J Am Acad Audiol 00:1–13 (2015)

A Phoneme Perception Test Method for High-Frequency Hearing Aid Fitting

DOI: 10.3766/jaaa.15037

Nicola Schmitt*
Alexandra Winkler†
Michael Boretzki*
Inga Holube†

Abstract

Background: Outcomes with hearing aids (HAs) can be assessed using various speech tests, but many tests are not sensitive to changes in high-frequency audibility.

Purpose: A Phoneme Perception Test (PPT), designed for the phonemes */s/* and */ʃ/*, has been developed to investigate whether detection and recognition tasks are able to measure individual differences in phoneme audibility and recognition for various hearing instrument settings. These capabilities were studied using two different fricative stimulus materials. The first set of materials preserves natural low-level sound components in the low- and mid-frequency ranges (LF set); the second set of materials attempts to limit the audibility to high-frequency fricative noise (nLF set). To study the effect on phoneme detection and recognition when auditory representations of */s/* and */ʃ/* are modified, a too strong nonlinear frequency compression (NLFC) setting was applied.

Research Design: Repeated measure design was used under several different conditions.

Study Sample: A total of 31 hearing-impaired individuals participated in this study. Of the 31 participants, 10 individuals did not own HAs but were provided with them during the study and 21 individuals owned HAs and were experienced users. All participants had a symmetrical sensorineural hearing loss.

Data Collection and Analysis: The present study applied a phoneme detection test and a recognition test with two different stimulus sets under different amplification conditions. The statistical analysis focused on the capability of the PPT to measure the effect on audibility and perception of high-frequency information with and without HAs, and between HAs with two different NLFC settings ("default" and "too strong").

Results: Detection thresholds (DTs) and recognition thresholds (RTs) were compared with respective audiometric thresholds in the free field for all available conditions. Significant differences in thresholds between LF and nLF stimuli were observed. The thresholds for nLF stimuli showed higher correlation to the corresponding audiometric thresholds than the thresholds for LF stimuli. The difference in thresholds for unaided and aided conditions was larger for the stimulus set nLF than for the stimulus set LF. Also, thresholds were similar in both aided conditions for stimulus set LF, whereas a large difference between amplifications was observed for the stimulus set nLF. When NLFC was set "too strong," DTs and RTs differed significantly for */s/*.

Conclusions: The findings from this study strongly suggest that measuring DTs and RTs with the stimulus set nLF is beneficial and useful to quantify the effects of HAs and NLFC on high-frequency speech cues for detection and recognition tasks. The findings also suggest that both tests are necessary because they assess audibility as well as recognition abilities, particularly as they relate to speech modification algorithms. The experiments conducted in this study did not allow for any acclimatization of the participants to increased high-frequency gain or NLFC. Further investigations should therefore examine the

---

*Sonova AG, Stäfa, Switzerland; †Institute of Hearing Technology and Audiology, Jade University of Applied Sciences and Cluster of Excellence "Hearing4All," Oldenburg, Germany

Corresponding author: Nicola Schmitt, Sonova AG, CH-8712 Stäfa, Switzerland; E-mail: nicola.schmitt@sonova.com

This study was partly funded by the program Jade2Pro of Jade University of Applied Sciences and partly funded by Sonova AG.

Preliminary results of this work were presented as a poster at the 2012 International Hearing Aid Research Conference, Lake Tahoe, NV, August 8–12, 2012.

impact on DTs and RTs in the PPT as well as the contrasting effects of strong setting of NLFC to DTs and RTs because of (re)learning of modified auditory representations of */s/* and */ʃ/* as caused by NLFC.

Key Words: fitting, hearing aids, hearing device evaluation, high frequency, phoneme, single-subject design, speech test

Abbreviations: ANOVA = analysis of variance; BTE = behind the ear; C = consonant; CV = consonant-to-vowel; dB CL = dB consonant level; DT = detection threshold; FF = free-field audiometry; HAs = hearing aids; LF = low-frequency cues; nLF = without low-frequency cues; NLFC = nonlinear frequency compression; PPT = Phoneme Perception Test; RMS = root-mean-square; RT = recognition threshold; SD = standard deviation; V = vowel; VC = vowel-to-consonant; VCV = vowel–consonant–vowel

INTRODUCTION

Elevated audiometric thresholds are most prominent in the high-frequency region in most patients experiencing hearing loss (Bisgaard et al, 2010). In spite of advanced active feedback cancellation systems and multichannel gain steering technology, hearing aids (HAs) often provide insufficient gain >5 kHz to achieve audibility (Kimlinger et al, 2015). For those where gain does not provide sufficient audibility, nonlinear frequency compression (NLFC) (Simpson et al, 2005), which shifts high-frequency sounds to lower frequency regions where hearing loss can be better compensated by amplification, could be beneficial.

Benefit for speech perception is typically tested using a recognition task: phonemes, syllables, words, or sentences. However, the prerequisites for recognition are audibility and discrimination. Audibility is the threshold for detecting a signal. Therefore, detection is used here as a term to describe the respective task. For discrimination, differences between phonemes have to be perceived to be able to distinguish between the presented items. However, to analyze factors that lead to low recognition scores, the detection and discrimination abilities of the listeners have to be determined. Govaerts et al (2006) introduced a test battery for all three tasks: detection, discrimination, and recognition, for cochlear implant testing in children. This test battery was adapted to German language and extended to adult cochlear implant patients (Arweiler-Harbeck et al, 2011). Within those test batteries, different stimuli were used for the different tasks, instead of evaluating detection, discrimination, and recognition for the same speech material.

Furthermore, it is still unclear how much high-frequency amplification is appropriate (see Hogan and Turner, 1998; Ching et al, 1998; 2001; Turner and Cummings, 1999; Stelmachowicz et al, 2001; Turner and Henry, 2002; Ricketts et al, 2008; Pittman, 2008; Hornsby et al, 2011) and how NLFC algorithms should be fitted (see McCreery et al, 2013). To evaluate these issues, everyday conversations or sentence tests might be inadequate because of the limited necessity of high-frequency input for speech recognition. The recognition of meaningful sentences is facilitated by language knowledge and redundancy. Therefore, low-frequency speech content dominates the frequency importance function for intelligibility for this speech material (Pavlovic, 1987). Much research has concentrated on the development of high-frequency-specific speech perception test material.

The first attempts were made by Beasley and Rosenwasser (1950) using meaningful words that focused on the low- and high-frequency regions. Further speech tests, including high-frequency items, were used by Pascoe (1975), Owens and Schubert (1977), and Foster and Haggard (1987). A more global approach was followed by Owens et al (1985), who developed the minimal auditory capability test battery for evaluating cochlear implants. This test battery includes phonemes, words, and sentences. Dillier and Spillmann (1992) also developed a minimal auditory capability test battery for the German language. For a detailed analysis, the consonants were classified using seven articulatory features that are used in the German language. They noted that classifying the results on acoustic, rather than articulatory, features could be more beneficial for fitting hearing devices.

This approach was followed by Heller (1992), who developed a high-frequency-specific monosyllabic word test with meaningful word pairs differing in one phoneme. One word always contained the phoneme /*s*/, whereas the other word in each pair contained another phoneme in the same position, leading to the largest spectral difference in the high-frequency region. Furthermore, the word pair was presented with a low-frequency noise masker to encourage listeners to use high-frequency speech cues when performing the discrimination task. To improve the sensitivity of the test for the perception of the phoneme /*s*/, Knoblach (1992) kept only those word pairs that changed their meaning when combining an initial /*s*/ (e.g., the German words “Seile” and “Eile”). To avoid the perception of /*s*/ based on the transitions between /*s*/ and the following phoneme (Streeter and Nigro, 1979; Stock, 1996), the same recording of /*s*/ (taken from the German word “Saal” without transition) was combined as the initial phoneme to several following words. These changes enabled differentiation of hearing loss configurations (Knoblach, 1992).

Another approach was followed by Glista et al (2012), Boretzki and Kegel (2009), and Boretzki et al (2011), who focused on the distinction between /*s*/ and /ʃ/. For German, English, and many other languages, the phonemes /*s*/ and /ʃ/ seem to be particularly well suited for analysis in the high-frequency region because of their unique spectral cues. Because of their unvoiced character, they consist of stationary noise and can be distinguished by their specific spectral peaks and bandwidths. The mean center frequencies of the spectral peaks of /*s*/ and /ʃ/ range from 3.2 to 8.4 kHz (Boothroyd and Medwetsky, 1992) and vary strongly between and within speakers (Boothroyd and Medwetsky, 1992; Newman et al, 2001); they can therefore be tailored to test specific frequency regions.

A nonsense syllable test that focuses on the high-frequency region was developed by Kuk et al (2010). They designed a long and a short list of CVCVC syllables spoken by a male and a female speaker. The syllables contained 25 consonants and 5 vowels in different serial positions. To evaluate HA processing in the high-frequency region, Kuk et al (2010) recommended the use of a female speaker. Unfortunately, recognition scores for the consonants, even for the female speaker, were hardly affected when the speech material was low-pass filtered at 4 kHz. This might have been due to the continued availability of low-frequency speech cues within the phonemes. This assumption is supported by Stelmachowicz et al (2001; 2002), who focused on the high-frequency region by studying the perception of /*s*/, /*f*/, and /*θ*/ or /*s*/ and /*z*/, respectively. Therefore, low-frequency speech cues might limit the ability to evaluate the effect of NLFC in the high-frequency domain.

One issue in evaluating NLFC is the possible training effects due to (re)learning of modified auditory representations, which varies strongly between individuals. Glista et al (2009) tested HAs equipped with conventional processing and NLFC outcomes using several speech perception tests. The detection test (Ling 6; Scollie et al, 2012) for /*s*/ and /ʃ/ and the University of Western Ontario Plurals Test for /*s*/ and /*z*/ (Glista and Scollie, 2012), as well as the high-frequency consonant recognition test, showed significant improvements in phoneme perception when using NLFC. On an individual level, multiple subjects experienced greater or lesser benefit than the candidacy predictors would lead one to expect. Wolfe et al (2011) studied the long-term effect of NLFC activated in HA processing for children having moderate hearing loss. The free-field thresholds of 4, 6, and 8 kHz, as well as the University of Western Ontario Plurals Test for /*s*/ and /*z*/ (Glista et al, 2009), and a previous version of the phoneme test used in this study (denoted as “Phonak Logatom Test”; Boretzki and Kegel, 2009) showed significant improvements with NLFC for high-frequency consonants. Although the audibility tests (detection of consonants and free-field thresholds) showed no further improvement after a six-month acclimatization period, performance on the recognition test increased during this period.

In summary, research indicates that a speech perception test should include detection, discrimination, and recognition tasks. In the case of high-frequency HA fittings, the speech cues of the material should be sensitive to this frequency region. The fricatives /*s*/ and /ʃ/ fulfill this criterion. Possible low-frequency cues (LF) resulting from low-frequency fricative noise and vowel-to-consonant (VC) or consonant-to-vowel (CV) transient could influence speech perception results because of the large differences in audibility between high and low frequencies in the presence of high-frequency hearing losses, and should therefore be restricted.

The objective of this study was to investigate whether the detection and recognition tasks of a Phoneme Perception Test (PPT) are able to measure individual differences in phoneme audibility and recognition for various hearing instrument settings. The capabilities were studied with two different fricative stimulus materials. The first set of materials preserved natural low-level sound components in the low- and mid-frequency ranges, and the second set of materials limited the audibility of high-frequency fricative noise. With the second set of materials, the hypothesis was that the PPT distinguishes better between amplification differences of HAs in the high-frequency region and the effect of NLFC. To study the effect on phoneme detection and recognition when auditory representations of /*s*/ and /ʃ/ are modified, “too strong” NLFC settings were applied.

MEASUREMENT PROCEDURES

This section first characterizes the participants and then describes the methods used in the experiments, that is, the PPT and the free-field audiometry (FF). Finally, the HA fitting and the experimental conditions are described.

Participants

A total of 31 (9 female, 22 male) hearing-impaired individuals participated in this study (mean age = 73.7 yr, standard deviation [SD] = 5.5 yr). All participants were provided with test behind-the-ear (BTE) HAs (see “Fitting of Test BTEs” below). Of the 31 participants, 10 individuals did not own an HA (the “novice HA users” group A), and 21 individuals owned an HA and were experienced users (the “experienced HA users” group B; mean duration of experience = 9.5 yr; SD = 8.6 yr). The private HAs were all BTE types, including receiver-in-the-canal types, of different manufacturers and fitted between four weeks and up to 10 yr ago (~4 yr on average). None of them included an algorithm for frequency modification (e.g., NLFC, frequency transposition). It was assumed that the HAs were fitted properly, because all HA users were satisfied with them. The fittings were not verified or modified.

All participants had a symmetrical sensorineural hearing loss. Table 1 gives the age, sex, and pure-tone average thresholds at 500 Hz, 1, 2, and 4 kHz. The mean air-conduction, pure-tone audiometric thresholds for both groups are shown in Figure 1. The participants were provided a small honorarium to offset expenses related to their participation (12 euro per hour).

PPT

The PPT consists of three subtests: a detection test, a distinction test, and a recognition test. The detection test measures the threshold for hearing the phonemes, the distinction test examines whether differences between two phonemes can be perceived, and the recognition test measures recognition thresholds (RTs) of corresponding nonsense words. In the present study, only the detection and the recognition tests were used. The PPT uses the detection thresholds (DTs) and RTs for /*ʃ*/ and /*s*/. In the recognition test, the /*s*/ and /*ʃ*/ stimuli were presented as vowel–consonant–vowel (VCV) tokens using the /*a*/ vowel sound (i.e., /*aʃa*/ and /*asa*/). The stimulus material and the test methods are described in the following subsections.

Stimuli

The audio material was recorded in a sound-attenuating booth with a separate recording and studio area. The distance of the speaker's aperture to the microphone was approximately 0.2 m. All recordings were performed in a single session. A Neumann U87 microphone (Georg Neumann GmbH, Berlin, Germany), a Studer Vista7 mixing console (Studer Professional Audio GmbH, Regensdorf-Zurich, Switzerland), a D19M input/output device (Studer Professional Audio GmbH), and the software ProTools (Avid Technology, Inc., Burlington, MA) were used to record the mono audio material with a resolution of 32 bits and a 48-kHz sampling rate.

Seven VCVs were spoken by a 28-yr-old female, native German speaker. The speaker was trained to speak all VCVs with a similar speed and pitch as much as possible. The first and the second vowels of the VCVs always consisted of /*a*/. The middle consonant in all words was one of the following: /*d*/, /*f*/, /*h*/, /*k*/, /*m*/, /*s*/, or /*ʃ*/ The first recorded VCV was *ama*, which served as a reference in loudness, pitch, and length during subsequent recordings. The speaker listened via Beyerdynamic MDR 7506 headphones (beyerdynamic GmbH & Co. KG, Heilbronn, Germany) to the reference VCV while all the other stimuli were spoken. Each VCV was recorded several times. From all recordings, one version of every VCV was selected, based on similarity in speaking rate and pitch.

Table 1. Mean Age, Sex, and Pure-Tone Average Thresholds at 500 Hz, 1, 2, and 4 kHz of Groups A and B

	Age (yr)	Sex	PTA4 (dB)
Group A	71.3	1 Female, 9 Male	44.6
Group B	75.0	8 Female, 13 Male	57.3

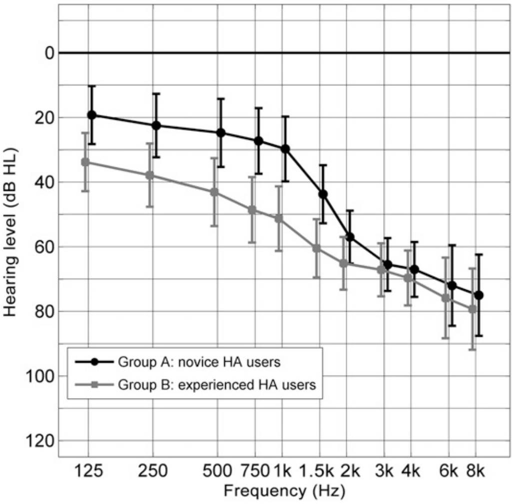

Figure 1. Pure-tone audiogram for groups A and B, means (±SD) as a function of frequency.

The PPT was designed to measure the perception of high-frequency consonants without the influence of random differences in pronouncing the /*a*/ in the VCVs (e.g., pitch, loudness, and length) that could serve as cues during the test. Hence, the initial and the final /*a*/ were replaced in all recorded VCVs by the initial and final /*a*/ of *ada* by cross-fading. Cross-fading was carried out by decreasing the level of the initial /*a*/ of *ada* to zero and increasing the level of the initial /*a*/ of the targeted VCV over a small period and just before the transition between /*a*/ and the following consonant. The same procedure was applied for the final /*a*/. This procedure preserves the VC and CV transient information.

To test detection and recognition of high-frequency speech cues from different speakers, the spectra of the two phonemes /*s*/ and /*ʃ*/ were adjusted. For this purpose, the phonemes were cut from *asa* and *asha* without the VC and CV transients. The original recordings of /*ʃ*/ and /*s*/ exhibited peak frequencies of 4.6 and 7.2 kHz, respectively. With Adobe Audition 3 (Adobe Systems Incorporated, San José, CA), the high-frequency noise bursts of /*ʃ*/ and /*s*/ were spectrally shaped with a graphic equalizer to two /*ʃ*/ with peak frequencies of 3 kHz (sh3) and 5 kHz (sh5) as well as two /*s*/ with peak frequencies of 6 kHz (s6) and 9 kHz (s9) (see Figure 2). The low-frequency components of /*ʃ*/ and /*s*/, that is, the spectra below ~1.5 kHz, and the transients were preserved for these stimuli and denoted by the suffix "LF" in the following descriptions.

Additional modifications were made to the stimuli to restrict the cues for /ʃ/ and /*s*/ to high frequencies. Specifically, the low frequency part of the phonemes were removed by high-pass filtering (Butterworth filter 8th order), and the VC and CV transients were replaced by the transients between /*a*/ and the stop consonant /*k*/ as taken from *aka*. The cutoff frequencies for high-pass filtering were 4.5, 6, 1.2, and 2.5 kHz for s6, s9, sh3, and sh5, respectively. This version of filtered (i.e., no) low-frequency information was denoted with the suffix "nLF" (i.e., without low-frequency cues).

In summary, two stimulus sets were available for the recognition task: one set with low-frequency information: asha3-LF, asha5-LF, asa6-LF, and asa9-LF, and one set without low-frequency information: asha3-nLF, asha5-nLF, asa6-nLF, and asa9-nLF. The stimuli *ada*, *afa*, *aha*, *aka*, and *ama* used in the recognition task as alternatives to *asa* and *asha* were the same when using LF or nLF. The detection test stimuli were created by cutting the /ʃ/ and /*s*/ consonants from the two sets without VC and CV transients. The two stimulus sets for the detection test are therefore as follows: (1) sh3-LF, sh5-LF, s6-LF, and s9-LF and (2) sh3-nLF, sh5-nLF, s6-nLF, and s9-nLF.

Calibration

Calibration was performed at the position of the listener by using a stationary white noise. The root-mean-square (RMS) levels of the stimuli were adjusted to be the same as the RMS level of the white noise. For this purpose, the RMS levels of the stimuli were calculated over the time duration of the consonants that were the whole stimuli in the detection test, but only during the middle phonemes of the VCVs used in the recognition test. In the following, these levels are referred to as "dB consonant level" (dB CL).

DT

To measure DTs, both stimulus sets for the detection test were presented in a range between 0 and 75 dB CL. The starting presentation level was 0 dB CL for all stimuli. The DT was determined by manual operation by the audiologist, using a modified version of the Hughson–Westlake procedure (Carhart and Jerger, 1959). The DT was measured using an ascending presentation method in two step sizes, 6 and 2 dB. First, the level was increased in steps of 6 dB until the stimulus was detected. Then, the level was decreased by 6 dB and increased in steps of 2 dB until the stimulus was detected again. This second phase was repeated until the stimulus was detected twice at the same level, and this final level was recorded as the participant's DT.

RT

The recognition test was designed to measure RTs for *asa* and *asha* in the range between 0 and 75 dB CL. The four target VCVs (two versions of *asa* and *asha*) in the recognition test were controlled individually, interleaved, and adaptive according to the participant's answers. The adaptive control principle was based on a 1-up-2-down staircase procedure to reach a correct response rate of 70.7% (Levitt, 1971). As a starting point, the presentation level of each VCV was set to 3 dB above the respective DT. Each stimulus was presented 10 times. The additional VCVs, *ada*, *afa*, *aha*, *aka*, and *ama*, were presented up to eight times randomly between the *asa* and *asha* stimuli and at random levels in the range of 30–82 dB SPL (RMS levels calculated over the whole duration of the VCVs).

The RT was calculated by averaging the last three presentation levels of each stimulus. An averaging over a certain number of reversals was abandoned because of the unpredictable number of reversals within ten presentations per stimulus. Once an *asa* or *asha* stimulus was presented twice at the minimum or the maximum level, that is, 0 or 75 dB CL, the adaptive procedure was terminated for this stimulus and the threshold was set to the respective minimum or maximum level. The number of presentations of the other VCVs varied

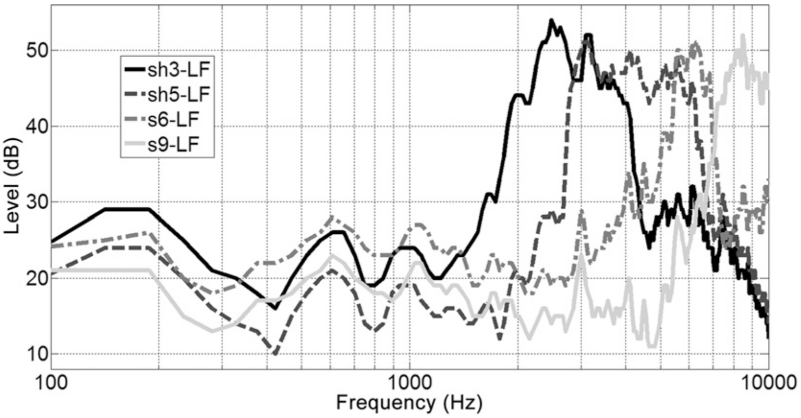

Figure 2. Spectra of the spectrally shaped /ʃ/ (sh3-LF and sh5-LF) and /*s*/ (s6-LF and s9-LF).

and was dependent on the RT determination for *asa* and *asha*. When all RTs were established or set to minimum or maximum level, the test stopped. For both stimulus sets, LF and nLF, the RTs were established and the confusions between the VCVs were recorded.

Seven consonant buttons labeled *d*, *f*, *h*, *k*, *m*, *s*, and *sch* (German for sh), as well as buttons for "repeat" ("Wiederholung" in German) and "?," were displayed on the touchscreen. Once the participant was ready and fully instructed, the test was started by the audiologist. The test interrupted automatically after a single VCV presentation and continued after a short break of ~0.5 sec when the participant clicked one of the response buttons. The participants were instructed to click the button for the corresponding consonant that they perceived between the two /*a*/ vowels. They were further instructed to click on the repeat button if self-created noise (e.g., body noise) occurred during stimulus presentation or when they were inattentive. In this case, the last presentation of a VCV stimulus was repeated. The "?" button could be clicked when the participants recognized a consonant that was not shown on the response list or when the presentation level of the VCV was too low. This unforced-choice method (Kaernbach, 2001) was used to increase the test efficiency, while reducing guessing.

FF

FF was carried out with a standard audiometry device (Unity 2, Siemens Audiologische Technik GmbH, Erlangen, Germany). The stimuli were pulsed warble tones at 0.5, 1, 2, 3, 4, 6, and 8 kHz. A measurement range from 0 to 100 dB HL was available for all frequencies, except for 8 kHz, where the measurement range was limited to 85 dB HL. The hearing thresholds were determined with the Hughson–Westlake procedure using a step size of 5 dB. The ascending procedure was started at a level of 20 dB HL or 20 dB below the previously measured frequency.

Fitting of Test BTEs

Two different pairs of BTE were provided to the participants, either Ambra micro P (n = 28) or Ambra SP (n = 3) (Phonak, Stäfa, Switzerland). Phonak Ambra SP was used for two participants who owned high-power HAs and one participant without a private HA but with a hearing loss of ≥90 dB from 3 kHz upward. They are labeled as "Test BTEs" in Figures 3, 5, and 6, and in Table 3. Both BTE models were equipped with NLFC. The BTEs were fitted to the individual earmolds of the participants, if available and applicable (n = 10). In all other cases, slim tubes with domes were used (n = 21). The BTEs were programmed by an audiologist using Phonak Target 2.1 fitting software. The BTEs' gain and output levels were calculated by the fitting prescription "Adaptive Phonak Digital," and the "Basic Tuning" parameter "Gain level" was set to 80%. This presetting was based on the audiometric thresholds of AudiogramDirect, an in-situ tone audiometric tool built into the fitting software. No further changes were made to the gain, output, or any adaptive parameter. The fitting parameters were copied to two manual programs. One of the programs, the "default" setting, remained unchanged. For the other manual program, the NLFC parameter was set to "too strong." The setting "too strong" was established according to the following procedure: the word "Mississippi," spoken by a recorded female, was repeatedly presented via one loudspeaker at a level of 50 dBA. The task of the subjects was to indicate as soon as Mississippi sounded lisping while the strength of NLFC was increased. This procedure was similar to determining the pure-tone threshold when level changes are replaced by changes in strength of NLFC.

Experimental Conditions

As visualized in Figure 3, group A was tested without HAs and group B was tested with their own HAs. Both groups were also tested using the test BTE with NLFC set to "default" and "too strong." Figure 3 also shows the different tests applied under the respective conditions. Testing occurred in three measurement sessions with durations of ~90 min each. The fitting of the test BTE was carried out for all participants in session one. The remaining tests, that is, PPTs and FF, were carried out in a randomized order in sessions two and three. Because of time constraints, free-field thresholds were determined only for 5 participants of group A and 18 participants of group B.

All measurements were run in the same sound attenuating booth at the Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, Oldenburg, Germany, using a spectrally calibrated audio

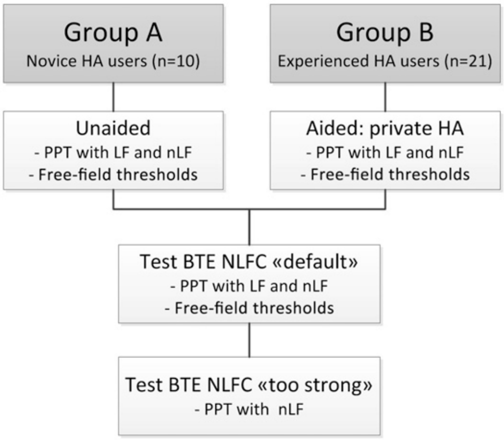

Figure 3. Overview of the tests that were carried out for each test condition and for both participant groups.

system. The sounds were presented via one loudspeaker (Genelec 8030A, Iisalmi, Finland) placed at 0° azimuth and at a distance of 1.5 m from the participants.

With this experimental design, the following characteristics of the PPT were investigated:

1. Relation between DT and RT and frequency-specific audibility as measured by conventional warble-tone, FF for stimulus sets nLF and LF.
2. Sensitivity of DT and RT to HA amplification, that is, a comparison between aided and unaided conditions for stimulus sets nLF and LF.
3. Sensitivity of DT and RT to differences in amplification for aided hearing, that is, a comparison between private and test HAs for stimulus sets nLF and LF.
4. Sensitivity of DT and RT to different settings of NLFC.

Statistica 10 (StatSoft Inc., Tulsa, OK) was used for all statistical analyses. Not all of the DT and RT data that were analyzed with the analysis of variances (ANOVAs) were normally distributed. However, the residuals showed a constant and independent variance. Therefore, ANOVA was performed for all DT and RT data.

The study was approved by the Ethics Committee of Oldenburg University.

RESULTS

Relation of DT and RT to Frequency-Specific Audibility in FF

The results of DTs and RTs were compared to the respective audiometric thresholds in FF for all available conditions (unaided, aided: private HA, test BTE NLFC "default") together. This aggregation resulted in 46 data points for DTs and RTs and all stimuli, respectively. The center frequencies of the consonants tested were used as corresponding FF frequency. If the center frequency did not match a specific FF frequency, the nearest FF frequency was chosen (i.e., 4 kHz for sh5 and 8 kHz for s9). Figure 4 shows the results for DTs using the stimuli with LF on the left side and nLF on the right side. For nLF, DTs tended to be higher than FF thresholds for low levels, where most of the thresholds for /ʃ/ are located. On the other hand, DTs tended to be lower than FF thresholds for high levels, where most of the thresholds for /*s*/ are located. This effect was clearly more prominent for LF stimuli.

Spearman's rank correlations to FF are given in Table 2 for DT as well as RT and the respective audiometric thresholds for both stimulus sets (nLF and LF). All correlations were positive and significant. The two /ʃ/ stimuli (sh3 and sh5) correlated with the respective FF frequency for both stimulus sets by approximately the same amount and the two /*s*/ stimuli (s6 and s9) were strongly correlated with the respective FF frequencies for stimulus set nLF. The /*s*/ stimuli of stimulus set LF were only weakly correlated to FF, as also shown in Figure 4. Table 2 also indicates that the results for both stimulus sets, nLF and LF, were significantly different for both /*s*/ stimuli (Wilcoxon matched pairs test using Bonferroni correction) with nLF showing higher thresholds than LF.

Sensitivity of DT and RT to HA Amplification

DTs and RTs of group A (n = 10) measured in the unaided and aided conditions with test BTE for all stimuli and both stimulus sets are shown in Figure 5. All thresholds measured in the aided condition were lower than those measured in the unaided condition. The differences between unaided and aided were larger for the stimulus set nLF than for the stimulus set LF. An increase in both DTs and RTs as a function of center frequency was observed only for stimulus set nLF. In addition, median RTs were similar or lower than the respective DTs for both stimulus sets.

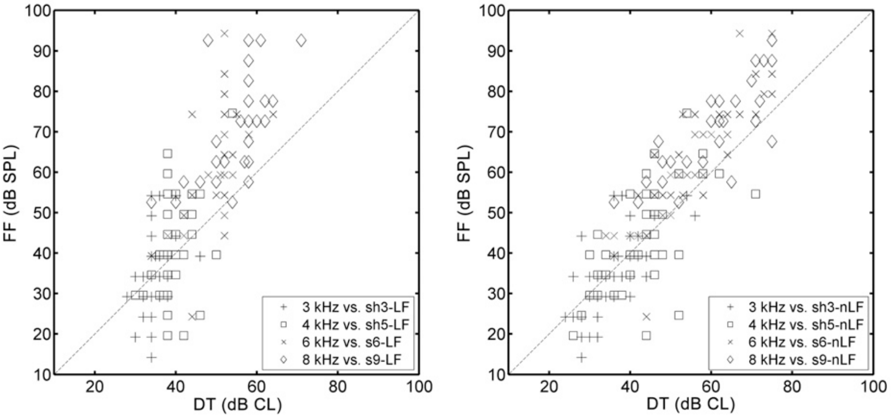

Figure 4. Scatter plot for DT and FF derived with the stimuli LF (left) and nLF (right). Symbols denote the respective frequency region between 3 and 9 kHz.

Table 2. Spearman Rank Correlations (*r*) and Wilcoxon Matched Pairs Results of DT and RT at the Respective FF Frequencies

		Rank Correlations		Wilcoxon
FF Frequency	DT and RT Stimuli	*r* (LF vs. FF)	*r* (nLF vs. FF)	*p* (LF vs. nLF)
3 kHz	DT: sh3	0.714	0.728	0.567
	RT: asha3	0.629	0.568	0.130
4 kHz	DT: sh5	0.634	0.619	0.034
	RT: asha5	0.588	0.552	0.011
6 kHz	DT: s6	0.590	0.905	0.001*
	RT: asa6	0.401	0.674	0.000*
8 kHz	DT: s9	0.550	0.907	0.000*
	RT: asa9	0.316	0.561	0.000*

Notes: n = 46. *Significant differences in the Wilcoxon test are indicated for $p \leq 0.006$ due to Bonferroni correction.

The ANOVA (see Table 3, second column) revealed significant differences between the tests (DT and RT) for the sets (nLF and LF), the HA condition (unaided and test BTE NLFC "default"), and the stimuli. Significant interactions were found for stimulus set and the HA condition, test and stimuli, and set and stimuli. Post hoc *t*-tests for paired samples with Bonferroni correction revealed significantly lower RTs and DTs for aided than unaided for both sets and stimuli except for s6-LF, s9-LF, asa6-LF, asa6-nLF, and asa9-nLF.

Sensitivity of DT and RT to Differences in Amplification

DTs and RTs of group B (n = 21) measured in the aided condition with private HA and with the test BTE in NLFC setting "default" are shown in Figure 6. Thresholds were similar in both aided conditions for stimulus set LF, whereas a large difference between both amplifications was especially observed for s9-nLF and asa9-nLF. In addition, the dependency of the thresholds on frequency was more pronounced for the stimulus set nLF.

The ANOVA (see Table 3, third column) revealed significant differences between tests (DT and RT), sets (nLF and LF), HA conditions (private HA and test BTE NLFC "default"), as well as stimuli. Significant interactions were found for test and set, HA condition and set, test and stimuli, set and stimuli, HA condition and stimuli, test and set and stimuli, and set and HA condition and stimuli. Post hoc *t*-tests for paired samples with Bonferroni correction revealed significantly lower

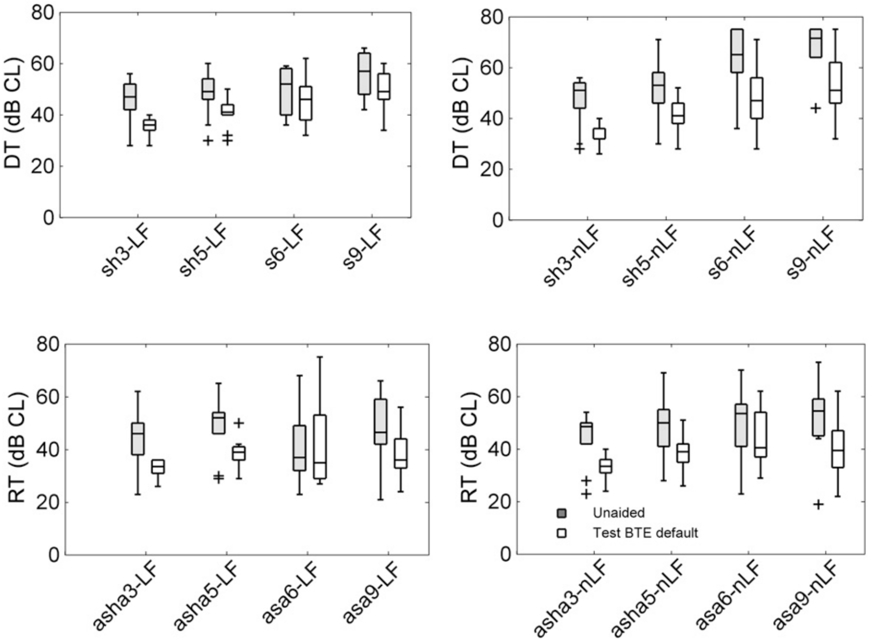

Figure 5. DTs (top) and RTs (bottom) for group A (n = 10) measured in the unaided condition and with the test BTE in NLFC setting "default" for stimulus set LF (left) and nLF (right). Boxes represent the interquartile range and are subdivided by the median. Whiskers indicate maximum and minimum values within 1.5 times the interquartile range.

Table 3. ANOVA Main Effects and Interactions: Test (DT and RT), Set (LF and nLF), and HA Condition (Unaided, Private HA, Test BTE NLFC "Default," and Test BTE NLFC "Too Strong")

	HA		NFLC
Variable	Unaided–Test BTE "Default"	Private HA–Test BTE "Default"	"Default"–"Too Strong"
TEST (DT–RT)	$F_{(1,9)} = 51.301$, $p < 0.001$*	$F_{(1,20)} = 21.34$, $p < 0.001$*	$F_{(1,30)} = 26.782$, $p < 0.001$*
SET (LF–nLF)	$F_{(1,9)} = 11.201$, $p = 0.009$*	$F_{(1,20)} = 51.12$, $p < 0.001$*	—
HA	$F_{(1,9)} = 48.924$, $p < 0.001$*	$F_{(1,20)} = 4.587$, $p = 0.045$*	—
STIMULI (Sh3, Sh5, S6, S9, Asha3, Asha5, Asa6, Asa9)	$F_{(3,27)} = 30.412$, $p < 0.001$*	$F_{(3,60)} = 97.118$, $p < 0.001$*	$F_{(3,90)} = 137.471$, $p < 0.001$*
TEST × SET	$F_{(1,9)} = 1.333$, $p = 0.278$	$F_{(1,20)} = 27.256$, $p < 0.001$*	—
TEST × HA	$F_{(1,9)} = 0.846$, $p = 0.382$	$F_{(1,20)} = 2.347$, $p = 0.141$	—
SET × HA	$F_{(1,9)} = 5.966$, $p = 0.037$*	$F_{(1,20)} = 9.193$, $p = 0.007$*	—
TEST × STIMULI	$F_{(3,27)} = 14.329$, $p < 0.001$*	$F_{(3,60)} = 11.138$, $p < 0.001$*	$F_{(3,90)} = 31.139$, $p < 0.001$*
SET × STIMULI	$F_{(3,27)} = 12.247$, $p < 0.001$*	$F_{(3,60)} = 29.534$, $p < 0.001$*	—
HA × STIMULI	$F_{(3,27)} = 2.282$, $p = 0.102$	$F_{(3,60)} = 8.83$, $p < 0.001$*	—
TEST × SET × HA	$F_{(1,9)} = 1.134$, $p = 0.315$	$F_{(1,20)} = 1.102$, $p = 0.306$	—
TEST × SET × STIMULI	$F_{(3,27)} = 0.564$, $p = 0.643$	$F_{(3,60)} = 6.454$, $p = 0.001$*	—
TEST × HA × STIMULI	$F_{(3,27)} = 2.422$, $p = 0.088$	$F_{(3,60)} = 0.06$, $p = 0.980$	—
SET × HA × STIMULI	$F_{(3,27)} = 1.007$, $p = 0.405$	$F_{(3,60)} = 7.652$, $p < 0.001$*	—
TEST × SET × HA × STIMULI	$F_{(3,27)} = 0.028$, $p = 0.993$	$F_{(3,60)} = 0.108$, $p = 0.955$	—
NLFC	—	—	$F_{(1,30)} = 4.541$, $p = 0.041$*
TEST × NLFC	—	—	$F_{(1,30)} = 98.874$, $p < 0.001$*
NLFC × STIMULI	—	—	$F_{(3,90)} = 7.354$, $p < 0.001$*
TEST × NLFC × STIMULI	—	—	$F_{(3,90)} = 41.778$, $p < 0.001$*

Note: *Indicates significant differences.

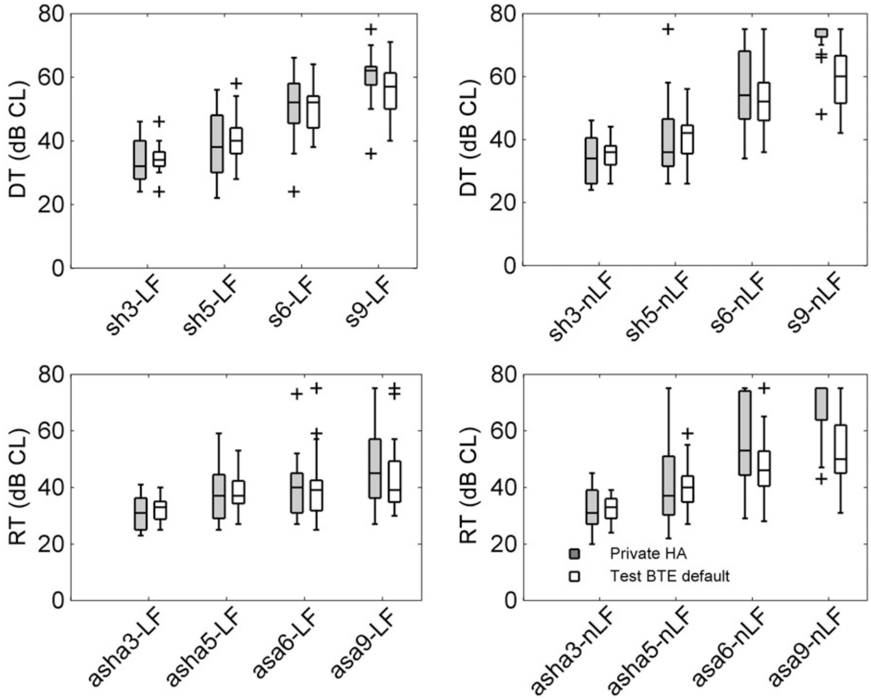

Figure 6. DTs (top) and RTs (bottom) for group B (n = 21) measured with participants' private HA and the test BTE in NLFC setting "default" for stimulus set LF (left) and nLF (right). Boxes represent the interquartile range and are subdivided by the median. Whiskers indicate maximum and minimum values within 1.5 times the interquartile range.

RTs and DTs for the test BTE in NLFC setting default than with private HA for s9-nLF and asa9-nLF.

Sensitivity of DT and RT to Different Settings of NLFC

DTs and RTs of all participants (n = 31) measured with the test BTE in the NLFC settings "default" and "too strong" with stimulus set nLF are shown in Figure 7. Most of the DTs were on a similar level, except for the higher DTs for stimulus /*s*/ for the NLFC setting "default." Similar results were observed for most of the RTs. The largest differences between RTs and DTs appeared for the /*s*/ stimuli in the NLFC setting "too strong," where many of the RTs were at maximum (75 dB CL).

The ANOVA (see Table 3, fourth column) revealed significant differences between the tests (DT and RT), the stimuli, and the HA conditions (test BTE NLFC setting "default" versus "too strong"). Significant interactions were found for test and NLFC setting, test and stimuli, NLFC setting and stimuli, and test and NLFC setting and stimuli. The post hoc *t*-test for paired samples with Bonferroni correction revealed significantly lower DTs but significantly higher RTs for NLFC setting "too strong" for the /*s*/ stimuli.

DISCUSSION

Relationship between DT and RT and FF

The present study applied a detection test and a recognition test included in the PPT with two different stimulus sets in different amplification conditions. The analysis focused on the capability of the PPT to measure the effect of mid-to-high frequency audibility on perception with and without HAs, with different HAs, and with different NLFC settings. Previous studies have shown that speech stimuli selected to avoid perception by low-frequency speech cues are more likely to reveal a benefit of amplification (e.g., Stelmachowicz et al, 2002; Turner and Henry, 2002). Therefore, in this study, two stimulus sets were created, one set (LF) with low-to-mid frequency cues (transients, low-frequency noise) and one set (nLF) without these cues. As intended by the stimulus modifications, significantly higher correlations were observed between FF thresholds and the DTs and RTs for the stimulus set nLF than between the free-field thresholds and the DTs and RTs for the stimulus set LF. These findings indicate that high-pass filtering of the consonants and replacement of VC and CV transient information increases frequency-specific audibility of DTs and RTs. For example, s6-nLF was detected due to the audibility of the 6 kHz spectral peak, and this cue was also used for the recognition of this phoneme.

When comparing DTs to RTs, lower thresholds for the latter task were observed in some cases. These results seem to be implausible, since detection should occur at a lower level than recognition (Miller et al, 1951). Elliott et al (1981) observed approximately 8 dB higher RT than DT for *ba*, *da*, and *ga* as expected, but could not rule out that the low thresholds for DT were dominated by the detection of /*a*/ instead of /*b*/, /*d*/, and /*g*/. The higher thresholds of DT compared to RT in this study might be an effect of the different procedures used for determining the thresholds. For DT, a modified version of the Hughson–Westlake procedure was used, which results in a threshold between 50% and 100% correct responses on the psychometric function (Marshall and Jesteadt, 1986). For RT, a 1-up-2-down staircase procedure was used, which results in a threshold at 71% correct responses on the psychometric function (Levitt, 1971). Marshall and Jesteadt (1986) found a 6.5 dB higher threshold for the Hughson–Westlake procedure used in pure-tone audiometry with a step size of 5 dB than for a two-interval forced-choice tasks that adapted to 71%. They related the difference to the defined observation intervals, the smaller step size, and the control for response bias in the two-interval forced-choice task. Another factor to consider is the number of alternatives when measuring RT. In this study, seven response alternatives were given on the touch screen. Miller et al (1951) reported that RTs increase with a larger number of response alternatives. On the other hand, Allen and Li (2009) reported that confusion alternatives are defined by speech cue similarities. For example, two phonemes differing in unique speech cues (e.g., /*s*/ and /*m*/) might not serve as confusion alternatives. Because the PPT was spoken by a German native speaker and since the German language contains only two unvoiced phonemes with a noise burst and a unique peak frequency in the high frequencies (/*s*/ and /*ʃ*/), adding more response alternatives would not necessarily increase RT. A third factor might be that the RT was calculated taking only the last three presentation levels into account. This might result in a different threshold than 71% correct for single listeners but should converge on average to this value. Taking the procedural differences between DT and RT into account, lower DT than RT seems to be possible but would result in equal differences for all stimuli. However, Figures 5–7 reveal that the differences between DT and RT are dependent on the stimuli and hearing instrument setting. This observation leads to the conclusion that the observed differences cannot be related to systematic procedural differences. Instead, the differences might indicate that more hints than just consonant audibility are used for VCV recognition. One possible hint is the alternatives in the

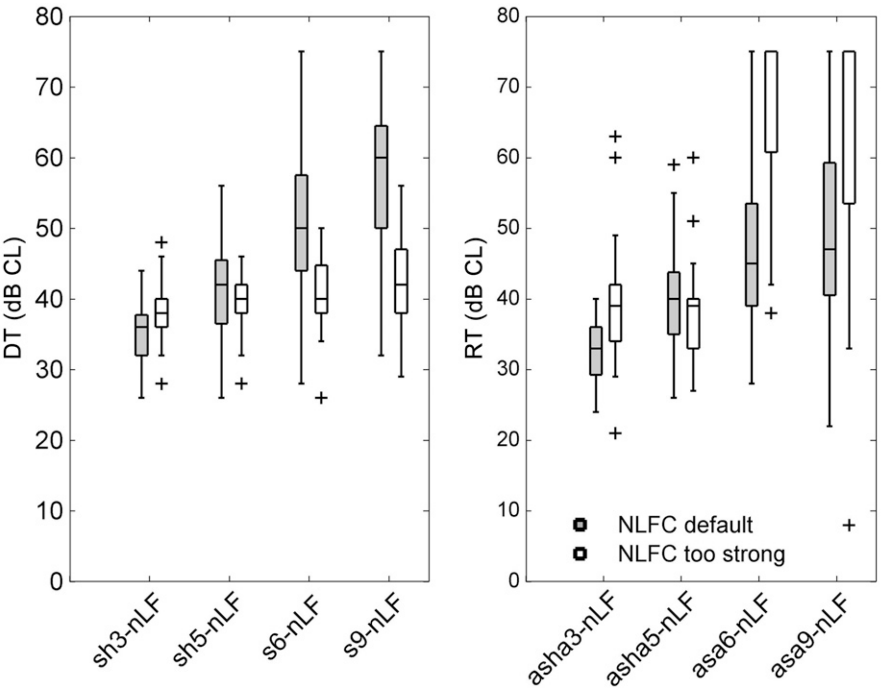

Figure 7. DTs (left panel) and RTs (right panel) of all participants (n = 31) derived with stimulus set nLF in the NLFC setting "default" and NLFC setting "too strong."

recognition test. Listeners may be more likely to select *asa* or *asha* in those cases without consonant audibility because they did not perceive any of the other consonants that contain more low-frequency energy.

Sensitivity of DT and RT to HA Amplification

Many hearing-impaired individuals exhibit elevated audiometric thresholds in the high-frequency region (Bisgaard et al, 2010). Consequently, adjusting the HA gain in this region is a very crucial task within the fitting process. Many studies have investigated the impact of high-frequency audibility on speech perception, with contradicting results. Some studies failed to show any benefit (Rankovic, 1991; Ching et al, 1998; Hogan and Turner, 1998), whereas in other studies, audible high-frequency speech cues improved speech perception (Stelmachowicz et al, 2001; Turner and Henry, 2002; Glista et al, 2009; Wolfe et al, 2011). In the present study, a subgroup of participants was tested under conditions with and without HA. The hearing loss of these participants was more dominant in the high-frequency than in the low-frequency region and the HA-applied, frequency-specific gain set accordingly. The results for DTs and RTs were dependent on the stimulus set. With stimulus set LF, only a small benefit for aided compared to unaided could be observed. With stimulus set nLF, the benefit in the aided condition compared to the unaided condition increased with the peak frequency of the stimulus, that is, both stimulus sets could distinguish between aided and unaided test conditions but the differences were more distinctive for stimulus set nLF than for stimulus set LF. These findings indicate that tailoring the stimuli to the high-frequency region makes it possible to establish the effect of amplification in this region on speech perception. This might be important when comparing small differences in high-frequency amplification.

Sensitivity of DT and RT to Differences in Amplification

The results for the novice HA users were replicated by the experienced users when comparing their own HA to the test BTE. The data indicate that high-pass filtering and transient replacement increased the sensitivity to frequency-specific audibility and could distinguish between both HAs. The differences between the two HAs at high frequencies could be generated by higher gain settings for the test BTE compared to the private HA or by NLFC, which was only available in the test BTE. The NLFC algorithm shifts part of the high-frequency information to lower frequencies. For these kind of algorithms, it is especially important to be as frequency specific as possible (Glista et al, 2009; Wolfe et al, 2011). The combination of amplification and information shift to lower frequencies might improve DTs and RTs for the stimuli s9-nLF and asa9-nLF besides pure amplification, which is limited to 50 dB HL (Ricketts et al, 2008).

Sensitivity of DT and RT to Different Settings of NLFC

The effect of NLFC on detection and recognition was investigated by Glista et al (2009) and Wolfe et al (2011). They reported high individual variability of initial recognition and detection performance and improvements after six months of acclimatization (Wolfe et al, 2011). These studies used different tests and stimulus materials for detection and recognition, resulting in difficult comparisons of the outcomes and the interactions. In the present study, two NLFC settings, "default" and "too strong," were tested with the same stimulus sets. The results showed that high compression of the consonant information to lower frequencies as compared to a moderate compression reduced DTs for s6-nLF and s9-nLF because of a reduced hearing loss and more available gain in the low-frequency region. In contrast, RTs for asa6-nLF and asa9-nLF increased significantly for the NLFC setting "too strong." This increase was caused by more confusions of *asa* with *asha* than with NLFC "default." Very strong NLFC settings move energy peaks of /*s*/ to lower frequencies where the energy peak frequency of /ʃ/ is normally located.

This study did not attempt to identify the optimal HA and NLFC setting for an individual participant. Instead, the goal was to select the most sensitive stimulus set to measure unaided, aided, and NLFC performance on detection and recognition tasks. The findings from this study suggest that measuring DTs and RTs with the stimulus set nLF is useful to quantify the effect of HAs and of NLFC on high-frequency speech cues. They also indicate that both tests are necessary for the evaluation of the NLFC settings, because they reveal audibility as well as recognition ability especially for speech modification algorithms.

Although the advantage of stimuli modifications for the evaluation of HA amplification was shown, several questions still remain. The experiments conducted in our study did not allow for any acclimatization of the participants to increased high-frequency gain or NLFC. Since Wolfe et al (2011) reported improvements in recognition but not in detection after an acclimatization period of six months, further studies of the impact of acclimatization to NLFC on DTs and RTs in the PPT as well as the contradictory effects of strong setting of NLRC to DTs and RTs are required.

Future research should also focus on speech perception near threshold and the impact of audibility-unrelated hints for recognition, which might resolve the reasons for lower RTs compared to the respective DTs observed in several cases. The results might also clarify recognition results for other speech tests under different amplification conditions.

In summary, the PPT seems to be a suitable tool to support HA fitting in the high-frequency region with respect to gain and NLFC settings. Nevertheless, the informational value is limited to the tailored frequency domain. For broader applicability in the improvement of HA fittings, other suitable phonemes should be selected, included in DT and RT testing, and evaluated.

Acknowledgments. The authors thank Stephanie Maaßen for her support on data collection and Trecia Brown, Gurjit Singh, and three anonymous reviewers for their comments. Language corrections on an earlier version of the article were carried out by www.stels-ol.de.

REFERENCES

Allen JB, Li F. (2009) Speech perception and cochlear signal processing. *IEEE Signal Process Mag* 26(4):73–77.

Arweiler-Harbeck D, Janeschik S, Lang S, Bagus H. (2011) Suitability of auditory speech sound evaluation (A§E®) in German cochlear implant patients. *Eur Arch Otorhinolaryngol* 268(9):1259–1266.

Beasley WC, Rosenwasser H. (1950) Determining factors in composing and analyzing speech-hearing tests. *Laryngoscope* 60(7): 658–679.

Bisgaard N, Vlaming MS, Dahlquist M. (2010) Standard audiograms for the IEC 60118-15 measurement procedure. *Trends Amplif* 14(2):113–120.

Boothroyd A, Medwetsky L. (1992) Spectral distribution of /s/ and the frequency response of hearing aids. *Ear Hear* 13(3):150–157.

Boretzki M, Kegel A. (2009) *The benefits of nonlinear frequency compression for people with mild hearing loss. AudiologyOnline*. http://www.audiologyonline.com/articles/benefits-nonlinear-frequency-compression-for-869-869.

Boretzki M, Schmitt N, Kegel A, Krueger H, Rehmann J, Eichhorn F, Meisenbacher K, Raether J. (2011) Future directions in evaluating frequency compression. In: *A Sound Foundation through Early Amplification. Proceedings of the International Pediatric Audiology Conference*, November 8–10, 2010 (pp. 201–203). Chicago, IL: Phonak AG.

Carhart R, Jerger JF. (1959) Preferred method for clinical determination of pure-tone thresholds. *J Speech Hear Disord* 24(4):330–345.

Ching TY, Dillon H, Byrne D. (1998) Speech recognition of hearing-impaired listeners: predictions from audibility and the limited role of high-frequency amplification. *J Acoust Soc Am* 103(2): 1128–1140.

Ching TY, Dillon H, Katsch R, Byrne D. (2001) Maximizing effective audibility in hearing aid fitting. *Ear Hear* 22(3):212–224.

Dillier N, Spillmann T. (1992) Deutsche Version der Minimal Auditory Capability (MAC)-Test-Batterie: Anwendungen bei Hörgeräte- und CI-Trägern mit und ohne Störlärm. *Moderne Verfahren der Sprachaudiometrie*. Heidelberg, Germany: Median-Verlag, 238–263.

Elliott LL, Longinotti C, Clifton L, Meyer D. (1981) Detection and identification thresholds for consonant-vowel syllables. *Percept Psychophys* 30(5):411–416.

Foster JR, Haggard MP. (1987) The four alternative auditory feature test (FAAF)—linguistic and psychometric properties of the material with normative data in noise. *Br J Audiol* 21(3):165–174.

Glista D, Scollie S. (2012) Development and evaluation of an English language measure of detection of word-final plurality markers: the University of Western Ontario Plurals Test. *Am J Audiol* 21(1):76–81.

Glista D, Scollie S, Bagatto M, Seewald R, Parsa V, Johnson A. (2009) Evaluation of nonlinear frequency compression: clinical outcomes. *Int J Audiol* 48(9):632–644.

Glista D, Scollie S, Sulkers J. (2012) Perceptual acclimatization post nonlinear frequency compression hearing aid fitting in older children. *J Speech Lang Hear Res* 55(6):1765–1787.

Govaerts PJ, Daemers K, Yperman M, De Beukelaer C, De Saegher G, De Ceulaer G. (2006) Auditory speech sounds evaluation (A§E): a new test to assess detection, discrimination and identification in hearing impairment. *Cochlear Implants Int* 7(2): 92–106.

Heller O. (1992) Frequenzspezifische Sprachaudiometrie. In: Kollmeier B, Ed., *Moderne Verfahren der Sprachaudiometrie*, 264–277. Heidelberg, Germany: Median-Verlag.

Hogan CA, Turner CW. (1998) High-frequency audibility: benefits for hearing-impaired listeners. *J Acoust Soc Am* 104(1):432–441.

Hornsby BW, Johnson EE, Picou E. (2011) Effects of degree and configuration of hearing loss on the contribution of high- and low-frequency speech information to bilateral speech understanding. *Ear Hear* 32(5):543–555.

Kaernbach C. (2001) Adaptive threshold estimation with unforced-choice tasks. *Percept Psychophys* 63(8):1377–1388.

Kimlinger C, McCreery R, Lewis D. (2015) High-frequency audibility: the effects of audiometric configuration, stimulus type, and device. *J Am Acad Audiol* 26(2):128–137.

Knoblach W. (1992) Methodischer Ansatz für einen hochfrequenzspezifischen Sprachtest. In: Kollmeier B, Ed., *Moderne Verfahren der Sprachaudiometrie*, 278–286. Heidelberg, Germany: Median-Verlag.

Kuk F, Lau CC, Korhonen P, Crose B, Peeters H, Keenan D. (2010) Development of the ORCA nonsense syllable test. *Ear Hear* 31(6): 779–795.

Levitt H. (1971) Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49(2):467–477.

Marshall L, Jesteadt W. (1986) Comparison of pure-tone audibility thresholds obtained with audiological and two-interval forced-choice procedures. *J Speech Hear Res* 29(1):82–91.

McCreery RW, Brennan MA, Hoover B, Kopun J, Stelmachowicz PG. (2013) Maximizing audibility and speech recognition with nonlinear frequency compression by estimating audible bandwidth. *Ear Hear* 34(2):e24–e27.

Miller GA, Heise GA, Lichten W. (1951) The intelligibility of speech as a function of the context of the test materials. *J Exp Psychol* 41(5):329–335.

Newman RS, Clouse SA, Burnham JL. (2001) The perceptual consequences of within-talker variability in fricative production. *J Acoust Soc Am* 109(3):1181–1196.

Owens E, Kessler DK, Raggio MW, Schubert ED. (1985) Analysis and revision of the minimal auditory capabilities (MAC) battery. *Ear Hear* 6(6):280–290.

Owens E, Schubert ED. (1977) Development of the California Consonant Test. *J Speech Hear Res* 20(3):463–474.

Pascoe DP. (1975) Frequency responses of hearing aids and their effects on the speech perception of hearing-impaired subjects. *Ann Otol Rhinol Laryngol* 84(5 pt 2):(Suppl 23):1–40.

Pavlovic CV. (1987) Derivation of primary parameters and procedures for use in speech intelligibility predictions. *J Acoust Soc Am* 82(2):413–422.

Pittman AL. (2008) Short-term word-learning rate in children with normal hearing and children with hearing loss in limited and extended high-frequency bandwidths. *J Speech Lang Hear Res* 51(3):785–797.

Rankovic CM. (1991) An application of the articulation index to hearing aid fitting. *J Speech Hear Res* 34(2):391–402.

Ricketts TA, Dittberner AB, Johnson EE. (2008) High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *J Speech Lang Hear Res* 51(1):160–172.

Scollie S, Glista D, Tenhaaf J, Dunn A, Malandrino A, Keene K, Folkeard P. (2012) Stimuli and normative data for detection of Ling-6 sounds in hearing level. *Am J Audiol* 21(2):232–241.

Simpson A, Hersbach AA, McDermott HJ. (2005) Improvements in speech perception with an experimental nonlinear frequency compression hearing device. *Int J Audiol* 44(5):281–292.

Stelmachowicz PG, Pittman AL, Hoover BM, Lewis DE. (2001) Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults. *J Acoust Soc Am* 110(4): 2183–2190.

Stelmachowicz PG, Pittman AL, Hoover BM, Lewis DE. (2002) Aided perception of /s/ and /z/ by hearing-impaired children. *Ear Hear* 23(4):316–324.

Stock A. (1996) Über die Verständlichkeitsrelevanz von Frikativtransitionen. In: *PhD Thesis University Würzburg*. Hamburg, Germany: Dr. Kovač-Verlag.

Streeter LA, Nigro GN. (1979) The role of medial consonant transitions in word perception. *J Acoust Soc Am* 65(6):1533–1541.

Turner CW, Cummings KJ. (1999) Speech audibility for listeners with high-frequency hearing loss. *Am J Audiol* 8(1):47–56.

Turner CW, Henry BA. (2002) Benefits of amplification for speech recognition in background noise. *J Acoust Soc Am* 112(4): 1675–1680.

Wolfe J, John A, Schafer E, Nyffeler M, Boretzki M, Caraway T, Hudson M. (2011) Long-term effects of non-linear frequency compression for children with moderate hearing loss. *Int J Audiol* 50(6):396–404.