



Published in final edited form as:

J Speech Lang Hear Res. 2011 August ; 54(4): 1211–1223. doi:10.1044/1092-4388(2010/09-0218).

Vowel identification by hearing-impaired listeners in response to variation in formant frequencies ^{a)}

Michelle R. Molis and **Marjorie R. Leek**

National Center for Rehabilitative Auditory Research, Portland VA Medical Center, 3710 SW US Veterans Hospital Road, Portland, Oregon 97207

Abstract

Purpose—This study examined the influence of presentation level and mild-to-moderate hearing loss on the identification of a set of vowel tokens systematically varying in the frequency locations of their second and third formants.

Method—Five normally-hearing (NH) and five hearing-impaired (HI) listeners identified synthesized vowels that represented both highly-identifiable and ambiguous examples of /ɪ/, /ʊ/, and /ɜ/.

Results—Response patterns of NH listeners showed significant changes with an increase in presentation level from 75 to 95 dB SPL, including increased category overlap. HI listeners, listening only at the higher level, showed greater category overlap than normal, and overall identification patterns that differed significantly from those of NH listeners. Excitation patterns based on estimates of auditory filters suggested smoothing of the internal representations, resulting in impaired formant resolution.

Conclusions—Both increased presentation level for NH listeners and the presence of hearing loss produced a significant change in vowel identification for this stimulus set. Major differences were observed between NH and HI listeners in vowel category overlap and in the sharpness of boundaries between vowel tokens. It is likely that these findings reflect imprecise internal spectral representations due to reduced frequency selectivity.

INTRODUCTION

Vowel perception by listeners with hearing loss is reported to be robust and only affected by the most severe losses (Nábělek, Czyzewaski, & Krishnan, 1992; Owens, Talbott, & Schubert, 1968; Pickett et al., 1970; Van Tasell, Fabry, & Thibodeau, 1987). The shape of the frequency spectrum is one of the prime determinants of vowel identity (Peterson & Barney, 1952), especially the location of peaks in the spectrum corresponding to vocal tract resonances, or formants (Molis, 2005). When vowel errors are noted, they are usually systematic confusions of adjacent categories that have similar formant frequencies (Hack & Erber, 1982; Owens et al., 1968; Richie Kewley-Port, & Coughlin, 2003; Van Tasell et al., 1987). A potential source of these confusions is the loss of spectral contrast in the internal cochlear representation of the vowel spectra (Leek, Dorman, & Summerfield, 1987).

^{a)}Portions of this research were presented at the 146th meeting of the Acoustical Society of America (J. Acoust. Soc. Am., 114, 2359) and the 2006 meeting of the American Auditory Society, Scottsdale, AZ.

Corresponding author: Michelle Molis, National Center for Rehabilitative Auditory Research, Portland VA Medical Center, 3710 SW US Veterans Hospital Road (NCRAR), Portland, Oregon 97207, michelle.molis@va.gov, Tel: 503-220-8262 x55574; Fax: 503-721-1402.

Changes in the function of the auditory periphery caused by hearing impairment include reduction of sensitivity and frequency selectivity, resulting in smaller amplitude differences between the peaks and valleys in the internal representation of the vowel spectra (Moore, 1995; Leek et al., 1987). Across-frequency masking patterns using vowels as maskers show less resolution of the frequencies associated with formant peaks in the maskers for hearing-impaired (HI) listeners than for normal-hearing (NH) listeners (Bacon & Brandt, 1982; Sidwell & Summerfield, 1985; Van Tasell et al., 1987). Turner and Henn (1989) assessed frequency resolution and vowel recognition, and determined that the poorer performance of HI listeners on identification of vowels was related to their poorer frequency selectivity. In another study of spectral peak resolution by HI listeners, Henry et al. (2005) reported a significant, albeit weak, relationship between discrimination of rippled noises, characterized by peaks and valleys in their spectra, and vowel and consonant identification. Performance was not significantly related to absolute thresholds. These authors also remarked that whereas speech understanding in quiet is not highly dependent on good frequency resolution, speech perception in background noise may be.

Impairments in frequency resolution produce a reduction in spectral contrasts that define the formant frequencies of vowels in the cochlear representation, resulting in a flatter-than-normal excitation pattern, and reducing or eliminating the representation of spectral peaks required for vowel identification (Leek et al., 1987; Alcántara & Moore, 1995; Leek & Summers, 1996). NH listeners typically require less contrast between spectral peaks and valleys for accurate identification or discrimination of synthetic, vowel-like stimuli than do HI listeners (Leek et al., 1987; Leek & Summers, 1993, 1996; Dreisbach, Leek, & Lentz, 2005), although the amount of this difference varies depending on a number of other stimulus and listener factors. The difference in performance between the two groups of listeners reflects the smearing and flattening of the internal spectrum. Leek et al. noted, however, that most naturally-produced vowels in real-life environments have more spectral contrast than has been tested in these studies.

The goal of this study was to measure the effect of hearing loss on identification of vowels that may or not be good category tokens. If vowel identification performance is assessed only with single, unambiguous vowel category tokens, much of the difference in vowel recognition between the two listener groups will be masked and any difficulty HI listeners have with vowel perception will be underestimated. Here, vowel identification by NH and HI listeners was measured for a densely-sampled stimulus space encompassing three vowel categories. The stimuli varied systematically in their second and third formants and represented either unambiguous examples of the response categories (defined as those consistently labeled by NH listeners) or ambiguous vowels that could reasonably be identified as belonging to multiple response categories by NH listeners. The relatively dense and quasi-uniform sampling of the stimulus space allowed for comparison of the identification response patterns a) of NH listeners at different presentation levels; and, b) between NH and HI listeners across a range of stimuli. In addition, in order to gauge the influence of impaired frequency resolution on vowel identification, excitation patterns were constructed based on estimates of auditory filter bandwidths of the HI listeners as a model of the internal spectral representation of the stimuli.

METHOD

Listeners

Five NH listeners (1 man and 4 women) and five HI listeners (2 men and 3 women) served as paid participants. No attempt was made to control for dialect differences. NH listeners ranged between the ages of 27 and 58 years ($M = 46$ yrs.). HI listeners ranged between the ages of 61 and 80 years ($M = 72$ yrs.) and had mild to moderate sensorineural hearing losses

in the test ear (i.e., air-conduction thresholds between 30 and 65 dB HL at audiometric frequencies from 0.25 to 4 kHz, air-bone gaps of ≤ 10 dB from 0.5 to 4 kHz, and a normal tympanogram). Figure 1 shows individual air-conduction thresholds for the HI listeners and average thresholds of the NH group, along with standard errors of the mean. All listeners provided written informed consent.

Vowel Identification Task

Stimuli—The stimuli for this study were used previously to explore vowel categorization by NH listeners and are described in greater detail elsewhere (Maddox, Molis, & Diehl, 2002; Molis, 2005). Briefly, the stimuli were fifty-four vowels with five steady formants, synthesized using a cascade resonance synthesizer implemented on a PC (Klatt & Klatt, 1990). The second and third formant frequencies (F2 and F3) varied in equal 0.4 Bark steps: 9.0 to 13.4 Bark (1081 Hz to 2120 Hz) for F2 and 10.0 to 15.2 Bark (1268 Hz to 2783 Hz) for F3. Exemplars of the American English vowel categories /ɪ/, /ʊ/, and /ɚ/ can be found in this stimulus region. All stimuli had the same first, fourth, and fifth formant frequencies, the same fundamental frequency contour and duration (225 ms), and were normalized for equal RMS amplitude.

Procedures—Listeners were tested individually in a sound attenuated room. Stimuli were presented monaurally over Beyer DT-100 headphones in blocked conditions of 75 and 95 dB SPL for NH listeners and 95 dB SPL for HI listeners. Three of the NH listeners heard the 95 dB stimuli first and two heard the 75 dB stimuli first. For each condition, two blocks of seven replications each (14 total) were presented (756 trials/listener) with a short break between blocks. On each trial, listeners were asked to identify a stimulus by pressing one of three response buttons on a button box labeled with the key words "hid", "hood", and "heard". They were assured that there were no "right" or "wrong" answers and were encouraged to label ambiguous stimuli as the category they felt was the most appropriate.

To verify that items in the stimulus set were reasonably discriminable, three of the HI listeners and all five NH listeners completed a preliminary discrimination task. Three stimuli from the extreme corners of the stimulus space, representing one instance of each of the three possible response categories, were selected and presented to listeners in thirty randomized pairings (including identical pairs) at a level of 95 dB SPL. Listeners were asked to indicate with a button press whether the stimuli in each pair were the same or different. All listeners were able to discriminate between the stimuli with at least 80% accuracy.

Statistical Analyses—The number of times a listener chose each response category out of 14 trials was used as an estimate of the probability of the listener's vowel identification at each F2 and F3 combination. The statistical objective was to model these probabilities in the F2/F3 plane. The approach taken was to treat this as a response surface regression problem (Neter, Kutner, Nachtsheim, & Wasserman, 1996). Tests could thus be conducted of the effect of either presentation level (Level-Effect model) or listener group (Group-Effect model) on the shapes of the mean choice probability surfaces for each response category at each F2/F3 combination. Furthermore, tests of differences at particular F2/F3 values could be conducted based on linear contrasts of the fitted model (Level- or Group-Effect).

For each general model (Level- or Group-Effect), the selection probabilities for each of the vowel categories (/ɪ/, /ʊ/, or /ɚ/) were modeled using multivariate response regression (Hastie, Tibshirani, & Friedman, 2009). This is analogous to MANOVA with the exception that the F2 and F3 values were treated as continuous covariates instead of categorical predictors. An arcsine transform [$Y' = 2\arcsine(\text{proportion})^{1/2}$] was used in order to stabilize

the variance of the proportional data (Neter et al., 1996). To allow for variability among listeners' response profiles, the models also included random effects. Specifically, we modeled the mean arcsine-square root proportion of each listener's responses that identified each stimulus as one of three vowel category options (/ɪ/, /ʊ/, or /ə/) using a general linear mixed model. The Level-Effect and Group-Effect models were each composed of three surfaces, one per response category. Each surface was composed of the fixed effects in the second-order response surface with F2, F3, F2×F3, F2², and F3² values. Presentation level effects or listener group effects and their interactions with response surface terms were included in the Level-Effect and Group-Effect models respectively. Random coefficients for the response surface components were included to allow for individual variability around the overall response surface. Correlation among the three response category surfaces was modeled using an unstructured covariance model for the residuals, which is identical to the MANOVA approach. The Level-Effect model fit to the NH subjects' data included 36 fixed effects parameters and 9 covariance parameters (total = 45 parameters) to model the 1,080 recorded responses. The Group-Effect model fit to the full sample included 36 fixed effects parameters and 16 covariance parameters (total = 52 parameters) to model the 1,620 recorded responses in both the NH and HI subjects.

Using the Level-Effect or Group-Effect model respectively, tests of presentation level effects within the NH group and tests of differences between listener groups were made using the likelihood ratio statistic. For example, an omnibus test (Neter et al., 1996) of any presentation level effects on the overall response surface among NH is equal to $2(\log-L_{\text{Full}} - \log-L_{\text{No Level Effects}})$, where $\log-L_{\text{Full}}$ and $\log-L_{\text{No Level Effect}}$ are the log-likelihoods of the fitted multivariate response regression models with and without the level effects, respectively. Under the null hypothesis of no level effect, this statistic has a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters between the full model and the model without any level effects. This statistic formally tests whether the model fits the data significantly more poorly if one ignores all effects of level among NH listeners. A similar omnibus test for listener group differences was constructed by replacing the level effect with the group effect in the likelihood ratio statistic.

Measurement of auditory filters and excitation pattern construction

Auditory filters were measured for the HI listeners at center frequencies of 1000 and 2000 Hz using a notched-noise masking method (Glasberg & Moore, 2000; Rosen & Baker, 1994). Filters at 500 Hz were also measured for two listeners (HI1 and HI4). Masked thresholds were submitted to a rounded exponential (ROEX) modeling procedure to extract equivalent rectangular bandwidths (ERB) of the measured auditory filters (Patterson, Nimmo-Smith, Weber, & Milroy, 1982).

Stimuli—Tones of 500, 1000, and 2000 Hz were used as signals. Steady-state durations of the tones were 300 ms, including 50 ms cosine-squared rise and fall ramps. The notched-noise maskers were created by adding together two bands of noise located on either side of the signal frequency. The width of each noise band was 0.4 times the signal frequency, and the bands were generated digitally for each noise presentation. Maskers were 400 ms in duration. When the signal was present along with the masker, it was temporally centered within the masker duration.

Thresholds at each signal frequency were measured in eight notched-noise conditions, with the maskers either centered symmetrically around the signal frequency (six maskers) or asymmetrically with one band closer to the signal frequency than the other (two maskers). Each notched-noise masker was constructed by adding together two noise bands, one on either side of the signal frequency, with a given frequency difference between the signal

frequency and the near edge of each noise band. These values are expressed as normalized frequency, defined as the absolute value of the ratio of the deviation of the near edges of the maskers from the signal frequency (Δf) and the signal frequency (f_s). Values of $|\Delta f/f_s|$ were 0, 0.05, 0.10, 0.20, 0.30, and 0.40 on each side of the signal frequency for symmetric notches. The two asymmetric notched noises were placed at $|\Delta f/f_s|=0.2$ and 0.4.

Instrumentation—Pure tone signals were generated by a Tucker-Davis-Technology (TDT) waveform generator (WG1), and gated on and off through a TDT SW2 cosine switch. Noises were constructed digitally and played through a TDT DD1 D-A converter, with a sample rate of 40,000 samples/second. Signals and noises were separately attenuated and played to listeners monaurally over a Beyer DT-100 earphone.

Procedure—Masked thresholds were measured for each listener with eight notched noise maskers at two or three signal frequencies. All thresholds at one signal frequency were measured before beginning testing on the next frequency, determined randomly for each listener. The signal level for each set of notched noise maskers was either 70 or 80 dB SPL, determined individually for each listener, depending on the quiet threshold level for the signal frequency. For all cases except one, if the quiet threshold was greater than 30 dB HL, the higher signal level was used. In one case, at a signal frequency of 500 Hz, an 80-dB signal was used even though the quiet threshold was only 25 dB HL. The masker level varied over trials according to an adaptive track.

The listener was seated in a sound-treated booth, wearing earphones and facing a touch screen response terminal. The threshold measurement procedure was a single-interval maximum likelihood procedure, originally described by Green (1993), modified as described in Leek et al. (2000). Briefly, each trial consisted of the presentation of the notched-noise and, with an 80% probability, the signal tone. On each trial, the listener was asked to indicate whether or not a tone was heard within the masking noise by touching a marked area on the terminal. After each trial presentation, a set of candidate psychometric functions was calculated based on the response to that trial and all previous trials in the block. The noise level on the next trial was taken from the 70% correct performance point on the candidate function that most closely reflected the data collected so far within the block. As the block continued, and more data became available, the choice of the successful candidate psychometric function converged on the one most likely to account for the data. At the end of the block of trials, the level of noise necessary to produce 70% correct detections of the signal at the specified level was calculated from the final estimated psychometric function, and this value was taken as the threshold estimate for that block of trials. The average of three such measurements was taken as the final threshold value for that experimental condition. If the standard deviation across the three estimates was larger than 5 dB, that set of thresholds was abandoned and a second set of three thresholds was measured. Typically, a set of three threshold estimates could be measured using this procedure in under five minutes.

Filter characteristics were estimated using a one-parameter ROEX (p) fitting procedure described by Patterson et al. (1982). The slope parameter, p , was allowed to differ on each side of the filter to permit asymmetric filters to be specified. The iterative procedure finds the best-fitting set of parameters to define the ROEX function given all eight threshold measurements for each center frequency. The frequency response of the earphone was taken into consideration in the fitting procedure. Comparison of the predicted thresholds based on the fitted filter parameters to the actual data indicated that the filters were well-fit, with an average RMS difference between the thresholds predicted by the fitted filter and the observed thresholds of about 1 dB. The ERB for each filter was calculated from the slopes of the filters determined by the fitting process.

RESULTS

Vowel Identification

Presentation Level—The vowel identification response patterns were similar among the NH listeners; therefore, their pooled identifications at each presentation level are shown in Figure 2. The response counts for each vowel category are displayed in separate panels and are arrayed according to second and third formant frequencies expressed in Bark. In each panel, observed response frequencies are displayed in terms of bubble size, i.e., larger bubbles indicate higher response counts, whereas smaller bubbles indicate less frequent responses. The light gray lines indicate the locations of linear boundaries between categories arising from the identifications by a previous NH group listening at 70 dB (Maddox et al., 2002) and are included to provide landmarks for comparison.

At the lower presentation level (left column), response distributions were similar to those for a different group of NH listeners (Maddox et al., 2002; Molis, 2005)—the three vowel categories were concentrated in three distinct regions of the stimulus space. At the higher presentation level (right column), the response regions expanded slightly as the overlap between categories increased.

Although there is no accepted goodness-of-fit test for continuous, multivariate outcomes modeled with random effects, the goodness-of-fit of multivariate response surface models can be verified through visual inspection of a graphical representation of the average observed selection rate and the predicted selection rate. This evaluation carried out for the three response categories among NH listeners at low and high presentation levels revealed that there was considerable overlap between the prediction and observed responses, indicating that the overall model fit was quite good and the overall pattern of responses at each stimulus level was well described by the model.

The model results were used to compute the likelihood ratio statistic for the null hypothesis of no level effects on the category response surfaces. The result is a statistic of 40.3, which, on 20 degrees of freedom, yields a p-value of 0.004, indicating that there were significant differences in the response surfaces for the two presentation levels among the NH listeners.

Listener Group—The HI listeners' response patterns were more diverse than those of the NH listeners; therefore, their response counts are presented individually in columns in Figure 3. The gray lines show the same category boundaries as Figure 2 reflecting performance of another NH group. The high-frequency pure tone average of thresholds at 1, 2, and 3 kHz (PTA_{123}) is also indicated for each listener. For the most part, the category identifications of HI1 and HI2 were concentrated in separate regions of the stimulus space, similar to the NH listeners but with more overlap between the categories. For the remaining HI listeners, there appeared to be even greater overlap among identifications of the three vowel categories, with quite variable responses in some of the categories for a few listeners (e.g., HI4 / *ɪ* / and HI5 / *ɜ* /).

The goodness-of-fit of the multivariate response surface model to the response category probabilities among both HI and NH listener groups at the 95 dB presentation level was verified graphically. The model showed reasonably good fit to the data for both listener groups. The likelihood ratio test yielded a statistic of 161.4 on 16 degrees of freedom ($p < 0.001$); therefore, there is strong evidence that NH and HI listeners differed in their vowel identification surfaces over these F2 and F3 combinations.

In order to visualize the differences in the group three-dimensional response surfaces ($F2 \times F3 \times$ response probability), the predicted response category probabilities were plotted as

two-dimensional equiprobability contours. The relationship between a three-dimensional response surface and its two-dimensional representation is shown in Figure 4. On the right is a plot of a fitted response surface for the NH listeners' /*ʊ*/ response and on the left is the projection of this surface onto the stimulus space. Lines on both plots connect formant values for which the predicted response probability is equal. Because the response surface is steeply sloping from the region of highest to lowest response probability, the equiprobability contours appear relatively closely spaced in the two-dimensional representation.

Figure 5 shows contour plots of the predicted probabilities for the three response categories. The response probabilities of the NH group are presented in the left column, and those of the HI group are on the right. This figure depicts the major difference between HI and NH listeners—HI listeners have more flattened response surfaces over the entire F2/F3 plane than do NH listeners. This flattening is observable by two characteristics of the response probability contours for all three categories: the maximum predicted probabilities of the HI group are not as great as those for the NH group (for example, the predicted HI response probabilities for /*ɪ*/ and /*ɚ*/ never reach 0.9 as they do for the NH probabilities); and the equiprobability contours are more spread out for the HI listeners than for the NH listeners, indicating more gradual slopes of the response surfaces of the HI listeners.

The difference in the predicted selection rates between the HI and NH listeners at each point on the F2/F3 space was assessed by means of tests derived from linear contrasts of the parameters in the fitted model. This resulted in a total of 162 tests (54 per response category), which had considerable risk of Type I error. In order to address this, a false discovery rate (FDR) p-value adjustment was used to guarantee that the proportion of Type I errors was no greater than 5% (Benjamini & Hochberg, 1995; Verhoeven, Simonsen, & McIntyre, 2005). This procedure is much less conservative than the usual Bonferroni adjustment, and allowed identification of regions in the stimulus space where differences between HI and NH listeners lay. For each response category, stimuli for which the predicted response probabilities were significantly different (FDR-adjusted $p < 0.05$) between the two listener groups are indicated by symbols placed at the F2/F3 stimulus locations on the panels depicting the HI responses on Figure 5.

The predicted response probabilities for /*ɪ*/ are shown in the top panels of Figure 5. There was little difference in the overall shape of the response surfaces for this category. For both listener groups the greatest response probability was in the upper right-hand corner where F2 and F3 had the highest values. There was no significant difference between the two groups for the stimuli located where the probability of an /*ɪ*/ response by HI listeners was at least 0.8. However, the point-by-point analysis reveals that there were significant differences across the portion of the stimulus space where the probability of an /*ɪ*/ response by HI listeners was between 0.2 and 0.8 (15 out of 54 stimulus comparisons). This is a result of the overall attenuation of the response surface for the HI listeners—the range between the highest and lowest probabilities is reduced. There were no differences in the areas of lowest response probability.

The predicted response probabilities for /*ʊ*/ are shown in the middle panels of Figure 5. Again, the maximum response probabilities were in roughly the same location for the two groups and there were not significant differences at the peak probability regions. For this category, the significant differences are in a boundary region between high and low response probability (13 out of 54 stimulus comparisons). As for the /*ɪ*/ response, there were no differences in the regions of lowest probability for high values of F3.

The biggest differences in the shapes of the response surfaces were observed for /*ɚ*/ (the bottom panels of Figure 5). For this category, there was a noticeable shift in the highest

probability response between the two groups from the extreme lower left for the NH listeners to a region with a higher F2 and F3 for the HI listeners. This pattern is observable in the raw data for a number of the HI listeners (see Figure 3). The shift in response pattern between the two groups resulted in significant differences at the locations of the peak in response for the NH listeners as well as for a large region of the stimulus space where the probability of an / σ / response was low for the NH listeners (34 out of 54 comparisons).

Auditory Filter Shapes and Excitation Patterns

Figure 6 displays estimates of auditory filter bandwidth for each HI listener at the frequencies tested and the best linear fits to these estimates. The line connecting the filled symbols indicates the auditory filter bandwidth estimates calculated for NH listeners at similar presentation levels reported by Glasberg and Moore (1990), showing that filters typically broaden with increasing frequency with a slope of about 0.11. Listener HI2 had a shallow slope (0.07); however, the bandwidths were broader than normal. For the remaining four HI listeners, filters broadened with frequency at a faster rate than for NH listeners, with slopes ranging from 0.29 to 0.80. The two listeners whose filters were also measured at 500 Hz (HI1 and HI4) had near-normal bandwidths at that frequency.

The auditory filter measurements may be used to calculate excitation patterns in response to the vowel stimuli, following procedures described by Glasberg and Moore (1990). For this analysis, the equations estimating the changes in ERB with auditory filter center frequency for each listener, shown in Figure 5, were used to construct a bank of auditory filters ranging from about 200 to about 5000 Hz. For simplicity, only symmetric auditory filters were used in this analysis. Each of the 54 stimuli was processed through the simulated filterbanks of the individual listeners. No accommodation was made within the excitation pattern calculation for changes in bandwidth with level because the nonlinearity simulated by varying the bandwidth with level is essentially lost in hearing-impaired cochleas (Moore, 2007).

Figure 7 compares the estimated excitation patterns above 600 Hz for a simulated NH listener (solid black line) and for two of the HI listeners, one with relatively little hearing loss (HI1: dashed black line) and one with more hearing loss (HI4: dashed-dotted black line). Although only frequencies in the region of about 600 to 5000 Hz are considered here, the NH excitation pattern for frequencies below 600 Hz is included as a reference (solid gray line). Shown are (a) a stimulus most often identified by NH listeners as / ɪ / (high F2/high F3); (b) a stimulus most often identified by NH listeners as / ʊ / (low F2/high F3); and (c) a stimulus most often identified by NH listeners as / ɚ / (low F2/low F3).

Formant peaks for F2 and F3 are easily identified through visual inspection of the excitation patterns of the simulated NH listener; however, the locations of the formant peaks in the excitation patterns of the HI listeners are not as easily determined. This is a reflection of the broader HI auditory filters. The lack of clear definition of formant peaks in the F2/F3 region probably is largely responsible for the generally poorer and less consistent identifications of the vowel sets.

DISCUSSION

In this study, listeners labeled both clearly-identifiable and ambiguous exemplars from three vowel categories. The category responses of the NH listeners at the 75 dB presentation level were confined to generally non-overlapping regions of the stimulus space. This is consistent with data from previous studies of NH listeners in response to these stimuli (Maddox et al., 2002; Molis, 2005). Response overlap among categories increased significantly with an increase in presentation level to 95 dB. The purpose of collecting data at the higher level

was to facilitate comparisons with the HI listeners, who required this level in order to assure audibility of the stimulus set. However, when the response patterns of the HI listeners were compared with those of the NH listeners at this presentation level, differences were still observed between groups.

Although HI listeners' responses to the 'best' category examples—those selected most often by NH listeners—were broadly similar to those of NH listeners, their response areas were less compact and there was greater category overlap. The increased overlap is a result of increased variability of the responses of the HI group and is observable in two characteristics of the fitted response probability surfaces for all three categories: the maximum response probabilities are not as high and the slopes of the response surface were not as steep for the HI group as for the NH group. For two of the vowel categories, /ɪ/ and /ʊ/, the two listener groups did not differ where the predicted response probabilities were the highest or were very low. Differences between the responses of the two groups occurred primarily in the boundary regions between the categories. For the remaining category, /æ/ an overall shift in the response pattern of the HI listeners resulted in more widespread differences between the two groups.

Possible influences on vowel perception

There are several possible sources of any of the differences observed between the response patterns of the two listener groups, including formant audibility, presentation level, frequency selectivity, and age.

Formant audibility—The overall presentation level for the HI listeners was 95 dB SPL. It was assumed that, at this level, F2 and F3 would be audible for most listeners; however, some frequencies could have been inaudible to some of the HI listeners for some of the stimuli. The peak in the excitation pattern for each tone at threshold was compared with the excitation level at that frequency for the excitation patterns of each stimulus and listener as described earlier. If the excitation in the formant regions for the stimuli was greater than the excitation level of the tone threshold, it was assumed that those formants were audible. This analysis revealed that F2 was always audible for every HI listener, and F3 was always audible for four of the five HI listeners. For the remaining HI listener, HI3, F3 was inaudible for three of the stimuli with a low F2 and a high F3, identified as /ʊ/ by NH listeners. However, the probable inaudibility of F3 for these stimuli would not necessarily be crucial to the identification of /ʊ/. Estimates of the effective second formant (Carlson, Granström, & Fant, 1970; Carlson, Fant, & Granström, 1975) have shown that back vowels such as /ʊ/ can be well approximated by F1 and F2 alone. For stimuli where F2 and F3 were relatively close—less than 2.6 Bark apart—both formants were audible for all HI listeners. This was true even when both formants were high in frequency—as for /ɪ/—since the levels of closely spaced formants generally increase with increasing proximity (Fant, 1956). Nevertheless, although F2 and F3 remained audible for all listeners, performance for /ɪ/ was very poor for listeners HI3 and HI4. Considered together, these observations indicate that specific formant audibility was not clearly predictive of identification performance and would likely not by itself account for the findings of this study.

High presentation level and frequency selectivity—It is possible that the high presentation level itself had a negative impact on performance for the HI listeners. Previous studies have found that high presentation levels are associated with decreased word and sentence intelligibility in quiet and in noise (French & Steinberg, 1947; Molis & Summers, 2003; Studebaker, Sherbecoe, McDaniel, & Gwaltney, 1999; Speaks, Karmen, & Benitez, 1967). However, studies that have specifically addressed vowel perception have found no detrimental effect of increased level. Coughlin et al. (1998) reported that young NH listeners

could identify four target vowels presented at 70 or 95 dB SPL with near-perfect accuracy. Elderly HI listeners performed better for vowel identification at 95 than at 70 dB SPL – perhaps an effect of reduced audibility at the lower level. In a subsequent study, neither an overall increase in presentation level from 60 to 95 dB SPL, nor shaped gain that was designed to keep the stimuli 15 dB above threshold, influenced vowel identification by HI listeners (Richie et al., 2003). It should be noted, however, that the response categories in each of these studies were represented by single vowel tokens.

In the current study, vowel categorization for the NH listeners changed significantly when the presentation level was raised from 75 to 95 dB SPL. And, in fact, their performance at the high level qualitatively resembled the performance of the HI listeners in certain respects: the category overlap was increased and boundaries were more gradual than for the lower presentation level. These changes are consistent with a reduction in spectral contrast caused by smearing of the peaks and valleys in the excitation patterns for NH listeners at high levels (e.g., Leek et al., 1987; Leek & Summers, 1996). The loss of spectral contrast is due to broadening of auditory filters as normal cochlear processing becomes more linear at high stimulus levels. A major characteristic of HI listeners is also a linearized system. Therefore, the patterns of responses for both the NH listeners at high levels and the HI listeners are likely due, at least in part, to the same underlying reduction of cochlear nonlinearity, although to a greater extent for the HI than the NH listeners.

All of the HI listeners had impaired frequency selectivity over at least a portion of the relevant frequency range as demonstrated by their broader auditory filter bandwidth estimates (Figure 6). Depending on the severity of the hearing loss, individual formant peaks can be either minimally identifiable or almost totally absent in the excitation patterns. For instance, in the excitation patterns based on the auditory filter estimates for HI4, the listener with the broadest filter estimates, there is very little differentiation for these stimuli in the F2/F3 frequency region. This listener's auditory filter bandwidth estimated at 2000 Hz was more than four times greater than for a normal auditory filter. The excitation patterns shown in Figure 7 illustrate the effect of the broad auditory filters in the frequency region of F2/F3. The formant peaks in that frequency region are clearly apparent for NH listeners, as well as for listener HI1, whose auditory filters were just over twice as broad as normal. Although somewhat smoothed by the poorer-than-normal frequency resolution, HI1's excitation pattern still reflects at least the second formant, although the higher formants are more problematic. In contrast, the excitation pattern modeled after listener HI4's auditory filters is almost flat in the F2/F3 region.

Age—The average age of the HI listener group was greater than that of the NH group (72 vs. 46 years). Previous research indicates that, after controlling for amount of hearing loss, vowel perception does not differ between older and younger listeners. Richie et al. (2003) compared the performance of a group of young HI listeners matched with a group of elderly HI listeners from an earlier study (Coughlin et al., 1998) and found no difference in vowel identification between the two listener groups when vowel categories were represented by single tokens. However, Nábělek (1988) found that age was significantly correlated with vowel identification for stimuli degraded by noise and reverberation, but not in quiet. Although in this study listeners performed the task without the penalizing effects of noise and reverberation, a case might be made that the stimuli presented here were in some respects “degraded”. These stimuli were presented with only static spectral information, with none of the other cues to vowel identity available from speech context, formant dynamics, or vowel duration. Moreover, task uncertainty was increased through the inclusion of ambiguous stimuli that could reasonably be identified as belonging to multiple response categories. The relatively advanced ages of the HI listeners might have contributed to the present findings, particularly related to these more “cognitive” degradations of the

stimuli, perhaps analogous to effects of the acoustical distortions created by noise and reverberation. It would be interesting to determine the similarities in type and degree of performance deficits related to these two very different types of perceptual degradations (cognitive versus acoustical), although the present data do not address that question.

Possible implications of ambiguity in vowel perception

The potential consequences for HI listeners of imprecise or ambiguous vowels encountered in everyday communication may be, at best, uncertainty about vowel identity, and at worst, misperception of vowel categories. In addition to simply misunderstanding, a consequence of increased perceptual uncertainty may be increased attentional effort required to process speech. In such cases, more emphasis must be placed on top-down processing as the importance of context becomes more critical. This may result in extra cognitive processing load, even if the word is eventually correctly recognized due to top-down processing. Pichora-Fuller et al. (1995) reported that elderly listeners with and without hearing loss made more use of contextual information in identifying sentence-final words presented in background noise. Further, elderly listeners had poorer recall of the words they had identified. Pichora-Fuller et al. concluded that extra cognitive resources were needed to process speech in noise because of deficits in central processing associated with aging, and that the combined effects of hearing loss and age resulted in even more effortful communication in background noise.

The increased ambiguity of vowels that are not “good” tokens, as used here, may add to the extra cognitive load required of HI listeners to support the accurate and timely analysis of everyday speech sounds. Rakerd, Seitz, and Whearty (1996) obtained measures of listening effort for HI listeners listening to speech while concurrently performing a digit memorization task. The speech listening effort was greater for HI listeners. The authors argued that a greater cognitive contribution was required for the HI listeners. Similarly, Gatehouse and Gordon (1990) reported that HI listeners had to expend greater cognitive effort to match performance on speech listening tasks than NH listeners. It appears, then, that HI listeners may have to work harder to listen to ongoing speech than NH listeners with intact auditory systems well adapted to their native speech.

CONCLUSION

By employing a densely-sampled stimulus set we were able to demonstrate significant differences in vowel identification patterns due to increases in presentation level for NH listeners and for listeners with hearing loss when compared with NH performance. For the most part, vowel tokens that were highly-identifiable by NH listeners were also categorized consistently by listeners with hearing loss. However, more ambiguous vowels showed different patterns of response for HI listeners relative to performance by NH listeners. In particular, HI listeners demonstrated greater overlap among response categories to different tokens, and less regularity in responses over repeated presentations. A number of factors may have contributed to difficulties in ambiguous vowel identification by HI listeners, including a loss in frequency selectivity, leading to smoothing of the internal representation of the spectrum.

Acknowledgments

This research was supported by grant number R01 DC 00626 from NIDCD. Some support for preparation of this article was provided by the Department of Veterans Affairs. The authors would like to thank Van Summers and Frederick Gallun for their helpful suggestions and discussions, and Garnett McMillan for statistical assistance. Portions of this work were carried out while the authors were affiliated with the Army Audiology & Speech Center, Walter Reed Army Medical Center, Washington, D.C. The opinions or assertions contained herein are the private

views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

References

- Alcántara JI, Moore BCJ. The identification of vowel-like harmonic complexes: Effects of component phase, level, and fundamental frequency. *Journal of the Acoustical Society of America*. 1995; 97:3813–3824. [PubMed: 7790659]
- Bacon SP, Brandt JF. Auditory processing of vowels by normal-hearing and hearing-impaired listeners. *Journal of Speech and Hearing Research*. 1982; 25:339–347. [PubMed: 7176605]
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*. 1995; 57:289–300.
- Carlson, R.; Fant, G.; Granström, B. Two-formant models, pitch and vowel perception. In: Fant, G.; Tatham, MAA., editors. *Auditory Analysis and Perception of Speech*. London: Academic Press; 1975. p. 55-82.
- Carlson R, Granström B, Fant G. Some studies concerning perception of isolated vowels. *STL-QPSR*. 1970; 2–3:19–35.
- Coughlin M, Kewley-Port D, Humes LE. The relation between identification and discrimination of vowels in young and elderly listeners. *Journal of the Acoustical Society of America*. 1998; 104:3597–3607. [PubMed: 9857518]
- Dreisbach LE, Leek MR, Lentz JJ. Perception of spectral contrast by hearing-impaired listeners. *Journal of Speech, Language, and Hearing Research*. 2005; 48:910–921.
- Fant, G. On the predictability of formant level and spectrum envelopes from formant frequencies. In: Lehiste, I., editor. *Readings in Acoustic Phonetics*. Cambridge, MA: MIT Press; 1967. p. 109-120.
- French NR, Steinberg JC. Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America*. 1947; 19:90–119.
- Gatehouse S, Gordon J. Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology*. 1990; 24:63–68. [PubMed: 2317602]
- Glasberg BR, Moore BC. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*. 1990; 47:103–138. [PubMed: 2228789]
- Glasberg BR, Moore BCJ. Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *Journal of the Acoustical Society of America*. 2000; 108:2318–2328. [PubMed: 11108372]
- Green DM. A maximum-likelihood method for estimating thresholds in a yes-no task. *Journal of the Acoustical Society of America*. 1993; 93:2096–2105. [PubMed: 8473622]
- Hack ZC, Erber NP. Auditory, visual, and auditory-visual perception of vowels by hearing-impaired children. *Journal of Speech and Hearing Research*. 1982; 25:100–107. [PubMed: 7087410]
- Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*. 2nd ed. New York: Springer Series in Statistics; 2009.
- Henry BA, Turner CW, Behrens A. Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. *Journal of the Acoustical Society of America*. 2005; 118:1111–1121. [PubMed: 16158665]
- Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*. 1990; 87:820–857. [PubMed: 2137837]
- Leek MR, Dorman MF, Summerfield Q. Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*. 1987; 81:148–154. [PubMed: 3819173]
- Leek MR, Dubno JR, He N, Ahlstrom JB. Experience with a yes-no single-interval maximum-likelihood procedure. *Journal of the Acoustical Society of America*. 2000; 107:2674–2684. [PubMed: 10830389]
- Leek MR, Summers V. The effect of temporal waveform shape on spectral discrimination by normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*. 1993; 94:2074–2082. [PubMed: 8227749]

- Leek MR, Summers V. Reduced frequency selectivity and the preservation of spectral contrast in noise. *Journal of the Acoustical Society of America*. 1996; 100:1796–1806. [PubMed: 8817905]
- Maddox WT, Molis MR, Diehl RL. Generalizing a neuropsychological model of visual categorization to auditory categorization of vowels. *Perception & Psychophysics*. 2002; 64:584–597. [PubMed: 12132760]
- Molis MR. Evaluating models of vowel perception. *Journal of the Acoustical Society of America*. 2005; 118:1062–1071. [PubMed: 16158661]
- Molis MR, Summers V. Effects of high presentation levels on recognition of low- and high-frequency speech. *Acoustics Research Letters Online*. 2003; 4:124–128.
- Moore, BCJ. *Perceptual Consequences of Cochlear Damage*. Oxford: Oxford University Press; 1995.
- Moore, BCJ. *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*. Chichester, England: John Wiley & Sons, Ltd.; 2007.
- Nábělek AK. Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss. *Journal of the Acoustical Society of America*. 1988; 84:476–484. [PubMed: 3170940]
- Nábělek AK, Czyzewski Z, Krishnan LA. The influence of talker differences on vowel identification by normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*. 1992; 92:1228–1246. [PubMed: 1401512]
- Neter, J.; Kutner, M.; Nachtsheim, C.J.; Wasserman, W. *Applied Linear Statistical Models*. 4th ed. Chicago: Irwin; 1996.
- Owens E, Talbot CB, Schubert ED. Vowel discrimination of hearing-impaired listeners. *Journal of Speech and Hearing Research*. 1968; 11:648–655. [PubMed: 5722492]
- Patterson RD, Nimmo-Smith I, Weber DL, Milroy R. The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *Journal of the Acoustical Society of America*. 1982; 72:1788–1803. [PubMed: 7153426]
- Peterson GE, Barney HL. Control methods used in the study of vowels. *Journal of the Acoustical Society of America*. 1952; 24:175–184.
- Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*. 1995; 97:593–608. [PubMed: 7860836]
- Pickett, J.M.; Martin, E.S.; Johnson, D.; Brand Smith, S.; Daniel, Z.; Willis, D.; Otis, W. On patterns of speech feature reception by deaf listeners. In: Fant, G., editor. *Speech Communication Ability and Profound Deafness*. Washington, DC: Alexander Graham Bell Association for the Deaf; 1970. p. 119-134.
- Rakerd B, Seitz PF, Whearty M. Assessing the cognitive demands of speech listening for people with hearing losses. *Ear & Hearing*. 1996; 17:97–106. [PubMed: 8698163]
- Richie C, Kewley-Port D, Coughlin M. Discrimination and identification of vowels by young, hearing-impaired adults. *Journal of the Acoustical Society of America*. 2003; 114:2923–2933. [PubMed: 14650026]
- Rosen S, Baker RJ. Characterising auditory filter nonlinearity. *Hearing Research*. 1994; 73:231–243. [PubMed: 8188552]
- Sidwell A, Summerfield Q. The effect of enhanced spectral contrast on the internal representation of vowel-shaped noise. *Journal of the Acoustical Society of America*. 1985; 78:495–506.
- Speaks C, Karmen JL, Benitez L. Effect of a competing message on synthetic sentence identification. *Journal of Speech and Hearing Research*. 1967; 10:390–395. [PubMed: 6082373]
- Turner CW, Henn CC. The relation between vowel recognition and measures of frequency resolution. *Journal of Speech and Hearing Research*. 1989; 32:49–58. [PubMed: 2704201]
- Van Tasell DJ, Fabry DA, Thibodeau LM. Vowel identification and vowel masking patterns of hearing-impaired subjects. *Journal of the Acoustical Society of America*. 1987; 81:1586–1597. [PubMed: 3584696]
- Verhoeven K, Simonsen K, McIntyre L. Implementing false discovery rate control: increasing your power. *Oikos*. 2005; 108:643–647.

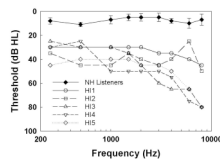


Figure 1. Air-conduction thresholds (in dB HL re: ANSI 1996) for the listeners in this study. Average NH thresholds, with standard errors of the mean, are indicated by filled symbols; individual thresholds for the HI listeners are indicated by open symbols.

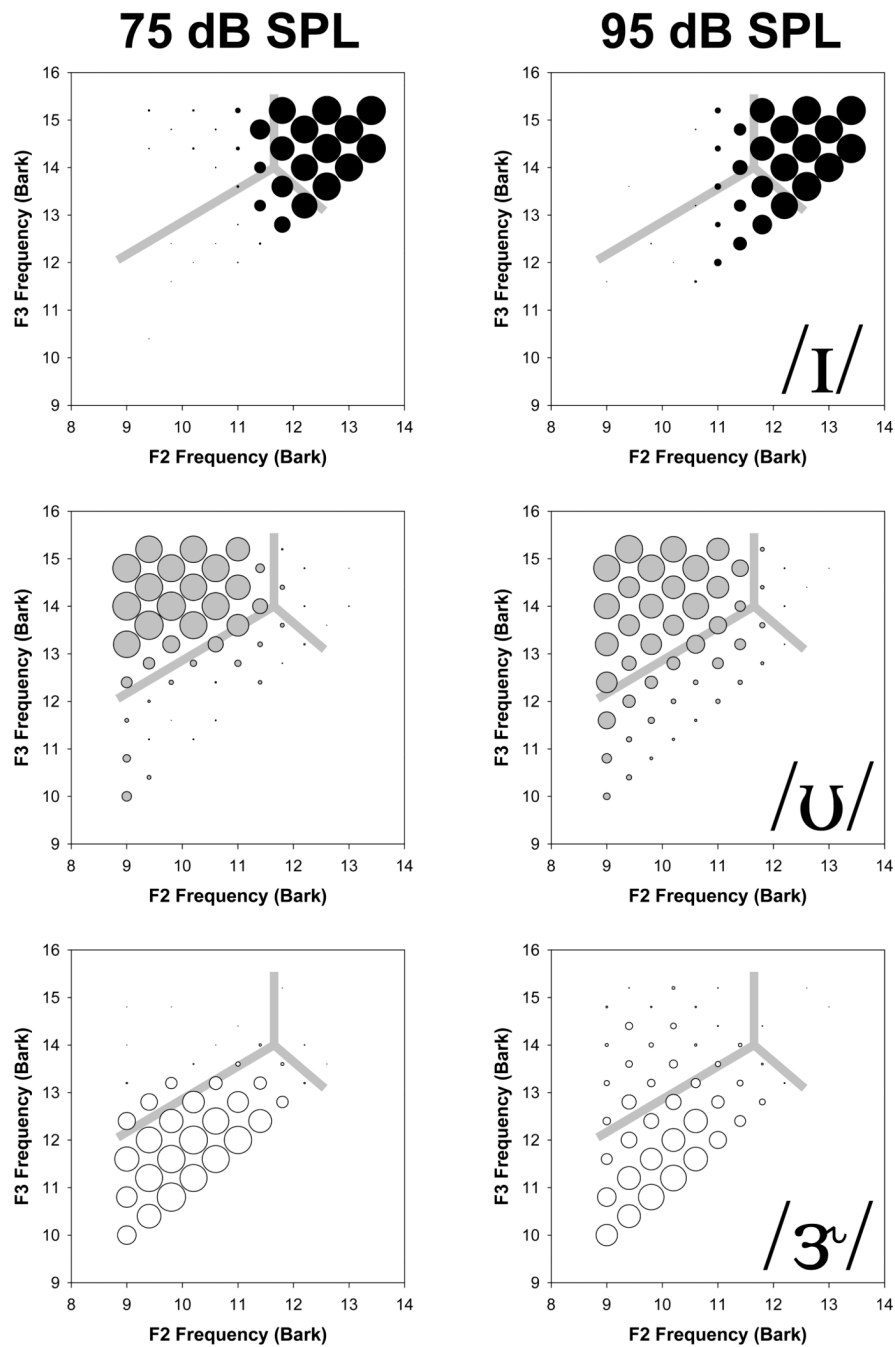


Figure 2. Bubble plots of pooled vowel identification for the five NH listeners. Bubble size corresponds to the total number of times a stimulus was identified as a member of that category. Each vowel category (top to bottom: / I /, / U /, / ɜ /) for each presentation level (left: 75 dB SPL; right: 95 dB SPL) is displayed on a separate panel. Light gray lines show average linear category boundaries calculated from data for twelve normal-hearing listeners in an earlier study (Maddox et al., 2002).

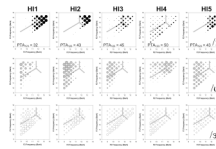


Figure 3.

Bubble plots of individual vowel identification for the five HI listeners. Each vowel category (top to bottom: / ɪ /, / ʊ /, / ə /) for each HI listener is displayed on a separate panel. Light gray lines indicate average NH category boundaries as in Fig 2. The pure tone threshold average of 1, 2, and 3 kHz (PTA_{123}) is also indicated.

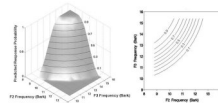


Figure 4. Fitted three-dimensional response surface for NH listeners' / σ / response (left) and corresponding two-dimensional projection onto the $F2 \times F3$ stimulus space (right). Lines on both plots indicate, in 0.1 steps, formant values for which the predicted response probability is equal.

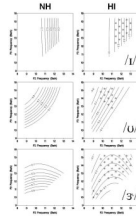


Figure 5. Predicted response probabilities of /ɪ/ (top row), /ʊ/ (middle row), and /ɜ:/ (bottom row) for NH (left) and HI (right) listeners for 95 dB SPL presentation level. Contour lines indicate equal probability of response selection in 0.1 steps. On the panels for the HI listeners, stimuli for which there was a significant difference in response probabilities between the two groups are indicated with an asterisk.

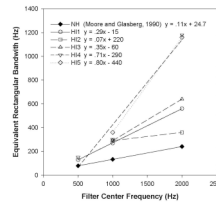


Figure 6. Auditory filter bandwidth estimates in ERB as a function of filter center frequency for individual HI listeners (open symbols) and linear fits to those estimates. Solid black symbols indicate auditory filter bandwidth estimates for normal-hearing listeners reported by Glasberg and Moore (1990).

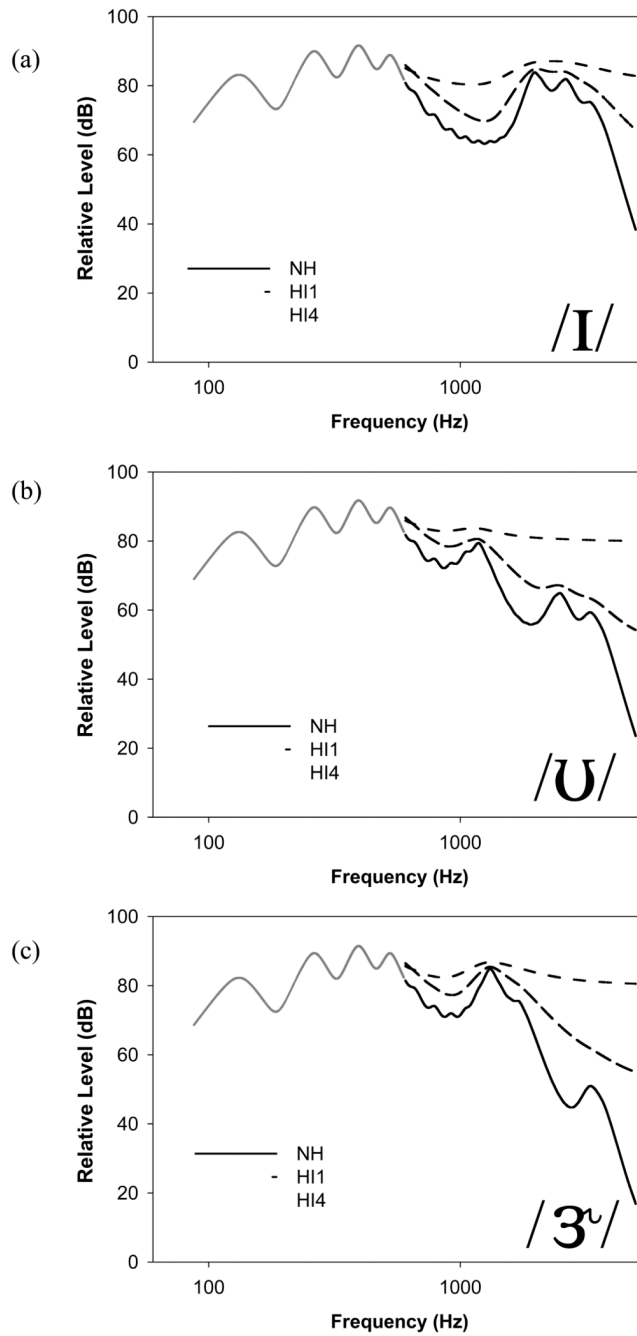


Figure 7. Excitation patterns for frequencies between 600 and 5000 Hz for the three target stimuli based on the auditory filter bandwidth estimates for HI1 (dashed black line) and HI4 (dashed-dotted black line), and for normal-hearing listeners (solid black line) using auditory filter bandwidth estimates reported by Glasberg and Moore (1990). The solid gray line, which shows the NH excitation pattern for frequencies below 600 Hz, is included as a reference.