Auditory tests for characterizing hearing deficits: The BEAR test battery

Raul Sanchez-Lopez[a,*], Silje Grini Nielsen[a], Mouhamad El-Haj-Ali[b], Federica Bianchi, Michal Fereczkowski[a,b,c], Oscar M Cañete[a], Mengfan Wu[b,c], Tobias Neher[b,c], Torsten Dau[a,] and Sébastien Santurette[a,d]

[a] *Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark.*

[b] *Institute of Clinical Research, University of Southern Denmark, Odense, Denmark.*
[c] *Research Unit for Oto-Rhino-Laryngology, Odense University Hospital, Odense, Denmark*

[d] *Centre for Applied Audiology Research, Oticon A/S, Smørum, Denmark*

* Corresponding author: Raul Sanchez-Lopez, Hearing Systems Section, Department of Health Technology, Ørsteds Plads 352, R-111, Technical University of Denmark (DTU), 2800 Kgs. Lyngby, Denmark.

Email: rsalo@dtu.dk

Auditory tests for characterizing hearing deficits: The BEAR test battery

Introduction: The Better hEAring Rehabilitation (BEAR) project aims to provide a new clinical profiling tool – a test battery – for hearing loss characterization. Whereas the loss of sensitivity can be efficiently measured using pure-tone audiometry, the assessment of supra-threshold hearing deficits remains a challenge. In contrast to the classical 'attenuation-distortion' model, the proposed BEAR approach is based on the hypothesis that the hearing abilities of a given listener can be characterized along two dimensions reflecting independent types of perceptual deficits (distortions). A data-driven approach provided evidence for the existence of different auditory profiles with different degrees of distortions. Design: Eleven tests were included in a test battery, based on their clinical feasibility, time efficiency and related evidence from the literature. The tests were divided into six categories: audibility, speech perception, binaural processing abilities, loudness perception, spectro-temporal modulation sensitivity and spectro-temporal resolution. Study sample: Seventy-five listeners with symmetric, mild-to-severe sensorineural hearing loss were selected from a clinical population. Results: The analysis of the results showed interrelations among outcomes related to high-frequency processing and outcome measures related to low-frequency processing abilities. Conclusions: The results showed the ability of the tests to reveal differences among individuals and their potential use in clinical settings.

200 words

Keywords: Hearing loss, supra-threshold auditory deficits, auditory processing, psychoacoustics

Introduction

In current clinical practice, hearing loss (HL) is diagnosed mainly on the basis of pure-tone audiometry (ISO 8253-1, 2010). The audiogram helps differentiate between conductive and sensorineural hearing losses and can characterize the severity of the hearing loss from mild to profound. However, the pure-tone audiogram only assesses the sensitivity to simple sounds, which is not necessarily related to listening abilities at supra-threshold sound pressure levels (e.g. a person's ability to discriminate speech in noise).

Pure-tone audiometry is often complemented by speech audiometry (ISO 8253-3, 2012), which is a test of word recognition performance in quiet. Although this test can provide information about supra-threshold deficits (Gelfand, 2009), measurements of speech understanding in noise have been found more informative (Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004; Nilsson, Soli, & Sullivan, 1994). Since improving speech intelligibility is usually the main goal of successful hearing rehabilitation, several auditory factors affecting speech intelligibility in noise have been investigated (e.g. Glasberg & Moore, 1989; Houtgast & Festen, 2008; Strelcyk & Dau, 2009). Audibility (in conditions with fluctuating maskers), frequency selectivity (in conditions with stationary noise), and temporal processing acuity (in conditions with speech interferers), have been identified as important factors affecting speech reception thresholds in noise (e.g. Desloge, Reed, Braida, Perez, & D'Aquila, 2017; Johannesen, Pérez-González, Kalluri, Blanco, & Lopez-Poveda, 2016; Oxenham & Simonson, 2009; Rhebergen, Versfeld, & Dreschler, 2006). Thus, a hearing evaluation that goes beyond pure-tone sensitivity and speech intelligibility in quiet would be expected to provide a more accurate characterization of a listener's hearing deficits.

In Denmark, the Better hEAring Rehabilitation (BEAR) project was initiated with the aim of developing new diagnostic tests and hearing-aid compensation strategies for audiological practice. Although the assessment of individual hearing deficits can be complex, new evidence suggests that the perceptual consequences of a hearing loss can be characterized effectively by two types of hearing deficits, defined as "auditory distortions" (Sanchez-Lopez, Bianchi, Fereczkowski, Santurette, & Dau, 2018). By analysing the outcomes of two previous studies (Johannesen et al., 2016; Thorup et al., 2016) with a data-driven approach, Sanchez-Lopez et al. (2018) identified high-frequency hearing loss as the main predictor of one of the distortions, whereas the definition of the second type of distortion was inconclusive. The mixed results obtained from these analyses were most likely due to differences between the two studies in terms of hearing loss profiles and outcome measures. Here, a new dataset was therefore collected based on a heterogeneous group of listeners with audiometric hearing losses ranging from very mild to severe and with a large range of audiometric profiles. To that end, the most informative tests resulting from the analysis of Sanchez-Lopez et al. (2018) were included, together with additional auditory tests that had shown potential for hearing profiling in other previous studies. The tests included in the current study are referred to as the BEAR test battery.

The characterization of hearing deficits beyond the audiogram was considered in several earlier studies (e.g. Brungart, Sheffield, & Kubli, 2014; Rönnberg et al., 2016; Santurette & Dau, 2012; Saunders, Field, & Haggard, 1992; Vlaming et al., 2011). Among them, the HEARCOM project (Vlaming et al., 2011) proposed an extended hearing profile formed by the results of several behavioural tests. These tests targeted various auditory domains, such as audibility, loudness perception, speech perception,

binaural processing, and spectro-temporal resolution, as well as a test of cognitive abilities. Importantly, while the auditory domains considered in the BEAR test battery are similar to the ones considered in the HEARCOM project, the BEAR project aims to additionally classify the patients in subcategories and to create a link between hearing capacities and hearing-aid parameter settings.

The tests included in the BEAR test battery were chosen based on the following criteria: 1) There is evidence from the hearing research literature that the considered test is informative and reliable; 2) The outcomes of the test may be linked to a hearing-aid fitting strategy; 3) The outcome measures are easy to interpret and to explain to the patient; 4) The task is reasonably time-efficient or can be suitably modified to meet this requirement (e.g., by changing the test paradigm or developing an out-of-clinic solution); 5) The test implementation can be done with equipment available in clinics; 6) The tasks are not too demanding for patients and clinicians; 7) Tests with several outcome measures are prioritized, and 8) The tests are language independent.

The selected test battery included measures of audibility, loudness perception, speech perception, binaural processing abilities, spectro-temporal modulation (STM) sensitivity and spectro-temporal resolution. It was implemented and tested in normal-hearing (NH) and hearing-impaired (HI) listeners. The goals of the study were: 1) To collect reference data from a representative sample of HI listeners for each of the selected tests, 2) to analyse the test-retest reliability of these tests, 3) to analyse the relationships between the different outcome measures, and 4) to propose a version of the test battery that can be implemented in hearing clinics.

General methods

Participants and general setup

Seventy-five listeners (38 females) participated in the study, who were aged between 59 and 82 years (median: 71 years). Five participants were considered NH (PTA ≤ 25 dB HL). PTA was defined as the pure-tone average between 0.5, 1 and 2 kHz. The HI group consisted of 70 participants with symmetric sensorineural hearing losses. Symmetric sensorineural hearing loss was defined as an interaural difference (ID) ≤ 15 dB HL at frequencies below 8 kHz and ID ≤ 25 dB HL at 8 kHz and air-bone gap < 10 dB HL. The pure-tone audiograms of the participants are shown in Figure 1.

The participants were recruited from the BEAR database (Wolff et al., 2020) at Odense University Hospital (OUH), from the patient database at Bispebjerg Hospital (BBH), and from the database at the Hearing Systems Section at the Technical University of Denmark (DTU). The tests were performed in a double-walled sound-insulated booth. Most of the tests were implemented using a modular framework for psychoacoustic experiments (AFC; Ewert, 2013) and the stimuli were presented through headphones. The study was approved by the Science-Ethics Committee for the Capital Region of Denmark H-16036391. All participants gave written informed consent and received financial compensation for their participation.

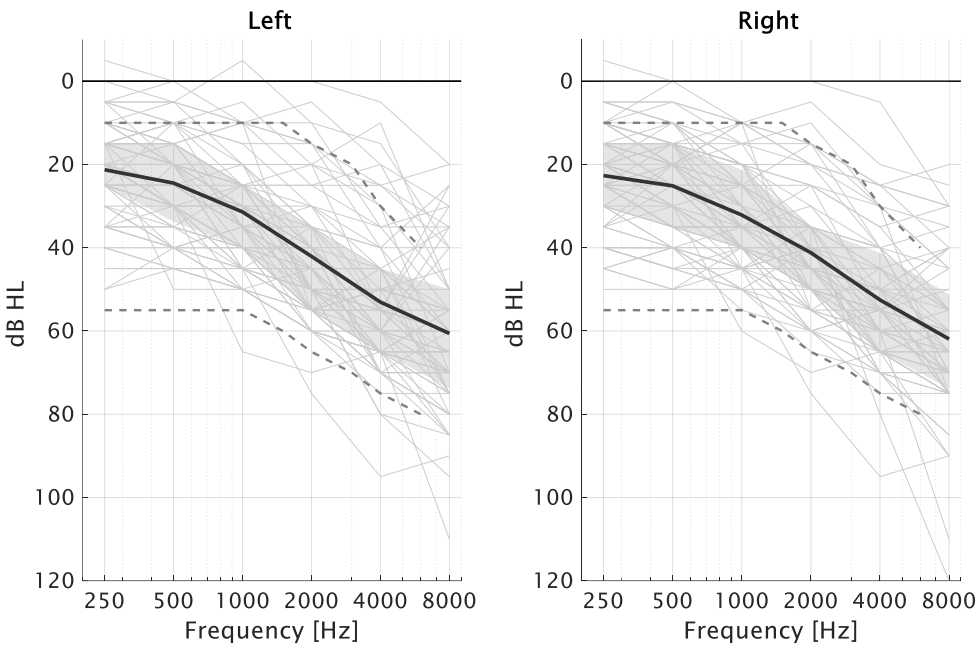

Figure 1: Audiograms of the 75 participants of the study together with the average for each ear (dark solid lines) and interquartile ranges (grey areas). The grey dashed lines correspond to the standard audiograms N1 and N4 from Bisgaard, Vlaming, & Dahlquist (2010).

Analysis of test reliability

The test-retest reliability of the test battery was assessed using intraclass correlation coefficients (ICC; Koo & Li, 2016) and the standard error of measurement (SEM; Stratford & Goldsmith, 1997). Test-retest measurements were performed with a subgroup consisting of seven HI and three NH participants for all tests of the test battery. The seven HI listeners had bilateral HL with a mean PTA of 31 dB HL. The retest session was conducted within four months after the first visit.

Overview of the test battery

The proposed tests are divided into six categories. Table 1 shows the tests and the corresponding auditory domains. The following sections present all tests individually, the experimental method and the summary statistics of the outcome measures presented

in Table 2. The dataset is publicly available in a Zenodo repository (Sanchez-Lopez et al., 2019). More details about the method can be found in the supplementary material and in the data repository.

Table 1: List of the tests included in the BEAR test battery and their corresponding auditory domains.

Test	***Test domain***
Pure-tone audiometry: *Fixed level frequency threshold (eAUD-HF)*	*Audibility*
Word recognition scores in quiet (WRS-4UFC): *Hearing in noise test (HINT)*	*Speech perception tests*
Maximum frequency for IPD detection (IPD_{fmax}) *Binaural pitch (Bpitch)* *Extended binaural audiometry in noise (eAUD-B)*	*Binaural processing abilities*
Adaptive categorical loudness scaling (ACALOS)	*Loudness perception*
Fast spectro-temporal modulation sensitivity (fSTM)	*Spectro-temporal modulation*
Extended audiometry in noise: *Tone in noise detection test (eAUD-N)* *Spectral masking release condition (eAUD-S)* *Temporal masking release condition (eAUD-T)*	*Spectro-temporal resolution*

Table 2: Summary statistics of the outcome measures of the BEAR test battery for the NH and HI group. The results are presented in terms of mean, standard deviation (SD) and the 1st (Q1) and 3rd quantiles (Q3) for the right ear (RE), left ear (LE) or both ears (Bin). In the case of frequency-specific examination, the frequency range is either low (LF) or high (HF).

			NH			***HI***		
Outcome measure	***Freq. Range***	***Ear***	***Mean (SD)***	***Q1***	***Q3***	***Mean (SD)***	***Q1***	***Q3***
SRT_Q (dB)		*LE*	*19.9 (7.1)*	*16.5*	*19.2*	*41.5 (13.5)*	*31.8*	*50.6*
		RE	*23.3 (8.9)*	*17.2*	*29.0*	*42.7 (12.6)*	*33.9*	*51.1*
Max DS (%)		*LE*	*99.2 (1.6)*	*100.0*	*100.0*	*97.2 (4.1)*	*95.3*	*100.0*
		RE	*97.2 (1.8)*	*95.5*	*97.6*	*93.9 (6.4)*	*92.1*	*98.4*
SRT_N (dB)		*LE*	*1.0 (0.7)*	*0.4*	*1.5*	*4.1 (3.4)*	*1.4*	*6.7*
		RE	*-0.5 (1.1)*	*-1.0*	*0.0*	*2.6 (3.8)*	*0.0*	*4.2*
$SScore^{+4dB}$ (%)		*LE*	*85.0 (11.7)*	*85.0*	*90.0*	*60.0 (26.6)*	*40.0*	*85.0*
		RE	*91.0 (9.6)*	*90.0*	*95.0*	*62.3 (24.0)*	*48.7*	*80.0*
MCL (dB HL)	*LF*	*LE*	*81.5 (14.8)*	*73.3*	*84.1*	*80.6 (8.4)*	*76.4*	*85.8*
		RE	*76.5 (13.2)*	*70.0*	*80.0*	*79.1 (7.9)*	*74.7*	*84.1*

			NH			*HI*		
Outcome measure	***Freq. Range***	***Ear***	***Mean (SD)***	***Q1***	***Q3***	***Mean (SD)***	***Q1***	***Q3***
	HF	*LE*	*79.0 (17.6)*	*66.6*	*90.8*	*82.7 (12.3)*	*75.8*	*90.0*
		RE	*73.8 (17.2)*	*65.0*	*80.0*	*80.3 (9.9)*	*74.7*	*87.5*
Slope (CU/dB)	*LF*	*LE*	*0.35 (0.1)*	*0.3*	*0.4*	*0.45 (0.1)*	*0.3*	*0.5*
		RE	*0.36 (0.1)*	*0.3*	*0.4*	*0.48 (0.2)*	*0.3*	*0.5*
	HF	*LE*	*0.45 (0.1)*	*0.3*	*0.4*	*0.84 (0.5)*	*0.5*	*0.9*
		RE	*0.41 (0.1)*	*0.3*	*0.4*	*0.81 (0.4)*	*0.5*	*0.9*
DynR (dB HL)	*LF*	*LE*	*91.5 (16.8)*	*78.3*	*97.5*	*76.7 (15.8)*	*64.5*	*88.3*
		RE	*91.1 (18.8)*	*79.1*	*100.0*	*73.9 (16.0)*	*61.6*	*86.8*
	HF	*LE*	*77.6 (18.2)*	*72.5*	*85.8*	*50.8 (15.1)*	*40.6*	*60.2*
		RE	*78.6 (17.9)*	*67.5*	*90.8*	*50.7 (15.5)*	*38.9*	*60.4*
sSTM -3dB (d')	*LF*	*Bin*	*2.6 (0.6)*	*2.4*	*3.0*	*1.7 (1.3)*	*0.4*	*3.0*
	HF		*1.6 (0.8)*	*1.1*	*2.4*	*0.6 (1.1)*	*-0.3*	*1.4*
fSTM (dB)	*LF*	*LE*	*-7.7 (1.8)*	*-9.0*	*-7.6*	*-2.8 (2.1)*	*-3.5*	*-0.8*
		RE	*-5.1 (3.1)*	*-7.2*	*-1.6*	*-1.6 (1.3)*	*-2.0*	*-0.6*
	HF	*LE*	*-8.0 (2.0)*	*-8.6*	*-6.2*	*-2.6 (2.4)*	*-3.8*	*-0.6*
		RE	*-5.6 (3.6)*	*-8.6*	*-2.1*	*-1.9 (1.5)*	*-2.0*	*-1.0*
IPD $_{fmax}$ (kHz)		*Bin*	*0.76 (0.26)*	*0.59*	*0.98*	*0.69 (0.27)*	*0.52*	*0.88*
Bin Pitch 20 (%)		*Bin*	*87.5 (25.0)*	*87.5*	*100.0*	*80.7 (30.9)*	*70.0*	*100.0*
eAUD-HF FLFT (kHz)		*LE*	*10.9 (1.2)*	*10.2*	*11.9*	*7.57 (2.7)*	*5.3*	*10.0*
		RE	*11.7 (1.1)*	*10.9*	*12.5*	*8.12 (2.3)*	*6.7*	*10.2*
eAUD-N (dB HL)	*LF*	*LE*	*70.4 (4.5)*	*68.0*	*71.5*	*71.8 (2.6)*	*70.2*	*73.2*
		RE	*69.2 (4.6)*	*65.2*	*72.5*	*72.0 (2.8)*	*69.6*	*74.3*
	HF	*LE*	*71.1 (2.5)*	*69.7*	*72.7*	*74.7 (3.4)*	*72.5*	*76.1*
		RE	*70.8 (3.6)*	*70.5*	*71.7*	*74.2 (3.1)*	*72.0*	*76.2*
TMR eAUD (N -T) (dB)	*LF*	*LE*	*7.5 (3.4)*	*6.0*	*7.5*	*7.7 (4.0)*	*6.1*	*10.1*
		RE	*5.2 (3.3)*	*4.0*	*7.6*	*8.3 (2.7)*	*6.5*	*10.3*
	HF	*LE*	*13.0 (0.6)*	*12.7*	*13.2*	*7.9 (5.0)*	*5.0*	*11.6*
		RE	*10.7 (3.1)*	*9.1*	*10.2*	*8.1 (5.2)*	*5.1*	*10.7*
SMR eAUD (N -S) (dB)	*LF*	*LE*	*19.3 (3.6)*	*16.5*	*21.7*	*19.6 (17.7)*	*17.7*	*23.2*
		RE	*18.8 (4.6)*	*17.0*	*21.2*	*20.0 (5.2)*	*16.5*	*23.8*
	HF	*LE*	*26.8 (4.5)*	*27.5*	*29.0*	*19.3 (9.5)*	*12.1*	*26.3*
		RE	*27.2 (3.7)*	*26.2*	*29.5*	*19.5 (9.9)*	*12.0*	*26.8*
BMR (S0N0 – S0Spi) (dB)		*Bin*	*16.5 (4.7)*	*13.5*	*17.5*	*14.7 (4.6)*	*12.2*	*17.5*

SRT_Q: Speech reception threshold in quiet / Max DS: Maximum speech discrimination score. // SRT_N: Speech reception threshold in noise / Score +4: Sentence recognition score at +4 dB SNR // MCL: Most comfortable level / Slope: Slope of the loudness function / DynR: Dynamic range // sSTM: Sensitivity for detecting a spectro-temporally modulated noise at 20log(m) = -3 dB, where m is the modulation depth / fSTM: Fast version of the STM test (Bernstein et al., 2016) // IPD fmax: Frequency threshold for detecting an interaural phase difference of 180°.

			NH			*HI*		
Outcome measure	***Freq. Range***	***Ear***	***Mean (SD)***	***Q1***	***Q3***	***Mean (SD)***	***Q1***	***Q3***

Bin pitch: Binaural pitch detection scores for 20 presentations // eAUD-HF: Fixed-level frequency threshold (FLFT) at 80 dB SPL // eAUD-N: Tone detection in TEN noise // TMR: Temporal masking release // SMR: Spectral masking release // BMR: Binaural masking release.

Speech perception in quiet

Methods

The word recognition score (WRS-4UFC) test was proposed as a systematic and self-administered procedure that allows the estimation of supra-threshold deficits in speech perception in quiet. The speech material was the same as the one used for standard speech audiometry (Dantale I; Elberling, Ludvigsen, & Lyregaard, 1989) in Danish. The self-administered procedure consisted of a 4-interval-unforced-choice paradigm (4UFC). After the presentation of each word, the target was placed randomly in one of four intervals. The other three words were also taken from the Dantale-I corpus. They were chosen based on the lowest Levenshtein phonetic distance (Sanders & Chin, 2009) from the target. Four lists of 25 words were presented at 40, 30, 20 and 10 dB above the individual PTA, in this order. A logistic function was fitted to the results from each individual ear and the speech reception threshold (SRT_Q) and maximum speech discrimination score (Max DS) were estimated using *psignifit* 4 software (Schütt, Harmeling, Macke, & Wichmann, 2016).

Results and discussion

The HI listeners' SRT_Q were, on average, 20 dB higher than the ones of the NH group. The interquartile range for the HI group was about 19 dB whereas for the NH group it

was 3 dB for the left ear (LE) and 11.8 dB for the right ear (RE). The Max DS for both groups was close to 100%. However, the HI listeners showed larger variability, especially in the right ear (SD= 6.42).

In the analysis of the test-retest variability, the WRS-4UFC test showed poor to moderate reliability especially at low levels (PTA + 10 dB; ICC = 0.25). However, at the higher presentation levels (i.e. individual PTA + 40 dB) the standard error of the measurement was only 4% (1 word). Regarding clinical applicability, the WRS-4UFC needs to be compared to traditional speech audiometry to explore the influence of using closed- vs. open-set and forced- vs. unforced-choice test procedures on the results.

Speech perception in noise

The Hearing in Noise Test (HINT; Nilsson et al., 1994) is an adaptive sentence recognition test carried out with speech-shaped noise. The following assumptions are considered in HINT: 1) Speech materials made of meaningful sentences yield a steep psychometric function; 2) Stationary noise with the same spectral shape as the average spectrum of the speech material makes the speech reception threshold in noise (SRT_N) less dependent of the spectral characteristics of the speaker's voice. Furthermore, the signal-to-noise ratio (SNR) between the target and masker is better defined across the frequency range; 3) The SRT_N is independent of the absolute noise level as long as the noise level is above the "internal noise" level. Therefore, it is recommended to present the noise at least 30 dB above the "internal noise". The internal noise is defined as the sum of the SRT in quiet of the tested listener and the SRT in noise for NH listeners, for a given speech material (Reinier Plomp, 1986).

Methods

The Danish HINT was used as in Nielsen & Dau (2011) to obtain the SRT_N. Additionally, a 20-sentence list was presented at a fixed signal-to-noise ratio of +4 dB and scored to obtain a sentence recognition score ($SScore^{+4dB}$). The presentation level of the noise was set between 65 and 85 dB SPL to ensure that the noise was always presented 30 dB above the individual PTA. Each ear was tested individually. All participants were tested using the same list with the same ear. However, for the test-retest reliability study, the list and ear presented were randomized, only using lists 6-10.

Results and discussion

The SRT_N for NH listeners were, on average, 2 dB higher than the ones reported Nielsen and Dau (2011). However, this might be explained by the fact that they used diotic presentation which can lead to a 1.5 dB improvement (Plomp & Mimpen, 1979). The results also showed a lower SRT_N (1.5 dB) and higher $SScore^{+4dB}$ (4%) for the right ear in both groups of listeners. According to Nielsen and Dau (2011), there was a significant main effect of test list. Such differences are seen mainly for lists 1-4, which were the lists used here. Therefore, the observed interaural difference can be ascribed to a list effect.

The ICC values (SRT_N: ICC= 0.61; $SScore^{+4dB}$: ICC = 0.57) indicated only moderate reliability of the HINT. The SRT_N showed an SEM = 1.02 dB, which is below the step size of the test (2 dB). The $SScore^{+4dB}$ showed an SEM value of 7.94%, which corresponds to an error in one of the sentences.

The use of speech-in-noise tests can be a useful tool for the characterization of the listener's hearing deficits that can be performed under different conditions, including monaural, binaural, unaided and aided stimuli presentations. While here the tests were

performed monaurally and unaided, a binaural condition as well as at least one aided measure (i.e., with hearing aids), could also be included in clinical practice.

Loudness perception

Loudness perception can substantially differ between NH and HI listeners and has been connected to the peripheral non-linearity (e.g. Jürgens, Kollmeier, Brand, & Ewert, 2011). While the growth of loudness shows a non-linear behaviour in a healthy ear, the results from HI listeners suggest that loudness perception becomes linear when outer-hair cell (OHC) function is affected (e.g. Moore, 2007). Besides, the possibilities of characterizing hearing deficits, loudness function can be used for fitting hearing aids (e.g., Oetting, Hohmann, Appell, Kollmeier, & Ewert, 2018). Adaptive categorical loudness scaling (ACALOS; Brand & Hohmann, 2002) is the reference method for the current standard (ISO 16832, 2006) for loudness measurements.

Methods

According to the ACALOS method, narrow-band noises were presented sequentially, and the participant had to judge the perceived loudness using a 13-category scale ranging from “not heard” to “extremely loud”. The raw results, which correspond to categorical units (CU) spanned between 0 and 50, were fitted to a model of loudness as described in Oetting, Brand, & Ewert (2014). The outcome measures of the ACALOS presented here are the most comfortable level (MCL), the slope of the loudness function (Slope), and the dynamic range (DynR) defined as the difference between uncomfortable level (50 CU) and the hearing threshold (0.5 CU). Low-frequency (LF) average corresponds to frequencies below 1.5 kHz, high-frequency (HF) average correspond to frequencies above 1.5 kHz

Results and discussion

The average MCL estimate ranged between 73 and 82 dB HL in both groups and for both frequency ranges. The average slope of the loudness growth was slightly steeper for the HI listeners in the low-frequency range (0.45 CU/dB HI vs. 0.35 CU/dB NH) and substantially steeper in the high-frequency range (0.8 vs 0.45). The average dynamic range was between 80 and 90 dB HL for the NH listeners, and smaller for the HI listeners, especially at high frequencies (50.8 dB).

Regarding the test-retest reliability, ACALOS showed an excellent reliability for estimating the hearing thresholds (ICC = 0.94; SEM = 4.5 dB), good reliability for estimating the MCL (ICC = 0.68, SEM = 6.5 dB) and very good reliability for estimating the slope (ICC = 0.82; SE M = 0.07 CU/dB). Overall, these results supported the inclusion of ACALOS in a clinical test battery, as it provides several outcomes (hearing thresholds, growth of loudness, MCL and dynamic range). ACALOS also showed a high time efficiency (around 10 min. per ear).

Spectro-temporal modulation sensitivity

A speech signal can be decomposed into spectral and temporal modulations. While speech-in-noise perception assessment leads to some confounds due to the variety of speech corpora, noise maskers, and test procedures that can all affect the results, the assessment of the sensitivity of simpler sounds might be of interest for characterizing a listener's spectro-temporal processing abilities. Bernstein et al. (2013) showed significant differences between NH and HI listeners for detecting STM in random noise. These differences corresponded to specific conditions that were also useful for the prediction of speech-in-noise performance in the same listeners. Lately, the assessment of STM sensitivity in these specific conditions gained an increasing interest due to its

potential for predicting speech intelligibility (Bernstein et al., 2016; Gallun et al., 2018; Zaar, Simonsen, Bherens, Dau, & Laugesen, 2019) and for assessing cochlear-implant candidacy (Choi et al., 2016). Here, STM sensitivity was assessed using a new test paradigm that may be more suitable for a clinical implementation. The test was performed in two conditions: a low-frequency condition (similar to the one previously used in Bernstein et al., 2016) and a high-frequency condition (Mehraei, Gallun, Leek, & Bernstein, 2014).

Methods

The stimuli were similar to those of Bernstein et al. (2016) and Mehraei et al. (2014), but a different presentation paradigm was employed. A sequence of four noises was presented in each trial. The first and third stimulus always contained unmodulated noise, whereas the second and fourth stimuli could be either modulated or unmodulated. After the sequence was presented, the listener had to respond whether the four sounds were different (‘yes’) or the same (‘no’). Two procedures involving catch trials were evaluated. The first test (sSTM -3 dB) was a screening test consisting of 10 stimuli modulated at -3 dB level and five unmodulated ones presented in random order. The outcome measure was the listener’s sensitivity (d’) in the task. The second test (fSTM) tracked the 80% threshold using the single-interval adjusted matrix (SIAM; Kaernbach, 1990) paradigm.

Results and discussion

The screening STM test shows the sensitivity in terms of d’, where the maximum value is d’ = 3, i.e. 10 modulated and 5 unmodulated stimuli correctly detected. In the hypothetical case when all the catch trials are detected, the lowest d’ value can be -0.3. The NH listeners showed a high sensitivity in the low-frequency condition (d’ = 2.6)

and a somewhat lower sensitivity in the high-frequency condition (d' = 1.63) corresponding to 65% correct responses. The HI listeners showed a higher variability and a lower sensitivity in the low-frequency condition (~70% correct) and substantially lower sensitivity in the high-frequency condition (0-50% correct responses). The threshold tracking procedure (fSTM) showed results between -9 and -6 dB in the NH group, whereas the HI listeners showed thresholds between -3.50 and -0.5 dB. Although the results of the fSTM low-frequency condition were consistent with Bernstein et al. (2016), the results in the high-frequency condition showed higher thresholds than the ones in Mehraei et al. (2014). This can be ascribed to the higher presentation level used in Mehraei et al. (2014) than in the current test procedure.

The fSTM showed an excellent reliability (ICC = 0.91; SEM = 0.93 dB) in the LF condition. However, several HI listeners were not able to complete the procedure for the HF condition. Overall, the use of the SIAM tracking procedure allowed us to obtain accurate thresholds, although additional repetitions were required, especially in the HF condition. This might be because the psychometric function for detecting the stimulus can be shallower in this condition or because the 100% detection could not be reached even in the fully-modulated trials. Therefore, a Bayesian procedure being able to estimate the threshold and slope of the psychometric function, such as the Bayes Fisher information gain (FIG; Remus & Collins, 2008), might be more suitable for this type of test.

Binaural processing abilities

Binaural hearing is useful for sound localization and the segregation of complex sounds (Darwin, 1997). Interaural differences in level or timing are processed for spatial hearing purposes in the auditory system. With hearing loss, the neural signal at the

output of the cochlea can be degraded which may lead to reduced binaural abilities typically connected to temporal fine structure (TFS) processing. Based on a method estimating the upper-frequency limit for detecting an interaural phase difference (IPD) of 180° (IPD_{fmax}; Neher, Laugesen, Søgaard Jensen, & Kragelund, 2011; Ross, Tremblay, & Picton, 2007; Sébastien Santurette & Dau, 2012b), Füllgrabe, Harland, Sęk, & Moore, (2017) recently proposed a refined test as a feasible way to evaluate TFS sensitivity. This paradigm was used in recent research that suggested that IPD_{fmax} might be related to non-auditory factors (Strelcyk, Zahorik, Shehorn, Patro, & Derleth, 2019) and affected by factors beyond hearing loss, such as musical training (Bianchi, Carney, Dau, & Santurette, 2019). Therefore, the IPD_{fmax} might be a task that requires auditory and non-auditory processing abilities beyond TFS sensitivity.

In contrast, binaural pitch detection assesses binaural processing abilities in a different manner. This test requires the detection of pitch contours embedded in noise, which are diotically or dichotically evoked. While the diotic condition can be resolved monoaurally, the dichotic condition requires the binaural processing abilities to be sufficiently intact to detect the contour. Previous studies showed that some listeners were unable to detect binaural pitch, regardless of the audiometric configuration (Sanchez-Lopez et al., 2018; Santurette & Dau, 2012). Therefore, it was of interest to compare the results of these two binaural processing tests.

Methods

The maximum frequency for detecting an IPD of 180° with pure-tones was obtained using a 2-AFC tracking procedure similar to the one used in Füllgrabe et al. (2017). The frequency threshold (IPD_{fmax}) was obtained from the average of two runs.

Binaural pitch detection scores were obtained using a clinical implementation of the test proposed by Santurette & Dau, (2012). A 3-minute sequence of noise was presented bilaterally. Ten diotic and 10 dichotic pitch contours, embedded in the noise, had to be detected by the listener. The tones forming the pitch contours were generated by adding frequency-specific IPDs to the presented noise (Cramer & Huggins, 1958). The outcome measure of the binaural pitch test was the percentage score averaged across two repetitions (BP20).

Results and discussion

The listeners in the NH and HI groups showed IPD_{fmax} thresholds around 700 Hz with a standard deviation (~270 Hz) and interquartile range (~370 Hz) similarly in both groups. These results are in line with the ones reported in Füllgrabe & Moore (2017). The IPD_{fmax} test showed excellent reliability (ICC = 0.95; SEM = 65.4 Hz), and the median time needed for two repetitions was 10 minutes. This suggests that IPD_{fmax} is a reliable measure of binaural processing abilities that can reveal substantial variability among both NH and HI listeners, which is valuable for highlighting individual differences among patients.

The overall results from the binaural pitch test for the NH listeners showed >87.5% correct detection, whereas the HI listeners' results showed a higher variability with an interquartile range from 70-100%. The test showed excellent reliability (ICC = 0.98; SEM = 4%). Listeners reported a positive experience due to the test being short and easy to understand.

Extended audiometry in noise (eAUD)

The extended audiometry in noise (eAUD) is a tone detection test intended to assess different aspects of auditory processing by means of a task similar to pure-tone

audiometry. The tone is presented either in noise or in quiet and the listener has to indicate whether the tone was perceived or not. The aspects of auditory processing assessed here are 1) high-frequency audibility, 2) spectral and temporal resolution and 3) binaural processing abilities.

High-frequency audibility

Recently, elevated thresholds at high frequencies (>8 kHz) have been linked to the concept of "hidden hearing loss" and synaptopathy (Liberman, Epstein, Cleveland, Wang, & Maison, 2016). However, the measurement of audiometric thresholds above 8 kHz is not part of the current clinical practice. The fixed-level frequency threshold (FLFT) has been proposed as a quick and efficient alternative to high-frequency audiometry (Rieke et al., 2017). The test is based on the detection of a tone presented at a fixed level. The frequency of the tone is varied towards high frequencies and the maximum audible frequency at the given level is estimated in an adaptive procedure. Here, a modified version of FLFT was used as the extended audiometry at high frequencies (eAUD-HF).

Spectro-temporal resolution

Frequency and temporal resolution are aspects of hearing that are fundamental for the analysis of perceived sounds. While NH listeners exhibit a frequency selectivity on the order of one third of an octave (from Glasberg & Moore, (1990), HI listeners have typically broader auditory filters leading to impaired frequency selectivity (e.g. Moore, 2007). Temporal resolution can be characterized by the ability to "listen in the dips" when the background noise is fluctuating based on the so-called masking release (Festen & Plomp, 1990). Schorn & Zwicker, (1990) proposed an elaborated technique for assessing both spectral and temporal resolution using two tests: 1) Psychoacoustical

tuning curves and 2) temporal resolution curves. In both cases, the task consists of detecting a pure tone that is masked by noise or another tone while the spectral or temporal characteristics of the masker are varied. Later, Larsby & Arlinger (1998) proposed a similar paradigm, the F-T test, which was successfully tested in HI listeners (van Esch & Dreschler, 2011). Here, the spectro-temporal resolution was assessed using a new test. This test is a tone-in-noise detection task consisting of three conditions as sketched in Figure 2.

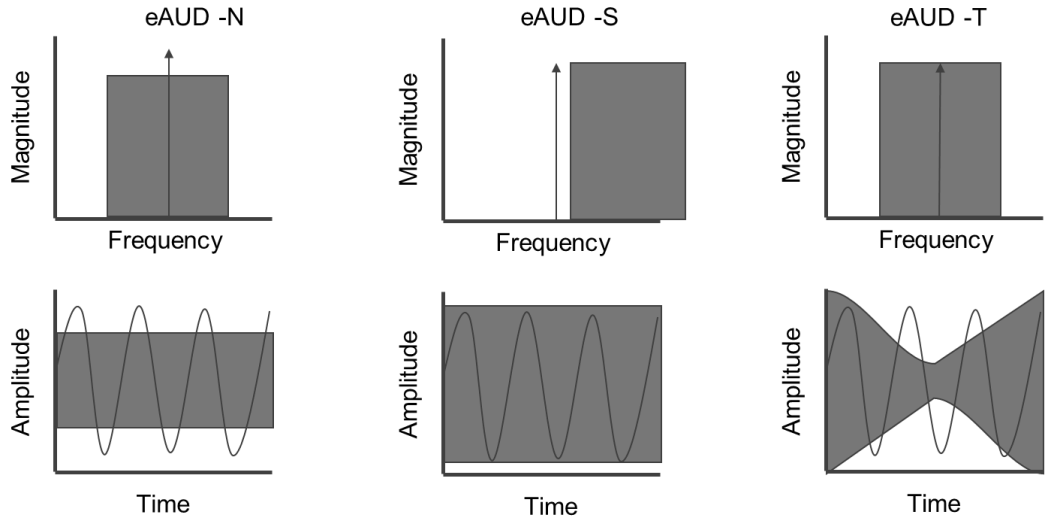

Figure 2: Sketch of the conditions of the spectro-temporal resolution measures of the extended audiometry in noise (eAUD). The top panel shows the spectrum of the noise and target pure-tone (delta), the bottom panel shows both signals in the time domain. Left panel: Tone in noise condition (eAUD-N). Middle panel: Spectral condition (eAUD-S). Right panel: Temporal condition (eAUD-T).

1) eAUD-N: The tone is embedded in a 1-octave-wide threshold equalizing noise (TEN; Moore, 2001). Because of the properties of the TEN, the tone detection threshold is comparable to the level of the noise in dB HL.
2) eAUD-S: The tone is embedded in a TEN that has been shifted up in frequency. In the spectral domain, this yields spectral unmasking of the tone, so the detection threshold is lower than in eAUD-N.
3) eAUD-T: The tone is embedded in a temporally-modulated noise with the same spectral properties as the one in eAUD-N. In the temporal domain, the modulations of the noise yield temporal unmasking, so the tone can be detected in the dips.

The outcome measures were focused on the temporal and spectral benefits expected in the eAUD-S and eAUD-T conditions compared to the eAUD-N condition. While in the noise condition (eAUD-N) the threshold is expected to be approximately at the level of the noise, in the temporal and spectral conditions the thresholds should be lower showing temporal masking release (TMR) and spectral masking release (SMR).

Binaural Masking Release

Besides the binaural tests presented previously, another approach for evaluating the binaural processing abilities is assessing binaural masking release (Durlach, 1963), which has been used in several studies (Neher, 2017; Strelcyk & Dau, 2009) and implemented in some commercial audiometers (Brown & Musiek, 2013). In this paradigm, a tone-in-noise stimulus is presented in two conditions: (1) a diotic condition where the tone is in phase in the two ears, and (2) a dichotic condition where the tone is in antiphase in the two ears. The difference between the two yields the benefit for tone detection due to binaural processing, the so-called binaural masking release (BMR).

Methods

The procedure used here was a yes/no task using a SIAM procedure (Kaernbach, 1990). As in traditional up-down procedures, the target can be presented in a given trial or not. If the target was detected, the target-presentation level is decreased according to a given step size; if it was not detected, the level is increased. If the stimulus was not presented (catch trial) but the listener provided a positive response, the level is decreased compared to the previous trial.

The target stimulus for all the conditions tested here was a warble tone. For each run, the first two reversals were discarded, and the threshold of each trial was calculated as the average of the four subsequent reversals. The low-frequency condition (LF) corresponds to the detection of a 0.5-kHz warble tone, whereas the high-frequency (HF) condition corresponded to a 2-kHz warble tone. The final threshold was calculated as the mean threshold of two repetitions. The outcome measures of the eAUD are 1) the high-frequency threshold (eAUD-HF), 2) the tone-in-noise threshold (eAUD-N), 3) the SMR, 4) the TMR, and 5) the BMR.

Results and discussion

The maximum frequency threshold for a tone presented at 80 dB SPL (eAUD-HF) was 11 kHz for the NH listeners and 8 kHz for the HI listeners. The HI group showed larger variability compared to the NH group (interquartile range: 6 kHz vs. 10 kHz). In contrast, the eAUD-N condition showed a larger variance for the NH group (SD = 4.5 dB HL) at low frequencies. The detection thresholds were in line with previous work with thresholds close to the noise presentation level (70 dB HL) (Vinay, Hansen, Raen, & Moore, 2017). The TMR shown by the NH group was larger at high frequencies (10 dB) than at low frequencies (7 dB). The HI group showed, on average, similar TMR

only at low frequencies. The SMR shown by the NH listeners was 19 dB for low frequencies and 26 dB for high frequencies. In contrast, for the HI listeners, the SMR was 7 dB lower only in the high-frequency condition. The BMR shown by both groups was around 15 dB, as expected from previous studies (Durlach, 1963).

The reliability of the eAUD was moderate for most of the conditions (ICC < 0.75). The eAUD-HF test showed very good reliability (ICC = 0.89; SEM = 495 Hz), and the eAUD-S at low frequencies showed good reliability (ICC = 0.85; SEM = 1.78 dB). The masking release estimates showed good reliability only for the high-frequency condition. The reason for this might be that masking release is a differential measure, and the cumulative error is, therefore, higher than that of each individual measure. The reduced reliability can be explained to some extent by the method used. To have a similar procedure as in pure-tone audiometry, the parameters of the SIAM tracking procedure were set accordingly. However, this made the test challenging and the listeners consistently missed several catch trials. Thus, extra trials were required to improve measurement accuracy. However, the standard error of the measurement was in most cases larger than the final step size (2 dB). As in the case of the fSTM, a different procedure, such as Bayesian adaptive methods, might increase measurement reliability.

Exploratory analysis

The collection of tests included in the test battery was intended to explore different and potentially independent aspects of hearing to obtain an auditory profile with controlled interrelations among the tests. A factor analysis performed in the HEARCOM study (Vlaming et al., 2011) based on data from 72 HI subjects revealed auditory dimensions: 1) high-frequency processing, 2) audibility, 3) low-frequency processing and 4) recruitment. In the current study, the results of the behavioural tests were analysed

further in order to explore possible interrelations between the various outcome measures.

Methods

First, the data were pre-processed as in Sanchez Lopez et al. (2018) to reduce the number of variables. The outcome variables of the frequency-specific tests were divided into LF (≤1 kHz) and HF (>1 kHz) variables. This decision was supported by a correlation analysis performed on the complete set of outcome variables, where the outcomes corresponding to 2, 4 and 6 kHz as well as the ones corresponding to 0.25, 0.5 and 1 kHz were highly intercorrelated. For the tests performed monaurally, the mean of the two ears was taken as the resulting outcome variable. The resulting dataset (BEAR3 dataset) contained 26 variables, divided into six groups corresponding to the six aspects of auditory processing considered here. The exploratory analysis consisted of a correlation analysis using Spearman correlations and factor analysis. The factor analysis was performed using an orthogonal rotation ("varimax") and the method of maximum likelihood. The number of components was chosen using parallel analysis, the resulting number of components was four.

Results

Figure 3 shows the results from the correlation analysis performed on the BEAR3 dataset. For convenience, the absolute value of the correlation was used when visualizing the data to show the strength of the correlation. The circles on the left-hand side of the figure depict significant correlations ($p < 0.00001$), and the correlation values are presented on the left-hand side of the figure. Two groups of correlated variables can be observed. The upper-left corner shows variables related to LF processing (dynamic range, the slope of the loudness function, and hearing thresholds)

and speech intelligibility in quiet. The bottom-right corner shows a larger group of correlated variables including HF processing, speech intelligibility in noise, and spectro-temporal resolution at high frequencies. The variables that are not significantly interrelated are shown in the middle part of Figure 3, including the three variables related to binaural processing abilities ($IPD_{fmax,}$ BP_{20} and BMR) which were not significantly correlated to each other. The speech reception threshold in quiet (SRT_Q) and the STM detection were correlated to various variables such as tone-in-noise detection, HF spectro-temporal resolution, LF hearing thresholds and speech-in-noise perception.

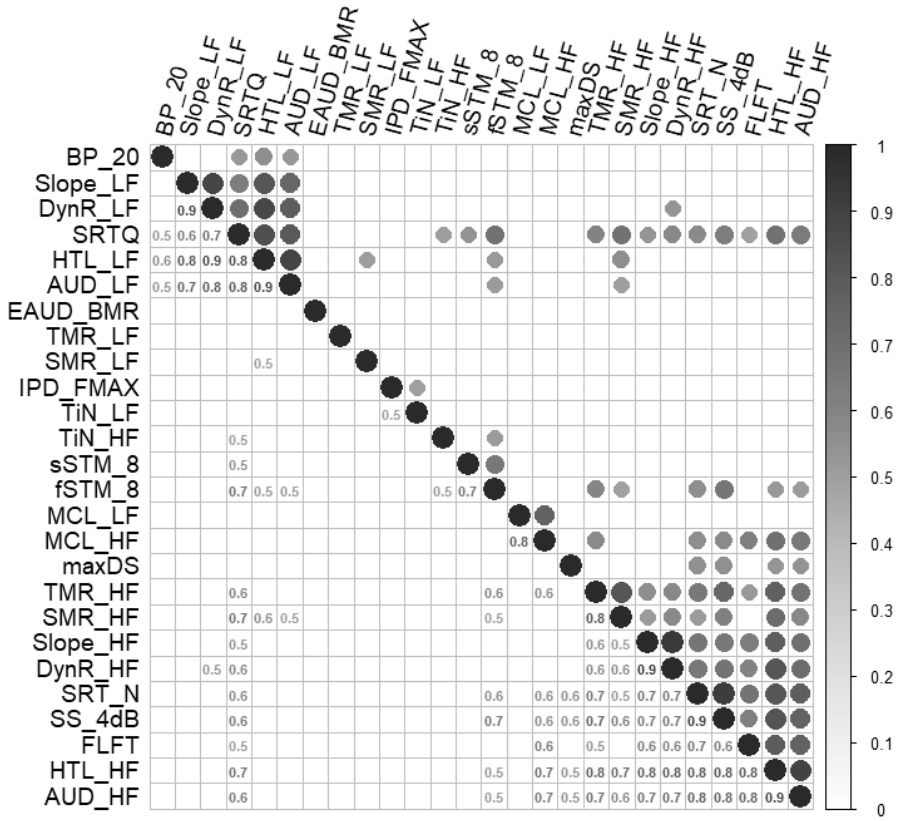

Figure 3: Correlation plot of the data set BEAR3. The upper part shows the significantly correlated variables as coloured circles. The lower panel shows the numeric correlation value.

The four factors resulting from the factor analysis showed 63% of explained cumulative variance. The variables with higher loadings (> 0.65) for each of the factors are shown in Table 3.

Table 3: Variables correlated to the four latent orthogonal factors resulting from the factor analysis with the method of maximum likelihood (ML). Columns are sorted in terms of the variance explained by each factor.

	ML2(19%)	***ML1(18%)***	***ML3(14%)***	***ML4(12%)***
HTL_LF	*0.93*			
DYNR_LF	*-0.90*			
AUD_LF	*0.82*			
SLOPE_LF	*0.81*			
SRTQ	*0.67*			
DYNR_HF		*-0.93*		
SLOPE_HF		*0.82*		
HTL_HF		*0.79*		
AUD_HF		*0.73*		
MCL_HF			*0.92*	
MCL_LF			*0.85*	
SRT_N				*0.77*
SSCORE_4DB				*-0.78*

The first factor, in terms of the amount of variance explained (19%), was associated with LF loudness perception and speech intelligibility in quiet, whereas the second factor (18% of variance explained) was associated with HF loudness perception.

Despite loudness perception being associated with the first and second factor, the MCL was associated, both at high and low frequencies, with the third factor, while the fourth factor was associated with speech intelligibility in noise.

General discussion

The first goal of the present study was to collect data of a heterogeneous population of HI listeners, reflecting their hearing abilities in different aspects of auditory processing. The current study was motivated by the need for a new dataset to refine the data-driven approach for auditory profiling. The dataset should contain a representative population of listeners and outcome measures (Sanchez-Lopez et al., 2018) to allow a refined definition of the two types of auditory distortions and to identify subgroups of listeners with clinical relevance. To refine the data-driven auditory profiling, the BEAR3 dataset fulfils all the requirements discussed in Sanchez-Lopez et al. (2018). Other datasets containing a large number of listeners (Gieseler et al., 2017; Rönnberg et al., 2016) or physiological measures (Kamerer, Kopun, Fultz, Neely, & Rasetshwane, 2019) could also be interesting for complementing the auditory profiling beyond auditory perceptual measures

Relationships across different aspects of auditory processing

The proposed test battery considers outcomes divided into six dimensions of auditory processing. One of the objectives of the study was to investigate the interrelations of different dimensions and measures. The present analysis showed two interesting findings. First, the correlation analysis shows two clusters of variables related to either low- or high-frequency audiometric thresholds. Speech-in-noise perception was associated with high-frequency sensitivity loss, temporal, and spectral masking release whereas speech-in-quiet was correlated with both low- and high-frequency hearing loss.

Several outcomes were not interrelated, especially the outcomes associated with binaural processing abilities. Second, factor analysis yielded latent factors related to low- and high-frequency processing, most comfortable level and speech in noise. Vlaming et al. (2011) showed four dimensions in the factor analysis of the HEARCOM project data corresponding to high-and low-frequency spectro-temporal processing, MCL and recruitment. In contrast, the current study showed that the slopes of the loudness growth, both at low and high frequencies, were not interrelated and contributed to the first and second latent factors. Additionally, the speech-in-noise test performed in HEARCOM was associated with the low-frequency processing, whereas, in the present study, speech-in-noise dominates the fourth factor and is significantly correlated with high frequencies. The reason for this discrepancy might be the use of different types of noise and test procedures in the two studies.

Overall, the data of the present study seem to be dominated by the audiometric profiles, with low- and high-frequency processing reflecting the main sources of variability in the data. However, binaural processing abilities, loudness perception and speech-in-noise outcomes showed a greater contribution to the variability of the supra-threshold measures than spectro-temporal processing outcomes.

Towards clinical feasibility of the tests

The test-retest reliability of the test battery was investigated based on the results of a subset of listeners who participated 2-5 months after the first visit. The analysis was based on the ICC and the SEM. Some of the tests, such as IPD_{fmax}, binaural pitch and FLFT showed good to excellent test-retest reliability with all ICC values above 0.9, while other tests, such as the extended audiometry in noise and speech intelligibility in quiet, showed poor reliability.

The selected tests were conducted in two sessions and the total time was, on average, three hours. In realistic clinical setups, a subset of tests with high reliability and a reasonably low difficulty would need to be prioritized. For a clinical version of the test battery, other tracking procedures such as Bayesian Functional information (Remus & Collins, 2008) might be adopted to improve the reliability and time-efficiency in some tasks such as STM and tone detection in noise. Moreover, if time-efficiency is crucial, testing some aspects of auditory processing out of the clinic, as other proposed test batteries for auditory research (Gallun et al., 2018), might be a solution for completing the patient's hearing profile.

A clinical test battery with the subset of tests that showed a good or excellent test-retest reliability should be evaluated in a large scale study. This should include several aspects of auditory processing and provide detailed information on the supra-threshold deficits of the patient. The tests that showed potential for the clinical implementation were ACALOS, HINT, fSTM, BP and IPD_{fmax}. Such a test battery could serve to identify clinically relevant subset of patients (auditory profiles) that may benefit from specific types of hearing rehabilitation towards a "stratified approach" (Lonergan et al., 2017) for audiology practice.

Conclusion

The analysis of the data showed that a reduced BEAR test battery has the potential for clinical implementation, providing relevant and reliable information reflecting several auditory domains. The proposed test battery showed good reliability, was reasonably time-efficient and easy to perform. The implementation of a clinical version of the test battery is publicly available and can be evaluated in future research, e.g. in a larger field study to further refine the auditory profiling approach. Moreover, the current data will

be re-analysed in a continuation study to better define the auditory profiles proposed in the data-driven approach and the two types of auditory distortions.

Declaration of interest

This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (Better hEAring Rehabilitation project), Oticon, GN Resound, Widex and other partners (Aalborg University, University of Southern Denmark, the Technical University of Denmark, Force, Aalborg, Odense and Copenhagen University Hospitals). No potential conflict of interest was reported by the authors. The data that support the findings of this study are openly available in Zenodo at http://doi.org/10.5281/zenodo.3459579. A clinical implementation of the test battery is publicly available at https://bitbucket.org/hea-dtu/bear-test-battery/src/master/.

Acknowledgements

We thank the staff from OUH, BBH and HEA, especially JH Schmidt, SS Houmøller, E Kjærbøl, RS Sørensen, and the student helpers from the MSc of Audiology at SDU. The funding and collaboration of all BEAR partners are sincerely acknowledged. We also want to show our gratitude to all the participants in the study.

References

Bernstein, J. G. W., Danielsson, H., Hällgren, M., Stenfelt, S., Rönnberg, J., & Lunner, T. (2016). Spectrotemporal Modulation Sensitivity as a Predictor of Speech-Reception Performance in Noise With Hearing Aids. *Trends in Hearing*, *20*, 1–17. https://doi.org/10.1177/2331216516670387

Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., & Leek,

M. R. (2013). Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *Journal of the American Academy of Audiology*, *24*(4), 293–306. https://doi.org/10.3766/jaaa.24.4.5

Bianchi, F., Carney, L. H., Dau, T., & Santurette, S. (2019). Effects of Musical Training and Hearing Loss on Fundamental Frequency Discrimination and Temporal Fine Structure Processing: Psychophysics and Modeling. *Journal of the Association for Research in Otolaryngology*, *20*(3), 263–277. https://doi.org/10.1007/s10162-018-00710-2

Bisgaard, N., Vlaming, M. S. M. G., & Dahlquist, M. (2010). Standard Audiograms for the IEC 60118-15 Measurement Procedure. *Trends in Amplification*, *14*(2), 113–120. https://doi.org/10.1177/1084713810379609

Brand, T., & Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. *The Journal of the Acoustical Society of America*, *112*(4), 1597–1604. https://doi.org/10.1121/1.1502902

Brown, M., & Musiek, F. (2013). Pathways: The Fundamentals of Masking Level Differences for Assessing Auditory Function. *The Hearing Journal*, *66*(1), 16. https://doi.org/10.1097/01.HJ.0000425772.41884.1d

Brungart, D. S., Sheffield, B. M., & Kubli, L. R. (2014). Development of a test battery for evaluating speech perception in complex listening environments. *The Journal of the Acoustical Society of America*, *136*(2), 777–790. https://doi.org/10.1121/1.4887440

Choi, J. E., Hong, S. H., Won, J. H., Park, H.-S., Cho, Y. S., Chung, W.-H., … Moon, I. J. (2016). Evaluation of Cochlear Implant Candidates using a Non-linguistic Spectrotemporal Modulation Detection Test. *Scientific Reports*, *6*(April), 35235. https://doi.org/10.1038/srep35235

Cramer, E. M., & Huggins, W. H. (1958). Creation of Pitch through Binaural Interaction. *The Journal of the Acoustical Society of America*, *30*(5), 413–417. https://doi.org/10.1121/1.1909628

Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, *1*(9), 327–333. https://doi.org/10.1016/S1364-6613(97)01097-8

Desloge, J. G., Reed, C. M., Braida, L. D., Perez, Z. D., & D'Aquila, L. A. (2017). Masking release for hearing-impaired listeners: The effect of increased audibility through reduction of amplitude variability. *The Journal of the Acoustical Society of America*, *141*(6), 4452–4465. https://doi.org/10.1121/1.4985186

Durlach, N. I. (1963). Equalization and Cancellation Theory of Binaural Masking-Level Differences. *The Journal of the Acoustical Society of America*, *35*(8), 1206–1218. https://doi.org/10.1121/1.1918675

Elberling, C., Ludvigsen, C., & Lyregaard, P. E. (1989). DANTALE: a new Danish speech material. *Scandinavian Audiology*, *18*(3), 169–175. https://doi.org/10.3109/01050398909070742

Ewert, S. (2013). AFC - A modular framework for running psychoacoustic experiments and computational perception models. *Proceedings of the International Conference on Acoustics AIA-DAGA 2013*, 1326–1329. Retrieved from www.aforcedchoice.com

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*, 1725–1736. https://doi.org/10.1121/1.400247

Füllgrabe, C., Harland, A. J., Sęk, A. P., & Moore, B. C. J. (2017). Development of a method for determining binaural sensitivity to temporal fine structure. *International Journal of Audiology*, *56*(12), 926–935. https://doi.org/10.1080/14992027.2017.1366078

Füllgrabe, C., & Moore, B. C. J. (2017). Evaluation of a Method for Determining Binaural Sensitivity to Temporal Fine Structure (TFS-AF Test) for Older Listeners With Normal and Impaired Low-Frequency Hearing. *Trends in Hearing*, *21*, 1–14. https://doi.org/10.1177/2331216517737230

Gallun, F. J., Seitz, A., Eddins, D. A., Molis, M. R., Stavropoulos, T., Jakien, K. M., …

Srinivasan, N. (2018). Development and validation of Portable Automated Rapid Testing (PART) measures for auditory research. *The Journal of the Acoustical Society of America*, *143*(3), 050002. https://doi.org/10.1121/2.0000878

Gelfand, S. A. (2009). *Essentials of Audiology* (4th ed., p. 582). 4th ed., p. 582. https://doi.org/10.1097/MAO.0b013e3181c99550

Gieseler, A., Tahden, M. A. S., Thiel, C. M., Wagener, K. C., Meis, M., & Colonius, H. (2017). Auditory and Non-Auditory Contributions for Unaided Speech Recognition in Noise as a Function of Hearing Aid Use. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00219

Glasberg, B. R., & Moore, B. C. (1989). Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scandinavian Audiology. Supplementum*, *32*, 1–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2711118

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1–2), 103–138. https://doi.org/"https://doi.org/10.1016/0378-5955(90)90170-T

Houtgast, T., & Festen, J. M. (2008). On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *International Journal of Audiology*, *47*, 287–295. https://doi.org/10.1080/14992020802127109

ISO 16832. (2006). *Acoustics - Loudness scaling by means of categories* (Vol. 2006, pp. 1–12). Vol. 2006, pp. 1–12. Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=32442

ISO 8253-1. (2010). Acoustics - Audiometric test methods - Part 1: Pure-tone air and bone conduction audiometry. *International Organization for Standardization*. Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=43601

ISO 8253-3. (2012). Acoustics. Audiometric test methods - Part 3: Speech audiometry. *International Organization for Standardization*. Retrieved from

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45101

Johannesen, P. T., Pérez-González, P., Kalluri, S., Blanco, J. L., & Lopez-Poveda, E. A. (2016). The Influence of Cochlear Mechanical Dysfunction, Temporal Processing Deficits, and Age on the Intelligibility of Audible Speech in Noise for Hearing-Impaired Listeners. *Trends in Hearing*, *20*, 233121651664105. https://doi.org/10.1177/2331216516641055

Jürgens, T., Kollmeier, B., Brand, T., & Ewert, S. D. (2011). Assessment of auditory nonlinearity for listeners with different hearing losses using temporal masking and categorical loudness scaling. *Hearing Research*, *280*(1–2), 177–191. https://doi.org/10.1016/j.heares.2011.05.016

Kaernbach, C. (1990). A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing. *The Journal of the Acoustical Society of America*, *88*(6), 2645–2655. https://doi.org/10.1121/1.399985

Kamerer, A. M., Kopun, J. G., Fultz, S. E., Neely, S. T., & Rasetshwane, D. M. (2019). Reliability of Measures Intended to Assess Threshold-Independent Hearing Disorders. *Ear and Hearing*, *40*(6), 1267–1279. https://doi.org/10.1097/AUD.0000000000000711

Killion, M. C., Niquette, P. a, Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *116*(October 2004), 2395–2405. https://doi.org/10.1121/1.2166610

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*. https://doi.org/10.1016/j.jcm.2016.02.012

Larsby, B., & Arlinger, S. (1998). A method for evaluating temporal, spectral and combined temporal-spectral resolution of hearing. *Scandinavian Audiology*, *27*(1),

3–12. https://doi.org/10.1080/010503998419641

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., & Maison, S. F. (2016). Toward a Differential Diagnosis of Hidden Hearing Loss in Humans. *PLOS ONE*, *11*(9), e0162726. https://doi.org/10.1371/journal.pone.0162726

Lonergan, M., Senn, S. J., McNamee, C., Daly, A. K., Sutton, R., Hattersley, A., … Pirmohamed, M. (2017). Defining drug response for stratified medicine. *Drug Discovery Today*. https://doi.org/10.1016/j.drudis.2016.10.016

Mehraei, G., Gallun, F. J., Leek, M. R., & Bernstein, J. G. W. (2014). Spectrotemporal modulation sensitivity for hearing-impaired listeners: Dependence on carrier center frequency and the relationship to speech intelligibility. *The Journal of the Acoustical Society of America*, *136*(1), 301–316. https://doi.org/10.1121/1.4881918

Moore, B. C. J. (2001). Dead Regions in the Cochlea: Diagnosis, Perceptual Consequences, and Implications for the Fitting of Hearing Aids. *Trends in Amplification*, *5*(1), 1–34. https://doi.org/10.1177/108471380100500102

Moore, B. C. J. (2007). Cochlear Hearing Loss. In *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*. https://doi.org/10.1002/9780470987889

Neher, T. (2017). Characterizing the binaural contribution to speech-in-noise reception in elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4976327

Neher, T., Laugesen, S., Søgaard Jensen, N., & Kragelund, L. (2011). Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities? *The Journal of the Acoustical Society of America*, *130*(3), 1542–1558. https://doi.org/10.1121/1.3608122

Nielsen, J. B., & Dau, T. (2011). The Danish hearing in noise test. *International Journal of Audiology*, *50*(3), 202–208. https://doi.org/10.3109/14992027.2010.524254

Nilsson, M., Soli, S. D., & Sullivan, J. a. (1994). Development of the Hearing In Noise

Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Oetting, D., Brand, T., & Ewert, S. D. (2014). Optimized loudness-function estimation for categorical loudness scaling data. *Hearing Research*, *316*, 16–27. https://doi.org/10.1016/j.heares.2014.07.003

Oetting, D., Hohmann, V., Appell, J. E., Kollmeier, B., & Ewert, S. D. (2018). Restoring Perceived Loudness for Listeners With Hearing Loss. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000521

Oxenham, A. J., & Simonson, A. M. (2009). Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *The Journal of the Acoustical Society of America*, *125*(1), 457–468. https://doi.org/10.1121/1.3021299

Plomp, R. (1986). A Signal-to-Noise Ratio Model for the Speech-Reception Threshold of the Hearing Impaired. *Journal of Speech, Language, and Hearing Research*, *29*(2), 146–154. https://doi.org/10.1044/jshr.2902.146

Plomp, R., & Mimpen, A. M. (1979). Improving the Reliability of Testing the Speech Reception Threshold for Sentences. *International Journal of Audiology*, *18*(1), 43–52. https://doi.org/10.3109/00206097909072618

Remus, J. J., & Collins, L. M. (2008). Comparison of adaptive psychometric procedures motivated by the Theory of Optimal Experiments: Simulated and experimental results. *The Journal of the Acoustical Society of America*, *123*(1), 315–326. https://doi.org/10.1121/1.2816567

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. a. (2006). Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. *The Journal of the Acoustical Society of America*, *120*(6), 3988–3997. https://doi.org/10.1121/1.2358008

Rieke, C. C., Clavier, O. H., Allen, L. V., Anderson, A. P., Brooks, C. A., Fellows, A.

M., … Buckey, J. C. (2017). Fixed-Level Frequency Threshold Testing for Ototoxicity Monitoring. *Ear and Hearing*, 1. https://doi.org/10.1097/AUD.0000000000000433

Rönnberg, J., Lunner, T., Ng, E. H. N., Lidestam, B., Zekveld, A. A., Sörqvist, P., … Stenfelt, S. (2016). Hearing impairment, cognition and speech understanding: exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. *International Journal of Audiology*, *55*(11), 623–642. https://doi.org/10.1080/14992027.2016.1219775

Ross, B., Tremblay, K. L., & Picton, T. W. (2007). Physiological detection of interaural phase differences. *The Journal of the Acoustical Society of America*, *121*(2), 1017–1027. https://doi.org/10.1121/1.2404915

Sanchez-Lopez, R., Bianchi, F., Fereczkowski, M., Santurette, S., & Dau, T. (2018). Data-Driven Approach for Auditory Profiling and Characterization of Individual Hearing Loss. *Trends in Hearing*, *22*(1), 233121651880740. https://doi.org/10.1177/2331216518807400

Sanchez-Lopez, R., Nielsen, S. G., El-Haj-Ali, M., Bianchi, F., Fereckzowski, M., Cañete, O., … Santurette, S. (2019). Data from "Auditory tests for characterizing hearing deficits: The BEAR test battery." https://doi.org/10.5281/zenodo.3459579

Sanders, N. C., & Chin, S. B. (2009). Phonological Distance Measures*. *Journal of Quantitative Linguistics*, *16*(1), 96–114. https://doi.org/10.1080/09296170802514138

Santurette, S., & Dau, T. (2012). Relating binaural pitch perception to the individual listener's auditory profile. *The Journal of the Acoustical Society of America*, *131*(4), 2968–2986. https://doi.org/10.1121/1.3689554

Saunders, G. H., Field, D. L., & Haggard, M. P. (1992). A clinical test battery for obscure auditory dysfunction (OAD): Development, selection and use of tests. *British Journal of Audiology*, *26*(1), 33–42. https://doi.org/10.3109/03005369209077869

Schorn, K., & Zwicker, E. (1990). Frequency selectivity and temporal resolution in patients with various inner ear disorders. *Audiology : Official Organ of the International Society of Audiology*, *29*(1), 8–20. https://doi.org/10.3109/00206099009081641

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, *122*, 105–123. https://doi.org/10.1016/j.visres.2016.02.002

Stratford, P. W., & Goldsmith, C. H. (1997). Use of the Standard Error as a Reliability Index of Interest : An Applied Example Using Elbow Flexor Strength Data. *Physical Therapy*, *77*(7), 745–750. https://doi.org/10.1093/ptj/77.7.745

Strelcyk, O., & Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *The Journal of the Acoustical Society of America*, *125*, 3328–3345. https://doi.org/10.1121/1.3097469

Strelcyk, O., Zahorik, P., Shehorn, J., Patro, C., & Derleth, R. P. (2019). Sensitivity to Interaural Phase in Older Hearing-Impaired Listeners Correlates With Nonauditory Trail Making Scores and With a Spatial Auditory Task of Unrelated Peripheral Origin. *Trends in Hearing*, *23*, 233121651986449. https://doi.org/10.1177/2331216519864499

Thorup, N., Santurette, S., Jørgensen, S., Kjærbøl, E., Dau, T., & Friis, M. (2016). Auditory profiling and hearing-aid satisfaction in hearing-aid candidates. *Danish Medical Journal*, *63*(10). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27697129

van Esch, T. E. M., & Dreschler, W. A. (2011). Measuring spectral and temporal resolution simultaneously: a comparison between two tests. *International Journal of Audiology*, *50*(7), 477–490. https://doi.org/10.3109/14992027.2011.572083

Vinay, Hansen, A. S., Raen, Ø., & Moore, B. C. J. (2017). Reference thresholds for the TEN(HL) test for people with normal hearing. *International Journal of Audiology*,

56(9), 672–676. https://doi.org/10.1080/14992027.2017.1307531

Vlaming, M. S. M. G., Kollmeier, B., Dreschler, W. A., Martin, R., Wouters, J., Grover, B., … Houtgast, T. (2011). HearCom: Hearing in the Communication Society. *Acta Acustica United with Acustica*, *97*(2), 175–192. https://doi.org/10.3813/AAA.918397

Wolff, A., Houmøller, S. S., Hougaard, D., Gaihede, M., Hammershøi, D., & Schmidt, J. (2020). Health-related Quality of Life in Hearing Impaired Danish Adults before and after Hearing Aid Rehabilitation. *International Journal of Audiology, (u. rev.).*

Zaar, J., Simonsen, L. B., Bherens, T., Dau, T., & Laugesen, S. (2019). Investigating the relationship between spectro-temporal modulation detection, aided speech perception, and noise reduction preference. *Proceedings of ISAAR 2019: Auditory Learning in Biological and Artificial Systems. Vol 7*. Nyborg (Denmark).