

Development and testing of artificial low-frequency speech codes

Charlotte M. Reed, PhD; Matthew H. Power, MS; Nathaniel I. Durlach, MA; Louis D. Braida, PhD;
Kristin K. Foss, MS; Jean A. Reid, BS; Susan R. Dubois, MS

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract—In a new approach to the frequency-lowering of speech, artificial codes were developed for 24 consonants (C) and 15 vowels (V) for two values of lowpass cutoff frequency F (300 and 500 Hz). Each individual phoneme was coded by a unique, nonvarying acoustic signal confined to frequencies less than or equal to F . Stimuli were created through variations in spectral content, amplitude, and duration of tonal complexes or bandpass noise. For example, plosive and fricative sounds were constructed by specifying the duration and relative amplitude of bandpass noise with various center frequencies and bandwidths, while vowels were generated through variations in the spectral shape and duration of a ten-tone harmonic complex. The ability of normal-hearing listeners to identify coded Cs and Vs in fixed-context syllables was compared to their performance on single-token sets of natural speech utterances lowpass filtered to equivalent values of F . For a set of 24 consonants in *C-/a/* context, asymptotic performance on coded sounds averaged 90 percent correct for $F=500$ Hz and 65 percent for $F=300$ Hz, compared to 75 percent and 40 percent for lowpass filtered speech. For a set of 15 vowels in */b/-V-/t/* context, asymptotic performance on coded sounds averaged 85 percent correct for $F=500$ Hz and 65 percent for $F=300$ Hz, compared to 85 percent and 50 percent for lowpass filtered speech. Identification of coded signals for $F=500$ Hz was also examined in CV syllables where C was selected at random from the set of 24 Cs and V was selected at random from the set of 15 Vs. Asymptotic performance of roughly 67 percent correct and 71 percent correct was obtained for C and V identification, respectively. These scores are somewhat lower than those obtained in the fixed-context experiments. Finally, results were obtained concerning the effect of token variability on the identification of lowpass

filtered speech. These results indicate a systematic decrease in percent-correct score as the number of tokens representing each phoneme in the identification tests increased from one to nine.

Key words: *acoustic signal, artificial low-frequency speech code, bandpass noise, lowpass filtered speech, signal processing, spectrogram.*

INTRODUCTION

Much of the previous work on frequency lowering has been based on a signal-processing approach in which natural speech was altered to achieve some type of lowering of the speech spectrum (5). The intelligibility of the resulting frequency-lowered signals for both normal and hearing-impaired listeners has generally been disappointing in that overall performance on lowered speech was found, at best, to be comparable to that obtained with lowpass filtering to equivalent bandwidths (4,16,27,28). It is not known, however, whether these negative results reflect the use of inappropriate signal-processing schemes or basic limitations of the auditory system (assuming that sufficient training has been provided).

In order to gain insight into this problem, we are temporarily relinquishing the constraint that the low-frequency representation of speech be created through signal processing of natural speech and determining the results that can be achieved by means of artificial low-frequency codes. To the extent that a low-frequency code can be created that does not degrade intelligibility too severely (after sufficient training), it will prove that the

auditory-speech reception system is capable of understanding speech that is confined to low frequencies. Such a result would not only serve as an "existence theorem" for the perception of low-frequency speech (and indicate that the search for effective signal-processing schemes that lower the frequency content of speech is not a waste of time), but it could also play an important role in the development of a new class of hearing aids. Consider a system consisting of two components: a speech recognizer (that has spoken speech as an input and phonemic symbols as an output) and an auditory coder (that has the phonemic symbols as input and acoustic signals as an output). The code in question could then be used in the auditory coder for listeners whose residual hearing is confined to low frequencies.

Results from several areas of research suggest that it should be possible to develop a low-frequency code that is substantially better than lowpass filtered normal speech. Examples include the classical work concerned with the learning of Morse code (6), work on the Tadoma method of speech communication (29), and work on reading aids for the blind (3,8,33). The results have shown that speech can be radically transformed in a variety of ways while maintaining reasonable levels of intelligibility (although communication rates may be slower than those achieved through normal speech communication). For example, in the work on reading aids for the blind (known as "Spell Talk"), each letter of written text was associated with a unique artificial sound code which corresponded to the phoneme most often resulting from that letter in English. Codes for vowels and semivowels were constructed through the addition of three damped sinusoids whose frequencies corresponded to the first three formants in the spectrum of naturally produced sounds, while consonant codes were derived using bandpass noise with different cutoff frequencies. The speech-synthesis system had a bandwidth of 8000 Hz. To aid subjects in word segmentation, pitch and loudness variations were used to signal the onset of a new word. After roughly 50 to 70 hours of practice on identification of sentences from a 50-word vocabulary, two subjects were able to perceive coded conversational speech from an unrestricted vocabulary with relatively few errors or repetitions at rates of roughly 175 words/min (33). In view of the rather modest amounts of training associated with these experiments, and the generally poor performance exhibited at the beginning of the training (which suggests that the code was sufficiently artificial to eliminate the possibility of immediate comprehension), these results are highly encouraging with respect to the ability of subjects to learn artificial codes.

Other studies of the perception of artificial speech codes have been undertaken to develop and test models of speech perception (2,30,31). These investigators have constructed coded speech from sine waves whose frequencies and amplitudes are modulated on the basis of measurements derived from spectrographic displays of natural speech. The results of these studies generally support the conclusion that highly simplified, abstract representations of speech can produce perceptual results that are qualitatively similar to those obtained with natural speech signals. Other investigators have studied various types of complex acoustic stimuli from the point of view of auditory perceptual learning (10,11,12,24,36). On the whole, these studies have been concerned with the relation among signal dimensionality, information transfer, and learning rates for stimulus identification. Although the results are generally consistent with the principle that information transfer increases with the number of stimulus dimensions (17), there appear to be no clear-cut rules governing the relationship between learning rates and stimulus construction.

Previous studies of low-frequency speech codes generally have not made major departures from natural speech utterances; for example, low-frequency codes for vowels have been derived from a linear-predictive analysis of natural speech (1). There has also been a number of attempts to recode the high-frequency energy typically associated with fricative sounds into the low-frequency region of residual hearing in listeners with high-frequency loss. In some of these schemes, which employ vocoding or transposition, the recoded information is added to the low-frequency region of the original speech signal (13,25,34). In another approach, which is limited to unvoiced fricatives, the original speech signal is turned off upon recognition of an unvoiced fricative and replaced by a low-frequency bandpass noise (14). Such schemes are capable of improving the intelligibility of high-frequency speech sounds; however, the overall performance of these systems on full sets of sounds generally is not much better than that observed with lowpass filtering.

Our current approach to creating low-frequency speech codes is more artificial than that of previous studies in that the codes are created independent of natural speech utterances. The development of these codes is based on considerations related to optimizing information transfer and minimizing the magnitude of the relearning problem. Under the constraints of a low-bandwidth system, the stimuli are created by varying frequency content (both center frequency and bandwidth), amplitude, and duration, while attempting to maintain abstract properties of natural speech signals to minimize the learning problem. The artificial codes are

constructed such that: 1) coding operates on individual phonemes rather than whole syllables; 2) each phoneme is assigned a unique, nonvariable auditory code; 3) the durations of the coded waveforms do not exceed the average durations of natural speech; and, 4) the frequency content of the waveforms is confined to frequencies less than or equal to a fixed cutoff frequency F .

We are optimistic that an artificial low-frequency code can be found that is superior to lowpass filtered speech (of the same cutoff frequency) because we believe that the latter is far from optimal. Not only do we disagree with the usual evolutionary argument that normal broadband speech is optimal for the normal ear (since there are many important biological constraints on speech-production mechanisms unrelated to communication, such as those concerned with breathing and eating), but there is no evidence that speech was evolved to optimize intelligibility under lowpass filtering conditions.

In this paper, we report the results of a series of identification experiments on normal-hearing subjects with sets of artificially coded consonants and vowels for two values of F (300 and 500 Hz). Performance on the coded sounds was evaluated with reference to performance achieved using natural, uncoded syllables that were lowpass filtered to the same cutoff frequency F . The values of F were chosen on the basis of studies of the intelligibility of lowpass filtered speech which indicate that performance is sufficiently low that room exists for improvement with coded stimuli (19). Inasmuch as the results obtained with the coded signals are not limited by variations due to multiple tokens, the reference tests with filtered natural syllables were based on single-token stimulus sets. When only one token of a syllable is included in the stimulus set, listeners can learn to identify that syllable based on the peculiarities of a particular recording rather than on the characteristics of the sounds being tested. As the number of tokens increases, listeners are less likely to use artifactual cues (due to memory limitations), and instead are forced to perform "class" identifications. To gain further understanding of the effects of token variation on speech intelligibility, an additional experiment was conducted with lowpass filtered natural speech in which intelligibility was studied as a function of the number of utterances representing each syllable.

The experiments reported here are derived from several studies conducted as student theses and projects (26).^{1,2} Despite the formal limitations of some of the experiments (e.g., the studies vary in such things as the number of subjects tested, the type of training provided, and the number of trials collected), we nevertheless feel that the results provide interesting data on important topics.

METHODS

Coded signals

Consonants (C). Codes for a set of 24 English consonants were constructed by specifying various abstract properties that characterize each of three major classifications of consonants (i.e., manner, place, and voicing). The specifications for each coded consonant (assuming Consonant-Vowel—CV, or Consonant-Vowel-Consonant—CVC context), in terms of frequency content, duration, relative amplitude, and interval of silence preceding or following the vowel are provided in **Table 1** for $F=300$ and 500 Hz. Orthographic symbols used to represent the coded consonant signals are also listed in **Table 1**. The stimuli were separated into the categories of plosives, fricatives, affricates, nasals, and semivowels. Plosive and fricative sounds were represented by bandpass noise whose duration was 30 msec for plosives and either 100 or 200 msec for fricatives. The affricate sounds were coded using two consecutive bands of noise with different center frequencies and durations. The semivowels consisted of three-tone complexes of 100-msec duration. Finally, the nasals were represented by 50-msec two-tone complexes. Voiceless Cs were characterized by a 100-msec silent interval between the C segment and the V segment whereas voiced Cs had contiguous C and V segments.

Vowels and Diphthongs (V). Codes for 10 vowels were generated by varying the spectral shape of a 10-tone harmonic complex with 50-Hz spacing for $F=500$ Hz and 30-Hz spacing for $F=300$ Hz. The relative amplitudes of the components for each vowel were derived from linearly predicted spectra of natural vowel utterances by measuring the amplitude of the spectrum at each multiple of 250 Hz between 250 and 2,500 Hz and then imposing this amplitude on a pure tone 5 times lower in frequency (for $F=500$ Hz) or 8.3 times lower in frequency (for $F=300$ Hz). Based on measurements of vowel durations in natural utterances (22), six of the vowels were coded as "long" (with a duration of 220 msec), and four of the vowels were coded as "short" (with a duration of 150 msec). The specification of each of 10 vowels in terms of overall duration and relative amplitude of the individual components for $F=500$ and 300 Hz is provided in **Table 2**, along with a set of orthographic symbols used to represent each coded vowel. The overall level of the individual vowels covers a range of roughly 13 dB.

Five diphthongs were generated by varying the frequency and amplitude of the individual components over the course of the stimulus. The stimulus began with a harmonic series whose amplitudes were based on those of the pure vowel from which the diphthong originated.

Table 1.
Specifications of coded consonants for F = 500 and 300 Hz.

Phoneme Category	Symbol for Coded Sound	Tone or Noise Content Frequency Range (Hz)		Overall Duration (ms)	Amplitude* (dB)	Silence Duration† (ms)
		F=500 Hz	F=300 Hz			
/p/	P	20-500	20-300	30	-9	100
/t/	T	300-500	180-300	30	-8	100
/k/	K	20-200	20-120	30	-8	100
/b/	B	20-500	20-300	30	-9	0
/d/	D	300-500	180-300	30	-8	0
/g/	G	20-200	20-120	30	-8	0
/f/	FF	20-500	20-300	200	-9	100
/θ/	TH	20-500	20-300	100	-29	100
/s/	S	300-500	180-300	200	-8	100
/ʃ/	SH	20-200	20-120	200	-8	100
/v/	VV	20-500	20-300	200	-9	0
/ʒ/	TX	20-500	20-300	100	-29	0
/z/	Z	300-500	180-300	200	-8	0
/ʒ/	ZH	20-200	20-120	200	-8	0
/h/‡	H	20-100	20-60	100	-16	100
		400-500	240-300	100	-17	
/hw/‡	WH	20-100	20-60	250	-6	100
		400-500	240-300	250	-7	
/tʃ/§	CH	300-500	180-300	20	-8	100
		20-200	20-120	80	-8	
/dʒ/§	J	300-500	180-300	20	-8	0
		20-200	20-120	80	-8	
/m/	M	50	30	50	0	0
		160	96	50	-10	
/n/	N	50	30	50	0	0
		320	192	50	-10	
/r/	R	60	36	100	0	0
		250	150	100	-10	
		350	210	100	-15	
/w/	W	60	36	100	0	0
		100	60	100	-10	
		350	210	100	-15	
/l/	L	60	36	100	0	0
		250	150	100	-10	
		500	300	100	-15	
/j/	Y	60	36	100	0	0
		400	240	100	-10	
		500	300	100	-15	

*The stimulus-component amplitudes given here represent rms D/A-converter output voltages V_{DA} relative to 0.3 volts. The D/A converter was followed by an amplifier: to ensure a comfortable listening level, each subject independently chose a (generally different) amplification setting α , and left α fixed throughout all conditions. α varied across subjects from 1.5 to 10.5 dB. Considering V_{DA} , α , and acoustic transfer functions, the overall level of the most intense stimulus varied across subjects from approximately 86 to 95 dB SPL.

†Defines the interval of silence between C and V segments in CV and CVC syllables.

‡The two noise bursts occurred simultaneously.

§The two noise bursts were non-simultaneous, with a 10-ms delay present between the onset of the first band and the onset of the second.

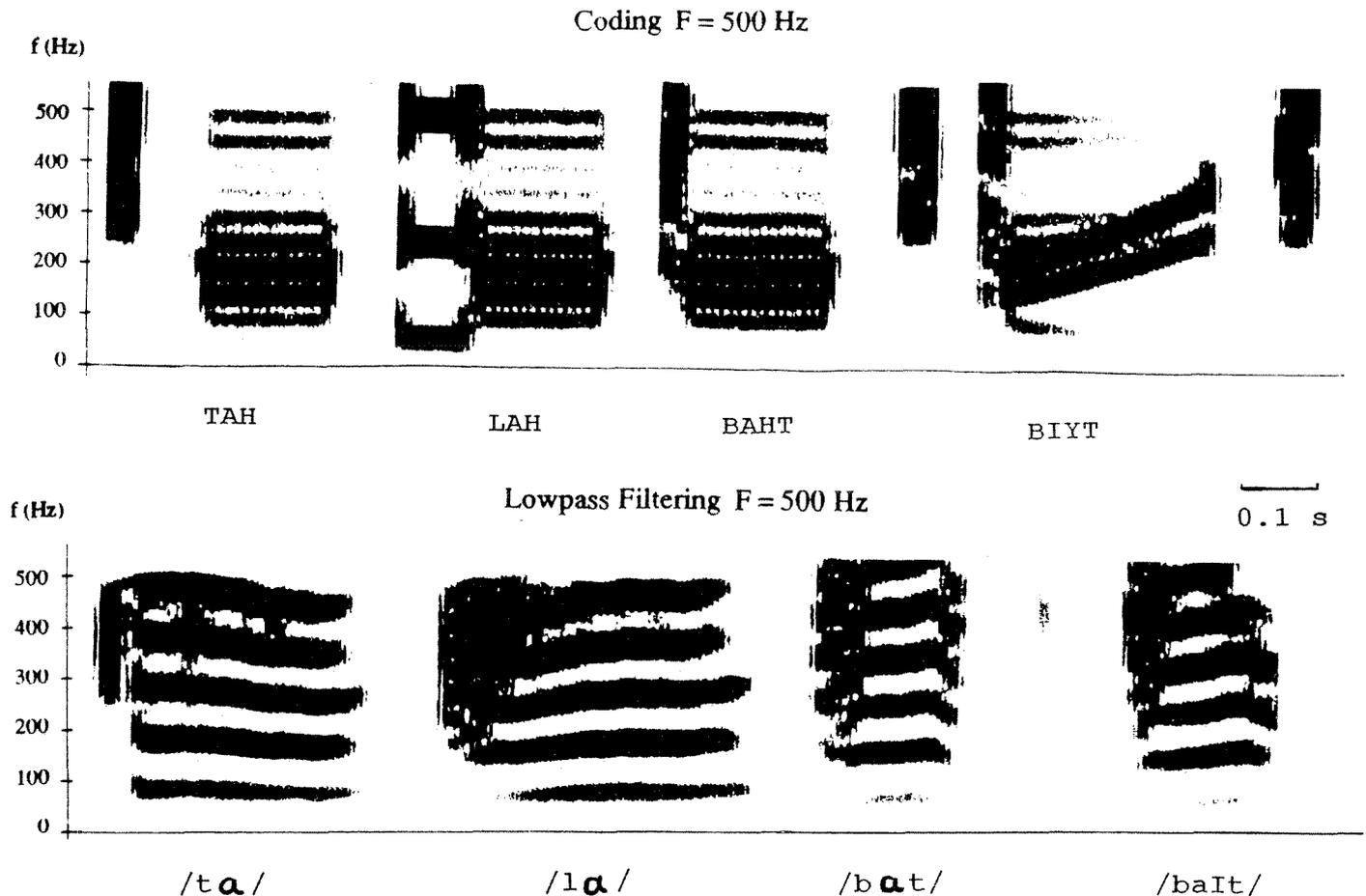


Figure 1.

Spectrograms comparing coded C-/AH/ and B-V-T syllables (upper panel) with lowpass filtered C-/a/ and /b-/V-/t/ syllables spoken by male talker BH (lower panel). These narrow-band (39-Hz) spectrograms were computed digitally on input signals with high frequency emphasis of 12 dB/oct.

The offset frequency of each harmonic component was determined from the formant-frequency ratios of the two pure vowels that make up the diphthong (22). The ratios of the first, second, and third formants of the two pure vowels determined the offset frequencies of components 1-2, 3-4, and 5-9, respectively. In addition, the amplitude of each component was linearly decreased by 10 dB over the 300-msec duration of each diphthong. The specification of each diphthong in terms of starting amplitude and offset frequency for $F=500$ and 300 Hz is provided in **Table 2**, along with a set of orthographic symbols to represent the coded diphthongs.

The coded signals were generated digitally using a sampling rate of 10 kHz. The waveforms were shaped with a Hanning window with rise/fall times of 10 msec. Coded CV and CVC syllables were constructed merely by concatenating the individual components (either with or without a 100-msec delay between segments depending on

whether or not the C was coded as unvoiced or voiced). Spectrograms of selected coded C and V signals are shown in **Figure 1**.

Natural speech signals

The natural speech signals were CV and /b-/V-/t/ utterances recorded by one male and one female talker. Recordings of the male talker were used in experiments comparing performance on coded versus lowpass filtered utterances (Experiments 1-5), while recordings of the female talker were used in a study of the effect of token variability on the perception of lowpass filtered speech (Experiment 6). The stimuli recorded by the male talker were three repetitions of 24 CV syllables composed of the 24 English consonants with the vowel /a/ and three repetitions of 15 CVC syllables composed of 15 English vowels in /b-/V-/t/ context. The stimuli recorded by the female talker were three utterances of each of 72 CV syllables com-

Table 2.
Specification of coded vowels and diphthongs for F = 500 and 300 Hz.

VOWELS													
Component No.	Frequency (Hz)		Phoneme Category Symbol for Code	Amplitude in dB*									
	F=500	F=300		/i/ EE	/I/ IH	/ɛ/ EH	/æ/ AE	/a/ AH	/ɔ/ AW	/U/ UU	/u/ OO	/ʌ/ UH	/ɜ/ ER
1	50	30		-12	-12	-8	-12	-17	-16	-24	-12	-8	-27
2	100	60		-34	-16	0	-8	-10	-8	-12	-14	0	-12
3	150	90		-45	-30	-17	-6	-8	-9	-21	-22	-12	-32
4	200	120		-52	-36	-22	-17	-8	-34	-21	-27	-14	-35
5	250	150		-54	-39	-25	-20	-21	-42	-33	-45	-14	-29
6	300	180		-46	-39	-23	-19	-30	-50	-44	-52	-30	-32
7	350	210		-55	-34	-10	-8	-38	-52	-47	-53	-35	-39
8	400	240		-50	-23	-23	-22	-41	-51	-50	-62	-36	-54
9	450	270		-36	-34	-24	-24	-38	-44	-42	-62	-34	-68
10	500	300		-46	-30	-25	-25	-40	-59	-57	-47	-37	-73
Duration (ms)				220	150	150	220	220	220	150	220	150	220

DIPHTHONGS																					
Phoneme Category	Code Symbol	Onset Amplitude in dB Component No.										Value of Frequency in Hz at Offset† Component No.									
		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
/eI/	AY	-12	-8	-6	-17	-20	-19	-8	-22	-24	-25	30	59	174	231	265	317	370	423	476	500
/aI/	IY	-17	-10	-8	-8	-21	-30	-38	-41	-38	-40	18	36	104	139	159	190	222	254	286	300
/aU/	OW	-17	-10	-8	-8	-21	-30	-38	-41	-38	-40	16	32	164	219	157	188	220	251	282	300
/oU/	OA	-16	-8	-9	-34	-42	-50	-52	-51	-44	-59	18	36	84	112	138	165	193	220	248	300
/ɔI/	OY	-16	-8	-9	-34	-42	-50	-52	-51	-44	-59	39	77	182	243	232	279	325	372	418	500
												23	46	109	146	139	167	195	223	251	300
												34	68	355	474	265	317	370	423	476	500
												21	41	213	284	159	190	222	254	286	300

*Amplitudes are specified as described in the footnote to Table 1.

†The upper row of frequency values for each diphthong is for F=500 Hz and the lower row for F=300 Hz.

posed of the 24 English consonants with the vowels /i a u/. The recordings were made in an anechoic chamber with an Electro-Voice RE-55 microphone placed 6 inches in front of and above the speaker's mouth. They were then low-pass filtered at 4.5 kHz and digitized at a 10-kHz sampling rate with 12-bit resolution.

The utterances were lowpass filtered at F=500 or 300 Hz. Three different methods of filtering were used over the course of the study: a cascade of two analog Butterworth lowpass filters (Krohn-Hite 3343R), a 16-pole Butterworth digital filter, and an FIR digital filter designed using a Kaiser window. The responses of the two Butterworth filters are essentially equivalent and provide attenuation of approximately 100 dB/oct over the 500-1000 Hz interval. The FIR filter, with a much sharper passband-stopband transition, is more effective at attenuating out-of-band signal

components, in spite of its somewhat smaller stopband attenuation (85 dB at 1000 Hz and 97 dB at 5000 Hz).

Spectrograms of selected lowpass filtered CV and CVC utterances are provided in **Figure 1**, where they can be compared to coded versions of the same syllables.

Experiments

A description of the six experiments is provided in **Table 3**. A one-interval forced-choice identification procedure with trial-by-trial correct-answer feedback was employed in all experiments. The subjects were provided with a closed set of response alternatives that corresponded to the stimulus set for a given experiment (see **Table 3**). An experimental run consisted of randomized presentations of the stimuli from the set being tested. Following each stimulus presentation, the subject was given unlimited time

in which to respond by typing one of the response alternatives into a computer terminal. If the answer was correct, the next stimulus was presented. If the response was incorrect, the correct response appeared on the screen and the stimulus was presented acoustically three times. The number of trials in an experimental run varied across experiments (see **Table 3**). The experiments were controlled by a VAX 11/750 computer. Digitized waveforms for both the coded and natural-speech signals were sent through a D/A converter, passed through an amplifier, and presented over TDH headphones to subjects seated in an IAC sound-treated booth. Subjects adjusted the overall level of the stimuli to a comfortable listening level, which generally ranged from roughly 85-95 dB SPL.

In Experiments 1, 3, and 4, subjects received training on subsets of the test stimuli prior to experiments with the full stimulus set for both coded signals and for the one-token set of lowpass filtered natural speech. For consonants, the training sets consisted of the fricatives /f θ s ʃ v ʒ z ʒ/, the plosives /p t k b d g/, the semivowels /w r l j/

plus the aspirated fricatives /h hw/, and a set of plosives, nasals, and affricates /p t k tʃ b d g d ʒ m n/. For vowels and diphthongs, the training sets consisted of the 10 pure vowels and the 5 diphthongs. Identification tests were performed on each subset until performance appeared to stabilize. After performance had stabilized on all the subsets for a given experiment, identification experiments were conducted with the complete stimulus set until asymptotic performance was achieved. In Experiments 2, 5, and 6, testing was conducted only with the complete stimulus set appropriate for a given experiment.

The subjects in these experiments were normal-hearing young adults who were paid for their participation in the study.

Data analysis

The data for individual subjects in each experiment were examined through learning curves in which percent-correct scores were plotted as a function of the cumulative number of trials for each stimulus type. Estimates of

Table 3.
Experimental conditions.

	F (Hz)*	Coded Stimulus Set	Filtered Stimulus Set	Talker	Method of † Filtering	Number of ‡ Trials per Run	Subjects
Exp 1	500	24 C-/AH/ syllables	1-token or 3-token set of 24 C-/a/ syllables	Male (BH)	FIR or KH	565-720	AH, JR KF, ST
Exp 2	300	24 C-/AH/ syllables	1-token or 3-token set of 24 C-/a/ syllables	Male (BH)	DBW	96-504	DP, JP, MP
Exp 3	500	15 isolated Vs or 15 B-V-T syllables	1-token or 3-token set of 15 /b/-V-/t/ syllables	Male (BH)	KH or DBW	100-300	DP, MP
Exp 4	300	15 isolated Vs or 15 B-V-T syllables	1-token or 3-token set of 15 /b/-V-/t/ syllables	Male (BH)	DBW	150-540	AA, DP, JP, LA, MP
Exp 5	500	CV syllables§ with random selection of C from set of 24 and V from set of 15; 24 C-/AH/ syllables; or 15 isolated Vs				50-360	DP, MP, PA
Exp 6	500		2 1-token sets of 24 C-/a/ syllables; 1 3-token set of 24 C-/a/ syllables; 1 9-token set of 24 CV syllables where V=/iaul/	Female (DC)	KH	648 or 720	SD

*F is lowpass cutoff value for frequency content of waveforms.

†Three methods of filtering were used: Krohn-Hite filters (KH), digital FIR filters (FIR), and digital Butterworth filters (DBW). See text for further description.

‡Within a given experiment, the range in number of trials per run arises either from computer failure before the completion of a full experimental run or from individual subjects' preferences concerning the length of runs.

§Experimental runs for the fixed-context and one-component-identified roving-context conditions were interleaved: on one run, subjects listened to fixed context; on the next run, subjects listened to roving-context. All C-identified runs preceded all V-identified runs. The both-components-identified roving-context conditions were completed before this interleaved set.

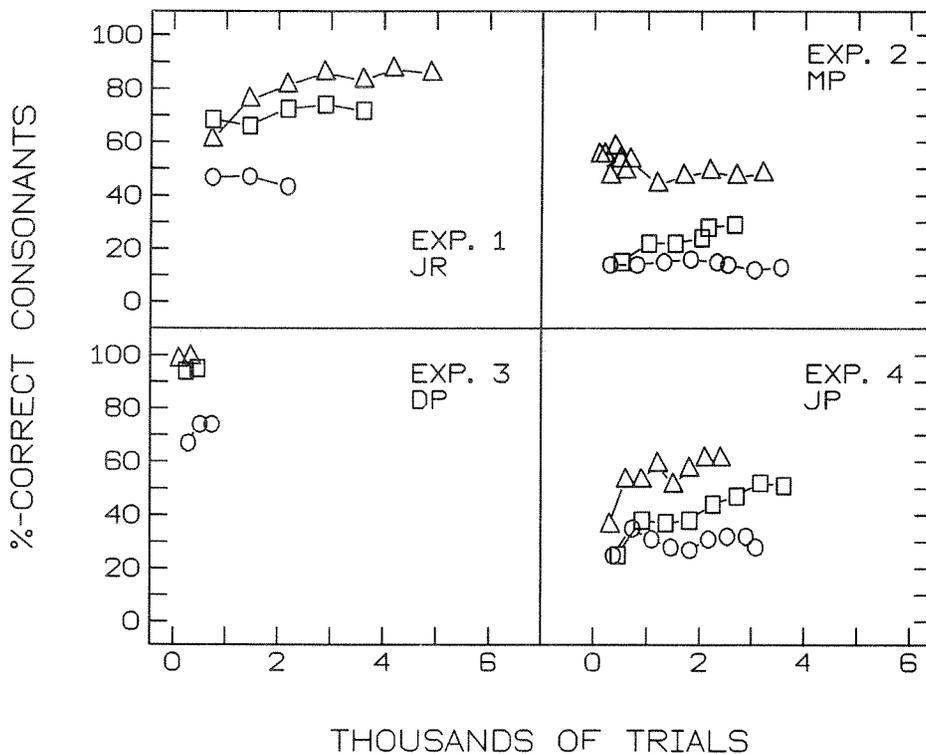


Figure 2.

Examples of learning curves from individual subjects in Experiments 1, 2, 3, and 4. Percent-correct identification scores are plotted as a function of cumulative number of trials.

Triangles represent results for coded stimuli; squares represent results with a 1-token set of lowpass filtered speech; and circles represent results with a 3-token set of lowpass filtered speech.

asymptotic performance were made through visual inspection of the learning curves. Examples of learning curves for individual subjects in Experiments 1 through 4 are presented in **Figure 2**. Confusion matrices were generated for each experimental condition and each subject from trials obtained at the end of the training periods. For a given experiment, the matrix for each stimulus set and each subject contained the same number of trials. The number of trials per matrix varied across experiments and was limited by the number of trials obtained near asymptotic performance for each subject. The percent-correct scores derived from the matrices for individual subjects in Experiments 1-6 are provided in **Table 4**.

For each experimental condition, matrices were collapsed across subjects and calculations were made of information transfer on various articulatory and acoustic features. Tables of feature definitions for consonants and vowels are provided in **Appendix 1** (7,19,38).

RESULTS

Experiments 1 and 2

For $F=500$ Hz, asymptotic performance on the coded sounds (which was similar across subjects and averaged roughly 90 percent correct) was higher than that on the 1-token set of filtered sounds (where performance averaged roughly 75 percent correct). Of the four subjects tested,

three showed advantages in the range of 20 to 30 percentage points for coded over 1-token filtered sounds, while one subject (ST) had similar performance for the two types of signals. For $F=300$ Hz, asymptotic performance on the coded sounds (averaging roughly 65 percent correct) was substantially higher than that for the 1-token set of filtered CVs (averaging roughly 40 percent correct). Subjects MP and JP exhibited similar performance, while the scores for DP were substantially higher. For all subjects at both values of F , performance on the 3-token set of filtered C-/a/ syllables was substantially lower than on the 1-token set. The average difference in asymptotic performance between the 1-token and 3-token sets was roughly 30 percentage points for $F=500$ Hz and 20 percentage points for $F=300$ Hz.

The percentage of transmitted information (TI) on a set of eight features is provided for the consonants in the upper half of **Figure 3**. The grouping of the bars within each feature is in the order: coded consonants in C-AH syllables, a 1-token set of filtered C-/a/ syllables, and a 3-token set of filtered C-/a/ syllables. For both values of F , performance on the coded stimuli exceeded that on the 1-token set of filtered stimuli for all features except nasality. In general, the responses to a given stimulus were more widely distributed for filtered stimuli than for coded stimuli, even for the single-token set of filtered speech sounds. The overall pattern of confusions for filtered sounds was consistent with expectations based on previous studies (19).

Table 4.

Percent-correct identification scores for the six experiments.

Experiment 1 N=720				
	Coded	1-Token Filtered	3-Token Filtered	
AH	94.6	65.8	29.4	
JR	85.6	71.7	43.3	
KF	89.9	74.0	35.7	
ST	92.6	93.8	56.4	
Experiment 2 N=576				
	Coded	1-Token Filtered	3-Token Filtered	
DP	88.0*	58.0	25.3	
JP	50.5	31.2	16.3	
MP	49.3	28.6	13.0	
Experiment 3 N=405				
	Coded Isolated	Coded /b/-V-/t/	Filtered 1-Token	Filtered 3-Token
DP	98.5	99.3	94.6	74.1
MP	82.7	69.9	77.5	61.7
Experiment 4 N=675				
	Coded	1-Token Filtered	3-Token Filtered	
AA	59.1	42.7	—	
DP	99.3†	72.4	40.1	
JP	62.1	51.9	30.5	
LA	40.3	24.6	—	
MP	46.1	37.2	26.5	
Experiment 5—Consonants N=642				
	C-AH	CV Respond C	CV Respond C + V	
DP	96.3	83.8	86.4	
MP	72.3	38.5	30.1	
PA	80.7	—	83.6	
Experiment 5—Vowels N=642				
	V Isolated	CV Respond V	CV Respond C + V	
DP	98.6‡	92.4§	89.3	
MP	87.4	49.1	43.6	
PA	75.2	—	78.7	
Experiment 6 N=2160				
	1-Token (Set A)	1-Token (Set B)	3-Token	9-Token
SD	77.7	70.1	62.5	57.4

The scores were derived from the confusion matrices compiled from the last N trials for each subject in a given experiment. The four exceptions to the N values provided for each experiment (indicated by footnotes) occur in the data of subject DP, who tended to reach asymptotic performance within fewer trials than the other subjects. In summing matrices across subjects, a weighting factor was applied to the data of DP.

*192 trials †450 trials ‡525 trials §500 trials

As expected, the features of voicing and nasality had the highest levels of TI relative to that of other features; the absolute levels of performance on these two features, however, appear to be somewhat higher in the Miller and Nicely study than for our filtered 3-token data. One major source of confusion in the coded stimuli (which was not observed in the filtered stimuli) was confusion of the nasals M and N with the voiced stop consonants, which accounts for the poor reception of the feature nasality for coded sounds. In addition, place errors within a given class of sounds and voicing errors within the same manner of production accounted for a large percentage of confusions of coded stimuli for both values of F.

Experiments 3 and 4

At F=500 Hz, a different pattern of results was obtained for the two subjects tested. For subject DP, performance was nearly perfect for coded vowels presented in isolation and in B-V-T context as well as for the 1-token set of filtered /b/-V-/t/ syllables. For subject MP, asymptotic performance on isolated coded vowels and on the 1-token set of filtered syllables was equivalent (80 percent correct), and was roughly 10 percentage points higher than that obtained on coded vowels presented in B-V-T context. For F=300 Hz, on the other hand, each of the five subjects exhibited higher asymptotic performance on the coded B-V-T syllables than on the 1-token set of filtered /b/-V-/t/ syllables. Averaged across subjects, asymptotic performance for coded vowels was roughly 65 percent compared to 50 percent on 1-token filtered sounds. Levels of performance were similar across subjects, with the exception of DP, who was able to identify the 300-Hz coded sounds perfectly. The effect of increasing the set size of the filtered syllables from 1 to 3 tokens was similar to that observed in the consonant identification experiments. The average difference in asymptotic performance between the 1-token filtered set and the 3-token filtered set was roughly 20 percentage points for F=500 Hz and F=300 Hz (for the three subjects who were tested with the 3-token set at F=300).

The percentage of information transfer on a set of seven vowel features is provided in the lower half of **Figure 3**. The grouping of the bars on each feature is in the order (for F=500 Hz): coded vowels in B-V-T syllables, coded vowels in isolation, a 1-token set of filtered /b/-V-/t/ syllables, and a 3-token set of filtered /b/-V-/t/ syllables. For F=300 Hz, data were not obtained for coded vowels in B-V-T syllables; the ordering of the bars within each feature is the same as that described above for F=500 Hz, with the omission of the initial condition. At F=500 Hz, overall information transfer was equivalent for coded Vs in isolation and in B-V-T syllables and for a single-token set of

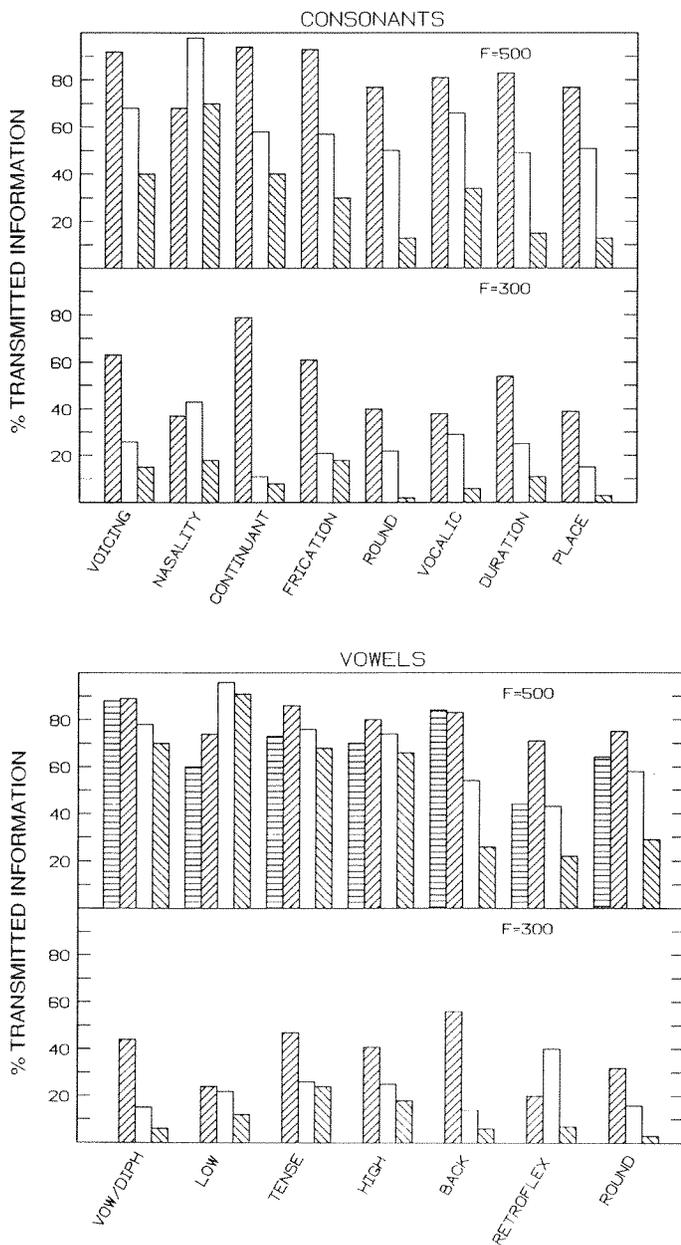


Figure 3.

Percent-correct transmitted information on a set of eight consonant features (upper half of figure) and on a set of seven vowel features (lower half of figure). Results are shown for F=500 and 300 Hz. See text for description of experimental conditions corresponding to the ordering of the bars within each feature.

filtered /b/-V-/t/ syllables. The information transfer associated with individual features (shown in **Figure 3**) was generally similar with the exception of the features low (better perceived for lowpass filtering) and back (better perceived for coding). At F=300 Hz, information transfer was higher for isolated coded vowels compared to one-token

filtered /b/-V-/t/ syllables. With the exception of the feature retroflexion, features were better perceived under coding than filtering. The confusion patterns observed under low-pass filtering were generally similar to those reported by Miller (18) for identification of vowels lowpass filtered to 670 Hz.

Experiment 5

The results of Experiment 5 for the identification of coded CV syllables at F=500 Hz are presented in the upper half of **Figure 4** for consonants, and in the lower half of **Figure 4** for vowels. Percent-correct consonant identification from a set of 24 coded consonants is shown for a fixed vowel context (C-AH) as well as for a roving vowel context (where V was drawn at random from the set of 15 coded vowels and diphthongs). For subjects DP and MP, results of the roving-vowel experiments are shown for two conditions: one in which subjects identified only the C component, and one in which subjects identified both the C and V components. For subject PA, results were obtained only for the condition in which both components were identified. For subject MP, consonant identification was clearly superior in fixed context (roughly 75 percent correct) as opposed to roving context (roughly 35 percent correct). With a roving vowel context, performance was similar whether this subject was asked to respond only to C or to both C and V. For subject DP, roving-vowel context appears to have had only a minor effect on consonant recognition, reducing the near-perfect scores obtained in fixed-vowel context to roughly 90 percent correct for both sets of instructions with roving-vowel context. Subject PA appears to perform equally well for fixed and roving-context conditions (with asymptotic performance of roughly 75 to 80 percent correct). In the lower half of **Figure 4**, percent-correct identification of the set of 15 coded vowels is shown for vowels presented in isolation as well as for vowels presented in roving-context CV syllables where C was selected at random from the set of 24 coded Cs. For the roving-context experiments, subjects MP and DP identified either V only or both C and V, while subject PA responded using the second set of instructions only. For each subject, the same general pattern of results was observed for vowel identification as for consonant identification. Again, subject MP shows a much larger effect for roving versus fixed context compared to subjects DP and PA.

For the roving-context conditions in which subjects were required to identify both the C and V components, percent-correct identification of phonemes (which is the average of correct identification of the C components and

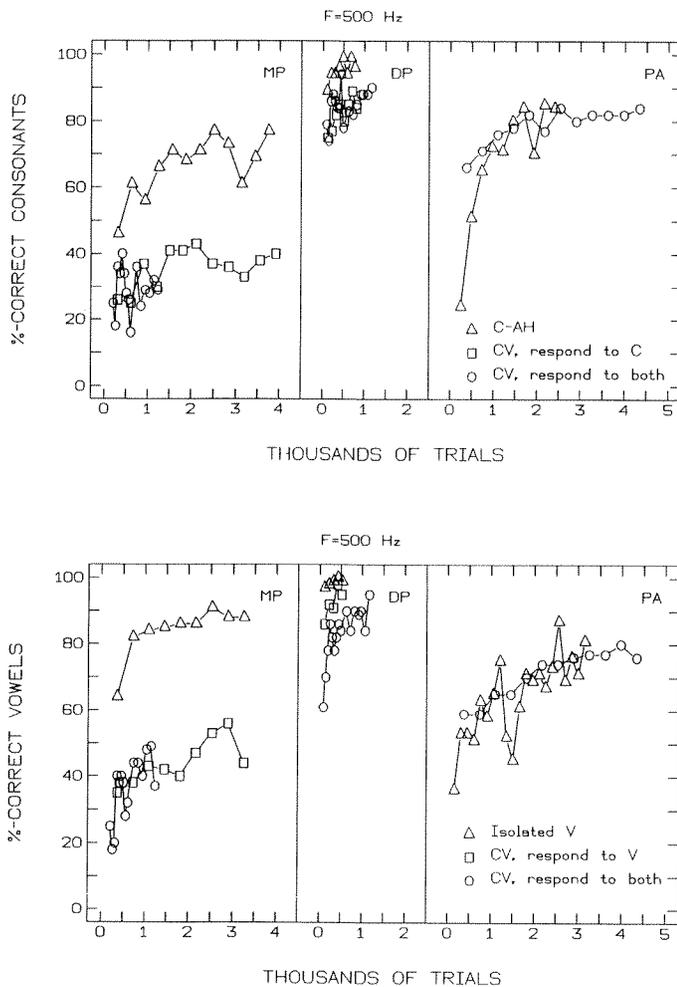


Figure 4. Percent-correct consonant (top half of figure) and vowel (bottom half of figure) identification for coded stimuli with $F=500$ Hz plotted as a function of number of trials. Results are shown for three individual subjects.

the V components), asymptoted at roughly 38 percent correct for MP, 80 percent for PA, and 90 percent correct for DP. The subjects' ability to identify individual syllables (requiring correct identification of both the C and V components in a given syllable) ranged from 10 percent for MP, to 70 percent for PA, to 80 percent for DP.

Experiment 6

Learning curves for the identification of consonants in lowpass filtered CV syllables ($F=500$ Hz) are presented in **Figure 5** for two different 1-token sets of C/a/ syllables, for a 3-token set of C/a/ syllables, and for a 9-token set of CV syllables (3 each of C/a/, C/i/, and C/u/). Results obtained on one subject indicate that asymptotic perfor-

mance decreases as the number of utterances representing each syllable increases. As the number of tokens increased from one to three to nine, asymptotic performance decreased from roughly 78 to 64 to 58 percent correct. Thus, the effect appears to be substantial, particularly for small numbers of tokens. Asymptotic performance on the two 1-token sets of syllables was similar, although not identical, and less training was required to reach asymptotic levels for the second set of 1-token tests. Initial performance on the first single-token set, the 3-token set, and the 9-token set was similar (roughly 55 percent correct); however, the level of asymptotic performance and the number of sessions required for stable performance were different for the different token sets.

DISCUSSION

The results of C and V identification tests in fixed-context nonsense syllables indicate an advantage for coded stimuli over single-token sets of lowpass filtered natural speech utterances. The result is somewhat more robust for Cs as opposed to Vs, in that an advantage for the coded signals was observed at both values of F (300 and 500 Hz) for Cs and only at the lower value of F for Vs. For Cs, the size of the advantage averaged roughly 15 percentage points at $F=500$ Hz and 25 percentage points for $F=300$ Hz. For Vs, performance was equivalent for the two types of signals at $F=500$ Hz; at $F=300$ Hz, an advantage of roughly 15 percentage points was observed for coded signals. Thus, when subjects trained to asymptotic performance on both the coded stimuli and on single-token sets of lowpass filtered utterances, they appeared to be able to extract a greater amount of information from the coded sounds. This result offers support (at the segmental level, at least) for the existence of artificial low-bandwidth codes that contain more information than natural speech at equivalent bandwidths.

The identification performance achieved with the coded signals appears to be superior to that obtained for frequency-lowered speech achieved through signal processing of natural speech sounds, even though these naturally lowered stimuli typically have substantially higher values of F (due primarily to limitations imposed by the signal-processing techniques). An example is the frequency-lowering of natural speech using a pitch-invariant nonuniform compression of the short-term spectral envelope to $F=1667$ and 1000 Hz (9). Normal-hearing subjects were trained to asymptotic performance on identifying consonants in multiple-token sets of CV syllables.³ Asymp-

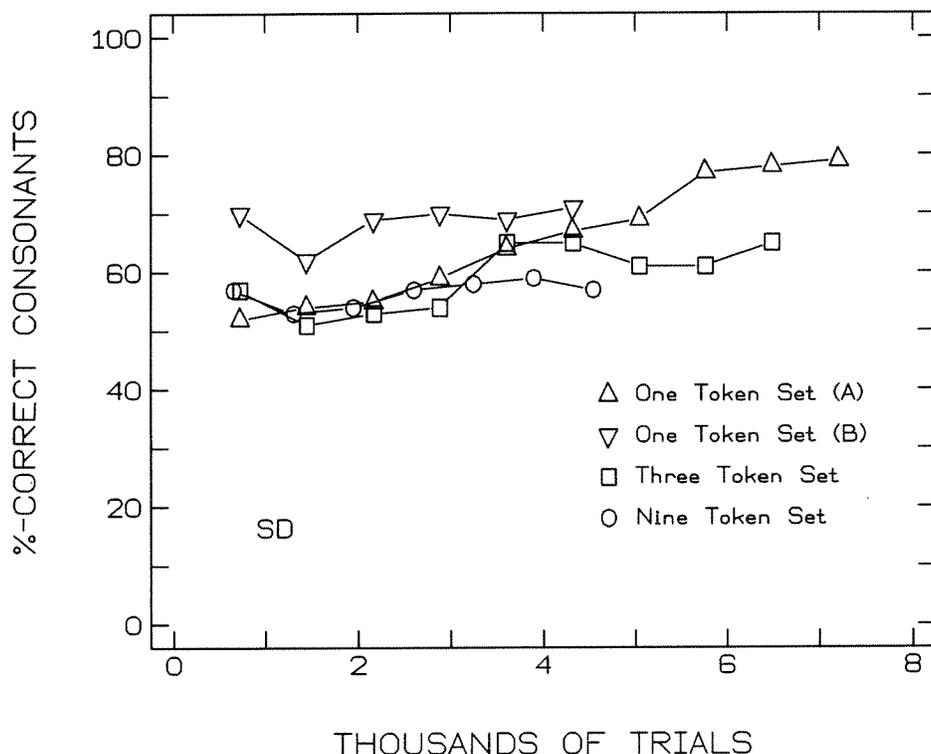


Figure 5. Percent-correct identification of consonants in lowpass filtered CV syllables with $F=500$ Hz plotted as a function of number of trials.

otic performance averaged 82 percent and 50 percent correct, respectively, for $F=1667$ and 1000 Hz. At $F=1000$ Hz, the score obtained with the frequency-lowered speech was substantially worse than that obtained for lowpass filtering to $F=1000$ Hz (67 percent correct). Furthermore, tests conducted with single-token stimulus sets showed no advantage for lowering over lowpass filtering: high asymptotic performance (identification scores greater than 95 percent correct) was achieved for both types of signals. The results obtained by Hicks³ represent some of the highest levels of identification performance achieved with lowering of natural speech signals (27,28). It appears that the results achieved for the coded signals at bandwidths of 500 and 300 Hz are promising relative to results achieved with lowering of natural speech.

Previous work on the development of low-frequency speech codes for vowels has been conducted (1). Various methods of coding to $F=1000$ Hz were derived from a linear-predictive analysis of natural speech and included variations in the number of components represented in the code, the type of lowering (linear or warped), and the transformation of fundamental frequency. The codes previously studied by Allen *et al.* (1) that most closely resemble the coded vowels used in the present study are their codes that introduce all the harmonic components present in a naturally produced utterance linearly lowered by a factor

of 4 (to $F=1000$ Hz) with F_0 lowered either linearly or by a somewhat smaller factor. Results of ABX word-discrimination tests indicated similar performance for these codes and for lowpass filtering to $F=1000$ Hz. Obvious differences between the signals of Allen *et al.* (1) and our coded vowels, are that their codes were limited to $F=1000$ Hz (compared to $F=500$ and 300 Hz studied here) and that their codes are more closely linked to natural speech utterances, whereas ours are generated by a more abstract set of rules.

A salient property of the coded signals used in the present study is that a fixed token was specified for each phoneme (i.e., the token-to-token variability that is characteristic of naturally produced speech is absent in these signals). In comparisons with lowpass filtered natural speech, token variability was artificially eliminated by the use of single-token sets of utterances. The results of Experiment 6, as well as comparisons of the 3-token sets with the single-token sets in Experiments 1-4, indicate a systematic decrease in identification performance as the number of tokens from a single talker representing each phoneme increases from one to three to nine. In addition, it appears that the particular set of single tokens chosen can influence performance. For example, identification performance on consonants ranged from 65-92 percent for three different single-token sets. Other studies of token

variability reported in the literature have been concerned with the effects of single versus multiple talkers on vowel perception (32,35). In these experiments, listeners were asked to identify vowels in lists where utterances were drawn either from multiple talkers (12 or 15 different talkers) or from a single talker. The size of the effect of increasing the number of talkers represented in a stimulus list on the perception of vowels presented in isolation or in CVC nonsense syllables ranged from 0-12 percentage points. More recently, Pisoni and his colleagues (15,20,21,23) have undertaken a series of studies on the effects of inter-talker variability on speech perception. These studies have demonstrated that performance on a variety of tasks (including recognition of degraded words, recall of spoken word lists, and naming procedures) is influenced by the number of talkers employed for stimulus presentation.

Further study of the effects of token variability on speech intelligibility is important for several reasons, including the contribution of such studies to basic speech science. Another motivation for such studies arises from the current work on low-frequency speech codes, and concerns understanding the extent to which the reduction in information transfer associated with token variability increases in importance as auditory capacity (auditory area and auditory resolution within the auditory area) is reduced. That is, what is the effect of token variability on intelligibility for hearing-impaired listeners relative to normal listeners? The fixed-token approach requires the use of a speech recognizer at its input. Thus, the effect of errors made by the recognizer on the selection of fixed-token codes must be considered relative to the effect of errors introduced by token variability in systems which process natural speech.

Identification of coded segments in variable-context CV syllables proved to be more difficult than in fixed-context syllables: scores for either C or V identification were somewhat lower than those obtained in the fixed-context experiments for $F=500$ Hz. These results are inconclusive for several reasons: 1) comparable studies were not performed with lowpass filtered speech; 2) different patterns of results were obtained for the three subjects tested; and, 3) no coarticulatory rules were applied in the formation of the syllables. Subject DP was able to recognize 90 percent of the coded phonemes in variable-context syllables compared to 40 percent for MP, and the difference in performance between the tests with fixed and variable context was much larger for MP than for DP and PA. Large intersubject differences appear to be common in studies of perceptual learning of complex auditory stimuli. For

example, a large variability in absolute performance among subjects as well as apparent differences in learning strategies employed by different subjects has been reported (11,12). Finally, the development of a set of coarticulatory rules for the formation of multicomponent syllables may offer additional cues for segment recognition which are absent in the simple concatenation approach used in the current experiments.

A major question that must be addressed in future research is whether listeners can learn to perceive these sounds at rates of presentation comparable to those that occur in normal speech. Of crucial concern is whether or not, with ample training, these sounds can be processed with the automaticity that is associated with the processing of speech by normal-hearing listeners. The highest reported rates for the reception of Morse code, for example, are roughly 180 to 200 letters/min (37), which is slow relative to normal speaking rates (roughly 200 words/min which translates into roughly 1000 letters/min). Future research with the artificial low-frequency codes will be concerned with the perception of coded segmental streams, and in particular, the relationship between identification scores and the rate at which components are presented.

ACKNOWLEDGMENT

This research was supported by Grant No. R01 DC00117 from the National Institutes of Health.

END NOTES

¹DuBois SR: A study of the intelligibility of low-pass filtered speech as a function of the syllable set type-token ratio. Project paper, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1984.

²Foss, KK: Identification experimentation on low frequency artificial codes as representation of speech. B.S. thesis, Massachusetts Institute of Technology, 1983.

³Hicks BL, Braida LD, Durlach NI: Identification of filtered and lowered CV monosyllables. Unpublished manuscript, 1981.

REFERENCES

1. Allen DR, Strong WJ, Palmer EP: Experiments on the intelligibility of low-frequency speech codes. *J Acoust Soc Am* 70:1248-1255, 1981.

2. **Bailey PJ, Summerfield Q, Dorman M:** On the identification of sine-wave analogues of certain speech sounds. *Status Report on Speech Research SR-51/52*, 1-25. New Haven, CT: Haskins Laboratories, 1977.
3. **Bellavia DC:** A prosthetic reading aid for the blind. PhD diss., Carnegie-Mellon University, 1970.
4. **Block von R, Boerger G:** Horverbessernde verfahren mit bandbreitenkompression. *Acustica* 45:294-303, 1980.
5. **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*, 1979.
6. **Bryan WL, Harter N:** Studies in the physiology and psychology of the telegraphic language: The acquisition of a hierarchy of habits. *Psychol Rev* 6:345-375, 1899.
7. **Chomsky N, Halle M:** *The Sound Pattern of the English Language*. New York: Harper & Row, Publishers, 1968.
8. **Detwiler JS:** A synthetic dialect of English for a reading machine for the blind. PhD diss., Carnegie-Mellon University, 1971.
9. **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*, 121-124. New York: IEEE, 1981.
10. **House AS, Stevens KN, Sandel TT, Arnold JB:** On the learning of speechlike vocabularies. *J Verbal Learn Verbal Behav* 1:133-143, 1962.
11. **Leek MR, Watson CS:** Learning a tonal vocabulary. *J Acoust Soc Am (Suppl. 1)* 72:93, 1982.
12. **Leek MR, Watson CS:** Learning to detect auditory pattern components. *J Acoust Soc Am* 76:1037-1044, 1984.
13. **Lippmann RP:** Perception of frequency lowered consonants. *J Acoust Soc Am (Suppl. 1)* 67:78, 1980.
14. **MacKinnon DA, Lee HC:** Realtime recognition of unvoiced fricatives in continuous speech to aid the deaf. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 586-589. New York: IEEE, 1976.
15. **Martin CS, Mullennix JW, Pisoni DB, Summers WV:** Effects of talker variability on recall of spoken word lists. *J Exp Psychol [Learn Mem Cogn]* 15:676-684, 1989.
16. **Mazor M, Simon H, Scheinberg J, Levitt H:** Moderate frequency compression for the moderately hearing impaired. *J Acoust Soc Am* 62:1273-1278, 1977.
17. **Miller GA:** The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol Rev* 63:81-97, 1956.
18. **Miller GA:** The perception of speech. *For Roman Jakobson*, 353-360, M. Halle, H. Hunt, H. Maclean (Eds.). The Hague: Mouton & Co., 1956.
19. **Miller GA, Nicely PA:** An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am* 27:338-352, 1955.
20. **Mullennix JW, Pisoni DB:** Stimulus variability and processing dependencies in speech perception. *Percept Psychophys* 47:379-390, 1990.
21. **Mullennix JW, Pisoni DB, Martin CS:** Some effects of talker variability on spoken word recognition. *J Acoust Soc Am* 85:365-378, 1989.
22. **Peterson GE, Barney HL:** Control methods used in a study of the vowels. *J Acoust Soc Am* 24:175-184, 1952.
23. **Pisoni DB:** Effects of talker variability on speech perception: Implications for current research and theory. In *Proceedings of the 1990 International Conference on Spoken Language Processing*, Kobe, Japan. H. Fujisaki (Ed.), 1990.
24. **Pollack I, Ficks L:** Information of elementary multidimensional auditory displays. *J Acoust Soc Am* 26:155-158, 1954.
25. **Posen MP:** Intelligibility of frequency-lowered speech produced by a channel vocoder. Masters thesis, Massachusetts Institute of Technology, 1984.
26. **Power MH:** Representation of speech by low frequency artificial codes. B.S. thesis, Massachusetts Institute of Technology, 1985.
27. **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* 74:409-419, 1983.
28. **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* 78:2139-2141, 1985.
29. **Reed CM, Rabinowitz WM, Durlach NI, Braida LD, Conway-Fithian S, Schultz MC:** Research on the Tadoma method of speech communication. *J Acoust Soc Am* 77:247-257, 1985.
30. **Remez RE, Rubin PE, Pisoni DB, Carrell TD:** Speech perception without traditional cues. *Science* 212:947-950, 1981.
31. **Remez RE, Rubin PE, Nygaard LC, Howell WA:** Perceptual normalization of vowels produced by sinusoidal voices. *J Exp Psychol* 13:40-61, 1987.
32. **Strange W, Verbrugge RR, Shankweiler DP, Edman TR:** Consonant environment specifies vowel identity. *J Acoust Soc Am* 60:213-224, 1976.
33. **Sullivan KJ:** Understanding machine generated Spelltalk: Adaptation to speech processing of a non-speech code. PhD diss., Carnegie-Mellon University, 1972.
34. **Velmans ML, Marcuson M:** A speechlike frequency transposing hearing aid for the sensory-neural deaf. Report to the Department of Health and Social Security (Contract No. R/E 1049/84) London, 1980.
35. **Verbrugge RR, Strange W, Shankweiler DP, Edman**

- TR: What information enables a listener to map a talker's vowel space. *J Acoust Soc Am* 60:198-212, 1976.
36. **Watson CS, Foyle DC, Kidd GR:** Limits of auditory pattern discrimination for patterns with various durations and numbers of components. *J Acoust Soc Am* 88:2631-2638, 1990.
37. **Watson CS:** Time course of auditory perceptual learning. *Ann Otol Rhinol Laryngol (Suppl. 74)* 89:96-102, 1980.
38. **Wickelgren W:** Distinctive features and errors in short-term memory for English consonants. *J Acoust Soc Am* 39:1248-1266, 1966.

APPENDIX

Appendix 1a. Classification of the 24 consonants on a set of eight features derived from Chomsky and Halle (1968), Miller and Nicely (1955), and Wickelgren (1966).

	p	t	k	b	d	g	f	θ	s	ʃ	v	ʒ	z	ʒ	tʃ	dʒ	m	n	r	w	l	j	h	hw	
	P	T	K	B	D	G	F	TH	S	SH	VV	TX	Z	ZH	CH	J	M	N	R	W	L	Y	H	WH	
VOCALIC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	
VOICING	0	0	0	1	1	1	0	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0
NASALITY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	
CONTINUANT	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	
ROUND	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	
FRICATION	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	
DURATION	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	
PLACE	0	1	4	0	1	4	0	1	2	3	0	1	2	3	3	3	0	1	1	0	2	3	4	4	

Appendix 1b follows on page 82.

Appendix 1b. Feature classification of the 15 vowels and diphthongs on a set of seven features derived from Chomsky and Halle (1968). A classification of "c" for diphthongs indicates that the definition of that feature changes from the onglide to the offglide portions of the diphthong.

	i	I	ɛ	ɛ̃	ɑ	ɔ	ʊ	u	ʌ	ɜ̃	eI	aI	aU	oU	ɔI
	EE	IH	EH	AE	AH	AW	UU	OO	UH	ER	AY	IY	OW	OA	OY
DIPHTHONG	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
LOW	0	0	0	1	1	1	0	0	0	0	0	c	c	0	0
TENSE	1	0	0	1	1	1	0	1	0	0	1	1	1	1	1
HIGH	1	1	0	0	0	0	1	1	0	1	0	c	c	0	c
BACK	0	0	0	0	1	1	1	1	1	1	0	1	1	1	c
RETRO	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
ROUND	0	0	0	0	0	1	1	1	0	1	0	0	c	1	c