of the perceptual learning process. They documented that learning of TC speech continued after adaptation to a few sentences and throughout training sessions. In addition, Banai and Lavner (2014) studied the effects of training length on the learning of TC speech and on its transfer to other conditions. Trained participants performed better than untrained participants in trained conditions. Furthermore, only the listeners exposed to the long training protocol showed a higher level of generalization to untrained conditions.

With respect to speech recognition tasks in noise, the results of Banai and Lavner (2012, 2014) suggest the transferability of learning of TC speech to original speech during the initial phase of learning. However, it is still not known whether the transfer of learning is possible for the second phase of further repetitions. The transfer of learning acquired during the initial phase should result in the attainment of better thresholds (i.e., lower SRTs) for participants who were trained with TC speech and subsequently presented with original speech than for naïve participants. The transfer of learning acquired during the second phase of further repetitions in participants who are trained with TC speech could even result in thresholds for subsequently presented original speech that are comparable to thresholds of participants who are trained with original speech. The remaining threshold difference for original speech of both groups of participants should allow for a distinction between the initial phase of brief adaptation on one hand and the subsequent learning during further repetitions on the other.

Research Questions of the Current Study

To investigate potential learning effects, the OLSA was conducted in repeated measures within consecutive sessions. NH and HI participants listened to the original sentences and to TC sentences. It was expected that learning of the speech would progress through an initial general phase to a subsequent, more prolonged and

stimulus-specific phase. Consequently, SRT values should improve with repetition within and between sessions. This pattern of results was expected to be the same for NH and HI listeners, as well as for the original and TC speech. The results were complemented by an additional analysis of the transfer of learning between TC and original speech.

Methods

Participants

Participants were assigned to the four groups outlined in Table 1, that is, NH and HI listeners, presented either with original or with TC speech material. Based on pure-tone audiometry, NH listeners had 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 listeners are depicted in Figure 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.

Stimuli

Speech recognition performance was measured using the German matrix test (OLSA, Wagener et al., 1999a). All sentences had the same structure: name, verb, numeral, adjective, and object. The sentences were generated from a random selection of 1 of the 10 words for each structural element of the sentence. The test consisted 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 had the same long-term spectrum as the speech (Wagener et al., 1999a).

For the presentation of TC speech with the original fundamental frequency, the OLSA sentences were processed with a pitch-synchronous, overlap-add procedure implemented in the software Praat (Boersma & Weenink, 2009). This approach analyzes the pitch of a speech signal, sets pitch marks, and segments the original signal into

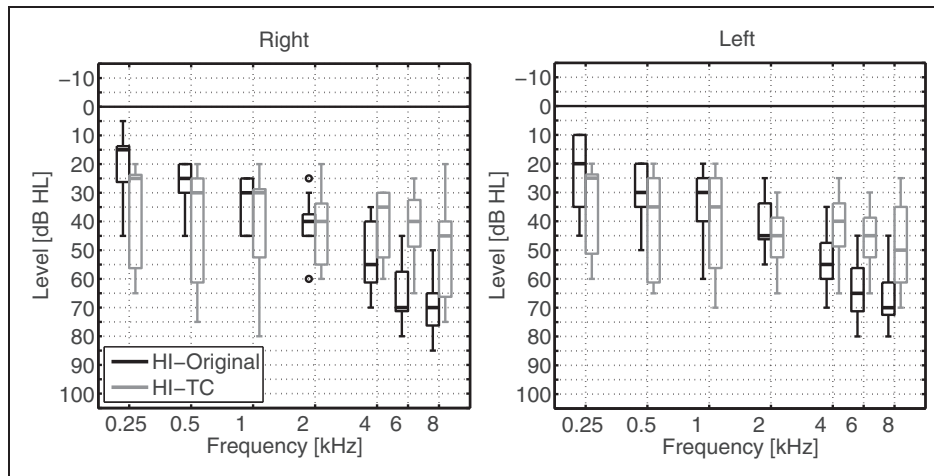


Figure 1. Box plots showing the results of pure-tone audiometry testing for the right and left ear of the HI participants, who listened to original (black) or TC (gray) speech material. Shown are the medians, interquartiles, whiskers (1.5 times interquartile ranges at most), and outliers.

Note. HI = hearing impaired; TC = time compressed.

windowed frames. Afterward, segments at regular intervals are deleted and the remaining segments are concatenated to the TC signal. The position and number of deleted segments are dependent only on the time-compression factor and are not influenced by speech characteristics. This compression differs from natural fast speech in that, for example, pauses and vowels are compressed the most when compared with other parts of the speech (e.g., Covell, Withgott, & Slaney, 1998). Schlueter, Lemke, Kollmeier, and Holube (2014) showed only very small differences between the long-term spectra of original speech and of TC speech as processed with Praat. Among other reasons, they recommended Praat for generating TC speech because it resulted in similar signal quality when compared with 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 based on Schlüter, Holube, and Lemke (2013), where younger NH scored 50% recognition for speech presented at an SNR of 1 dB and compressed to about 30% of its original length. HI showed equal speech recognition scores at a compression to about 50% and SNRs of 1 dB or higher.

Measurements

After brief anamnesis, otoscopic examination, and pure-tone audiometric testing, the OLSA measurements were performed in an acoustically shielded 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 to the NH listeners. The HI listeners were presented with a noise level that depended on their hearing threshold at both 0.5 and 1 kHz. The level correction was based on the half-gain rule but without frequency shaping. The corrected overall presentation level, l , was calculated using Equation (1).

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

If HI listeners 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 for HI was 77 dB SPL (range: 70–85 dB SPL). The measurement of the SRT was conducted with an adaptive procedure, adjusting the presentation 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 the previous sentence, the target recognition of 50%, the slope of the assumed discrimination function, and the rate of convergence (Brand & Kollmeier, 2002). The first sentence of each list was presented at 0 dB SNR or 10 dB SNR for

normal and TC speech, respectively. After each test list, the SRT was estimated using a maximum likelihood method based on all SNRs and sentence-specific recognition scores within that test list (Brand & Kollmeier, 2002).

All participants took part in five sessions, which were arranged at 1- to 3-day intervals. During each session, they completed six lists with randomly selected list numbers of the OLSA. In the last session, participants of the two groups who had listened in all previous measurements to the TC speech condition completed an additional, seventh, list of original speech material. The study was approved by the Ethical Committee of the Carl von Ossietzky University, Oldenburg.

Results

NH Participants

General results. Figure 2(a) provides an overview of the SRT results for six successively measured lists and for each of the five sessions. These results are presented separately for the NH-Original and NH-TC groups. In general, a trend of decreasing SRTs within and between sessions was observed, clearly showing higher SRTs for TC than for original speech. Also, the overall improvement in SRT across measurements and sessions was larger for TC than for original speech.

Intrasession learning. As described in the Introduction, Wagener et al. (1999b) recommended two training lists (or up to 60 sentences) for the matrix test. Based on their recommendation, initial intrasession learning effects were studied using the differences between the first and third list of the first session. These differences are shown in Figure 3 and reflect the improvement that can be achieved through performing the recommended training. The results of NH-Original group (see Figure 3(a)) 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. As shown in Figure 3(a), NH listeners achieved smaller improvements for original (median difference of SRTs: NH-Original 1.2 dB) than for TC speech (median difference of SRTs: NH-TC 7.8 dB). These improvements were significantly different (U -Test; NH: $p < .001$).

For further analysis, it was assumed that the recommended two training lists (Wagener et al., 1999b) had been conducted, that is, data collection started with the third list. This approach allowed the study of ongoing learning within a session. As no statistically significant differences were found between results from the fifth and sixth lists for all sessions and groups (Wilcoxon test for all sessions and groups: $p > .078$), the mean of the fifth and sixth lists served as a representation of the SRT at the end of a session. For all five sessions, ongoing

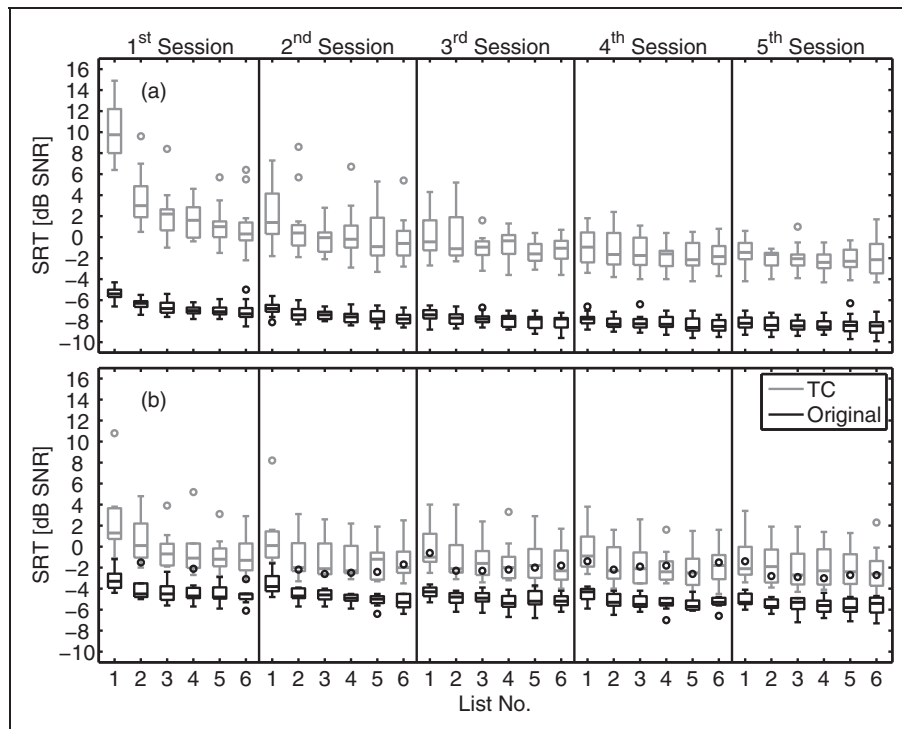


Figure 2. Box plots of SRT values obtained in five sessions with six successively measured lists. Results are from groups of (a) NH and (b) HI as well as original (black) and TC (gray) speech material.

Note. SRT = speech reception threshold; NH = normal hearing; HI = hearing impaired; TC = time compressed.

intrasession learning effects were calculated using the individual differences between the results of the third list and the end of each session (mean of the fifth and sixth lists). Median differences for this measure of intrasession learning effects are shown in Table 2; statistically significant differences are denoted by an asterisk. These intrasession learning effects were not significantly different between sessions when analyzed for each NH group

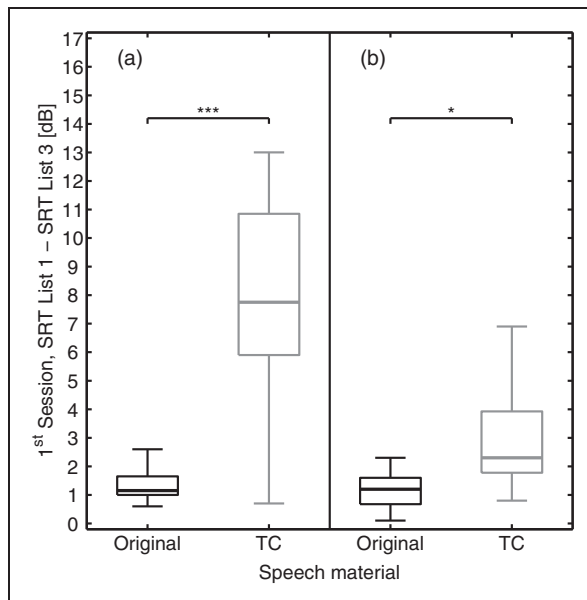


Figure 3. Intrasession learning. SRT differences between the first and third lists in the first session for (a) NH and (b) HI, as well as for original (black) and TC (gray) speech material. Significance was analyzed with an *U*-Test and significant *p*-values are displayed (**p* < .05, ****p* < .001).

Note. SRT = speech reception threshold; NH = normal hearing; HI = hearing impaired; TC = time compressed.

separately (Friedman, NH-Original: *p* = .058, NH-TC: *p* = .167).

To determine significant differences of lists within sessions, Friedman tests were performed with Bonferroni correction and comparing all SRTs of each session. NH-Original group showed significantly different results within all sessions (*p* < .008), whereas significance for NH-TC group was limited to the first three sessions (*p* < .019). To determine the number of lists after which the SRT was not significantly different, Wilcoxon tests with Bonferroni correction were calculated for the first to fourth lists in comparison to the end of each session (mean of fifth and sixth lists). For NH-Original group and in the first three sessions, the first, second, and third lists were significantly different from the end of a session (*p* < .020). In the fourth session, only the first and second list were significantly different from the session's end (*p* < .049), and in the fifth session, the Wilcoxon test only obtained significance for the first list compared with the end (*p* = .007). For NH-TC group, the first, second, and third list were significantly different from the end of the first session (*p* < .024). Furthermore, the first and second list of the second session reached significance when compared with the end (*p* < .048). In the third session, only the first list was different to the end (*p* = .024). In the fourth and fifth session, none of the lists were significantly different from the end of the sessions.

Intersession learning. Intersession learning effects were assessed by comparing the mean SRTs of the fifth and sixth lists for all sessions. These results are depicted in Figure 4(a) for NH listeners and show decreasing SRTs for all sessions that are significantly different for NH-Original and NH-TC groups (Friedman test, NH-Original: *p* < .001, NH-TC: *p* < .001). To determine the session after which all other sessions are not significantly different from the fifth and last session, Wilcoxon tests

Table 2. Intrasession and Intersession Learning.

Learning	session	NH-Original	NH-TC	HI-Original	HI-TC	All
Intrasession	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)
Intersession		1.4 dB*** (0.7 dB)	3.0 dB* (2.2 dB)	0.9 dB* (0.7 dB)	1.2 dB* (1.0 dB)	

Note. NH = normal hearing; HI = hearing impaired; TC = time compressed; SRT = speech reception threshold. Intrasession learning effects are described by the median SRT differences between the third and the mean of fifth and sixth lists (and the interquartile ranges). For intersession learning effects, mean SRTs of the fifth and sixth lists were first computed for each participant and session. Subsequently, the median differences (and the interquartile range) between the first and fifth sessions were calculated. Results are displayed for NH and HI groups who listened to original or TC speech, as well as the results for all participants and speech signals (All). Statistical testing consisted of a Wilcoxon test with Bonferroni correction, which compared the values used for the calculated differences. Significant *p*-values are displayed (**p* < .05, ***p* < .01, ****p* < .001).

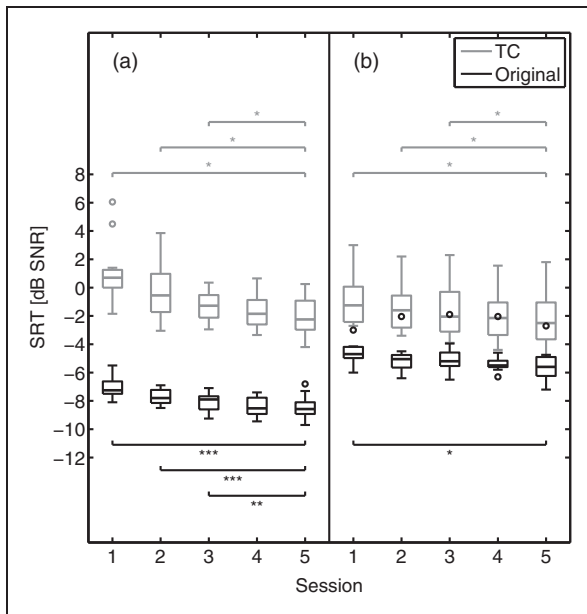


Figure 4. Interession learning. Mean SRTs of the fifth and sixth lists for five sessions performed by (a) NH and (b) HI using original (black) and TC (gray) speech material. Statistical testing consisted of a Wilcoxon test with Bonferroni correction and significant p -values are displayed (* $p < .05$, ** $p < .01$, *** $p < .001$).

Note. SRT = speech reception threshold; NH = normal hearing; HI = hearing impaired; TC = time compressed.

with Bonferroni correction were performed. For NH-Original group as well as NH-TC group, significant differences of the SRT were found for the first three sessions in comparison to the fifth session (NH-Original: $p < .003$, NH-TC: $p < .023$). The fourth session was not significantly different from the fifth. In addition, Table 2 depicts the median difference between the first and fifth session for all groups. Significantly larger interession learning effects were found for TC than for original speech (U -test, $p < .001$).

Transfer effects. Participants of the group NH-TC completed one additional, final, list using the original speech material. Figure 5(a) shows the SRTs from this list, together with the last measurement of the group NH-Original. NH-TC listeners who trained on TC speech during the experiment achieved higher SRTs for original speech than NH-Original listeners, who trained on original speech during the experiment (Figure 5(a)). An U -test confirmed the results ($p < .001$).

Hearing-Impaired Participants

General results. Figure 2(b) gives an overview of SRT results for the repeated measurements of HI listeners. The general trend of decreasing SRTs within and

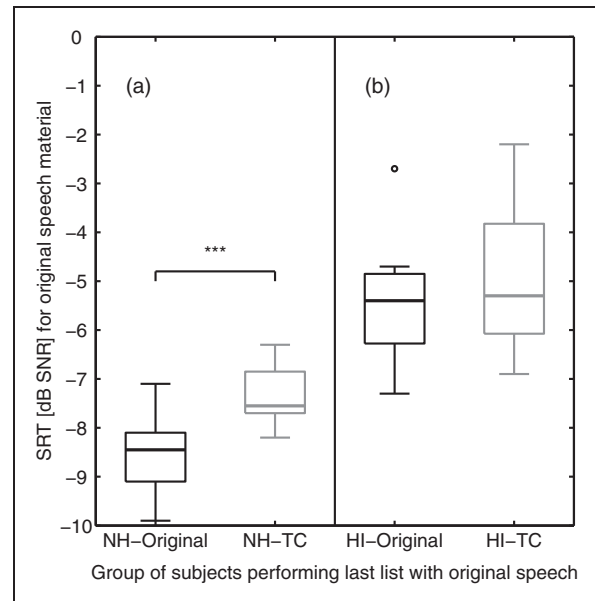


Figure 5. SRTs for original speech. Shown are the results of the sixth list within the fifth session of NH/HI Original and the results of a seventh list within the fifth session of NH/HI TC. Significance was analyzed with an U -Test and significant p -values are displayed (*** $p < .001$).

between sessions, as well as higher SRTs for TC than for original speech, was also found in HI listeners.

Intrasession learning. For the first session, differences between the first and third lists for HI listeners are shown in Figure 3(b) and reflect the improvement resulting from using the two recommended training lists. As for NH, HI listeners showed a smaller improvement for original (median difference of SRTs between first and third list: HI-Original 1.2 dB) than for TC speech (median difference of SRTs between first and third list: HI-TC 2.3 dB). These improvements were significantly different (U -Test $p = .012$).

As for NH listeners, and because of no significant differences (Wilcoxon test for all sessions and groups: $p > .067$), the mean of the fifth and sixth lists was used to represent the SRT at the end of a session. Again, intrasession learning effects were examined (individual differences between results of the third lists, the list after two recommended training lists, and mean of fifth and sixth lists; medians listed in Table 2). When analyzed for each HI-group separately, the differences in intrasession learning effects between sessions did not reach statistical significance (Friedman, HI-Original: $p = .422$, HI-TC: $p = .082$).

As described for the NH listeners, Friedman tests with Bonferroni correction were calculated to determine whether differences within sessions were observable, when including all SRTs in each session.

For HI-Original group, significant differences were found, except for the fifth sessions ($p < .030$). For HI-TC group, significant differences were obtained for all sessions ($p < .016$).

Wilcoxon tests with Bonferroni correction were then performed to determine the lists with significantly different SRTs compared to the end of each session (mean of fifth and sixth lists). For HI-Original group, the first list was significantly different from the end of the session for the first session ($p = .030$) and the fourth session ($p = .031$). Significant differences were also obtained for the first and second list of the second session ($p < .043$). Within the third and the fifth session, none of the lists were significantly different from the end. For HI-TC group, the first and the second list showed significant differences compared with the end of the first session ($p = .031$). Additionally, only the first list was significantly different from the end for the second to the fifth session ($p < .046$).

Intersession learning. Intersession learning effects (mean SRTs of the fifth and sixth lists) of HI listeners are depicted in Figure 4(b) and show decreasing SRTs for all sessions. For both groups, sessions showed significant differences (Friedman test, HI-Original: $p = .005$; HI-TC: $p < .001$). Wilcoxon tests with Bonferroni correction were used to analyze differences between each session and the end of the measurements (fifth session). Results of HI-Original group showed that the first and second session were significantly different from the fifth session ($p < .051$). For HI-TC group, the first to third sessions were different to the end ($p < .043$). In addition, Table 2 depicts the median difference between the first and fifth session for all HI-groups. Intersession learning effects did not differ significantly for TC speech compared with original speech (U -test: $p = .222$).

Transfer effects. As for NH-TC group, participants of the group HI-TC completed one additional, final list using the original speech material. Figure 5(b) shows these SRTs together with the last measurement of group HI-Original. HI listeners showed comparable SRTs, whether they listened to original or to TC speech during the experiment (U -test, $p = .566$).

Discussion

In general, recognition in NH and HI listeners was poorer for TC than for original speech, resulting in higher SRT values for TC speech. The results of the current study showed a general trend of decreasing SRT values after practice even beyond the previously observed short-term learning effect when using the OLSA. Although in detail, participants started each new session at higher SRT values than those 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 readapt in order to reach the SRT values of previous sessions.

The observed learning effects have an impact on speech-in-noise tests that involve a matrix structure. In clinical applications of speech-in-noise tests, the number of repetitions of sessions is commonly smaller than in scientific research and, as a result, intrasession and intersession learning effects can affect these situations differently. In the following, results are compared with previous studies, using speech tests for original and TC speech separately. The discussion focuses on the number of necessary training lists and the possibility of comparing measurements across sessions.

Original Speech

Number of training lists. As expected from results presented by Wagener et al. (1999b) and Wagener and Brand (2005), different SRT values for NH and HI listeners were measured for original speech. Nevertheless, the initial intrasession learning effect of NH and HI in the first session showed a similar median improvement of the SRT values of about 1 dB for the first two lists. Our results confirm Wagener et al.'s (1999b) findings and support their recommendation for the use of two training lists 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. recommended one to two training lists (with 20 sentences each) or up to 60 sentences dependent on the required measurement accuracy for the purpose of training. They posited that such training facilitates both familiarization with the measurement procedure and adaptation to the structure of the OLSA sentences. In the current study, although NH listeners showed significant differences for the first three lists compared with the end of the first session, differences after two training lists were as small as 0.4 dB. This result is in agreement with Wagener et al., who observed an test-retest accuracy of about 0.5 dB for SRT measurements after initial training using two lists (of 20 sentences each) in one session, the same as for HI listeners in this study.

Beyond previous results for the OLSA, the current study examined learning effects for further sessions. For NH listeners, the subsequent sessions again showed significant differences for up to three lists (for the second and third session) compared with the end of the sessions. However, the differences after two training lists compared with the end of the sessions are generally smaller than 0.4 dB. Therefore, the recommendation of two training lists also seems to be useful for subsequent

sessions for NH listeners. For HI listeners, only the second session SRTs showed significant differences for the first two lists compared with the end of a session. All other sessions showed significant differences to the end of the respective session for the first list only or no significance between the lists. The differences after two training lists were smaller than 0.6 dB for all sessions. For HI listeners, the question remains whether one training list is sufficient for measurements in consecutive sessions. As all current results were conducted with a high number of measurements within one session, learning effects in subsequent sessions were partly influenced by the large number of lists presented in the previous sessions and could therefore be more pronounced when a smaller number of measurements per session is used.

A lack of training has negative implications in the process of fitting a hearing aid, because the required improvement of 2 dB in SRT for speech in noise (that is supposed to indicate a beneficial hearing aid fitting; see Gemeinsamer Bundesausschuss, 2014) might be generated through learning. For scientific research, even smaller improvements and therefore a high accuracy in speech-in-noise tests might be of interest. Hence, two training lists are recommended for both NH and HI listeners, for each session, which also serves the common request to use identical protocols for both subject groups.

Comparison across sessions. Regarding intersession learning effects, statistically significant improvements of SRT values were observed for the first to the third session (NH) or the first to second session (HI), when they were compared with the last session. Additionally, the NH listeners showed a larger median improvement of the SRT than the HI listeners from the first to the fifth session. These results are in contrast to Hagerman (1984) and Hagerman and Kinnefors (1995), who observed more learning with poorer hearing ability. The intersession learning effect observed in this study, however, does confirm results by Wagener and Brand (2005), who found larger SRT improvements over time for NH listeners than for HI listeners. However, the improvements due to intersession learning observed over five sessions in the current study exceeded those of Wagener and Brand, who compared results of only two subsequent sessions. The current results call attention to the observation that for both NH and HI listeners, the learning process continues after two sessions. This assumption is based on the comparison of results for different sessions while using the subject groups as their own controls. Independent control groups with less training were not included in the study. Therefore, the impact of additional training, for example, in the second to fourth session, on the performance in the fifth session cannot be compared to reference values obtained without this additional training.

Nevertheless, it can be concluded that 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. However, the fact that none of the observed SRTs for NH listeners were below -10 dB SNR (which is approximately 3 dB below the normative value for native listeners after initial training) points toward a floor effect, that is, the decrease in SRT due to learning might not continue if even more training is employed.

Therefore, even though the long-term training effect may be limited to 3 dB at maximum, 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 or crucial comparisons should be grouped within one session. These recommendations are in line with Hernvig and Olsen (2005) for the Danish matrix test. In addition to these suggestions, the test conditions should be randomized across participants and sessions.

TC Speech

Banai and Lavner (2012, 2014) reported initial learning effects during early sessions and additional ongoing learning effects in later sessions for TC speech. This pattern is comparable to the results of the current study obtained with a matrix test, where larger learning effects were observed during early measurements and ongoing learning effects were observed within and between sessions. In the following, the necessary training for TC speech is discussed similarly to the discussion on original speech.

Number of training lists. Both the NH-TC and HI-TC groups showed larger initial intrasession learning effects within the first session than for original speech and also showed further improvements within each session. Comparisons of the lists to the end of a session for the NH listeners revealed significant differences for up to three lists, whereas significant differences for HI listeners were limited to one or two lists. The improvements after the first two training lists were about 1 dB for the first session and smaller than 0.4 dB for following sessions. Especially, the results of the first session for NH listeners might argue for a higher number of training lists when using TC speech rather than for original speech. Therefore, in summary, two training lists are also advised for TC speech, to ensure efficient test administration for NH and HI listeners. Prior training with the original speech test might be advantageous for NH listeners.

Comparison across sessions. The results were obtained during five sessions, which were arranged

at 1- to 3-day intervals. Significant differences of SRTs compared with the last session were observed for NH and HI listeners until the third session. The results of intrasession and intersession learning showed an explicit adaptation to TC speech and confirmed the observations of, for example, Dupoux and Green (1997) and Golomb et al. (2007). Similar to the results for uncompressed speech, the learning was limited in the current study due to an assumed floor effect at an SRT of approximately -5 dB dependent on the applied time compression.

As the results of different sessions were significantly different for TC speech, comparisons of SRT values measured in different sessions should be avoided. It is therefore advisable to design studies using measurements within one single session and to randomize test conditions across participants. Participants with previous experience in listening to TC speech can be included, as long as differences between measurement conditions are examined within one session. However, the degree of experience with TC speech is important for the determination of absolute SRT values, as recognition correlates positively with the time of exposure to TC speech (Gordon-Salant & Friedman, 2011).

Comparison between NH and HI groups. Different time compression, 30% and 50% of original length, was selected for NH and HI listeners in order to reach comparable speech recognition for both groups after the initial learning process. This goal was achieved, as only the first two lists within the first session were significantly different between NH-TC and HI-TC groups (*U*-test with Bonferroni correction; first session first list: $p = .005$, first session second list: $p = .041$). All subsequent measurements resulted in similar SRTs. Hence, differences of learning between NH-TC and HI-TC groups were observed in the first session. Furthermore, the number of initial lists that differ significantly from the end of a session varied between both groups. Intersession results showed that the effect of learning over all sessions was larger for NH than for HI listeners. These differences occurred even though similar intelligibility of the speech was achieved after two lists by the different time compression of the speech material. One factor for the different amount of learning for NH-TC and HI-TC groups might be the applied differing time compressions themselves. The group NH-TC experiences more time compression and shows worse SRT at the beginning of the experiment compared with the group HI-TC. But then, the SRT of both groups seems to be limited at about -5 dB SNR for TC speech, that is, lower SRTs are not achievable even after several sessions. Listeners of the group NH-TC might show larger learning effects because they have trouble coping with the high amount of time compression at the beginning of the experiment. It is still unclear whether the listeners of group HI-TC

would show the same amount of learning as the NH-TC group if they would be tested with the same time compression of 30% or if the NH-TC group would also be tested with a time compression of 50%. However, different time compression will most probably be used in practical applications for NH and HI listeners to achieve similar scores in speech tests.

In addition, it remains unclear whether the observed differences of learning found in the current study for TC speech presented to NH and HI listeners are related to the hearing ability or the age of participants. Previous studies have documented a deterioration of recognition for fast speech that is associated with age. Older NH listeners performed less well in speech recognition tasks with TC speech than did younger NH listeners (Golomb et al., 2007). However, the current study used an incomplete design; it was conducted with younger NH and older HI listeners. Thus, no conclusion on the differential effects of hearing ability and age can be drawn and, therefore, detailed statistical comparisons between NH and HI listeners were omitted. In summary, as less training is necessary in this study for older HI than for young NH listeners, and the same training protocol is recommended for young NH and old HI listeners, the recommended protocol is expected to also be suitable for older NH and younger HI listeners.

Transfer effects. In addition, transfer of learning from TC to original speech was explored. All participants who listened to the TC speech condition (NH-TC, HI-TC) performed one final measurement with original speech. Their SRT results for original speech were compared with results of NH and HI listeners who participated in the original speech condition (NH-Original, HI-Original). As transfer effects were only tested from TC speech to original speech, observations cannot be supported by transfer effects from original to TC speech. Additionally, the performance of the NH-TC and HI-TC groups using original speech without learning, and whether it is comparable to the performance of NH-Original or HI-Original groups, is unknown. Nevertheless, the results, especially of the NH-TC group, exhibited trends. NH-TC group showed lower SRT values for their last measurement with original speech than naïve NH listeners within their first measurement of original speech. But they show comparable median SRT values to those measured with the original OLSA for NH listeners after two initial training lists (Wagner et al., 1999b). Therefore, participants were able to apply the learning of the initial phase. According to Wagner et al. (1999b), this first phase included customization to the measurement procedure and to the formal structure of the sentences, both being kinds of information that are independent of the speech rate. The comparison of the current data to the

thresholds observed by Wagener et al. indicates that the initial phase of brief adaptation was completed after about two initial training lists and the second, more prolonged and stimulus-specific phase of learning dominated the effects subsequently observed (see also Banai and Lavner, 2012). This is supported by the observation that NH-TC group failed to completely transfer learning. They showed nearly 1 dB poorer SRT values for original speech than NH-Original group who were trained with original speech. The absence of a total transfer indicates that the learning processes for original and TC speech did not fully accumulate. This might also be valid for further alterations of the speech signal or for matrix tests using different languages which is in line with observations made by Pallier et al. (1998) and Sebastián-Gallés et al. (2000), who showed that the adaptation to TC native speech was supported by previous exposure to TC, nonnative speech that resembled native speech. For this adaptation to take place, however, the listener did not necessarily need to know the foreign language. Difficulties in the transfer occurred for signals that were dissimilar.

The nonsignificant difference of the SRT values for original speech between TC-trained and Original-trained HI groups was unexpected and different reasons might explain these results. First, the variability of measured SRT values was larger for HI than for NH listeners and this might conceal the differences in SRT due to different learning mechanisms. Second, the two hearing-impaired groups that were trained with original or TC speech showed different hearing levels in pure-tone audiometric testing. At high frequencies between 4 and 8 kHz, HI-TC group exhibited a lower average hearing loss than HI-Original group. Even though the effect of high-frequency hearing loss on SRT in noise is limited, this might still 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 groups. Nonetheless, possible differences between the two groups of HI listeners remain speculative.

Conclusions

- Speech recognition 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 six lists and five sessions). However, the maximum learning effect of the second phase seems to be limited by a floor effect, thus yielding a maximum SRT decrease of 3 dB for normal listeners and uncompressed speech.
- NH and HI listeners exhibit larger improvements in SRT for TC than for original speech.
- If matrix tests are used with original or TC speech, two training lists should be administered in each session before measuring SRTs.
- In scientific applications that use original speech and that aim to compare small differences or results of different sessions, the potential effects of intrasession and intersession learning require a careful randomization of test situations across sessions. The recruiting of experienced listeners may be beneficial.
- If tests are conducted with TC speech, comparison of results of different sessions is not recommended; instead, measurements should be performed within one session.

Acknowledgments

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 www.stels-ol.de for English language support.

Author Note

Parts of the results were presented at the 15th Annual Conference of the Deutsche Gesellschaft für Audiologie (2012) in Erlangen, Germany, and the International Hearing Aid Conference (2012) in Lake Tahoe, USA.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: 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).

References

- Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage*, *49*(1), 1124–1132.
- Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America*, *126*(5), 2649–2659.
- Banai, K., & Lavner, Y. (2012). Perceptual learning of time-compressed speech: More than rapid adaptation. *PLoS One*, *7*(10), e47099.
- Banai, K., & Lavner, Y. (2014). The effects of training length on the perceptual learning of time-compressed speech and its generalization. *The Journal of the Acoustical Society of America*, *136*(4), 1908–1917.
- Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved from www.praat.org, (Last viewed: 08/02/2011).

- Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, *111*, 2801–2810.
- Covell, M., Withgott, M., & Slaney, M. (1998). MACH1: Nonuniform time-scale modification of speech. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal processing, Seattle, WA, Vol. 1, pp. 349–352.
- Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology. Human Perception and Performance*, *23*(3), 914–927.
- 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.
- Gemeinsamer Bundesausschuss. (2014). Richtlinie des Gemeinsamen Bundesausschusses über die Verordnung von Hilfsmitteln in der vertragsärztlichen Versorgung (Hilfsmittel-Richtlinie/Hilfsm-RL), *Bundesanzeiger*. Retrieved from https://www.g-ba.de/downloads/62-492-934/Hilfsm-RL_2014-07-17.pdf.
- Golomb, J. D., Peelle, J. E., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *The Journal of the Acoustical Society of America*, *121*(3), 1701–1708.
- Gordon-Salant, S., & Friedman, S. A. (2011). Recognition of rapid speech by blind and sighted older adults. *Journal of Speech Language and Hearing Research*, *54*(2), 622–631.
- Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology*, *11*(2), 79–87.
- Hagerman, B. (1984). Clinical measurements of speech reception threshold in noise. *Scandinavian Audiology*, *13*(1), 57–63.
- Hagerman, B., & Kinnefors, C. (1995). Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. *Scandinavian Audiology*, *24*(1), 71–77.
- Hervig, L. H., & Olsen, S. Ø. (2005). Learning effect when using the Danish Hagerman sentences (Dantale II) to determine speech reception threshold. *International Journal of Audiology*, *44*(9), 509–512.
- Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., & Kollmeier, B. (2012). A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International Journal of Audiology*, *51*(7), 536–544.
- Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology*, *54*(Suppl 2), 3–16.
- Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report. *American Journal of Audiology*, *7*(2), 21–25.
- Ortiz, J. A., & Wright, B. A. (2009). Contributions of procedure and stimulus learning to early, rapid perceptual improvements. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 188–194.
- Ozimek, E., Warzybok, A., & Kutzner, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International Journal of Audiology*, *49*(6), 444–454.
- Pallier, C., Sebastián-Gallés, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory and Cognition*, *26*(4), 844–851.
- Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1315–1330.
- Schlüter, A., Holube, I., & Lemke, U. (2013). Verfahren zur Bestimmung der Zeitkompressionsschwelle von Sprache im Störgeräusch. In *16. Jahrestagung der Deutschen Gesellschaft für Audiologie*. Rostock, Germany.
- Schlueter, A., Brand, T., Lemke, U., Nitzschner, S., Kollmeier, B., & Holube, I. (2015). Speech perception at positive signal-to-noise ratios using adaptive adjustment of time compression. *The Journal of the Acoustical Society of America*, *138*(5), 3320–3331.
- Schlueter, A., Lemke, U., Kollmeier, B., & Holube, I. (2014). Intelligibility of time-compressed speech: The effect of uniform versus non-uniform time-compression algorithms. *The Journal of the Acoustical Society of America*, *135*(3), 1541–1555.
- Sebastián-Gallés, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception and Psychophysics*, *62*(4), 834–842.
- Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology*, *26*, 183–196.
- Wagener, K. C., & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology*, *44*(3), 144–156.
- Wagener, K., Brand, T., & Kollmeier, B. (1999b). Entwicklung und Evaluation eines Satztests für die deutsche sprache III: Evaluation des Oldenburger satztests. *Zeitschrift für Audiologie*, *38*, 86–95.
- Wagener, K., Kühnel, V., & Kollmeier, B. (1999a). Entwicklung und evaluation eines satztests für die deutsche sprache I: Design des oldenburger satztests. *Zeitschrift für Audiologie*, *38*, 4–15.
- Wagener, K., Josvassen, J. L., & Ardenkjær, R. (2003). Design, optimization and evaluation of a Danish sentence test in noise. *International Journal of Audiology*, *42*, 10–17.