Experiments on the intelligibility of low-frequency speech codes

Donald R. Allen, William J. Strong, and E. Paul Palmer

Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602
(Received 31 March 1980; accepted for publication 21 July 1981)

The intelligibility of unprocessed and of low-pass filtered speech was compared to that of speech processed using three versions of an all-harmonic code consisting of many harmonic sinusoids, a largest harmonic code consisting of four harmonic sinusoids closest to the formants, and a formant code consisting of three sinusoids scaled from the formant frequencies. Fundamental frequency and formant frequencies were scaled down in frequency by different amounts in the various codes. Normal-hearing subjects were tested on three different types of tests. The Diagnostic Rhyme Test (DRT) was used on the two speech varieties and on codes that were not frequency lowered, a Diagnostic Discrimination Test (DDT) was used on frequency-lowered speech codes, and a prosodic test was run on all versions of the speech and speech codes. Results of each test are presented and compared for the various talker, speech, and speech-code combination; they show that the low-pass-filtered speech was always more intelligible than any low-frequency speech code tested.

PACS numbers: 43.70.Dn, 43.66.Sr, 43.66.Ts, 43.70.Ep

INTRODUCTION

Communication is a difficult task for the hundreds of thousands of people who are severely hearing-impaired. There have been many efforts to design specially coded speech information for the hearing-impaired because signing and speech reading (lip reading) are not adequate in many instances. The coded speech may be presented tactually, visually, or auditorally. (See Braida *et al.*, 1979; Levitt, 1973; Pickett, 1977; or Strong, 1975, for reviews of various schemes.) We have chosen to work with low-frequency auditory coding of speech because the use of hearing frees the tactual and visual senses to perform their normal functions. Auditory codes may also be useful in situations such as telephoning where visual communication is unavailable.

Several methods have been devised in attempts to recode speech within the usable frequency range of the hearing-impaired. A four-to-one reduction of the speech bandwidth is indicated under the assumption that speech energy above 4 kHz is relatively unimportant and because many severely hearing-impaired have some residual hearing out to 1 kHz. Early attempts at frequency lowering involved slow playback of recorded speech. When the speech was slowed to one-fourth speed (to produce a frequency compression of four to one), temporal cues were grossly distorted. This technique is not useful in real time since the signal is four times as long as the original. More recently, improvements have been made in slow-playback coding by using pitch-synchronous coding and by discarding three-fourths of the signal (Schreiner, 1977). This restores some temporal cues, permits the scheme to run in real time, and has been shown to be moderately effective even at this severe compression. Reed *et al.* (1978) have modified linear frequency-lowering by "warping" the spectrum so that the lower frequency components are lowered very little while the higher components (typically the noise components) are lowered much more. Experiments have shown that the warping improves the code and increases speech reception scores.

Another method of low-frequency coding is to overlay the original speech signal below 1000 Hz with the frequency-compressed portions above 1000 Hz. Thus the low frequencies, so important for vowels and prosodics, are preserved and the high-frequency information important for fricatives is available when needed (Braida, 1979; Guttman and Nelson, 1968; Risberg, 1965).

Two pilot studies of low-frequency codes produced unreasonably disparate results. Reeder *et al.* (1977) obtained discriminability scores of 91% for normal-hearing subjects using a three component nonharmonic formant code in which no attempt was made to preserve voice pitch. Stewart *et al.* (1977) obtained discriminability scores of 61% for normal-hearing subjects using a four-component, largest-harmonic code in which an attempt was made to preserve pitch information. The present study had three main objectives. First, the results of the pilot experiments were to be checked by using similar coding techniques and the same listener task. Second, three new coding schemes, which have more spectral information than the two above, were to be tested. One was similar to the largest harmonic code, but all harmonics were used instead of only four. A second was similar to that of Schreiner (1977) in which the fundamental frequency was scaled down by the same factor as the spectrum whereas this was not true with the other codes (see Sec. I below). A third was similar to that of Reed *et al.* (1978) in which the spectrum was warped instead of linearly lowered. Finally, the ability of frequency-lowered codes to transmit prosodic information was to be tested.

An ultimate goal of research on recoding is to see how well the codes work with impaired listeners. However, only normal-hearing subjects were used in this study; the assumption was made that the differences in hearing capability of the listeners could be overlooked so that the study could concentrate on the effects of the codes. The tacit assumption is made that if normal listeners cannot discriminate the code, then hearing-impaired listeners cannot do so either. This overlooks the possibility that experience with the

 0001-4966/81/111248-08$00.80

code might affect the results differently between normal and impaired listeners.

I. CODING TECHNIQUES

A linear prediction analysis scheme (Markel and Gray, 1976) was chosen because of its ease of implementation and relatively good preservation of speech intelligibility. To produce the code, natural speech was low-pass filtered at 4.5 kHz, digitized at a 10-kHz sampling rate, and stored on disk for further processing. The speech was analyzed through a 256-point Hamming window at 10-ms intervals. The analysis provided reflection coefficients, a voice–unvoiced decision, a gain factor, and fundamental frequency at each 10-ms interval and these were stored on disk. The linear prediction analysis and synthesis resulted in a degradation of the intelligibility of original speech from 95% to 87% which is still relatively high. This degradation is primarily related to the fixed analysis frame rate of 10 ms and the effects of windowing the signal which smear the transients and distort the spectrum. Errors occurring in analysis are propagated to the speech codes.

The speech codes were synthesized from the stored analysis data. For voiced speech, the harmonic frequencies were determined as multiples of the fundamental; the level of each harmonic was determined from a spectral envelope generated from the reflection coefficients and the gain factor. The harmonics were added with zero phase to produce one period of the output waveform as determined from the fundamental frequency. The analysis parameters were then linearly interpolated between the two adjacent frames to obtain new parameters for the next period of the output. For the extended pitch periods (those longer than the analysis interval of 10 ms) some information was lost since some frames were skipped.

For unvoiced speech, "noisy sinusoids" were generated by passing white noise through filters centered at multiples of 100 Hz and each having an 80-Hz bandwidth. The amplitudes for these noisy sinusoids were determined from the spectral envelope, and they were added with random phase to produce 10 ms of the desired noise signal. The speech codes were D/A converted to cassette tape without further processing.

Five different coding techniques were compared in this study, all using the synthesis technique described above. They differed in the way in which the components were chosen and in how many were used. Three of the codes used an all-harmonic spectrum and the others used a restricted spectrum.

The first all-harmonic code (AH) used a compression factor of four to one for the fundamental and the spectrum. Thus an original fundamental of 100 Hz was reduced to 25 Hz and a nominal 4-kHz spectral bandwidth was reduced to 1 kHz. The resulting signal has higher spectral density [compare Fig. 1(a) and (b)] and poorer time resolution compared to the original. A four-to-one reduction causes the coded signal to be four times as long as the original. To maintain the proper time relations three-fourths of the signal must be discarded

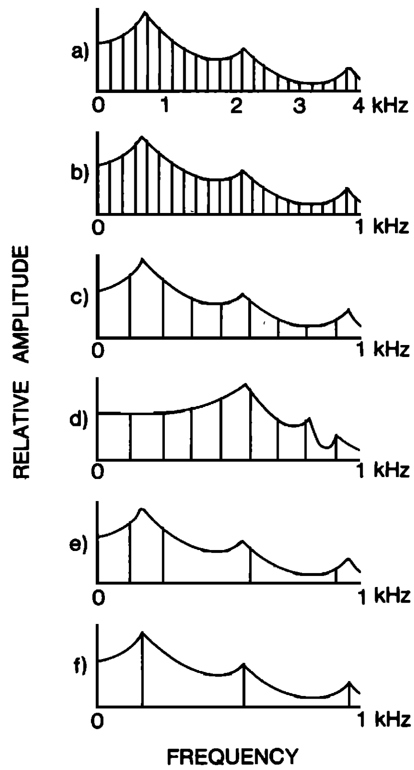

FIG. 1. Idealized spectra of the synthesized speech codes. (a) Original speech. (b) AH—all harmonic; fundamental scaled with spectrum. (c) AM—all harmonic, modified fundamental scaling. (d) AW—all harmonic; warped spectrum. (e) LM—largest harmonic. (f) FC—formant code. Note 4-kHz bandwidth in (a) but only 1 kHz in all others.

which causes a loss of some important transient information. An advantage of this type of coding is that it is fairly straightforward to implement in real time (Schreiner, 1977).

For the second all-harmonic code (all harmonic with modified fundamental, AM) the spectrum was compressed by a factor of four but the fundamental was compressed by a smaller factor according to the *ad hoc* formula

$$FN = 200 \times FO/[(200)^2 + (FO)^2]^{1/2}, \quad (1)$$

where FN is the new fundamental and FO is the original fundamental. For example, a new fundamental of 89 Hz results from an original of 100 Hz, while an original of 200 Hz becomes a fundamental of 141 Hz. This modification of the fundamental frequency resulted in better time resolution and lower spectral density [see Fig. 1(c)] than was the case in the AH code. The output waveform was generated by adding components corresponding to the harmonics of the modified fundamental with amplitudes which were determined from the compressed spectral envelope.

The third all-harmonic code (all harmonic with warped spectrum, AW) resulted from warping the spectrum. The fundamental was lowered as in the second all-harmonic code and then the amplitudes of other harmonics were chosen from the warped spectrum [see Fig. 1(d)]. A detailed description of the warping technique will be found in Picheny (1977).

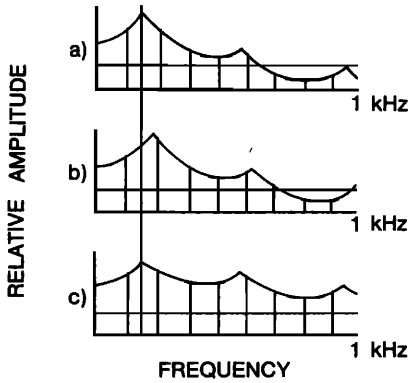

FIG. 2. Spectral effects of biasing and equalization using idealized spectra of AMT code (all-harmonic, modified-fundamental scaling, frequency lowered) (a) AMT, (b) AMTB, biased by 100 Hz, and (c) AMTL linear equalization. (Lines drawn on the graphs are references for comparison.)

The largest-harmonic code (LM) used only four of the largest components of the AM code. The two largest harmonics near the first formant were added to the largest harmonic near each of the second and third formants [see Fig. 1(e)]. Because of the harmonic relationship between the components, pitch was preserved. With only four components, the spectral density of the output is minimized. There was some difficulty in estimating the formants which, along with their movement, resulted in computer artifacts which degraded the signal. Artifacts came at a rate of one to five per second depending on speaker and the particular word spoken.

The formant code (FC) used three nonharmonic sinusoids with frequencies corresponding to the formant peaks [see Fig. 1(f)]. This was the easiest code to generate, but it was clearly the least speechlike since it did not preserve pitch. It was subject to the artifacts described for the LM code.

Frequency adjustments, or biasing, and amplitude equalization were used in an attempt to lessen masking effects in the low-frequency codes. The masking effects are primarily due to the upward spread of masking and are important in the frequency-lowered codes since the formants and the harmonics are closer together than in the original speech. A fixed-frequency bias (B) [compare unbiased in Fig. 2(a) and biased in Fig. 2(b)] moved the spectral envelope upward by 100 Hz which helped keep the important first formant in a usable frequency range. A variable bias (V) in which the biasing amount varied with the fundamental frequency was used optionally with the FC code in an attempt to restore pitch information. Otherwise pitch information is not available in the formant code since its components are nonharmonic. A linear equalization (L) scheme changed the original component levels to reduce the overall dynamic range and to increase the lower level components according to the relation

$$LN = 40 + LO/2\,, \qquad (2)$$

where LN is the new level in dB and LO is the original level [compare linear equalization in Fig. 2(c) and unequalized in Fig. 2(a)]. Thus an original sound level of 60 dB becomes 70 dB and 100 becomes 90.

II. TESTING PROCEDURE

A speech code must preserve the most salient aspects and qualities of speech if it is to be usable. Three different tests were used in this study to determine how well various speech attributes are preserved by the speech codes. Each test was given randomly to 10 listeners (11 in the case of the prosodic tests) and the results compiled. The coded speech was generated, stored on disk, and then D/A converted to cassette tape. The tests were administered in a relatively quiet environment over close fitting headphones (Koss 4AAA). The level was loud (approximately 95 dB), but not uncomfortably so, to partially simulate the effects that the coded speech would have for the hearing-impaired. All conditions used are summarized in Table I.

The Diagnostic Rhyme Test (DRT) (Voiers *et al.*, 1973) was run on the natural speech, speech low-pass filtered at 900 Hz, and on each of the codes which were not frequency compressed. The AH, AM, and AW codes are identical in the uncompressed case. Two talkers were used, a female with an average fundamental frequency of 220 Hz, and a male with an average fundamental frequency of 120 Hz. These tests (DRT) provided us with a control group from which to judge the effects of the coding methods.

A prosodic test was run on all versions of the codes to check the ability of the codes to convey pitch, stress, and rhythm in speech. Two adult males [with fundamental frequency averages of 96 (male 96) and 125 (male 125) Hz] and two adult females [with fundamen-

TABLE I. Summary of versions of speech and speech codes tested. Untransposed codes carry a two letter designation, transposed codes carry the two letter designation plus a T and any other suffix which might apply (e.g., AHTB is all harmonic, transposed, with fixed bias).

Speech and speech codes	
Natural	Speech A/D converted, D/A converted, recorded on audio tape.
LPF900	Speech low-pass-filtered at 900 Hz.
LPF700	Speech low-pass-filtered at 700 Hz.
LPC12	Analyzed and synthesized using 12 predictor coefficients.
AH	All harmonic, fundamental, and spectrum scaled the same.
AM	All harmonic, modified fundamental scaling.
AW	All harmonic, warped spectrum, modified fundamental scaling.
LM	Largest harmonic, modified fundamental scaling.
FC	Formant code.
Suffixes	
T	Frequency compressed (transposed) version.
B	Fixed bias of 100 Hz.
V	Variable bias scaled to fundamental.
L	Linear amplitude equalization.

tal frequency averages of 187 (female 187) and 200 (female 200) Hz] were used as the speakers in this test. Three factors (fundamental frequency, intensity, and duration) affect prosodics, and different talkers use them in differing amounts. Informal listening indicated that the male 96 seemed to use mostly fundamental frequency so we would expect lower scores on his formant codes which do not provide these cues. The female 187 seemed to use duration almost exclusively so we would expect higher scores on her voice for all the tests since they all convey duration fairly well. The other two speakers used a more uniform combination of the three factors.

The test consisted of two repetitions of each of the following five versions of the sentence "John drove to the store":

John drove to the store. (unstressed)
John drove to the store. (John stressed)
John *drove* to the store. (drove stressed)
John drove to the *store*. (store stressed)
John drove to the store? (question)

The sentences were presented randomly and the listeners were asked to determine which version was heard.

A Diagnostic Discrimination Test (DDT), using the same two talkers as the DRT, was run on the most promising codes as determined from the prosodic study. The coded DRT words were presented in an ABX format, with the listener asked to determine which of the first two words the last one resembled most. All four possible combinations of A and B (ABA, ABB, BAA, and BAB) were used to reduce any effects due to word order.

III. TEST RESULTS AND DISCUSSION

A. Intelligibility test

The DRT results for natural speech and the speech codes without frequency compression appear in Table II and Fig. 3. Looking first at overall scores, we see that all of the speech codes and speech low-pass filtered (65 dB per octave) at 900 Hz (LPF900) were significantly less intelligible than natural speech. The best code used in this study (AH) was somewhat better than the poorest code (FC) with the LPF900 speech midway between them. Speech from the same two talkers synthesized with 12 predictor coefficients (LPC12) (Smith *et al.*, 1981) was marginally better than the best code, but it was still significantly poorer than the natural speech. Clearly, speech information is lost in the coding process prior to frequency compression. All speech codes and LPF900 are within one standard deviation of each other.

Looking at the differences between the male and female talker, we see that there are some distinct intertalker differences. For the LM, FCV, and FC codes, and LPF900 speech, the speech produced by the female talker shows consistently lower scores than speech produced by the male talker for the speech feature compactness. That trend is reversed for the graveness feature on the FC and FCV codes where the male speech is more than 15% lower than the female speech. The presence of more harmonics seems to narrow the difference be-

TABLE II. Results using Diagnostic Rhyme Test (DRT) with untransposed speech and codes. Mean values and standard deviations of percent correct responses are given for individual speech features and talkers. M—male, F—female, Avg.—average for both talkers.

		Voicing	Nasality	Sustention	Sibilation	Graveness	Compactness	Overall
Natural	M	96.9 ± 4.2	98.7 ± 2.5	91.8 ± 4.9	97.5 ± 4.2	86.9 ± 4.4	98.7 ± 3.7	95.1 ± 2.2
	F	95.0 ± 7.3	100.0 ± 0	98.1 ± 4.0	95.6 ± 5.6	83.7 ± 8.0	99.4 ± 1.9	95.3 ± 2.2
	Avg.	96.0	99.4	95.0	96.6	85.3	99.1	95.2
LPC12 (Smith *et al.*, 1981)	M	93.7 ± 9.3	96.8 ± 3.3	83.1 ± 8.9	96.2 ± 4.4	66.8 ± 12.5	95.6 ± 5.2	88.7 ± 4.8
	F	90.6 ± 7.9	96.9 ± 3.3	70.6 ± 12.2	85.6 ± 5.9	79.4 ± 10.6	93.1 ± 5.5	86.0 ± 3.8
	Avg.	92.2	96.9	76.9	90.9	73.1	94.4	87.4
AH, AM, AW	M	88.1 ± 9.0	96.9 ± 5.8	63.1 ± 16.4	88.7 ± 6.7	72.5 ± 9.8	86.2 ± 7.8	82.6 ± 6.8
	F	86.9 ± 15.4	98.1 ± 4.0	75.6 ± 12.3	85.0 ± 11.9	72.5 ± 11.9	84.4 ± 8.9	83.8 ± 7.7
	Avg.	87.5	97.5	69.4	86.9	72.5	85.3	83.2
LM	M	85.6 ± 10.1	95.6 ± 4.9	55.0 ± 12.4	79.4 ± 19.6	67.5 ± 10.0	90.0 ± 7.0	78.9 ± 7.3
	F	88.1 ± 7.6	93.7 ± 4.0	78.1 ± 9.4	90.6 ± 6.4	73.7 ± 11.1	81.2 ± 8.4	84.3 ± 4.6
	Avg.	86.9	94.7	66.6	85.0	70.6	85.6	81.6
LPF900	M	90.6 ± 16.4	97.5 ± 7.5	78.1 ± 13.2	71.9 ± 12.9	50.6 ± 13.2	85.6 ± 11.5	79.1 ± 10.3
	F	91.8 ± 10.5	95.6 ± 6.3	78.7 ± 11.6	75.6 ± 10.2	54.4 ± 11.9	70.6 ± 12.2	77.8 ± 6.9
	Avg.	91.2	96.6	78.4	73.8	52.5	78.1	78.5
FCV	M	81.2 ± 9.7	88.7 ± 7.3	69.4 ± 13.8	77.5 ± 12.9	48.1 ± 11.5	86.9 ± 7.1	75.3 ± 5.7
	F	90.0 ± 12.9	81.8 ± 14.9	70.6 ± 11.2	80.6 ± 11.7	67.5 ± 8.3	74.4 ± 10.3	77.5 ± 7.9
	Avg.	85.6	85.3	70.0	79.1	57.8	80.7	76.4
FC	M	76.9 ± 10.8	88.7 ± 9.6	65.0 ± 15.1	70.6 ± 9.3	42.5 ± 13.6	86.9 ± 9.0	71.8 ± 6.7
	F	86.2 ± 13.6	76.2 ± 16.7	73.1 ± 16.3	78.1 ± 12.9	58.7 ± 7.5	73.7 ± 9.6	74.4 ± 9.9
	Avg.	81.6	82.5	69.1	74.4	50.6	80.3	73.1

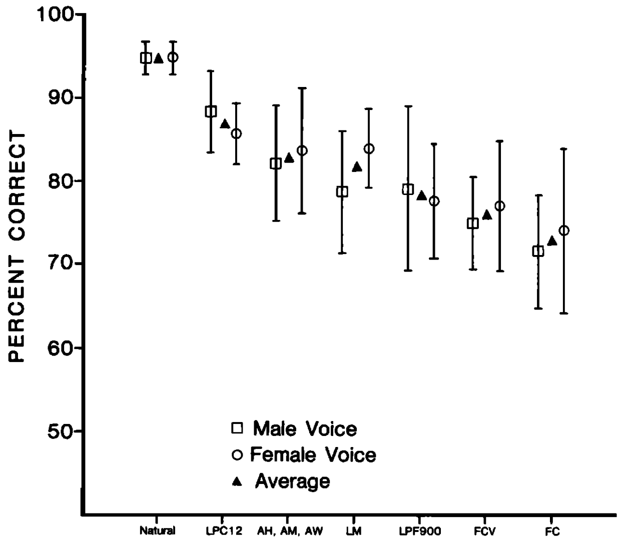

FIG. 3. Diagnostic Rhyme Test (DRT) intelligibility scores for various versions of speech and untransposed speech codes.

tween male and female as can be seen in the LM and AM codes. The sustention feature was more seriously degraded by the LM codes for the male, the scores being 23% lower than for the female; the male sustention score is only 92% in the natural speech so this may be a peculiarity of this voice. All of the other cases are within one standard deviation of each other so they can be considered approximately equivalent.

B. Prosodics test

The results of the prosodic study are summarized in Table III. Considering first the prosodic scores averaged over four talkers for natural speech, LPF900 speech, and untransposed codes, we note that low-pass filtering had no effect on prosodics as determined by this test. This seems reasonable because most prosodic information is carried by the fundamental frequency, intensity, and duration, only two of which are affected at all by filtering. The harmonic codes showed only minor degradation in the prosodic score but the formant codes showed substantial degradation even when variable bias was included to cue pitch.

Looking next at the transposed codes we note that the best version of an all-harmonic code (AMTBL) has a score of 88.4% which is as good as the untransposed scores. The best version of the largest-harmonic code (LMTL) at 84.6% is only slightly poorer than the best all-harmonic code. The best formant code (FCTV) was down 7% at 78.2%, but one version of the formant code had the worst overall score at 56.4%. The versions of the formant code with variable bias did considerably better than those without (compare FCTB and FCTV, or FCTBL and FCTVL). The all-harmonic codes in which the fundamental is lowered by a factor of four (AHT and AHTL) are 10% lower than their corresponding AM codes (AM and AMTL) with modified fundamentals. This indicates that high spectral density

TABLE III. Results of prosodic tests showing percent correct and standard deviation. Results for individual talkers and overall averages given. Average fundamental frequency for each talker is given for the natural-speech case.

Code	Male	Male	Female	Female	Overall average
Fundamental frequency	96	125	187	200	
Untransposed					
Natural	90.0	94.5	89.1	90.9	91.1 ± 10.5
LPF900	90.9	95.5	93.6	85.5	91.4 ± 9.9
AH, AM, AW	85.5	86.4	86.4	90.9	87.3 ± 11.2
LM	77.3	93.6	82.7	85.5	84.8 ± 12.2
FC	47.3	62.7	61.8	89.1	65.2 ± 21.3
FCV	46.4	62.7	65.5	80.0	63.6 ± 18.6
Transposed					
AMT	78.2	79.1	88.2	75.5	80.2 ± 14.0
AMTB	82.7	90.0	87.3	82.7	85.7 ± 13.7
AMTL	80.0	89.1	90.9	75.5	83.9 ± 14.8
AMTBL	96.4	90.9	82.7	83.6	88.4 ± 10.9
AHT	50.0	70.9	79.1	79.1	69.8 ± 18.0
AHTL	61.8	70.0	87.3	74.5	73.4 ± 15.4
LMT	71.8	81.8	87.3	64.5	76.4 ± 14.8
LMTB	78.2	87.3	85.5	78.2	82.3 ± 13.8
LMTL	80.9	96.4	92.7	68.2	84.6 ± 14.7
LMTBL	81.8	88.2	90.0	73.6	83.4 ± 14.5
FCT	36.4	58.2	57.3	73.6	56.4 ± 18.8
FCTB	36.4	67.3	61.8	73.6	59.8 ± 21.5
FCTV	59.1	81.8	88.2	83.6	78.2 ± 18.6
FCTBL	50.0	64.5	75.5	73.6	65.9 ± 16.1
FCTVL	49.1	78.2	88.2	81.8	74.3 ± 17.6

TABLE IV. Mean values and standard deviations of percent correct responses for the Diagnostic Discrimination Test (DDT) with results for individual speech features and talkers. M—male, F—female, Avg.—average for both talkers.

		Voicing	Nasality	Sustention	Sibilation	Graveness	Compactness	Overall
LPF900	M	85.9 ± 12.4	91.6 ± 9.3	71.5 ± 15.9	72.8 ± 12.5	58.7 ± 16.1	76.2 ± 13.0	76.1 ± 12.2
	F	97.2 ± 4.1	94.0 ± 6.8	75.6 ± 8.9	77.5 ± 12.2	54.4 ± 12.4	61.5 ± 11.7	76.7 ± 7.7
	Avg.	91.6	92.8	73.6	75.2	56.6	68.9	76.4
AWT	M	85.3 ± 5.5	88.8 ± 13.7	73.6 ± 10.7	62.9 ± 15.8	58.0 ± 11.5	76.3 ± 9.3	74.2 ± 9.7
	F	88.5 ± 13.3	72.4 ± 16.5	66.1 ± 19.7	71.9 ± 15.3	54.1 ± 12.3	61.5 ± 11.5	69.1 ± 14.1
	Avg.	86.9	80.6	69.9	67.4	56.1	68.9	71.7
LPF700 (Stewart *et al.*, 1977)	Avg.	94	86	57	56	36	65	66
AMTL	M	83.7 ± 11.2	79.1 ± 19.2	65.9 ± 14.1	69.1 ± 18.2	60.0 ± 17.0	75.0 ± 17.7	72.1 ± 15.2
	F	84.7 ± 12.5	50.9 ± 20.4	50.6 ± 18.3	65.6 ± 13.8	42.7 ± 17.6	41.6 ± 13.0	55.3 ± 14.2
	Avg.	84.2	65.0	58.3	67.4	51.4	58.3	63.7
FCTVL	M	76.9 ± 16.3	73.7 ± 16.1	52.5 ± 20.0	65.0 ± 20.3	54.7 ± 20.2	66.9 ± 18.0	64.9 ± 17.6
	F	85.9 ± 16.6	59.1 ± 21.2	57.8 ± 17.6	70.6 ± 13.6	42.8 ± 21.4	57.5 ± 15.2	62.3 ± 14.7
	Avg.	81.4	66.4	55.2	67.8	48.8	62.2	63.6
LMTBL	M	79.1 ± 19.3	69.1 ± 17.6	60.6 ± 17.5	60.9 ± 22.6	54.7 ± 14.4	75.6 ± 23.3	66.7 ± 17.7
	F	89.1 ± 15.1	49.7 ± 17.9	55.0 ± 18.3	71.9 ± 10.6	42.5 ± 13.9	52.5 ± 17.9	60.1 ± 13.9
	Avg.	84.1	59.4	57.8	66.4	48.6	64.1	63.4
AMTBL	M	79.4 ± 18.1	78.4 ± 21.2	68.4 ± 17.9	65.9 ± 21.1	55.3 ± 17.9	72.8 ± 14.4	70.0 ± 17.0
	F	82.8 ± 13.8	46.2 ± 19.7	50.9 ± 17.6	60.0 ± 19.9	33.4 ± 13.7	45.0 ± 18.9	53.1 ± 15.5
	Avg.	81.1	62.3	59.7	63.0	44.4	58.9	61.6
AHTL	M	78.4 ± 14.6	60.9 ± 17.6	52.8 ± 20.1	70.3 ± 17.8	53.7 ± 22.4	67.5 ± 17.2	64.0 ± 14.8
	F	85.0 ± 18.3	50.6 ± 20.7	53.1 ± 23.1	67.5 ± 17.5	45.8 ± 12.9	55.0 ± 18.4	58.8 ± 17.7
	Avg.	81.7	55.8	53.0	68.9	49.8	61.3	61.4
FCTBL	M	72.8 ± 16.4	67.2 ± 17.9	52.8 ± 17.4	62.2 ± 19.5	50.3 ± 14.9	65.9 ± 22.2	61.9 ± 16.9
	F	83.4 ± 13.3	51.9 ± 15.9	55.0 ± 16.1	67.8 ± 13.8	48.4 ± 16.7	54.7 ± 18.2	60.2 ± 13.2
	Avg.	78.1	59.6	53.9	65.0	49.4	60.3	61.1
AMTB	M	68.7 ± 20.9	74.3 ± 18.0	45.3 ± 25.9	63.7 ± 18.7	53.7 ± 23.0	60.9 ± 22.6	59.9 ± 21.3
	F	79.1 ± 23.9	46.6 ± 21.5	49.7 ± 17.4	61.5 ± 23.7	24.1 ± 14.0	44.0 ± 17.6	50.8 ± 18.2
	Avg.	73.9	60.5	47.5	62.6	38.9	52.5	55.4

and poor time resolution tend to obscure the prosodic cues. Perception of pitch changes is especially affected since the fundamental (which ranged from 18–33 Hz for the male 96 and 45–75 for the female 200) is so low.

Biasing provided small gains in most cases, but biased and unbiased versions were always within one standard deviation. The combination of biasing and equalization tends to provide an overall benefit beyond either effect alone.

Looking at specific speaker differences, we see that the scores for the female 187 are consistently higher than the average. The female 200 results seem to be fairly constant over all the techniques and do not exhibit the trends of the overall averages. The male 96 was particularly susceptible to prosodic errors, and the trend of low scores for the formant codes is clearly evident.

C. Discrimination test

To reduce the number of tests required of the subjects, some of the transposed codes in the prosodics test were eliminated before testing in the discrimination task. We chose the codes with biasing and linear equalization so that version of each of the four synthesis techniques was used; the AMTB and AMTL versions were used to verify the effects of biasing and equalizations. Finally, the LPF900 version was used in the discrimination test so that scores of the DRT and DDT could be compared to establish a standard between them. The same listeners were used on the DRT and DDT tests but only some were common with the prosodics listeners.

The results of the discrimination test are summarized in Table IV and Fig. 4. The first apparent result is the similarity between the totals of all the coding techniques excepting AWT and AMTB. This occurred in spite of some differences between the codes as evidenced by the DRT and prosodics tests. This result was surprising since several listeners commented on how different the various transposed codes sounded. The warped code (AWT) was 8% better than the best linearly lowered code and only 5% below low-pass-filtered (LPF900) speech. The filtered-speech condition of Stewart *et al.* (1977) (LPF700) has been included to indicate that our results are consistent with theirs; their filtering to 700 Hz has decreased the intelligibility by 10%. The warped lowering shows an improvement over linear lowering, although the standard deviations overlap. Biasing and equalization both improved the

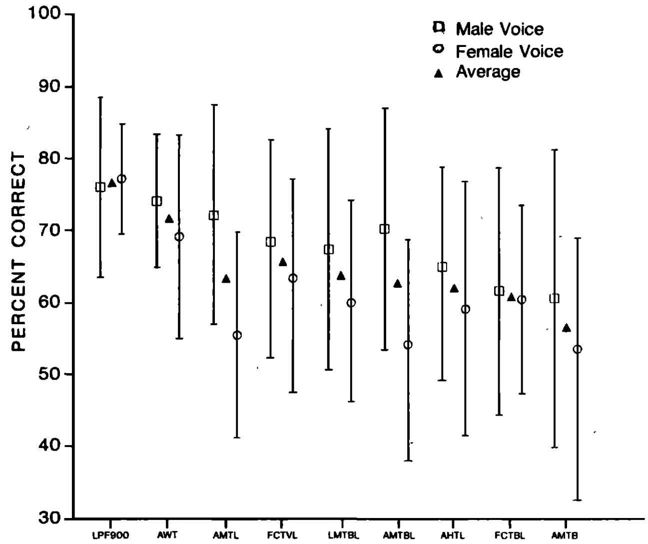

FIG. 4. Diagnostic Discrimination Test (DDT) discriminability scores for LPF900 speech and various transposed speech codes.

transposed code for prosodics, but biasing decreased the scores slightly (compare AMTL and AMTBL) whereas linear equalization raised the scores somewhat more (compare AMTB and AMTBL) for discrimination; equalization by itself is better than biasing and equalization together. For all traits, the worst code is at most 12% below the best linearly lowered code which is well within the standard deviations of this test.

If we look at the overall results of the tests on linear codes for each talker we see that the female has scores which are consistently below those of the male. The male's highest score is for AMTL with AMTBL, LMTBL, FCTVL, AHTL, and FCTBL following in order. For the female, the AM and FC codes exchange places, with FCTVL being the highest followed by FCTBL, LMTBL, AHTL, AMTL, and AMTBL. Two things might contribute to this difference. First, modifying the fundamental frequency may change the spectral density to the point where the formant peaks are ill defined for the female speech. The formant code has well defined formants even though the components are not harmonically related. The male speech has sufficient components to adequately specify the peaks and also benefit from the harmonic structure of the AM code. Second, the pitch is lowered more for the female voice than for the male which could cause perceptual errors. This indicates that more spectral density is required without losing timing cues as occurs for the AHT code.

These results and those of past studies using similar techniques appear in Table V. The Stewart *et al.* (1977) result of 61% compares well with the score of 63.4% for the LMTBL code used in this study. The Schreiner (1977) score of 63% is similar to the 61.4% for the AHTL code. (The Schreiner score for correct responses of 75% was corrected for guessing by subtracting half the score for incorrect responses of 25% because the test had three response alternatives.) Reed *et al.* (1978) obtained scores of 89% for her warped code but used a different listener task which resulted in scores of 90% for low-pass-filtered speech. This compares with the scores in this study of 71% for AWT and 76% for low-pass filtering. The biggest discrepancy occurs between the Reeder *et al.* (1977) study and this study. They obtained discrimination scores of 91% compared to 61% for the comparable FCTBL code. Although the coding technique and tests were similar, the Reeder *et al.* (1977) results were not replicated. Their testing method allowed four repetitions of each discrimination pair before a same or different judgment was required. Informal listening tests using multiple repetitions of the test pair words with a few of our subjects resulted in little change in scores. The possibility exists that the Reeder *et al.* subjects were inadvertently given cues during the testing process (Reeder, 1979).

IV. CONCLUSIONS

Although the results of this study are not encouraging there are a few important conclusions to be drawn. The nearly constant 63% results obtained on the discrimination tests for most of the linearly lowered speech codes compared with 76% for low-pass-filtered speech indicate that none of the coding techniques restored speech cues lost by bandlimiting the signals. Warping preser-

TABLE V. Comparison of results obtained in the present study with results of other studies. All results are corrected for guessing.

Researcher or code	Overall percent correct	Type of speech test
Stewart *et al.* (1977)	61	DDT on DRT words
LMTBL	63.4	DDT on DRT words
Schreiner (1977)	63	DDT on CVC and CV
AHTL	61.4	DDT on DRT words
Reed *et al.* (1978)	89	AB–BA on CV
Reed *et al.* (1978) LPF900	90	AB–BA on CV
AWT	71.7	DDT on DRT words
LPF900	76.4	DDT on DRT words
Reeder *et al.* (1977)	91	DDT–4 repetitions, on DRT words
FCTBL	61.1	DDT on DRT words

ved some speech cues lost by linear lowering, but in this study it still performed more poorly than low-pass-filtered speech. These low scores may indicate some basic perceptual limits of the ear for discrimination of signals bandlimited in this manner.

The prosodics study demonstrates that if a speech code is to convey information concerning stress and intonation, then considerable harmonic structure should be preserved. When the components are inharmonic (FC) or when only a few harmonics are included (LM), the performance of the listeners is reduced. Also it seems important to avoid a too high spectral density such as occurs when all the original harmonics are scaled down (AH).

Some further study is warranted of warped lowering schemes. The addition of equalization and improvement of the unvoiced portions of speech may make this code more discriminable than low-pass-filtered speech.

Braida, L. D., Durlach, N. I., Lippmann, R. P., Hicks, B. L., Reed, C. M., and Rabinowitz, W. M. (1979). "Hearing Aids—A Review of Past Research on Linear Amplification, Amplitude Compression, and Frequency Lowering," ASHA monogr. 19.

Guttman, N., and Nelson, J. R. (1968). "An Instrument that Creates some Artificial Speech Spectra for the Severely Hard of Hearing," Am. Ann. Deaf 113, 295–302.

Levitt, H. (1973). "Speech Processing Aids for the Deaf: An Overview," IEEE Trans. Audio Electroacoust. AU-21, 269–273.

Markel, J. D., and Gray, A. H. (1976). *Linear Prediction of Speech* (Springer–Verlag, New York).

Picheny, M. A. (1977). "The Effect of Frequency Lowering on Vowel Perception," unpublished masters thesis, Massachusetts Institute of Technology.

Pickett, J. M. (Ed.) (1977). *Proceedings of the Research Conference on Speech Aids for the Deaf* (Gallaudet College, Washington, DC).

Reed, C. M., Hicks, B. L., Braida, L. D., and Durlach, N. I. (1978). "The Effect of Frequency Lowering on Consonant Discrimination," J. Acoust. Soc. Am. Suppl. 1 63, S80.

Reeder, K., Strong, W. J., and Palmer, E. P. (1977). "Preliminary Study of a Low-Frequency, Formant-Based Speech Code for the Severely Hearing Impaired," J. Acoust. Soc. Am. 61, 1379–1381.

Reeder, K. (1979). Personal communication.

Risberg, A. (1965). "The Transposer and a Model for Speech Perception," STL–QPSR 4, 26–30, KTH, Stockholm.

Schreiner, P. G. (1977). "A Speech Processing Aid for Persons with Profound Sensorineural Hearing Loss: An Application of Bandwidth Compression and Dynamic Range Reduction," Ph. D. dissertation, The University of Michigan.

Smith, M., Robinson, K. E., and Strong, W. J. (1981). "Intelligibility and Quality of Linear Predictor and Eigenparameter Coded Speech," IEEE Trans. Acoust. Speech Signal Process. ASSP–29, 391–395.

Stewart, S. E., Strong, W. J., and Palmer, E. P. (1977). "Comparison of Some Auditory Speech Codes for the Severely Hearing Impaired," see Pickett (1977).

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

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