Development of the Adaptive Music Perception Test

Martin J. Kirchberger,[1,2] and Frank A. Russo[3,4]

Objectives: Despite vast amount of research examining the influence of hearing loss on speech perception, comparatively little is known about its influence on music perception. No standardized test exists to quantify music perception of hearing-impaired (HI) persons in a clinically practical manner. This study presents the Adaptive Music Perception (AMP) test as a tool to assess important aspects of music perception with hearing loss.

Design: A computer-driven test was developed to determine the discrimination thresholds of 10 low-level physical dimensions (e.g., duration, level) in the context of perceptual judgments about musical dimensions: meter, harmony, melody, and timbre. In the meter test, the listener is asked to judge whether a tone sequence is duple or triple in meter. The harmony test requires that the listener make judgments about the stability of the chord sequences. In the melody test, the listener must judge whether a comparison melody is the same as a standard melody when presented in transposition and in the context of a chordal accompaniment that serves as a mask. The timbre test requires that the listener determines which of two comparison tones is different in timbre from a standard tone (ABX design). Twenty-one HI participants and 19 normal-hearing (NH) participants were recruited to carry out the music tests. Participants were tested twice on separate occasions to evaluate test–retest reliability.

Results: The HI group had significantly higher discrimination thresholds than the NH group in 7 of the 10 low-level physical dimensions: frequency discrimination in the meter test, dissonance and intonation perception in the harmony test, melody-to-chord ratio for both melody types in the melody test, and the perception of brightness and spectral irregularity in the timbre test. Small but significant improvement between test and retest was observed in three dimensions: frequency discrimination (meter test), dissonance (harmony test), and attack length (timbre test). All other dimensions did not show a session effect. Test–retest reliability was poor (<0.6) for spectral irregularity (timbre test); acceptable (>0.6) for pitch and duration (meter test), dissonance and intonation (harmony test), melody-to-chord ratio I and II (melody test); and excellent (>0.8) for level (meter test) and attack (timbre test).

Conclusion: The AMP test revealed differences in a wide range of music perceptual abilities between NH and HI listeners. The recognition of meter was more difficult for HI listeners when the listening task was based on frequency discrimination. The HI group was less sensitive to changes in harmony and had more difficulties with distinguishing melodies in a background of music. In addition, the thresholds to discriminate timbre were significantly higher for the HI group in brightness and spectral irregularity dimensions. The AMP test can be used as a research tool to further investigate music perception with hearing aids and compare the benefit of different music processing strategies for the HI listener. Future testing will involve larger samples with the inclusion of hearing aided conditions allowing for the establishment of norms so that the test might be appropriate for use in clinical practice.

Key Words: Adaptive test, Hearing impairment, Hearing loss, Music perception.

(Ear & Hearing 2014;XX;00–00)

[1]Department of Health Sciences and Technology, ETH Zürich, Zurich, Switzerland; [2]Phonak AG, Stäfa, Switzerland; [3]Department of Psychology, Ryerson University, Toronto, Ontario, Canada; and [4]Communication Team, Toronto Rehabilitation Institute, Toronto, Ontario, Canada.

INTRODUCTION

Motivation

Hearing impairment affects the perception of sound leading to reductions in speech intelligibility and the enjoyment of music. Extensive research into the effects of hearing loss on speech intelligibility (e.g., Humes 1991; Barreto & Ortiz 2008) has led to various attempts to improve speech perception through hearing aids (Levitt 2001; Tawfik et al. 2009; Cornelis & Moonen 2012). By comparison, research into the effects of hearing loss on music perception is in its infancy. Hearing impaired (HI) listeners—unaided and aided—have specified various ways in which the enjoyment of music is affected: instruments are hard to distinguish, sounds are distorted, and melodies are difficult to recognize (Neukomm 2006; Sarchett 2006; Leek et al. 2008). There are no standardized tests available that have been designed to assess music perception in individuals with hearing loss adaptively. A large number of music perception tests are specifically designed to assess musicality or the skills of healthy and formally trained individuals (Woodruff 1983; Don et al. 1999). Another group of music tests focuses on individuals with cochlear implants (Gfeller et al. 2002; Gfeller et al. 2005; Medel Medical Electronics 2006; Cooper et al. 2008; Looi et al. 2008; Spitzer et al. 2008; Kang et al. 2009). The first music perception test battery to be specifically developed for hearing-aid users was the Music Perception Test (Uys & van Dijk 2011). This test battery includes 11 subtests that assess perception in four dimensions: rhythm (4 subtests), timbre (2 subtests), pitch (2 subtests), and melody (3 subtests). As the tests are not adaptive, its sensitivity to differences between NH and HI listeners could be somewhat limited.

Concept

Music can be broken down into a core set of components such as meter, harmony, melody, and timbre. Meter is the division of a rhythmic pattern according to equal periods or measures (Limb 2006). Strong and weak beats in the pattern are grouped into larger units (Oh 2008), typically categorized as duple or triple according to whether the measure is organized in groups of two or three beats. The perception of harmony corresponds to the vertical organization of pitch (Piston 1987). This organization concerns simultaneous pitches as in a chord or in counterpoint. Another important aspect of harmony concerns the sequencing of chords, as in a harmonic progression. A melody is a linear succession of tones, which is perceived as a single entity (Limb 2006). The succession of tones can be specified with regard to melodic contour and interval size (Davies & Jennings 1977). The former refers to the direction of the pitch change (Schubert 2006) and the latter refers to its magnitude. Timbre is often defined as the attribute of auditory perception whereby a listener can judge two sounds as dissimilar using any criterion other than pitch, loudness, and duration (ANSI 1960). This definition is somewhat unsatisfactory in that it defines timbre through what it is not

0196/0202/14/XXXX-0000/0 • Ear & Hearing • Copyright © 2014 by Lippincott Williams & Wilkins • Printed in the U.S.A.

<zdoi; 10.1097/AUD.0000000000000112>

rather than what it is (Sethares 2005). An alternative approach to defining timbre uses multidimensional scaling techniques to extract the underlying physical dimensions that contribute to its perception (Grey 1977; Krumhansl 1989; Krimphoff et al. 1994; McAdams et al. 1995). Brightness and attack are two dimensions that have been consistently identified as important using multidimensional scaling techniques (Grey 1977; Wessel 1979; Krumhansl 1989; McAdams & Cunibile 1992; Krimphoff et al. 1994; McAdams et al. 1995). The conclusions regarding a third dimension, however, differ. Within the framework of our study, we have defined a third dimension in accordance with the conclusions of Krimphoff et al. (1994) as spectral irregularity. The spectral irregularity is determined by deviation from linearity in the amplitude envelope of the harmonics. A clarinet would have high spectral irregularity because it lacks energy in the even harmonics. A trumpet tone by contrast would have low spectral irregularity as it possesses odd and even harmonics with a linear spectral roll-off.

Objective

We aimed to develop a computerized test that is adaptive and thus capable of assessing music perception in hearing as well as HI populations. The Adaptive Music Perception (AMP) test has been designed to investigate the perception of important music dimensions such as meter, timbre, harmony, and melody. These mid-level dimensions are constructed from 10 low-level dimensions D1 to D10 such as attack, level, and pitch (Fig. 1), which may ultimately be modified in a fairly direct manner through changes to the hearing aid technology. While the assessment is not entirely culturally neutral, there is no reason to expect that performance will vary across individuals who primarily consume music in the Western harmonic idiom regardless of genre preferences. Each test of the battery asks the listener to make a response concerning a mid-level dimension while adaptively manipulating the underlying low-level dimensions, thereby allowing for the determination of the respective low-level discrimination thresholds. Thus, thresholds are determined within a musically relevant framework. We contend that the use of these tests corresponds more directly to functional hearing in music than a conventional assessment of discrimination thresholds.

Hypothesis

Survey studies of individuals with hearing loss have revealed a wide range of problems encountered following the onset of hearing loss including melody recognition and level changes (Feldman & Kumpf 1988; Leek et al. 2008). We assume that these subjective impressions of problems can be objectively tracked with adaptive listening tests that are presented within a musical framework. We predict that individuals with hearing loss will have higher discrimination thresholds than individuals with NH.

MATERIALS AND METHODS

The AMP test comprises four subtests including meter, harmony, melody, and timbre discrimination.

General Test Design

Each subtest was designed to determine the discrimination thresholds of two or more low-level physical dimensions. In each trial, one of these low-level dimensions was manipulated

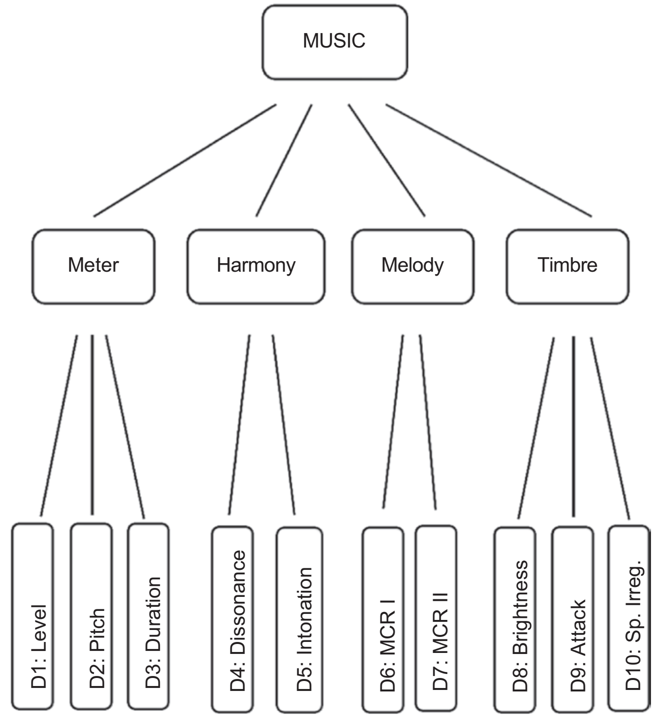

Fig. 1. The mid level and low-level dimensions of the Adaptive Music Perception test. D1–D10 indicates dimension 1–10; MCR, melody-to-chord ratio; Sp. Irreg., spectral irregularity.

so as to instantiate the mid level dimension. For example, the downbeat in the meter subtest could be conveyed by a difference in level, duration, or pitch. The adaptive process of each low-level dimension was interleaved to obscure the structure of the test and to discourage analytic modes of listening. We pseudorandomized the position of the correct answer with the constraint that the right answer was equally distributed among both answer buttons.

Adaptation

An experimental procedure is adaptive if the selection of the stimuli is determined by the results of the preceding trials (Treutwein 1995). Our motivation for making the AMP test adaptive was to measure performance in a manner that was fast and accurate for all degrees of hearing loss and music ability. The AMP test applies the transformed two-alternative-forced-choice method using a two-down one-up rule (Levitt 1971), which is commonly used in tests of auditory thresholds (e.g., Jesteadt et al. 1977; Wier 1977; Moore et al. 1984; Bochner et al. 1988). The presentation and adaptation of the low-level dimensions in each subtest were interleaved. The adaptation process was stopped, once a minimum of six reversal points were reached on each dimension. The standard step sizes were halved after the first and second reversals in an effort to accelerate the determination of thresholds without compromising accuracy. Discrimination thresholds were calculated by averaging the values of the last four reversals.

Sound Stimuli

Stimuli used across all four subtests were digitally synthesized. A spectrogram of the stimuli is displayed in Figure 2. The reference tone that was used for calibration and throughout the tests was a complex tone with the fundamental frequency $f_0 = 220$ Hz (pitch A3) and 40 harmonics. The starting phases of the 40 harmonics were randomized once and equal for all participants. The amplitude of each harmonic $(n = 0, 1, \ldots 39)$ was attenuated by $2\ \text{dB} \times n$ in relation to the amplitude of the fundamental f_0. The resulting spectral shape corresponds roughly with that of a trumpet tone (Krimphoff et al. 1994). The test stimuli were played back at 40 dB SL which is consistent with previous psychoacoustic studies (Jesteadt et al. 1977; Wier 1977; Moore et al. 1984; Bochner et al. 1988; Arehart 1994).

Test Procedure

The AMP test was implemented in MATLAB R2009a. Before testing with a participant commenced, the hearing thresholds were determined for the reference tone using a two-alternative forced-choice procedure. The presentation level of the reference tone in the test was then individually set for each participant at 40 dB SL. On average, the meter and timbre subtests took 10 min each; the harmony subtest took 12 min; and the melody subtest took 11 min. Each subtest could be interrupted and continued at any time. Each subtest was preceded by training trials with feedback to ensure that the task was understood. The initial discrimination differences of the low-level dimensions were made high enough so that the nature of the tasks did not generate confusion. To assess test–retest reliability for each subtest, participants were asked to complete testing on two separate sessions.

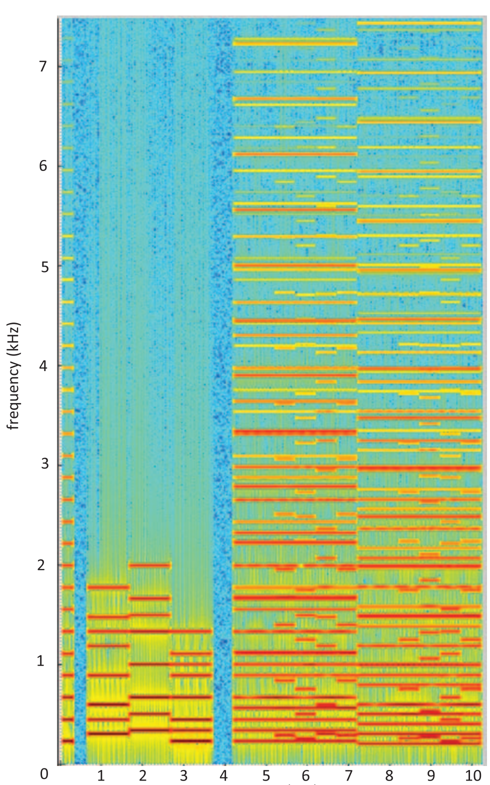

Fig. 2. Spectrogram of the Adaptive Music Perception test audio stimuli. The reference tone (0–0.4 sec) in the calibration, meter subtest, and timbre subtest is followed by the reference cadence of the harmony subtest (0.8–3.8 sec), and by one example of the melody subtest (4.2–10.2 sec).

Meter Subtest: Level, Pitch, and Duration

A sequence of six harmonic tones was played isochronously in each trial. The participant was asked to indicate whether the sequence constituted a duple or triple meter. The interonset interval between isochronous beats was 600 msec. The upbeats had the spectral characteristics of the calibration tone (fundamental frequency: 220 Hz, roll-off: −2 dB per harmonic) and a duration of 280 msec. As mentioned above, the initial phases are defined and originate from a one-time randomization before the test development. The downbeats are realized by adaptively introducing a difference from the upbeats with respect to level, pitch, or duration. Correspondingly, the dimensions for which thresholds are determined are labeled level (D1), pitch (D2), and duration (D3).

Rationale for Stimulus Design

If the interonset interval is lower than 100 msec, listeners perceive a sequence of tones as a single, continuous event. For intervals greater than 1.5 sec, grouping gets more difficult as the sounds seem disconnected (Fraisse 1978). In between these two limits, the distance of two events is overestimated for small intervals and underestimated for bigger intervals. The point of subjective equality, where subjective judgments corresponded to the objective duration, is reached at approximately 600 msec (Krumhansl 2000).

The sensation of pulse is also most salient for intervals of 600 msec (Parncutt 1994). To increase the sensitivity of the meter subtest, we therefore chose to use an interonset interval of 600 msec. The perceived meter of a sequence is generally established quite early after a couple of beats (Parncutt 1994). With regard to the test design, it is beneficial to keep the test duration short to avoid fatigue. In the meter subtest, only six beats are played back in each trial. This is the least common multiple of duple (two) and triple (three) meter and allows for three or two full measures respectively. According to participants' feedback in the pilot study, six beats are sufficient to detect the meter if the downbeat is audible.

Harmony Subtest: Dissonance and Intonation

Two cadences composed of three-dyad chords (root note and fifth) were played back in succession: subdominant (IV), dominant (V), and tonic (I) (Fig. 3). The participants were asked to indicate the cadence that appeared to be less resolved. The reference cadence was perfectly in tune, the second cadence was detuned. We used just intonation to define the frequencies of the reference cadence: the ratios of two notes in any interval were defined by the harmonic series and are related by small whole numbers. The frequency ratio of the root note and the fifth of every dyad was 3 : 2, such that the respective harmonics

Fig. 3. The subdominant (IV), dominant (V), and tonic (I) dyad of the harmony subtest.

coincide. Had we used equal temperament, the frequency ratio of the root note and the fifth would be $(\sqrt[12]{2})^7 = 1.4983$, and thus the harmonics would not coincide (Fig. 4). The frequency ratio between the subdominant and the tonic was $4:3$.

The root note of every dyad was composed of 6 harmonics with a roll-off factor of −2 dB per harmonic. The fifth of every dyad was composed of 4 harmonics with a roll-off factor of −3 dB per harmonic. The first harmonic of the fifth was attenuated by 1 dB compared to the first harmonic of the root note. By this means, the coinciding harmonics (root: harmonic 3/fifth: harmonic 2, root: harmonic 6/fifth: harmonic 4) were equal in amplitude. The initial phases of all harmonics were pseudorandomized with the provision that coinciding harmonics differed by 120°. In one of the two cadences, the tonic dyad was manipulated in one of two different ways: either the pitch of the fifth was increased or the pitch of the whole dyad was shifted. In the first case, the vertical structure of the tonic dyad was detuned causing higher sensory dissonance (Terhardt 1984; Tufts et al. 2005). In the second case, the vertical structure of the tonic dyad was in tune; however, the presented pitch did not correspond to the expected pitch in the given tonality (Piston 1987; Krumhansl 1990; Lerdahl 2001). The adaptive modifications that were applied to the pitch of the last dyad were within the bounds of a semitone. Thus, the basic meaning of the cadence was preserved (Blackwood 1985) and the pitch modifications implied a change in intonation. Accordingly, the dimensions for which thresholds were determined in the harmony subtest were labeled dissonance (D4) and intonation (D5).

Melody Subtest: Melody-to-Chord Ratio

Two 4-note melodies were played back and accompanied by a harmonic chord, which served as a mask. The task was to identify whether the contours of the two melodies were the same or different. The two melodies presented during the trial always started on different pitches (i.e., second melody was transposed). The transposition (pitch shift by constant interval) of the melodies was implemented to ensure that focusing on absolute pitches could not circumvent the melody recognition task. If the contours of the two melodies were the same, the second melody would be a copy of the first melody chromatically transposed by one whole tone. The contour of either the first or the second melody was made different by interchanging the second and the third tone of the respective melody.

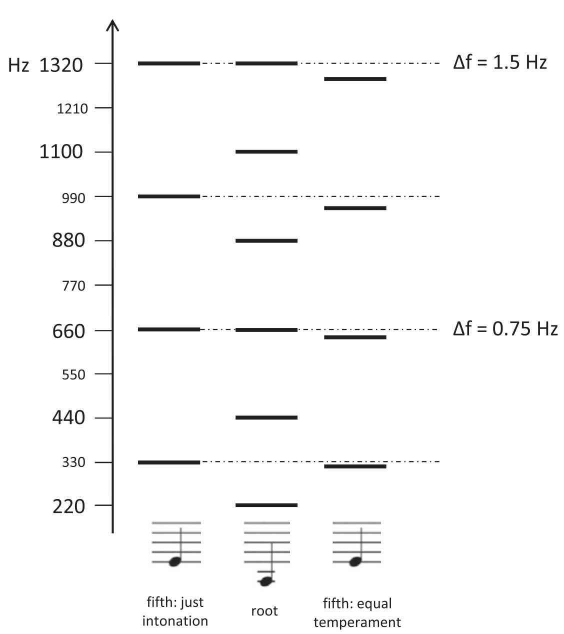

Fig. 4. A schematic plot of the harmonic spectrum of the root note and the fifth in just intonation or equal temperament. Deviations are exaggerated for better visibility.

Two sets of melodies, a lower and a higher one, were included in the subtest. The two sets as well as the accompanying A-major and G-major chords are depicted in Figure 5.

During the subtest, the sound level of the chords was kept constant and the levels of the melodies were adapted. For both melodies, the melody-to-chord ratio (MCRs) were determined indicating the point at which the melodies were perceptually masked by the chords. The dimensions are labeled MCR I (D6) and MCR II (D7) correspondingly.

Timbre Subtest: Brightness, Attack, and Spectral Irregularity

Three tones were played back in succession. Participants were asked to decide whether the third tone sounded like the first or the second (ABX design).

Stimuli Design

The tones differed in one of the three timbral dimensions: rapidity of attack, brightness, or spectral fine structure (Krimphoff et al. 1994). The deviations were introduced to a standard tone that matches the calibration tone in its spectral characteristics: the fundamental frequency (220 Hz), the number of harmonics (40) including the initial phases, and the roll-off factor (−2 dB). The duration of the tones was set to 1 sec. For the purposes of this test, attack was defined by the duration of the linear fade-in, brightness by the roll-off factor of the spectrum, and spectral irregularity by extent of attenuation in the even harmonics (Krimphoff et al. 1994; Jensen 2001). The dimensions of the timbre subtest were labeled brightness (D8), attack (D9), and spectral irregularity (D10).

Stimuli Calibration

To avoid perceptual interactions between the underlying low-level dimensions, a calibration session preceded the timbre subtest. The participant had to equalize the loudness of the standard tone with calibration tones covering the expected range of the difference values of the respective dimensions that are adapted in the test. By means of interpolation, the necessary amplification was calculated and applied to the modified tones before being played back. By this means, we intended to minimize loudness differences so that the presented tones were primarily distinguished by timbral characteristics.

Music Experience

Each participant completed a questionnaire to assess years of musical training, listening, and performing habits as well as a self-assessment of their musicality (Table 1, column 1 and 2). Points were allocated according to the answers (Table 1, column 3) and weighted with respect to their relative importance. For example, years of music training was deemed most important and thus received the greatest weight (Kang et al. 2009). The points were then summed to determine an overall measure of

Fig. 5. Two examples of the melody subtest. The upper bars depict two melodies from melody set 1 that have the same contour (up-down-up). The lower bars depict two melodies from melody set 2 that have a different contour (up-down-up; down-up-down).

music experience (ME). The theoretical range of ME is [−5;6]. The range and average of the normal-hearing (NH) participants was [−3;4] and 0.32 respectively; the range and average of the HI participants was [−3;4] and 0.21, respectively. A *t* test confirmed that the two groups did not differ with regard to ME: $F_{1,39} = 0.023$, $p = 0.881$.

Participants

Nineteen NH participants (P1–P19) and 21 HI participants (P20–P41) were recruited for this study (Table 2). Degree of hearing loss was determined according to the American Speech-Language-Hearing Association (i.e., average value of the hearing thresholds at 0.5 kHz, 1 kHz, 2 kHz, and 4 kHz across both ears). Participants with hearing loss less than 25 dB HL were assigned to the NH group, participants with hearing loss greater than or equal to 25 dB HL were assigned to the HI group. All audiometric data was assessed within 1 year of the test date. The age of the NH participants ranged from 27 to 67 years (M = 47.8 years), whereas the age of the HI participants ranged from 38 to 83 years (M = 64.7 years). As expected, the age difference between both groups was statistically significant: $t = 4.15$, $p < 0.001$.

RESULTS

On each low-level dimension D1–D10, a two factor, analysis of covariance (ANCOVA) was performed. The between-subjects factor (independent variable) was the hearing loss group (NH, HI). ME was entered as the covariate. The results are shown in Table 3. In addition, test–retest reliability was assessed using interclass correlation.

TABLE 1. The music questionnaire and the respective points allocation to determine an objective music experience measure for each participant

Question	Answer	Allocated Points
1 How many years of music training have you received? (a) Please indicate your instrument/voice training; (b) Please indicate your music theory or ear training; (c) Please indicate if and how long you have ever taught yourself to play an instrument? (Answer accumulated)	<1 yr	−4
	1–3 yrs	−2
	3–10 yrs	0
	>10 yrs	2
	Professional musician	4
Are you currently active musically (i.e., within the last year)?	No	0
	Yes	1
Do you consider yourself musical?	No	−0.5
	Somewhat	0
	Yes	0.5
How many hours do you listen to music (average per day)?	<1 hr	−0.5
	1–3 hrs	0
	>3 hrs	0.5

The NH and HI group not only differed in hearing loss but also in age. To account for this age difference, we ran a forced-entry multiple regression analysis with age, hearing loss, and ME as independent variables. Test results are displayed in Table 4. Because of the correlation of the predictor variables HL and age in the multiple regression analysis, we performed a multicollinearity diagnostic. Kutner et al. (2004) proposed values above 10 as critical. In all four subtest data sets, the variance inflation factor values were smaller than 2.5. Thus, we can assume that multicollinearity does not play a role in the multiple regression analysis of our test data. With the exception of results that are of particular theoretical interest, only those results that reached significance are reported.

Meter Subtest

All 19 NH and 21 HI participants completed the meter subtest in both session. Figure 6 depicts the results obtained for level (D1), pitch (D2), and duration (D3) as boxplots. The NH group performed better on D2 (pitch): $F_{1,39} = 15.57, p < 0.001$. The HI group performed marginally better on D1, however, this trend did not reach significance: $F_{1,39} = 1.95$, $p = 0.170$. The ANCOVA revealed a significant effect of ME on D1, $F_{1,39} = 9.09$, $p = 0.005$ and D2, $F_{1,39} = 6.033$, $p = 0.019$, but not on D3, $F_{1,39} = 1.87, p = 0.180$. A session effect was observed on D2 $F_{1,39} = 9.62, p = 0.004$. The interclass correlations were 0.904 (D1), 0.753 (D2), and 0.739 (D3). The multiple regression confirms a significant effect of hearing loss on pitch discrimination (D2: $p_{HL} = 0.008$, $\beta_{HL} = 0.498$). The multiple regression also confirms ME as a significant factor for level discrimination (D1: $p_{ME} = 0.001$, $\beta_{ME} = -0.508$) and pitch discrimination (D2: $p_{ME} = 0.014, \beta_{ME} = -0.329$).

Harmony Subtest

Four NH participants and two HI participants were excluded. One NH participant was excluded because they did not understand the task even after extensive training. Three NH participants could not complete either the test or the retest because of scheduling conflicts. Two HI persons indicated that their power of concentration had decreased over the course of testing. Fifteen NH participants and 19 HI participants conducted the harmony subtest in both sessions. Figure 7 depicts the results obtained for dissonance (D4) and intonation (D5) as boxplots. NH participants outperformed HI participants on D4, $F_{1,33} = 4.87$, $p = 0.035$, and D5, $F_{1,33} = 8.63$, $p = 0.006$.

TABLE 2. Characteristics of the test participants: sex, audiometric data, and HL as defined by American Speech-Language-Hearing Association, ME, and age

		Left Ear				Right Ear						
Participants	Sex	0.5k	1k	2k	4k	0.5k	1k	2k	4k	HL	ME	Age
P1	M	−10	−5	0	−5	−10	0	−5	−5	−5.0	−1	28
P2	M	−5	−5	−5	0	0	0	0	−5	−2.5	2	27
P3	M	0	−5	0	−5	−5	−5	−5	10	−1.9	1	30
P4	M	0	0	−5	0	0	0	0	5	0.0	−1	33
P5	M	5	5	0	10	10	0	−5	−5	2.5	−2	50
P6	M	5	5	5	0	5	5	5	0	3.8	−1	36
P7	F	0	0	0	10	5	5	10	10	5.0	0	51
P8	M	5	5	10	10	0	0	10	5	5.6	2	33
P9	M	5	0	5	20	5	5	0	15	6.9	2.5	51
P10	M	5	5	15	15	5	0	10	5	7.5	4	45
P11	M	15	5	5	15	10	5	5	20	10.0	1.5	47
P12	M	10	10	10	40	5	5	5	20	13.1	3	59
P13	M	15	10	15	15	15	5	15	25	14.4	−2	45
P14	M	5	0	5	25	15	5	10	60	15.6	−0.5	67
P15	M	5	5	25	30	5	5	25	30	16.3	2	63
P16	M	20	15	15	30	20	20	15	35	21.3	−2	55
P17	M	0	5	25	70	5	5	20	50	22.5	0	65
P18	M	5	5	10	65	10	10	15	65	23.1	0.5	67
P19	M	15	20	20	45	10	15	20	40	23.1	−3	57
P20	M	5	15	35	60	10	10	30	50	26.9	−1	61
P21	M	15	15	35	45	10	25	30	45	27.5	3	66
P22	M	15	15	20	60	15	20	20	60	28.1	2	51
P23	M	5	5	30	70	5	5	60	75	31.9	−2	68
P24	M	20	20	30	60	20	25	35	50	32.5	−0.5	68
P25	M	20	30	45	50	20	30	40	55	36.3	3.5	76
P26	M	30	20	55	75	30	20	5	60	36.9	1	63
P27	F	40	40	30	45	45	40	25	30	36.9	0	45
P28	M	25	30	40	50	25	35	35	55	36.9	−2.5	69
P29	M	5	0	0	45	45	65	75	65	37.5	4	76
P30	F	30	35	40	30	30	35	50	50	37.5	2.5	74
P31	M	30	40	50	60	30	35	45	45	41.9	−0.5	77
P32	M	50	50	60	85	15	15	25	45	43.1	−3	64
P33	M	15	30	60	70	10	35	60	70	43.8	1	73
P34	M	45	50	45	55	40	40	35	45	44.4	−1.5	47
P35	M	5	15	45	75	15	50	70	85	45.0	−3	38
P36	M	25	35	55	60	25	30	55	80	45.6	−2.5	64
P37	F	45	55	70	85	25	35	40	50	50.6	3.5	68
P38	M	50	60	60	50	45	60	60	55	55.0	1	50
P39	M	55	55	60	55	55	65	65	65	59.4	−1.5	77
P40	M	55	65	65	80	50	55	65	75	63.8	1	83

HL, degree of hearing loss; ME, music experience.

ME was positively related to performance on D5: $F_{1,33} = 4.26$, $p = 0.047$. The effect of session was significant on D4, $F_{1,33} = 4.50$, $p = 0.042$. The interclass correlations were 0.805 (D4) and 0.824 (D5). The multiple regression reaffirms a significant effect of hearing loss and ME on intonation (D5: $p_{HL} = 0.010$, $\dagger_{HL} = -0.499$; $p_{ME} = 0.02$, $\beta_{ME} = -0.352$).

Melody Subtest

Five NH participants and seven HI participants were excluded because they found the task too challenging. These participants could not detect reliably whether the melodies were different even in the absence of background noise. Fourteen NH participants and 14 HI participants completed the melody subtest in both sessions. Figure 8 depicts the MCRs obtained for both melody sets (D6 and D7) as boxplots. HI participants performed worse in melody set I, $F_{1,27} = 7.87, p = 0.010$ and melody set II, $F_{1,27} = 10.29, p = 0.004$. There were no significant effects of ME or session. The interclass correlations were 0.809 (D6) and 0.838 (D7). The multiple regression confirms a significant effect of hearing loss on both MCRs (D6: $p_{HL} = 0.034$, $\beta_{HL} = -0.565$; D7: $p_{HL} = 0.003$, $\beta_{HL} = -0.777$).

Timbre Subtest

One NH person was excluded because the participant could not conduct the retest session because of a scheduling conflict. Eighteen NH and 21 HI participants completed the timbre subtest in both sessions. Figure 9 depicts the discrimination thresholds obtained for brightness (D8), attack (D9), and spectral irregularity (D10) as boxplots. NH participants outperformed HI participants on D8, $F_{1,38} = 7.43$, $p = 0.010$ and D10,

TABLE 3. Results of the repeated-measure analysis of covariance for each dimension of the tests

	Hearing Loss	Music Experience	Session Effect	Hearing Loss x Session
D1: Level	$F_{1,39} = 1.95$ $p = 0.170$	$F_{1,39} = 9.09$ $p = 0.005$	$F_{1,39} = 2.21$ $p = 0.146$	$F_{1,39} = 0.68$ $p = 0.414$
D2: Pitch	$F_{1,39} = 15.57$ $p < 0.001$	$F_{1,39} = 6.033$ $p = 0.019$	$F_{1,39} = 9.62$ $p = 0.004$	$F_{1,39} = 2.97$ $p = 0.093$
D3: Duration	$F_{1,39} = 1.11$ $p = 0.300$	$F_{1,39} = 1.87$ $p = 0.180$	$F_{1,39} = 0.51$ $p = 0.480$	$F_{1,39} = 0.098$ $p = 0.756$
D4: Dissonance	$F_{1,33} = 4.87$ $p = 0.035$	$F_{1,33} = 1.41$ $p = 0.244$	$F_{1,33} = 4.50$ $p = 0.042$	$F_{1,33} = 0.511$ $p = 0.480$
D5: Tonality	$F_{1,33} = 8.63$ $p = 0.006$	$F_{1,33} = 4.26$ $p = 0.047$	$F_{1,33} = 3.78$ $p = 0.061$	$F_{1,33} = 0.013$ $p = 0.911$
D6: MCR I	$F_{1,27} = 7.87$ $p = 0.010$	$F_{1,27} = 0.58$ $p = 0.455$	$F_{1,27} = 0.05$ $p = 0.824$	$F_{1,27} = 0.005$ $p = 0.945$
D7: MCR II	$F_{1,27} = 10.29$ $p = 0.004$	$F_{1,27} = 1.93$ $p = 0.178$	$F_{1,27} = 1.95$ $p = 0.175$	$F_{1,27} = 0.156$ $p = 0.697$
D8: Brightness	$F_{1,38} = 7.43$ $p = 0.010$	$F_{1,38} = 0.034$ $p = 0.856$	$F_{1,38} = 0.764$ $p = 0.388$	$F_{1,38} = 0.001$ $p = 0.971$
D9: Attack	$F_{1,38} = 1.16$ $p = 0.289$	$F_{1,38} = 3.77$ $p = 0.060$	$F_{1,38} = 4.80$ $p = 0.035$	$F_{1,38} = 0.495$ $p = 0.486$
D10: Spect. Irreg.	$F_{1,38} = 5.98$ $p = 0.019$	$F_{1,38} = 1.88$ $p = 0.180$	$F_{1,38} = 1.78$ $p = 0.190$	$F_{1,38} = 0.383$ $p = 0.540$

The condition hearing loss was the dependent variable, music experience the covariate.
MCR, melody-to-chord ratio; Sp. Irreg., spectral irregularity.

$F_{1,38} = 5.98, p = 0.019$. The effect of session was significant on D9, $F_{1,38} = 4.80, p = 0.035$. The interclass correlations were 0.821 (D8), 0.937 (D9), and 0.712 (D10). The multiple regression analysis did not yield any significant results.

DISCUSSION

We have developed an Adaptive Music Perception test for HI listeners. Results of the AMP test showed a significant effect of hearing loss on the perception of important musical aspects such as meter, harmony, melody, and timbre. In 7 of the 10 adapted underlying low-level dimensions, the ANCOVA revealed that HI participants performed significantly worse than their NH counterparts: pitch (D2, meter subtest), dissonance (D4, harmony subtest), intonation (D5, harmony subtest), MCR I (D6, melody subtest), MCR II (D7, melody subtest), brightness (D8, timbre subtest), and spectral irregularity (D10, timbre subtest). There was no significant effect of hearing loss on level (D1, meter subtest), duration (D3, meter subtest), or attack (D9, timbre subtest). After taking age into account using a multiple regression approach to analyses, hearing loss remained a significant factor in pitch discrimination (D2, meter subtest), intonation (D5, harmony subtest), and melody detection (D6 & D7, melody subtest). The next section of the discussion considers thresholds obtained for each low-level dimension and how they compare to existing findings obtained using conventional threshold testing—that is threshold tests that are not embedded in a musical task as they are here.

D1: Level, Meter Subtest

Jesteadt et al. (1977) used sinusoidal tones to determine level discrimination thresholds. An average threshold of 1 dB was obtained for NH participants at 40 dB SL. As in other studies (Schacknow & Raab 1973; Penner et al. 1974; Fastl & Schorn 1981), the threshold was found to be independent of frequency. In the present study, the stimuli were also presented at a sensation level of 40 dB and we obtained an average

TABLE 4. Forced-entry multiple linear regression for each dimension tested

	R^2 adj	p	p_{HL}	β_{HL}	p_{AGE}	β_{AGE}	p_{ME}	β_{ME}
D1	0.255	0.003	0.114	−0.323	0.829	0.044	0.001	−0.508
D2	0.420	0.000	0.008	0.498	0.510	0.666	0.014	−0.329
D3	−0.026	0.573	0.729	−0.082	0.389	−0.206	0.225	−0.210
D4	0.044	0.234	0.349	0.228	0.735	0.081	0.322	−0.189
D5	0.445	0.000	0.010	0.499	0.657	0.081	0.020	−0.352
D6	0.289	0.011	0.034	0.565	0.834	0.054	0.985	−0.003
D7	0.388	0.002	0.003	0.777	0.296	−0.251	0.222	−0.197
D8	0.227	0.007	0.288	0.215	0.076	0.371	0.830	−0.032
D9	0.038	0.230	0.629	0.108	0.784	0.062	0.078	−0.304
D10	0.122	0.056	0.181	0.290	0.671	0.092	0.128	−0.249

Predictors were HL, age, and ME.
HL, degree of hearing loss; ME, music experience.

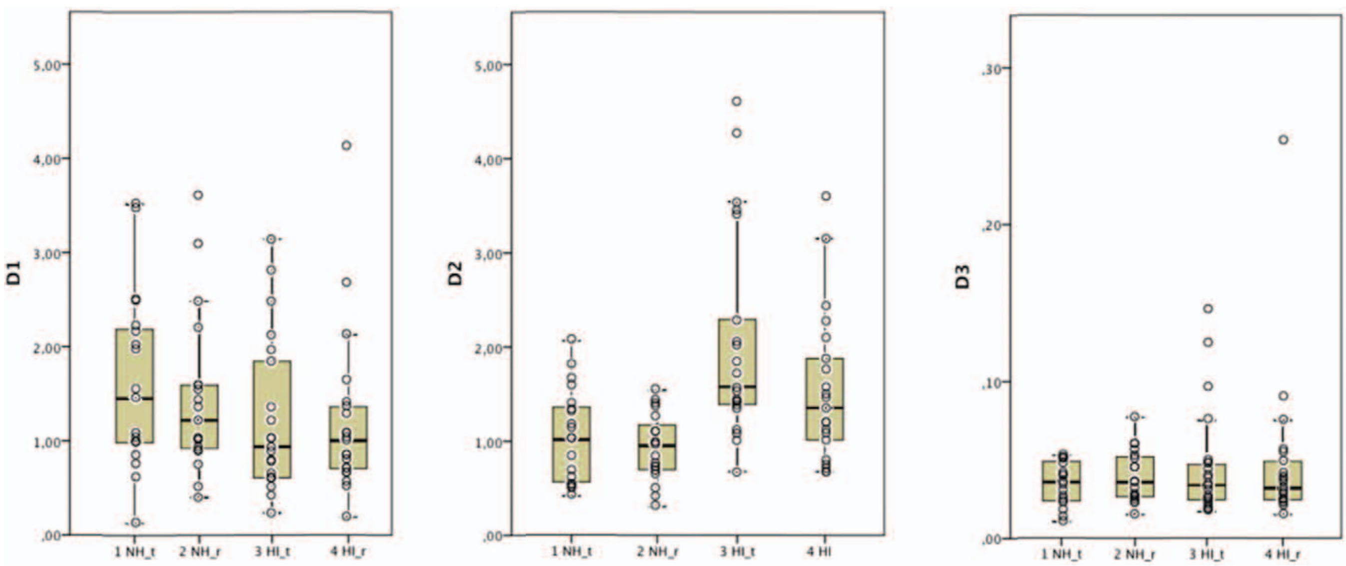

Fig. 6. Box-plot showing the level D1 (dB), the pitch D2 (Hz), and the duration D3 (sec) thresholds of the meter subtest by group and session. 1 NH_t indicates normal hearing/test session; 2 NH_r, normal hearing/retest session; 3 HI_t, hearing impaired/test session; 4 HI_r, hearing impaired/retest session.

threshold of 1.5 dB for NH participants and 1.2 dB for HI participants. The slightly larger thresholds obtained in the current study are likely due to the use of complex tones and the presentation of a stimuli within a musical task. Several authors have proposed that level discrimination should be better in HI listeners because of recruitment and abnormal loudness growth (Luscher & Zwislocki 1949; Denes & Naunton 1950). Findings on this question, however, have been mixed. Fastl and Schorn (1981) investigated level discrimination thresholds of HI listeners with different kinds of hearing disorders such as conductive, noise-induced presbycusis at two distinct frequencies 400 Hz and 5 kHz. In all groups, average thresholds were higher than those of the NH control group. In our study, the average level threshold of the HI group was marginally smaller than the threshold of the NH group but it did not reach significance, due in part to large individual variability that is typically found in level discrimination tasks (Rabinowitz et al. 1976; Houtsma et al. 1980).

D2: Pitch, Meter Subtest

Several psychophysical studies of complex-tone pitch discrimination have shown that discrimination thresholds in NH listeners are lower than they are in HI listeners. (Hoekstra & Ritsma 1977; Hoekstra 1979; Horst 1987; Moore & Glasberg 1988, 1990; Moore & Peters 1992). Discrimination thresholds for NH listeners range from 0.2 Hz (0.1%) (Hoekstra 1979) to 1.7 Hz (0.85%) (Moore & Peters 1992) depending on the method. The results of our study are comparable with an average discrimination threshold of 1.0 Hz (0.45%) for NH listeners and 1.8 Hz (0.8%) for HI listeners. Nevertheless, some caution is advisable in the use of the pitch discrimination dimension of the meter subtest given the significant session effect.

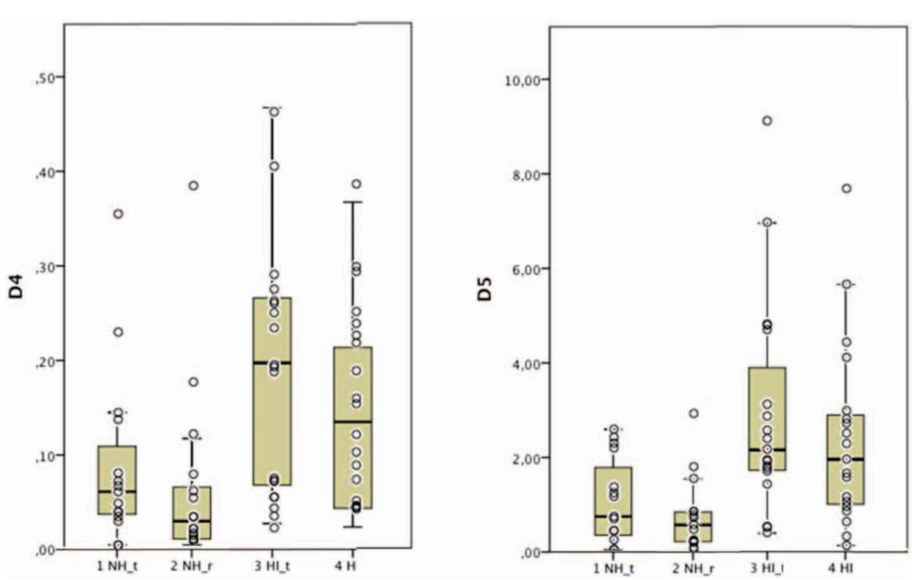

Fig. 7. Box-plot showing the dissonance D4 (Hz) and the intonation D5 (Hz) thresholds of the harmony subtest by group and session. 1 NH_t indicates normal hearing/test session, 2 NH_r, normal hearing/retest session, 3 HI_t, hearing impaired/test session, 4 HI_r, hearing impaired/retest session.

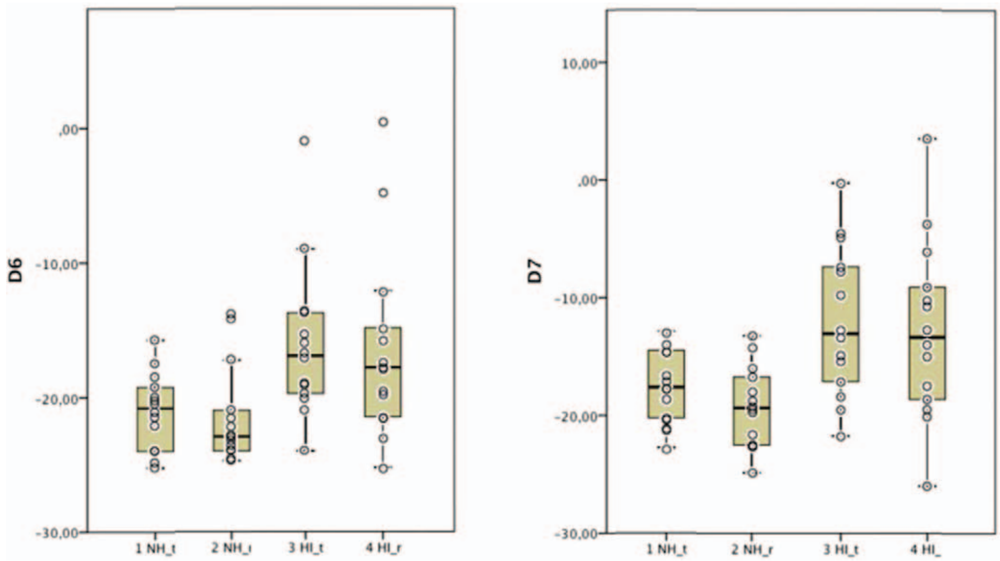

Fig. 8. Box-plot showing the melody-to-chord ratio I D6 (dB) and the melody-to-chord ratio II D7 (dB) thresholds of the melody subtest by group and session. 1 NH_t indicates normal hearing/test session; 2 NH_r, normal hearing/retest session;3 HI_t, hearing impaired/test session; 4 HI_r, hearing impaired/retest session.

D3: Duration, Meter Subtest

A trend was observed for the average discrimination threshold of the HI listeners to be larger than that of the NH listeners. The trend, however, was not significant. These results align with Ruhm et al. (1966) who concluded that hearing impairment has no effect on the discrimination of acoustic signals on the basis of duration. Tyler et al. (1982), however, found that duration discrimination was generally poorer in HI than in NH listeners. Bochner et al. (1988) conducted experiments to compare the temporal sensitivity of NH and HI listeners for static and transient sounds. Depending on the type of sound, conclusions regarding duration discrimination thresholds vary. Whereas NH listeners performed significantly better for static sounds shorter than 125 msec, the groups did not differ for transient sounds or static sounds longer than 200 msec. Our experiment falls into the latter category given that it utilized static sounds with a minimum duration of 280 msec. Like Bochner et al. (1988), hearing loss did not have a significant effect on discrimination thresholds for duration.

D4: Dissonance, Harmony Subtest

Discrimination thresholds for dissonance were significantly larger for HI than for NH listeners according to the ANCOVA analysis. This finding is consistent with results reported by Tufts et al. (2005). Stimuli in the Tufts et al. study resembled those used in the present study, that is, dyads composed of complex tones with 5 to 6 harmonics. The discrimination thresholds in our study were remarkably small: the average thresholds for HI and NH were 0.16 and 0.08 Hz respectively. The detection of beats due to amplitude modulation accounts for these thresholds. We use natural tuning for the reference dyad, that is the frequency ratio between the fifth and the root is 2:3 and the

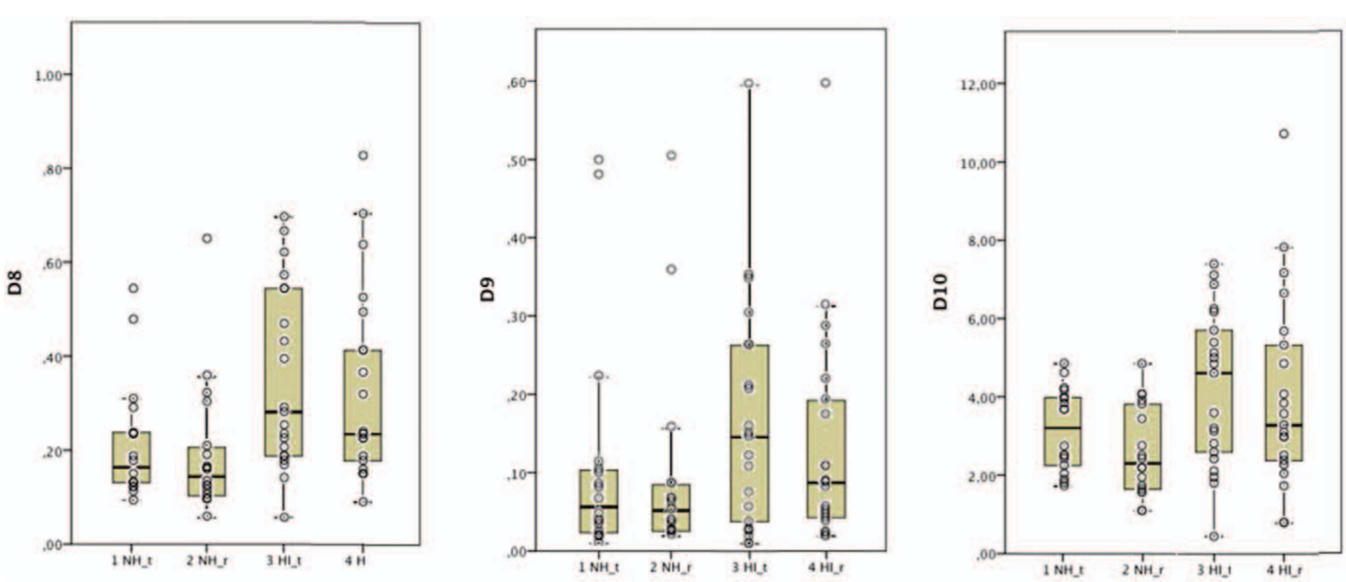

Fig. 9. Box-plot showing the brightness D8 (dB), attack D9 (Hz), and spectral irregularity D10 (dB) thresholds of the timbre subtest by group and session. 1 NH_t, normal hearing/test session; 2 NH_r, normal hearing/retest session; 3 HI_t, hearing impaired/test session; 4 HI_r, hearing impaired/retest session.

harmonics completely coincide. Thus, the dyad is very sensitive to modifications compared to a dyad in equal temperament whose harmonics do not coincide (c.f. Fig. 4). The forced-entry multiple regression analysis, however, does not reveal a significant effect of hearing loss on the perception of dissonance in the test results. On the basis of this result and the significant session effect, we would advise caution in the interpretation of the test results from the dissonance dimension (D4).

D5: Intonation, Harmony Subtest

Sensitivity to intonation seems to be adversely affected by hearing loss. The HI group had significantly larger thresholds than the NH group and the multiple regression analysis revealed hearing impairment as the main predictor of the intonation thresholds. We assume that the reduction in sensitivity stems from poorer F0 discrimination (Moore & Peters 1992).

D6 and D7: MCR I and II, Melody Subtest

Thirty percent of the participants were not able to carry out the melody subtest. For those participants who were able to do so, hearing loss had a detrimental effect. The average MCR of the HI group was 5 dB (MCR I) and 6 dB (MCR II) higher than the NH group. This outcome appears to be compatible with speech-in-noise thresholds. Listeners with moderate hearing loss (40–60 dB HL) have signal-to-noise ratio losses between 2 and 20 dB (Killion et al. 2004). Because poorer frequency selectivity contributes to more masking (Moore 2003), HI listeners need higher MCRs than their NH counterparts.

D8: Brightness, Timbre Subtest

The HI group was less sensitive to changes in brightness than the NH group. The multiple regression analysis revealed the same trend, but did not yield a significant effect of hearing impairment on brightness perception. Emiroglu (2007) also conducted experiments comparing brightness perception of NH and HI listeners. The just noticeable differences (JNDs) of listeners possessing hearing loss with steep slope were significantly higher than the JNDs of NH listeners. HI listeners possessing a flat or shallow slope also had higher JNDs but the trend did not reach significance. Nevertheless, Emiroglu (2007) also observed cases where HI listeners performed better than NH listeners. According to her theory, the internal representation of brightness can be enlarged due to recruitment so that smaller differences may be perceived more easily. While recruitment may contribute positively to brightness perception, the broadening of auditory filters in HI listeners should detract (Leek & Summers 1993).

D9: Attack, Timbre Subtest

Hearing loss does not seem to effect the discrimination of attack. Like duration (D3), attack has a temporal component in that we modify attack by adapting the attack time. However, it differs from D3 in that a change in duration also implies a change in the steepness of the loudness curve. Therefore, it may be assumed, that a higher sensitivity to loudness changes might increase the sensitivity to attack length. Consequently, this might lead us to expect that a HI person with smaller level discrimination thresholds would also perform better in this dimension. However, the variability in the data is quite high and the expected trend is not evident.

D10: Spectral Irregularity, Timbre Subtest

The HI listeners were significantly less sensitive to changes in spectral irregularity than their NH counterparts. We attribute this difference to reduced frequency selectivity (Oxenham & Bacon 2003). The multiple regression analysis, however, did not yield a significant effect for hearing loss. Thus it remains possible that the effect of hearing impairment was driven by age differences across the groups.

The results of this investigation show that ME had a positive effect on a subset of the dimensions tested. Music training requires the listener to become sensitive to subtle changes in loudness and pitch. Therefore, it is quite understandable that the effect of ME was significant in the level (D1) and pitch (D2) dimensions of the meter subtest as well as the intonation dimension (D5) of the harmony subtest. The quality of the test or the statements regarding the effect of hearing loss, however, are not compromised by this effect of ME. First, both the ANCOVA analysis and the multiple regression analysis separate the effect of ME from the effect of hearing loss. More importantly, these tests are designed to be used to assess different hearing aid algorithms for their applicability for music processing. For instance, we intend to improve the gain model for hearing aids and evaluate the benefit of new designs and parameterizations by using the AMP test. Therefore, under typical use, the results will be analyzed within and not between subjects. We acknowledge that the sample recruited for this study was not ideal. In particular, the age difference between groups limits the interpretability of the ANCOVA results. However, the regression analysis indicates that hearing loss independently contributed to variance in four of the test dimensions (D2, D5, D6, D7). Moreover, the proportion of variance accounted for by hearing loss in these dimensions was larger than that accounted for by age.

Test–retest reliability was excellent (>0.8) for 2 of 10 dimensions: level (D1) and attack (D9). Acceptable test–retest reliability (>0.6) was found in 7 of 10 dimensions: pitch (D2), duration (D3), dissonance (D4), intonation (D5), MCR I & II (D6&D7), and brightness (D8). Poor test–retest reliability (<0.6) was only found for spectral irregularity (D10). In retrospect, it seems that extending our training session would have been beneficial, leading to fewer instances of significant differences between test and retest. We observed a significant session effect for the dimension pitch (D2), dissonance (D4), and attack (D9). With regard to pitch, we assume that the fact that the meter subtest is conducted first, might contribute to the session effect. The participants are not used to a laboratory environment and they might be distracted and nervous in the beginning. Level (D1) and duration (D3) of the meter subtest, however, did not show a session effect. We therefore assume that pitch (D2) needs longer training or even an additional training session beforehand to minimize practice effects. With regard to dissonance (D4), we assume that participants need more time to fully grasp the task that the dissonant chord is less resolved. Again, more training might help to get the participants more familiar with the task even in the first session. Regarding attack (D9), we assume that it might not have been obvious for each participant in the first session, that this dimension is part of timbre perception. It differs from the other two timbre dimensions—brightness (D8) and spectral irregularity (D10)—in that it focuses on the temporal aspect rather than the spectral aspect of the stimuli. To minimize a potential session effect, we advise to counterbalance the presentation order of the different conditions under test. If

AMP is used to compare the benefit of two algorithms for music perception, half of the participants should start with algorithm 1, the other half with algorithm 2.

Future versions of this test may reduce the number of subtests so as to focus on those dimensions that are most relevant to music listening with an impairment and to make administration of the test more time efficient. On the basis of the current results, it seems that the duration dimension in the meter subtest and the attack dimension in the timbre subtest may be expendable. Neither dimension is affected by hearing loss. As a consequence of cutting these dimensions, the meter and timbre subtest would also become shorter allowing for more reversals to establish thresholds. Additional reversals should help to make thresholds more precise and to reduce test–retest variability.

We acknowledge that the melody subtest may be challenging for a significant proportion of individuals. Nevertheless, it is certainly an interesting tool for those who are very concerned with music. The main difficulty for those who could not grasp the task was to understand the concept of transposition. Future use of this test might circumvent the problem by omitting the transposition. We introduced this concept to ensure that participants did not focus on absolute pitches but listened to the melody as a whole.

We assume that the state-of-the-art hearing aid signal processing is not optimal for music listening as it is designed to improve speech intelligibility. At various points in the signal processing chain, we imagine that simple changes could be made to improve music processing. The gain model, for example, might be optimized to enhance the perception of timbre by emphasizing spectral differences. Neighboring bands of the filter bank could be steered less independently to allow the preservation of spectral contrast. The compression ratio and the time constants of the gain model decisively change the perception of the attack of a signal. Level changes are also inherently modified by the compression ratios. Pitch/F_0 discrimination is enhanced if audibility, especially of the higher frequencies, is ensured (Arehart 1994). The AMP test may be used to systematically assess the efficacy of these types of changes in signal processing strategy.

We developed an adaptive test battery that assesses the perception of important music aspects such as meter, harmony, melody, and timbre objectively. We were able to reveal differences in music perception between HI and NH participants. The AMP test can be used as a research tool to further investigate music perception with hearing aids and compare the benefit of different music processing strategies for the HI listener. Future testing will involve larger samples with the inclusion of hearing aided conditions allowing for the establishment of norms so that the test might be appropriate for use in clinical practice.

ACKNOWLEDGMENTS

The research protocol for this study was approved by grant (EK 2013-N-19) from the Ethics Commission of the ETH Zurich.

This research was supported by Phonak AG.

The authors declare no other conflict of interest.

Address for correspondence: Martin J. Kirchberger, Department of Health Sciences and Technology, ETH Zürich, 8092 Zürich, Switzerland. E-mail: kirmarti@student.ethz.ch

Received January 17, 2014; accepted August 22, 2014.

REFERENCES

ANSI (1960). *USA Standard Acoustical Terminology*. New York, NY: American National Standards Institute.

Arehart, K. H. (1994). Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners. *J Acoust Soc Am*, *95*, 3574–3585.

Barreto, S., & Ortiz, K. (2008). Intelligibility measurements in speech disorders: A critical review of literature. *Pró Fono Revista de Atualização*, *20*, 201–206.

Blackwood, E., (1985). *The Structure of Recognizable Diatonic Tunings*. Princeton, NJ: Princeton University Press.

Bochner, J. H., Snell, K. B., MacKenzie, D. J. (1988). Duration discrimination of speech and tonal complex stimuli by normally hearing and hearing-impaired listeners. *J Acoust Soc Am*, *84*, 493–500.

Cooper, W. B., Tobey, E., Loizou, P. C. (2008). Music perception by cochlear implant and normal hearing listeners as measured by the montreal battery for evaluation of amusia. *Ear Hear*, *29*, 618–626.

Cornelis, B., Moonen, M., Wouters, J. (2012). Speech intelligibility improvements with hearing aids using bilateral and binaural adaptive multichannel Wiener filtering based noise reduction. *J Acoust Soc Am*, *131*, 4743–4755.

Davies, J. B., & Jennings, J. (1977). Reproduction of familiar melodies and the perception of tonal sequences. *J Acoust Soc Am*, *61*, 534–541.

Denes, P., & Naunton, R. F. (1950). The clinical detection of auditory recruitment. *J Laryngol Otol*, *64*, 375–398.

Don, A. J., Schellenberg, E.G., Rourke, B. P. (1999). Music and language skills of children with Williams syndrome. *Child Neuropsychol*, *5*, 154–170.

Emiroglu, S. S. (2007). *Timbre Perception and Object Separation with Normal and Impaired Hearing*. Oldenburg, Germany: University of Oldenburg.

Fastl, H., & Schorn, K. (1981). Discrimination of level differences by hearing-impaired patients. *Audiology*, *20*, 488–502.

Feldmann, H., & Kumpf, W. (1988). [Listening to music in hearing loss with and without a hearing aid]. *Laryngol Rhinol Otol (Stuttg)*, *67*, 489–497.

Fraisse, P. (1978). Time and rhythm perception. In E. C. Carterette & M. P. Friedmann (Eds.), *Handbook of Perception* (Vol. 8) (pp. 203–254), New York, NY: Academic Press.

Gfeller, K., Olszewski, C., Rychener, M., et al. (2005). Recognition of “real-world” musical excerpts by cochlear implant recipients and normal-hearing adults. *Ear Hear*, *26*, 237–250.

Gfeller, K., Witt, S., Woodworth, G., et al. (2002). Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Ann Otol Rhinol Laryngol*, *111*, 349–356.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *J Acoust Soc Am*, *61*, 1270–1277.

Hoekstra, A. (1979). *Frequency Discrimination and Frequency Analysis in Hearing*. Ph.D. thesis, Groningen, The Netherlands: University of Groningen.

Hoekstra, A., & Ritsma, R. J. (1977). Perceptive hearing loss and frequency selectivity. In E. F. Evans & J. P. Wilson (Eds.), *Psychophysics and Physiology of Hearing* (pp. 263–271). London, UK: Academic Publisher.

Horst, J. W. (1987). Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects. *J Acoust Soc Am*, *82*, 874–885.

Houtsma, A. J., Durlach, N. I., Braida, L. D. (1980). Intensity perception XI. Experimental results on the relation of intensity resolution to loudness matching. *J Acoust Soc Am*, *68*, 807–813.

Humes, L. E. (1991). Understanding the speech-understanding problems of the hearing impaired. *J Am Acad Audiol*, *2*, 59–69.

Jensen, K. (2001). *The Timbre Model*. Retrieved from: http://www.researchgate.net/publication/2918235_The_Timbre_Model/file/9fcfd50eb03fcd92c8.pdf.

Jesteadt, W., Wier, C. C., Green, D. M. (1977). Intensity discrimination as a function of frequency and sensation level. *J Acoust Soc Am*, *61*, 169–177.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Percept Psychophys*, *49*, 227–229.

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

Kang, R., Nimmons, G. L., Drennan, W., et al. (2009). Development and validation of the University of Washington clinical assessment of music perception test. *Ear Hear*, *30*, 411–418.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., et al. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, *116*(4 Pt 1), 2395–2405.

Krimphoff, J., McAdams S., Winsberg S. (1994). Caractérisation du timbre des sons complexes. ii. analyses acoustiques et quantification psychophysique. *J Phys IV France*, *4*, 625–628.

Krumhansl, C. (1989). Why is musical timbre so hard to understand? In S. Nielzen & O. Olsson (Eds). *Structure and Perception of Electroacoustic Sound and Music* (pp.43–54). New York, NY: Excerpta Medica.

Krumhansl, C. L. (1990). Cognitive Foundations of Musical Pitch. New York, NY: Oxford University Press.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychol Bull*, *126*, 159–179.

Kutner, M. H., Nachtsheim, C. J., Neter J., (2004). Allied Linear Regression Models (4th ed). New York: McGraw-Hill Irwin.

Lee, C. (1991). The perception of metrical structure: Experimental evidence and a model. In P. Howell, R. West, I. Cross (Eds). *Representing Musical Structure*. London, UK: Academic Publisher.

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

Leek, M. R., Molis, M. R., Kubli, L. R., et al. (2008). Enjoyment of music by elderly hearing-impaired listeners. *J Am Acad Audiol*, *19*, 519–526.

Leek, M. R., & Summers, V. (1993). Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise. *J Acoust Soc Am*, *94*, 3127–3137.

Lerdahl, F. (2001). *Tonal Pitch Space*. New York, NY: Oxford University Press.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, *49*(Suppl 2), 467+.

Levitt, H. (2001). Noise reduction in hearing aids: a review. *J Rehabil Res Dev*, *38*, 111–121.

Limb, C. J. (2006). Structural and functional neural correlates of music perception. *Anat Rec A Discov Mol Cell Evol Biol*, *288*, 435–446.

Looi, V., McDermott, H., McKay, C., et al. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear Hear*, *29*, 421–434.

Luscher, E., & Zwislocki, J. (1949). A simple method for indirect monaural determination of the recruitment phenomenon (difference limen in intensity in different types of deafness). *Acta Oto-Laryngol*, (Suppl 78), 156–168.

McAdams, S., & Cunible, J. C. (1992). Perception of timbral analogies. *Philos Trans R Soc Lond B Biol Sci*, *336*, 383–389.

McAdams, S., Winsberg, S., Donnadieu, S., et al. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychol Res*, *58*, 177–192.

Medel Medical Electronics. (2006). *Mu.S.I.C Perception Test User Guide*. St Asaph, UK: H & D Fitzgerald.

Minghella, D., Russo, F., Pichora-Fuller, M. K. (2007). Effect of age on sensitivity to tonality. *Canadian Acoust*, *35*, 58–59.

Moore, B. C. J., & Glasberg, B. R. (1988). Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral cochlear impairments. In H. Duifhuis, J. W. Horst, H. P. Wit (Eds). Basic Issues in Hearing. London, UK: Academic Press.

Moore, B. C. J., & Glasberg, B. R. (1990). Frequency selectivity in subjects with cochlear hearing loss and its effects on pitch discrimination and phase sensitivity. In F. Grandor, G. Cianfrone, D. T. Kemp (Eds). *Cochlear Mechanisms and Otoacoustic Emissions*. Basel, Switzerland: Karger.

Moore, B. C., Glasberg, B. R., Shailer, M. J. (1984). Frequency and intensity difference limens for harmonics within complex tones. *J Acoust Soc Am*, *75*, 550–561.

Moore, B. C., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *J Acoust Soc Am*, *91*, 2881–2893.

Moore, B. C. J., (2003). An Introduction to the Psychology of Hearing (5th ed). New York, NY: Academic Press.

Mener, D. J., Betz, J., Genther, D. J., et al. (2013). Hearing loss and depression in older adults. *J Am Geriatr Soc*, *61*, 1627–1629.

Neukomm, E. (2006). *Musikhören bei Schwerhörigkeit mit und ohne Hörgerät*. Retrieved from: http://www.pro-audito.ch/fileadmin/daten/Neukommpas_musikstudie.pdf.

Oh, J. (2008). *Resolving Conflicting Linguistic and Musical Cues in Metric & Beat-Strength Perception of Songs*. Ph.D. thesis, Stanford, CA: Stanford University.

Oxenham, A. J., & Bacon, S. P. (2003). Cochlear compression: Perceptual measures and implications for normal and impaired hearing. *Ear Hear*, *24*, 352–366.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Percept*, *11*, 409–436.

Penner, M. J., Leshowitz, B., Cudahy, E. et al (1974). Intensity discrimination for pulsed sinusoids of various frequencies. *Percept Psychophys*, *15*, 568–570.

Piston, W. (1987). *Harmony* (5th ed). New York, NY: W. W. Norton & Company.

Rabinowitz, W. M., Lim, J. S., Braida, L. D., et al. (1976). Intensity perception. VI. Summary of recent data on deviations from Weber's law for 1000-Hz tone pulses. *J Acoust Soc Am*, *59*, 1506–1509.

Ruhm, H. B., Mencke, E. O., Milburn, B., et al. (1966). Differential sensitivity to duration of acoustic signals. *J Speech Hear Res*, *9*, 371–384.

Schubert, E. (2006). The effect of implied harmony, contour and musical expertise on judgments of similarity of familiar melodies. *J New Music Res*, *35*, 161–174.

Sarchett, W. (2006). *Hearing aid issues for the hearing impaired audiophile*. Retrieved from: http://www.hometheaterhifi.com/volume_13_2/hearing-aids-6-2006-part-1.html.

Schacknow, P. N., & Raab, D. H. (1973). Intensity discrimination of tone bursts and the form of the Weber function. *Percept Psychophys*, *14*, 449–450.

Sethares, W. (2005). *Tuning, Timbre, Spectrum* (2nd ed). New York, NY: Springer.

Spencer, L. J., & Oleson, J. J. (2008). Early listening and speaking skills predict later reading proficiency in pediatric cochlear implant users. *Ear Hear*, *29*, 270–280.

Spitzer, J. B., Mancuso, D., Cheng, M. Y. (2008). Development of a clinical test of musical perception: appreciation of music in cochlear implantees (AMICI). *J Am Acad Audiol*, *19*, 56–81.

Owen, S. V., & Froman, R. D. (1998). Uses and abuses of the analysis of covariance. *Res Nurs Health*, *21*, 557–562.

Tawfik, S., El Danasoury I., Moussa, A., et al. (2009). Enhancement of speech intelligibility in digital hearing aids using directional microphone/noise reduction algorithm. *Int Adv Otol*, *6*, 74–82.

Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music Percept*, *1*, 276–295.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Res*, *35*, 2503–2522.

Tufts, J. B., Molis, M. R., Leek, M. R. (2005). Perception of dissonance by people with normal hearing and sensorineural hearing loss. *J Acoust Soc Am*, *118*, 955–967.

Tyler, R. S., Summerfield, Q., Wood, E. J., et al. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. *J Acoust Soc Am*, *72*, 740–752.

Uys, M., & van Dijk, C. (2011). Development of a music perception test for adult hearing-aid users. *S Afr J Commun Disord*, *58*, 19–47.

von Békésy, G. (1947). A new audiometer. *Acta Oto-Laryngol*, *35*, 411–422.

von Helmholtz, H. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. New York, NY: General Books.

Wessel, D. (1979). Timbre space as a musical control structure. *Comput Music J*, *3*, 42–52.

Wier, C. C., Jesteadt, W., Green, D. M. (1977). Frequency discrimination as a function of frequency and sensation level. *J Acoust Soc Am*, *61*, 178–184.

Woodruff, L. C. (1983). *A Predictive Validity Study of the Primary Measures of Music Audiation*. Ph.D thesis, Philadelphia, PA: Temple University.