Department of Veterans Affairs

Journal of Rehabilitation Research
and Development Vol. 30 No. 1, 1993
Pages 26–38

Intelligibility of frequency-lowered speech produced by a channel vocoder

Miles P. Posen, MS; Charlotte M. Reed, PhD; Louis D. Braida, PhD
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract—Frequency lowering is a form of signal processing designed to match speech to the residual auditory capacity of a listener with a high frequency hearing loss. A vocoder-based frequency-lowering system similar to one studied by Lippmann was evaluated in the present study. In this system, speech levels in high frequency bands modulated one-third-octave bands of noise at low frequencies, which were then added to unprocessed speech. Results obtained with this system indicated, in agreement with Lippmann, that processing improved the recognition of stop, fricative, and affricate consonants when the listening bandwidth was restricted to 800 Hz. However, results also showed that processing degraded the perception of nasals and semivowels, consonants not included in Lippmann's study. Based on these results, the frequency-lowering system was modified so as to suppress the processing whenever low frequency components dominated the input signal. High and low frequency energies of an input signal were measured continuously in the modified system, and the decision to process or to leave the signal unaltered was based on their relative levels. Results indicated that the modified system maintained the processing advantage for stops, fricatives, and affricates without degrading the perception of nasals and semivowels. The results of the present study also indicated that training is an important consideration when evaluating frequency-lowering systems.

Key words: *frequency-lowering systems, speech intelligibility, vocoder evaluation.*

Address all correspondence and requests for reprints to: Charlotte M. Reed, Room 36-751, Massachusetts Institute of Technology, 77 Commonwealth Avenue, Cambridge, MA 02139.

INTRODUCTION

One general class of approaches toward lowering of the speech spectrum for individuals with high frequency hearing impairment is based on detecting high frequency information and recoding it through the use of low frequency signals which take advantage of the residual hearing of the listener. Two methods that have been employed to achieve such signal processing include frequency transposition and channel vocoding. In schemes employing frequency transposition (1,2), information in a specified high frequency band is shifted downward using amplitude modulation or nonlinear distortion. In schemes employing channel vocoding (3,4,5,6), the high frequency speech content is analyzed by a bank of filters whose output envelopes are used to control the amplitude of signals from low frequency synthesis filters. Braida, et al. (7) reviewed studies of transposition or vocoding prior to 1979 and concluded that, in general, the benefits to speech reception with these lowering schemes were generally small and restricted to a narrow class of speech sounds. Among the reasons offered by Braida, et al. for this lack of success were inappropriate selection of both the frequency range used to analyze high frequency information and the form of the recoded signals, as well as the choice of the level of the recoded signals relative to the normal low frequency speech components. For example, in many of the vocoder systems evaluated in the past, the analysis filter outputs controlled the levels of low frequency sinewaves, which were the sole signal presented to

the listener (3,5,6). Not only do such systems fail to distinguish between voiced and unvoiced sounds, they also eliminate suprasegmental speech cues available in the low frequency speech range. Several recent studies employing either transposition or vocoding, however, have reported improved identification of fricative and affricate sounds over that obtained through low-pass filtering (8,9,10,11).

Velmans (9) described a transposer-based system in which high frequency information in the range of 4,000–8,000 Hz was shifted downward into the range of 0–4,000 Hz using a balanced modulator, and then combined with the linearly amplified signal. In evaluations of consonant reception by listeners with normal-hearing, conducted with or without frequency transposition, the speech signal was low-pass filtered to 900 Hz and combined with high-pass white noise. Consonant identification was superior for the transposer over low-pass filtering, both alone and in conjunction with speechreading (by roughly 13 percentage points without speechreading and 8 percentage points when speechreading was also available). Velmans and Marcuson (10) presented data which indicated that the device proved beneficial to impaired listeners whose residual hearing extended to 1,000 Hz (or beyond) but who had little or no residual hearing above 4,000 Hz. These listeners showed improvements in transposed consonant identification ranging from 11 to 30 percent, although identification of only those consonants with strong high frequency speech cues (/s, ʃ ,z, ʒ , tʃ ,dʒ,t/) was investigated. Further results obtained with a group of 25 students with hearing impairment in schools for the deaf (11) indicated that benefits for the transposition device were larger in children with severe to profound high frequency losses than in those with mild to moderate losses in this range. Again, evaluations were restricted to consonants with strong high frequency content.

In a vocoding system developed by Lippmann (8), speech sounds in the 1,000–8,000 Hz range were analyzed with a bank of band-pass filters whose outputs controlled the levels of low frequency bands of noise in the 400–800 Hz range. The noise signals were added to the original speech signal, which was low-pass filtered to 800 Hz. Preliminary discrimination tests conducted with a variety of choices for the analysis and synthesis filters led to the selection of a system with four analysis and four synthesis bands for further study. Specifically, this system consisted of four analysis bands (two of which were two-third-octave wide and two of which were one-octave wide) and four synthesis bands composed of four one-third-octave bands of noise whose center frequencies ranged from 400 to 800 Hz. Consonant identification tests using CVC syllables composed from 16 consonants (C) and 6 vowels (V) were conducted on 7 subjects with normal hearing for both frequency-lowered speech and linearly amplified speech with an 800-Hz bandwidth. Overall percent-correct identification increased by 10 percentage points from 36 percent correct with linear amplification to 46 percent with frequency lowering. For individual subjects, increases ranged from 4 to 14 percentage points.

Several factors may underlie the positive results observed by Velmans and Lippmann relative to those reported in other studies of transposition and vocoding for frequency lowering (7). First, in both studies the recoded signals were presented at levels that did not significantly interfere with normal low-frequency speech sounds. Second, the frequency range selected to analyze high frequency information and the specific form of the recoded low frequency signals may have contributed to improved performance.

The current study was concerned with extending the investigation of the effects of vocoder-based frequency-lowered systems on the reception of consonants. In Experiment 1, the performance of listeners with normal hearing was evaluated using a system modeled after that described by Lippmann (8). In Experiment 2, a modified system designed to improve upon results obtained in Experiment 1, was evaluated. The results of the current study are compared with those of other relevant studies, and theoretical predictions are presented for performance through the frequency-lowered system in combination with speechreading.

EXPERIMENT 1: INITIAL EVALUATION OF A VOCODER-BASED FREQUENCY-LOWERING SYSTEM

Method

System Description. A block diagram of the system used in Experiment 1, modeled after that of Lippmann (8), is shown in **Figure 1**. This system

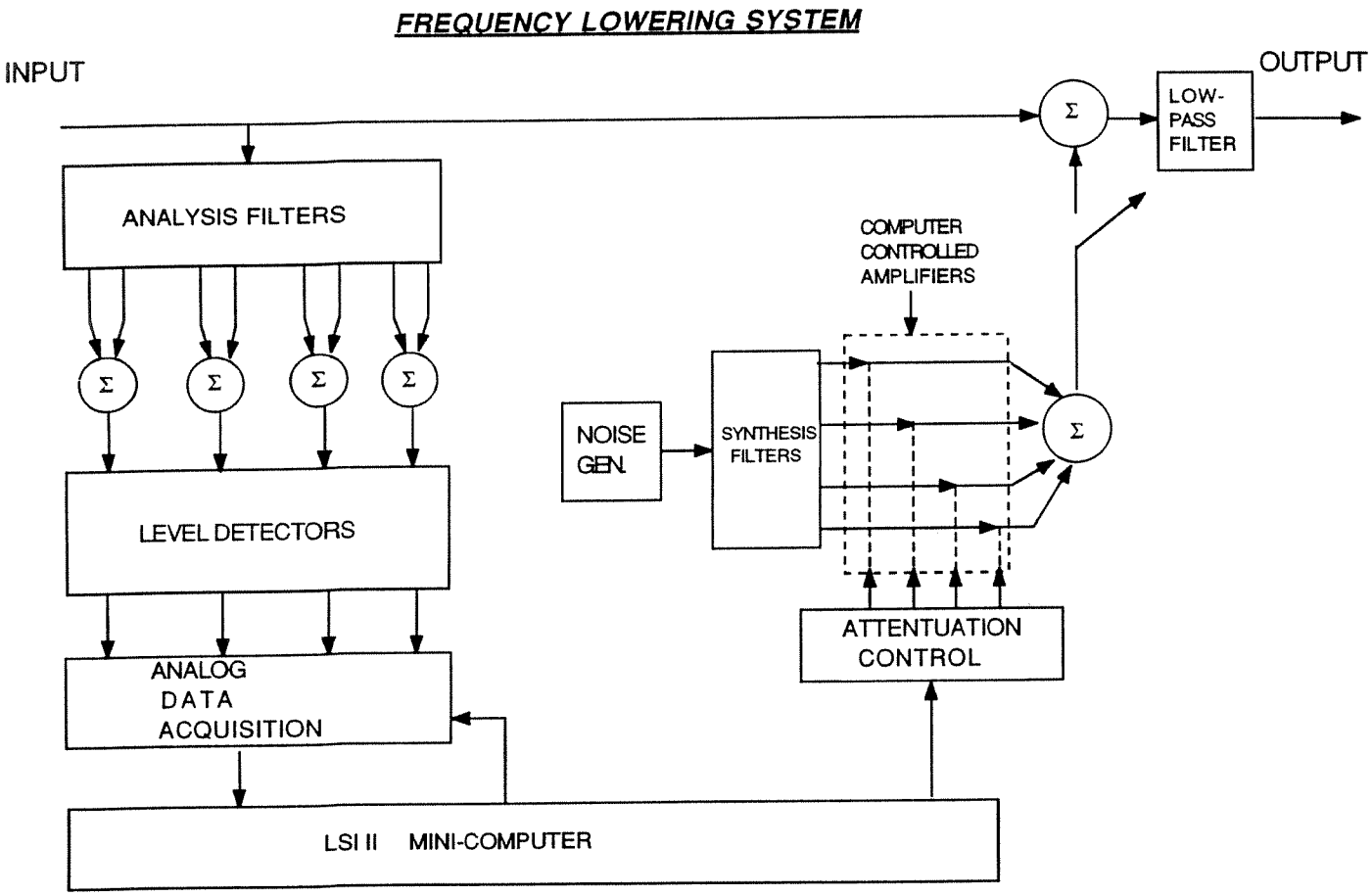

Figure 1.
Block diagram of the vocoder-based frequency-lowering system employed in Experiment 1.

made use of a computer-controlled multiband speech processor (12). High frequency speech information was analyzed by first passing speech through a bank of eight contiguous one-third-octave filters (General Radio 1925 Multifilter) with standard center frequencies in the range of 1.0 to 5.0 kHz (1.0, 1.25, 1.6, 2.0, 2.5, 3.15, 4.0, 5.0). The outputs of adjacent filters were combined using analog summing amplifiers to form four analysis bands with rejection rates of 48 dB/octave. The output levels of these bands were measured using rms level detectors that had logarithmic conversion, averaging times of 20 ms, and dynamic ranges of 65–70 dB. These levels were made available to an LSI-11 minicomputer by a multiplexed analog-to-digital converter, which had a 10 μs conversion time and a 12-bit capability. The sampling algorithm was based on that developed for controlling an amplitude compressor (12). The sampling periods ranged from 0.58 msec for the two lowest-frequency analysis bands to 0.14 msec for the highest-frequency band. The sample values were used to determine the output levels of low frequency narrow-band noise signals, which were summed and added to the original speech signal. These noise-band signals were generated by passing wide-band noise through four contiguous one-third-octave filters (GR 1925) whose center frequencies ranged from 400 to 800 Hz. The high frequency analysis bands and the low frequency synthesis filters were monotonically related in that the lowest analysis channel controlled the lowest synthesis band, the second-lowest analysis channel controlled the second-lowest synthesis band, and so on. The levels of the noise bands were controlled by computer-controlled amplifiers that could vary gain with 0.5 dB resolution independently for each noise band. The output level of a noise band was linearly related to the output level of its analysis band in that a 1 dB increase in the signal level in an analysis band caused a 1 dB increase in the level of the correspond-

ing low frequency noise-band signal. In addition, the levels of the noise bands were adjusted so that normal low frequency speech sounds were not significantly masked. Based on Lippmann's design as well as on subjective impressions from informal listening, the system was adjusted such that the 10 percent cumulative level of each noise band was 12 dB below the 10 percent cumulative level of speech measured using the same one-third-octave filter used to generate the noise signal.

The original speech signal combined with the noise signal was passed through two cascaded TTE miniature low-pass filters (series number J87E) to simulate a sharply sloping high frequency loss. The overall filter characteristics included 300 dB/octave rolloff, cutoff frequency of 800 Hz, and at least 80 dB attenuation in the stop band. These characteristics were chosen to provide an idealization of a hearing loss in which no high frequency speech cues are available. A toggle switch was included in the system to allow transmission of either unprocessed low-pass filtered speech or low-pass filtered speech with added low frequency noise-band signals.

Materials. The speech materials used for testing and training consisted of consonant-vowel (CV) nonsense syllables. The syllables were composed of 24 English consonants (stops /p t k b d g/; fricatives /f θ s ʃ v ð z ʒ /; affricates / tʃ dʒ /; semivowels /hw w l r j/; nasals /m n/; whisper /h/) with 3 vowels (/ɑ i u/), resulting in a total of 72 syllables. Each syllable was spoken 3 times by each of 4 speakers, 2 male and 2 female, resulting in 864 tokens. Recordings were made in an anechoic chamber with the microphone situated 6 inches in front of the speaker's mouth. The recorded syllables were passed through a 4.5 kHz anti-aliasing filter with 140 dB/octave rolloff and converted to 12-bit digital samples at a sampling rate of 10 kHz. The digitized waveforms were normalized to equal rms levels and stored on a large disk memory. The waveforms were thus accessible for automatic computer presentation.

The stimuli were divided into two groups: a test set and a training set. Four tokens of each CV syllable were included in the test set used before and after training, one token spoken by each speaker. The materials used for training consisted of the remaining 576 tokens (2 tokens per speaker).

Subjects. The subjects were two normal-hearing students in their early 20s, one male (MP) and one female (JR). The subjects had no more than 10 dB HL in the audiometric frequencies and were both native speakers of English with standard dialect.

Procedure. The subjects were trained and tested on the identification of frequency-lowered speech and low-pass filtered speech. On each trial a stimulus was selected at random from the available set and played over a high-quality audio-output channel. The signal was then passed through the processing system before being presented monaurally through TDH-39 headphones at a level of roughly 88 dB SPL.

For each pre- and posttraining test, each token from the test-token stimulus set was presented a total of 4 times, resulting in a total of 1,152 trials. Correct-answer feedback was not provided for either test. Training was achieved using the training-token stimulus set and stimulus-response procedure described above with correct-answer feedback. During a training session, each token from the stimulus set was presented twice, resulting in a total of 1,152 trials. When a subject responded incorrectly, the correct response was displayed visually and repeated aurally three times. Training sessions were continued until the learning curves of the subjects appeared to level off. The criterion for asymptotic performance was the observance of at least three consecutive training scores that were within three percentage points of each other. Each training session lasted approximately 2 hours. Subject MP began testing and training on frequency-lowered speech, while subject JR began the task with low-pass filtered speech.

Data Analysis. Confusion matrices were constructed from the pretraining and posttraining experimental runs for each subject under each experimental condition. The matrices were analyzed through calculations of percent-correct performance on various subsets of stimuli as well as through calculations of the percentage of unconditional information transfer (13) for the features described in **Table 1**, which were derived from definitions provided by Miller and Nicely (13) and Chomsky and Halle (14).

Results

Learning curves for each subject under low-pass filtering (circles) and the original vocoder-based lowering system (squares) are available in **Figure 2**. (Results from a third system, described in Experi-

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

Table 1.
Classification of the 24 consonants on a list of eight features derived from Miller and Nicely (13) and Chomsky and Halle (14).

	p	t	k	b	d	g	f	θ	s	ʃ	v	ð	z	ʒ	tʃ	dʒ	h	m	n	hw	w	r	j	l
Voicing	−	−	−	+	+	+	−	−	−	−	+	+	+	+	−	+	−	+	+	−	+	+	+	+
Nasal	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	−
Fricative	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	−	−	+	−	−	−	−
Affricate	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−
Round	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−
Vocalic	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+
Duration	−	−	−	−	−	−	−	−	+	+	−	−	+	+	−	−	−	−	−	−	−	−	−	−
Place	1	2	3	1	2	3	1	2	2	3	1	2	2	3	3	3	3	1	2	3	1	2	3	2

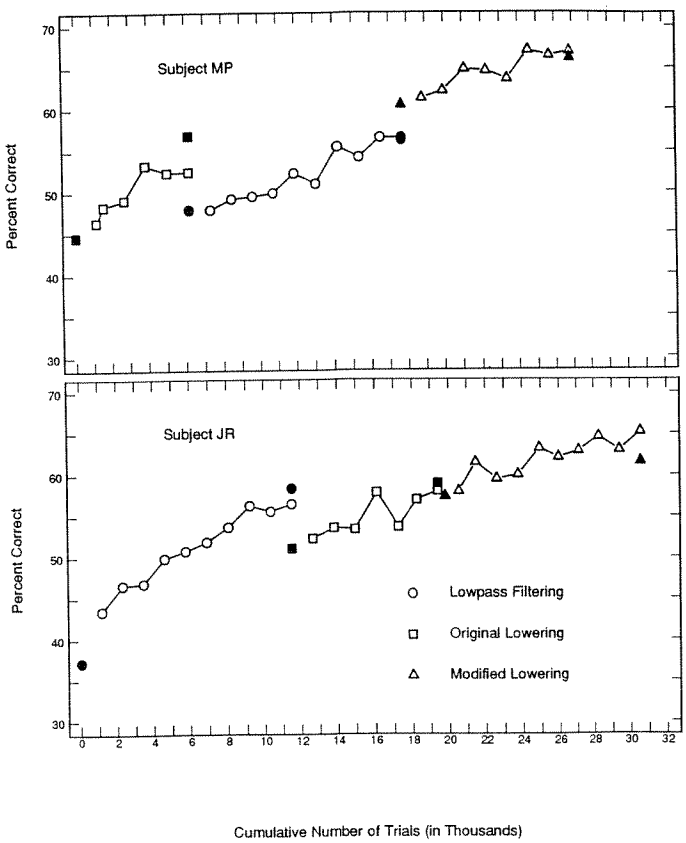

Figure 2.
Learning curves for each of three experimental conditions (low-pass filtering, original frequency-lowering system, and modified lowering system) for subject MP (top panel) and subject JR (bottom panel). Percent-correct score is plotted as a function of cumulative number of trials with feedback; filled symbols represent scores from test sessions and unfilled symbols, training sessions.

ment 2 below, are also included in the figure and will be discussed later in this article.) Filled symbols indicate results of pre- and posttraining test sessions (conducted without correct-answer feedback), while unfilled symbols indicate results from training sessions (in which trial-by-trial correct-answer feedback was provided). Posttraining test results of consonant identification using low-pass filtering and frequency lowering for subject MP showed gains of 8 and 12 percentage points, respectively, when compared with the pretraining test results. Similarly, subject JR showed improvements of 21 and 7 percentage points from training on low-pass filtered speech and frequency-lowered speech, respectively. Each subject achieved greater improvements with training for the system on which testing and training was begun (frequency lowering for subject MP and low-pass filtering for subject JR). Asymptotic performance on a given condition, however, appears not to depend on initial baseline performance: for example, both subjects leveled off at performance of roughly 56 percent correct for the lowering system, even though the pretraining score for subject MP was 47 percent compared with 37 percent for JR.

Posttraining scores obtained from each subject for these two conditions are shown in **Figure 3** for various groupings of consonants. The overall intelligibility of the full set of 24 consonants was not improved by the frequency-lowering system. Performance ranged from 56–58 percent correct across subjects and systems for the set of 24 consonants. When including only those consonants used by

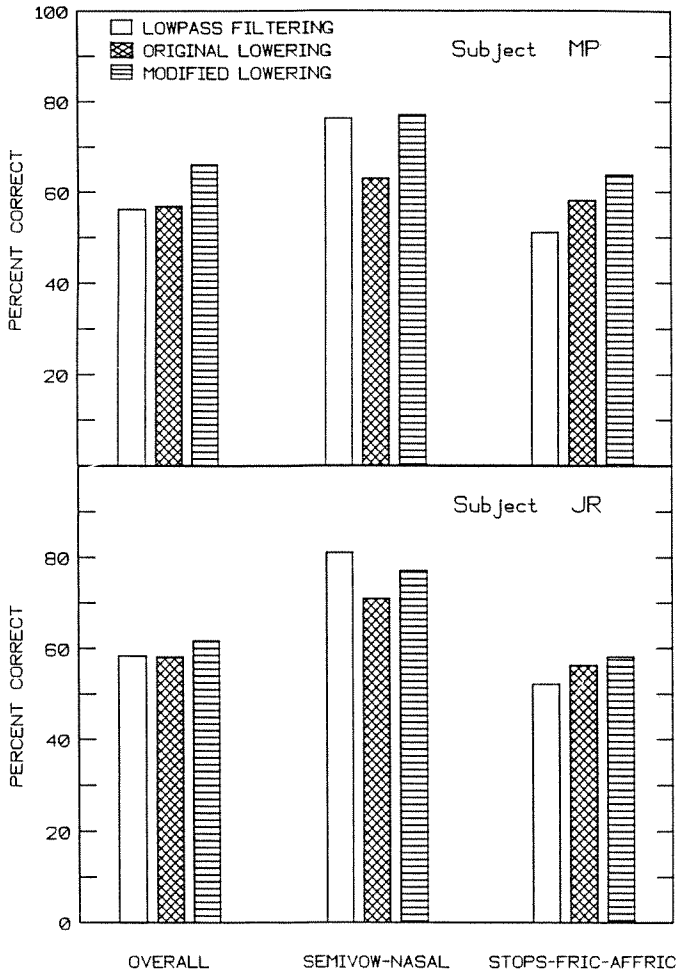

Figure 3.
Percent-correct score for each of three experimental conditions for subject MP (top panel) and subject JR (bottom panel). Scores are shown for the overall set of 24 consonants and for two subsets.

Lippmann (stops, fricatives, affricates), the percentage of consonants correctly identified by each subject was 4–7 percentage points higher with frequency lowering. This result is in close agreement with Lippmann's finding of a 4–14 percent range of improvement for seven normal-hearing listeners tested. When including only semivowels and nasals, correct identification scores for subject MP dropped 13 points from 76 percent with low-pass filtering to 63 percent with frequency lowering. Similarly, the scores of subject JR dropped 10 points, from 81 percent to 71 percent. It is apparent from the confusion matrices that processing degraded the intelligibility of nasals and (particularly) semivowels while improving the intelligibility of other sounds, in particular the fricatives and affricates. This result is confirmed by the analysis of unconditional informational transfer on individual features (displayed in **Figure 4**), which indicates that the features, affrication and duration, were better perceived under frequency lowering, while nasality and round were better perceived under low-pass filtering.

Identification of the vowels (/*a*,i,u/) was unaffected by the processing system. Each subject correctly identified 89–91 percent of the vowels presented with processing and low-pass filtering, and errors were based solely on confusions of /i/ and /u/.

EXPERIMENT 2: EVALUATIONS WITH A MODIFIED FREQUENCY-LOWERED SYSTEM

The results of Experiment 1 indicated that, while the frequency-lowering technique studied by Lippmann improved the recognition of stop, fricative, and affricate consonants, it also had the effect of degrading the perception of nasals and semivowels. Processing appeared to degrade consonantal signals characterized by low frequency energy. The processing of second and third formant vowel information may have interfered with the perception of low frequency cues of consonants in the CV syllables. Specifically, the lowered speech sounds of processed consonants may have been masked by the lowered speech sounds resulting from vowel processing. In an attempt to increase the intelligibility of frequency-lowered speech, the frequency-lowering system discussed in Experiment 1 was modified to suppress the processing when low frequency energy is predominant in the speech signal (as is the case for nasals, semivowels, and vowels) and to proceed with processing when high frequency energy predominates (as is the case for plosives, fricatives, and affricates).

Method

System Description. The modified system differs from that shown in **Figure 1** in that two additional signals were supplied to the minicomputer: high and low frequency energy of the input signal. These additional signals were formed by first passing speech through a bank of contiguous one-

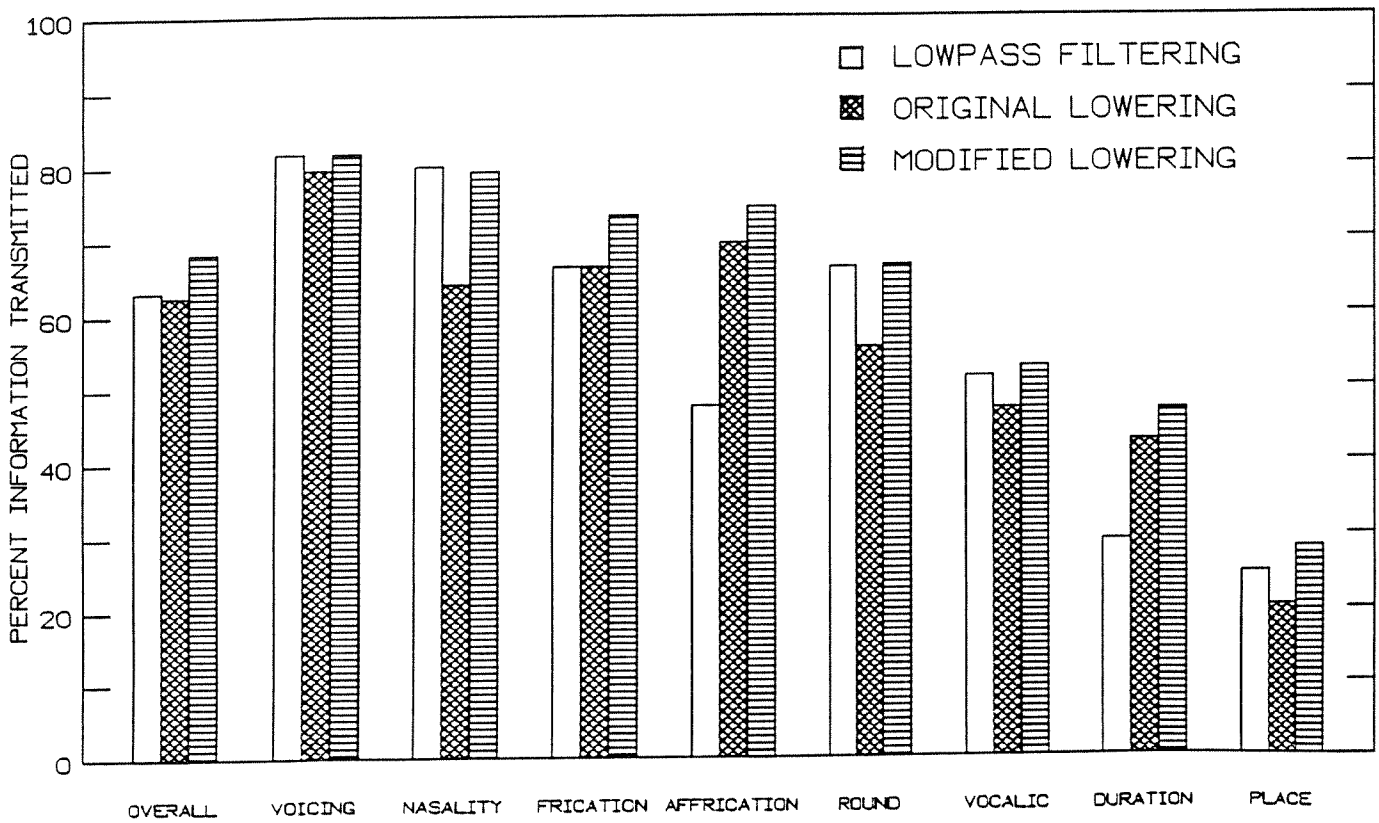

Figure 4.
Percentage of information transfer on the overall set of 24 consonants and on each of eight features for the three experimental conditions. Results were computed from confusion matrices combined across subjects MP and JR.

third octave filters with center frequencies of 125 to 5,000 Hz. The outputs of the filters with center frequencies of 125 to 1,250 Hz were combined using analog summing amplifiers to form the low frequency signal. Likewise, the high frequency signal was formed by summing the outputs of the filters with center frequencies of 1,600 to 5,000 Hz. (The selection of an appropriate corner frequency for defining low and high frequency components was based on samples of spectra of each consonant and vowel spoken by one male and one female talker.) The logarithms of the rms levels of the low and high frequency bands were determined, in addition to the output levels of the four analysis bands.

The low frequency noise was added to the original speech signal only when the power in the high frequency band exceeded that in the low frequency band less 3 dB. This threshold value was determined from subjective listening by choosing a value for which processing appeared to be suppressed for most vowels, nasals, and semivowels and activated for fricatives and affricates. Thus, when the low frequency power was greater than the high frequency power plus 3 dB, processing was suppressed by setting each variable gain amplifier to a maximum attenuation. Otherwise, processing proceeded in the same manner as that used in the preliminary experiment except each analysis-band level was sampled only once and each noise-band attenuation set accordingly. After all variable gain amplifiers were set (to maximum attenuation or attenuation values determined from corresponding analysis-band levels), the program repeated the above cycle. If processing was performed on each cycle, each analysis-band level was sampled every 0.35 ms. Otherwise, each analysis band was sampled every 0.21 ms. Sampling each band at the same rate, as opposed to sampling higher-frequency bands more frequently (as was done in the preliminary experiment), had negligible effects on the process-

ing. When processing occurred, the levels of the noise bands were controlled in the same manner as that described for the original system.

The same two subjects who participated in Experiment 1 were trained and tested on the identification of frequency-lowered speech produced by the modified system. Both subjects were aware of the modifications made to the system. The same stimulus sets and experimental procedure from the preliminary experiment were used. Results obtained were compared with the low-pass filtering results obtained previously and also to results obtained using the frequency-lowering system of Experiment 1.

Results

Learning curves for each subject for the modified lowering system (triangles) are shown, along with those for low-pass filtering (circles) and the original lowering system (squares), in **Figure 2**. Each subject's pretraining test score on the modified system was roughly equivalent to the posttraining score on the original system. Posttraining test results showed gains of 4 and 6 percentage points for subjects JR and MP respectively, when compared with the pretraining test results.

Posttraining test scores for various groupings of consonants are shown in **Figure 3** for each of the two subjects for each experimental condition. For subject MP, the overall percentage of consonants identified correctly using the modified system was 66 percent, an increase of 10 percentage points over low-pass filtering and an increase of 9 percentage points over frequency lowering using the original system. Subject JR identified 62 percent of the consonants correctly using the modified system, an increase of 4 percentage points over both systems of the preliminary experiment. Although we did not obtain repeated measures, the differences in scores between the two systems for each subject are larger than would be expected on the basis of Bernoulli fluctuations. At performance levels of roughly 60 percent correct and 1,000 trials, the standard deviation is 1.5 percentage points, and $2\sqrt{2}\sigma$ is approximately 4 percentage points.

As predicted, the perception of nasals and semivowels under the modified system was similar to that observed under low-pass filtering, due presumably to the selective processing achieved by the modified system. For subject MP, performance on nasals and semivowels was equivalent for the modified system and for low-pass filtering, and was superior to the performance on identification of these consonants for the original system (an improvement of 14 percentage points). Subject JR also perceived nasals and semivowels better using the modified system over the original system (a 6 percentage-point improvement) but still perceived these consonants better with low-pass filtering (her identification score was 4 percentage points higher with low-pass filtering than with the best frequency-lowering system). When including only stop, fricative, and affricate consonants, the percentage of consonants correctly identified by each subject was greatest for the modified system. The improvement in perception of these consonants using the modified system over the original system (6 percentage points for subject MP and 2 percentage points for subject JR) offers some support for the notion that processed vowel formant information in the original system may have interfered with the perception of these consonants.

The identification of vowels for each subject was similar in all three systems tested. Each subject correctly identified 90 percent of the vowels, using the modified system, compared with 89–91 percent with the other two systems.

To compare consonant confusions across the three systems, confusion matrices were generated by combining the post-training test results for both subjects for each system tested. The matrices appear in **Table 2** for low-pass filtering, **Table 3** for frequency lowering for the original system, and **Table 4** for frequency lowering for the modified system. Overall, performance on the original frequency-lowering system and low-pass filtering were equivalent (57 percent correct consonant identification). Performance on the modified system (64 percent correct) was 7 percentage points higher than on the other two systems (a significant difference based on Bernoulli statistics). A comparison of the percentage of unconditional information transfer on individual features across the three systems tested is shown in **Figure 4**. The perception of each feature related to nasals and semivowels (voicing, nasality, round, vocalic, place) under the modified system was as good as that obtained with low-pass filtering. The modified frequency-lowering system maintained the large improvement in the amount of information

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

Table 2.
Confusion matrix compiled across subjects for post-training data on the low-pass filtering condition.

	p	t	k	b	d	g	f	θ	s	ʃ	v	ð	z	ʒ	tʃ	dʒ	h	m	n	hw	w	r	j	l	
p	64	24	7	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	96
t	23	45	24	–	–	–	–	–	–	–	–	–	–	–	4	–	–	–	–	–	–	–	–	–	96
k	11	25	48	–	–	–	–	–	–	–	–	–	–	–	12	–	–	–	–	–	–	–	–	–	96
b	–	–	–	57	30	6	–	–	–	–	–	2	–	–	–	1	–	–	–	–	–	–	–	–	96
d	–	–	–	7	62	27	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	96
g	–	–	–	6	20	60	–	–	–	–	–	–	–	–	–	9	–	–	–	–	–	–	1	–	96
f	–	–	–	–	–	–	24	29	29	7	–	–	–	–	–	–	7	–	–	–	–	–	–	–	96
θ	1	–	–	–	–	–	5	50	22	4	7	1	–	–	2	4	–	–	–	–	–	–	–	–	96
s	–	–	–	–	–	–	8	26	45	13	1	–	–	–	1	–	2	–	–	–	–	–	–	–	96
ʃ	–	–	–	–	–	–	–	7	38	40	–	–	1	2	1	–	6	–	–	1	–	–	–	–	96
v	–	–	–	–	–	1	2	3	–	–	36	25	15	5	1	6	–	–	–	–	–	1	1	—	96
ð	–	–	–	3	7	2	1	8	–	–	13	28	13	2	–	6	–	–	–	–	–	6	7	–	96
z	–	–	–	–	–	–	–	–	–	–	13	30	22	25	–	2	–	–	–	–	–	–	4	–	96
ʒ	–	–	–	–	–	–	–	–	–	–	8	24	17	37	–	1	–	–	–	–	–	–	5	4	96
tʃ	–	9	10	–	–	–	–	–	–	1	–	–	–	–	73	1	2	–	–	–	–	–	–	–	96
dʒ	–	3	3	–	–	14	–	1	–	–	–	1	–	–	16	58	–	–	–	–	–	–	–	–	96
h	2	1	–	–	–	–	1	3	1	2	–	–	–	–	4	2	73	–	–	6	1	–	–	–	96
m	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	36	53	–	–	2	1	4	96
n	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	8	78	–	–	4	–	6	96
hw	–	–	–	–	–	–	–	–	–	1	–	–	–	–	7	–	4	–	–	83	1	–	–	–	96
w	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	2	76	8	8	2	96
r	–	–	–	–	–	–	–	–	–	–	4	3	1	6	–	3	–	–	–	3	7	53	11	5	96
j	–	–	–	–	–	–	–	–	–	–	–	2	1	1	–	1	–	–	–	–	3	–	88	–	96
l	–	–	–	–	–	–	–	–	–	–	1	1	–	–	–	–	–	5	3	–	–	–	–	86	96
	101	107	92	73	119	110	41	127	135	68	83	117	70	78	121	94	95	49	134	95	88	74	126	107	2304

Table 3.
Confusion matrix compiled across subjects for post-training data on the original lowering system.

	p	t	k	b	d	g	f	θ	s	ʃ	v	ð	z	ʒ	tʃ	dʒ	h	m	n	h	w	r	j	l	
p	48	16	25	–	–	1	–	2	1	–	–	–	–	–	–	–	3	–	–	–	–	–	–	–	96
t	28	55	9	–	–	–	–	–	–	–	–	–	–	–	3	1	–	–	–	–	–	–	–	–	96
k	30	34	24	–	–	–	–	–	–	–	–	–	–	–	6	1	1	–	–	–	–	–	–	–	96
b	–	–	–	74	12	9	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	96
d	–	–	–	13	56	27	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	96
g	–	–	–	16	9	68	–	–	–	–	–	2	–	–	–	–	–	–	–	–	–	–	1	–	96
f	–	1	–	–	–	–	28	15	26	3	–	–	–	–	–	–	21	–	–	2	–	–	–	–	96
θ	2	1	–	–	–	–	11	43	26	1	6	–	–	–	–	5	1	–	–	–	–	–	–	–	96
s	–	–	–	–	–	–	5	24	43	18	–	–	–	2	–	–	4	–	–	–	–	–	–	–	96
ʃ	–	–	–	–	–	–	–	1	18	72	–	–	–	3	1	–	1	–	–	–	–	–	–	–	96
v	–	–	–	–	1	–	8	–	–	–	41	29	4	5	–	–	1	–	–	2	–	3	2	–	96
ð	–	–	–	2	5	5	–	1	–	–	21	40	2	5	–	–	–	–	–	–	2	4	9	–	96
z	–	–	–	–	–	–	–	–	–	–	13	24	29	21	–	2	–	–	–	–	–	–	7	–	96
ʒ	–	–	–	–	–	–	–	–	1	–	7	10	15	60	–	1	–	–	–	–	–	–	2	–	96
tʃ	–	13	–	–	–	–	–	–	–	1	–	–	–	–	80	2	–	–	–	–	–	–	–	–	96
dʒ	–	3	–	–	2	6	–	–	–	–	–	–	–	–	12	73	–	–	–	–	–	–	–	–	96
h	2	–	–	–	–	–	3	4	1	1	1	–	–	–	–	–	71	–	–	8	4	1	–	–	96
m	–	–	–	–	–	–	–	–	–	1	–	2	–	–	–	–	–	35	47	–	1	2	–	8	96
n	–	–	–	–	–	–	–	–	–	–	1	3	–	–	–	–	–	15	54	–	–	3	2	18	96
hw	–	–	1	–	–	–	–	1	3	–	–	–	–	–	–	–	7	–	–	76	6	–	2	–	96
w	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	4	65	11	8	7	96
r	–	–	–	–	–	–	–	1	–	–	6	–	2	–	–	1	1	–	–	1	8	55	13	8	96
j	–	–	–	–	2	–	–	–	–	–	9	2	–	1	–	–	–	–	3	1	5	20	51	2	96
l	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	6	–	–	2	4	–	84	96
	110	123	59	105	87	116	55	92	119	97	106	113	52	97	102	86	111	56	104	94	93	103	97	127	2304

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

Table 4.
Confusion matrix compiled across subjects for posttraining data on the modified lowering system.

	p	t	k	b	d	g	f	θ	s	ʃ	v	ð	z	ʒ	tʃ	dʒ	h	m	n	hw	w	r	j	l	
p	56	8	25	–	–	–	–	–	–	–	–	–	–	–	–	5	2	–	–	–	–	–	–	–	96
t	11	63	21	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	96
k	24	15	53	–	–	–	–	–	–	–	–	–	–	–	4	–	–	–	–	–	–	–	–	–	96
b	–	–	–	71	16	9	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	96
d	–	–	–	4	42	50	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	96
g	–	–	–	13	11	67	1	–	–	–	–	1	–	–	–	1	–	–	–	–	–	–	2	–	96
f	–	1	–	–	–	–	35	10	31	4	–	–	–	–	–	–	15	–	–	–	–	–	–	–	96
θ	–	–	–	–	1	1	12	38	29	2	4	1	–	–	–	6	2	–	–	–	–	–	–	–	96
s	–	–	–	–	–	–	3	12	53	25	–	–	1	1	1	–	–	–	–	–	–	–	–	–	96
ʃ	–	–	–	–	–	–	–	–	10	86	–	–	–	–	–	–	–	–	–	–	–	–	–	–	96
v	–	–	–	1	–	–	7	1	–	–	41	27	10	–	–	3	2	–	–	–	–	2	1	1	96
ð	–	–	–	3	6	5	1	4	–	–	27	26	9	1	–	–	–	–	–	–	–	3	11	–	96
z	–	–	–	–	–	–	–	–	–	–	13	24	38	15	–	–	–	–	–	–	–	–	6	–	96
ʒ	–	–	–	–	–	–	–	–	–	2	–	8	16	57	–	13	–	–	–	–	–	–	–	–	96
tʃ	–	6	1	–	–	–	–	–	–	1	–	–	–	–	88	–	–	–	–	–	–	–	–	–	96
dʒ	–	2	–	–	–	1	–	–	–	–	–	–	–	1	10	82	–	–	–	–	–	–	–	–	96
h	1	–	1	–	–	–	8	–	1	1	–	–	–	–	–	1	81	–	–	2	–	–	–	–	96
m	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	42	49	–	–	3	1	1	96
n	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	12	71	–	–	3	–	9	96
hw	3	–	–	–	–	–	2	–	–	–	–	–	–	–	–	–	3	–	–	84	4	–	–	–	96
w	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	4	67	19	6	–	96
r	–	–	–	–	–	–	1	–	–	–	4	4	3	1	–	–	–	–	–	3	9	59	9	3	96
j	–	–	–	–	–	–	–	–	–	–	2	1	1	–	–	–	–	–	3	–	1	2	86	–	96
l	–	–	–	–	–	–	–	–	–	–	–	4	–	–	–	–	–	4	1	–	–	1	–	86	96
	95	95	101	92	76	133	70	65	124	121	91	97	78	76	104	111	105	58	124	93	81	92	122	100	2304

transferred under the original lowering system for the affrication and duration features and also showed a substantial gain in the amount of information transferred for the frication feature. Thus, the modified frequency-lowering system managed to maintain a significant processing advantage for fricatives and affricates without degrading the perception of the nasals and semivowels.

DISCUSSION

The modified frequency-lowering system performed significantly better than the Lippmann-based system evaluated in the preliminary experiment. Each subject found the nasals and semivowels to be more intelligible with the modified system, in addition to a number of the stop, fricative, and affricate consonants, as predicted earlier. Compared with low-pass filtering, the modified system performed equally well in handling nasals and semivowels, while improving the intelligibility of stop, fricative, and affricate consonants.

The effects of training on the reception of lowered speech were also studied by Reed, et al. (15) for frequency lowering accomplished through a pitch-invariant non-uniform compression of the

short-term spectral envelope (16). Using a training procedure similar to the one employed here, consonant identification was studied with a major subset of the same CV productions utilized for the present study. For speech lowered to bandwidths ranging from 1,000-1,250 Hz, posttraining scores of 60-67 percent correct were observed for two listeners with normal-hearing and one listener with high frequency sensorineural loss. The number of trials required to achieve asymptotic performance for the spectrally compressed speech, however, was nearly double that required for the vocoder-based processing. Reed, et al. (17) reported on the discriminability of consonants processed by the frequency-lowering system used by Hicks. Their results were similar to the results of the present study in that overall performance under frequency lowering was roughly comparable to that with low-pass filtering, but the perception of various consonants was different in the two systems tested; the perception of nasals and semivowels was better for filtering than for lowering, while the perception of fricatives and affricates was superior under lowering conditions than for low-pass filtering.

In the current study, evaluations of performance were limited to auditory presentation of low-pass-filtered or frequency-lowered speech alone, and did not include the more realistic condition of supplementing the auditory stimulus with lipreading information. An analysis of the confusions for the modified vocoder-based lowering system indicates that a high percentage of the errors (>75 percent) stems from place confusions within a given class of sounds. For example, a high error rate is observed among the three voiced stops as well as among the three unvoiced stops. It is highly likely that such confusions may be resolved through the addition of speechreading, for which place of articulation is better perceived than voicing or manner of articulation.

Braida (18) describes a model of audiovisual integration that can be used to estimate the identification scores that would be obtained in conjunction with speechreading for each type of processing. In the "prelabeling" model, identification in each modality is described in terms of a multidimensional Thurstonian process, and audiovisual identification is assumed to reflect the orthogonal composition of auditory and visual cues. Predicted audiovisual scores were derived by fitting auditory and visual confusion matrices using separate three-dimensional cue-spaces, and computing the highest value of accuracy corresponding to the six-dimensional audiovisual cue space. Predictions were derived for visual confusions reported by Busacco (19)—20 consonants, /a/ context—and Owens and Blazek (20)—23 consonants, /a/ and /u/ contexts, hearing-impaired listeners—combined with auditory confusions from each of the three systems tested in the present study. As expected, the predicted audiovisual scores were higher than both auditory and visual scores (roughly 95 percent correct for the /a/ context, 85 percent for /u/), and the range of scores was smaller than in the auditory case. Predicted AV scores for low-pass filtered materials were roughly equal to those for the unmodified processing. A small but significant advantage was observed for the modified processing, particularly under the /u/ context where predicted AV scores were four points higher than for low-pass filtering. These results suggest that the improvements in auditory performance obtained with the modified processing are likely to be seen in audiovisual consonant recognition tests, although the size of the improvement may be reduced.

SUMMARY AND CONCLUSIONS

The following points summarize the major findings of the present study:

1. Results obtained on two subjects with normal-hearing with a vocoder-based frequency-lowering system showed that the intelligibility of fricative, stop, and affricate consonants was improved by an average of 6 percentage points over low-pass filtering, while the intelligibility of semivowels and nasals was degraded by 12 percentage points on the average. The present study replicated results obtained by Lippmann (8), who used only fricative, stop, and affricate sounds and found an average improvement of 10 percentage points over low-pass filtering. In addition, the preliminary study indicated a degradation in performance for vowel-like consonants.
2. Based on the results obtained with the original frequency-lowering system, this system was modified so as to suppress the processing whenever low frequency components dominated the input signal. In this manner, it was

hoped that the modified system would maintain the processing advantage for stops, fricatives, and affricates while not degrading semivowels and nasals. This modification was achieved by measuring and comparing two additional signals, the high and low frequency energies of the input signal. When the low frequency energy exceeded the high frequency energy by a specified threshold amount, the processing was suppressed; otherwise, the system performed as before. Results indicated that the modified system further improved the intelligibility of stop, fricative, and affricate consonants by an average of 9 percentage points over low-pass filtering. In addition, the intelligibility of semivowels and nasals was similar for processing (with the modified system) and filtering.

3. The ability to use spectral and temporal high frequency cues available in the lowered speech was facilitated by training. Each subject improved his or her ability to perceive lowered speech with training and required an average of 15 hours of training to reach a stable performance level for each system tested.

ACKNOWLEDGMENT

This research was supported by Grant No. RO1 DC00117 from the National Institutes of Health.

REFERENCES

1. Johansson B. The use of the transposer for the management of the deaf child. Internat Audiol 1966;5:362-73.
2. Risberg A. The transposer and a model for speech perception. STL-QPSR 4:26-30, Stockholm, 1965.
3. Ling D, Doehring DG. Learning limits of deaf children for coded speech. J Speech Hear Res 1969;12:83-94.
4. Ling D, Druz WS. Transposition of high-frequency sounds by partial vocoding of the speech spectrum: its use by deaf children. J Aud Res 1967;7:133-44.
5. Pimonow L. Technical and physiological problems in the application of synthetic speech to aural rehabilitation. Am Ann Deaf 1968;113:275-94.
6. Takefuta Y, Swigart E. Intelligibility of speech signals spectrally compressed by a sampling-synthesizer technique. IEEE Trans Aud Electroacoust 1968;AU-16:271-4.
7. Braida LD, Durlach NI, Lippmann RP, Hicks BL, Rabinowitz WM, Reed CM. Hearing aids: a review of past research on linear amplification, amplitude compression, and frequency lowering. ASHA Monograph No. 19. Rockville (MD): American-Speech and Hearing Association, 1979.
8. Lippmann RP. Perception of frequency lowered consonants. J Acoust Soc Am 1980;67(1 Suppl):78.
9. Velmans ML. Speech imitation in simulated deafness, using visual cues and 'recoded' auditory information. Lang Speech 1973;16:224-36.
10. Velmans ML, Marcuson M. A speechlike frequency transposing hearing aid for the sensory-neural deaf. Report to the Department of Health and Social Security. London: The Department, Contract No. R/E 1049/84, 1980.
11. Velmans M, Marcuson M, Grant J, Kwiatkowski R, Rees R. The use of frequency transposition in the language acquisition of sensory-neural deaf children. Report of the Department of Psychology. London: Goldsmiths' College, University of London, 1988.
12. Coln MCW. A computer controlled multiband amplitude compression system [thesis]. Cambridge (MA): Massachusetts Institute of Technology, 1979.
13. Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. J Acoust Soc Am 1955;27:338-52.
14. Chomsky N, Halle M. The sound pattern of English. New York: Harper & Row, 1968.
15. Reed CM, Schultz KI, Braida LD, Durlach NI. Discrimination and identification of frequency-lowered speech in listeners with high-frequency hearing impairment. J Acoust Soc Am 1985;78:2139-41.
16. Hicks BL, Braida LD, Durlach NI. Pitch invariant frequency lowering with nonuniform spectral compression. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing; 1981; Atlanta (GA): 121-4.
17. Reed CM, Hicks BL, Braida LD, Durlach NI. Discrimination of speech processed by low-pass filtering and pitch-invariant frequency lowering. J Acoust Soc Am 1983;74:409-19.
18. Braida LD. Crossmodal integration in the identification of consonant segments. Quart J Exp Psych 1991;43A(3): 647-77.
19. Busacco DA. The effect of age on the benefit derived from visual cues in auditory-visual speech recognition by the hearing impaired [dissertation]. New York: Columbia Univ., 1988.
20. Owens E, Blazek B. Visemes observed by hearing-impaired and normal-hearing adult viewers. J Speech Hear Res 1985;28:381-93.