Phonetic identification by elderly normal and hearing-impaired listeners

M. F. Dorman, K. Marton, and M. T. Hannley

Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287

J. M. Lindholm

Department of Psychology, Arizona State University, Tempe, Arizona 85287

(Received 8 December 1983; accepted for publication 1 October 1984)

Young normal-hearing listeners, elderly normal-hearing listeners, and elderly hearing-impaired listeners were tested on a variety of phonetic identification tasks. Where identity was cued by stimulus duration, the elderly hearing-impaired listeners evidenced normal identification functions. On a task in which there were multiple cues to vowel identity, performance was also normal. On a /b d g/ identification task in which the starting frequency of the second formant was varied, performance was abnormal for both the elderly hearing-impaired listeners and the elderly normal-hearing listeners. We conclude that errors in phonetic identification among elderly hearing-impaired listeners with mild to moderate, sloping hearing impairment do not stem from abnormalities in processing stimulus duration. The results with the /b d g/ continuum suggest that one factor underlying errors may be an inability to base identification on dynamic spectral information when relatively static information, which is normally characteristic of a phonetic segment, is unavailable.

PACS numbers: 43.71.Ky, 43.66.Sr, 43.70.Hs

INTRODUCTION

In order to extract phonetic information from the stream of speech in a normal manner, listeners must be sensitive to the duration of acoustic segments in that stream. For example, the relative durations of silent intervals may signal: (1) the distinction between fricative and affricate; (2) the presence or absence of a stop consonant in a consonant cluster; (3) the voicing of a stop consonant in word-medial position; and (4) the distinction between single and double stop consonants (see Dorman et al., 1979; Lisker, 1957; Pickett and Decker, 1960; Repp et al., 1978). The duration of fricative noise also can signal the distinction between fricative and affricate (Gerstman, 1957; Dorman et al., 1980). The duration of formant transitions can signal the distinction between stop consonant and semivowel (Liberman et al., 1954; Schwab et al., 1981). It is clear from this partial listing that sensitivity to acoustic segment duration is important for normal phonetic identification. In adverse listening situations, segment duration may be an especially important cue to phonetic identity (Wardrip–Fruin, 1982).

Phonetic identification often is impaired in listeners with sensorineural hearing loss and may become more so with advancing age (see Hayes, 1980). In order to understand the causes of poor identification, previous investigators have assessed, among other aspects of auditory function, the ability of hearing-impaired listeners to make temporal judgments.

Listeners with sensorineural hearing impairment have been found, generally, to have poorer-than-normal acuity for temporal events. For example, Tyler et al. (1982) found that this group of listeners had larger average temporal difference limens than normal listeners (42 vs 18 ms), larger gap difference limens (78 vs 51 ms), and larger gap detection thresholds (12 vs 7 ms). Similar results have been reported by Fitzgibbons and Wightman (1982), and Irwin et al. (1981).

In the several experiments on temporal acuity, the differences between hearing-impaired and normal listeners were found to be relatively small when compared to the magnitude of differences in acoustic segment duration which signal phonetic contrasts. For example, the elevation in gap difference limen as reported by Tyler et al. (1982) was 27 ms. In speech, when a silent interval signals the presence of a stop consonant in a cluster (e.g. “split”), that interval is on the order of 80 ms or longer. The absence of a stop is signaled by intervals of less than 20 ms. This comparison suggests that, although hearing-impaired listeners have poorer than normal temporal acuity, the magnitude of the disability is such that phonetic identification may not be affected.

The available evidence is contradictory on this point. Tyler et al. (1982) found no difference between normal and hearing-impaired adults in the identification of stimuli from a voiced–voiceless consonant continuum when the stimuli differed in the temporal feature of voice onset time (VOT). Parady et al. (1981), and Johnson et al. (1984) have reported a similar outcome for children with mild and moderate hearing impairment. On the other hand, Godfrey and Millay (1978) found that about half of their sample of hearing-impaired listeners were unable to identify stimuli from a /be/ to /we/ continuum appropriately when the stimuli varied in the duration of formant transitions. Abnormalities in identification have also been reported for stimuli from continua in which vowel duration and fricative noise duration were manipulated (Ginzel et al., 1982). In these studies elderly hearing-impaired listeners required a slightly longer vowel duration to identify the long member of a short–long vowel
contrast and a slightly longer fricative noise duration to identify /s/ in a /t/-/s/ contrast.

The differences in outcome of these studies may be due simply to the different stimuli used. Identification on the basis of relative voicing onset (in the voiced–voiceless identification task) is quite different psychophysically from identification based on the duration of rapid formant transitions (in the stop–semivowel identification task). Thus a general statement on the ability of hearing-impaired listeners to process differences in the duration of acoustic events relevant to phonetic perception can be made only after sampling from a range of tasks based on different stimulus characteristics.

Whatever tasks we choose, it is important to know more than whether or not hearing-impaired listeners can identify a given naturally produced phone. As we noted above, natural tokens of contrastive phones may differ greatly in acoustic signal duration. In “slit” versus “split,” for example, the differences would not be expected to strain the resolving power of even a damaged auditory system. However, if we create a series of signals which vary in the duration of the relevant acoustic segment from one extreme to another in small steps, then by assessing the slope of the identification function and where the phonetic boundary (i.e., the point of 50% identification accuracy) lies along the function, we can assess in greater detail the damaged auditory system’s temporal processing ability. If the slope of the identification function is abnormally shallow, then we may infer that the listeners have reduced sensitivity to the dimension under consideration. If the slope is normal but the boundary is shifted, then we may infer an abnormal criterion for that dimension, or an abnormal weighting of other acoustic cues.

These considerations led us to create three continua in which certain temporal parameters of speech signals were varied. The stimuli from these continua were used in identification experiments with normal-hearing listeners and with elderly hearing-impaired listeners. In one continuum, the duration of a fricative noise was varied over the range of 60–140 ms to create a “chop” to “shop” continuum. In another, a silent interval was varied over the range of 120–20 ms to create a “slit” to “split” continuum. In yet another, the duration of formant transitions and correlated amplitude envelope rise time was varied over the range 40–90 ms to create a /ba/ to /we/ continuum.

In order to contrast performance on these tasks with performance on tasks in which acoustic cues other than temporal ones were crucial for phonetic identification, we created two other test sequences. In one, a two-formant /ba/-/da/-/ga/ continuum, the relevant acoustic cue underlying identification was the initial frequency (and, hence, direction of change) of a brief (40 ms) second-formant transition. In the other, a vowel identification task in /bVt/ format, the cues were the loci of the formant frequencies in the region of the syllabic nuclei, and presumably the formant transitions into and out of the nuclei (see Verbrugge et al., 1976).

I. METHODS
A. Subjects

Three groups of listeners were used in each of the five listening tasks—a group of young (less than 40 years old) listeners with pure-tone thresholds lower than 20 dB HL at 0.5, 1, 2, and 4 kHz; a group of elderly (over 60 years) listeners with pure-tone thresholds at 0.5, 1, and 2 kHz lower than 20 dB HL and a 4-kHz threshold less than 30 dB HL; and a group of elderly listeners with sensorineural hearing impairment. The listeners in the latter group reported an onset of hearing loss after 50 years of age. Since subjects with histories of extensive noise exposure were eliminated from the sample, we presume the losses to be age related. The number of subjects in each group varied among the tasks (minimum = 10, maximum = 43; see Sec. II for exact numbers). A cadre of ten young, normal-hearing listeners, ten elderly, normal-hearing listeners, and ten elderly, hearing-impaired listeners participated in each of the tasks. The pure-tone thresholds of the listeners within a group did not differ as a function of the task (p > 0.05). Thus the mean audiograms of the listeners who participated in the “shop”–“chop” task, shown in Fig. 1, are representative of the audiograms of the listeners in each of the other tasks.

B. Stimuli

1. “Shop–chop” continuum

To create a continuum in which the duration of a steady-state signal varied, the word “shop,” spoken by a male, was first sampled at 20 kHz and then stored in digital form in computer memory. Using a waveform editor, the “sh” noise was removed from the signal and stored independently of the vocalic portion. The rise time of the noise was edited to 50 ms by trimming the onset. The fall time was edited to 10 ms. The peak energy in the fricative noise was at 2240 Hz, as determined by LPC analysis. “Center cuts” were then made in the noise, as close as possible to zero crossings, so that when the initial and final portions of the noise were rejoined, nine stimuli were created with durations of 60–140 ms in 10-ms intervals. By removing portions from the middle of the noise segment, we left its onset and offset unchanged. The nine fricative noises were then rejoined with the vocalic portion of the word to form a set of stimuli in which only the duration of the fricative noise varied. The stimuli with brief fricative noise were intended to be heard as “chop” while those with longer noise were intended to be heard as “shop.” Ten tokens of each stimulus were randomized with a 4-s interstimulus interval to form a test sequence. A familiarization sequence of five tokens of “shop” (140-ms noise duration) and five tokens of “chop” (60-ms noise duration) preceded the test sequence.

2. “Ba–we” continuum

To create the starting point for a continuum in which the duration of a nonsteady-state signal varied, the syllable /ba/ was synthesized with two formants using the parallel-resonance synthesizer at Haskins Laboratories. The first formant began at 304 Hz and reached a steady-state value of 756 Hz over 40 ms. The second formant began at 1193 Hz at time +10 ms and rose to the steady-state value of 1527 over 40 ms. All transitions were linear. F 1 was synthesized at +3 dB relative to F 2. F 1 bandwidth was 60 Hz; F 2 bandwidth was 120 Hz. All formants rose linearly at a rate of 2240 Hz/ms. By removing portions from the middle of the noise segment, we left its onset and offset unchanged. The nine fricative noises were then rejoined with the vocalic portion of the word to form a set of stimuli in which only the duration of the fricative noise varied. The stimuli with brief fricative noise were intended to be heard as “chop” while those with longer noise were intended to be heard as “shop.” Ten tokens of each stimulus were randomized with a 4-s interstimulus interval to form a test sequence. A familiarization sequence of five tokens of “shop” (140-ms noise duration) and five tokens of “chop” (60-ms noise duration) preceded the test sequence.
To create the stimuli for the vowel identification task, the words "beet, bit, bet, bait, bat, bought, boat, but, Bert, boot," spoken by a male, were sampled at 10 kHz and stored in digital form in computer memory. The stimuli varied along several parameters—the shape of the formant transitions of /b/, the formant frequencies at the syllabic nucleus, the shape of the formant transitions for /t/, and the overall signal duration. Ten tokens of each stimulus were randomized into a test sequence. Three repetitions of the sequence described above, i.e., "beet . . . boot," constituted the practice sequence.

5. "Ba-da-ga" continuum

To create the starting point for a continuum in which only the initial frequency and direction of change of a brief formant transition were varied, the syllable /ba/ was synthesized with two formants. The first formant began at 100 Hz and reached a steady state at 765 Hz. The second formant began at 619 Hz and rose to a steady state at 1230 Hz. Eleven more stimuli were then created by varying the starting point of the second formant transition from 759–2298 Hz in approximately equal steps. For all stimuli the first formant was synthesized at +3 dB relative to the second formant; the transition duration of both the first and second formants was 40 ms; all transitions were linear; F1 bandwidth was 60 Hz, F2 bandwidth 90 Hz; and the fundamental frequency fell from 114–86 Hz over the 200-ms syllable duration.

C. Procedure

All listeners were first given a brief audiologic examination, consisting of pure-tone audiometry and immittance measures with contralaterally and ipsilaterally elicited acoustic reflexes. Performance on the phonetic identification tasks was then assessed, generally over a 2-day span, with 1.5 h of testing each day. The speech stimuli were reproduced via a Crown 800 series tape deck, Crown D–50 amplifier, Hewlett-Packard attenuators, and Telex 1470 earphones. The subjects were seated in booths in a six-position listening station located in a sound-treated room. For all listeners in all conditions the signals were presented at 90 dB SPL. The presentation level was calibrated to peak vowel amplitude. In each of the experimental conditions the listeners were told the response options—e.g., "shop" or "chop"—and were told to write their responses on a response sheet.

II. RESULTS

A. Contrasts cued by acoustic segment duration

The average identification functions of the three groups of listeners for the "chop"—"shop" continuum are shown in Fig. 2. One-way analyses of variance indicated that neither the slopes [F(2,61) = 0.77, p > 0.05] nor the location of the phonetic boundaries [young normal (n = 10), mean = 96 ms; elderly normal (n = 22), mean = 95 ms; elderly hearing-impaired (n = 32), mean = 92 ms; F(2,61) = 0.93, p > 0.05] differed among the groups.

The identification functions of the three groups for the /ba/-/we/ continuum are shown in Fig. 3. The slopes of the functions did not differ [F(2,46) = 1.51, p > 0.05], nor did the phonetic boundaries [young normal (n = 10), mean = 77 ms; elderly normal (n = 19), mean = 75 ms; elderly hearing-impaired (n = 20), mean = 74 ms; F(2,46) = 0.43, p > 0.05].
The identification functions of the three groups for the "slit" to "split" continuum are shown in Fig. 4. The slopes of the functions did not differ \( F(2,81) = 0.40, p > 0.05 \). The phonetic boundaries, however, differed significantly [young normal \((n = 18)\), mean = 77 ms; elderly normal \((n = 23)\), mean = 72 ms; elderly hearing-impaired \((n = 43)\), mean = 65 ms; \( F(2,81) = 5.6, p < 0.02 \)]. Post hoc Scheffe tests revealed that only the young normal group differed significantly from the elderly hearing-impaired group. The boundaries for the young normal listeners were distributed evenly over the range 65–95 ms with but a single outlier at 51 ms. In contrast, for the elderly hearing-impaired group 12 of the 43 boundaries fell between 51 and 53 ms. To assess the possibility that the differences in boundary location among the hearing-impaired listeners were due to differences in auditory sensitivity either in the frequency domain of the fricative noise or in the domain of the \( F_1/F_2 \) onset frequencies, we examined the audiograms of the 12 subjects with the shortest boundary locations and those of 12 subjects with the longest locations. The mean audiograms of the two groups did not differ significantly at any of the five frequencies tested.

C. Stop consonant identification

The average identification functions for the stimuli from the /bdg/ continuum are shown in Fig. 5. The young normal listeners identified stimuli with a rising second formant transition as "ba" and stimuli with transitions falling more than 305 Hz as "ga." A /da/ category centered on the stimulus with a 152-Hz fall in \( F_2 \) was also present. Although asymptotic performance within this category was relatively low, the category was well defined in the sense that less than 10% "da" responses were elicited by stimuli at the ends of the continuum.

The performance of the elderly normal-hearing \((n = 22)\) and elderly hearing-impaired listeners \((n = 21)\) differed from that of the young normal listeners \((n = 10)\) in several respects. Six elderly normal-hearing listeners and five elderly hearing-impaired listeners did not have a well-defined /ba/ category in the sense that they made "ba" responses to stimuli at the /ga/ end of the continuum. The performance of the remaining listeners in response to stimuli from the /ba/ category was analyzed for location of the phonetic boundary and the slope of the identification function. Marginally significant main effects were found for the measures of phoneme boundary \([F(2,39) = 3.10, p = 0.06]\) and slope \([F(2,39) = 3.07, p = 0.06]\). Post hoc comparisons indicated that the boundary of the young normal listeners differed from that of the elderly hearing-impaired listeners. The
slope of the identification function differed between the two groups of elderly listeners.

An analysis of responses to the stimuli from the /gə/ category in terms of slope of the identification function and phoneme boundary was inappropriate due to the shallow slopes of the identification functions. However, it is clear from inspection of Fig. 5 that the elderly listeners made “da” responses to stimuli from the /gə/ category. An analysis of variance on “gə” responses to the four stimuli with the greatest fall in F2 indicated a main effect for stimulus number \( F(3,150) = 5.41, p = 0.001 \), a main effect for groups \( F(2,50) = 19.81, p < 0.001 \), and a significant linear interaction component \( F(2,50) = 19.19, p = 0.004 \).

Subsequent tests indicated that both of the elderly groups differed from the young normal group. The elderly hearing-impaired group, however, did not differ from the elderly normal group. The performances of the subjects in the hearing-impaired group were similar, i.e., only one of the 20 listeners was able to identify /gə/ well. In contrast, half of the elderly normal listeners identified the most extreme stimulus in the series as /gə/ at least 75% of the time. Thus, two subgroups of elderly normal-hearing listeners could be identified. One group performed with as low identification accuracy as the elderly hearing-impaired listeners, while another group performed in a manner similar, but not identical, to young, normal-hearing listeners.

D. Frequency selectivity and identification of /gə/

Frequency selectivity is the ability to detect one signal in the presence of another. Hearing impaired listeners show reduced frequency selectivity (e.g., Pick et al., 1977). As a consequence, spectral peaks in the speech signal should be less effectively resolved by hearing-impaired listeners than by normal-hearing listeners and differences among hearing-impaired listeners in frequency selectivity may be related to differences in accuracy of speech identification. Since in the present experiment there were wide differences in the ability of the elderly listeners to identify /gə/, we determined the correlation between frequency selectivity at 2000 Hz and percent /gə/ responses.\(^1\) The 2000-Hz probe was chosen because it falls near the middle of the region of F2 starting frequencies for the /gə/ category. The listeners were a subset of the elderly normal-hearing (n = 11) and elderly hearing-impaired (n = 12) listeners. Because frequency selectivity is associated with hearing loss, we also determined the correlation between /gə/ identification and threshold in quiet at 2000 Hz and the slope of the masking pattern for a 2000-Hz probe. The analyses indicated a small (\(-0.457\)) but significant (\(p < 0.05\)) correlation between threshold at 2000 Hz and percent /gə/ identification, a small (0.384) nonsignificant correlation between the slope of the masking pattern and percent /gə/ identification, and a small (\(-0.404\)) nonsignificant correlation between threshold at 2000 Hz and the slope of the masking pattern.

III. DISCUSSION

Measures of identification accuracy have been obtained for stimuli in which stimulus duration or the location of formant frequencies varied. The results indicate that elderly hearing-impaired listeners are not impaired in their ability to identify signals on the basis of temporal features or of the location of quasi steady-state formants. However, elderly hearing-impaired listeners and some elderly normal-hearing listeners are greatly impaired in their ability to identify signals which vary in the onset frequency and direction of change of brief formant transitions. The implications of these results are discussed below.

A. Phonetic contrasts cued by acoustic segment duration

In the frequency domain of the formant transitions which cued the /bæ/–/wæ/ contrast, the elderly hearing-impaired listeners showed only a mild hearing impairment. Perhaps it is not surprising that normal identification func-
ations were obtained. Moreover, the contrast was not cued by the duration of formant transitions alone. Rather, the rise time of the amplitude envelope covaried with the duration of the transitions. Thus two kinds of cues to phonetic identity—transition duration and amplitude rise time—were available to listeners.

A better index of hearing-impaired listeners’ ability to make phonetic judgments based on a single temporal dimension is provided by the results of the “chop” versus “shop” identification task, since only one stimulus dimension, fricative noise duration, varied. These results also give us a better view of the effects of hearing loss on phonetic judgments and in the present study, the poor performance of elderly hearing-impaired listeners was not due to exceptionally poor frequency selectivity. The poor identification accuracy shown by the elderly normal-hearing listeners in the present study indicates that factors other than related to sensitivity loss play a role in determining the identification accuracy of elderly hearing-impaired listeners. We can describe one factor as an inability to base phonetic identification on dynamic spectral cues when other cues, which normally characterize a phonetic segment, are unavailable for use. This description follows from the analysis presented below.

The relatively better identification accuracy for /b/ (than for /d/ and /g/) shown by the elderly listeners may be the consequence of two-formant synthesis providing more of the distinctive acoustic attributes of labial than of alveolar or velar stops. In natural speech /b/ is characterized by rising formant transitions. Two-formant synthesis with a rising F2 provides this attribute. In contrast, two-formant synthesis provides neither the characteristic spectral peak at 3–4 kHz and “uptilted” spectrum of /d/ (Blumstein and Stevens, 1979) nor the relatively long midfrequency peak and spreading formant transitions of /g/ (Kewley-Port, 1983; Stevens, 1975). There is little converging evidence on which features of /d/ are necessary for recognition by hearing-impaired listeners. For /g/, however, converging evidence exists. Turek et al. (1980) have reported poor identification of /g/ synthesized with a spreading formant pattern alone. It follows that the information lacking in the synthesis was the relatively static feature of a long duration midfrequency peak provided by the presence of a release burst. If so, then elderly hearing-impaired listeners as well as some elderly normal-hearing listeners rely on this feature to a greater extent than young normal-hearing listeners.

**ACKNOWLEDGMENTS**

This work was supported by NIH grant NS 15457. We thank Mary Clear Ormsen for assistance in data collection.

---

1 Frequency selectivity was determined using a masking technique (Jerger et al., 1960). The probe stimulus was a 200-Hz tone pulsed twice a second on a 50% duty cycle, having a total on-time of 250 ms, including 25-ms rise and decay times. The maskers were narrow bands of noise having center frequencies of 0.5, 1.0, 1.5, 2.0, and 4.0 kHz. The noise had a uniform band width of 300 Hz measured at the half-power point and was attenuated by 48 dB per octave on each side of the center frequency. The subject’s unmasked threshold for the probe was first determined using an adaptive technique with a 70% criterion. Three 250-ms tone bursts were presented at each 2-dB step; three reversals were used to bracket each threshold. Continuous narrow-band noise at 0.5 kHz was then introduced ipsilaterally at 90 dB SPL. The threshold for the probe was determined with noise and tone presented simultaneously using the same adaptive technique. The procedure was repeated for each of the other noise bands.


