Record: 1

Title: The role of spectral and temporal cues in vowel identification by listeners with impaired hearing.

Subject(s): DEAFNESS -- Research

Author(s): Van Summers, W.; Leek, Marjorie R.

Source: Journal of Speech & Hearing Research, Oct92, Vol. 35 Issue 5, p1189, 11p, 2 charts, 8 graphs

Abstract: Investigates whether sensorineural hearing impairment affects the relative weighting of formant frequency and vowel duration information in determining vowel identification judgments. Frequency resolution, temporal resolution, and vowel identification tasks; Relation of psychoacoustic measures with labeling performance.

AN: 9306107052

ISSN: 0022-4685

Persistent link to this record: http://search.epnet.com/direct.asp?an=9306107052&db=tth

Cut and Paste: <A href="http://search.epnet.com/direct.asp?an=9306107052&db=tth">The role of spectral and temporal cues in vowel identification by listeners with impaired hearing.</A>

Database: TOPICsearch

THE ROLE OF SPECTRAL AND TEMPORAL CUES IN VOWEL IDENTIFICATION BY LISTENERS WITH IMPAIRED HEARING

Contents

This study examined the use of duration and formant frequency in the labeling of synthetic CVC stimuli forming a beet-bit continuum. Durational and F2 frequency cues to vowel identity varied systematically across stimuli. Subjects with normal hearing tended to rely primarily on F2 frequency in vowel labeling, whereas subjects with impaired hearing relied less on F2 information. This group difference was observed even for stimuli with large F2 differences, which were easily discriminated by all subjects. The effect of vowel duration on labeling was similar for both groups, with long-duration stmuli receiving more "beet" responses than short-duration stimuli across the F2 range. Psychoacoustic measures of frequency resolution and temporal resolution were poor predictors of a subject's use of formant information and duration information in labeling.

KEY WORDS: speech perception, vowel identification, hearing impairment, frequency resolution, temporal resolution

The acoustic speech signal provides a great deal of redundant information in specifying various phonetic features of a message. For example, Lisker (1978) catalogued 18 potential acoustic cues for the p/b voicing distinction in word-medial position. For listeners with normal hearing, segmental intensities, durations, and spectral composition all provide important cues to the identity of various segments. However, the relative salience of these cues may be altered for individuals with impaired hearing because of auditory processing deficits. The present study examined the use of formant frequency and vowel duration in vowel identification by subjects with impaired hearing and subjects with normal hearing. The study focused on how the use of these cues might relate to the spectral and temporal processing capabilities of a subject.

Vowel identification is generally thought to be based on resolution of formant frequency cues, particularly F1 and F2 (Nearey, 1989; Peterson & Barney, 1952). Given that listeners with impaired hearing often show impaired frequency resolution (Tyler, 1986), they might be expected to have difficulty in identifying or discriminating vowels. However, these listeners perform well on tests of vowel perception using naturally produced utterances (Dorman, Marton, Hannley, & Lindholm, 1985; Owens, Talbott, & Schubert, 1968). One explanation for the relatively normal performance of listeners with impaired hearing is that differences in formant frequencies between vowels are large enough so that they are easily resolved, even by listeners with poorer-than-normal frequency resolution. A study by Turner and Henn (1989), however, does not support this explanation. When subjects with impaired hearing were provided with only steady-state formant frequencies and asked to make vowel identifications, vowel confusions resulted that were well predicted by the subjects' frequency-resolving capabilities.

An alternative explanation for the high levels of performance demonstrated by listeners with impaired hearing on naturally produced vowels is that these listeners make use of the redundancy of cues within the speech signal to compensate for reductions in the availability or reliability of formant frequency cues. One such redundant cue for vowels is duration. In English, vowels with similar formant frequencies tend to form tense/lax or long/short pairs that differ in inherent duration (e.g., the vowels in heed vs. hid). Perceptually, vowel duration has its greatest influence on vowel identification when formant frequency differences between vowels are fairly small, as in tense/lax pairs (Ainsworth, 1972; Bennett, 1968; Klatt, 1976). While vowel duration clearly plays a secondary role relative to formant frequency information in vowel perception by listeners with normal hearing, the relative weighting of these cues may be altered for listeners in whom spectral and/or temporal processing is impaired.

Recent work by Walden, Montgomery, Prosek, and Hawkins (1990) suggests that listeners with impaired hearing may rely more on visual speech cues than listeners with normal hearing when both visual and acoustic information are available. These authors argued that listeners with impaired hearing may develop a tendency to make greater use of the clearly perceived visual speech cues as an adaptation to their hearing impairment. The present study examined whether a similar adaptation may develop with regard to the use of spectral versus temporal acoustic cues to vowel identity.

Either or both of these cues may be inadequately processed by listeners with impaired hearing. Auditory sensitivity has been shown to be fairly directly related to frequency-resolving capability (Dubno & Schaefer, 1989; Humes, Espinoza-Varas, & Watson, 1988). However, temporal-resolving ability does not appear to be as closely related to auditory threshold (Dorman & Hannley, 1985; Humes et al., 1988). Although listeners with impaired hearing, as a group, often show poorer temporal resolution than listeners with normal hearing (Arlinger & Dryselius, 1990; Bochner, Snell, & MacKenzie, 1988; Nelson & Freyman, 1987), substantial intersubject variability exists. Not infrequently, listeners with impaired hearing perform as well as or better than listeners with normal hearing on temporal tasks (Jeateadt, Bilger, Green & Patterson, 1976; Ruhm, Mencke, Milburn, Cooper, & Rose, 1966; Tyler, Summerfield, Wood, & Fernandes, 1982).

Tyler et al. (1982) reported that the performance of listeners with impaired hearing on two temporal resolution tasks correlated significantly with their performance on a task requiring speech identification in noise. These correlations remained significant even when pure-tone auditory thresholds were partialled out. These findings suggest that temporal resolution exerts an influence on speech intelligibility independent of auditory sensitivity.

The significant correlation between temporal resolution and speech identification in noise for listeners with impaired hearing suggests that these listeners may make considerable use of temporal information in making phonetic decisions. Johnson, Whaley, and Dorman (1984) studied the processing of temporal and spectral cues to initial consonant voicing for listeners with impaired hearing and listeners with normal hearing. Their data suggest that listeners with impaired hearing may rely more on temporal, and less on spectral, information in making voicing judgments than do listeners with normal hearing.

In the present study, we tested subjects with normal hearing and subjects with sensorineural impaired hearing for frequency and temporal resolution capabilities. We then collected vowel identification data from these subjects for synthetic vowel stimuli differing in formant frequency and vowel duration. The focus of the study was whether sensorineural hearing impairment affects the relative weighting of formant frequency and vowel duration information in determining vowel identification judgments. For each listener, vowel labeling behavior was examined in light of hearing sensitivity, frequency resolution, and temporal resolution capability.

Method

Subjects

Eleven subjects with normal hearing and 11 with sensorineural hearing impairment participated in the study. All subjects were native English speakers. Listeners with impaired hearing had sensitivity thresholds of 30 dB HL (ANSI, 1969) or greater at 2000 Hz and had been aware of their impairment for at least 3 years. Five additional listeners with impaired hearing could not consistently label the clearest exemplars of each vowel category and did not complete the vowel identification portion of the study (see below). Mean audiometric thresholds for the listeners with impaired hearing and those with normal hearing who completed the study are presented in Figure 1.

Frequency Resolution Task

Frequency resolution was measured at 2000 Hz using the notched-noise method and ROEX auditory-filter fitting procedures described in Patterson and Moore (1986) and Glasberg and Moore (1990). This center frequency was selected because the formant frequency differences in the synthetic speech stimuli are in the 2000-Hz region.

Stimuli. A 2000-Hz tone with a duration of 300 msec was the signal stimulus. The signal was gated with 50-msec cosine onset and offset ramps and was temporally centered in a 400-msec burst of masking noise (notched noise). Maskers consisted of two bands of noise positioned (symmetrically or asymmetrically) about the signal with a spectral gap of a specified width separating them.

Eleven notched-noise maskers were tested. Five of these were positioned symmetrically about the 2000-Hz signal, with notch widths of 0, 800, 1600, 2400, and 3200 Hz. Six notched noises were positioned asymmetrically about the signal so that the near edge of one noise band was 400 Hz closer to the signal than the near edge of the other band. For three of these six maskers, the high-frequency noise band was positioned closer to the signal. For the other three, the low-frequency band was closer. The near edge of the noise band closer to the signal frequency was set 400,800, or 1200 Hz away from the signal. Each noise band within a notched-noise masker was 800 Hz wide, with two exceptions where the 0-Hz boundary limited the bandwidth of the low-trequency band. For the two maskers with the largest notch width between target tone and low-frequency noise, the low-frequency edge of the notch fell at 400 Hz. As a result, the low-frequency noise band for these stimuli had a bandwidth of 400 Hz (noise from 0 to 400 Hz). Notched-noise maskers were presented at a pressure spectrum level of 55 dB.

Procedure. Thresholds for the 2000-Hz tone in the presence of each notched-noise masker were determined in an adaptive tracking procedure targeting the 79% correct performance level on the psychometric function (Levitt, 1971). On each trial, two noise presentations were made, one of which also included the 2000-Hz tone. The subject pressed a button to indicate which noise presentation included the tone. Correct answer feedback was provided.

At the start of a run, the tone was set to a level approximately 10 dB above the estimated threshold. The level of the tone was decreased by 5 dB after three consecutive correct responses and increased by 5 dB after each incorrect response. This 5-dB step-size was reduced to 2 dB after four reversals in the direction of the adaptive track. The run then continued until 10 additional reversals had occurred. The mean signal level at these final 10 reversals was taken as the threshold estimate.

The threshold values for the complete set of notched noises were used to estimate the characteristics of auditory filters in the 2000-Hz region based on rounded-exponential (ROEX) filter-fitting techniques described in Patterson and Moore (1986) and recently updated in Glasberg and Moore (1990). The modeling procedure uses the threshold data to determine the best fitting ROEX filter, using parameters p_u, p_l and r. The p_u and p_l parameters are associated with the slopes of the upper and lower skirts of the filter, respectively. The r parameter is associated with the dynamic range of the filter. Following Glasberg and Moore (1990), the p_u and p_l parameters were used to estimate the equivalent rectangular bandwidth (ERB) and symmetry of the auditory filter for each subject.

Temporal Resolution Task

A temporal difference limen (TDL) task was used to assess temporal resolution. This task was chosen for two reasons. First, Tyler et al. (1982) reported a significant correlation between TDL's and performance on a speech identification task. Second, in the present study, vowel identification based on duration requires discrimination between the vowel durations of the two stimulus series. This involves the detection of differences in the duration of filled intervals, which is what the TDL task also requires.

Stimuli. Tyler et al. (1982) examined TDL's for 500 Hz and 4000 Hz narrow-band noises and reported little effect of stimulus frequency on performance. In keeping with the present focus on the 2000 Hz frequency region, third-octave narrow-band noises centered at 2000 Hz were used as stimuli.

Procedure. Temporal resolution was measured with an adaptive TDL task. On each trial, subjects were asked to identify which of two noise bursts was longer' A 30-msec stimulus served as the standard on each trial. The standard was paired with a 130-msec stimulus on the first trial. In the initial part of the run, the durational difference between standard and comparison stimuli was increased by a factor of 1.25 when the subject responded incorrectly and decreased by a factor of 0.8 after three consecutive correct responses. These step-sizes were adjusted to 1.1 and 0.91 respectively after four reversals in the direction of the adaptive track. The run continued until 10 additional reversals had occurred. The geometric mean of differences in duration between signal and comparison stimuli at these final 10 reversals was taken as the TDL estimate. Presentation level was randomly varied over a 5-dB range (90-95dB SPL) within each pair of bursts in order to reduce loudness as a cue to durational differences. For each subject, two blocks of trials were presented and the results averaged.

Vowel Identification Task

Each listener was asked to identify a series of stimuli as beet or bit. Synthetic stimuli were presented in which formant-frequency cues and vowel-duration cues to /i/ and /I/ were varied orthogonally. Stimuli varied in vowel duration and F2 frequency by amounts approximating the range of values generally seen for /i/ and /I/ in everyday speech.

Stimuli. Two 7-step series of vowel stimuli were created using the cascade software synthesizer developed by Klatt (1980). Vowels were synthesized in a "say /b/-vowel-/t/ again" context. Each /b/-vowel-/t/ stimulus consisted of an initial burst 10 msec before voicing onset, a vocalic (vowel) region, a 75-msec closure interval before release of the final /t/, and a 35-msec final burst. With the exception of F2, formant frequencies for all stimuli in each series were set to values intermediate between those for /i/ and /I/. Within each of the two series, F2 steady-state values ranged from 1640 Hz to 2300 Hz in 110-Hz steps. This range of F2 frequencies is slightly greater than the range nominally associated with the /I/-/i/ contrast (Peterson & Barney, 1952; Stevens & House, 1963). The two series of stimuli differed in total vowel duration. Stimuli from the "short" series were 70 msec in duration, and stimuli from the "long" series were 120 msec long. This 50-msec difference in vowel duration approximates the durational differences between tense and lax vowels generally present in everyday speech (Crystal & House, 1988).

Procedure. A pretest was used to familiarize the subject with the task and to assure that the subject could reliably label the stimuli based on F2 frequency, vowel duration, or both. Eight repetitions of the clearest beet exemplar (120-msec duration-2300-Hz F2 frequency) were randomized with eight repetitions of the clearest bit exemplar (70 msec-1640 Hz) and presented in a single block of trials. On each trial, subjects labeled a stimulus as beet or bit by pressing one of two buttons on a keyboard. Subjects were required to label each endpoint stimulus accurately on at least 7 of 8 presentations. Five subjects with impaired hearing who could not consistently label the endpoint stimuli were not tested further.(n1) All subjects with normal hearing met the criterion for continuation in the study.

Following the pretest, all stimuli from both series were randomized and presented for identification in a single block of trials. Each of the 14 stimuli was presented 15 times for a total of 210 trials. Stimuli were again labeled as beet or bit on each presentation. Presentation level was randomly varied over a 5-dB range (90-95dB SPL) in order to reduce potential loudness cues to vowel identity.

After completion of the full identification task, listeners were asked to demonstrate their ability to discriminate between the F2-endpoint stimuli within each durational series in a same-diflerent paradigm. This was done in order to determine whether differences in reliance on F2 frequency in labeling were directly related to differences in discrimination based on F2. Eight stimulus pairs were constructed for the discrimination task. Four pairs were formed by combining the high-F2 (2300-Hz) endpoint stimulus from each durational series with the low-F2 (1640-Hz) endpoint from the same series and manipulating order within these pairs (i.e., HIF2-LOF2 and LOF2-HIF2 pairs for each durational series). Four pairs were formed by pairing each F2-endpoint stimulus with itself. These eight stimulus pairs were randomized for presentation in a single block of trials. Ten repetitions of each pair were presented for a total of 80 trials. The "say again" context was not used in this discrimination task in order to reduce memory load effects and give a clearer indication of discrimination ability. Presentation level was randomly varied over a 5-dB range (90-95dB SPL) within each pair.

Results

Frequency Resolution

Auditory filter characteristics derived from the notched-noised data were examined as an indicator of frequency resolution for subjects with impaired hearing and subjects with normal hearing. Figure 2 shows auditory filter bandwidths (ERB) in the 2000-Hz region plotted as a function of audiometric thresholds at 2000 Hz for each subject. Bandwidths in Hz are shown on the left ordinate while relative bandwidths (BW_{Hz}/f_c) are indicated on the right ordinate to assist in comparisons with other studies. As expected, subjects with normal hearing demonstrated significantly narrower auditory filter ERBs than subjects with impaired hearing ($t(20) = -5.54$, $p < .0001$).

The symmetry of the auditory filter was estimated as the ratio of the p_u and p_l parameters as suggested by Glasberg, Moore, Patterson, and Nimmo-Smith (1984). A value of 1.0 for this ratio indicates a perfectly symmetric auditory filter. The lower this value below 1.0, the more skewed the filter, with the upper filter skirt more shallow than the lower skirt. Values greater than 1.0 indicate a filter with a steep upper skirt relative to the lower skirt. Listeners with impaired hearing generally showed more asymmetric auditory filters than listeners with normal hearing (see Figure 3). For most of the listeners with impaired hearing, the high-frequency skirt of the filter was much steeper relative to the low-frequency skirt than for listeners with normal hearing. Glasberg and Moore (1990) noted that when the ratio p_u/p_l exceeds 2.0, the slope of the steeper skirt (the high-frequency skirt here) is poorly defined by the fitting procedure. Seven of the 11 listeners with impaired hearing had p_u/p_l, values of 2.0 or greater (shown by the dotted line in Figure 3). As a result, the ERB measure of frequency resolution, which is based on the slopes of the two filter skirts, may not be reliable for these subjects. Therefore, an additional ERB measure was determined, based only on the low frequency skirt of each subject's auditory filter (ERB_L). As already reported for the more traditional ERB measure, subjects with normal hearing showed significantly narrower ERB_L values than subjects with impaired hearing ($t(20) = -5.54$, $p < .0001$).

Temporal Resolution

Most of the listeners with impaired hearing performed well on the temporal difference limen (TDL) task. TDL values are plotted against audiometric thresholds at 2000 Hz for each subject in figure 4. Most of the subjects with impaired hearing performed within the range shown by the listeners with normal hearing. The temporal discrimination of three listeners with impaired hearing was poorer than normal. Overall, TDLs did not significantly differ across groups (1(20) = - 1.58, $p > .05$).

The results of the psychoacoustic tasks suggested that this group of listeners with impaired hearing differed from listeners with normal hearing more in terms of frequency resolution than temporal resolution. The following sections present each groups' performance on the vowel perception tasks and examine how performance on the psychoacoustic tasks might relate to the use of frequency and duration in vowel labeling.

F2-Endpoint Discrimination

Recall that after participating in the vowel-labeling task, subjects were tested on their ability to discriminate between the endpoint stimuli within each series (the 1640-Hz and 2300-Hz stimuli). All subjects performed well on this discrimination task. Table 1 reports subjects' performance in terms of the overall discriminability of the stimulus set (d')(n2) and a maximum percent correct score based on d' assuming no response bias [P_{max} (C)] (Green & Swets, 1974). Subjects with normal hearing performed at or near 100% accuracy in most cases. Although subjects with impaired hearing rarely showed 100% accuracy, all subjects performed with at least 90% accuracy in terms of P_{max} (C). These levels of discrimination performance were high enough so that subjects who were heavily influenced by the F2 cue were expected to show consistent labeling of endpoint stimuli.

Vowel Identification

Mean labeling data for each group for each series of vowel stimuli are plotted in the panels of Figure 5. The data for each group are plotted in separate panels in the upper part of the figure. The lower panels replot the same data with the panels now separating the data for the long-duration series from data for the short-duration series. Probit analysis (Finney, 1971) was used to fit a normal ogive to the mean labeling data from each vowel series. The probit fits to each labeling function are plotted in the figure.

The midpoint of each probit function represents the estimated F2 steady-state frequency necessary for a stimulus from a given series to produce an equal number of beet and bit responses. The frequency difference between the midpoints of the two durational series reflects the influence of vowel duration on vowel identification: the amount of F2 change necessary to counterbalance the effect of duration and maintain performance at 50% beet responses. This measure will be referred to as DIFF50. Group differences in DIFF50 scores are best seen by comparing the separation of the probit functions plotted in the upper panels of Figure 5. The separation of the functions is similar across groups, suggesting a similar influence of vowel duration on labeling responses. This issue will be examined further in analyses described below.

The slope of each function is determined by the standard deviation of each probit fit (probit SD). Differences in slope indicate differences in the degree to which responses to the F2 frequency cues were categorical. That is, steep labeling slopes indicate that, for most of the F2 values tested, F2 frequency provided unambiguous, primary cues to vowel identity. Shallow slopes indicate that a number of F2 frequencies provided ambiguous or secondary cues. The lower panels of Figure 5 allow comparisons of labeling slopes for listeners with normal hearing versus listeners with impaired hearing for each durational series. For both series, listeners with normal hearing showed steeper labeling slopes than listeners with impaired hearing.

In addition to the probit fits to the group data, probit techniques were also used to model each individual subject's labeling data. The midpoints and standard deviations of these individual probit fits were used to compare the performance of subjects with normal hearing and subjects with impaired hearing. As will be seen, for some of the data on listeners with impaired hearing, the probit procedures provided poor fits to the actual data. Therefore, these comparisons were supplemented with additional measures.

Chi-square goodness-of-fit tests showed that in all cases the probit analysis did an acceptable job of fitting the labeling functions of each subject with normal hearing. However, these subjects showed several different response patterns. Seven of the 11 subjects with normal hearing consistently labeled low-F2 stimuli as bit and high-F2 stimuli as beet regardless of vowel duration. For these subjects the three stimuli in the middle of each series (i.e., the 1860-Hz, 1970-Hz, and 2080-Hz stimuli) were the only stimuli for which vowel duration had a clear influence on labeling (at least a 20% change in number of beet responses). One subject was influenced by the duration cue over a wider range of F2 frequencies. For this subject stimuli from the long-duration series received at least 20% more beet responses than matching stimuli from the short-duration series at five of the seven F2 frequencies tested (all but the 1640-Hz and 2300-Hz endpoints).

Three subjects with normal hearing did not hear any of the stimuli from the long-duration series as clear exemplars of bit. For these subjects, for the long-duration series, lowering F2 from 2300 Hz to 1640 Hz resulted in a fairly gradual change in responses from near 100% beet to near 50% beet. The long-vowel-duration cue appears to have had an important influence on labeling regardless of F2 frequency. The data from these subjects influenced the overall pattern for the group with normal hearing so that the labeling function for the long-duration series was slightly shallower than for the short-duration series. Further, there was an influence of duration on responses to low-F2 stimuli in addition to the mid-range F2 stimuli.

Chi-square goodness-of-fit tests showed significant departures from the probit model for 5 of the 22 labeling functions produced by the listeners with impaired hearing (11 subjects * 2 functions/subject). For 2 of these 5 functions, acceptable probit fits were obtained by limiting the fit to six of the seven data points. In the following analyses, which were based on the individual fits, the remaining 3 functions were excluded because the poor fits did not support a reliance on the fitted parameters.

The probit SD's were used to compare the groups in terms of reliance on F2 frequency in vowel labeling. Recall that probit SD's are directly related to function slopes. Data were available from all subjects for the short-duration series.(n3) For this series, probit SD's were significantly smaller (i.e., labeling slopes were steeper) for subjects with normal hearing than for subjects with impaired hearing ($t(20) = -2.24$, $p < .05$). For the long-duration series, probit SD's were also smaller for the group with normal hearing than for the 8 subjects with impaired hearing whose data were adequately modeled by the probit analysis. However, this difference was not statistically significant ($t(17) = -1.66$, $p > .05$). The three labeling functions not used in this analysis were among the shallowest in either group. Thus there may have been significant group differences in the slopes of labeling functions for the long-duration series as well if the probit fits had been adequate for all listener with impaired hearing.

Recall that the difference in frequency between the midpoints of the long-duration and short-duration probit functions (DIFF50) was taken as an indicator of the effect of vowel duration on labeling. The mean difference in F2 across series midpoints did not differ significantly between the group with normal hearing and the 8 listeners with impaired hearing for whom the probit analysis provided an adequate data fit ($t(17) = - 1.35$, $p > .10$).

A repeated measures ANOVA was conducted as a further check on the influence of F2 frequency and vowel duration on labeling. This additional analysis provided a method of comparing performance for all subjects regardless of whether the probit procedures provided an adequate fit to a given subject's data. Group membership (nominal hearing, impaired hearing) was used as a between-subjects variable in the analysis, whereas F2 frequency and vowel duration were within-subjects variables. As expected, the analysis showed significant main effects for F2 frequency and vowel duration on labeling [$F(6,120) = 92.17$ and $P(1,20) = 69.61$ respectively, $ps < .0001$]. There was a significant interaction between F2 frequency and vowel duration that reflects the greater effect of duration for stimuli with intermediate F2 values than for high-F2 or low-F2 stimuli ($F(6,120) = 4.26$, $p < .001$). The main effect of group was not significant, suggesting no clear group differences in terms of overall bias towards beet or bit responses [$F(1,20) = .001$, $p > .10$]. Of more interest was a significant interaction between F2 and group membership [$F(6,120) = 5.09$, $p < .001$]. This interaction reflects the greater effect of a change in F2 on responses for the group with normal hearing compared to the group with impaired hearing. The group-by-duration interaction was not significant, suggesting that duration had a similar effect on responses for both groups [$F(1,20) = .49$, $p > .10$]. This similarity can be seen by comparing the amount of separation of the two functions in the upper panels of Figure 5 (particularly at low and intermediate F2 frequencies up to 2080 Hz). The three-way interaction between group, F2 frequency, and vowel duration was not significant.

The results reported thus far suggest differences between listeners with normal hearing and those whose hearing is impaired in the use of F2 frequency in vowel labeling. These results pertain to labeling of each vowel series in its entirety. It was also of interest to focus on labeling of the F2-endpoint stimuli from each series in particular Recall that all subjects demonstrated high levels of discrimination between the 1640-Hz and 2300-Hz stimuli from each durational series. It was therefore of particular interest to determine whether the group differences seen for the full series were present in the labeling of these endpoint stimuli. This result would suggest that the listeners with impaired hearing made less use of F2 frequency cues in assigning a vowel label regardless of how distinct that information might be for a particular token.

A repeated measures ANOVA was carried out similar to the one previously described but including only the labeling performance for F2-endpoint stimuli. The results were nearly identical to those based on the full set of data. Once again, there were significant main effects of vowel duration and F2 frequency on labeling, whereas the main effect of group was not significant. The F2 frequency-by-group interaction was significant, reflecting a greater influence of F2 on labeling for subjects with normal hearing than those with impaired hearing. As before, the vowel duration by group interaction was not significant.

Relating Psychoacoustic Measures With Labeling Performance

On both the frequency- and temporal-resolution tasks, subjects with impaired hearing showed a range of performance, with some subjects performing within (or approaching) the normal range and some performing more poorly than subjects with normal hearing. In this section, possible relationships between subjects' performance on the psychoacoustic tasks and their reliance on frequency and duration in labeling of the vowel stimuli are examined.

Table 2 displays correlation coefficients between psychoacoustic measures and vowel-labeling measures of performance by all subjects. Recall that as a group, subjects with impaired hearing showed shallower labeling functions than subjects with normal hearing for each durational series. Given that group membership was based on audiometric threshold at 2000 Hz, it is not surprising that thresholds at 2000 Hz were significantly correlated with probit SD's (which reflect labeling slopes) for both long-and short-duration series. That is, slopes of the functions tended to become more shallow as threshold increased. Analyzing the results for the listeners with impaired hearing separately should tend to reduce the correlation coefficients over those calculated for all subjects because of the decreased range of results. Nevertheless, the correlation between threshold at 2000 Hz and probit SD was higher for the group with impaired hearing alone than for the combination of subjects.(n4) This reflects the absence of any consistent change in slopes across the 0 dB to 10 dB range of thresholds seen for subjects with normal hearing.

A straightforward explanation for these significant correlations might be that the degraded frequency resolution experienced by listeners with impaired hearing caused them to rely less on the F2 frequency in labeling. This would lead to an expectation that the psychoacoustic measures of frequency resolution (auditory filter ERB and ERB_L) might bear an even stronger relation than 2000-Hz thresholds to probit standard deviations. However, correlations between ERB and probit SD and between ERB_L and probit SD were generally smaller than the previously reported correlations between thresholds and SD's. Partialing out the effects of threshold reduced these correlations further. Thus the present measures of frequency resolution were no better than simple audiometric thresholds in predicting reliance on F2 frequency in labeling.

Temporal resolution performance did not correlate significantly with probit SD for either vowel series. These results are not surprising for two reasons. First, since F2 frequency is generally viewed as a primary cue in vowel labeling, it should be relied upon regardless of temporal processing ability (given adequate frequency resolution). Second, the similar performance of the two groups on the temporal task made it unlikely that the individual group differences in probit SD's would correlate with performance on the temporal task.

The right side of Table 2 displays correlations between the psychoacoustic measures and the DIFF50 measure of the effect of vowel duration on labeling Although the group analyses did not demonstrate significant group differences in reliance on duration for labeling, the correlational analyses did show a relationship between threshold at 2000 Hz and the DIFF50 measure of reliance on duration. This correlation was particularly strong only for the group with impaired hearing. The correlation fell just short of statistical significance ($p < .07$) for both groups combined. In each instance, increases in threshold were correlated with increased reliance on duration in labeling. As was seen for labeling slopes, there did not appear to be appreciable change in DIFF50 across the 0 dB to 10 dB threshold range of the subjects with normal hearing.

Increased reliance on the vowel duration cue may be the result of reduced access to (the primary) F2-frequency cues as a result of impaired frequency resolution. If so, it would be expected that measures of frequency resolution (ERB, ERB,) would be significantly correlated with reliance on duration (DIFF50). This expectation was confirmed in the case of ERB_L and DIFF50 in the overall group data. In general, however, correlations between the frequency resolution measures and DIFF50 were no greater than correlations between threshold at 2000 Hz and DIFF50. When the effects of threshold were partialled out, the correlations between the frequency resolution measures and DIFF50 scores were reduced even further.

It seems reasonable to expect that temporal resolution ability might bear a direct relationship with reliance on vowel duration in vowel labeling. This would mean the TDL values should correlate negatively with DIFF50 scores. These correlations did not significantly differ from zero in the present data. The limited range of variability in temporal resolution performance across subjects is probably the cause of this result. Only 3 of 11 subjects with impaired hearing showed TDL values that were not within the range of values for subjects with normal hearing. Two of these 3 subjects' data could not be used in computing the correlation between TDL and DIFF50 because of inadequate fits of subjects' labeling functions by the probit procedures.

Discussion

Listeners with normal hearing made greater use of F2 formant frequency in vowel labeling than did listeners with impaired hearing. For each durational Series, listeners with normal hearing showed steeper labeling functions across the F2 range than subjects with impaired hearing. Group differences in the labeling of F2-endpoint stimuli were of particular interest. Listeners with impaired hearing, although having little trouble resolving frequency differences between F2 endpoint stimuli, did not label these stimuli as consistently as listeners with normal hearing. Thus it appears that listeners with impaired hearing treated F2 frequency as a less reliable cue to vowel identity than did their counterparts even in cases where F2 cues were readily available.

Although all subjects were highly accurate in discriminating between F2-endpoint stimuli, mere was a small difference between groups in discrimination performance, with subjects with normal hearing performing with slightly greater accuracy (98.5% vs.94.7% mean $Pm_{max}C$). It may be that this difference in discrimination accuracy is related to the group differences in labeling. However, differences between groups in labeling the F2-endpoint stimuli were far larger than would be expected based on a 4% difference in discriminability. For subjects with normal hearing F2 variation (from 1640 to 2300 Hz) produced changes in percentage beet responses of 85% (120-msec series) and 98% (70-msec series). For subjects with impaired hearing the changes were 62% and 67% for me two durational series. Averaging across the durational series, the F2 manipulation produced approximately a 90% change in beet responses for subjects with normal hearing and only a 65% change for subjects with impaired hearing. It seems unlikely that the small (4%) differences in discrimination between groups could be the entire explanation for these much larger group differences in labeling. Rather, these data suggest an adjustment of the perceptual weight assigned to F2 frequency cues by listeners with impaired hearing. If, in everyday speech, F2 frequency cues are often ambiguous and unreliable for listeners with impaired hearing because of impaired frequency resolution, these listeners may reorganize their perceptual weighting of the cues to vowel identity and come to rely more upon other cues (see Besing, 1988; Lacroix and Harris, 1979, for similar discussions).

If listeners with impaired hearing are able to alter the perceptual weight assigned to various cues as an adjustment to their impairment, they might be expected to show nearnormal perceptual performance in "low-stress" (e.g., quiet, nonreverberant, single-talker) environments in which a multiplicity of cues is available. That is, when redundant cues are available there may be little perceptual impairment incurred by relying on cues which are thought to be secondary but which are processed well. However, these listeners might perform poorly in more stressful environments in which cues they have come to rely on are compromised. This is exactly the pattern of performance generally reported in comparisons of speech perception in quiet and in noise by listeners with impaired hearing (Dubno, Dirks, & Morgan, 1984; Plomp & Mimpen, 1979). When multiple cues are available, as in low-stress environments, many processing strategies may produce high levels of performance. Conversely, as cues are reduced in more stressful environments, acuity in processing the specific cues that remain available may become critical. This is consistent with previous research in which performance on specific psychoacoustic tasks (e.g., frequency resolution, temporal resolution) has shown a higher correlation with speech perception in noise than in quiet (Fester & Piomp, 1983; Glasberg & Moore, 1989).

Inherent vowel duration plays a secondary role in cuing vowel identity for listeners with normal hearing (Strange, 1989). In the present study it was thought that vowel duration might play an important role in cuing vowel identity for listeners with impaired hearing, particularly those listeners with good temporal resolution abilities. The data did not support this expectation. Although subjects with impaired hearing used F2 frequency less than subjects with normal hearing in vowel labeling, reliance on duration did not systematically increase as reliance on F2 decreased Cues other than F2 frequency and vowel duration [e.g., cues associated with F1 (DiBenedetto, 1989a, 1989b)] may play a more central role in vowel perception by these subjects with impaired hearing. Our stimuli neutralized all cues to vowel identity other than F2 frequency and vowel duration. Therefore, reliance on other cues could not be examined in the present data.

Because formant-frequency cues provide primary information for vowel identity, it was expected that when F2 frequencies were set to unambiguous values, F2 would outweigh duration as a vowel cue for listeners with normal hearing. This pattern was seen for most, but not all, listeners. Three of the subjects with normal hearing did not hear any members of the long-duration series as clear beet stimuli. For these subjects, stimuli from the 120-msec duration series received at least 40% beet responses even at the lowest F2 frequencies. Thus it appears that the relative contribution of F2 frequency and vowel duration in vowel labeling may vary somewhat even across listeners with normal hearing. While individual variability in the use of various segmental cues has been previously reported, it has generally been seen for cues associated with features of consonantal segments (e.g., burst and formant transition cues to place of articulation [Walley & Carrell, 1983; see Hazan & Rosen, 1991 for further examples).

A possible explanation for the reduced use of F2 frequency in vowel labeling by listeners with impaired hearing is their impaired frequency resolution. As a group, these listeners showed impaired frequency resolution and less reliance on F2 frequency in vowel labeling when compared with listeners with normal hearing. However, at the individual level, performance on the notched-noise frequency resolution task did not strongly correlate with the influence of F2 on labeling. The present measure of frequency resolution examines performance in only one frequency region at a time. Performance on this task may be too far removed from the frequency-resolving task required for vowel labeling for a clear relationship to emerge. Perhaps frequency resolution performance measured with more complex speechlike stimuli would be more indicative of an individual subject's reliance on F2-frequency in labeling. Ochs, Humes, Ohde, and Grantham (1989) advanced a similar argument to explain why their measure of frequency discrimination did not predict deficits in perception of formant transitions by listeners with impaired hearing.

Audiometric thresholds at 2000 Hz were as accurate as frequency resolution measures (ERB, ERBL) as predictors of reliance on F2 frequency in vowel labeling. In addition, ERB and ERB, were poor predictors of cue use once the portion of variance that could be accounted for by threshold was removed (partialled out). These results, however, should not be taken as evidence that frequency resolution had little effect on subjects' use of F2-frequency in labeling. Another possible interpretation is that frequency resolution had an important influence on cue use and that the influence of threshold was only by way of its correlation with frequency resolution. Both Stelmachowicz, Jesteadt, Gorga, and Mott (1985) and Glasberg and Moore (1989) have previously raised this point in the interpretation of correlations among auditory capabilities.

Conclusions

The results of this study are consistent with a reamed adjustment on the part of listeners with impaired hearing to rely less on formant-frequency cues in vowel labeling than do listeners with normal hearing. This was the case even though formant frequency differences were easily discriminable by all listeners. Presumably this adjustment is based on past experience in which formant-frequency information has not always proved reliable, possibly as a sequela of poor frequency resolution. Although most of the subjects with impaired hearing in this study demonstrated near-nominal temporal difference limens, there was no evidence of a greater than normal reliance on temporal cues to vowel identity for the group with impaired hearing. While audiometric thresholds did correlate with reliance on second-formant frequency and vowel duration in labeling, efforts to make individual predictions of cue use based on psychoacoustic measures of frequency resolution and temporal resolution were largely unsuccessful.

Acknowledgments

Support for this research was provided by NIH grant DC 00626 and by the Department of Clinical Investigation, Walter Reed Army Medical Center. All subjects provided written consent before participating in the study. 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.

(n1) Thresholds at 2000 Hz and temporal resolution performance for these 5 subjects were within the ranges shown by the subjects with impaired hearing who could reliably label the vowel endpoints. However, frequency resolution performance was poorer for the 5 subjects who failed the pretest than for any of them who passed it. As a result of this poor frequency resolution, the F2 frequency differences present in the endpoint stimuli may not have been discriminable for the subjects who failed the pretest.

(n2) Calculation of d' cannot be carried out in instances of 100% correct performance on either "same" or "different" trials. For this reason, cases of 1 100% hits and 0% false alarms were adjusted to 99.5% and

0.5% respectively.

(n3) Slope estimates for 2 subjects were based on probit fits to six of the seven date points.

(n4) For the long-duration series, although the Pearson r was greater for the group with impaired hearing taken separately, the correlation was only statistically significant when the data from the subjects with normal hearing were included. The reduced degrees of freedom for the test of the group with impaired hearing alone presumably led to this result.

TABLE 1. F2-endpoint discrimination results for subjects with normal hearing and impaired hearing.

Subjects	d′	$P_{max}(C)$
Normal hearing		
MS	5.15	99.5
CV	4.54	98.8
TR	4.54	98.8
PT	5.15	99.5
SA	4.54	98.8
HW	5.15	99.5
DK	4.54	98.8
SB	4.02	97.8
MC	4.54	98.8
CO	3.24	94.8
EM	4.54	98.8
AVG	4.54	98.5
Impaired hearing		
SS	4.54	98.8
JS	2.93	92.8
WS	4.54	98.8
LL	2.88	92.5
FE	3.92	97.5
HR	3.61	96.4
WL	3.24	94.8
TJ	3.09	93.8
CM	3.09	93.8
CS	2.93	92.8
RL	2.58	90.2
AVG	3.40	94.7

TABLE 2. Pearson product moment correlation coefficients between psychoacoustic measures and vowel-labeling measures.

		Probit standard deviation			
	Subjects	short-duration series		long-duration series	
Threshold	All	.57[*]		.48[*]	
	I-H	.67[*]		.66	
ERB	All	.49[*]	(.04)	.43	(.10)
	I-H	.30	(-.05)	.36	(.05)
ERB_L	All	.48[*]	(.01)	.49[*]	(.19)
	I-H	.26	(.22)	.45	(.02)
TDL	All	-.00	(.22)	-.22	(-.34)
	I-H	-.18	(-.16)	-.43	(-.26)

	Differences between midpoints (DIFF50)	
Threshold	.43	
	.77[*]	
ERB	.44	(.18)
	.59	(.37)
ERB_L	.52[*]	(.32)
	.69	(.38)
TDL	-.11	(-.20)
	-.19	(.17)

Note. Within each cell: upper entries represent the correlation coefficients for subjects with normal hearing and those with impaired combined; lower entries represent subjects with impaired hearing only; entries in parentheses represent correlations with absolute threshold at 2000 Hz partialled out. The psychoacoustic measures listed in the left column all reflect performance in the 2000-Hz region.

[*] $p < .05$.

GRAPH: FIGURE 1. Mean audiometric thresholds tor subjects with normal hearing and impaired hearing. Error bare indicate +/- 1 standard deviation about the mean.

GRAPH: FIGURE 2. Equivalent rectangular bandwidth of auditory filters at 2000 Hz for listerners with normal hearing and impaired hearing as a function of audiometric threshold at 2000 Hz Left ordinate shows filter bandwidth In Hz; right ordinate shows bandwidth as proportion of center frequency.

GRAPH: FIGURE 3. Auditory filter asymmetry (p_u/p_l) for listeners with normal hearing and impaired hearing as a function of audiometric threshold at 2000 Hz.

GRAPH: FIGURE 4. Temporal difference limen as a function of threshold at 2000 Hz for listeners with normal hearing and impaired hearing.

GRAPHS: FIGURE 5. Mean percent beet responses to each vowel stimulus tor listeners with normal hearing and Impaired hearing. Stimulus F2 frequency is plotted on the abscissa. Solid lines represent best fitting normal ogives to data for each series based on probit analysis. The upper panels separate the data by group (normal hearing-upper left, Impaired hearing-upper right). The lower pane-is replot the data, separating results for the 120 msec vowel series (lower left) from results for the 70-msec series (lower right).

References

Ainsworth, W. A. (1972). Duration as a cue in the recognition of synthetic vowels. Journal of the Acoustical Society of America, 51, 648-651.

American National Standards Institute. (1970). Specifications for audiometers (ANSI S3.6-1969). New York: ANSI.

Alinger, S., & Dryselius, H. (1990). Speech recognition in noise, temporal and spectral resolution in normal and impaired hearing. Acta Otolaryngol, 46, 30-37.

Bennett, D.C. (1968). Spectral form and duration cues in the recognition of English and German vowels. Language and Speech, 11, 65-85.

Besing, J. (1988). Aspects of vowel perception by normal hearing and hearing-impaired listeners. Doctoral dissertation, Louisiana State University, Baton Rouge, LA.

Bochner, J. H., Snell, K. B., & MacKenzie, D. J. (1988). Duration discrimination of speech and tonal complex stimuli by normally hearing and hearing-impaired listeners. Journal of the Acoustical Society of America, 84, 493-500.

Crystal, T. H., & House, A. S. (1988). The duration of American-English vowels: an overview. Journal of Phonetics, 16, 263-284.

DiBenedetto, M. (1 (1989a). Vowel representation: Some observations on temporal and spectral properties of the first formant frequency. Journal of the Acoustical Society of America, 86, 55-66.

DiBenedetto, tel. (1989b). Frequency and time variations of the first formant: Properties relevant to the perception of vowel height. Journal of the Acoustical Society of America, 86, 67-77.

Dorman, M. F., & Hannley, M. (1985). Idenfitication of speech and speechlike signals by hearing impaired listeners. In R. Daniloff (Ed.), Speech science: Recent advances (pp. 111-153). Boston: College Hill.

Dorman, M. F., Marton, K., Hannley, tel. T., & Lindholm, J. M. (1985). Phonetic identification by elderly normal and hearing-impaired listeners. Journal of the Acoustical Society of America, 77, 664 670.

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on recognition of speech in noise. Journal of the Acoustical Society of America, 76, 87-96.

Dubno, J. R., & Schaefer, A. e. (1989). Frequency resolution for broadband-noise masked normal listeners. Journal of the Acoustical Society of America, 86, 5122.

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

Finney, D.J. (1971). Probit Analysis. Cambridge, U.K.: Cambridge University Press.

Glasberg, e. R., & Moore, e. c. J. (1989). Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. Scandinavian Audiology, 32, 1-25.

Glasberg, B. R., & Moore, e. c. J. (1990). Derivation of auditory filter shapes from notched-noise data. Hearing Research, 47, 103-138.

Glasberg, B. R., Moore, B.C.J., Patterson, R. D., & Nimmo-Smith, I. (1984). Dynamic range and asymmetry of the auditory filter. Journal of the Acoustical Society of America, 76, 419-427.

Green, D. M., & Swets, J. A. (1974). Signal detection theory and psychophysics. New York: Krieger.

Hazan V., & Rosen, S. (1991). Individual variability in the perception of cues to place contrasts in Initial stops. Perception and Psychophysics, 49, 187-200.

Humes, L. E., Espisnoza-Varas, B., & Watson, C. S. (1988). Modeling sensorineural hearing loss. I. Model and retrospective evaluation. Journal of the Acoustical Society of America, 83, 188-202.

Jesteadt, W., Bilger, R. C., Green, D. M., & Patterson, J. H. (1976). Temporal acuity in listeners with sensorineural hearing loss. Journal of Speech and He&ring Research, 19, 357-370.

Johnson D., Whaley, P., & Dorman, M. (1984). Processing of cues for stop consonant voicing by young hearing-impaired listeners. Journal of Speech and Hearing Research, 27, 112-118.

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America, 59, 1208-1221.

Klatt, D., (1980). Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America, 67, 971-995.

Lacroix, P. G., & Harris, J. D. (1979). Effects of high-frequency cue reduction on the comprehension of distorted speech. Journal of Speech and Hearing Disorders, 44, 236-246.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. Journal of the Acoustical Society of America 49, 467-477.

Lisker, L. (1978). Rapid vs. rabid: A catalogue of acoustic features that may cue the distinction. Haskins Laboratories Status Report on Speech Research, SR-54, 127-132.

Nearey, T. M. (1989). Static, dynamic, and relational properties of vowel perception. Journal of the Acoustical Society of America, 85, 2088-2113.

Nelson, D. A., & Freyman, R. L. (1987). Temporal resolution in sensorineural hearing-impaired listeners. Journal of the Acoustical Society of America, 81, 709-720.

Ochs, M.T., Humes, L. E., Ohde, R. N., & Grantham, D. W. (1989). Frequency discrimination and stop consonant identification in normally hearing and hearing-impaired subjects. Journal of Speech and Hearing Research, 32, 133-142.

Owen, E., Talbott, C. a., & Schubert, E. D. (1968). Vowel discrimination of hearing-impaired listeners. Journal of Speech and Hearing Research, 11, 648-655.

Patterson, R. D., & Moore, e. c. J. (1986). Auditory filters and excitation patterns as representations of frequency resolution. In B. C. J. Moore (Ed.) Frequency selectivity in hearing (pp. 123177). London: Academic.

Peterson, G. E., & Barney, H. L. (1952). Control methods in the study of vowels. Journal of the Acoustical Society of America, 24, 175-184.

Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. Journal of the Acoustical Society of America, 66, 1333-1342.

Ruhm, H. B., Mencke, E. O., Milburn, R., Cooper, W. A., & Rose, D. (1966). Differential sensitivity to duration of acoustic signals. Journal of Speech and Hearing Research, 9, 371-384.

Stelmachowicz, P. G., Jesteadt, W., Gorga, M. P., & blots, J. (1985). Speech perception ability and psychophysical tuning curves in hearing-impaired listeners. Journal of the Acoustical Society of America, 77, 620 627.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An accoustical study. Journal of Speech and Hearing Research, 6, 111-128.

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentences. Journal of the Acoustical Society of America, 85, 2135-2153.

Turner, C. W., & Henn, C. C. (1989). The relation between vowel recognition and measures of frequency resolution. Journal of Speech and Hearing Research, 32, 49 58.

Tyler, R. S. (1986). Frequency resolution in hearing-impaired listeners. In B. C. J. Moore (Ed.) Frequency selectivity in hearing (pp. 309-371). London: Academic.

Tyler, R. S., Summerfield, A. Q., Wood, E., & Fernandes, M. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. Journal of the Acoustical Society of America, 72, 740-752.

Walden, B. E., Montgomery, A. A., Prosek, R. A., & Hawkins, D. B. (1990). Visual biasing of normal and impaired auditory speech perception. Journal of Speech and Hearing Research, 33, 163 173.

Walley, A. C., & Carrell, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. Journal of the Acoustical Society of America, 73, 1011-1022.

Received December 15, 1991

Accepted March 17, 1992

Contact author: W. Van Summers, PhD, Army Audiology and Speech Center, Walter Reed Army Medical

Center, Washington, DC 20307-5001.

~~~~~~~~

By W. Van Summers and Marjorie R. Leek Walter Reed Army Medical Center Washington, DC

Source: Journal of Speech & Hearing Research, Oct92, Vol. 35 Issue 5, p1189, 11p, 2 charts, 8 graphs.
Item Number: 9306107052

Top of Page

Back
~~~~~~~~