In summary, the data from the present experiment indicate that the hierarchy of MLD's is the same in simultaneous and temporal masking; however, there are differences in the amount of additional masking obtained across monaural and binaural conditions in the combined forward–backward masking procedure. These results were viewed as indicating that the temporal (phase) and intensive information associated with temporally separated maskers and signals combine within the nervous system differently for binaural processing than for monaural processing.

*This research was supported by an individual research grant from the National Institute of Health (NIH), an NIH Program Project Grant, and an NIH training grant, Grants Nos. NS 06 459, NS 05 475, and 111 57-02.

Berg, K., and Yost, W. A. (1976). "Temporal Masking of a Click by Noise in Diotic and Dichotic Listening Conditions," J. Acoust. Soc. Am. 60, 173–177.

Blodgett, H. D., Jeffress, L. A. and Taylor, R. W. (1958). "Relation of Masked Thresholds to Signal Duration for Various Interaural Phase-Combinations," Am. J. Psychol. 71, 283–290.

Deatherage, B. H., and Evans, T. R. (1969). "Binaural Masking: Backward, Forward, and Simultaneous Effects," J. Acoust. Soc. Am. 46, 362–371.

Dolan, T. R., and Trahiotis, C. (1972). "Binaural Interaction in Backward Masking," Percept. Psychophys. 11, 92–94.

Green, D. M., and Yost, W. A. (1976). "Binaural Analysis," in *Handbook of Sensory Physiology*, (edited by X. Keidel and and Y. Neff (Springer– Verlag, Berlin), Vol. V(2).

Lakey, J. R. (1976). "Temporal masking-level differences: The effect of mask duration," J. Acoust. Soc. Am. 59, 1434–1433.

Levitt, H. (1971). "Transformed Up-Down Methods in Psychoacoustics," J. Acoust. Soc. Am. 49, 467–477.

McFadden, D. (1966). "Masking-Level Differences with Continuous and with Burst Masking Noise," J. Acoust. Soc. Am. 40, 1414–1419.

Punch, J., and Carbart, R. (1973). "Influence of interaural phase on forward masking," J. Acoust. Soc. Am. 54, 897–904.

Robinson, C. E., and Pollack, I. (1973). "Interaction between forward and backward masking: A measure of the integrating period of the auditory system," J. Acoust. Soc. Am. 53, 1313–1316.

Robinson, D. E., and Trahiotis, C. (1972). "Effects of Signal Duration and Masker Duration on Detectability under Diotic and Dichotic Listening Conditions," Percept. Psychophys. 12, 333–334.

Trahiotis, D., Dolan, T. R., and Miller, T. H. (1972). "Effect of "Backward" Masker Fringe on the Detectability of Pulsed Diotic and Dichotic Tonal Signals," Percept. Psychophys. 12, 335–339.

Small, A. M., Boggess J., Klich, R., Kuehn, D., Thelin, J., and Wiley, T. (1972). "MLD's in Forward and Backward Masking," J. Acoust. Soc. Am. 51, 1365–1367 (L).

Wightman, F. (1973). "Binaural detection in forward and backward masking," J. Acoust. Soc. Am. 54, 309 (A).

Preliminary study of a low-frequency, formant-based speech code for the severly hearing impaired

Kaye Reeder, William J. Strong, and E. Paul Palmer

Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602
(Received 12 October 1976; revised 6 January 1977)

This correspondence describes a computer-derived low-frequency auditory speech code. The code employs only three low-frequency tones consisting of sine waves, or narrow-band noise, plus overall amplitude control. The frequencies of the tones are a function of the three highest-amplitude formants in the speech signal.. A report of preliminary tests of the ability of normal-hearing and hearing-impaired subjects to learn common words and to discriminate words of a diagnostic rhyme test is given.

PACS numbers: 43.70.Gr, 43.66.Ts, 43.70.Dn, 43.70.Jt.

INTRODUCTION

Simple amplification is not adequate for aiding all hearing-impaired persons—particularly those who are severely impaired. We subscribe to the view that wearable speech-coding reception aids hold the most promise for providing severely hearing-impaired users meaningful speech-related inputs. Coded speech can be presented via visual, tactile, or auditory modes. (See Refs. 1–5 for examples and discussion of various of these codes.) We have chosen an auditory code over visual or tactile codes because it is comparatively easy to implement and it frees the other senses to perform their normal functions. To be successful, low-frequency auditory codes must be sufficiently complete to allow the discrimination of significant distinctions in the language. We have addressed the problem of obtaining such a speech code, but the question of the utility of the code by severely hearing-impaired subjects is still open because of lack of extensive training and testing.

I. DESCRIPTION OF CODE AND EXPERIMENT

Formant frequencies and amplitudes and voiced–unvoiced information which have been used to synthesize intelligible speech[6] were chosen as the speech param-

TABLE I. Pure-tone audiograms for better ear of sensorineural hearing-impaired subjects. (Losses in dB below normal threshold [ISO].)

Subject	Frequency (Hz) 250	500	1000	2000	4000
SN1 (oral)	10	30	55	80	80
SN2 (oral)	65	80	90	90	90
SN3 (oral)	55	70	95	100	...
SN4 (oral)	60	85	100	95	100
SN5 (manual)	85	90	100	100	...
SN6 (manual)	85	90	100	110	...

eters to be coded. Tape-recorded speech was low-pass filtered at 4.5 kHz, digitized at 10 000 samples/sec, and stored on disk. A 25.6-msec sample was analyzed every 10 msec and the analysis parameters stored. A linear prediction method was used to obtain a smoothed spectrum, and the peaks picked by a second derivative method.[7] The frequencies of the three highest-amplitude peaks were chosen to represent the desired formants. Each formant frequency was divided by 4 to place it in a low-frequency range, and a frequency of 100 Hz was added to each to bring the lowest frequencies up where better time resolution and better coupling to the ear are obtained. This resulted in changing a nominal formant frequency range of 200–4000 Hz to a range of 150–1100 Hz. Three computer-generated sine waves then represented the three transposed formant frequencies which were smoothly interpolated from one 10-msec interval to the next. For unvoiced speech, each sine wave was modulated with low-pass-filtered noise. The voice–noise decision was made using zero crossings and slope changes. The overall amplitude of the three waves was varied according to the overall amplitude of the input speech signal, but the individual formant amplitudes were maintained in the fixed ratio $A_1:A_2:A_3=1.0:0.5:0.33$ because informal listening tests of the transposed code by normal-hearing subjects showed no apparent advantage gained by using a variable ratio between the amplitudes. The three signals were added and stored for later output via D/A converter and headphones. No effort was made to incorporate fundamental frequency in the code.

Preliminary tests were chosen to provide an indication of the possible utility of the transposed speech code, but at the same time avoiding extensive training with hearing-impaired subjects. Interactive testing was not possible because the computer system used to generate the code runs in several hundred times real time. The first test required normal-hearing subjects to learn 20 most-common, isolated, English words. The second test employed the 96 rhyming word pairs of Form IV of the Diagnostic Rhyme Test[8] which uses different initial consonants to test the presence or absence of six speech attributes: voicing, nasality, sustention, sibilation, graveness, and compactness. This test was administered to both normal-hearing and sensorineural subjects. (Table I shows the pure-tone audiograms for the better ear of the sensorineural subjects. Subjects SN5 and SN6 had no spoken language and the other four had poor spoken English.) To avoid the necessity of learning new sounds, the test was used in a form where two rhyming words were presented and the subject merely indicated whether the two were the same or different. A pair such as "feel-veal" was presented as one of four combinations (feel-feel, feel-veal, veal-feel, veal-veal) chosen randomly. The coded words were presented to the subjects on a random basis via headphones at a level of 80 dBA for normals and a most-comfortable level for sensorineurals. Only one token of each word for each talker was used.

II. RESULTS AND DISCUSSION

Six normal-hearing subjects were able to learn the coded 20 most-common words for a male talker to an intelligibility level of 85% in an average of 4.3 h per subject. Learning the same words produced by a second male talker required only 1.7 h, and for a female talker required only 1.8 h. Informal tests showed that the coded words are retained so that after a few hours training, then a month's absence, most of the words are remembered, and brief review brings the level of remembering up to previous levels. Results of the discrimination test appear in Table II. The speech codes for rhyming words produce differences that are mostly discriminated by normal-hearing subjects as shown in the second column. (We have included the intelligibility scores of uncoded speech by normal-hearing subjects in the first column for comparison. However, direct comparison is not entirely valid as one would normally expect discrimination to be better than intelligibility.)

Since the low-frequency code tested here is not recognizable as speech, it is not immediately obvious how much speech information has been lost in the coding process. (The 91% discrimination score for the code versus a 95% intelligibility score for uncoded speech gives some indication.) The coding process is most degrading for nasality and least degrading for voicing as can be seen by comparing the discrimination and intelligibility scores for normals. However, it is significant that coded words can be distinguished, learned, and remembered and that learning is at least partially transferable from speaker to speaker. Thus, it appears that a significant amount of speech information is incorporated in the code.

The ability of the hearing-impaired subjects to discriminate the speech attributes is much poorer than for the normals, as seen in column 3 of Table II. The reasons for the differences are not completely clear.

TABLE II. Intelligibility of speech and discriminability of speech code.

Speech attribute	Intelligibility[a] of speech by normals	Discriminability of code by normals	Discriminability of code by sensorineurals
Voicing	99	96	81
Nasality	99	89	65
Sibilation	98	92	65
Sustention	96	90	63
Compactness	95	90	58
Graveness	86	88	65
Overall	95	91	66

[a]Taken from Keeler *et al.* 1976.

TABLE III. Overall discriminability of speech code for individual hearing-impaired subjects over several trials. (When not all 96 word pairs were used in the test, the number of pairs used are shown in parentheses.)

Trial subject	1	2	3	4	5	6	7
SN1	38	60	86(15)	76(25)	74(50)	68	78
SN2	38	34(15)	44(15)	60[a]	62	...	...
SN3	66	66	74	...	...	...	...
SN4	56	...	...	...	...	...	...
SN5	56	48	...	...	...	...	...
SN6	60	74	78	...	...	...	...

[a]This score occurred after $1\frac{1}{2}$ h of training on the ten most-common words.

The fact that pure-tone and complex-sound frequency discriminations are poorer for the severely impaired is a probable explanation.[9] However, other factors might enter in as hearing-impaired people are not trained by experience to make careful sound discriminations; their ultimate capabilities might be better than is indicated here. Practice produces significant improvement as the sounds and the task become more familiar as can be seen in the data of Table III. It should be noted that the hearing-impaired subjects were not tested to their asymptotic limits. Considerably more training on the discrimination task might be expected to enhance the discrimination scores of the deaf subjects, as often happens in discrimination experiments.

Several significant tasks need to be undertaken. First, an attempt should be made to improve the code so that discrimination scores on the code are equal to intelligibility scores on uncoded speech for normals. Second, hearing-impaired subjects need to be tested to asymptotic limits. Third, a real-time speech coder needs to be used so that many tokens of each word can be used to avoid artifacts. Fourth, the code should be tailored and tested for individual users because the discrimination scores for the sensorineurals are very disparate compared to normals.

[1]L. D. Braida, N. I. Durlach, B. L. Hicks, C. M. Reed, and R. P. Lippmann, "Matching Speech to Residual Auditory Function. III. Review of Previous Research on Frequency Lowering" (unpublished).

[2]S. Englemann and R. Rosov, "Tactual Hearing Experiment with Deaf and Hearing Subjects," Exceptional Children 7, 243–253 (1975).

[3]M. H. Goldstein and R. E. Stark, "Modification of vocalizations of preschool deaf children by vibrotactile and visual displays," J. Acoust. Soc. Am. **59**, 1477–1481 (1976).

[4]F. A. Saunders, W. A. Hill, and C. A. Simpson, "Speech Perception Via the Tactile Mode," Conference Record 1976 IEEE Int. Conf. Acoust. Speech Signal Process. 594–497 (1976).

[5]W. J. Strong, "Speech Aids for the Profoundly/Severely Hearing-Impaired: Requirements, Overview and Projections," The Volta Review 77, 536–556 (1975).

[6]L. O. Keeler, G. L. Clement, W. J. Strong, and E. P. Palmer, "Two Preliminary Studies of the Intelligibility of Predictor Coefficient and Formant-Coded Speech," IEEE Trans. Acoust. Speech Signal Process. **ASSP-24**, 429–432 (1976).

[7]R. L. Christensen, W. J. Strong, and E. P. Palmer, "Comparison of Three Methods of Extracting Resonance Information from Predictor Coefficient Coded Speech," IEEE Trans. Acoust. Speech Signal Process. **ASSP-24**, 8–14 (1976).

[8]W. D. Voiers, A. D. Sharpley, and C. J. Hehmsoth, *Research on Diagnostic Evaluation of Speech Intelligibility* (Tracor Inc., Austin, TX, 1973).

[9]J. M. Pickett and E. S. Martin, "Some Comparative Measurements of Impaired Discrimination for Sound Spectral Differences," Am. Ann. Deaf 113, 259–267 (1968).

Erratum: "Small-vibration theory of the clarinet" [J. Acoust. Soc. Am. 35, 305–313 (1963)] and a discussion of air-column parameters

John Backus

University of Southern California, Los Angeles, California 90007

(Received 20 January 1977)

An error in an earlier article is corrected and its consequences assessed. Useful expressions for air-column parameters are given, expressed in terms of the Q of the column.

PACS numbers: 43.75.Ef, 43.20.Mv, 43.10.Vx, 99.10.+g

In an earlier article[1] there was developed a theory to account for the fact that the playing frequency of the clarinet is somewhat below the corresponding resonance frequency of the air column of the instrument. The purpose of this letter is to correct a small but fundamental error in the derivation in that paper so that it will not be perpetuated further in other work, and also to use the opportunity to give useful expressions for quantities of importance in calculating air-column resonances and impedances.

