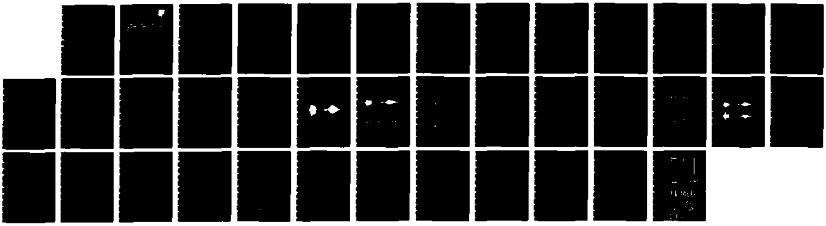
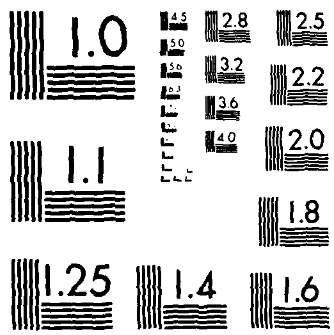
AD-A163 307   A COMPARATIVE ANALYSIS OF WHISPERED AND NORMALLY   1/1
PHONATED SPEECH USING AN LPC-10 VOCODER(U) ROME AIR
DEVELOPMENT CENTER GRIFFISS AFB NY J B WILSON ET AL
UNCLASSIFIED   DEC 85 RADC-TR-85-264   F/G 17/2   NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

RADC-TR-85-264
In-House Report
December 1985

A COMPARATIVE ANALYSIS OF WHISPERED AND NORMALLY PHONATED SPEECH USING AN LPC-10 VOCODER

AD-A163 307

Johnny B. Wilson, PhD and James D. Mosko, PhD

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, NY 13441-5700

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-85-264 has been reviewed and is approved for publication.

APPROVED:

THADEUS J. DOMURAT
Chief, Signal Intelligence Branch
Intelligence & Reconnaissance Division

APPROVED:

WALTER J. SENUS
Technical Director
Intelligence & Reconnaissance Division

FOR THE COMMANDER:

RICHARD W. POULIOT
Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (IRAA) Griffiss AFB NY 13441-5700. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A163 307

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS N/A
2a. SECURITY CLASSIFICATION AUTHORITY N/A 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE N/A	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S) RADC-TR-85-264	5. MONITORING ORGANIZATION REPORT NUMBER(S) N/A
6a. NAME OF PERFORMING ORGANIZATION Rome Air Development Center 6b. OFFICE SYMBOL (If applicable) IRAA	7a. NAME OF MONITORING ORGANIZATION N/A
6c. ADDRESS (City, State, and ZIP Code) Griffiss AFB NY 13441-5700	7b. ADDRESS (City, State, and ZIP Code) N/A
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Rome Air Development Center 8b. OFFICE SYMBOL (If applicable) IRAA	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N/A
8c. ADDRESS (City, State, and ZIP Code) Griffiss AFB NY 13441-5700	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 62702F PROJECT NO. 4594 TASK NO. 15 WORK UNIT ACCESSION NO. 91

11. TITLE (Include Security Classification)
A COMPARATIVE ANALYSIS OF WHISPERED AND NORMALLY PHONATED SPEECH USING AN LPC-10 VOCODER

12. PERSONAL AUTHOR(S)
Johnny B. Wilson, PhD, James D. Mosko, PhD

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
In-House	FROM ______ TO ______	December 1985	40

16. SUPPLEMENTARY NOTATION
N/A

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
17	02		Speech Processing; LPC-10
05	07		Linguistics;
			Phonetics

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The determination of the performance of an LPC-10 vocoder in the processing of adult male and female whispered and normally phonated connected speech was the focus of this study. The LPC-10 vocoder's analysis of whispered speech compared quite favorably with similar studies which used sound spectrographic processing techniques. Shifting from phonated speech to whispered speech caused a substantial increase in the phonomic formant frequencies and formant bandwidths for both male and female speakers. The data from this study showed no evidence that the LPC-10 vocoder's ability to process voices with pitch extremes and quality extremes was limited in any significant manner. A comparison of the unprocessed natural vowel waveforms and qualities with the synthesized vowel waveforms and qualities revealed almost imperceptible differences. An LPC-10 vocoder's ability to process linguistic and dialectical suprasegmental features such as intonation, rate and stress at low bit rates should be critical issue of concern for future research. Keywords:

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. James D. Mosko	22b. TELEPHONE (Include Area Code) (315) 330-4024 22c. OFFICE SYMBOL RADC (IRAA)

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

ACKNOWLEDGEMENT

The authors extend their gratitude to Captain John Ferrante for his computer wizardry and intellectual stimulation during the course of the investigation. Other colleagues, particularly Mr. Edward J. Cupples, provided daily assistance and encouragement.

One of us, Dr. John Wilson, resided at the Rome Air Development Center under the auspices of the Summer Faculty Program of the Southeastern Center for Electrical Engineering Education and the sponsorship of the Air Force Office of Scientific Research. He thanks both groups for the opportunity in this professional experience.

I. INTRODUCTION

The determination of the most efficacious method of transmitting speech signals in narrowband form has been a major focus of speech researchers and communications engineers back as far as 1928 when Homer Dudley introduced a device called a "vocoder"[1]. The development of efficient speech coding methods continues to be an area of critical concern, especially for the Air Force, because the ability to manipulate speech signals over communications transmission channels through efficient and versatile coding procedures will enhance air-to-air and air-to-ground reconnaissance and surveillance.

Significant advancements have been made in speech coding technology since Dudley's prototype vocoder. Moreover, the narrowband linear predictive coding (LPC) technique has become widely used in both civilian and military applications[2]. However, in spite of its recent widespread use, in spite of the many improvements made in it and in spite of its adoption as the military standard, the LPC vocoder still has not undergone a systematic and comprehensive evaluation for general and practical applications.

The present project was designed to test an LPC-10 vocoder's performance during the processing of speakers whose voices represent vocal extremes in respect to pitch and quality. The impetus for this study emanated from the following general hypotheses and findings reported in recent signal and speech transmission literature:

1. A speaker's voice quality has a clear effect on the subsequent LPC quality[3].

2. The efficacy of LPC analysis has been shown to be dependent on the fundamental frequency of the voiced signal[4].

3. Narrowband LPC notoriously lacks robustness[5].

4. LPC has difficulties with the voices of women and children[6].

5. The presence of whisper and nasality has a negative influence on intelligibility for both male and female LPC quality[7].

Although a few studies have examined the performance of LPC vocoders for design purposes, the "robustness," i.e., their ability to process a variety of voice types and speaking styles during practical user applications is yet to be determined. Information pertaining to the robustness of LPC vocoders is of utmost importance to signal transmissions personnel and to communications engineers because such information should aid in continued development and improvement of speaker independent speech recognition technology.

II. OBJECTIVES

The overall objective of the present project was to determine the robustness of the LPC-10 vocoder by assessing its ability to process voices with pitch and quality extremes. The specific objectives were:

1. To evaluate the ability of the LPC-10 vocoder to process phonated and whispered speech.

2. To evaluate the ability of the LPC-10 vocoder to process voices representing different sexes, i.e., voices with fundamental frequency and formant frequency extremes.

3. To determine the acoustic characteristics, i.e., the formant frequencies, formant bandwidths and formant amplitudes of male and female phonated and whispered vowels during connected speech.

Whispered and phonated voice characteristics were selected as the focus of this study because (1) they represent two extremes of voice production or laryngeal activity[8], (2) one of the concomitants of human fatigue is dysphonia which may range from partial to complete loss of voice, and (3) previous research has indicated the formant frequencies of whispered isolated vowels are higher than those of phonated isolated vowels[9]. Male and female voices were used because the habitual pitch for female voices is almost twice the habitual pitch for male voices.

The LPC technique was of particular concern to this study because, as pointed out earlier, it has been adopted as the military standard. Purportedly, this technique is especially attractive to military communications personnel because (1) only a short segment of speech is needed to yield accurate results, (2) it is suitable for analyzing high pitched voices, such as women's and children's, and (3) data rate can be reduced to approximately 2400 bits/sec. without producing degradation in speech quality[10].

III. METHOD

Subjects

Seven young adult males and four young adult females who demonstrated normal speech and hearing characteristics produced the speech samples for this study. Their ages ranged from 19-32 years and 17-24 years, respectively. The mean age for males was 23.6 years and the mean age for females was 21.3 years. All of the subjects grew up in the northeastern region of the United States and spoke General American English dialect.

Speech Sample

Each of the 11 subjects was instructed to first normally phonate a list of five sentences and secondly to whisper the same list. Embedded in each of the five stimulus sentences was an "h_d" word in which were embedded the experimental vowels: /i/, /ae/, /u/, /a/ and / / (see Figure 1).

These particular vowels were selected because they represent the cardinal or articulatory extremes for English vowel production (see Figure 3).

Instrumentation

The LPC-10 vocoder was the speech processing device used in the present study. The LPC-10 is a time-domain device that analyzes and synthesizes speech using the principles of linear predictive coding. Linear predictive coding is a speech modeling technique which approximates a given speech signal as a linear combination of past samples of a hypothetical input to a system whose output is the given speech signal. The predictor coefficients which become the parameters of the digital analysis filter are determined by minimizing the squared differences between the actual speech samples and the linearly predicted ones.

Since the vocoder analyzes a given speech sample before it reconstructs it from the analysis data, it is possible to remove information redundancy or to compress a speech sample if desired.

It should be noted that the device used in the present study was computer simulation of the LPC-10 vocoder. For testing purposes, the computer simulated version of the vocoder actually allowed more flexibility.

Recording Procedure

Each of the subject's speech samples was stored by recording it on a Uher Model 4000 Report IC reel-to-reel magnetic tape recorder while the subject was seated in an IAC sound attenuating booth. Each subject was allowed to practice the list of sentences under the whispered and phonated conditions until he felt comfortable with his productions. An Uher Model M518 Dynamic Microphone was positioned 5 inches in front of each subject's mouth, slightly below the level of the lips during the recording of the samples. Any samples that the investigator judged as faulty during his monitoring of the recording sessions were repeated by the subject until an acceptable sample was produced.

Analysis Procedure

Prior to analyzing the speech samples, each of the vowel-embedded words was isolated from the carrier sentences by mechanical hand splicing. One foot of leader tape was spliced onto each side of each of the vowel-embedded words so that the vowel samples could be fed continuously into the computer during analysis. During the splicing procedure, the phonated and whispered productions of a given vowel were paired so that they could be subsequently analyzed and displayed together.

Thus each of the 11 subjects had a total of 5 pairs of vowel-embedded words which were fed from the tape recorder output directly into the analog-to-digital preprocessor (see Figure 2). For example, a phonated /i/ and a whispered /i/ were digitized and analyzed together, and so on.

The digitization process consisted of low pass filtering each of the sample pairs at 5000 Hz with a sampling rate of 10000 Hz. Once the sample pairs were digitized, they were stored on a disk as a sampled data file or primary file.

At this point, the linear predictive coding analysis phase was initiated by giving the computer the API command which instructed the simulated vocoder to analyze and model the spectral characteristics of the input data (the digitized sample pair in the sampled data file) using the linear predictive coding method. In addition, the excitation of fundamental frequency status was determined using a modified cepstral processing technique.

Upon completion of the LPC analysis via the API execution the analysis data were stored on a disk and designated as the secondary file or analysis file. From the analysis file, a variety of acoustic parameters could be extracted (See Figure 2) once the appropriate command was given.

The first data analysis output program was the speech spectrogram (SGM). This program computed a frequency spectrum of the speech samples at specified points in time by performing a Fast Fourier Transform (FFT) on the coefficients that were determined during the LPC analysis. When SGM was completed, the formant frequencies, formant bandwidths and formant amplitudes became available in the analysis file.

Following the SGM program, a Formant Tracking (FTR) program was executed which essentially provided a trace of the formant trajectories on the previously completed digital spectrogram (see Figure 5).

At this point, the fundamental frequency, formant frequency, formant amplitude and formant bandwidth data were printed out by activating the Interactive Editing (IAE) mode. Thus, for each of the speech samples, the fundamental frequencies, formant frequencies, formant bandwidths, and formant amplitudes were printed out for 8 frames or points in time (duration of each frame=6.4 ms) from the center of each of the vowel samples.

From the analysis data file, three additional graphic displays of the phonated and whispered speech samples were produced: SPL (Spectral Plot), FPL (Frequency Plot), FDI (Raw Spectral Plot) and VTR (Vocal Tract Plot). Examples of these displays may be observed in Figures 6-10, respectively. The SPL display is a three-dimensional plot of formant frequency and amplitude as functions of time. The FPL display is essentially the same as the SPL display except that the spectrum of each frame is displayed separately allowing a more detailed observation of formant frequency and amplitude variations over time. The FDI display is a raw spectral display of the harmonics as well as the formants for a selected segment of a given sample. A smoothing curve (SSP) may be superimposed on the raw spectrum to show actual formant peaks and peak amplitudes.

The VTR display shows the vocal tract configurations that produced the various formant frequencies at specific points in time.

V. RESULTS

Twenty (20) phonated and 20 whispered speech samples were identified as meeting the target vowel productions for the female subjects while 35 phonated and 35 whispered samples were identified as meeting the target vowel productions for the male subjects. Thus 40 adult female vowels and 70 adult male vowels taken from connected speech provided the analysis data for this study. The mean values obtained from the LPC acoustic analyses are tabulated in Tables 1-9.

Males

An observation of the data in Table 1 indicates that F1 is higher in frequency for all five vowels when they are whispered than when they are phonated. For the whispered vowels, F1 was, on the average, 147 Hz higher than F1 for the phonated vowels.

Table 2 indicates that the F1 bandwidths followed a similar trend. The F1 bandwidths for the whispered vowels were, on the average, 59 Hz wider than the F1 bandwidths for the phonated vowels.

The F1 amplitudes followed a reverse trend as can be observed in Table 3. The F1 amplitudes were, on the average, 9 dB greater in magnitude for the phonated vowels than the whispered vowels. However, Table 3 also indicates a shift in energy toward F3 for the whispered vowels whereas most of the spectral energy was concentrated at F1 and F2 for the phonated vowels. The F3 amplitudes are, on the average, 2 dB greater in magnitude for the whispered vowels than F3 for the phonated vowels. It may also be observed that F1, F2 and F4 amplitudes were greater for the phonated vowels for the male speakers. One of the more salient differences between phonated and whispered vowels then, was the shift of spectral energy toward F3 for whispered vowels. Figures 6, 7 and 10 provide graphic examples of this trend.

The same trends reported for F1 center frequencies for phonated and whispered vowels may be observed for F1, F3 and F4. Formants 2, 3 and 4 for the vowel /i/, however, were higher in frequency for the phonated vowels than for the whispered vowels.

No discernible pattern of difference can be observed for F2, F3 and F4 bandwidths, although the F2 bandwidths were wider for all of the whispered vowels than for the corresponding phonated vowels with the exception of the vowel /u/ (see Table 2).

As shown in Table 3, all of the formant amplitudes were greater for the phonated vowels except the Formant 3 amplitudes. As was noted earlier, the majority of the spectral energy was concentrated at F3 for the whispered vowels.

It can also be observed in Table 1 that the mean fundamental frequency (F_o) for the male phonated vowels was 126 Hz while the mean F_o for the whispered vowels was 0 Hz. The 0 values for the whispered vowels verified that indeed the whispered vowels were not voiced or phonated (whisper is defined as the production of normally phonated sounds without voice[11]). The 126 Hz mean F_o is also very close to the normative value given in the literature for adult male habitual pitch.

Females

An examination of Table 4 indicates that the mean F1 frequencies for females were on the average 167 Hz higher for the whispered vowels than for the phonated vowels. The same trend was observed earlier for the male subjects.

The F1 and F2 bandwidths tabulated in Table 5 were wiser for the whispered than for the phonated /i/, /u/ and /a/ while the opposite trend for the vowels /ae/ and / / is indicated. The F1 bandwidths were on the average 57 Hz wider for the whispered vowels than for the phonated vowels while the F2 bandwidths were on the average 96 Hz wider.

As was observed with the male speakers, F2 center frequencies were higher for all of the whispered vowels than for the phonated vowels with the exception of the vowel /i/ (see Table 4). No discernible pattern of difference was observed for F3 and F4 center frequencies, bandwidths and amplitudes.

The mean fundamental frequency for the female phonated vowels was 217 Hz and the mean F_o for the whispered vowels was 0 (Table 4). As was pointed out for the males, the 0 Hz F_o for the female whispered vowels verified, by definition, the absence of glottal activity in female whispered vowels.

The 217 Hz mean F_o is almost identical to the 220 Hz normative habitual pitch commonly reported in the literature for young adult females.

VI. DISCUSSION AND CONCLUSIONS

A comparison of the male and female phonated and whispered vowel spectra was produced by the LPC-10 vocoder indicated higher F1 center frequencies under the whispered condition than under the phonated condition for all of the experimental vowels. The actual mean differences were 147 Hz and 167 Hz for the males and females, respectively. The identical finding was reported by Kallail and Emanuel (1984). However, the present study used male and female vowels in connected speech wherein the Kallail and Emanuel study used female vowels spoken only in isolation.

The increased F1 center frequency for whispered vowels can be attributed to the greater vertical tongue constriction used when producing whispered vowels. Figures 7 and 8 provide graphic evidence of greater tongue height and greater constriction in the pharyngeal area for whispered vowels than for phonated vowels. It is well known that vowel F1 center frequency is associated with movements of the tongue within the vertical plane of the oral cavity. It is also well known that the frequency of F1 is raised by constriction of the pharynx[12].

A comparison of the adult male phonated vowel formants with those reported by Peterson and Barney[13] (P & B) and Fairbanks and Grubbs[14] indicates very close agreement (see Table 7). The mean differences between the Peterson and Barney male F1, F2 and F3 frequencies and the LPC-10 male phonated F1, F2 and F3 frequencies were 45 Hz, 116 Hz and 53 Hz, respectively.

Similarly, Table 8 shows that adult female phonated vowel formants compared quite favorably with the adult female formants reported by Peterson and Barney. The differences between the P & B adult female F1 frequencies and the LPC adult female phonated F1 frequencies were 71 Hz, 126 Hz and 220 Hz, respectively.

These results would suggest that the linear predictive coding fundamental frequency and formant frequency extraction techniques are adequate when compared with the data obtained from conventional sound spectrographic techniques. It should be pointed out however that the adult females' F3 values as reported by the present study and those reported by Peterson and Barney are different enough to suggest some measurement disparity. The present data did not permit the specification of the direction of the disparity.

In an effort to assess the LPC vocoder's ability to process whispered speech, the LPC measured whispered vowel formants as produced by adult females were compared to the female whispered vowel formants reported by Kallail and Emanuel (K & E) which were measured with the sound spectrograph. These data are presented in Table 9. The mean differences between LPC measured whispered F1, F2 and F3 and the spectrographic whispered F1, F2 and F3 were 89 Hz, 282 Hz and 311 Hz, respectively. Although these results compare favorably, all of the vowel F3 values measured with the sound spectrograph were higher than the F3 values that were obtained with the LPC vocoder. A similar result was noticed earlier when the Peterson and Barney female F3 values were compared with the LPC F3 values. This trend suggests one of the two measurement techniques underestimated or overestimated F3 center frequency values. Additional research is needed to clarify this issue.

In summary, the data obtained from this study suggests the following:

1. The LPC-10 vocoder's performance in processing whispered and phonated speech when compared with conventional sound spectrographic techniques measured up quite favorably for the first two formants in adult male and female speakers. The LPC vocoder's ability to process the higher formants needs further research.

2. The LPC-10 vocoder processed voices with pitch extremes quite well. Its fundamental frequency tracking precision compared very closely with previously reported spectrographic data for male and female vowels.

3. The most salient acoustic difference between phonated and whispered vowels in connected speech was an increase in the frequency of the first three formants, the greatest increase occurring at F1 for males and females.

4. Whispered vowels had wider formant bandwidths than phonated vowels for male and female spoken vowels.

5. For phonated vowels, most of the spectral energy was concentrated at·the first two formants for males and females while the majority of the spectral energy shifted from the lower two formants to the F3 area for male and female whispered vowels.

6. The LPC-10 vocoder's synthesized phonated and whispered vowel waveforms and qualities and unprocessed phonated and whispered vowel waveforms were barely discernible, although the differences between the synthesized and unprocessed whispered vowel waveforms and qualities were slightly more noticeable than the differences between the unprocessed and synthesized phonated vowel waveforms and qualities (see Figure 11).

VII. RECOMMENDATIONS

Although the results of the LPC-10 vocoder's processing of phonated and whispered speech and the data obtained from previous spectrographic analysis compared rather well, there is still a need for further research which systematically tests the LPC vocoder's ability to process more intermediate qualities such as nasality, harshness, stridency, etc.

It would also be exremely important to determine the influence of dialectal differences (regional and foreign) on the processing ability of the LPC-10 vocoder. There is a dearth of research relative to this issue.

Finally, there is a need for research to study the influence of various suprasegmental or prosodic speech features such as intonation, rate and duration on LPC vocoding at different bit rates, particularly at 2400 bits/second and lower.

PHONATED

1. The syllable heed is the word.
2. The syllable had is the word.
3. The syllable who'd is the word.
4. The syllable hod is the word.
5. The syllable hud is the word.

WHISPERED

1. The syllable heed is the word.
2. The syllable had is the word.
3. The syllable who'd is the word.
4. The syllable hod is the word.
5. The syllable hud is the word.

Figure 1. The stimulus material used by each of the 11 subjects to produce the experimental vowel samples. Each of the vowels was produced with the same consonantal environment while the vowel embedded words were contained in the same carrier sentences.

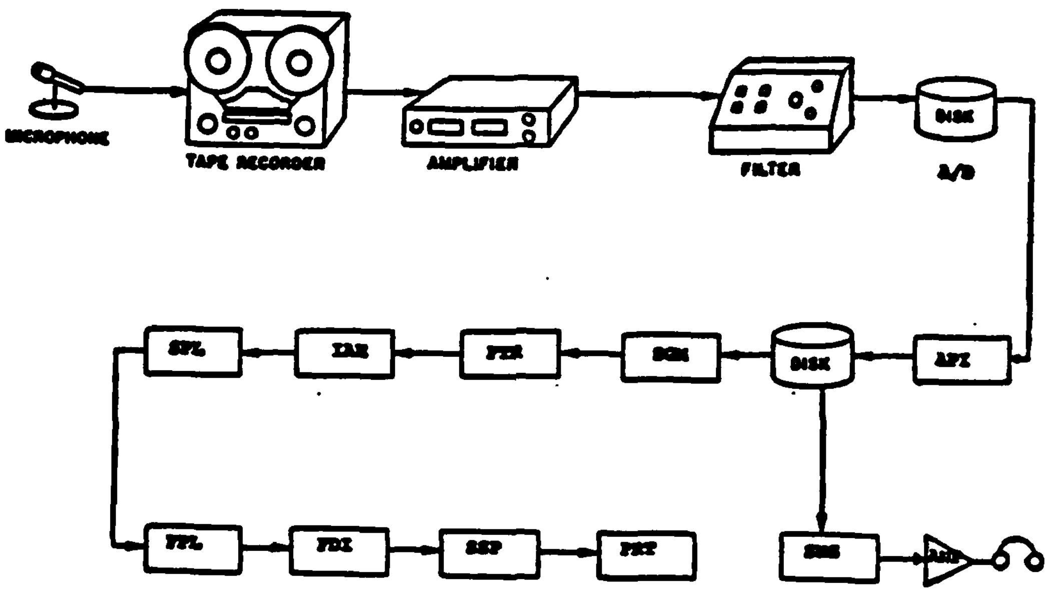


Figure 2. Block diagram of the computer simulation of the LPC-10 Vocoder used to analyze and synthesize the phonated and whispered speech samples.

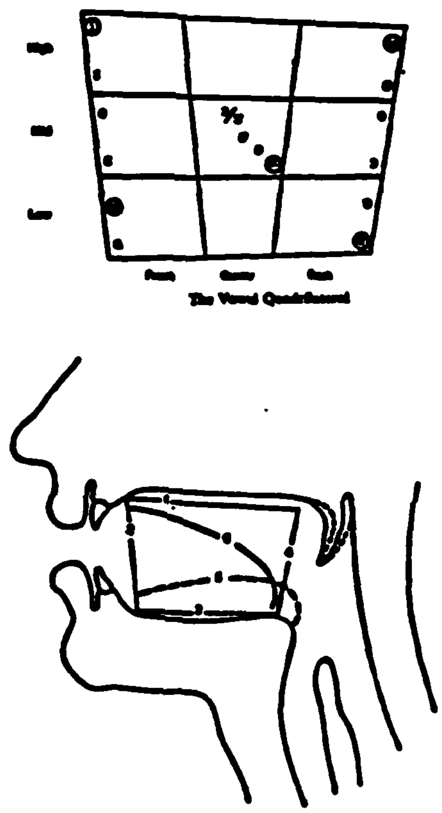


Figure 3. A diagram of the relative tongue positions in the oral cavity used in the General American English production of the five (5) experimental vowels: /i/,/æ/,/u/,/U/,/ʌ/.

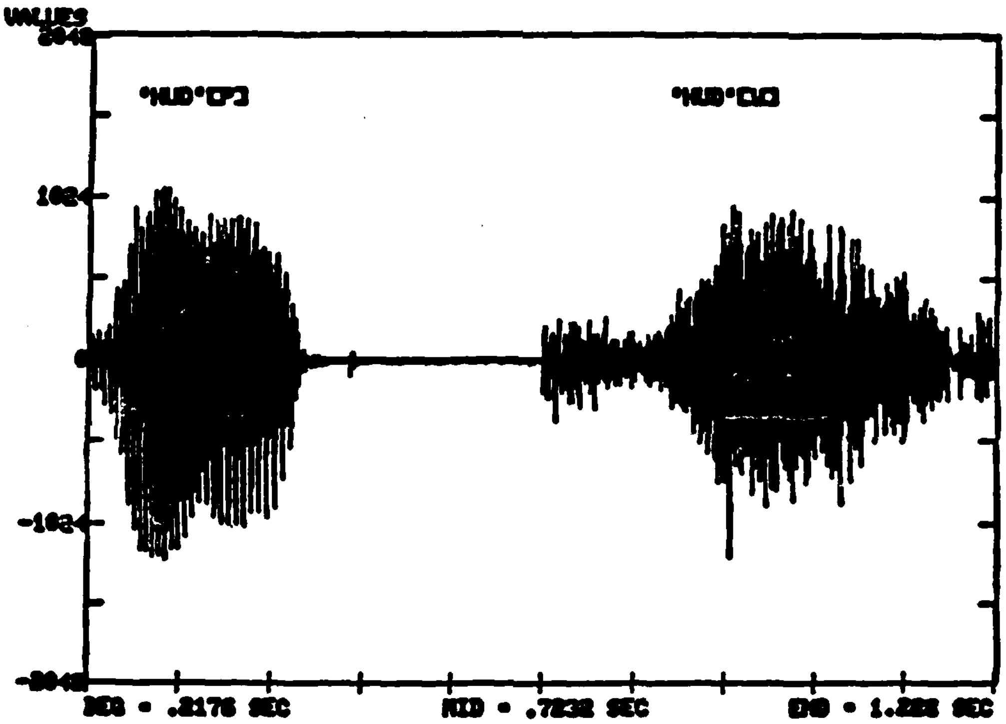


Figure 4. An oscillographic display of an adult male's production of a phonated [P] and whispered [W] vowel /ʌ/ prior to LPC processing. Note that the vowel is embedded in the word "Hud".

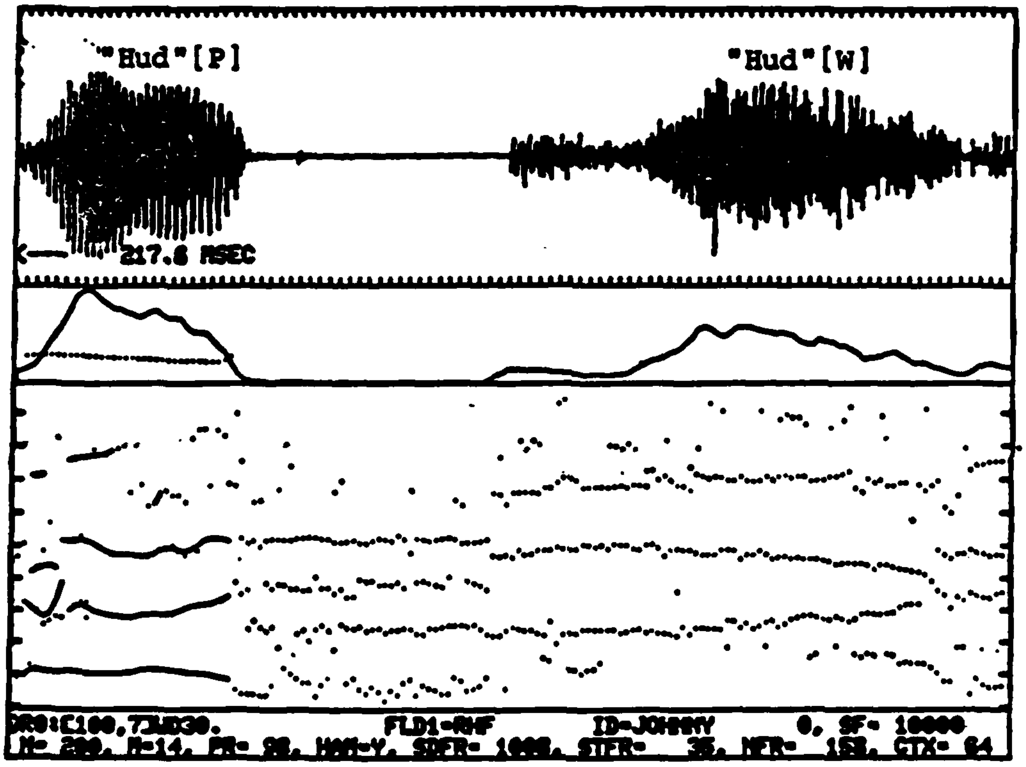
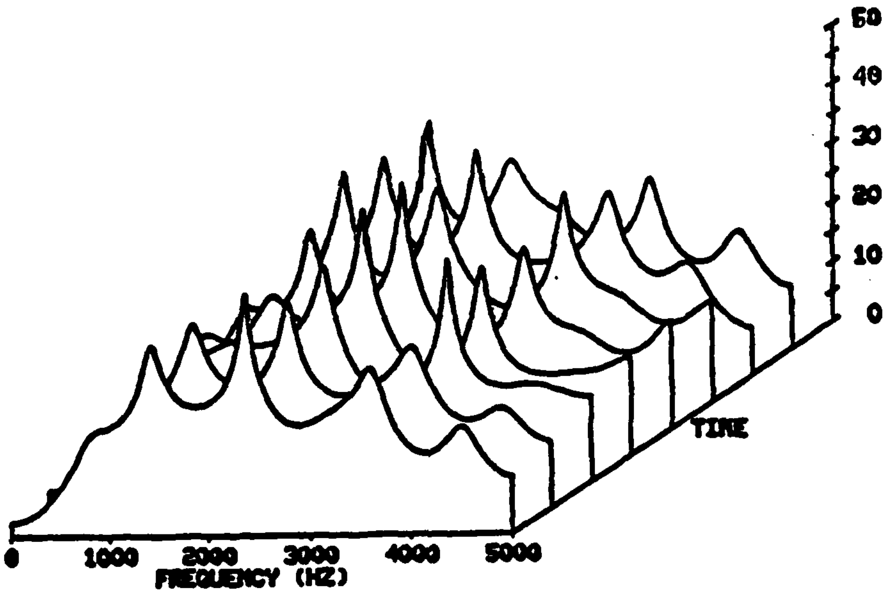
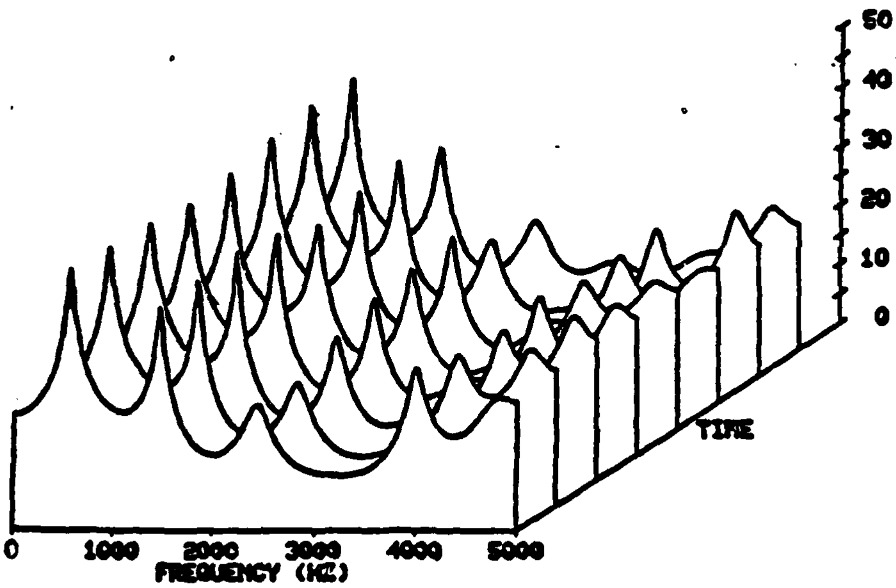


Figure 5. A digital spectrogram [SGM] with formant tracking [FTR] for an adult male speaker's production of a phonated [P] and whispered [W] vowel /ʌ/ embedded in the word "Hud".

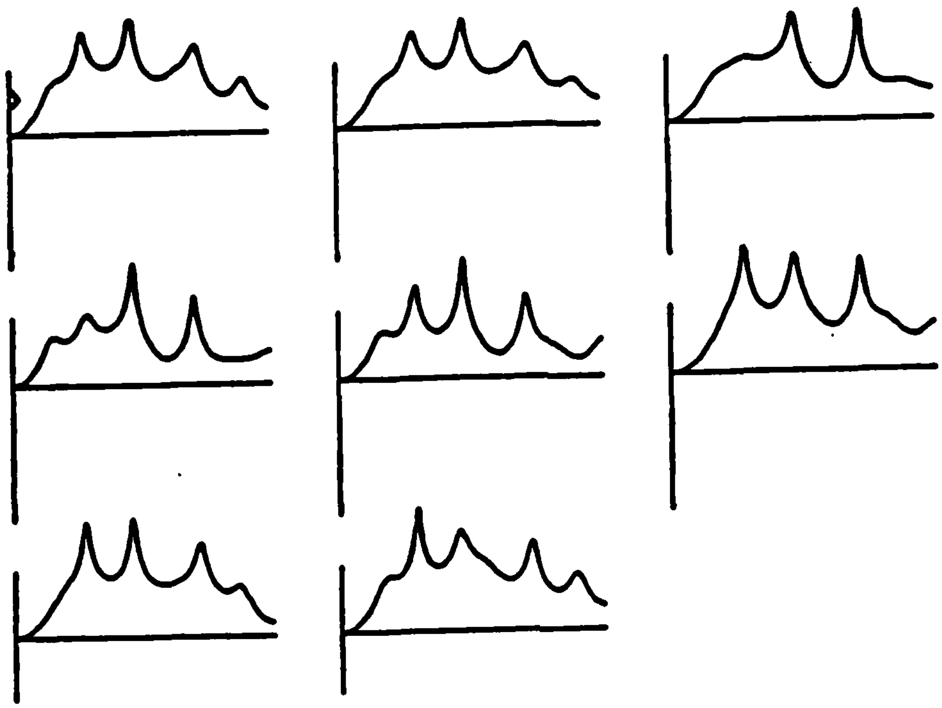


(b)

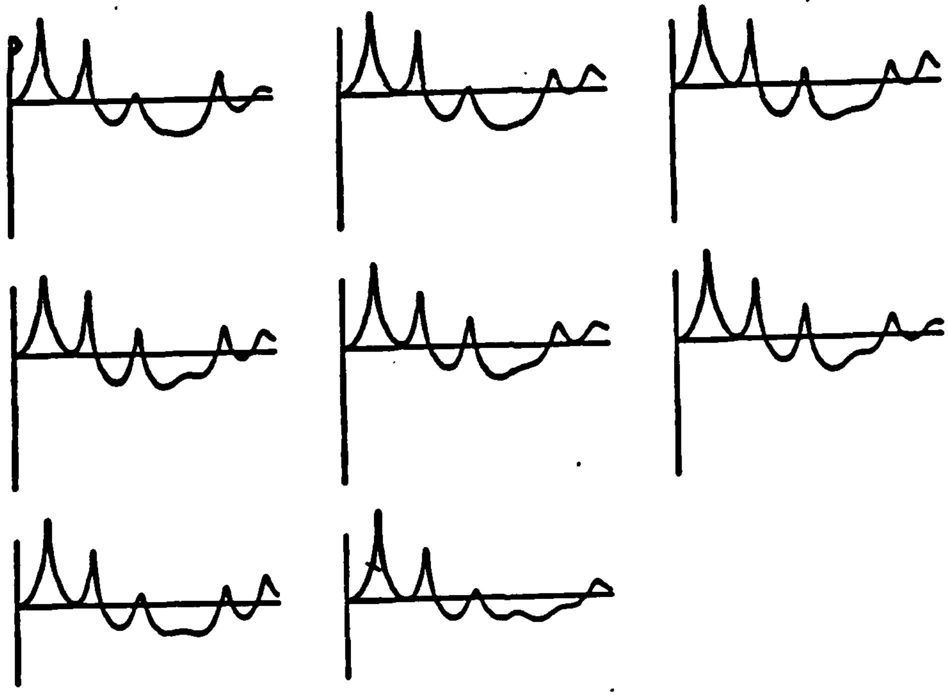


(a)

Figure 6. A 3-dimensional plot [SPL] of 8 frames of formant frequency and amplitude as a function of time for an adult male's (a)phonated and (b)whispered production of the vowel /ʌ/.



(b)



(a)

Figure 7. Frequency plot (FPL) of 8 frames of an adult male speaker's spectral tilt variations during the production of (a) phonated /ʌ/ and /b/ whispered /ʌ/.

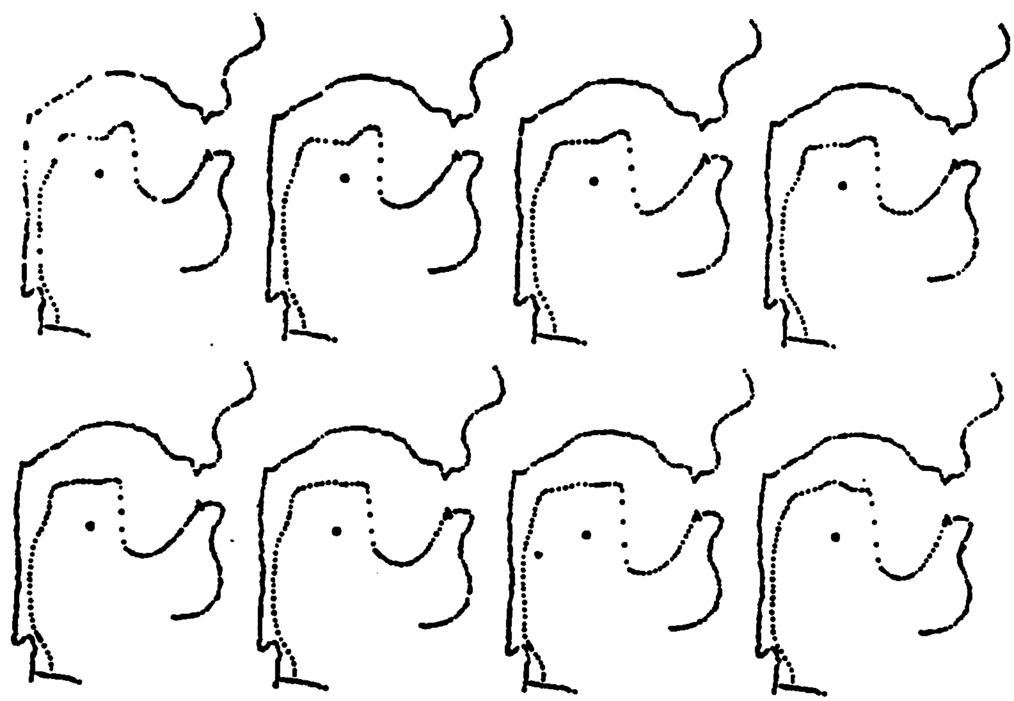


Figure 8. The vocal tract [VTR] configurations over 8 frames during an adult male's phonated production of the vowel /ʌ/.

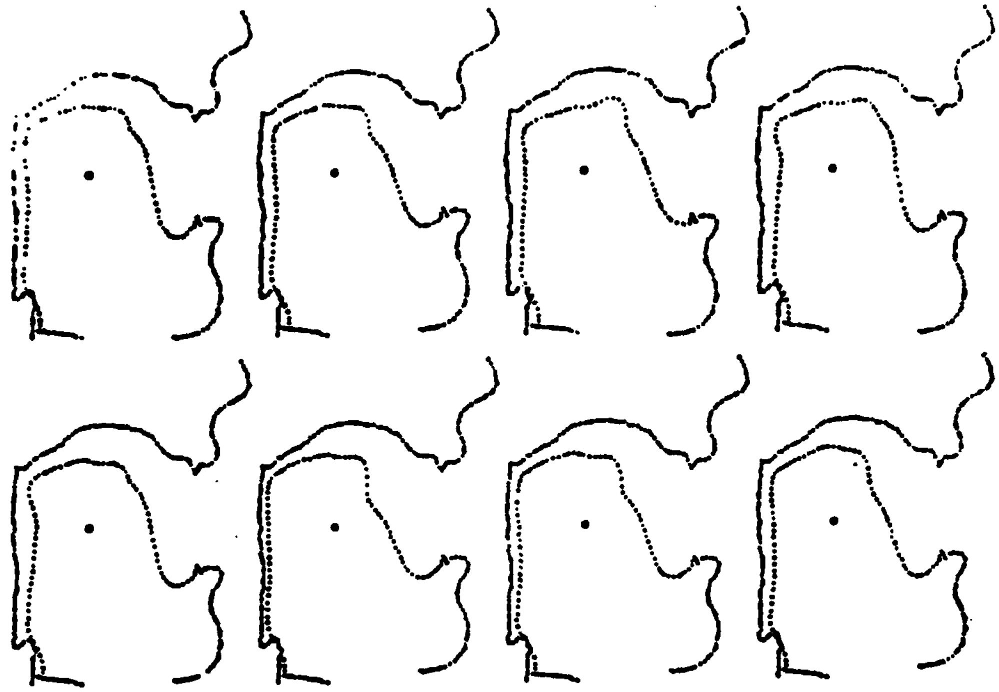
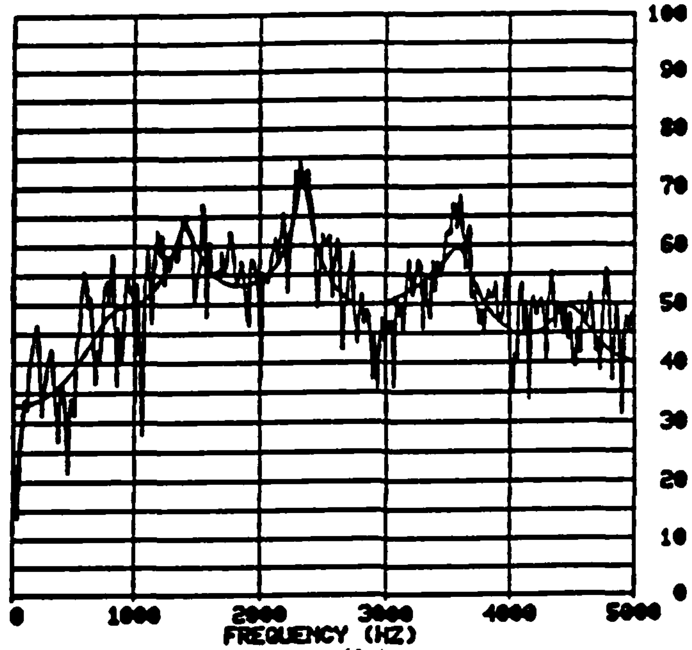
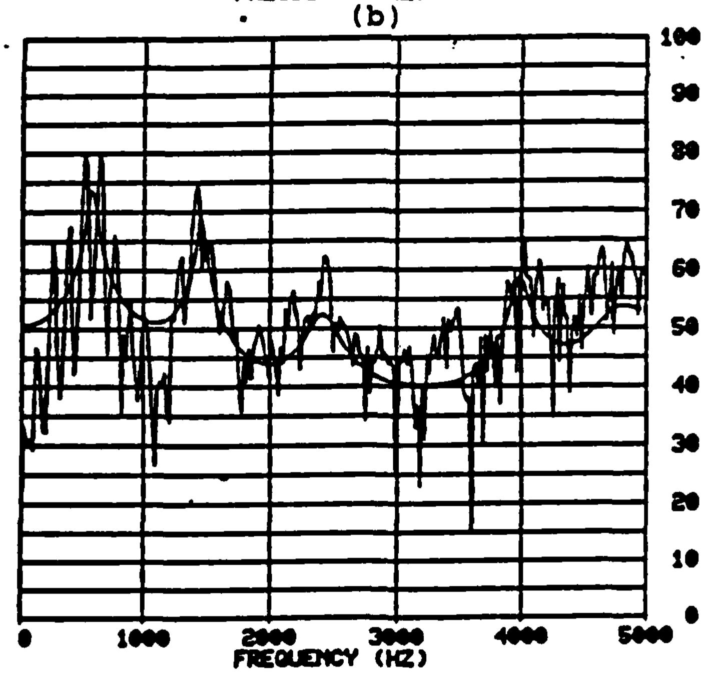


Figure 9. The vocal tract (VTR) configurations over eight frames during an adult male's whispered production of the vowel /ʌ/.



(b)



(a)

Figure10. An FFT raw spectrum (FDI) and smooth spectrum (SSP) of an adult male's (a) phonated and (b) whispered production of the vowel /ʌ/.

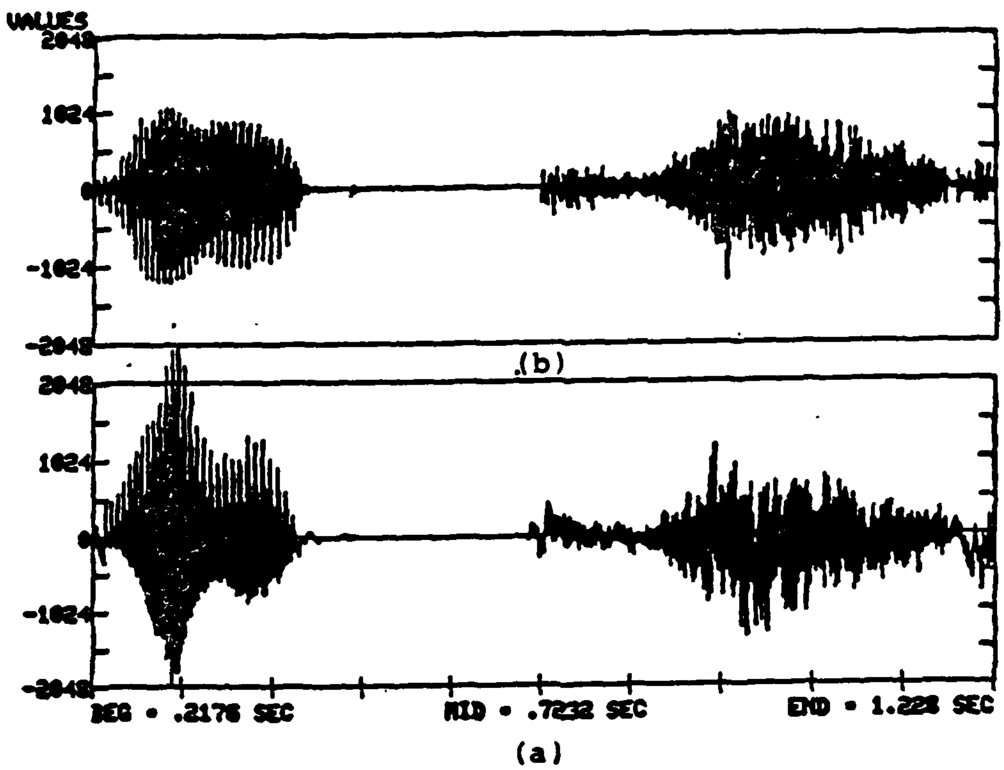


Figure 11. An oscillographic display of an adult male's (a) natural production of the phonated and whispered vowel /ʌ/ and (b) the synthesized waveform of the adult male's production of the phonated and whispered vowel /ʌ/.

Table 1. Average formant frequency and fundamental frequency values for male phonated and whispered productions of the experimental vowels. Standard deviations are given in parentheses below each of the mean values. Negative differences indicate that the whispered values are larger than the phonated values.

	n=7	*F1	F2	F3	F4	F_0	
PHONATED	$\bar{X}$	337	2341	3045	3797	132	[i]
	S.D.	(34)	(153)	(285)	(369)	(18)	
WHISPERED	$\bar{X}$	458	2061	2945	3585	0	
	S.D.	(98)	(392)	(230)	(227)		
DIFFERENCE		-121	280	100	212	132	
PHONATED	$\bar{X}$	620	1753	2495	3613	123	[æ]
	S.D.	(41)	(262)	(209)	(213)	(13)	
WHISPERED	$\bar{X}$	779	1869	2624	3655	0	
	S.D.	(108)	(208)	(272)	(431)		
DIFFERENCE		-159	-116	-129	-42	123	
PHONATED	$\bar{X}$	359	983	2250	3366	135	[u]
	S.D.	(35)	(150)	(206)	(291)	(14)	
WHISPERED	$\bar{X}$	419	1125	2478	3419	0	
	S.D.	(155)	(147)	(156)	(188)		
DIFFERENCE		-60	-142	-228	-53	135	
PHONATED	$\bar{X}$	721	1325	2431	3517	116	[ɑ]
	S.D.	(78)	(190)	(147)	(335)	(10)	
WHISPERED	$\bar{X}$	905	1518	2461	3610	0	
	S.D.	(76)	(141)	(169)	(186)		
DIFFERENCE		-184	-193	-30	-93	116	
PHONATED	$\bar{X}$	588	1340	2517	3587	123	[ʌ]
	S.D.	(35)	(96)	(158)	(181)	(10)	
WHISPERED	$\bar{X}$	799	1418	2565	3609	0	
	S.D.	(52)	(114)	(205)	(291)		
DIFFERENCE		-211	-78	-48	-22	123	

F1=Formant 1 F3=Formant 3
F2=Formant 2 F4=Formant 4
F_0=Fundamental Frequency

*All formant and fundamental frequency values are in Hertz (Hz).

Table 2. Average formant bandwidth values for male phonated and whispered productions of the experimental vowels. Standard deviations are given in parentheses below each of the mean values. Negative differences indicate that the whispered values are greater than the phonated values.

	n=7	*B1	B2	B3	B4	
PHONATED	X̄	138	158	251	270	[i]
	S.D.	(19)	(22)	(57)	(111)	
WHISPERED	X̄	237	262	246	300	
	S.D.	(65)	(102)	(83)	(51)	
DIFFERENCE		-99	-104	5	-30	

PHONATED	X̄	142	167	223	248	[æ]
	S.D.	(22)	(63)	(64)	(82)	
WHISPERED	X̄	170	232	235	292	
	S.D.	(43)	(107)	(68)	(67)	
DIFFERENCE		-28	-65	-12	-44	

PHONATED	X̄	130	181	248	355	[u]
	S.D.	(15)	(39)	(108)	(183)	
WHISPERED	X̄	232	179	295	246	
	S.D.	(73)	(39)	(152)	(79)	
DIFFERENCE		-102	2	-47	109	

PHONATED	X̄	144	190	226	347	[ɑ]
	S.D.	(15)	(68)	(103)	(120)	
WHISPERED	X̄	171	193	207	294	
	S.D.	(24)	(28)	(58)	(73)	
DIFFERENCE		-27	-3	19	53	

PHONATED	X̄	132	197	203	298	[ʌ]
	S.D.	(6)	(71)	(60)	(131)	
WHISPERED	X̄	171	210	187	238	
	S.D.	(19)	(40)	(45)	(69)	
DIFFERENCE		-39	-13	16	60	

B1=Bandwidth 1 B3=Bandwidth 3
B2=Bandwidth 2 B4=Bandwidth 4

***All bandwidth values are in Hertz (Hz).**

Table 3. Average formant amplitude values for male phonated and whispered productions of the experimental vowels. Standard deviations are given in parentheses below each of the mean values. Negative differences indicate that the whispered values are greater than the phonated values.

Vowel (n=7)		*A1	A2	A3	A4
[i] PHONATED	X̄	18	14	11	11
	S.D.	(5)	(2)	(4)	(4)
WHISPERED	X̄	10	13	12	10
	S.D.	(4)	(3)	(6)	(3)
DIFFERENCE		8	1	-1	1
[æ] PHONATED	X̄	19	15	10	8
	S.D.	(4)	(3)	(2)	(3)
WHISPERED	X̄	11	14	13	6
	S.D.	(3)	(2)	(3)	(2)
DIFFERENCE		8	1	-3	2
[u] PHONATED	X̄	23	16	8	8
	S.D.	(9)	(3)	(3)	(2)
WHISPERED	X̄	14	15	8	8
	S.D.	(6)	(3)	(2)	(2)
DIFFERENCE		9	1	0	0
[ɑ] PHONATED	X̄	24	17	8	8
	S.D.	(4)	(5)	(3)	(3)
WHISPERED	X̄	13	16	11	7
	S.D.	(5)	(4)	(5)	(2)
DIFFERENCE		11	1	-3	1
[ʌ] PHONATED	X̄	23	14	8	8
	S.D.	(2)	(3)	(2)	(2)
WHISPERED	X̄	12	12	10	7
	S.D.	(3)	(2)	(3)	(3)
DIFFERENCE		11	2	-2	1

A1=Amplitude 1 A3=Amplitude 3
A2=Amplitude 2 A4=Amplitude 4

*All amplitude values are in decibels (dB).

Table 4. Average formant frequency and fundamental frequency values for female phonated and whispered productions of the experimental vowels. Standard deviations are given in parentheses below each of the mean values. Negative differences indicate that the whispered values are greater than the phonated values.

[i]

	n=4	*F1	F2	F3	F4	F_0
PHONATED	$\bar{X}$	386	2787	3592	4020	226
	S.D.	(104)	(305)	(415)	(643)	(25)
WHISPERED	$\bar{X}$	570	1640	2789	3537	0
	S.D.	(107)	(376)	(273)	(161)	
DIFFERENCE		-184	1147	803	483	226

[æ]

		*F1	F2	F3	F4	F_0
PHONATED	$\bar{X}$	773	1936	2804	3681	209
	S.D.	(70)	(209)	(114)	(427)	(18)
WHISPERED	$\bar{X}$	896	2034	2889	3702	0
	S.D.	(87)	(334)	(345)	(319)	
DIFFERENCE		-123	-98	-85	-21	209

[u]

		*F1	F2	F3	F4	F_0
PHONATED	$\bar{X}$	432	1164	2434	3725	231
	S.D.	(46)	(319)	(187)	(243)	(23)
WHISPERED	$\bar{X}$	526	1294	2536	3597	0
	S.D.	(155)	(215)	(240)	(194)	
DIFFERENCE		-94	-130	-102	128	231

[ɑ]

		*F1	F2	F3	F4	F_0
PHONATED	$\bar{X}$	842	1442	2778	3864	201
	S.D.	(43)	(232)	(209)	(206)	(16)
WHISPERED	$\bar{X}$	999	1557	2561	3541	0
	S.D.	(62)	(76)	(76)	(255)	
DIFFERENCE		-157	-115	217	323	201

[ʌ]

		*F1	F2	F3	F4	F_0
PHONATED	$\bar{X}$	662	1321	2276	3273	218
	S.D.	(89)	(421)	(609)	(660)	(18)
WHISPERED	$\bar{X}$	941	1627	2676	3712	0
	S.D.	(42)	(214)	(415)	(365)	
DIFFERENCE		-279	-306	-400	-439	218

F1=Formant 1 F3=Formant 3
F2=Formant 2 F4=Formant 4
F_0=Fundamental Frequency

*All formant and fundamental frequency values are in Hertz (Hz).

Table 5. Average formant bandwidth values for female phonated and whispered productions of the experimental vowels. Standard deviations are given in parentheses below each of the mean values. Negative differences indicate that the whispered values are greater than the phonated values.

Vowel	n=4		*B1	B2	B3	B4
[i]	PHONATED	$\bar{X}$	132	172	348	228
		S.D.	(15)	(70)	(205)	(91)
	WHISPERED	$\bar{X}$	208	462	235	229
		S.D.	(40)	(50)	(51)	(18)
	DIFFERENCE		-76	-290	113	-1
[æ]	PHONATED	$\bar{X}$	174	237	191	351
		S.D.	(46)	(109)	(44)	(56)
	WHISPERED	$\bar{X}$	171	269	227	350
		S.D.	(30)	(51)	(35)	(97)
	DIFFERENCE		3	-32	-36	1
[u]	PHONATED	$\bar{X}$	122	167	344	245
		S.D.	(7)	(20)	(142)	(110)
	WHISPERED	$\bar{X}$	261	213	371	348
		S.D.	(104)	(74)	(14)	(81)
	DIFFERENCE		-139	-46	-27	-103
[ɑ]	PHONATED	$\bar{X}$	177	203	347	395
		S.D.	(34)	(71)	(128)	(196)
	WHISPERED	$\bar{X}$	230	251	239	267
		S.D.	(32)	(42)	(20)	(9)
	DIFFERENCE		-53	-48	108	128
[ʌ]	PHONATED	$\bar{X}$	172	192	182	221
		S.D.	(57)	(60)	(49)	(45)
	WHISPERED	$\bar{X}$	159	256	256	240
		S.D.	(21)	(17)	(29)	(28)
	DIFFERENCE		13	-64	-74	-19

B1=Bandwidth 1 B3=Bandwidth 3
B2=Bandwidth 2 B4=Bandwidth 4

*All bandwidth values are in Hertz (Hz).

Table 6. Average formant amplitude values for female phonated and whispered productions of the experimental vowels. Standard deviations are given in parentheses below each of the mean values. Negative differences indicate that the whispered values are greater than the phonated values.

	n=4	*A1	A2	A3	A4	
PHONATED	$\bar{X}$	20	15	11	9	[i]
	S.D.	(8)	(4)	(4)	(7)	
WHISPERED	$\bar{X}$	7	10	11	13	
	S.D.	(2)	(2)	(1)	(3)	
DIFFERENCE		13	5	0	-4	
PHONATED	$\bar{X}$	18	13	8	7	[æ]
	S.D.	(4)	(6)	(2)	(4)	
WHISPERED	$\bar{X}$	15	10	11	6	
	S.D.	(5)	(2)	(3)	(2)	
DIFFERENCE		3	3	-3	1	
PHONATED	$\bar{X}$	27	14	9	9	[u]
	S.D.	(7)	(5)	(1)	(2)	
WHISPERED	$\bar{X}$	14	16	6	9	
	S.D.	(5)	(4)	(1)	(1)	
DIFFERENCE		13	-2	3	0	
PHONATED	$\bar{X}$	20	17	6	8	[ɑ]
	S.D.	(3)	(4)	(4)	(2)	
WHISPERED	$\bar{X}$	12	14	10	6	
	S.D.	(5)	(2)	(2)	(2)	
DIFFERENCE		8	3	-4	2	
PHONATED	$\bar{X}$	19	16	10	7	[ʌ]
	S.D.	(3)	(3)	(4)	(2)	
WHISPERED	$\bar{X}$	18	13	8	5	
	S.D.	(5)	(5)	(3)	(2)	
DIFFERENCE		1	3	2	2	

A1=Amplitude 1 A3=Amplitude 3
A2=Amplitude 2 A4=Amplitude 4

*All amplitude values are in decibels (dB).

Table 7. The mean formant frequencies and fundamental frequencies for adult male phonated vowels as determined by the LPC vocoder and the mean formant frequencies and fundamental frequencies for adult male vowels reported by Peterson and Barney (P & E) as measured with a sound spectrograph.

Vowel	Source		F1	F2	F3	F_o
[i]	LPC	$\bar{x}$ n=7	337	2341	3045	132
	P & B	$\bar{x}$ n=33	270	2290	3010	136
	Difference		67	51	35	-4
[æ]	LPC	$\bar{x}$ n=7	620	1753	2495	123
	P & B	$\bar{x}$ n=33	660	1720	2410	127
	Difference		-40	33	85	-4
[u]	LPC	$\bar{x}$ n=7	359	983	2250	135
	P & B	$\bar{x}$ n=33	300	870	2240	141
	Difference		59	113	10	-6
[ɑ]	LPC	$\bar{x}$ n=7	721	1325	2431	116
	P & B	$\bar{x}$ n=33	730	1090	2440	124
	Difference		-9	235	-9	-8
[ʌ]	LPC	$\bar{x}$ n=7	588	1340	2517	123
	P & B	$\bar{x}$ n=33	640	1190	2390	130
	Difference		-52	150	127	-7

F1=Formant 1 F3=Formant 3
F2=Formant 2 F_o=Fundamental Frequency

*All formant frequency values are in Hertz (Hz).

Table 8. The mean formant frequencies and fundamental frequencies for adult female phonated vowels as measured with the LPC vocoder and the mean formant frequencies and fundamental frequencies for adult female phonated vowels as reported by Peterson and Barney (P & B) as measured with a sound spectrograph.

		F1	F2	F3	Fo	
LPC	$\bar{X}$	386	2787	3592	226	[i]
P & B	$\bar{X}$	310	2790	3310	235	
Difference		76	-3	282	-9	
LPC	$\bar{X}$	773	1936	2804	209	[æ]
P & B	$\bar{X}$	660	2050	2850	210	
Difference		113	-114	-46	-1	
LPC	$\bar{X}$	432	1164	2434	231	[u]
P & B	$\bar{X}$	370	950	2670	231	
Difference		62	214	-236	0	
LPC	$\bar{X}$	842	1442	2778	201	[ɑ]
P & B	$\bar{X}$	850	1220	2810	212	
Difference		-8	222	-32	-11	
LPC	$\bar{X}$	662	1321	2276	218	[^]
P & B	$\bar{X}$	760	1400	2780	221	
Difference		-98	-79	-504	-3	

F1=Formant 1 F3=Formant 3
F2=Formant 2 F4=Formant 4

*All formant frequency values are in Hertz (Hz).

Table 9. Mean formant frequencies for adult female whispered vowels reported by Kallail and Emanuel (K & E) and the mean formant frequencies for adult female whispered vowels as measured with the LPC vocoder. Negative differences indicate that the LPC values are greater than the K & E values.

		F1	F2	F3	
LPC	$\bar{X}$ n=20	570	1640	2789	[i]
K & E	$\bar{X}$ n=4	423	2778	3232	
Difference		147	-1138	-443	
LPC	$\bar{X}$ n=20	896	2034	2889	[æ]
K & E	$\bar{X}$ n=4	1057	2006	2989	
Difference		-161	28	-100	
LPC	$\bar{X}$ n=20	526	1294	2536	[u]
K & E	$\bar{X}$ n=4	478	1393	2853	
Difference		48	-99	-317	
LPC	$\bar{X}$ n=20	999	1557	2561	[ɑ]
K & E	$\bar{X}$ n=4	1073	1411	2959	
Difference		-74	146	-398	
LPC	$\bar{X}$ n=20	941	1627	2676	[ʌ]
K & E	$\bar{X}$ n=4	956	1624	2971	
Difference		-15	3	-295	

F1=Formant 1 F3=Formant 3
F2=Formant 2 F4=Formant 4

*All formant frequency values are in Hertz (Hz).

REFERENCES

1. Schroeder, M.R., "Vocoders: Analysis and Synthesis," Proc. IEEE, Vol. 54, pp. 720-734, 1966.

2. Kang, G.S. and Everett, S., "Improvement of the LPC Analysis," ICASSP, Boston, pp. 89-92, 1983.

3. Kahn, M. and Garst, P., "The Effects of Five Voice Characteristics on LPC Quality," ICASSP, pp. 531-534, 1983.

4. Hermansky, H., H. Fujisak and Sato, Y., "Analysis and Synthesis of Speech Based on Spectral Transform Linear Predictive Method," ICAASP, pp. 777-780, 1983.

5. Kang, G.S. and Everett, S., "Improvement of the LPC Analysis," ICASSP, Boston, pp. 89-92, 1983.

6. Berney, C.L. and Harshman, C., "Voiceware Does It Differently," Mini-Micro System, pp. 1-6, 1982.

7. Kahn, M. and Garst, P., "The Effects of Five Voice Characteristics on LPC Quality," ICASSP,. pp. 531-534, 1983.

8. Lashley, C.O. Vibratory Action of the Vocal Folds During Whisper. Unpublished Master's Thesis, University of Florida, 1984.

9. Kallail, K.J. and Emanuel, F.W., "Formant-Frequency Differences Between Isolated Whispered and Phonated Vowel Samples Produced by Adult Female Subjects," J. Speech and Hearing Res., Vol. 27, pp 245-251, 1984.

10. Atal, B.S. and Hanauer, S.L., "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acous. Soc. Amer., Vol. 50, pp. 637-655, 1971.

11. Lashley, C.O., Vibratory Action of the Vocal Folds During Whisper. Unpublished Master's Thesis, University of Florida, 1984.

12. Pickett, J.M., The Sounds of Speech Communication. Baltimore, University Park Press, 1980.

13. Peterson, G.E. and Barney, H.L., "Control Methods Used in the Study of Vowels," J. Acoust. Soc. Amer., Vol. 24, pp. 175-184, 1952.

REFERENCES (Continued)

14. Fairbanks, G. and Grubbs, P., "A Psychophysical Investigation of Vowel Formants," J. Speech and Hearing Dis., Vol. 4, pp. 203-219, 1961.

15. Kallail, K.J. and Emanuel, F.W., "Formant-Frequency Differences Between Isolaed Whispered and Phonated Vowel Samples Produced by Adult Female Subjects," J. Speech and Hearing Dis., Vol. 27, pp. 245-251, 1984.

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control, Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of C^3I systems. The areas of technical competence include communications, command and control, battle management, information processing, surveillance sensors, intelligence data collection and handling, solid state sciences, electromagnetics, and propagation, and electronic, maintainability, and compatibility.

END

FILMED

2-86

DTIC