Reprinted from THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, Vol. 48, No. 4, (Part 2), 993–998, October 1970

Received 6 October 1969

9.5, 9.7

Effects of Filtering and Vowel Environment on Consonant Perception

THOMAS GAY

Haskins Laboratories, New Haven, Connecticut 06510

The purpose of this experiment was to determine the effects of filtering and vowel environment on consonant perception. Sixteen consonants in consonant–vowel (CV) combination with seven vowels were recorded on tape, low-pass filtered, and played back to a group of listeners. In general, the results indicated that /t, k, b, d, g, s, f, z, w, r, n/ are affected by filter cutoff points, that /k, g, f, v, m/ show multivowel effects, and that /p, b, d, j, n/ show consistently lower scores only when followed by /i/. As expected, error types were predominantly "place" with "manner," voicing," and "nasality" errors occurring only at the less favorable cutoff frequencies. The results were discussed in terms of the predictability of the effects as a function of CV-transition characteristics and the suitability of small sample phonetically balanced word lists for assessing speech discrimination of individulas with high-frequency hearing loss.

INTRODUCTION

Since their introduction in the 1940's, phonetically balanced (PB) word lists have been used extensively for testing speech discrimination in both the clinic and research laboratory. The list's main features are that the words are common, familiar, easy to administer, and of course, are in "phonetic balance." Originally, the aim of phonetic balancing was to provide a list of words whose phonemic content occurred with the same frequency of occurrence as the phonemes found in everyday speech. This was accomplished simply by assigning a certain over-all proportion to each phoneme in the list, without regard for the internal phonemic makeup of the words. It has since been recognized, however, that coupling effects exist for different consonant and vowel sequences, with the articulatory and acoustic properties of a given phoneme often depending on those of its neighbor. In this sense then, it is not unreasonable to suspect that conditions may exist where the perception of a given sound might be either enhanced or degraded by the coarticulation effects of the adjacent phoneme. The most likely condition, of course, would be one in which the spectral characteristics of the phoneme are either altered or eliminated, as in filtering or, on a physiological level, a hearing impairment. In both cases, important cue information provided by the consonant–vowel (CV) transition might be reduced by varying degrees, depending on the amount of the transition eliminated by the distortion.

The experiment reported here attempts to describe some of these effects, specifically, the extent to which various vowel environments influence the identification of consonants in CV syllables, heard under conditions of low-pass filtering. Although coarticulation effects in real speech extend beyond simple CV sequences, the data obtained from this experiment can be considered a first step in determining the extent of these effects. These data will be examined in two ways: first, as strictly normative and second, since low-pass filtering somewhat resembles a high-frequency hearing loss, as a basis for speculating on certain clinical speech-discrimination problems.

I. PROCEDURES

The general procedure was to construct lists of various consonant-vowel syllables and record, filter, and play back these lists to a group of listeners.

The stimuli consisted of the 16 consonants, /p,t,k, b,d,g,s,f,z,v,w,j,r,l,m,n/, each in CV combination with the seven vowels, /i,ɛ,æ,a,ʌ,ɔ,u/. The total number of syllables was 112. These items, each repeated three times, were randomized into a master list. Three such randomizations were made, one for each of the three speakers. The speakers were three adult males whose speech was typical of the New York City dialect area. Recordings were made on one track of an Ampex,

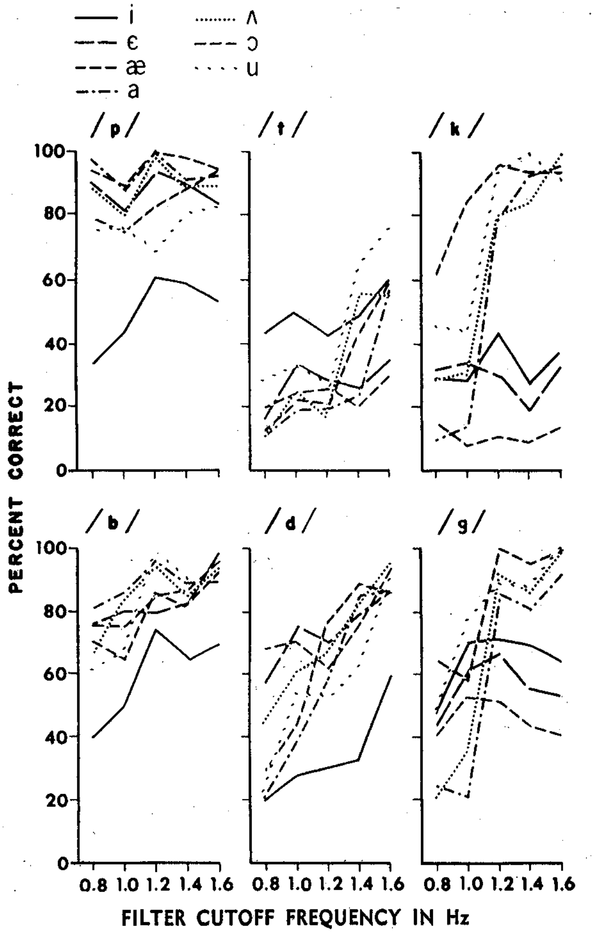

FIG. 1. Mean correct scores for six stops and five filtering conditions.

model AG-500 two-track tape recorder through an Electrovoice, model 654 microphone. The items were recorded at approximate 3-sec intervals with longer rest periods occurring after groups of 10. The carrier word "write" preceded each utterance. Gain levels for each speaker were adjusted so that the vowel /ɔ/ peaked at zero on the tape recorder's VU meter. Other than that, no attempts were made to equalize within-list gain levels. This meant, of course, that due to normal vowel-level differences, relative intensities among the tokens differed by as much as 8 dB. The master tape, then, contained all stimuli, each repeated three times by each of three speakers for a total of 1008 items (112×3×3).

This tape was edited into five different randomizations, one for each of five low-pass filter conditions. Filter cutoff points were 800, 1000, 1200, 1400, and 1600 Hz. Exploratory work showed these settings to cover the range between apparent chance responses and unmeaningfully high scores. The filtering was accomplished by playing the tapes back on one Ampex, AG-500, through two Allison, model 2B variable filters connected in series and re-recording the tapes on a second Ampex, AG-500. The filters provided a roll-off of approximately 60 dB/oct.

Listeners were seven normal-hearing undergraduate and graduate college students. Each was told about the makeup of the lists in only general terms. The response mode was open-set with the listeners free to choose any phonemically permissible CV combinations. Twenty-five practice items preceded each filter condition. The tapes were played back to the subjects (random-order presentation) binaurally, through Telephonics TDH-39 earphones in a quiet but not fully sound-treated room. Playback levels for all lists were adjusted to approximately 80 dB over-all SPL, as measured on a Brüel & Kjær audiometer calibration unit.

II. RESULTS

As would be expected, vowels were highly intelligible under all filter conditions and, except for /i,u/, exceeded 95% in all cases. Not unexpectedly, /i,u/ were sometimes confused with each other (consistently more /u/confusion for /i/ rather than vice versa), but with no observable consonant influence. Also, although evidence of occasional speaker influences for a small number of tokens existed, there were no consistent trends, and, thus, all data were averaged over the three speakers. As expected then, the major effects are those of filter condition and vowel environment.

The percent-correct scores for all consonants, under each filter condition, are plotted separately for each consonant category in Figs. 1–4.

A. Stops

Figure 1 shows the mean-percent scores for the group of stop consonants, /p, t, k, b, d, g/. As the graphs show, each consonant is somewhat differently affected by filter cutoff point and vowel environment. For /p/, there is no consistent filter cutoff effect, as the curves run moderately flat across the five filter conditions. The most conspicuous vowel effect is for /i/, where scores are consistently lowest. This might be explained somewhat by the fact that since the second formant for /i/ is somewhere in the vicinity of 2200 Hz, much of the information-bearing second formant transition rising to this level is probably eliminated by the filtering. (Similar /i/ effects occur for four of the remaining 15 consonants.) Like /p/, /b/ shows no real vowel effect (except for /i/), but scores generally increase with the more favorable filter conditions.

Unlike their labial counterparts, /t, d/ bear little similarity to each other. Scores for /t/ followed by /i/ are clearly higher, except at the two highest cutoff points. A cutoff effect exists at only 1400 Hz for three of the seven vowels. /d/, on the other hand, is characterized by a sharp increase across the cutoff points along with the deleterious effect of a following /i/.

/k, g/ are the most interesting of the stops. Here, both pronounced vowel and cutoff effects occur. For both consonants, back-vowel combinations show an increase in intelligibility at cutoff points of 1200 Hz and higher. A ready explanation of this occurrence can be found in the synthetic speech work of Delattre, Liberman, and Cooper,[1] who found that the course of the formant transitions for /k, g/ originates at two different starting points in frequency. The theoretical starting point or locus of a /k, g/ transition for a back vowel was found to be approximately 1200 Hz, while the locus for a front-vowel transition was at about 3000 Hz. The filtering effects found here, then, can be explained by the fact that information for /k, g/ preceding a back vowel does not appear below frequencies of 1200 Hz (thus, the lower scores for filter points below 1200 Hz) and that significant information for /k, g/ preceding a front vowel does not appear at frequencies below 3000 Hz (with lower scores expected for all cutoff points).

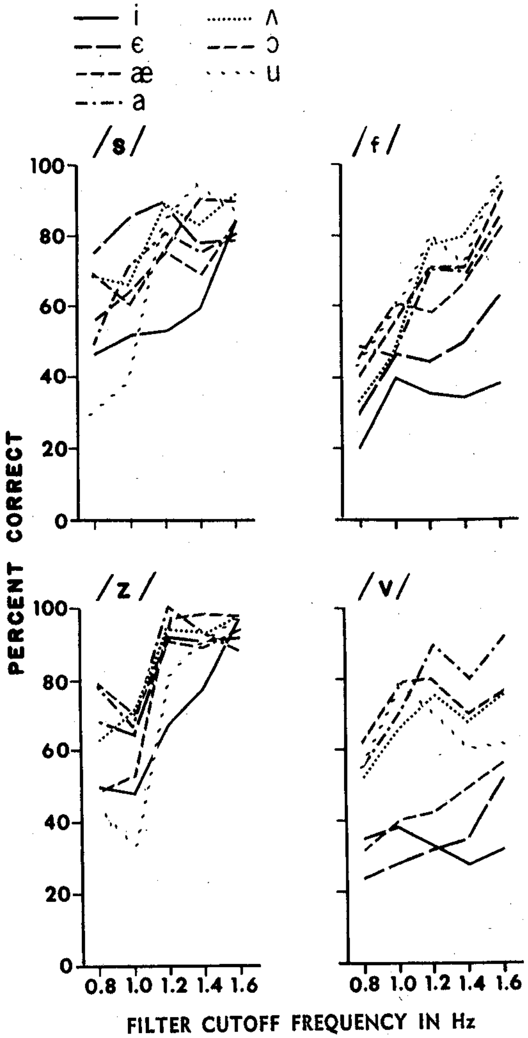

FIG. 2. Mean correct scores for four fricatives.

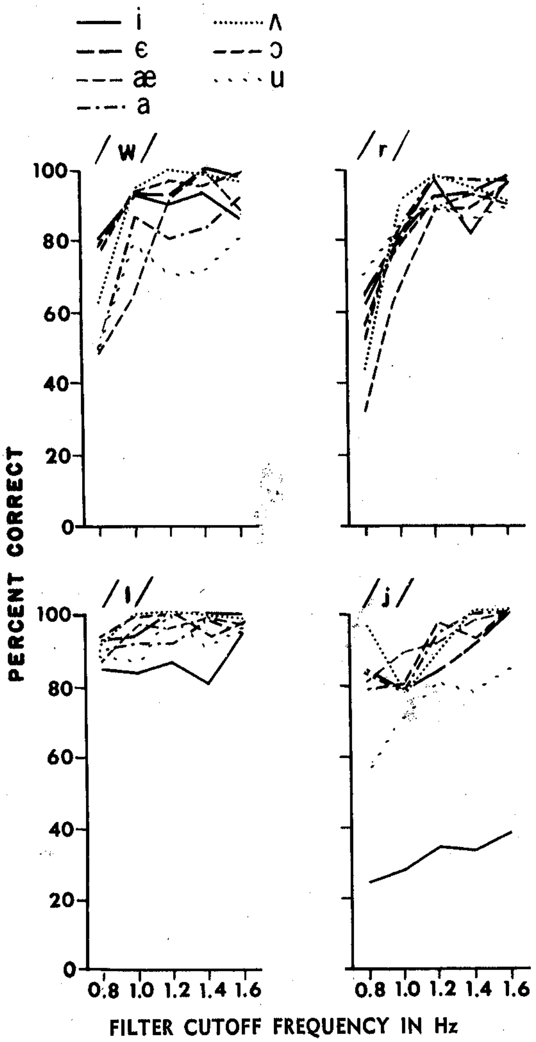

FIG. 3. Mean correct scores for four semivowels.

B. Fricatives

The fricatives, /s, z, f, v/, like the stops, show certain individual peculiarities (Fig. 2). For /s/, the vowel effects are inconsistent, but the curves generally rise across the filter cutoff points. In general, /s/ followed by /i/ shows the lowest scores. /z/, on the other hand, although showing no real vowel effects (except for /i/), shows a sharp increase in intelligibility beginning at the 1200-Hz position.

Both cut-off frequency and vowel environment affect /f, v/ identification. For both consonants, but especially /v/, back-vowel combinations are more intelligible than front-vowel combinations. These effects are superimposed upon the increases across cutoff points. The behavior of the fricatives might be explained by the fact that, while /s, z/ are identified primarily by their noise characteristics, /f, v/ are cued more by their second formant transitions.[2,3] The assumption here is that more transition information remains intact for

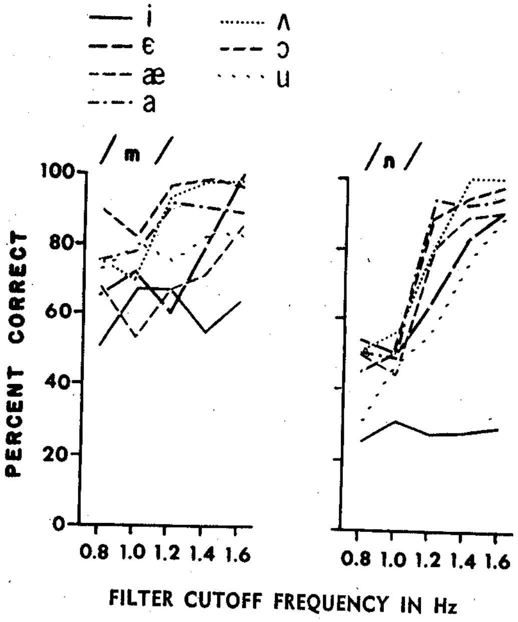

FIG. 4. Mean correct scores for two nasals.

back-vowel combinations as the transitions extend down to lower frequencies.

C. Semivowels

The results for the semivowels, /w, r, l, j/, are shown in Fig. 3. Filter cutoff effects occur for all consonants except /l/, whose intelligibility is highest of all consonants, regardless of filter cutoff conditions. /w, r/ show the greatest cutoff effects. The vowel effects for /w/ are somewhat unusual in that higher intelligibility generally accompanies front vowels, especially at the lowest cutoff points. /r/ also shows vowel effects at the lowest cutoff point. There are no real vowel effects for /l/ (except for a slight decrease in scores for /i/). The /i/ effect for /j/ is the most marked of any consonant.

D. Nasals

The results for /m, n/ are plotted in Fig. 4. Both sounds show marked (though complicated) vowel and cutoff effects, with strong vowel-filter interactions most evident for /m/. In general though, front-vowel curves are somewhat lower than back-vowel curves. Except for /i/, the only significant vowel and cutoff effects for /n/ occur at 1200 Hz. However, no special vowel-group preference emerges.

E. Error Types

Figure 5 summarizes the types of confusions that occurred for each of the six different consonant categories and five low-pass filter conditions.[4] As would be expected, place errors accounted for most of the confusions, regardless of filter cutoff frequency.[5] At the lower cutoff frequencies, however, additional error types occur, generally in the order of manner, voicing, and nasality. As can be seen from the graphs, only the voiceless stops /p, t, k/ are characterized almost wholly by place errors.

To summarize the above results briefly: the effects of filter cutoff frequency and vowel environment on consonant perception are complicated. Some consonants are affected by filter cutoff points, others are not. Those affected are /t, k, b, d, g, s, f, z, w, r, n/. Likewise, some consonants are affected by vowel environment, while others are not. The greatest multivowel effects occur for /k, g, f, v, m/. Of the sixteen consonants, /p, b, d, j, n/ show consistently lower scores when followed by the vowel /i/. Error types were predominantly "place" with "manner," "voicing," and "nasality" errors occurring at only the less-favorable cutoff points.

III. DISCUSSION

A. Filter-Transition Relationships

As was mentioned at the outset, a reasonable basis exists for predicting the perceptual effects of certain consonants heard under conditions of low-pass filtering. This, of course, is based on the cue information provided by the CV transition and the extent to which it

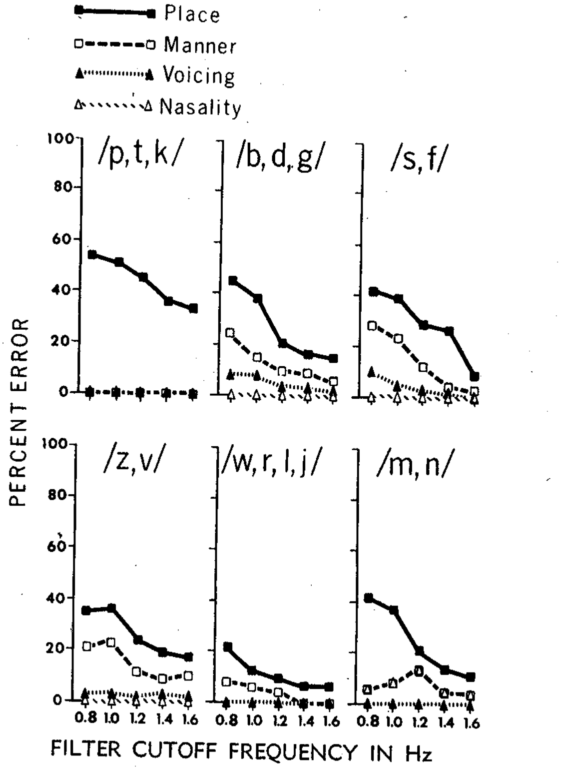

FIG. 5. Mean error scores showing type of error for six groups of consonants and five filtering conditions. Note: all manner errors for /m,n/ are also, by definition, nasality errors.

is eliminated by the filtering. These cutoff and vowel effects were most clearly demonstrated in the /k, g/ data, which supported Delattre, Liberman, and Cooper's[6] notion of a variable locus for these phonemes. The perception of some of the other sounds, however, including the remaining four stops, is not so easily explained. If a fixed locus for the labials (720 Hz) and dentals (1800 Hz) is assumed, then a lower level of intelligibility would be expected for those stimuli containing vowels with a higher frequency *F*2, since more of the transition is eliminated by the filtering (it is assumed that virtually all *F*3 information is missing under these conditions).[7] This, however, is not always the case.

For both /p, b/, a following /i/ might be expected to degrade the consonant's intelligibility, and this, indeed, is borne out by the data. The remaining vowels, on the other hand, do not follow in this order. The over-all picture is one rather of a grouping of the remaining curves without any hierarchical vowel preference. The data for /t, d/ are perhaps even more unusual. For /t/, at all but the two highest cutoffs, a following /i/ provides the highest-intelligibility levels, whereas for /d/, a following /i/ is accompanied by the lowest-intelligibility levels, at all cutoff points.

There are perhaps three explanations for the variability found for both sets of stops. First, certain unfiltered segments of the transition might, in one way or another, provide the necessary place cues; second, supplementary cue information might be contained in the burst segment of the phoneme; and third, perceptually significant variations might exist in the transition starting points, or even loci, of these phonemes. Support for this last possibility can be found in a recent experiment of Fant,[8] whose measurements for Swedish stops in CV syllables showed some large variation in *F*2 and *F*3 transition starting points, depending on the following vowel.

The behavior of the fricatives is generally straightforward, with a minimal vowel effect for /s, z/ and an important, predictable one for /f, v/. As was mentioned earlier, this can be explained by the fact that /f, v/ are cued primarily by their transitions, which remain more intact when extending down to the lower *F*2 back vowels. /s, z/, on the other hand, are cued more by their noise segments, the major portions of which are located above the filter cutoff points. Cutoff effects occur for all consonants, with those for /s, z/ apparently due to the increased presence of the frication. The /f, v/ filter effects, like the vowel effects, are more consistent, with increases for all vowels occurring with each increase in cut-off frequency.

Except for /ji/ (and perhaps /li/), the semivowels show few consistent vowel effects. This is not unusual, as these phonemes are distinguished from one another by the onset frequencies of their *F*1, *F*2, and *F*3 transitions. Somewhat unusual, however, are the cutoff effects for /w, r/. This is especially true of /w/, which is presumed to be cued by low frequency *F*1 and *F*2, in contrast, for example, to the higher-frequency starting points of /l/, which show no cutoff effects.[9]

The place cues for /m, n/ are generally considered to be identical to those of the stops and, thus, might be expected to behave somewhat like their labial and dental counterparts. Unfortunately, the data for /m/ are not very clear, although it might be suggested that the front-vowel stimuli, as a whole, are less intelligible than the back-vowel stimuli. The curves for /n/ seem to be similar to those of /d/, but with sharper slopes.

In summary then, the filtering effects for four of the six stop consonants (/p, t, b, d/) can not be related clearly to the course and extent of their CV transitions. Fricative behavior is generally straightforward, but unexplainable are the cutoff effects for the semivowels /w, r/ and, perhaps, the lack of them for /l/.

B. Clinical Implications

Although the results of this experiment are essentially normative, they can be applied to certain speech-discrimination problems due to hearing impairment. This is not to say, however, that low-pass filtering produces the same effects as a high-frequency hearing loss.[10] The comparisons made here are based only on the fact that similar portions of the spectrum are eliminated by the two conditions and that this might produce some similar perceptual effects. In this sense then, if these, or similar, vowel and cutoff effects exist for those with hearing impairments, then the use of a small sample word list, such as the W-22's, for testing speech discrimination would suggest the possibility of certain perceptual biases caused by the presence or absence of a given phoneme sequence. This assumes, of course, that common phoneme sequences are not adequately represented in the W-22 distributions. As was mentioned earlier, the W-22 frequencies, originally based on those of Dewey,[11] involved only over-all frequencies of occurrence. Not until 1963, with the publication of Denes'[12] data, was there any detailed information available on CV, VC, or CC syllable frequencies. When the present W-22 lists are analyzed according to these frequencies, however, the following can be noted: first, many familiar CV and VC syllables are not represented in the W-22 lists, and second, between 20%–25% of the W-22 words contain consonant clusters, most of which are hardly common in everyday speech. The significance of especially the latter is that the acoustical characteristics and, consequently, perceptual cues of many consonants are quite different when in CC or CV position. Specifically, the first element of a cluster is no longer characterized by its often perceptually significant second formant transition.

Apparently then, the internal phonemic make-up of the present PB words is not adequate. Although adequate representation can be built into a list, the job would be difficult and the results cumbersome. Indeed, it could also be argued that, in the clinical sense, such representation might not even be necessary. Since certain consonants and vowels are highly resistant or even insensitive to most hearing-loss conditions, these phonemes might be replaced in a list by those that show more complicated effects or interactions. This approach would probably provide a more detailed, less redundant account of an individual's speech-discrimination ability. Although lists of this nature are not as yet available, some presently existing lists can be adapted. For example, both Fairbanks'[13] Rhyme Test and House, Williams, Hecker, and Kryter's[14] closed response CVC lists control phoneme environment and, in addition, have the added advantage of allowing an inventory of specific phoneme errors to be easily made.

Finally, it might be mentioned that the data of this experiment can also be applied to the selection and use of speech materials for clinical auditory training. They might be useful in providing a basis for determining the degree of difficulty for various syllables and words used in clinical sessions, especially beginning ones.

ACKNOWLEDGMENTS

The author wishes to thank Mrs. Frieda Toback and Mrs. Celia Dorrow for their assistance in the collection and analysis of the data.

---

[1] P. C. Delattre, A. M. Liberman, and F. S. Cooper, "Acoustic Loci and Transitional Cues for Consonants," J. Acoust. Soc. Amer. **27**, 769–773 (1955).

[2] K. S. Harris, "Cues for the Discrimination of American English Fricatives in Spoken Syllables," Language and Speech **1**, 1–7 (1958).

[3] J. M. Heinz and K. N. Stevens, "On the Properties of Voiceless Fricative Consonants," J. Acoust. Soc. Amer. **33**, 589–596 (1961).

[4] Error types were classified as place, manner, voicing, and nasality. Multiple-type errors were counted in each appropriate category; e.g., if a /p/ was heard as a /d/, the error would be classed as both a place and voicing error.

[5] G. A. Miller and P. E. Nicely, "An Analysis of Perceptual Confusions among Some English Consonants," J. Acoust. Soc. Amer. **27**, 338–353 (1955).

[6] Ref. 1.

[7] The over-all higher intelligibility of /p,b/ over the rest of the stops can be interpreted in much the same way; that is, since the labials are characterized by a lower-frequency transition starting point, they are less vulnerable to missing higher-frequency components.

[8] G. Fant, "Stops in CV-Syllables," STL-QPSR (April 1969), pp. 1–25.

[9] J. O'Connor, L. J. Gerstman, A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Acoustic Cues for the Perception of Initial /w,j,r,l/ in English," Word **13**, 24–43 (1957).

[10] Besides the lack of evidence supporting a comparison of filtering with the pure-tone audiogram, filtering does not take into account factors such as recruitment, equal loudness contour effects, or any other nonlinear distortion that might accompany a high-frequency hearing loss.

[11] G. Dewey, *Relative Frequency of English Speech Sounds* (Harvard U. P., Cambridge, Mass., 1923).

[12] P. B. Denes, "On the Statistics of Spoken English," J. Acoust. Soc. Amer. **35**, 892–904 (1963).

[13] G. Fairbanks, "Test of Phonemic Differentiation: The Rhyme Test," J. Acoust. Soc. Amer. **30**, 596–600 (1958).

[14] A. S. House, C. E. Williams, M. H. L. Hecker, and K. D. Kryter, "Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set," J. Acoust. Soc. Amer. **37**, 158–166 (1965).