Relationship between Consonant Recognition in Noise and Hearing Threshold

Journal:	*Journal of Speech, Language, and Hearing Research*
Manuscript ID:	JSLHR-H-10-0239.R2
Manuscript Type:	Research Article
Date Submitted by the Author:	26-Jun-2011
Complete List of Authors:	Yoon, Yang-soo; House Ear Institute, Human communication sciences and devices Allen, Jont; University of Illinois at Urbana−Champaign, Electrical and Computer Engineering Gooler, David; University of Illinois at Urbana−Champaign, Speech and Hearing Science
Keywords:	consonant confusions, audiometric hearing threshold, signal-to-noise ratio

Relationship between Consonant Recognition in Noise and Hearing Threshold

Yang-soo Yoon[1], **Jont B. Allen**[2], and **David M. Gooler**[3]

yyoon@hei.org, jontalle@illinois.edu, dgooler@illinois.edu

[1]Communication and Neuroscience Division,

House Ear Institute, 2100 W. 3rd St., Los Angeles, CA 90057

[2]Department of Electrical and Computer Engineering,

University of Illinois at Urbana–Champaign

[3]Department of Speech and Hearing Science,

University of Illinois at Urbana–Champaign

Send correspondence to:

Yang-soo Yoon
House Ear Institute
2100 West Third Street
Los Angeles, CA 90057
Email: yyoon@hei.org

Abstract

Purpose: Although poorer understanding of speech in noise by hearing-impaired (HI) listeners is known not to be directly related to audiometric threshold [*HT* (f)], grouping HI listeners with *HT* (f) is widely practiced. In this study, the relationship between consonant recognition and *HT* (f) was considered over a range of signal-to-noise ratios (SNRs).

Method: Confusion matrices (CMs) from 25 HI ears were generated in response to 16 consonant-vowel syllables presented at 6 different SNRs. Individual Differences SCALing (INDSCAL) was applied to both feature-based matrices and CMs in order to evaluate the relationship between *HT* (f) and consonant recognition among HI listeners.

Results: The results showed no predictive relationship between the percent error scores [*Pe*] and *HT* (f) across SNRs. The multiple regression models showed that the *HT* (f) accounted for 39% of the total variance of the slopes of the *Pe*. Feature-based INDSCAL analysis showed consistent grouping of listeners across SNRs, but not in terms of *HT* (f). Systematic relationship between measures was also not defined by CM-based INDSCAL analysis across SNRs.

Conclusions: *HT* (f) did not account for the majority of the variance (39%) in consonant recognition in noise when the complete body of the CM was considered.

KEY WORDS: consonant confusions, audiometric hearing threshold, signal-to-noise ratio

Introduction

Pure-tone audiometry is a well established component of the audiometric test battery that measures behavioral hearing threshold to tones of different frequencies. For clinical and research purposes, many attempts have been made to test the correlation between speech recognition performance for hearing-impaired (HI) listeners and hearing thresholds. The results of such comparisons have generally shown little predictive value, particularly when speech recognition is measured in background noise (Festen and Plomp, 1983; Plomp, 1978; Smoorenburg, Latt, & Plomp, 1982). Evidence of a poor predictive relationship between hearing threshold and sentence recognition performance in noise is well documented (Bentler and Duve, 2000; Killion, 2004a, b; Lyregaard, 1982; Smoorenburg, Latt, & Plomp, 1982; Smoorenburg, 1992; Tschopp and Zust, 1994). The lack of correlation between the measures may be related to differences in the simple acoustic signals used for pure-tone audiometry and the complex nature of speech recognition even though frequency-specific audibility deficits are known to affect speech perception (Bamford et al., 1981; Carhart and Porter, 1971). Perception of running discourse may take advantage of increased information from complex signals and contextual and linguistic properties of speech as well as the linguistic experience of the listener.

In contrast to using meaningful sentences, some studies have investigated the relationship between speech recognition and hearing threshold using nonsense syllables (Bilger and Wang, 1976; Danhauer and Lawarre, 1979; Dubno, Dirks, & Langhofer, 1982; Gordon-Salant, 1987; Reed, 1975; Walden and Montgomery, 1975; Walden, Montgomery, Prosek, & Schwartz, 1980;

Wang, Reed, & Bilger, 1987). Using nonsense syllables is essential if investigators are interested in reducing the influence of contextual and linguistic factors so that recognition relies more on the use of acoustic features (Allen, 2005; Boothroyd and Nittrouer, 1988).

Previous studies have reported an inconsistent association between audiometric pure-tone thresholds and nonsense syllable recognition under different experimental methodologies. Four studies (Bilger and Wang, 1976; Dubno, Dirks, & Langhofer, 1982; Reed, 1975; Wang, Reed, & Bilger, 1978) showed a systematic relationship associating better performance with lower thresholds, but another four studies (Danhauer and Lawarre, 1979; Gordon-Salant, 1987; Walden and Montgomery, 1975; Walden, Montgomery, Prosek, & Schwartz, 1980) supported no such relationship. A number of different approaches to analysis were applied across the studies.

In three studies that showed a systematic relationship with pure-tone threshold (Bilger and Wang, 1976; Reed, 1975; Wang, Reed, & Bilger, 1978), the relationship was evaluated with the results of a Sequential INFormation Analysis (SINFA). SINFA provides the information for perceptual features embedded in confusion matrices (CMs) and determines the proportion of the information transmitted that is attributed to a given set of phonological features (Wang and Bilger, 1973). The procedure for constructing a (dis)similarity matrix for each subject can be summarized as follows. The results of a single SINFA were coded as a weighted vector for each of the stimulus features. The feature identified in the first iteration received the highest weight; the feature identified in the last iteration received the lowest weight; and the features not identified in the analysis received zero weight. Whenever the number of features identified exceeded the maximum weight, the lowest ranking features were all assigned weights of one. The similarity between any two subjects was defined as the sum of the products of corresponding feature weights. Finally the similarity matrices were submitted to Johnson's (1973) pair-wise multidimensional scaling procedure to represent the similarities among subjects spatially. Using this SINFA-based approach, the three studies showed a systematic relationship between phoneme recognition and configuration of the pure-tone threshold, distinguishing listeners with normal thresholds, those with a flat hearing loss, and hearing loss with sloping audiometric configurations (Bilger and Wang, 1976; Reed, 1975; Wang, Reed, & Bilger, 1978).

Unlike the SINFA-based approach, a similarity judgment task was applied in another three studies in which no systematic relationship between performance and audiometric thresholds was reported (Danhauer and Lawarre, 1979; Walden and Montgomery, 1975; Walden, Montgomery, Prosek, & Schwartz, 1980). In the similarity judgment task the subject was asked to rate the similarity between a pair of syllables using equal interval scaling (i.e., one being very similar; seven being very dissimilar). Similarity judgment allows the listener to consider perceptual qualities of the phonemes being compared in addition to recognition. For example, a HI listener can judge different speech sounds to be perceptually similar because they were correctly recognized as different phonemes but judged to be perceptually similar, or because they were incorrectly recognized and judged to be the same speech sound. The results of the similarity judgment were used as input for the INDSCAL (INdividual Difference SCALing) model, a multidimensional scaling technique (Carrol and Chang, 1970). Using the similarity judgment the three studies showed no unique association between measures (Danhauer and Lawarre, 1979; Walden and Montgomery, 1975; Walden, Montgomery, Prosek, & Schwartz, 1980). It should be noted that even though Walden and Montgomery (1975) reported a systematic relationship between measures, the INDSCAL analysis with three-dimensional solutions revealed ambiguous subject space, particularly between sibilant and sonorant dimensions (See Fig. 2, page 451, Walden and Montgomery, 1975).

Two studies analyzed phoneme recognition performance using raw CMs and compared the results with audiometric thresholds (Dubno, Dirks, & Langhofer, 1982; Gordon-Salant, 1987). Dubno, Dirks, & Langhofer (1982) assessed consonant confusions at a fixed +20 dB SNR (in cafeteria noise) in 38 HI listeners. A systematic relationship between consonant confusions and hearing threshold existed when the same consonant was given in error most commonly for a given target across all three HI listener groups but with differences in error probability. That is, given a target /sa/, /θa/ was confused with the target at an error rate of 28.6% by the steeply sloping group, 10.4% by the gradually sloping group, and 4.2% by the flat group. However, the greatest percentage of errors was not consistently associated with a particular group. Moreover, the three HI groups were not completely separable when the complete CM was taken into account for acoustic feature (manner and place) analyses. Gordon-Salant (1987) measured CMs for consonant identification at +6 dB SNR (12 talkers babble) for three groups of elderly listeners (10 NH, 10 gradual sloping, and 10 steep sloping listeners). The INDSCAL analysis of these raw CMs revealed no unique relationship between consonant confusions and the audiometric characteristics.

In summary, the results of four studies (Bilger and Wang, 1976; Dubno, Dirks, & Langhofer, 1982; Reed, 1975; Wang, Reed, & Bilger, 1978) lead to the conclusion that consonant confusions are systematically related to audiometric hearing threshold. Another four studies (Danhauer and Lawarre, 1979; Gordon-Salant, 1987; Walden and Montgomery, 1975; Walden, Montgomery, Prosek, & Schwartz, 1980) support the opposite conclusion.

An important distinction between the studies discussed above is the use of different input structures to the INDSCAL model. If SINFA-based (dis)similarity matrices (Bilger and Wang, 1976; Reed, 1975; Wang, Reed, & Bilger, 1978) or partial raw CMs (Dubno, Dirks, & Langhofer, 1982) were used for the INDSCAL, a systematic relationship between syllable perception and pure-tone audiometric threshold was obtained. In contrast, when similarity judgment measures (Danhauer and Lawarre, 1979; Walden and Montgomery, 1975; Walden, Montgomery, Prosek, & Schwartz, 1980) or complete raw CMs (Gordon-Salant, 1987) were used as input to the INDSCAL, no systematic relationship was observed. Similarity judgment measures are directly used as input for the INDSCAL. In contrast, SINFA-based measures should be carefully derived from raw CMs, and phonological features should be pre-selected by experimenters as input to the model. Consequently it is unclear how perceptual confusions embedded in CMs are reflected in SINFA-based (dis)similarity matrices. It is also unclear how the relationship between phoneme recognition and hearing threshold is impacted by these different input structures for the INDSCAL model.

Another issue in the previous studies is that CMs were measured in quiet (Bilger and Wang, 1976; Danhauer and Lawarre, 1979; Reed, 1975, Walden and Montgomery, 1975; Walden, Montgomery, Prosek, & Schwartz, 1980; Wang, Reed, & Bilger, 1978) or at +20 dB SNR (Dubno, Dirks, & Langhofer, 1982) and at +6 dB SNR (Gordon-Salant, 1987), which provided only partial information regarding the relationship between audiometric threshold and nonsense syllable recognition in noise. Finally, using nonsense syllables in noise provides the opportunity to evaluate performance with less use of contextual cues (e.g., meaning, grammar, prosody, etc.). These cues can increase speech understanding, particularly in noisy conditions, while not necessarily improving speech perception (Boothroyd & Nittrouer, 1988).

In the present study, the relationship between audiometric threshold and nonsense syllable recognition was evaluated with both SINFA- and CM-based INDSCAL analyses over a range of SNRs. The data evaluated here were previously studied for a separate analysis (Phatak, Yoon, Gooler, & Allen, 2009) that provided a new method to quantify the degree of consonant perception loss relative to normal hearing listeners over a range of SNRs. During the analyses, it was found that consonant confusions were not hearing-threshold specific, which led to motivation for this study. In the present study, the relationship between audiometric thresholds and syllable recognition in noise was evaluated in (1) mean performance-intensity functions, (2) correlation and multiple regression models having hearing threshold as predictors, and (3) SINFA-based and CM-based similarity matrices applied as inputs to the INDSCAL model.

Methods

Participants

The 22 paid participants had sensorineural hearing loss, were native speakers of American English, and were between the ages of 18 to 64 years old. Three listeners had bilateral hearing loss; hence each ear was tested separately (left and right ear identified as L and R) resulting in a total of 25 ears tested. Descriptive information for listeners is given in Table 1.

Participants were recruited on the basis of screening preexisting audiograms (Department of Otolaryngology, Carle Clinic Association, Urbana, IL) and only those showing a 3 frequency pure-tone average (PTA; 0.5 kHz, 1 kHz, and 2 kHz) between 30 dB hearing level (HL) to 70 dB HL were recruited. Listeners whose hearing threshold was greater than 70 dB HL at $f \geq 2$ kHz were not enrolled in the study because of high mean error rates in preliminary testing (see Procedures). The pure-tone audiograms of all participants were also measured for this study and are shown in the upper panel of Figure 1.

All procedures were approved by both the University of Illinois Institutional Review Board and the Carle Medical Research Institutional Review Board.

Test Materials

Sixteen naturally-spoken nonsense CV syllables composed of 16 American English consonants with the common vowel /a/ as in “father” were used as stimuli (Fousek, Svojanovsky, Grezl, & Hermansky, 2004). The 16 consonants presented were [/b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/, /s/, /t/, /v/, /ð/, /ʃ/, /θ/, /ʒ/, /z/]. One half of these syllables were spoken by five talkers and the remaining syllables spoken by another 5 talkers, resulting in 80 tokens [(5 talkers x 8 CVs) + (5 talkers x 8 CVs)] in total. The purpose of dividing syllables among talkers was to create a diversity of talkers and simultaneously shorten experiment time. The use of multiple utterances from several talkers also offers some assurance about the generality of the analyses beyond the experimental stimuli.

The CVs were presented in speech-weighted noise with no spectral correction (gain) as a function of SNR [−12 dB, −6 dB, 0 dB, 6 dB, 12 dB, and in quiet (Q)]. Each token was level-normalized before presentation using VU-meter software (Lobdell and Allen, 2007). No filtering was applied to the stimuli. The masker was a steady-state noise with an average speech-like power spectrum, identical to that used by Phatak and Allen (2007). For each CV, the RMS level of this noise was adjusted according to the level of the CVs to achieve the desired SNRs.

Stimuli were computer-controlled and delivered via an external USB audio card (Mobile-Pre, M-Audio), and presented monaurally via an Etymotic™ ER-2 insert earphone. Sound levels were controlled by an attenuator and headphone buffer (TDT™ system III) so that stimuli were presented at the most-comfortable-listening level (MCL) for each listener. The MCL was determined by each listener’s self rating with the Cox loudness rating scale (Cox, 1995) in response to 30 CVs with no error in quiet. System calibration estimates that CV presentation levels in the ear canal were between 75 and 85 dB SPL.

Procedures

The ear canal was inspected otoscopically and pure-tone audiometry was performed to measure hearing thresholds and to confirm type of hearing loss for each listener. Each participant was seated in a sound-treated room (Industrial Acoustics Company) for audiometry, practice, and experimental sessions. Stimuli were presented to a test ear via an insert earphone. Environmental sound to the other ear was attenuated using a foam earplug.

CV syllables were presented while participants viewed the graphical user interface that listed the 16 CVs with example words alphabetically. Participants were asked to select the button on the interface to identify the perceived CV.

A calibrate button was included so that the presentation level (MCL) could be determined by a subject's response to playing 30 CV syllables in quiet. In addition, pause and repeat buttons were available so that listeners could control the rate of stimulus presentation and could repeat the same stimulus without limit prior to responding. Our preliminary results with a few HI listeners showed no distinct influence of target repetition on performance.

Participants first performed a 30-minute, two practice-block (120 trials/block) session on CV identification in quiet with feedback. The eligibility to participate was determined by requiring the average percent error to be less than 50% across two practice blocks in quiet. If percent error score [Pe (SNR)] was ≥ 50% on the two practice blocks, two additional blocks were given to further consider eligibility for participation. Listeners became eligible to participate if *Pe* (SNR) was ≤ 50% on the second pair of practice blocks, but they remained ineligible if *Pe* (SNR) continued to be ≥ 50%.

The consonant identification test was administered to measure confusion matrices for CVs in speech-weighted noise as a function of SNR. For each presentation a CV and SNR were selected and presentation randomized from the array of 16 CVs and 6 SNR indices (including Q). The set of individual stimuli [(8 CVs x 5 talkers) + (8 CVs x 5 talkers) x 6 SNRs = 480, named a set] was repeated 6 times (480 x 6 = 2880 trials in total), yielding 30 [2880 / (16 CV x 6 SNR)] repetitions of each CV at each SNR.

Each set (480 trials) was evenly distributed into four blocks, (120 trials each) allowing participants to rest between blocks. No direct feedback about performance was provided for each CV presented. Percent correct feedback for each block was provided on the screen at the end of each block. The total number of trials and CVs already played were also provided on the screen.

Confusion matrices for each participant were plotted as a function of SNR. Any CV utterance, produced by a particular talker that showed > 20% error in quiet for NH listeners was considered mispronounced and was removed from data analysis (Phatak and Allen, 2007). Total participation time to complete all protocols (pure-tone audiometry, CV practice, CV test, and break time) was about 6 hours and was performed in two visits.

Results

Audiometric Analysis and Pe (SNR)

The pure-tone audiograms were separated into one of two overall audiometric configurations, to form a sloping group (n=18, Fig. 1, top left panel) and flat group (n=7, Fig. 1, top right panel). This classification was based on an historical scheme for describing the configuration of hearing threshold, *HT* (f), from the pure-tone audiogram (Bamford, Wilson, Atkinson, & Bench, 1981; Clark, 1981; Goodman, 1965; Margolis and Saly, 2007; Yoshioka and Thornton, 1980). This classification scheme suggests that audiogram profiles can be classified by threshold configuration such as normal, flat, and sloping curves. In some studies the sloping curve is further divided into two subgroups, for example, sloping curves with a slope ≤ 20 dB/oct or ≥ 30 dB/oct for 1 kHz $\leq f \leq$ 4 kHz (Clark, 1981; Goodman, 1965). Similarly, the flat curve could further be divided into two subgroups, flat curves with a slope ≤ 15 dB/oct or ≥ 25 dB/oct for 1 kHz $\leq f \leq$ 4 kHz (Margolis and Saly, 2007; Yoshioka and Thornton, 1980). In our sloping group, only 2 out of 18 ears (denoted by the dotted line in Fig. 1, top-left panel) showed audiogram configurations with a slope ≥ 30 dB/oct for 1 kHz $\leq f \leq$ 4 kHz, therefore subgroups were not defined. However, any trends indicated by the data points for these two listeners will be noted. For the flat group, all 7 listeners fell into a group having slopes ≤ 15 dB for 1 kHz $\leq f \leq$ 4 kHz. The mean hearing thresholds (denoted as thick lines in Fig. 1 upper panels) differed significantly between the sloping and flat configuration groups [$F(1,23)=6.7$, $p<0.05$]. At frequencies < 2 kHz, HT (*f*) for listeners with sloping hearing loss was approximately 20 dB better than for listeners with flat configuration, while at frequencies > 3 kHz, HT (*f*) for the flat group is 15 dB better than for the sloping group.

A comparison of the percent error scores *Pe* (SNR) between the two audiometric groups demonstrates a strong overlap in CV recognition performance with the range of *Pe* (SNR) exceeding 35% at each SNR. The lower panel of Figure 1 shows the *Pe* (SNR) across 16 CVs for individual listeners, coded according to the corresponding the *HT* (f). The mean *Pe* (SNR) for each group is shown by a thick line. A two-way repeated-measure ANOVA showed no significant difference between the mean *Pe* (SNR) of the two groups [$F(1,23)=0.2$, $p>0.05$]. The main effect of SNR was significant [$F(5,115)=505$, $p<0.001$]. The error scores for the two listeners with slopes ≥ 30 dB/oct for 1 kHz $\leq f \leq$ 4 kHz generally showed poorest performance among the listeners with sloping hearing loss.

In summary, we conclude that the audiogram-based listener grouping is poorly associated with the mean *Pe* (SNR) for nonsense CV recognition in noise. The results shown in Figure 1 indicate that the likelihood of demonstrating representative and distinctive descriptions of speech recognition performance across a range of HI listeners would be low if built upon the audiogram-based listener grouping.

Regression Model

To better understand the contribution of frequency-dependent audibility to CV recognition, we investigated the extent to which thresholds of individual audiometric frequencies are associated with overall recognition of nonsense CVs in noise. Specifically, we determined the extent to which listener's hearing thresholds account for the variance in *Pe* (SNR). To study this question, a multiple regression model was tested with the slopes of *Pe* (SNR) forming the dependent variable and with the *HT* (f) at standard audiometric frequencies, as the independent variable. These slopes were computed for each listener, based on a sigmoid fit without transformation.

The *HT* (f) at standard audiometric octave frequencies, namely, *x1* through *x6* for 0.25 kHz to 8 kHz were used as predictors. The best model was determined by testing all combinations of the 6 predictors. The search for the best combination of the predictors was finalized by finding the smallest sum of least square errors and the highest adjusted R^2 values. As a result, *HT* (f) at .25 kHz, 2 kHz, and 4 kHz were included in the model as predictors. In the case of using multiple predictors, it is possible the predictors do not operate independently, but reveal multicollinearity, preventing an indication of the influence of individual predictors. Multicollinearity is > 0.1 for all six predictors, which indicates no violation of the assumption that predictors operated independently (Tabachnick and Fidell, 1989). Other important assumptions were addressed appropriately[1].

The final linear regression model showed an insignificant relationship between the three predictors (*HT* (f)) and the slopes of the *Pe* (SNR) [$F(6,18) = 1.93$, $p>0.05$]. This model explained 39% of the total variance, suggesting that the balance of the variance is associated with other unmeasured variables. Based on weights (β coefficients) for the model, the order of effects on the slope of *Pe* (SNR) from greatest to least is for thresholds at 2 kHz, 4 kHz, and .25 kHz.

INDSCAL Analysis

In this section, an attempt was made to utilize listeners' perceptual errors to identify the relationship between audiometric threshold and consonant confusions in noise. To display this relationship across subjects, the INDSCAL model was used (Carrol and Chang, 1970). The INDSCAL model takes each listener's (dis)similarity matrix (measured in a CM or similarity judgment) as its input, transforms each CM into Euclidean distances, and iterates a process of estimating individual subject differences by applying individual sets of weights to the dimensions of a common group space. In the subject space, each listener is represented as a point, and the location of a listener in the subject space is adjusted by that subject's weights, indicating the particular salience to each of the dimensions of the space. In the present study, two-dimensional solutions were retained for both SINFA-based similarity matrices and raw CMs for each subject and for each SNR. A scree plot, a graph presenting a lack-of-fit INDSCAL model relative to dimensions, supports 3-dimensional solutions as the optimal number of dimensions, but a squared correlation index, the proportion of variance of the optimally-scaled data, with 2-dimensional solutions is also acceptable (Takane, Young, & de Leeuw, 1977). Another reason for choosing 2–dimensional solutions is to avoid the complexity of interpreting stimulus features across additional dimensions. The squared correlation index for the 2-dimensional solutions revealed that the model accounts for a variance of 72% to 97% over the

SNRs tested.

1. SINFA-based INDSCAL Model

A. Subject space

Evaluation of subject weight in 2-dimensional space, derived from the SINFA-based INDSCAL analysis, demonstrated no systematic relationship between stimulus features across SNRs and pure-tone threshold groups (Fig. 2). This result differs from that of Bilger and Wang (1976) despite obtaining (dis)similarity matrices using the same approach. However, two clear subject groups were identified, particularly by dimension (Dim.) 1. For example, 7 subjects with flat hearing loss (ID is given) are consistently separated into two groups across SNRs except SNR = –12 dB; 3R and 117R were separated from other five subjects (4L, 4R, 76L, 113R, and 216L) by Dim. 1. The two listeners whose audiograms showed a slope ≥ 30 dB/oct for 1 kHz ≤ *f* ≤ 4 kHz from the sloping group were also consistently classified in the same group across SNRs. This SINFA-based INDSCAL analysis revealed consistent groups of listeners across SNRs, but grouping was not consistent with the flat and sloping audiometric configurations.

To assess the consistency of these groups across SNRs, a retaining rate, the percentage of listeners remaining in the same group across SNRs, was computed, and the results are shown in top portion of Table 2. Overall retaining rate is constant across SNRs except –12 dB SNR. The retaining rates are given in each cell with the number of listeners in parentheses. Percentages oriented diagonally along the bottom of each column indicate the retaining rate between adjacent

SNRs. For example, 23 listeners (92%) out of 25 maintained their groups between Q and 12 dB SNR, and 16 (64%) out of 25 listeners were retained by their groups between –6 dB to –12 dB SNRs. Other cells indicate retaining rates for composite SNRs. For instance, 92% of listeners in the Q row remained in the same groups between Q and 12 dB SNR, but the retaining rate decreased to 68% when groups were considered over 12 dB, 6 dB, –6 dB, and –12 dB SNRs. The retaining rate decreases considerably at –12 dB SNR.

Two of the three subjects whose performance was measured separately for the right and left ears (4L/R in the flat group; 200L/R in the sloping group) were consistently categorized in the same group across SNRs (results were shown only for 4L/R in Figs. 2 and 4). The third subject who was tested bilaterally (2L/R in the sloping group) was categorized in the same group at –12 dB and 0 dB SNRs, but not at other SNRs.

2. CM-based INDSCAL Model

A. Stimulus space

The group stimulus space illustrated in Figure 3 provides a graphical representation of the stimulus coordinates derived by CM-based INDSCAL. This group stimulus space depicts the perceptual proximities of the stimuli presumed to underlie all listeners’ confusions. The dimensions are interpreted as the consonant features that can best account for the arrangement of the stimuli along each axis.

The stimuli appear to be arranged in two clusters along Dimension 1 (Dim. 1): the duration consonants (/s/, /ʃ/, /ʒ/, and /z/) are distinguished from the other 12 CVs at three lower SNRs (Fig. 3, top panels), whereas the fricative consonants (/f/, /s/, /v/, /ð/, /ʃ/, /ʒ/ and /z/) best define clusters at higher SNRs (Fig. 3, bottom panels). The feature labeled as duration is adopted from Miller and Nicely (1955) to distinguish four fricative consonants that are characterized by long duration, and intense, high-frequency noise. The presence of a long frication noise appears to be the important feature for defining Dim. 1 at lower SNRs.

Dimension 2 (Dim. 2) shows that the nasals (/m/ and /n/) are separated from the other 14 CVs at the four higher SNRs (Fig. 3). A misplacement of /ʒ/ is observed for Dim. 2 at +12 dB SNR. At −6 dB and −12 dB SNRs, the consonants on Dim. 2 are arranged in a single cluster, which precludes defining that dimension with an interpretable feature. For Dim. 2, the manner of articulation clearly serves as the common perceptual dimension at the four higher SNRs.

2. Subject space

The subject weights on 2−dimensional solutions of the CM-based INDSCAL process with 25 HI ears are shown in Figure 4. A dimension weight reflects the strength of the dimensional property in accounting for the confusions made by each subject at each SNR. That is, the weights reflect the types of confusions made by subjects at each SNR. For example, if the confusions are mainly between stimuli that share stimulus features specified by a dimension, then subject weights will be relatively high for that dimension. Where the confusions are mainly between stimuli that do not share stimulus features described by the dimension, the subject weights for that dimension are relatively low.

The result of the CM-based INDSCAL analysis shows no discernible categorization of listeners between the two audiometric groups at any SNR, including the quiet condition (Fig. 4). At each SNR, listeners were grouped (A, B, and C), based on differences in weighted Euclidean distances, although actual subjects within each cluster vary according to SNR. This CM-based group seems to be mainly dependent on the SNR, suggesting that confusions are a function of SNR, not of audiometric configurations. One noticeable pattern in the subject space is that listeners who had higher weights along Dim. 1 also had higher weights along Dim. 2. The variability of weights on both dimensions was noted for the two lowest SNRs. A distinct segregation of the two sloping group listeners with slopes ≥ 30 dB/oct for 1 kHz $\leq f \leq$ 4 kHz is demonstrated, particularly at 0 dB, 6 dB, and 12 dB SNRs, but any unique separation from other sloping group listeners across SNRs is not obvious. Because no consistency in HI grouping was found, plots of audiograms verse CM-based groups are not presented. In addition, the subjects with both ears tested (4L/R in the flat group; 2L/R and 200L/R in the sloping group) were consistently categorized in the same group at the four lower SNRs, but not for the two higher SNRs.

The retaining rate for CM-based listener grouping is shown in the bottom portion of Table 2. The retaining rate is proportional to SNR, that is, as a SNR decreases, the retaining rate also decreases, particularly at a SNR < 0 dB. The rate is also largely poorer than that for the SINFA-based grouping. 13 listeners (52%) out of 25 maintained their groups between Q and 12 dB SNR, and 11 (44%) out of 25 listeners were retained by their groups between −6 dB to −12 dB SNRs. Other cells indicate retaining rates for composite SNRs. For instance, 52% of listeners in Q row remained in the same groups between Q and 12 dB SNR, but the retaining rate decreased to 24% when groups were considered over to 12 dB, 6 dB, and −6 dB SNRs. Finally only 2 listeners (8%) remained in the same groups over four SNRs from 12 dB to −12 dB. This retaining rate would vary with the SNR step size being compared. If equal step sizes are compared on the diagonals formed along the bottom of columns in Table 2, then a U-shaped function is more apparent with the best retention rate for comparisons at 0 dB SNR. Indeed, 0 dB SNR has the highest retention rate.

Discussion

The goal of the present study was to determine the extent to which audiometric hearing threshold is associated with nonsense CV recognition in noise. The results revealed that the *Pe* (SNR) does not seem to be directly associated with the *HT* (f), as shown in Figure 1. However, in the quiet condition, the scores for the sloping hearing threshold group (20%) and for the flat hearing threshold group (25%) are similar to those reported by Dubno, Dirks, & Langhofer (1982). Dubno, Dirks, & Langhofer (1982) reported that errors among the listeners with sloping hearing loss were the lowest (22% error), whereas those with flat hearing loss were somewhat poorer (30% error), and those labeled steep hearing loss showed the highest error (50%) on CV or VC syllables in cafeteria noise at +20 dB SNR.

The multiple regression models revealed non-significant associations between the slopes of *Pe* (SNR) and *HT* (f). *HT* (f) contributed 39% of the total variances of the slope of *Pe* (SNR). The weights (β coefficients) for the model showed that effect on the slope of *Pe* (SNR) was the greatest for thresholds at 2 kHz. Carhart and Porter (1971) showed a similar finding for spondees: adding a threshold at 1 kHz (except in the group with marked high-frequency loss) for the regression model was highly correlated with speech reception threshold (SRT), but adding threshold at 2 kHz to the model improved the prediction slightly. However, adding thresholds at 4 kHz and 0.25 kHz did not produce practical improvement in predictability for spondee SRT. Bamford et al. (1981) correlated pure-tone audiograms with the slope of sentence perception performance in quiet for 150 HI children. Poor correlation ($r = 0.329$) was reported. It was also reported that the correlation between measures was highly affected by the degree of hearing loss, particularly from severe to profound hearing loss.

SINFA-based listener grouping (Fig. 2) showed no unique relationship of audiometric characteristics with consonant confusions even though two distinct groups were consistently defined across SNRs. This poor relationship is related to two technical issues in the SINFA analysis.

First, the SINFA requires prior knowledge about unknown perceptual features embedded in CMs. In the SINFA procedure, phonological features are selected by the experimenter with some unknown assumption about the perceptual features. The analysis of SINFA-based

INDSCAL provides only subject spaces without names of dimensions because experimenters select features for the model in advance. This is the reason that all studies that used SINFA never presented subject dimensions because the approach does not permit identification of that information. In addition, requirement of prior knowledge of perceptual features is a fundamental violation for INDSCAL model because the core concept of the INDSCAL model is to reveal unknown perceptual dimensions embedded in CMs or (dis)similarity matrices.

Another concern about using SINFA is related to the procedure for obtaining (dis)similarity matrices. As discussed in the Introduction, the feature identified in the first iteration received the highest weight; the feature identified in the last iteration received the lowest weight; and the features not identified in the analysis received zero weight. Whenever the number of features identified exceeded the maximum weight, the lowest ranking features were all assigned weights of one. The similarity between any two subjects was defined as the sum of the products of corresponding feature weights. This means that a similarity matrix for one subject might be very similar to that of another subject even though their features were identified in very different orders. For example subject A has ratings from 6 to 1 for the same set of features, but subject B has ratings from 1 to 6. The sum of the products between subjects A and B is 56. Another two subjects C and D have two top ratings (6 and 5) in common, but ratings for other features are not in common. The sum of the products between subjects C and D is 61. It is highly likely that the SINFA-based INDSCAL model would consider these two pairs of subjects similar even though their feature perception is completely different.

As shown in the 2-dimensional subject space derived from the CM-based INDSCAL (Fig. 4), no unique relationship between audiometric thresholds and perceptual confusions was evident, across SNRs including the quiet condition. This CM-based INDSCAL grouping seems to be a function of SNR, not of audiometric configuration.

Our CM-based INDSCAL solutions for perception in quiet are consistent with results of other studies (Danhauer and Singh, 1975; Danhauer and Lawarre, 1979; Walden, Montgomery, Prosek, & Schwartz, 1980) despite differences in some experimental conditions including (in their studies): a single talker, listener's demographics, stimulus context (CV-CV pairs), and response mode (similarity judgment using 7-point equal-appearing interval scaling). Danhauer and Singh (1975) found that subject weights in the 3-dimensional solutions generated by INDSCAL were neither obvious nor related to three different audiometric configurations. Danhauer and Lawarre (1979) also found that HI listeners represented in 3-dimensional solutions could not be clustered into distinct subgroups according to three different configurations of hearing loss. Walden, Montgomery, Prosek, & Schwartz (1980) also reported no consistent differences in feature weights between two HI listener groups represented by INDSCAL in 4-dimensional solutions. This result is somewhat in disagreement with those of Walden and Montgomery (1975) who reported distinct HI listener groupings in 3-dimensional subject space determined by the INDSCAL analysis. In contrast to a conclusion made by the authors, the INDSCAL analysis with three-dimensional solutions revealed ambiguous subject space, particularly between sibilant and sonorant dimensions (See Fig. 2, Walden and Montgomery, 1975). Compared with INDSCAL groupings from other studies, subject space in the study by Walden and Montgomery (1975) did not support distinct subject groups.

A study of CV perception in HI listeners by Bilger and Wang (1976) provides a particularly important comparison with the current study because the complete body of information from both diagonal and off-diagonal cells in CMs was fully taken into account for the analysis. Whereas 14 CVs used by Bilger and Wang (1976) were identical to those used in the current study, some details of the experimental conditions differed. For example, the number of talkers and vowels used differed (a single talker and three vowels [/i/, /a/, /u/] in Bilger and Wang (1976); 10 talkers and a single vowel /a/ in the current study). However, it has been demonstrated that differences in the vowel accompanying the consonant have little effect on the patterns of consonant confusions (Gordon-Salant, 1985; Phatak and Allen, 2007).

Grouping of pure-tone audiogram configurations as defined by SINFA-based INDSCAL of CV confusions in quiet revealed different patterns between the study by Bilger and Wang (1976, Fig. 5, bottom panels) and the current study (Fig. 5, top panels). Bilger and Wang (1976) found three distinct subgroups in 2-dimensional space. The data of Bilger and Wang (1976) revealed differences in the average configuration of hearing thresholds that appear clearly discernible (Fig. 5, bottom right panel). The NH/gradual group had a slope < 20 dB for 1 kHz ≤ f ≤ 4 kHz. For the same range of frequencies, the flat group had a slope < 5 dB, and the steep group had a slope > 30 dB. For the current study, the slopes of the average hearing thresholds showed great overlap across groups defined by SINFA-based INDSCAL (Fig. 5, top right panel).

For example, for 0.25 kHz≤ f ≤2 kHz, the slopes of group 1 and 2 are somewhat different (<15 dB), but for 2 kHz≤ f ≤4 kHz, the slopes of the two groups are similar (<10 dB). Listener grouping, defined by CM-based INDSCAL in the current study (Fig. 5, middle panel) was different from that defined by SINFA-based INDSCAL in the current study and in Bilger and Wang (1976). For the current study, the slopes of the average hearing thresholds showed great overlap across groups defined by CM-based INDSCAL (Fig. 5, middle right panel). Specifically for 1 kHz ≤ f ≤ 4 kHz, the slopes of groups A and B are similar (< 20 dB), and the slope of group C is < 30 dB.

The cause for the discrepancy in the results defined by SINFA-based INDSCAL between the current study and that of Bilger and Wang (1976) might be talker variation. For the present study, 16 CVs were produced by 10 different talkers, whereas for the Bilger and Wang (1976) study all CVs were produced by a single talker. It has been shown that perceptual confusions are clearly influenced by talker variation (Phatak, Lovitt, & Allen, 2008; Regnier and Allen, 2008). Phatak et al. (2008) showed that different utterances of the same consonant can produce a significant variability in performance scores and confusion patterns. The consonant most often confused with a given target consonant varied depending on the talker. The reason for using multiple talkers in the present study was to measure confusions under more realistic listening conditions. Such conditions may yield results that are more readily generalized, but more complex in that the confusions are more distributed even for the same utterance. Thus, it is likely that talker variation is one of the variables that can spread the effect of the audiometric difference across subject space, resulting in inconsistent groupings for performance in quiet as shown in the current study.

The cause for the discrepancy in the results between CM-based INDSCAL in the current study and SINFA-based INDSCAL in the study by Bilger and Wang (1976) appears to result from a difference in input structures for the INDSCAL model. In the study by Bilger and Wang (1976), (dis)similarity matrices for the INDSCAL model were constructed from the indices of feature perception, determined by the SINFA (Wang and Bilger, 1973), whereas for the current study (dis)similarity matrices were normalized, raw CMs. Details of how (dis)similarity matrices for the INDSCAL model were constructed, based on the results of SINFA were given in the Introduction. The differences in the structure of (dis)similarity matrices directly alter the iteration process from an arbitrary initial configuration of subject space in the INDSCAL model, resulting in a different estimated configuration of subject spaces (Jones and Young, 1972; MacCallum, 1977; Takane, Young, & de Leeuw, 1977). One of two conclusions made by Wang and Bilger (1973), about identifying distinct perceptual features for CVs from CMs measured in both in quiet and noise, is that similar information transmission for features does not guarantee similar consonant confusion patterns or vice versa. Thus, it is possible, based on the systematic differences between the present study and that of Bilger and Wang (1976) that the dissimilarity matrices, constructed from information transmission for features (SINFA), are more reflective of audiometric threshold differences than of confusion matrices.

For both Bilger and Wang (1976) and the current studies, audibility might be one of the factors, affecting internal structure of perceptual confusions. Bilger and Wang (1976) used a presentation level of 40 dB above the subject's SRT and a MCL (75 dB ~ 85 dB SPL) was used for the current study. Using data given in the study of Bilger and Wang (1976), the average presentation levels were computed for each of the HI groups categorized by SINFA-base analyses as follows: 54.6 dB HL for the NH/gradual group (Fig. 5, A panel), 67.6 dB HL for the flat group (Fig. 5, B panel), and 67.0 dB HL for the steep group (Fig. 5, C panel). For the current study, using the minimum audibility curve (ANSI-1969), the presentation levels of 75 and 85 dB SPL would be equivalent to 62.5 and 72.4 dB HL. The presentation levels in dB HL for both studies were comparable. However, by inspecting the audiograms given in Fig. 5 for both studies it is clear that sensation level is too low for some subjects. For example, in a study by Bilger and Wang (1976) three subjects in the NH/gradual group had sensation level less than 10 dB at frequencies >3 kHz; 2 and 4 subjects in the flat and steep groups showed the same results. In the current study, 6 subjects in the A group (Fig. 5, middle panel) had sensation level of less than 10 dB at 3 kHz; this result was similarly evident in 1 and 4 subjects in the B and C groups, respectively (Fig. 5, middle panel). Lower sensation level at high frequencies might affect perception of some consonants such as /sa/ and /ʃa/, but it is unclear how such a lack of audibility affects the confusion patterns and consequently it is difficult to predict how listener's groupings observed in both studies will be affected. It would be interesting to see how the relationship between consonant confusions and hearing threshold will be affected if a spectral compensation procedure such as NAL-R is applied to adjust frequency response based on the loudness equalization for each CV.

Another possible influence on the grouping observed in the current study is the characteristics of the noise masker (speech-shaped noise). That is, the presence of a noise stimulus might change the effective hearing loss configuration, making it more similar than different for persons with different losses. If this is the case, then the result of grouping in quiet for the current study should be different from that in noise. This was not the case for the results of the current study. For example, in Figure 2, seven subjects with flat hearing loss were consistently separated into two groups when syllables were presented in both noise and quiet. Irrespective of presence of noise, the two listeners with steeply sloped hearing loss (threshold ≥ 30 dB/oct for 1 kHz $\leq f \leq$ 4 kHz) were also consistently classified in the same group. In addition, 3R and 117R in the sloping group were separated from five other subjects in the same group (4L, 4R, 76L, 113R, and 216L) across SNR including the quiet condition. Based on this evidence, it is unlikely that the presence of a noise stimulus changes the effective hearing loss configuration and makes persons with different losses more similar than different.

The results of the present study might be useful for hearing aid fitting algorithm research. For most current hearing aid fitting algorithms, the pure-tone audiogram is the primary input even though the audiogram does not account for the majority of the variance in performance of speech perception in noise. Our results suggest that patients with similar audiometric configurations may require different hearing aid strategies.

Conclusions

A clear predictive relationship between the percent error scores, *Pe* (SNR), and audiometric hearing threshold, *HT* (f), was not found for syllable recognition both in noise and quiet. The result of a multiple regression model showed that 39% of total variance of the *Pe* (SNR) was contributed by the *HT* (f). The result of SINFA-based INDSCAL analysis revealed consistent grouping of listeners across SNRs, but groupings were not consistent with two configurations of pure-tone thresholds. The CM-based INDSCAL analysis showed no systematic relationship between the consonant confusions and the *HT* (f) at any SNRs, including the quiet condition. Thus, audiometric threshold does not account for the majority of the variance in performance of nonsense-syllable perception in noise when complete CMs were considered.

Acknowledgments

The data collection expenses were covered through research funding provided by Etymotic Research. We thank all members of the Human Speech Recognition group at the Beckman Institute, University of Illinois at Urbana-Champaign for their help. Our special thanks go to Feipeng Li and Sandeep Phatak for their careful reading and valuable suggestions. We thank Andrew Lovitt for writing the Matlab data collection code. We want to thank Justin Aronoff, associate researcher at House Ear Institute for his comments on MDS computation and interpretation of the results and thank Laurel Fisher for valuable comments on the regression analyses. We thank three anonymous reviewers for valuable comments on an earlier version of this manuscript.

Endnotes

1. An assumption of normality of residual errors was tested by checking histograms for the residuals as well as normal probability plots. The linearity assumption between variables was verified by plotting bivariate scatter plots of the variables. In practice these assumptions can never be fully confirmed; however in this case linearity was read from these scatter plots.

References

Allen, J. B. (2005). *Articulation and intelligibility, synthesis lecture in speech and audio processing*, series editor B. H. Juang (Morgan and Claypool).

ANSI (1968). Specifications for audiometers, S3.6-1969 (Am. Natl. Stand. Inst., New York).

Bamford, J. M., Wilson, I. M., Atkinson, D., & Bench, J. (1981). Pure tone audiograms from hearing-impaired children. II. predicting speech-hearing from the audiogram. *British Journal of Audiology*, 15(1), 3–10.

Bentler, R. A. & Duve, M. R. (2000). Comparison of hearing aids over the 20th century. *Ear and Hearing,* 21(6), 625–639.

Bilger, R. C. & Wang, M. D. (1976). Consonant confusions in patients with sensorineural hearing loss. *Journal of Speech Hearing Research*, 19(4), 718–748.

Boothroyd, A. & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America,* 84(1), 101–114.

Carrol, J. D. & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika*, 35, 283–319.

Carhart, R. & Porter. L. S. (1971). Audiometric configuration and prediction of threshold for spondees. *Journal of Speech and Hearing Research*, 14, 486-495.

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *American Speech-Language-Hearing Association*, 23(7), 493–500.

Cox, R. M. (1995). Using loudness data for hearing aid selection: the IHAFF approach. *Hearing Journal*, 48(2), 39–44.

Danhauer, J. L. & Lawarre, R. M. (1979). Dissimilarity ratings of English consonants by normally-hearing and hearing-impaired individuals. *Journal of Speech Hearing Research*, 22(2), 236–246.

Danhauer, J. L. & Singh, S. (1975). *Multidimensional speech perception by the hearing impaired: A treatise on distinctive features* (University Park Press: Baltimore).

Dubno, J. R., Dirks, D. D., & Langhofer, L. R. (1982). Evaluation of hearing-impaired listeners using a nonsense-syllable test. II. syllable recognition and consonant confusion patterns. *Journal of Speech Hearing Research*, 25(1), 141–148.

Festen, J. M. & Plomp, R. (1983). Relations between auditory functions in impaired hearing. *Journal of the Acoustical Society of America,* 73(2), 652–662.

Fousek, P., Svojanovsky, P., Grezl, F. & Hermansky, H. (2004). New nonsense syllables database-Analyses and preliminary ASR experiments, in *Proceedings of the International Conference on Spoken Language Processing*, 2749-2752.

Goodman, A. (1965). Reference zero levels for pure-tone audiometer. *American Speech-Language-Hearing Association,* 7, 262–263.

Gordon-Salant, S. (1985). Phoneme feature perception in noise by normal-hearing and hearing-impaired subjects. *Journal of Speech Hearing Research*, 28(1), 87–95.

Gordon-Salant, S. (1987). Consonant recognition and confusion patterns among elderly hearing-impaired subjects. *Ear Hear*, 8(5), 270-276.

Jones, L. E. & Young, F. W. (1972). Structure of a social environment: Longitudinal individual differences scaling of an intact group. *Journal of Personality and Social Psychology*, 24, 108–121.

Killion, M. (2004a). Myths about hearing in noise and directional microphones. *The Hearing Review*, 11(2), 14–19, 72–73.

Killion, M. (2004b). Myths that discourage improvements in hearing aid design. *The Hearing Review*, 11(1), 32–40.

Lobdell, B. & Allen, J. B. (2007). Modeling and using the VU-meter (volume unit meter) with comparisons to root-mean-square speech levels. *Journal of the Acoustical Society of America,* 121(1), 279–285.

Lyregaard, P. E. (1982). Frequency selectivity and speech intelligibility in noise. *Scandinavian Audiology Supplement*, 15, 113–122.

MacCallum, R. C. (1977). Effects of conditionality on INDSCAL and ALSCAL weights. *Psychometrika*, 42, 297–305.

Margolis, R. H. & Saly, G. L. (2007). Toward a standard description of hearing loss, *International Journal of Audiology*, 46(12), 746–758.

Marshall, L. & Bacon, S. P. (1981). Prediction of speech discrimination scores from audiometric data. *Ear and Hearing,* 2(4), 148–155.

Miller, G. A. & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America,* 27(2), 338–352.

Phatak, S. & Allen, B. J. (2007). Consonant and vowel confusions in speech-weighted noise. *Journal of the Acoustical Society of America,* 121(4), 2312–2326.

Phatak, S. A., Lovitt, A., & Allen, J. B. (2008). Consonant recognition in white noise. *Journal of the Acoustical Society of America, 1*24(2), 1220–1233.

Phatak, S. A., Yoon, Y. S., Gooler, M. D., & Allen, J. B. (2009). Consonant recognition loss in hearing impaired listeners. *Journal of the Acoustical Society of America, 1*26(5), 2683-2694.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *Journal of the Acoustical Society of America,* 63(2), 533–549.

Reed C. (1975). Identification and discrimination of vowel-consonant syllables in listeners with sensorineural hearing loss. *Journal of Speech and Hearing Research*. 18(4), 773-794.

Regnier, M. S. & Allen, J. B. (2008). A method to identify noise-robust perceptual feature: application for consonant /t/. *Journal of the Acoustical Society of America,* 123(5), 2801–2814.

Smoorenburg, G. F. (1992). Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *Journal of the Acoustical Society of America,* 91(1), 421–437.

Smoorenburg, G. F., de Latt, J. A., & Plomp, R. (1982). The effect of noise-induced hearing loss on the intelligibility of speech in noise. *Scandinavian Audiology Supplement*, 16, 123–133.

Tabachnick, B. & Fidell, L. (1989). *Using multivariate statistics* (Harper and Row Publishers).

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, 42, 7–67.

Tschopp, K. & Zust, H. (1994). Performance of normally hearing and hearing-impaired listeners using a German version of the SPIN test. *Scandinavian Audiology*, 23(4), 241–247.

Walden, B. E. & Montgomery, A. A. (1975). Dimensions of consonant perception in normal and hearing-impaired listeners. *Journal of Speech Hearing Research*, 18(3), 444–455.

Walden, B. E., Montgomery, A. A., Prosek, R. A., & Schwartz, D. M. (1980). Consonant similarity judgments by normal and hearing-impaired listeners. *Journal of Speech Hearing Research*, 23(1), 162–184.

Wang, M. D. & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America*, 54(5), 1248–1266.

Wang, M. D., Reed C. M., & Bilger R. C. (1987). A comparison of the effects of filtering and sensorineural hearing loss on patterns of consonant confusions. *Journal of Speech and Hearing Research*, 21(1), 5-36.

Yoshioka, P. & Thornton, A. R. (1980). Predicting speech discrimination from the audiometric thresholds, *Journal of Speech Hearing Research*, 23(4), 814–827.

Table 1. Descriptive information for listeners. Each listener is identified by ID number + ear tested ([R: Right or L: Left]). Three listeners whose performance was tested monaurally in both ears are indicated by ID plus R/L. Differences in number of listeners, ears, and audiograms vary because of the 3 listeners who were tested bilaterally. Listeners were divided into two groups on the basis of audiometric configuration: sloping group (18 ears) and flat group (7 ears).

	Sloping group						Flat group			Total
	ID	Gender	Age	ID	Gender	Age	ID	Gender	Age	
	1L	F	21	148L	M	60	3R	M	21	
	2L/R	F	59	170R	M	53	4L/R	F	63	
	12L	F	39	188R	M	64	76L	F	62	
	39L	M	63	195L	F	60	113R	M	48	
	48R	M	62	200L/R	F	52	177R	F	39	
	71L	M	60	208L	F	54	216L	F	58	
	112R	F	54	300L	M	54				
	134L	F	52	301R	M	58				
# of listener	16						6			22
# of ear	18						7			25

Table 2. Retaining rate for SINFA-based (top) and CM-based (bottom) INDSCAL groups over SNRs. First row and second column indicate SNRs. The proportion of listeners that remained in the same group out of 25 listeners is given with the number of listeners in parenthesis. The set of diagonal cells formed along the bottom of each column specifies the retaining rate for adjacent SNRs and Q.

	SNR	12 dB	6 dB	0 dB	−6 dB	−12 dB
SINFA-based INDSCAL	Q	92% (23)	92% (23)	80% (20)	92% (23)	68% (17)
	12 dB		92% (23)	72% (18)	92% (23)	56% (14)
	6 dB			72% (18)	100% (25)	64% (16)
	0 dB				72% (18)	60% (15)
	−6 dB					64% (16)
CM-based INDSCAL	SNR	12 dB	6 dB	0 dB	−6 dB	−12 dB
	Q	52% (13)	40% (10)	40% (10)	24% (6)	8% (2)
	12 dB		68% (17)	60% (15)	36% (9)	12% (3)
	6 dB			76% (19)	44% (11)	12% (3)
	0 dB				60% (15)	20% (5)
	−6 dB					44% (11)

Figure Captions

Figure 1: Hearing thresholds [*HT* (f)] and the percent error scores [*Pe* (SNR)] for the two audiogram-based groups. The upper panels are audiograms, categorized by configuration of *HT* (f): sloping group (18 ears, left panel) and flat group (7 ears, right panel). The average *HT* (f) is indicated by a thick line. The lower panel shows the *Pe* (SNR) per listener. The mean *Pe* (SNR) for each group is shown by a thick line. For both top and bottom panels, data of two listeners having an audiogram with a slope of ≥ 30 dB/oct for 1 kHz ≤ *f* ≤ 4 kHz from the sloping group are indicated by thin-dotted lines. *Ce* is chance performance.

Figure 2: Subject cluster, defined by SINFA-based INDSCAL analysis for each SNR. Subjects in the sloping group (n=18) are represented by open circles, while subjects in the flat group (n=7) is are represented by their IDs. Two listeners with PTA > 30 dB/oct are represented by thicker circles. Groups 1 and 2 were assigned in the quiet condition (lower right panel) for comparison with a study of Bilger and Wang (1976).

Figure 3: Group stimulus, derived by the CM-based INDSCAL model at each SNR. Dimension 2 is not precisely determined at SNRs = −12 dB and −6 dB.

Figure 4: Listener distributions in subject distance space, assessed by the CM-based INDSCAL model at each SNR. Members of the sloping group are denoted with open circles, while flat

group members are denoted with their IDs. At each SNR, each data cluster is labeled as A, B, and C, although actual subjects within each cluster vary according to SNRs. Two sloping group listeners with slopes ≥ 30 dB/oct for 1 kHz ≤ f ≤ 4 kHz are denoted by thicker circles. For better visualization of grouping, the abscissa and the ordinate are scaled differently in each SNR panel, but ranges of both axis limits are constant across SNRs.

Figure 5: Audiograms, categorized by INDSCAL. The top panels are audiograms from the current study, defined by the SINFA-based INDSCAL model in quiet. There are 12 and 13 listeners in panels 1 and 2. Middle panels are audiograms from the current study that are grouped by the CM-based INDSCAL model in quiet. There are 15, 5, and 5 listeners in panels A, B, and C, respectively. The bottom panels are pairwise multidimensional scaling-based HI groups for CVs presented in quiet, reported by Bilger and Wang (1976). Eight ears were classified as belonging to the NH/gradual group, 6 ears as the flat group, and 9 ears as the steep group. Average thresholds for all three groups are shown in the panels to the right for purpose of comparison.

Figures

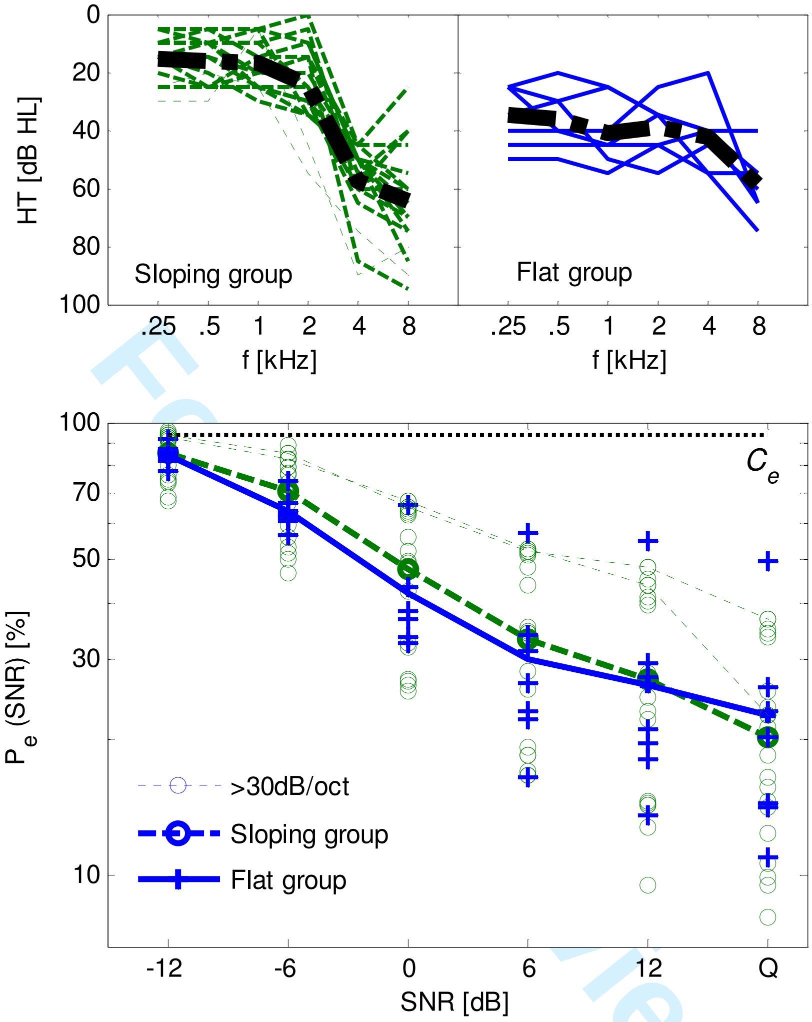


Figure 1

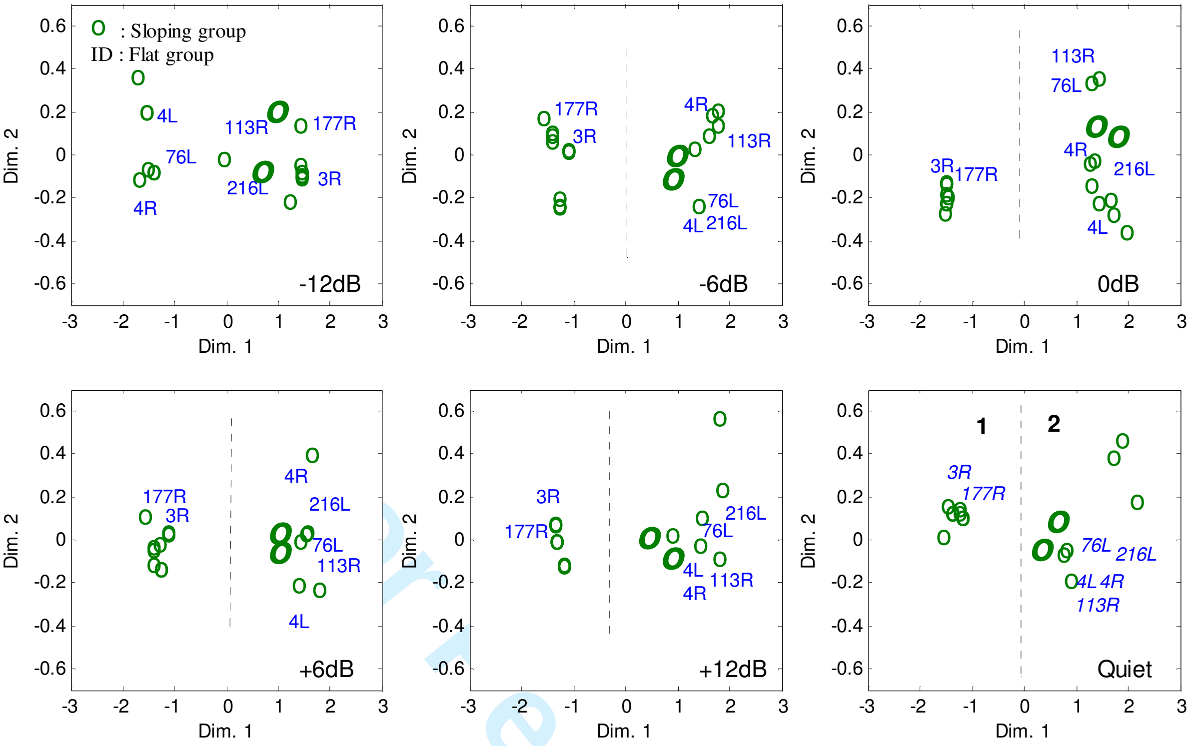

O : Sloping group
ID : Flat group
-12dB
-6dB
0dB
+6dB
+12dB
Quiet
1
2
Dim. 1
Dim. 2
4L
4R
76L
113R
177R
216L
3R

Figure 2

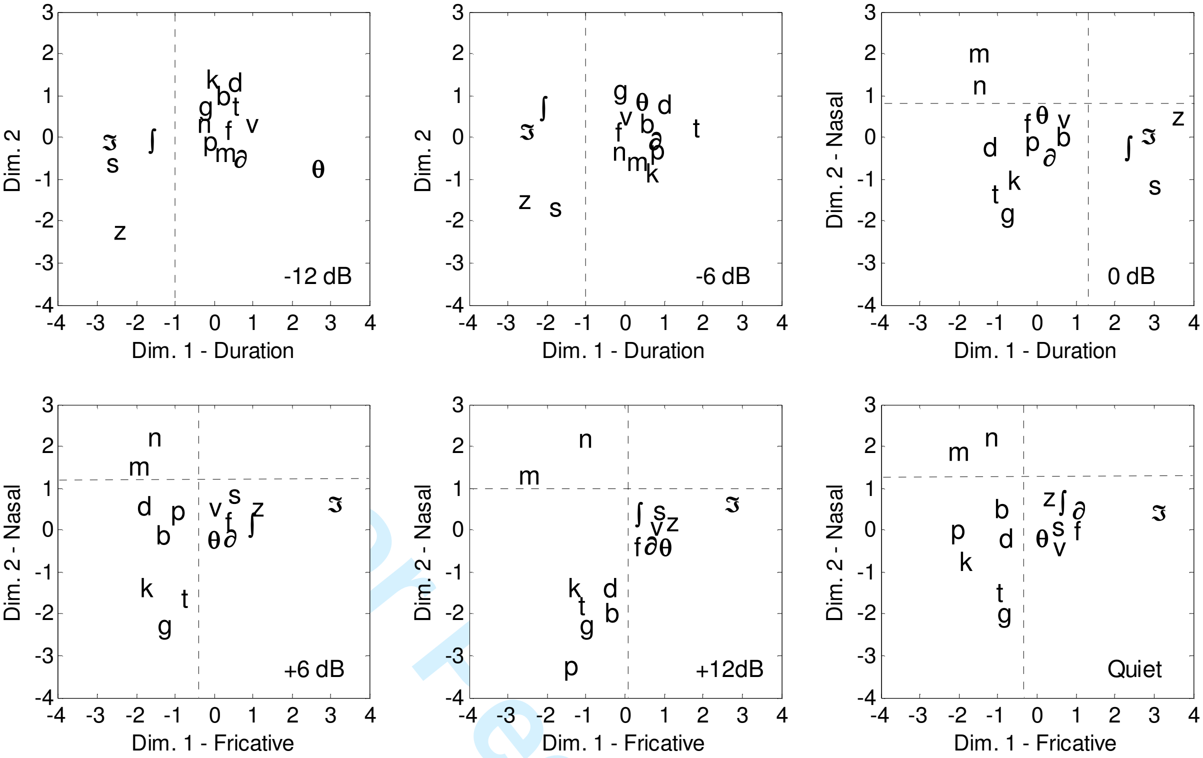

Dim. 2
Dim. 1 - Duration
-12 dB
k d b t g n f v p m ∂ θ ʒ ∫ s z
-6 dB
g θ d v b f ∂ t n m p k ∫ ʒ z s
Dim. 2 - Nasal
0 dB
m n f θ v p b d ∂ k t g ∫ ʒ z s
Dim. 1 - Fricative
+6 dB
n m d p b v s f z ∫ θ ∂ ʒ k t g
+12dB
n m ∫ s v z f ∂ θ ʒ k d t b g p
Quiet
m n b z ∫ ∂ ʒ p d θ s f v k t g
3
2
1
0
-1
-2
-3
-4
-4 -3 -2 -1 0 1 2 3 4

Figure 3

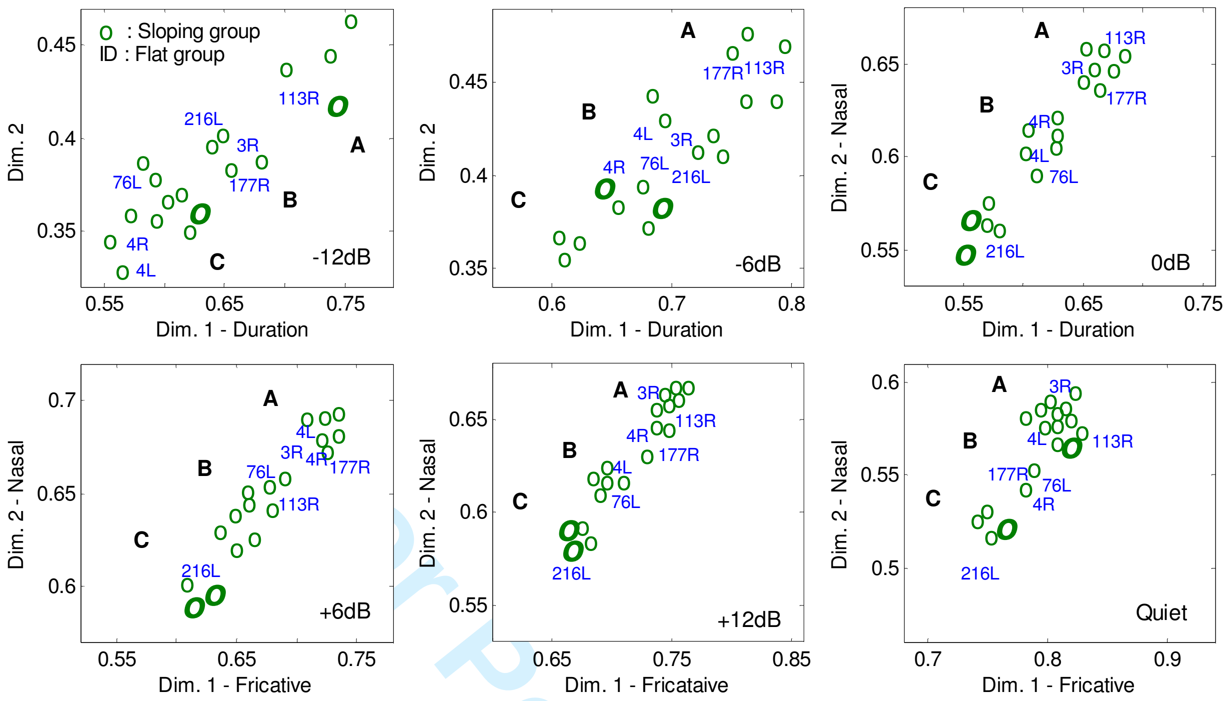
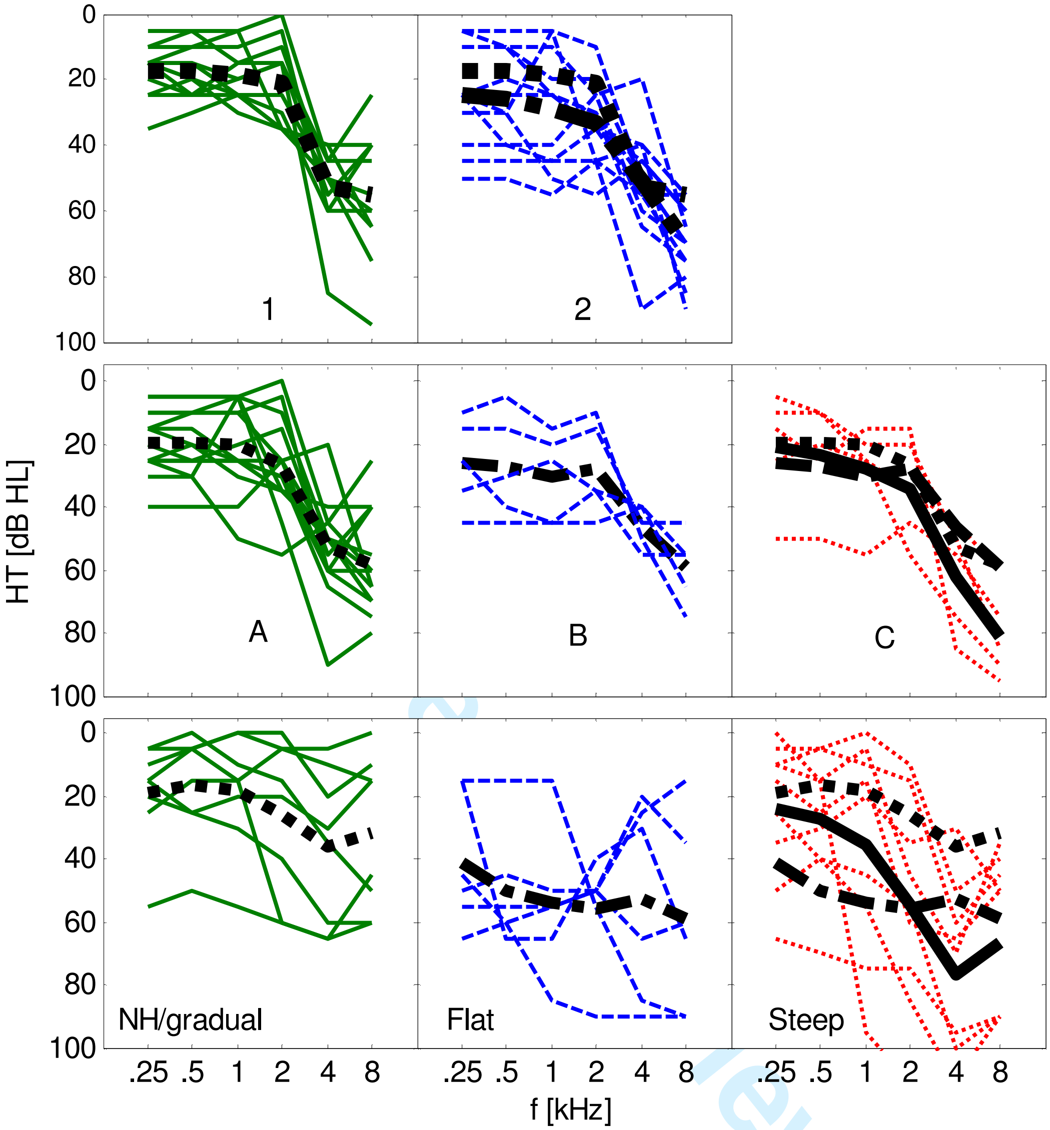

O : Sloping group
ID : Flat group
Dim. 2
Dim. 1 - Duration
-12dB
-6dB
0dB
Dim. 2 - Nasal
Dim. 1 - Fricative
Dim. 1 - Fricataive
+6dB
+12dB
Quiet
A
B
C
113R
216L
3R
177R
76L
4R
4L
0.35
0.4
0.45
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.5

Figure 4



764

765 Figure 5

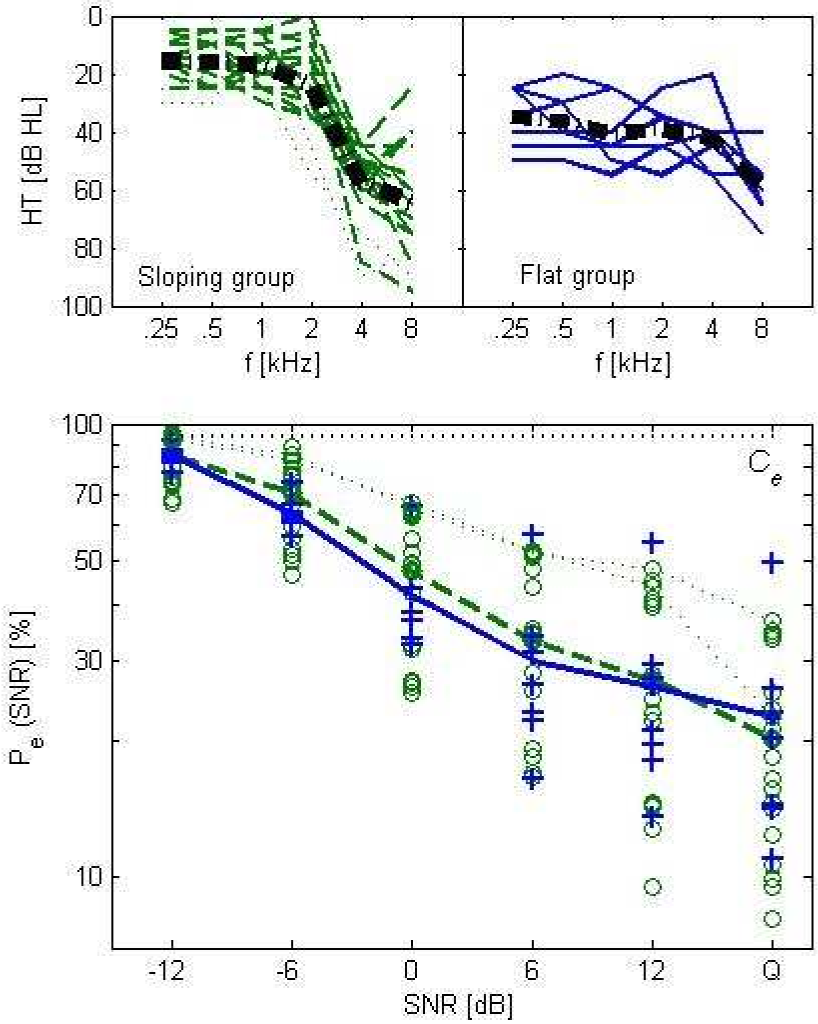


Figure 1: Hearing thresholds [HT (f)] and the percent error scores [Pe (SNR)] for the two audiogram-based groups. The upper panels are audiograms, categorized by configuration of HT (f): sloping group (18 ears, left panel) and flat group (7 ears, right panel). The average HT (f) is indicated by a thick line. The lower panel shows the Pe (SNR) per listener. The mean Pe (SNR) for each group is shown by a thick line. For both top and bottom panels, data of two listeners having an audiogram with a slope of ≥ 30 dB/oct for 1 kHz ≤ f ≤ 4 kHz from the sloping group are indicated by thin-dotted lines. Ce is chance performance.
150x185mm (72 x 72 DPI)

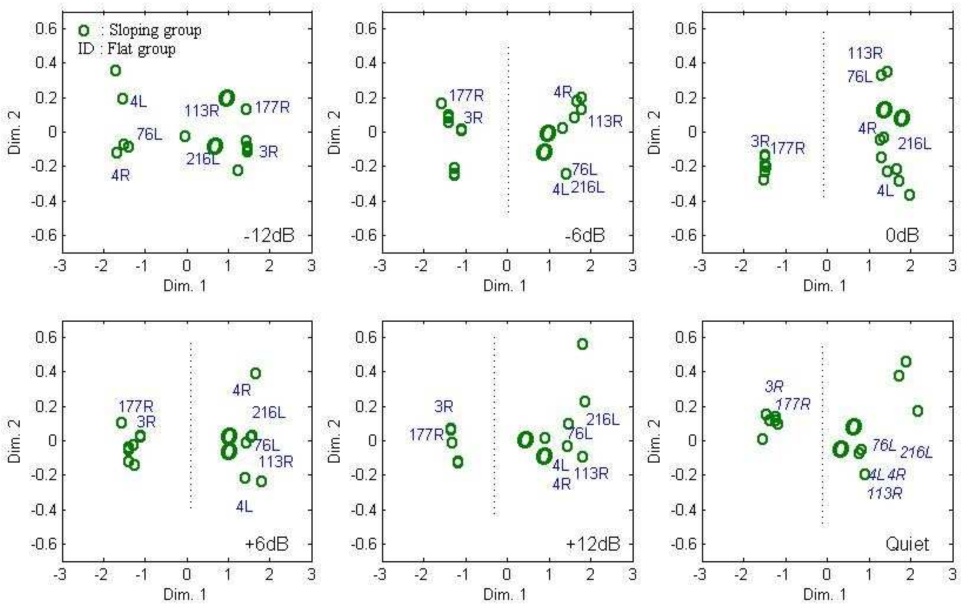


Figure 2: Subject cluster, defined by SINFA-based INDSCAL analysis for each SNR. Subjects in the sloping group (n=18) are represented by open circles, while subjects in the flat group (n=7) is are represented by their IDs. Two listeners with PTA > 30 dB/oct are represented by thicker circles. Groups 1 and 2 were assigned in the quiet condition (lower right panel) for comparison with a study of Bilger and Wang (1976).
300x185mm (72 x 72 DPI)

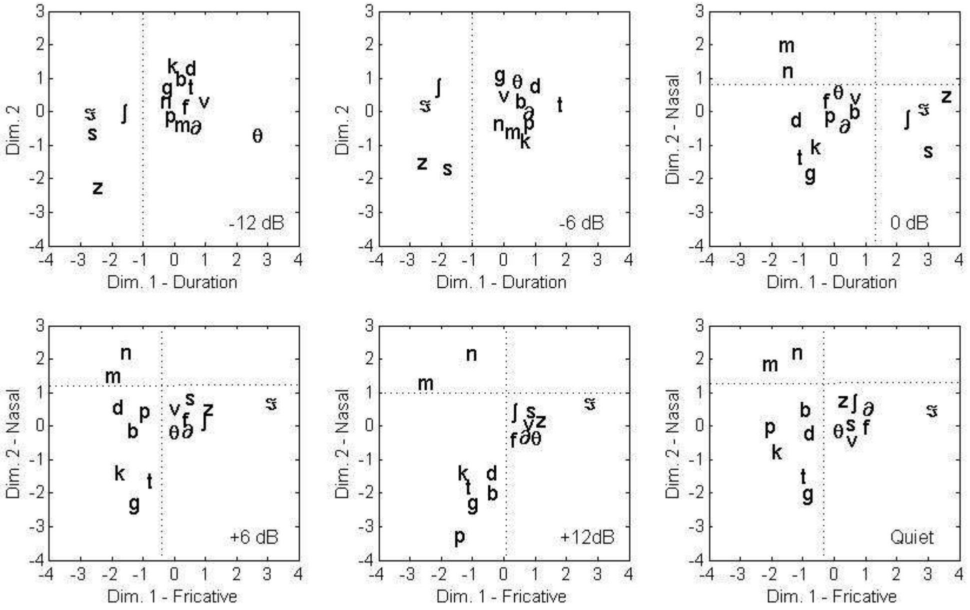


Figure 3: Group stimulus, derived by the CM-based INDSCAL model at each SNR. Dimension 2 is not precisely determined at SNRs = −12 dB and −6 dB.
300x185mm (72 x 72 DPI)

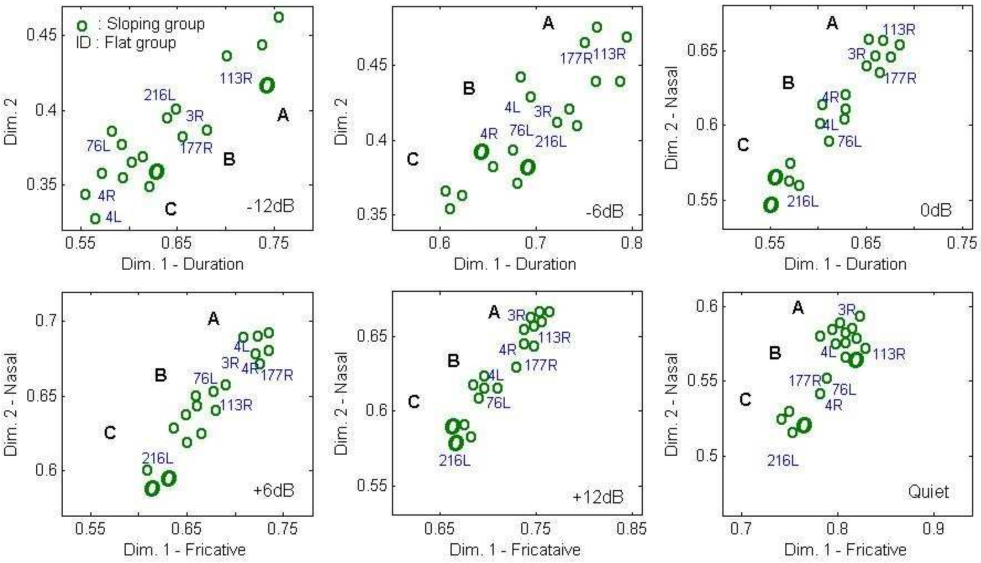


Figure 4: Listener distributions in subject distance space, assessed by the CM-based INDSCAL model at each SNR. Members of the sloping group are denoted with open circles, while flat group members are denoted with their IDs. At each SNR, each data cluster is labeled as A, B, and C, although actual subjects within each cluster vary according to SNR. Two sloping group listeners with slopes ≥ 30 dB/oct for 1 kHz ≤ f ≤ 4 kHz are denoted by thicker circles. For better visualization of grouping, the abscissa and the ordinate are scaled differently in each SNR panel, but ranges of both axis limits are constant across SNR.
319x182mm (72 x 72 DPI)

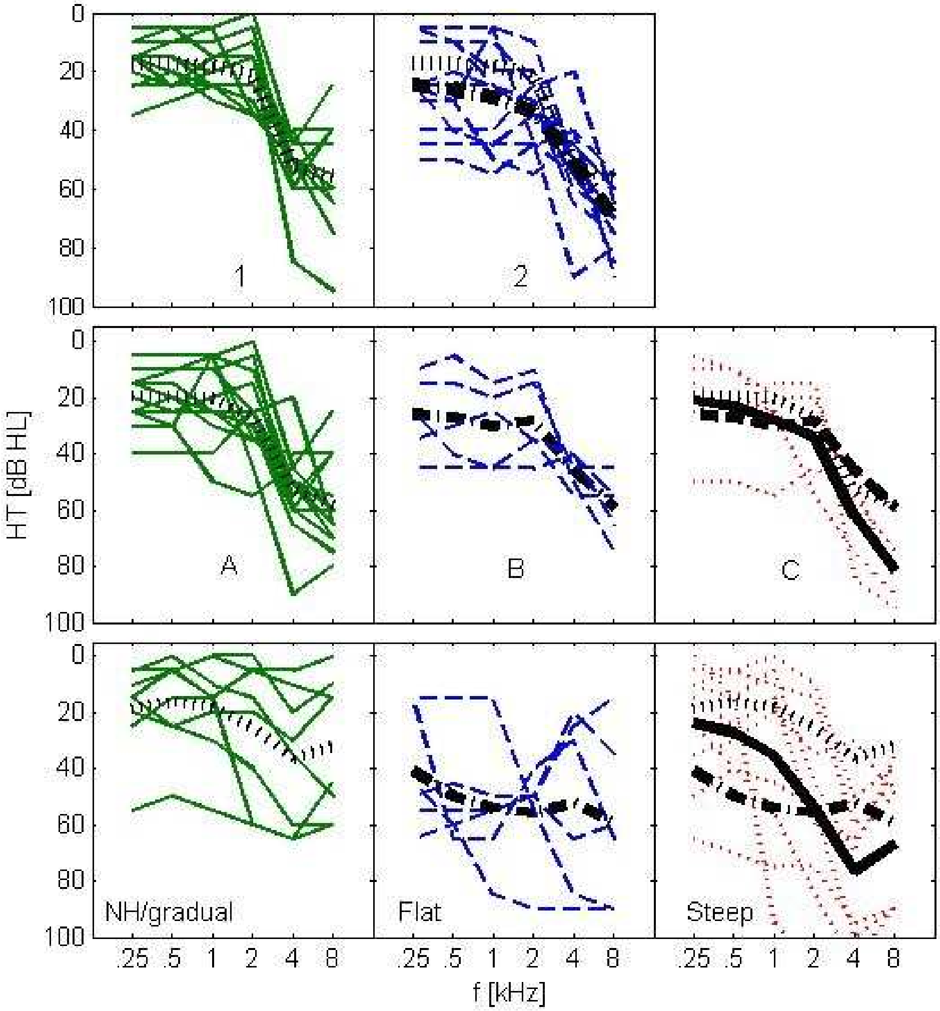


Figure 5: Audiograms, categorized by INDSCAL. The top panels are audiograms from the current study, defined by the SINFA-based INDSCAL model in quiet. There are 12 and 13 listeners in panels 1 and 2. Middle panels are audiograms from the current study that are grouped by the CM-based INDSCAL model in quiet. There are 15, 5, and 5 listeners in panels A, B, and C, respectively. The bottom panels are pairwise multidimensional scaling-based HI groups for CVs presented in quiet, reported by Bilger and Wang (1976). Eight ears were classified as belonging to the NH/gradual group, 6 ears as the flat group, and 9 ears as the steep group. Average thresholds for all three groups are shown in the panels to the right for purpose of comparison.
197x207mm (72 x 72 DPI)

Table 1. Descriptive information for listeners. Each listener is identified by ID number + ear tested ([R: Right or L: Left]). Three listeners whose performance was tested monaurally in both ears are indicated by ID plus R/L. Differences in number of listeners, ears, and audiograms vary because of the 3 listeners who were tested bilaterally. Listeners were divided into two groups on the basis of audiometric configuration: sloping group (18 ears) and flat group (7 ears).

	Sloping group						Flat group			Total
	ID	Gender	Age	ID	Gender	Age	ID	Gender	Age	
	1L	F	21	148L	M	60	3R	M	21	
	2L/R	F	59	170R	M	53	4L/R	F	63	
	12L	F	39	188R	M	64	76L	F	62	
	39L	M	63	195L	F	60	113R	M	48	
	48R	M	62	200L/R	F	52	177R	F	39	
	71L	M	60	208L	F	54	216L	F	58	
	112R	F	54	300L	M	54				
	134L	F	52	301R	M	58				
# of listener	16						6			22
# of ear	18						7			25

Table 2. Retaining rate for SINFA-based (top) and CM-based (bottom) INDSCAL groups over SNRs. First row and second column indicate SNRs. The proportion of listeners that remained in the same group out of 25 listeners is given with the number of listeners in parenthesis. The set of diagonal cells formed along the bottom of each column specifies the retaining rate for adjacent SNRs and Q.

	SNR	12 dB	6 dB	0 dB	−6 dB	−12 dB
SINFA-based INDSCAL	Q	92% (23)	92% (23)	80% (20)	92% (23)	68% (17)
	12 dB		92% (23)	72% (18)	92% (23)	56% (14)
	6 dB			72% (18)	100% (25)	64% (16)
	0 dB				72% (18)	60% (15)
	−6 dB					64% (16)
CM-based INDSCAL	SNR	12 dB	6 dB	0 dB	−6 dB	−12 dB
	Q	52% (13)	40% (10)	40% (10)	24% (6)	8% (2)
	12 dB		68% (17)	60% (15)	36% (9)	12% (3)
	6 dB			76% (19)	44% (11)	12% (3)
	0 dB				60% (15)	20% (5)
	−6 dB					44% (11)