Veterans Administration

Journal of Rehabilitation Research and Development, Vol. 23, No. 1, 1986 pages 89-94

Some Rehabilitative Considerations for Future Speech-Processing Hearing Aids[a]

S. G. REVOILE, Ph. D.
L. D. HOLDEN-PITT, M.S.
D. M. EDWARD, M.E.D.
J. M. PICKETT, Ph. D.

Sensory Communication Research Laboratory
Gallaudet College
Washington, D. C. 20002

Abstract—Improvements in speech-recognition and speech-processing technology promise eventual ways of assisting the hearing impaired by automatically enhancing the audibility of critical speech segments or distinctive features. Some results of applying enhancement techniques are summarized and procedures are proposed for selecting (i) the speech sounds requiring enhancement, (ii) the degree of amplification, and (iii) the training that hearing-impaired listeners might need.

INTRODUCTION

With advances in computer recognition of speech, future technology may offer real-time modifications of individual acoustic segments in speech. Hearing aids could then be developed that would house miniature computers capable of altering particular acoustic segments of speech, as described by Haggard and Trinder in 1983 (1) and by Schafer in 1982 (2). These speech-processing hearing aids might include two major stages: in one stage, given speech acoustic segments would be "recognized" by the computer; in a subsequent stage those particular segments would be modified or enhanced by the computer to increase their salience for persons with deficient auditory reception for those speech segments: see Haggard, 1980 (3). Such hearing aids could be particularly useful for hearing-impaired persons with reduced speech comprehension due to poor discrimination of particular acoustic patterns.

Aside from the performance quality of future speech-processing hearing aids, several other factors may affect the success of such devices. These factors are similar to concerns addressed in the initial stages of aural rehabilitation for hearing-impaired persons using conventional hearing aids. For speech-processing aids, it will first be necessary to identify the particular segments of speech to be enhanced for a given hearing-impaired person. Second, it must be determined how much enhancement is needed. Third, it will have to be known what kind of training procedure best facilitates accurate perception of speech with enhanced acoustic segments. Some research at Gallaudet College and at other institutions is currently aimed at providing preliminary answers to these questions.

Prior to descriptions of aural rehabilitation for persons using future speech-processing hearing aids, it is important to examine the work done thus far, which demonstrates that speech perception by the hearing-impaired can be facilitated through enhancements of speech acoustic segments.

[a]This work was supported by the Gallaudet Research Institute and by Grant No. G008103979 to Gallaudet College from the National Institute of Handicapped Research (NIHR) for the program of the Rehabilitation Engineering Center for the Deaf and Hearing-Impaired.

An oral version of this research report was presented at the Second International Conference on Rehabilitation Engineering in Ottawa, June 1984.

THE QUESTION TO BE ANSWERED

Recent studies of hearing-impaired listeners in this laboratory show that enhancements of individual acoustic segments in speech can improve some listeners' perception of the consonants associated with those speech segments. We have studied speech enhancements for the perceptual feature called the "voicing distinction" for the final consonants / s /, / z /, / f /, and / v / in the syllables BASS, BAZZ, BAFF, and BAV. When a listener can distinguish voicing for these consonants, it means only that the listener distinguishes the / s / versus the / z /, and the / f / versus the / v /. In our prior work we have found that some hearing-impaired persons depend mostly on certain temporal cues in speech (such as the vowel duration preceding final consonants) for distinguishing consonant voicing. Other hearing-impaired persons rely on both temporal and spectral cues: see Revoile et al., 1982 (4) and 1984 (5). (An example of an important spectral cue is the presence of low-frequency energy that may occur when the / z / and / v / are spoken; in contrast, this energy is absent when the / f / and / s / are spoken.)

In our prior studies (4, 5), cue-dependence by the hearing-impaired listeners was established through hearing tests using syllables that have had various consonant cues systematically neutralized or deleted. Since deletion of certain cues degraded the listeners' consonant perception, we wondered whether exaggeration or enhancement of these same cues might increase their salience and consequently improve consonant perception for listeners with deficient hearing for speech. The cues chosen for enhancement were those identified in our prior work as important for impaired listeners' perception of the voicing distinction for final / f / versus / v /, and / s / versus / z /; viz., vowel duration and spectral properties of the consonants.

METHODS

The enhancements of the temporal and spectral acoustic cues were carried out on different utterances of each of the test syllables: BASS, BAZZ, BAFF, BAV. These utterances had been spoken by a female and subsequently converted from analog to digital form (16.67 kHz sampling rate) for acoustical measurements and for processing of the cue enhancements. (The acoustical measurements included certain duration and spectral measurements of the vowel and consonant segments of each utterance.)

For the study on temporal cue enhancements, the vowel preceding the final / s /, / z /, / f /, or / v / was altered in duration: see Revoile et al., 1984 (6). In syllables with final / z / or / v / the preceding vowel was lengthened, while in syllables with final / s / or / f / the preceding vowel was shortened.

For the study on spectral cue enhancements, the final consonants' friction noises were altered spectrally: see Revoile et al., 1984 (7). The / f / and / s / frictions were low-pass-filtered and then intensified. The / v / and / z / frictions were replaced by waveform segments that occurred at the end of the vowel in each utterance. Those segments contained some consonant friction noise. The segments were low-pass-filtered, and also high-pass-filtered to reduce the contribution of the vowel. Finally, the segments were lengthened and intensified. For all of the syllables, the consonants' friction noises were altered without modification to the preceding vowels or the initial / b /'s.

In both studies, undergraduate students at Gallaudet College participated as listeners (N = 46). About one-half of the listeners had profound losses (≥90 dB Hearing Level for 0.5, 1, and 2 kHz threshold average). The other listeners had either severe or moderate losses.

The listeners' perception for the consonants was tested using syllable identification trials. The syllables were presented at an amplified sound level such as would be the case if the listener were wearing a hearing aid: this level was set individually for each listener. During the trials, each test utterance was presented singly. The listener was required to choose which of the four syllables (BASS, BAZZ, BAFF, or BAV), was presented. No feedback of response accuracy was provided.

Blocks of trials were tested first for the syllables without cue alterations. Subsequently, the syllables with enhanced cues were tested. Generally, the pairs of syllables (those without versus those with the enhanced cues) were tested in separate listening sessions. Most sessions also included some syllable-discrimination procedures, which facilitated the listeners' perceptual use of the consonant voicing distinctions. The listening sessions occurred twice weekly for more than 2 months, over which time repeated tests of consonant perception were administered.

RESULTS

In both experiments, the majority of listeners showed considerable performance improvements

for the consonants in syllables with enhanced cues, compared to consonant perception for the unenhanced syllables.

For the study of syllables with vowel-duration enhancements, consonant voicing perception is summarized in Table 1 for 15 of the listeners who demonstrated the largest improvements in performance (N = 25). The scores represent total percent-correct voicing perceptions of the consonants, averaged across all of the syllable utterances presented in a test. The values shown were based on at least three administrations of each test for most listeners. For the syllables with unenhanced cues, average performance was 63 percent for this group of listeners. (For reference, note that normal-hearing listeners would obtain 100 percent correct perception of the consonants.) Thus, these listeners' voicing perception for the consonants was rather poor when the syllable's vowel durations were not enhanced. In contrast, the listeners' consonant voicing perception was quite good when the syllables' vowel durations were enhanced—the groups' average performance was 92 percent, which is close to normal performance for the unenhanced syllables. This result reveals that enhanced vowel durations can improve voicing perception of final consonants for some hearing-impaired persons who have poor perception of those consonants in syllables without vowel-duration enhancements.

TABLE 1
Median and quartile percent-correct voicing for final fricatives in syllables unenhanced and in syllables with enhanced vowel duration (Listener N = 15).

	Unenhanced syllables (% correct)	Syllables with enhanced vowel duration (% correct)
Third quartile	70	96
Median	63	92
First quartile	55	89

TABLE 2
Median and quartile percent-correct voicing for final fricatives in syllables unenhanced and in syllables with enhanced friction cues (Listener N = 14).

	Unenhanced syllables (% correct)	Syllables with enhanced friction cues (% correct)
Third quartile	69	87
Median	62	85
First quartile	56	77

For the syllables with enhanced spectral cues, Table 2 summarizes the results. The scores shown represent tests of the 14 listeners who showed the largest improvement between their performance on the unenhanced and the cue-enhanced syllables (N = 21). For the syllables without spectral cue enhancements these listeners' consonant voicing perception averaged 62 percent. However, a sizable performance improvement, to 85 percent, was seen when the same listeners were tested with the syllables containing enhanced spectral cues in the consonant friction noises. This finding indicates that emphasis of certain spectral cues in consonant friction noises can improve voicing perception of those consonants for some hearing-impaired persons.

In both experiments, some listeners showed little change in performance between the cue-enhanced versus the unenhanced syllables. These were found to be listeners whose voicing perception for the consonants in the unenhanced syllables was either very good or very poor. For the listeners with good performances, only minor improvements were seen with the cue-enhanced syllables, since these listeners showed near-normal perception for the consonants when the syllables were not enhanced. For the listeners showing very poor perception, several explanations may be offered for their lack of improvement with the cue-enhanced syllables. One possibility is that they may have been experiencing tactile rather than auditory perception of the syllables. It is known that the temporal and spectral resolving power of the tactile sense is inferior to that of hearing.

These studies have shown that speech acoustic segments enhanced via computer can facilitate the perception of consonants by hearing-impaired listeners. Other work suggests that enhancements to facilitate speech perception may be accomplished merely by a careful speaking style: see Picheny and Durlach, 1979 (8). A talker, speaking in two different styles (casual or conversational versus more-careful enunciation) recorded identical sentences in both styles. For the sentences spoken more clearly, hearing-impaired listeners demonstrated better perception of consonants than for the sentences spoken casually. This effect may have been due to

the heightened salience of some acoustic segments in the clearly-spoken sentences. Research in speech science indicates that certain acoustic segments of speech may be minimized in conversational speech, in contrast to the same acoustic segments of individually spoken words.

DISCUSSION

The authors believe these studies indicate that perception of consonants by hearing-impaired listeners can be affected by the salience of acoustic segments associated with those consonants: consonant perception improves for heightened salience or enhancement of certain acoustic segments. Much additional work is needed to demonstrate the effectiveness and feasibility of enhancing acoustic segments in conversational speech.

In continuing this work, we will also study the procedures to be used for future rehabilitation of hearing-impaired persons using devices that enhance speech acoustic segments. Some of the procedures examined thus far are described below in the discussion of the rehabilitation process.

Selecting Acoustic Segments for Enhancement

In the summary above of rehabilitation concerns for future speech-processing hearing aids, it was said that an early step will be to identify the speech segments that are inaudible or indiscriminable for the hearing-impaired person. Identifying them might be a two-phase procedure; first, to learn generally which speech sounds (i.e., consonants and vowels) are misperceived by a hearing-impaired listener. Those results could provide overall information about the speech acoustic segments that may require enhancement for that listener. The second phase would involve tests of speech sound perception using words or syllables with particular acoustic segments systematically eliminated. Those tests should show in more detail which speech acoustic segments do or do not contribute to the listener's perception of specific speech sounds.

The initial tests of speech sound perception could be structured to reveal whether particular classes of speech sounds are misperceived. The tests would consist of word-recognition trials. A different word would be presented for each trial and the listener would choose an answer from a set of printed words. Recent examples of such tests are seen in Levitt and Resnick, 1978(9); Owens and Schubert, 1977 (10); Dubno et al., 1982 (11); and Boothroyd, 1984 (12).

Certain classes of consonants or of vowels have some acoustic patterns in common. The tests of speech sound perception would be conducted with the sounds grouped according to particular similarities and differences in acoustic patterns of the sounds. The listener's errors or confusions among speech sounds in certain classes would show generally the types of acoustic patterns that may be inaudible or indiscriminable for a given hearing-impaired listener.

More-detailed understanding of a listener's reception for acoustic segments could be obtained from consonant perception tests (via word-recognition trials) in which the physical values of segments associated with particular consonants are systematically altered. Using a given group of recorded words, for example, several test versions could be prepared that would each contain some different alterations to acoustic segments associated with the test consonants. Certain individual segments could be altered by waveform deletion or by spectral filtering, while other acoustic segments of the same test consonants would be left unaltered. A listener might show reduced consonant recognition for a test version in which a particular type of segment was deleted. This would reveal that the listener's recognition for the test consonants is dependent on that type of acoustic segment. Conversely, other acoustic segments of the test consonants, that were unaltered, might seem not to contribute to the listener's perception for the consonants. Thus, these segments could be selected for enchancement to facilitate the listener's consonant perception by providing cue-redundancy equal to that which is available for normal listeners. In previous work, this laboratory has used such tests to study the contribution of acoustic segments to consonant perception by hearing-impaired persons (4).

Assessing Enhancement Values

To determine the amount of enhancement needed, both consonant perception and auditory discrimination might be tested. The consonant-perception tests would be designed to show whether accurate perception could be obtained for enhancement values that simulate those found in carefully articulated speech. These enhancement values would be applied by computer to the acoustic segments of test consonants in words to be presented for recognition trials. Listeners achieving good consonant perception of these stimuli would be those for whom such enhancement values might be simulated in a speech-processing hearing aid. Further, such listeners would probably require little or no perceptual training for speech with enhanced acoustic segments, since their perception had improved upon

initial presentation of the enhanced speech.

For these same stimuli, other listeners might show poor perception, which possibly could indicate that greater enhancement and/or training would be required to facilitate their consonant perception. Auditory discrimination tests would be administered to these listeners, using artificial (computer-generated) speech containing acoustic segments with various types of modification. These modifications could be made to vary considerably relative to the range of characteristics of acoustic segments in natural speech. Discrimination tests with these stimuli would show the listener's auditory ability to distinguish among speech-like stimuli with widely varying characteristics. As mentioned above, synthetic speech stimuli have been used in recent work to assess perception of certain consonants by hearing-impaired listeners: see Johnson et al., 1984 (13), and Godfrey and Millay, 1978 (14). Such stimuli may also be used in clinical assessment of speech perception, as described by Fourcin in 1980 (15).

Training for Perception of Enhanced Speech

For hearing aids that enhance speech segments, the training used for habilitation with these aids may depend on various factors. The extent to which a hearing-impaired person has previously used hearing for speech communication will certainly influence the amount and types of training required. For some hearing-impaired persons, only minimal training may be necessary if the hearing aid provides enhanced speech that closely simulates natural speech. This possibility is suggested by the results seen for a few listeners with reduced consonant perception in our studies described above and in (5). Some of these listeners showed nearly perfect perception of consonants with enhanced segments upon the first presentation of those consonants. For such listeners, perception training for enhanced speech would probably be unnecessary.

For listeners who will require training to facilitate perception of consonants with enhanced acoustic segments, we are still uncertain as to the procedures that should be employed. There is a dearth of research that demonstrates the effectiveness of auditory training used for aural rehabilitation, according to Bamford, 1981 (16), and few studies have compared different types and adaptations of auditory training procedures. Consequently, there is little robust data to provide definitive direction for the selection of auditory training procedures.

In this laboratory's experiments thus far, we have used training procedures that require listeners to discriminate and identify syllables presented in succession, with feedback following each response. These procedures were presented during the experiments described above, with the enhancements to facilitate perception of the voicing distinction among the consonants / f /, / s /, / v /, and / z /.

One of the training procedures was a "syllable-group" identification task. On each trial, three syllables were presented in sequence; the first and third syllables were identical, while the second syllable was different, but only in its final consonant. For given blocks of trials, different utterances were used of two syllables; e.g., BASS and BAZZ. Thus, only two responses could occur for a block of trials with these syllables, viz. BASS BAZZ BASS or BAZZ BASS BAZZ. Prior to each of the first eight trials of a block, a blinking light showed the listener which syllable group would be presented. That prompting helped the listener to associate the correct identification labels with the syllable group presented per trial. After each of the listener's responses, a blinking light provided feedback of the correct answer.

The second training procedure was a paired-comparison task. On each trial, a pair of syllables presented, either two different utterances of the same syllable, or one utterance each of two different syllables. For a given block of trials, utterances were presented for only two syllables, e.g., BASS and BAZZ. Four response choices were available: BASS BASS, BAZZ BAZZ, BAZZ BASS, and BASS BAZZ. Following each response, the listener received feedback of the correct answer.

Although these training procedures were used in the experiments described above, control was inadequate to determine the effectiveness of the training procedures for improving consonant perception: see Walden et al., 1981 (17), and (16). Some of the listeners did show considerable improvement in consonant perception, but we believe improvement was due more to the enhancement of the acoustic segments associated with the consonants than to the training procedures used. As noted above, further study of enhancement of speech acoustic segments should be accompanied by study of procedures for training perception of enhanced speech by hearing-impaired listeners.

SUMMARY

In the future, hearing aids may be available that can selectively enhance acoustic segments in speech. These aids could be useful for hearing-impaired persons who have deficient auditory perception for speech acoustic segments. For those

with reduced speech perception, speech with enhanced segments might facilitate perception. Preliminary studies have shown improvements in perception of certain consonants for some hearing-impaired persons who were presented with syllables that had enhancements of the consonants' acoustic segments. Further study of different types of enhancements for various consonants is being carried on. Research is needed to develop effective procedures to use in the rehabilitation of persons who will be fitted with future speech processing hearing aids ■

REFERENCES

1. Haggard MP, Trinder JR: Problems in Realizing the Potential of Integrated Circuits in Future Aids for the Hearing-Impaired. **In** High Technology Aids for the Disabled, W. J. Perkins, Ed., London: Butterworth, 1983.
2. Schafer RW: Speech Processing for the Hearing Impaired. **In** The Vanderbilt Hearing-Aid Report: State-of-the-Art Research Needs. G.A. Studebaker and F. H. Bess, Eds. Pages 130-131. Upper Darby, Pennsylvania: Monographs in Contemporary Audiology. 1982.
3. Haggard MP: Speech Processing and Hearing-Aid Design: Some Critical Reflections. **In** Disorders of Auditory Function III. I. G. Taylor and A. Markides, Eds. London: Academic Press. 1980.
4. Revoile S, Pickett JM, Holden L, Talkin D: Acoustic cues to final stop voicing for impaired- and normal-hearing listeners. J Acoust Soc Am 72:1145-1154, 1982.
5. Revoile S, Holden-Pitt L. Pickett JM: Perceptual cues to the voiced-voiceless distinction of final fricatives for listeners with impaired or with normal hearing. J Acoust Soc Am, 77: 1263-1265, 1985.
6. Revoile S, Holden-Pitt L, Pickett JM: The effects of duration adjustments of preceding vowels on fricative voicing perception by hearing-impaired listeners. J Acoust Soc Am, Suppl 1, S5(A), 1984.
7. Revoile S, Holden-Pitt L, Pickett JM, Edward D, Brandt F: Improved Perception by the Hearing Impaired for Syllables Processed with Enhanced Consonant Spectral Cues. Paper presented at Second International Conference on Rehabilitation Engineering, Ottawa, Canada, 1984.
8. Picheny MA, Durlach NI: Speaking clearly for the hard of hearing. J Acoust Soc Am, Suppl 1, 65, S135(A), 1979.
9. Levitt H, Resnick SB: Speech reception by the hearing-impaired: Methods of testing and the development of new tests. **In** Sensorineural Hearing Impairment and Hearing Aids, C. Ludvigsen and J. Barfod, Eds. Scand Audiol, Suppl 6, 1978.
10. Owens E, Schubert E: Development of the California consonant test. J Speech Hear Res 20:463-474, 1977.
11. Dubno J, Dirks D, Langhofer L: Evaluation of hearing-impaired listeners using a nonsense-syllable test. II: Syllable recognition and consonant confusion patterns. J Speech Hear Res 25:144-148, 1982.
12. Boothroyd A: Auditory perception of speech contrasts by subjects with sensorineural hearing loss. J Speech Hear Res 27:134-144, 1984.
13. Johnson D, Whaley P, Dorman M: Processing of cues for stop consonant voicing by young hearing-impaired listeners. J Speech Hear Res 27:112-118, 1984.
14. Godfrey J, Millay K: Perception of rapid spectral change in speech by listeners with mild and moderate sensorineural hearing loss. J Am Audiol Soc (now Ear and Hearing) 3:200-208, 1978.
15. Fourcin AJ: Speech Pattern Audiometry (Oxford Medical Engineering Series). Oxford, England: Clarendon Press, 1980.
16. Bamford J: Auditory training. Br J Audiol 15:75-78, 1981.
17. Walden B, Erdman S, Montgomery A, Schwartz D, Prosek R: Some effects of training on speech recognition by hearing-impaired adults. J Speech Hear Res 24:207-216, 1981.