Software for a cascade/parallel formant synthesizer

Dennis H. Klatt

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 1 July 1979; accepted for publication 12 October 1979)

A software formant synthesizer is described that can generate synthetic speech using a laboratory digital computer. A flexible synthesizer configuration permits the synthesis of sonorants by either a cascade or parallel connection of digital resonators, but frication spectra must be synthesized by a set of resonators connected in parallel. A control program lets the user specify variable control parameter data, such as formant frequencies as a function of time, as a sequence of ⟨time, value⟩ points. The synthesizer design is described and motivated in Secs. I–III, and FORTRAN listings for the synthesizer and control program are provided in an appendix. Computer requirements and necessary support software are described in Sec. IV. Strategies for the imitation of any speech utterance are described in Sec. V, and suggested values of control parameters for the synthesis of many English sounds are presented in tabular form.

PACS numbers: 43.70.Jt, 43.70.Qa

INTRODUCTION

A need exists in psychology and the speech sciences for a flexible research tool in order to study speech perception through the synthesis of speech and speech-like sounds. Since most perceptual experiments are now performed under control of a general purpose digital computer, there are advantages to the use of the same laboratory computer for generating speech-like stimuli (rather than the purchase of special-purpose hardware).

The cascade/parallel formant synthesizer to be described can be simulated on a general-purpose digital computer in the manner depicted in Fig. 1. Synthesizer control parameter data such as the frequency motions of the first formant as a function of time are specified by the experimenter, using the synthesizer control program HANDSY.FOR. As many as 20 control parameters may be varied as a function of time to serve as input to the waveform generating synthesizer subroutines PARCOE.FOR and COEWAV.FOR. These three FORTRAN programs appear in Appendix B. Output waveform samples are computed in nonreal time and stored on a disk for subsequent playback through a digital-to-analog converter, analog low-pass filter, and loudspeaker.

Software simulation versus hardware construction

The advantages of a software implementation over the construction of analog hardware are substantial. The synthesizer does not need repeated calibration, it is stable, and the signal-to-noise ratio (quantization noise in the case of a digital simulation) can be made as large as desired. The configuration can easily be changed as new ideas are proposed. For example, the voices of women and children can be synthesized with appropriate modifications to the voicing source and cascade vocal tract configuration. Display terminals are usually available in a computer facility and can be programmed to view control parameter data or selected portions of the output speech waveform (see Sec. IV). Short-time spectra can also be computed and displayed in order to make detailed spectral comparisons between natural and synthetic waveforms (see Sec. V).

Formant synthesis versus articulatory synthesis

Speech synthesizers fall into two broad categories: (1) articulatory synthesizers that attempt to model faithfully the mechanical motions of the articulators and the resulting distributions of volume velocity and sound pressure in the lungs, larynx, and vocal and nasal tracts (Flanagan, Ishizaka, and Shipley, 1975), and (2) formant synthesizers which derive an approximation to

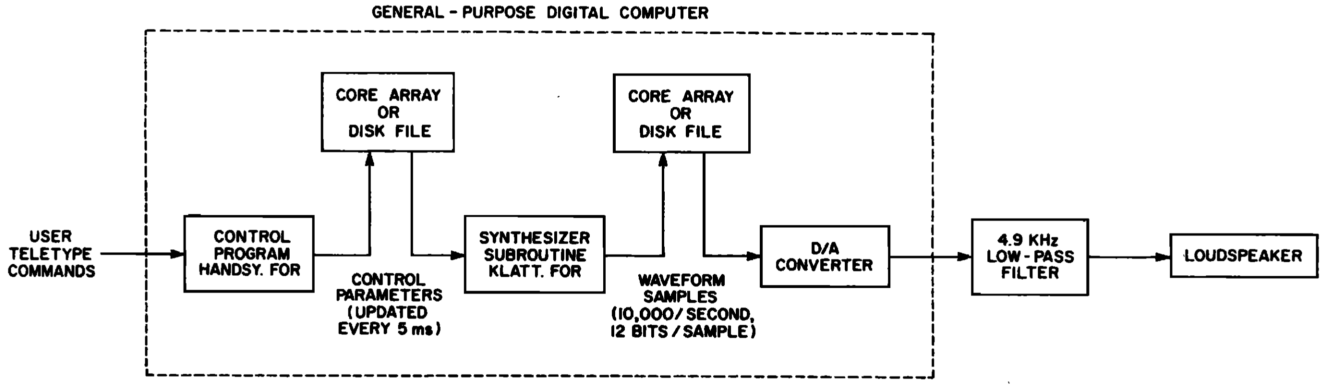
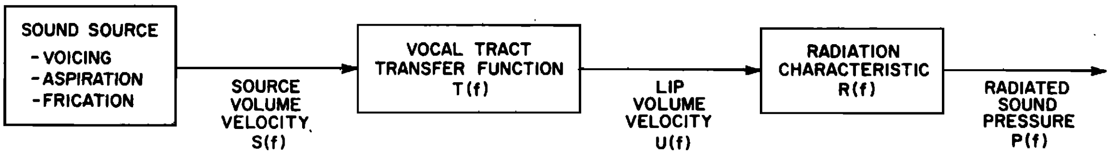

GENERAL - PURPOSE DIGITAL COMPUTER
USER TELETYPE COMMANDS
CONTROL PROGRAM HANDSY. FOR
CORE ARRAY OR DISK FILE
CONTROL PARAMETERS (UPDATED EVERY 5 ms)
SYNTHESIZER SUBROUTINE KLATT. FOR
CORE ARRAY OR DISK FILE
WAVEFORM SAMPLES (10,000/ SECOND, 12 BITS / SAMPLE)
D/A CONVERTER
4.9 KHz LOW - PASS FILTER
LOUDSPEAKER

FIG. 1. Relation of the software synthesizer to the hardware and supporting software of a small general-purpose digital computer.

0001-4966/80/030971-25$00.80


SOUND SOURCE
- VOICING
- ASPIRATION
- FRICATION
SOURCE VOLUME VELOCITY S(f)
VOCAL TRACT TRANSFER FUNCTION T(f)
LIP VOLUME VELOCITY U(f)
RADIATION CHARACTERISTIC R(f)
RADIATED SOUND PRESSURE P(f)

$$P(f) = S(f) * T(f) * R(f)$$

FIG. 2. The output spectrum of a speech sound, $P(f)$, can be represented in the frequency domain as a product of a source spectrum $S(f)$, a vocal tract transfer function, $T(f)$, and a radiation characteristic, $R(f)$.

a speech waveform by a simpler set of rules formulated in the acoustic domain. The present paper is concerned only with formant models of speech generation since current articulartory models require several orders of magnitude more computation, and the resultant speech output cannot be specified with sufficient precision for psychophysical experimentation.

The synthesizer design is based on an acoustic theory of speech production presented in Fant (1960), and is summarized in Fig. 2. According to this view, one or more sources of sound energy are activated by the buildup of lung pressure. Treating each sound source separately, we may characterize it in the frequency domain by a source spectrum, $S(f)$, where f is frequency in Hz. Each sound source excites the vocal tract which acts as a resonating system analogous to an organ pipe. Since the vocal tract is a linear system, it can be characterized in the frequency domain by a linear transfer function, $T(f)$, which is a ratio of lip-plus-nose volume velocity, $U(f)$, to source input, $S(f)$. Finally, the spectrum of the sound pressure that would be recorded some distance from the lips of the talker, $P(f)$, is related to lip-plus-nose volume velocity, $U(f)$, by a radiation characteristic, $R(f)$, that describes the effects of directional sound propagation from the head.

Each of the above relations can also be recast in the time (waveform) domain. This is actually how a waveform is generated in the computer. The synthesizer includes components to simulate the generation of several different kinds of sound sources (described in Sec. I), components to simulate the vocal tract transfer function (Sec. II), and a component to simulate sound radiation from the head (Sec. III).

Cascade versus parallel

A number of hardware and software speech synthesizers have been described (Dudley, Riesz, and Watkins, 1939; Cooper, Liberman, and Borst, 1951; Lawrence, 1953; Stevens, Bastide, and Smith, 1955; Fant, 1959; Fant and Martony, 1962; Flanagan, Coker, and Bird, 1962; Holmes, Mattingly, and Shearme, 1964; Epstein, 1965; Tomlinson, 1966; Scott, Glace, and Mattingly, 1966; Liljencrants, 1968; Rabiner *et al.*, 1971; Klatt, 1972; Holmes, 1973). They employ different configurations to achieve what is hopefully the same result: high-quality approximation to human speech. A few of the synthesizers have stability and calibration problems, and a few have design deficiencies that make it impossible to synthesize a good voiced fricative, but many others have an excellent design. Of the best synthesizers that have been proposed, two general configurations are common.

In one type of configuration, called a parallel formant synthesizer (see, e.g., Lawrence, 1953; Holmes, 1973), the formant resonators that simulate the transfer function of the vocal tract are connected in parallel, as shown in the lower portion of Fig. 3. Each formant resonator is preceded by an amplitude control that determines the relative amplitude of a spectral peak (formant) in the output spectrum for both voiced and voiceless speech sounds. In the second type of configuration, called a cascade formant synthesizer (see, e.g., Fant, 1959; Klatt, 1972), sonorants are synthesized using a set of formant resonators connected in cascade, as shown in the upper part of Fig. 3.

The advantage of the cascade connection is that the relative amplitudes of formant peaks for vowels come out just right (Fant, 1956) without the need for individual amplitude controls for each formant. The disadvantage is that one still needs a parallel formant configuration for the generation of fricatives and plosive bursts (the vocal tract transfer function cannot be modeled adequately by five cascaded resonators when the sound source is above the larynx) so that cascade synthesizers are generally more complex in overall structure.

A second advantage of the cascade configuration is that it is a more accurate model of the vocal tract transfer

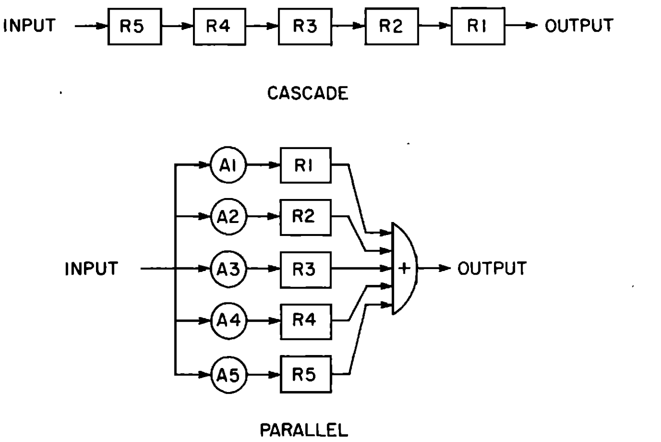


FIG. 3. The transfer function of the vocal tract may be simulated by a set of digital formant resonators R connected in cascade (the output of one feeding into the input of the next), or by a set of resonators connected in parallel (where each resonator must be preceded by an amplitude control A).

function during the production of non-nasal sonorants (Flanagan, 1957). As will be shown, the transfer functions of certain vowels are difficult to match using a parallel formant synthesizer. Although not optimal, a parallel synthesizer is particularly useful for generating stimuli that violate the normal amplitude relations between formants or if one wishes to generate, e.g., single-formant patterns.

The software simulation to be described has been programmed for normal use as a hybrid cascade/parallel synthesizer [Fig. 4(a)] or alternatively for special-purpose use as a strictly parallel synthesizer [Fig. 4(b)]. The experimenter must decide beforehand which configuration is to be employed. The change in configuration depends on the state of a single switch, and the program is smart enough to avoid performing unnecessary computations for resonators that are not used. To the extent that it is possible, the synthesizer has been adjusted so as to generate about the same output waveform whether the cascade/parallel configuration or the all-parallel configuration is selected.

Waveform sampling rate

Most of the sound energy of speech is contained in frequencies between about 80 and 8000 Hz (Dunn and White, 1940). However, intelligibility tests of band-pass filtered speech indicate that intelligibility is not measurably changed if the energy in frequencies above about 5000 Hz is removed (French and Steinberg, 1947). Speech low-pass filtered in this way sounds perfectly natural. Thus we have selected 5000 Hz (10000 samples per second) as the digital sampling rate of the synthesizer.

Parameter update rate

Control parameter values are updated every 5 ms. This is frequent enough to mimic even the most rapid of formant transitions and brief plosive bursts. If desired, the program can be modified to update parameter values only every 10 ms with relatively little decrement in output quality.

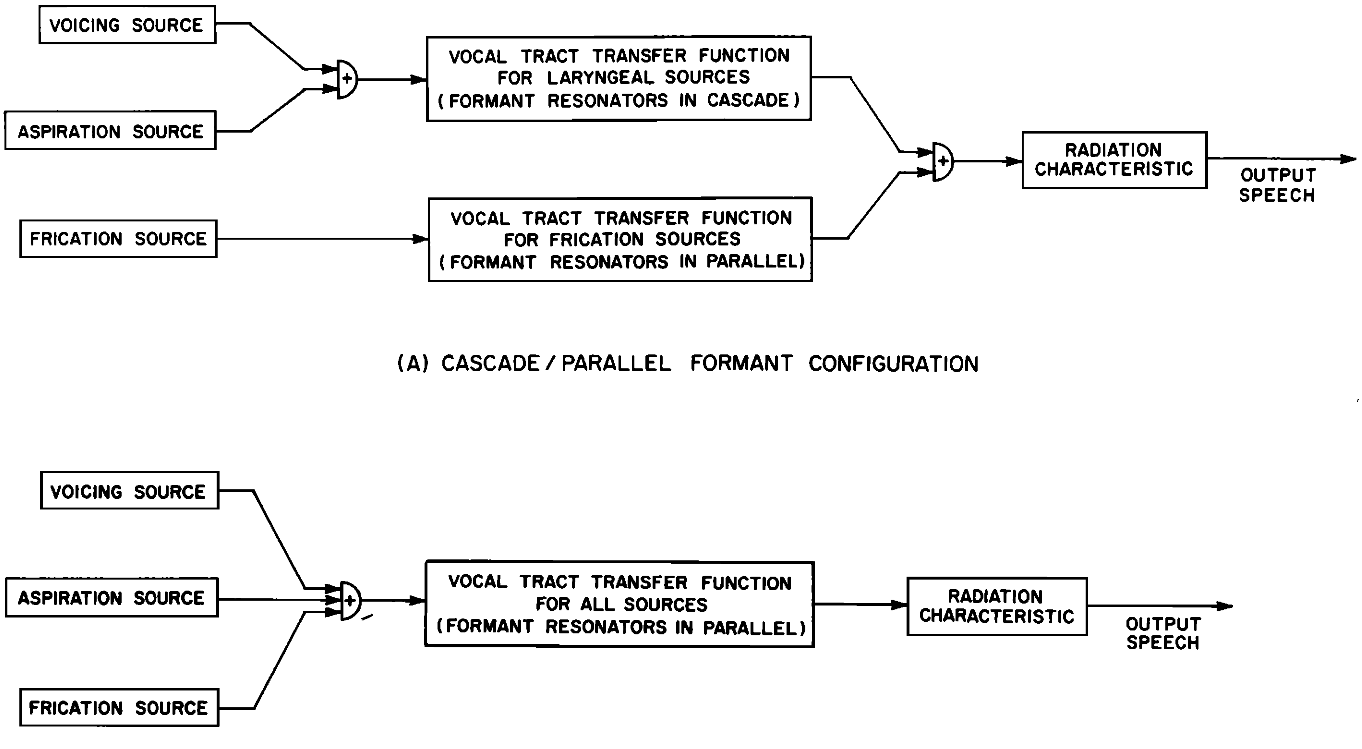
Digital resonators

The basic building block of the synthesizer is a digital resonator having the properties illustrated in Fig. 5. Two parameters are used to specify the input–output characteristics of a resonator, the resonant (formant) frequency F and the resonance bandwidth BW. Samples of the output of a digital resonator, $y(nT)$, are computed from the input sequence, $x(nT)$, by the equation

$$y(nT)=Ax(nT)+By(nT-T)+Cy(nT-2T), \qquad (1)$$

where $y(nT-T)$ and $y(nT-2T)$ are the previous two sample values of the output sequence $y(nT)$. The constants A, B, and C are related to the resonant frequency F and the bandwidth BW of a resonator by the impulse-invariant transformation (Gold and Rabiner, 1968)

$$C=-\exp(-2PI\,BW\,T),$$
$$B=2\exp(-PI\,BW\,T)\cos(2PI\,F\,T), \qquad (2)$$
$$A=1-B-C,$$


VOICING SOURCE
ASPIRATION SOURCE
FRICATION SOURCE
+
VOCAL TRACT TRANSFER FUNCTION FOR LARYNGEAL SOURCES (FORMANT RESONATORS IN CASCADE)
VOCAL TRACT TRANSFER FUNCTION FOR FRICATION SOURCES (FORMANT RESONATORS IN PARALLEL)
+
RADIATION CHARACTERISTIC
OUTPUT SPEECH

(A) CASCADE / PARALLEL FORMANT CONFIGURATION

VOICING SOURCE
ASPIRATION SOURCE
FRICATION SOURCE
+
VOCAL TRACT TRANSFER FUNCTION FOR ALL SOURCES (FORMANT RESONATORS IN PARALLEL)
RADIATION CHARACTERISTIC
OUTPUT SPEECH

(B) SPECIAL-PURPOSE ALL-PARALLEL FORMANT CONFIGURATION

FIG. 4. The synthesizer is normally used in a cascade/parallel configuration shown at the top, but may be used in an all parallel version shown at the bottom if one wishes to exercise independent control over formant amplitudes for vowels.

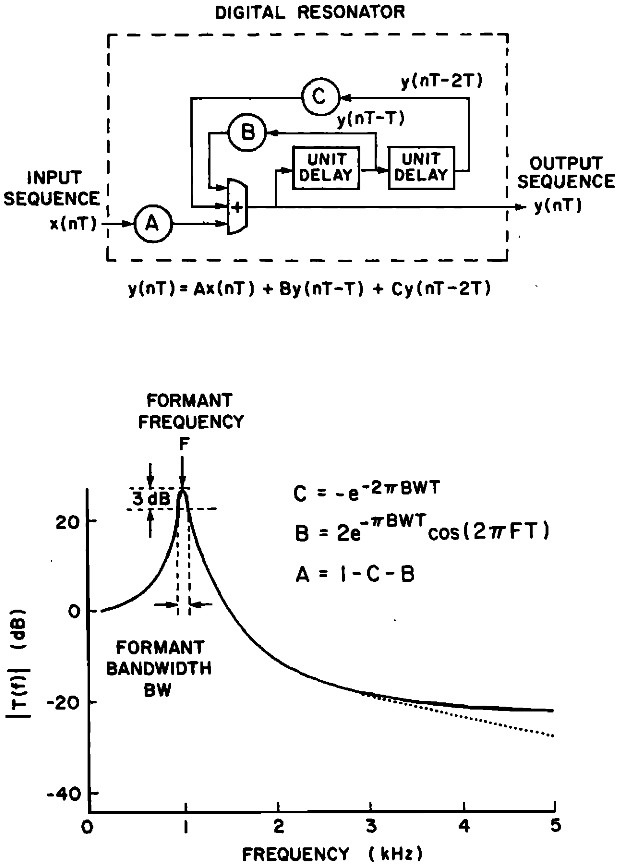


FIG. 5. The digital resonator shown in the form of a block diagram in the upper part of the figure has a transfer function (magnitude of the ratio of output to input in the frequency domain) as shown. In this example, $F=1000$ Hz and $BW=50$ Hz. The transfer function of a corresponding analog resonator is shown by the dotted line.

where PI is the familiar ratio of the circumference of a circle to its diameter. The constant T is one over the sampling rate and equals 0.0001 s in the present 5-kHz simulation.

The values of the resonator control parameters F and BW are updated every 5 ms, causing the difference equation constants to change discretely in small steps every 5 ms as an utterance is synthesized. Large sudden changes to these constants may introduce clicks and burps in the synthesizer output. Fortunately, acoustic theory indicates that formant frequencies must always change slowly and continuously relative to the 5-ms undate interval for control parameters.

A digital resonator is a second-order difference equation. The transfer function of a digital resonator has a sampled frequency response given by

$$T(f)=\frac{A}{1.-Bz^{-1}-Cz^{-2}} , \qquad (3)$$

where $z=\exp(j\,2PI\,f\,T)$, j is an imaginary number corresponding to the square root of -1, and f is frequency in Hz and ranges from 0 to 5 kHz. The transfer function has a (sampled) impulse response identical to a corresponding analog resonator circuit at sample times nT (Gold and Rabiner, 1968), but the frequency responses of an analog and digital resonator are not exactly the same, as can be seen in Fig. 5.

Digital antiresonator

An antiresonance (also called an antiformant or transfer-function zero pair) can be realized by slight modifications to these equations. The frequency response of an antiresonator is the mirror image of the response plotted in Fig. 5(i.e., replace dB by −dB). An antiresonator is used in the synthesizer to shape the spectrum of the voicing source and another is used to simulate the effects of nasalization in the cascade model of the vocal tract transfer function.

The output of an antiformant resonator, $y(nT)$, is related to the input $x(nT)$ by the equation

$$y(nT)=A'x(nT)+B'x(nT-T)+C'x(nT-2T), \qquad (4)$$

where $x(nT-T)$ and $x(nT-2T)$ are the previous two samples of the input $x(nT)$, the constants A', B', and C' are defined by the equations:

$$A'=1.0/A,\quad B'=-B/A,\quad C'=-CA\ , \qquad (5)$$

and where A, B, and C are obtained by inserting the antiresonance center frequency F and bandwidth BW into Eqs. (2).

Low-pass resonator

As a special case, the frequency F of a digital resonator can be set to zero, producing, in effect, a low-pass filter which has a nominal attentuation skirt of −12 dB per octave of frequency increase and a 3 dB down break frequency equal to $BW/2$. The voicing source contains a digital resonator RGP used as a low-pass filter that transforms a glottal impulse into a pulse having a waveform and spectrum similar to normal voicing. A second digital resonator RGS is used to low-pass filter the normal voicing waveform to produce the quasi-sinusoidal glottal waveform seen during the closure interval for an intervocalic voiced plosive.

Synthesizer block diagram

A block diagram of the synthesizer is shown in Fig. 6. There are 39 control parameters that determine the characteristics of the output. The name and range of values for each parameter are given in Table I. As can be seen from the table, one might wish to vary as many as 20 of the 39 parameters to achieve optimum matches to an arbitrary English utterance. The constant parameters in Table I have been given values appropriate for a particular male voice, and would have to adjusted slightly to approximate the speech of other male or female talkers. The list of variable control parameters is long compared with some synthesizers, but the emphasis here is on defining strategies for the synthesis of high quality speech. We are not concerned with searching for compromises that would minimize the information content in the control parameter specification.

I. SOURCES OF SOUND

There are two kinds of sound sources that may be activated during speech production (Stevens and Klatt,

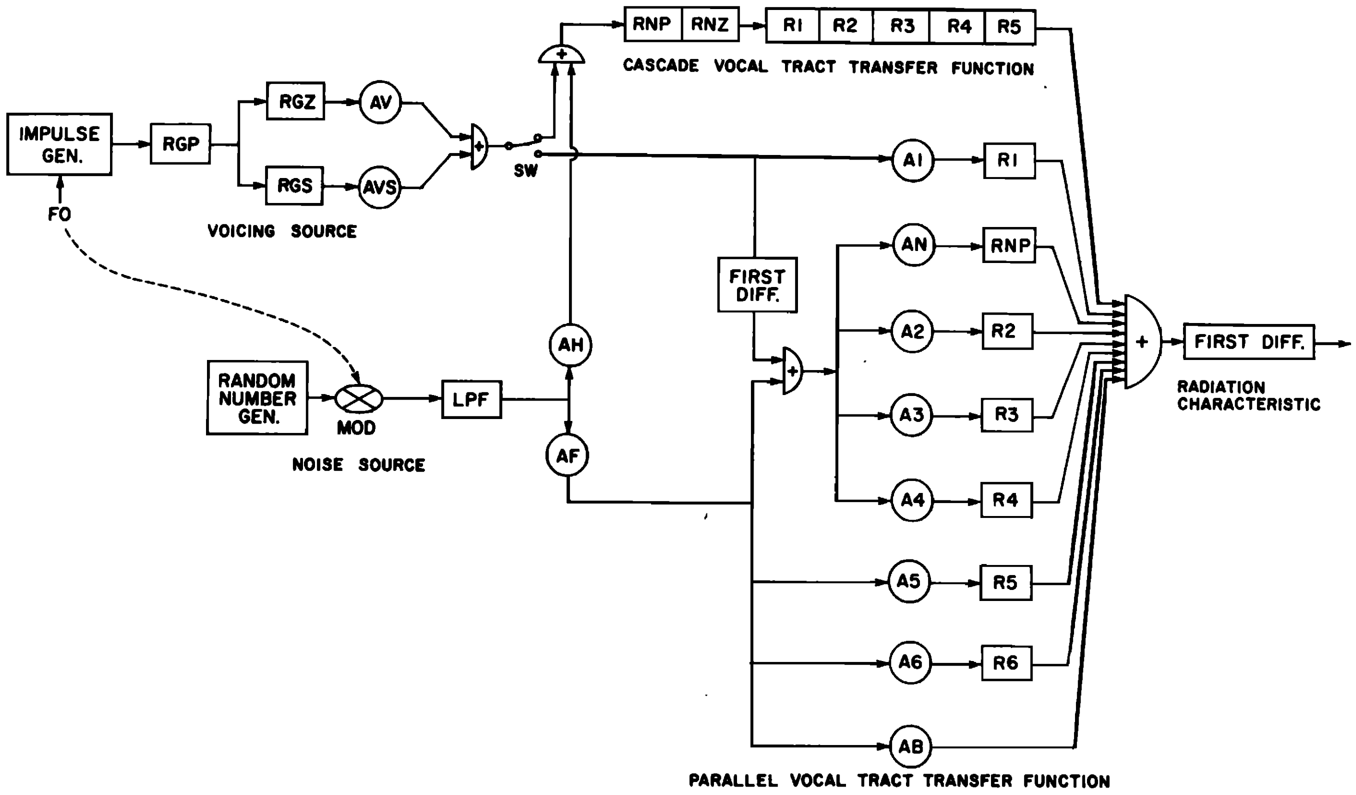


FIG. 6. Block diagram of the cascade/parallel formant synthesizer. Digital resonators are indicated by the prefix *R* and amplitude controls by the prefix *A*. Each resonator *Rn* has an associated resonant frequency control parameter *Fn* and a resonance bandwidth control parameter *Bn*.

1972). One involves quasi-periodic vibrations of some structure, usually the vocal folds. Vibration of the vocal folds is called voicing. (Other structures such as the lips, tongue tip, or uvula may be cuased to vibrate in sound types of some languages, but not in English.)

The second kind of sound source involves the generation of turbulence noise by the rapid flow of air past a narrow constriction. The resulting noise is called aspiration if the constriction is located at the level of the vocal folds, as for example during the production of the sound [*h*]. If the constriction is located above the larynx, as for example during the production of sounds such as [s], the resulting noise is called frication noise. The explosion of a plosive release also consists primarily of frication noise.

When voicing and turbulence noise generation co-exist, as in a voiced fricative such as [z] or a voiced [h], the noise is amplitude modulated periodically by the vibrations of the vocal folds. In addition, the vocal folds may vibrate without meeting in the midline. In this type of voicing, the amplitude of higher frequency harmonics of the voicing source spectrum is significantly reduced and the waveform looks nearly sinusoidal. Therefore the synthesizer should be capable of generating at least two types of voicing waveforms (normal voicing and quasi-sinusoidal voicing), two types of frication waveforms (normal frication and amplitude-modulated frication), and two types of aspiration (normal aspiration and amplitude-modulated aspiration). These are the only kinds of sound sources required for English, although trills and clicks of other languages may call for the addition of other source controls to the synthesizer in the future.

A. Voicing source

The structure of the voicing source is shown at the top left in Fig. 6. Variable control parameters are used to specify the fundamental frequency of voicing (*F*0), the amplitude of normal voicing (AV), and the amplitude of quasi-sinusoidal voicing (AVS).

An impulse train corresponding to normal voicing is generated whenever *F*0 is greater than zero. The amplitude of each impulse is determined by AV, the amplitude of normal voicing in dB. AV ranges from about 60 dB in a strong vowel to 0 dB when the voicing source is turned off. Fundamental frequency is specified in Hz; a value of $F0=100$ would produce a 100-Hz impulse train. The number of samples between impulses, *T*0, is determined by *SR*/*F*0, e.g., for a sampling rate of 10 000 and a fundamental frequency of 200 Hz, an impulse is generated every 50th sample.

Under some circumstances, the quantization of the fundamental period to be an integral number of samples might be perceived in a slow prolonged fundamental frequency transition as a sort of staircase of mechanical sounds (similar to the rather unnatural speech one gets by setting *F*0 to a constant value in a synthetic utterance), but the problem is not sufficiently serious to merit running the source model of the synthesizer at a higher sampling rate. If desired, some aspiration noise can be added to the normal voicing waveform to

TABLE I. List of control parameters for the software formant synthesizer. The second column indicates whether the parameter is normally constant (*C*) or variable (*V*) during the synthesis of English sentences. Also listed are the permitted range of values for each parameter, and a typical constant value.

N	*V/C*	Sym	Name	Min	Max	Typ
1	*V*	AV	Amplitude of voicing (dB)	0	80	0
2	*V*	AF	Amplitude of frication (dB)	0	80	0
3	*V*	AH	Amplitude of aspiration (dB)	0	80	0
4	*V*	AVS	Amplitude of sinusoidal voicing (dB)	0	80	0
5	*V*	F0	Fundamental freq. of voicing (Hz)	0	500	0
6	*V*	F1	First formant frequency (Hz)	150	900	450
7	*V*	F2	Second formant frequency (Hz)	500	2500	1450
8	*V*	F3	Third formant frequency (Hz)	1300	3500	2450
9	*V*	F4	Fourth formant frequency (Hz)	2500	4500	3300
10	*V*	FNZ	Nasal zero frequency (Hz)	200	700	250
11	*C*	AN	Nasal formant amplitude (dB)	0	80	0
12	*C*	A1	First formant amplitude (dB)	0	80	0
13	*V*	A2	Second formant amplitude (dB)	0	80	0
14	*V*	A3	Third formant amplitude (dB)	0	80	0
15	*V*	A4	Fourth formant amplitude (dB)	0	80	0
16	*V*	A5	Fifth formant amplitude (dB)	0	80	0
17	*V*	A6	Sixth formant amplitude (dB)	0	80	0
18	*V*	AB	Bypass path amplitude (dB)	0	80	0
19	*V*	B1	First formant bandwidth (Hz)	40	500	50
20	*V*	B2	Second formant bandwidth (Hz)	40	500	70
21	*V*	B3	Third formant bandwidth (Hz)	40	500	110
22	*C*	SW	Cascade/parallel switch	0(CASC)	1(PARA)	0
23	*C*	FGP	Glottal resonator 1 frequency (Hz)	0	600	0
24	*C*	BGP	Glottal resonator 1 bandwidth (Hz)	100	2000	100
25	*C*	FGZ	Glottal zero frequency (Hz)	0	5000	1500
26	*C*	BGZ	Glottal zero bandwidth (Hz)	100	9000	6000
27	*C*	B4	Fourth formant bandwidth (Hz)	100	500	250
28	*V*	F5	Fifth formant frequency (Hz)	3500	4900	3750
29	*C*	B5	Fifth formant bandwidth (Hz)	150	700	200
30	*C*	F6	Sixth formant frequency (Hz)	4000	4999	4900
31	*C*	B6	Sixth formant bandwidth (Hz)	200	2000	1000
32	*C*	FNP	Nasal pole frequency (Hz)	200	500	250
33	*C*	BNP	Nasal pole bandwidth (Hz)	50	500	100
34	*C*	BNZ	Nasal zero bandwidth (Hz)	50	500	100
35	*C*	BGS	Glottal resonator 2 bandwidth	100	1000	200
36	*C*	SR	Sampling rate	5000	20 000	10 000
37	*C*	NWS	Number of waveform samples per chunk	1	200	50
38	*C*	G0	Overall gain control (dB)	0	80	47
39	*C*	NFC	Number of cascaded formants	4	6	5

partially alleviate the problem and create a somewhat breathy voice quality.

B. Normal voicing

Ignoring for the moment the effects of RGZ, we see that the train of impulses is sent through a low-pass filter, RGP, to produce a smooth waveform that resembles a typical glottal volume velocity waveform (Flanagan, 1958). The resonator frequency FGP is set to 0 Hz and BGP to 100 Hz. The filtered impuses thus have a spectrum that falls off smoothly at approximately −12 dB per octave above 50 Hz. The waveform thus generated does not have the same phase spectrum as a typical glottal pulse, nor does it contain spectral zeros of the kind that often appear in natural voicing. These differences may be of some perceptual importance (for, e.g., naturalness of voice quality), in which case future versions of the program should provide for a more flexible voicing source specification.

The antiresonator RGZ is used to modify the detailed shape of the spectrum of the voicing source for particular individuals with greater precision that would be possible using only a single low-pass filter. The values chosen for FGZ and BGZ in Table I are such as to tilt the general voicing spectrum up somewhat to match the vocal characteristics of the author. The waveform and spectral envelope of normal voicing that is produced by sending an impulse train through RGP and RGZ are shown in Fig. 7(a).

C. Quasi-sinusoidal voicing

The amplitude control parameter AVS determines the amount of smoothed voicing generated during voiced fricatives, voiced aspirates, and the voicebars present in intervocalic voiced plosives. An appropriate wave shape for quasi-sinusoidal voicing is obtained by low-pass filtering an impulse by low-pass digital resonators RGP and RGS. The frequency control of RGS is set to zero to produce a low-pass filter, and BGS = 200 determines the cutoff frequency beyond which harmonics are

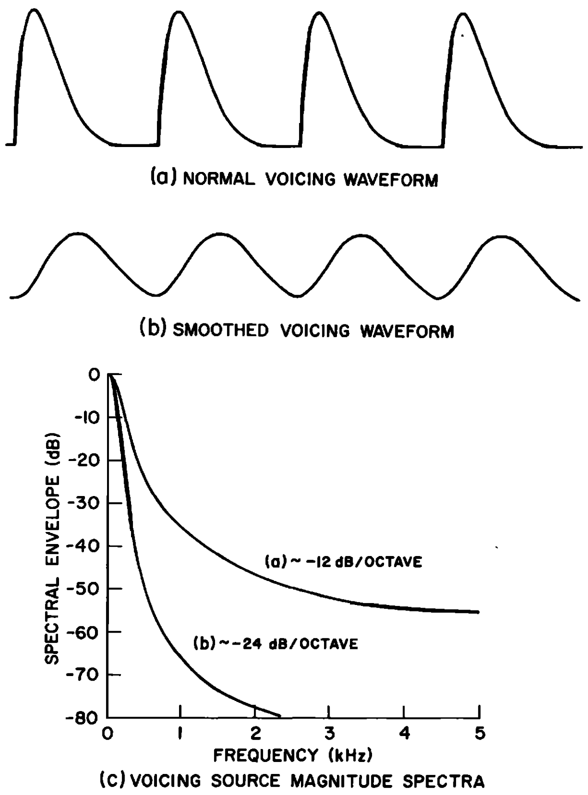


FIG. 7. Four periods from the synthetic waveforms of (a) normal voicing and (b) quasi-sinusoidal voicing are shown at the top, and the envelopes of the two resulting line spectra are shown in (c).

strongly attenuated.

The waveform and spectral envelope of quasi-sinusoidal voicing are shown in Fig. 7(b). After the effects of the vocal tract transfer function and radiation characteristic are imposed on the source spectrum, the output waveform of quasi-sinusoidal voicing contains significant energy only at the first and second harmonics of the fundamental frequency. AVS ranges from about 60 dB in a voicebar or strongly voiced fricative to 0 dB if no quasi-sinusoidal voicing is present. Some degree of quasi-sinusoidal voicing can be added to the normal voicing source (in combination with aspiration noise) to produce a breathy voice quality (e.g., AH = AV − 3, AVS = AV − 6).

D. Frication source

A turbulent noise source is simulated in the synthesizer by a pseudo-random number generator, a modulator, an amplitude control AF, and a −6 dB/octave low-pass digital filter LPF, as shown previously in Fig. 6. The spectrum of the frication source should be approximately flat (Stevens, 1971), and the amplitude distribution should be Gaussian. Signals produced by the random number generator have a flat spectrum, but they have a uniform amplitude distribution between limits determined by the value of the amplitude control parameter AF. A pseudo-Gaussian amplitude distribution is obtained in the synthesizer by summing 16 of the numbers produced by the random number generator.

In theory, the noise source is an ideal pressure source. The volume velocity of the frication noise depends on the impedance seen by the noise source. Since the vocal tract transfer function $T(f)$ relates source volume velocity to lip volume velocity, one must estimate noise volume velocity to determine lip output. In the general case, this is a complex calculation, but we will assume that source volume velocity is proportional to the integral of source pressure (an excellent approximation for a frication source at the lips because the radiation impedance is largely inductive, but only an approximation for other source locations). The integral is approximated by a first-order low-pass digital filter LPF that is shown in Fig. 6. Output samples from this filter, $y(nT)$, are related to the input sequence, $x(nT)$, by the equation

$$y(nT) = x(nT) + y(nT - T) \ .$$

As will be seen later, the radiation characteristic is a digital high-pass filter that exactly cancels out the effects of LPF. (For computational efficiency, the radiation characteristic can be moved into both the voicing source and the noise source; then the combination of radiation characteristic and the low-pass filter LPF can be removed from the noise source.)

An example of synthetic frication noise volume velocity that was generated in this way is shown in Fig. 8. The spectrum of this sample of noise fluctuates randomly about the expected long-term average noise spectrum (dashed line). Short samples of noise vary in their spectral properties due to the nature of random processes.

The output of the random number generator is amplitude modulated by the component labeled "MOD" in Fig. 6 whenever the fundamental frequency $F0$ and the amplitude of voicing AV are both greater than zero. Voiceless sounds (AV = 0) are not amplitude modulated because the vocal folds are spread and stiffened, and do not vibrate to modulate the airflow. The degree of

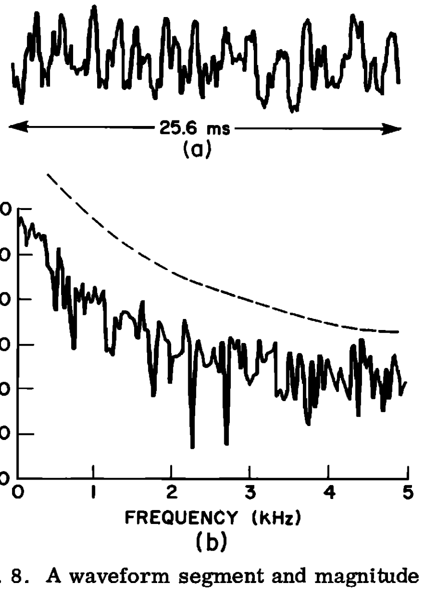


FIG. 8. A waveform segment and magnitude spectrum are shown of a 25.6-ms sample of frication noise. The expected long-term average spectrum of the output of the frication source is shown by the dashed curve. The dashed curve has been shifted up by 10 dB for clarity.

amplitude modulation is fixed at 50% in the synthesizer. The modulation envelope is a square wave with a period equal to the fundamental period. Experience has shown that it is not necessary to vary the degree of amplitude modulation over the course of a sentence, but only to ensure that it is present in voiced fricatives and voiced aspirated sounds.

The amplitude of the frication noise is determined by AF, which is given in dB. A value of 60 will generate a strong frication noise, while a value of zero effectively turns off the frication source.

E. Aspiration source

Aspiration noise is essentially the same as frication noise, except that it is generated in the larnyx. In a strictly parallel vocal tract model, AF can be used to generate both frication and aspiration noise. However, in the cascade synthesizer configuration, aspiration noise is sent through the cascade vocal tract model (since the cascade configuration is specially designed to model vocal tract characteristics for laryngeal sound sources), while fricatives require a parallel vocal tract configuration. Therefore separate amplitude controls are needed for frication and aspiration in a cascade/parallel configuration. The amplitude of aspiration noise sent to the cascade vocal tract model is determined by AH, which is given in dB. A value of 60 will generate strong aspiration, while a value of zero effectively turns off the aspiration source. Since frication and aspiration are generated by an identical process in the synthesizer, Fig. 8 describes the characteristics of the aspiration source as well.

F. Control of source amplitudes

Parameter values specifying source amplitudes AV, AVS, AF, and AH are adjusted by the user to new values every 5 ms. However, AV and AVS only have an effect on the synthetic waveform when a glottal impulse is issued. The reason for adjusting voicing amplitudes discontinuously at the onset of each glottal period is to prevent the creation of pops and clicks due to waveform discontinuities introduced by the sudden change in an amplitude control in the middle of a voicing period.

The noise amplitudes AF and AH are used to interpolate the intensity of the noise sources linearly over the 5-ms (50-sample) interval. (Thus there is a 5-ms delay in the attainment of a new amplitude value for a noise source.) Interpolation permits a more gradual onset for a fricative or [h] than would otherwise be possible. There is, however, one exception to this internal control strategy. A plosive burst involves a more-rapid source onset than can be achieved by 5-ms linear interpolation. Therefore, if AF increases by more than 50 dB from its value specified in the previous 5-ms segment, AF is (automatically) changed instantaneously to its new target value. We are presently evaluating the desirability of also injecting a step excitation of the vocal tract transfer function at this plosive release time so as to simulate the acoustic effect of a sudden release of the oral pressure behind the plosive occlusion.

G. Control of fundamental frequency

At times it is desired to specify precisely the timing of the first glottal pulse (voicing onset) relative to a plosive burst. For example, in the syllable [pa], it might be desired to produce a 5-ms burst of frication noise, 40 ms of aspiration noise, and voicing onset exactly 45 ms from the onset of the burst. Usually, a glottal pulse is issued in the synthesizer at a time specified by one over the value of the fundamental frequency control parameter value extant when the last glottal pulse was issued. However, if either AV or F0 is set to zero, no glottal pulse is issued during this 5-ms time interval; in fact no glottal pulses are issued until precisely the moment that both the AV and $F0$ control parameters become nonzero. In the case of the [pa] example above, both AV and $F0$ would normally be set to zero during the closure interval, burst, and aspiration phase, and AV would be set to about 60 dB and $F0$ to about 130 Hz at exactly 45 ms after the synthetic burst onset.

Since the update interval in the synthesizer is set to 5 ms, voice onset time can be specified exactly in 5-ms steps. If greater precision is needed, it would be necessary to change the parameter update interval from 5 ms (NWS = 50) to say 2 ms (NWS = 20).

H. Control of noise samples in a stimulus continuum

A pseudo-random number generator is used to generate both burst and aspiration for a plosive such as [pa]. The spectrum and intensity of a long sample of noise produced by the pseudo-random number generator can be expected to have the desired amplitude and spectral characteristics, but short samples of noise will vary considerably due to the random nature of pseudo-random numbers (recall Fig. 8). A particular brief noise sequence may have greater or lesser total intensity, or a peculiar spectral peak or valley not shared by other samples of noise that are used to generate a set of stimuli varying in voice onset time or burst frequency.

These random fluctuations in noise characteristics can cause some stimuli in a supposed continuum to stand out as different. When performing psychological experiments involving stimuli generated by the pseudo-random number generator, there are two ways to get around this problem. One is to use the same random number sequence in the generation of each member of the continuum by reinitializing the random number function, as is done by default if the synthesizer program is reloaded each time that a new stimulus is to be generated. The other way to minimize response fluctuations due to random noise is to generate several tokens of each stimulus type with different initial values given to the random number generator, and average listener responses over each type to try to wash out token variations.

II. VOCAL TRACT TRANSFER FUNCTIONS

The acoustic characteristics of the vocal tract are determined by its cross-sectional area as a function of distance from the larynx to the lips. The vocal tract forms a nonuniform transmission line whose behavior can be determined for frequencies below about 5 kHz by solving a one-dimensional wave equation (Fant, 1960). (Above 5 kHz, three-dimensional resonance modes would have to be considered.) Solutions to the wave equation result in a transfer function that relates samples of the glottal source volume velocity to output volume velocity at the lips.

The synthesizer configuration in Fig. 6 includes components to realize two different types of vocal tract transfer function. The first, a cascade configuration of digital resonators, models the resonant properties of the vocal tract whenever the source of sound is within the larynx. The second, a parallel configuration of digital resonators and amplitude controls, models the resonant properties of the vocal tract during the production of frication noise. The parallel configuration can also be used to model vocal tract characteristics for laryngeal sound sources, although the approximation is not quite as good as in the cascade model, see below.

A. Cascade vocal tract model

Assuming that the one-dimensional wave equation is a valid approximation below 5 kHz, the vocal tract transfer function can be represented in the frequency domain by a product of poles and zeros. Furthermore, the transfer function contains only about five complex pole pairs and no zeros in the frequency range of interest, as long as the articulation is non-nasalized and the sound source is at the larynx (Fant, 1960). The transfer function conforms to an all-pole model because there are no side-branch resonators or multiple sound paths. (The glottis is partially open during the production of aspiration so that the poles and zeros of the subglottal system are often seen in aspiration spectra; the only way to approximate their effects in the synthesizer is to increase the first formant bandwidth to about 300 Hz. The perceptual importance of the remaining spectral distortions caused by the poles and zeros of the subglottal system is probably minimal).

Five resonators are appropriate for simulating a vocal tract with a length of about 17 cm, the length of a typical male vocal tract, because the average spacing between formants is equal to the velocity of sound divided by half the wavelength, which works out to be 1000 Hz. A typical female vocal tract is 15 to 20% shorter, suggesting that only four formant resonators be used to represent a female voice in a 5 kHz simulation (or that the simulation should be extended to about 6 kHz). It is suggested that the voices of women and children be approximated by setting the control parameter NFC to 4, thus removing the fifth formant from the cascade branch of the block diagram shown in Fig. 6. For a male talker with a very long vocal tract, it may be necessary to add a sixth resonator to the cascade branch. As currently programmed, NFC can be set to four, five, or six formants in the cascade branch. Any change to NFC implies a change in the effective length of the vocal tract. NFC should not be used simply to remove a formant already present below 5 kHz because the spectrum of the resulting sound is tilted down in an in appropriate way: stimuli with a reduced number of formants must be generated using the all-parallel synthesizer configuration. (So such changes must be made with care.)

Ignoring for the moment the nasal pole resonator RNP and the nasal zero anti-resonator RNZ, the cascade model of Fig. 6, consisting of five formant resonators, has a volume velocity transfer function that can be represented in the frequency domain as a product of transfer functions identical to Eq. (3) (Gold and Rabiner, 1968):

$$T(f)=\prod_{n=1}^{5}\frac{A(n)}{1.-B(n)z^{-1}-C(n)z^{-2}}\,, \tag{6}$$

where the constants $A(n)$, $B(n)$, and $C(n)$ are determined by the values of the nth formant frequency $F(n)$ and nth formant bandwidth $BW(n)$ by the relations given earlier in Eq. (2). The constants $A(n)$ in the numerator of Eq. (6) insure that the transfer function has a value of unity at zero frequency, i.e., the dc airflow is unimpeded. The magnitude of $T(f)$ is plotted in Fig. 9 for several values of formant frequencies and formant bandwidths

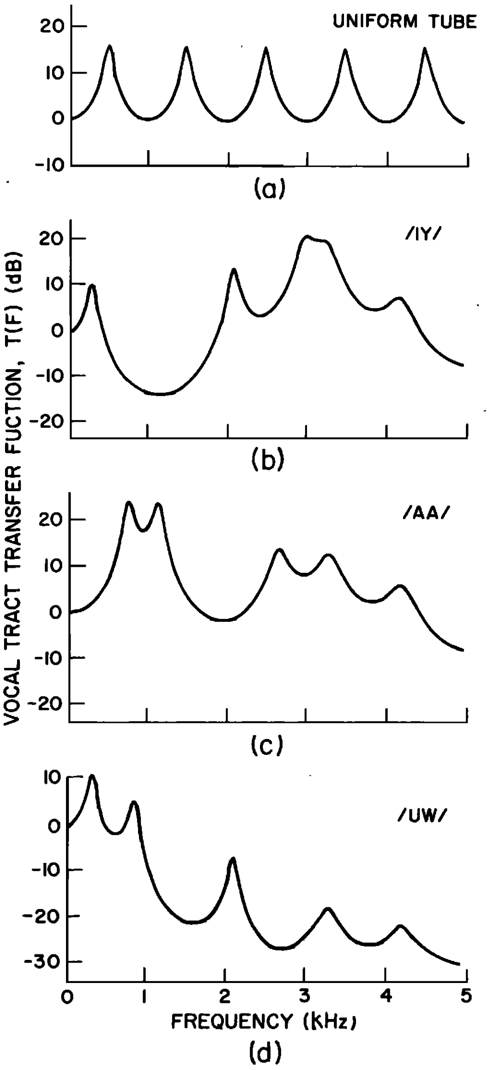


FIG. 9. The magnitude of the vocal tract transfer function is plotted for an ideal uniform vocal tract, and for the vowels [i], [a], and [u].

B. Relation to analog models of the vocal tract

The transfer function of the vocal tract can also be expressed in the continuous world of differential equations. Equation (6) is then rewritten as an infinite product of poles in the Laplace transform s plane:

$$T(f) = \prod_{n=1}^{\infty} \frac{s(n)\, s^*(n)}{[s+s(n)][s+s^*(n)]} \,, \qquad (6a)$$

where $s = j\ 2\,PI\ f$, and the constants $s(n)$ and $s^*(n)$ are determined by the values of the nth formant frequency $F(n)$ and the nth formant bandwidth $BW(n)$ by the relations

$$s(n) = PI\ BW(n) + j\,2PI\,F(n),$$

$$s^*(n) = PI\,BW(n) - j\,2PI\,F(n)\ .$$

The two formulations (6) and (6a) are exactly equivalent representation of the transfer function for an ideal vocal tract configuration corresponding to a uniform tube closed at the glottis and having all formant bandwidths equal to, e.g., 100 Hz. The two formulations are indistinguishable at representing vocal tract transfer functions below 5 kHz. However, in a practical synthesizer, the infinite product of poles can only be approximated [e.g., by building five electronic resonators and a higher-pole correction network (Fant, 1959)].

C. Formant frequencies

Each formant resonator introduces a peak in the magnitude spectra shown in Fig. 9. The frequency of formant peak "n" is determined by the formant frequency control parameter Fn. (The amplitude of a formant peak depends not only on Fn and the formant bandwidth control parameter BWn, but also on the frequencies of the other formants, as will be discussed below.)

Formant frequency values are determined by the detailed shape of the vocal tract. Formant frequency values associated with different phonetic segments in the speech of the author will be presented in Sec. V. The frequencies of the lowest three formants vary substantially with changes to articulation (e.g., the observed range of $F1$ is from about 180 to 750 Hz, of $F2$ is 600 to 2300 Hz, and of $F3$ is 1300 to 3100 Hz for a typical male talker). The frequencies and bandwidths of the 4th and 5th formant resonators do not vary as much and could be held constant with little decrement in output sound quality. These higher frequency resonators help to shape the overall spectrum, but otherwise contribute little to intelligibility for vowels. The particular values chosen for the fourth and fifth formant frequencies (Table I) produce an energy concentration around 3 to 3.5 kHz and a rapid falloff in spectral energy above about 4 kHz, which is a pattern typical of many talkers.

D. Formant bandwadths

Formant bandwidths are a function of energy losses due to heat conduction, viscosity, cavity-wall motions, radiation of sound from the lips, and the real part of the glottal source impedance. Bandwidths are difficult to deduce from analyses of natural speech because of irregularities in the glottal source spectrum. Bandwidths have been estimated by other techniques such as using a sinusoidal swept-tone sound source (Fujimura and Lindqvist, 1971). Results indicate that bandwidths vary by a factor of 2 or more as a function of the particular phonetic segment being spoken. The primary perceptual effect of a bandwidth change is an increase or decrease in the effective intensity of a formant energy concentration [see Fig. 11(b) below] because formant bandwidths are narrower than a critical band (Carlson, Granstrom, and Klatt, 1979). Bandwidth variation is small enough that all formant bandwidths might be held constant in some applications, in which case only $F1$, $F2$ and $F3$ would be varied to simulate the vocal tract transfer functions for non-nasalized vowels and sonorant consonants.

E. Nasals and nasalization of vowels

It is not possible to approximate nasal murmurs and the nasalization of vowels that are adjacent to nasals with a cascade system of five resonators alone. More than five formants are often present in these sounds and formant amplitudes do not conform to the relations inherent in a cascade configuration because of the presence of transfer function zeros (Fujimura, 1961; 1962). Typical transfer functions for a nasal murmur and for a nasalized [$\tilde{I}$] are shown in Fig. 10. Nasalization introduces additional poles and zeros into the transfer function of the vocal–nasal tract due to the

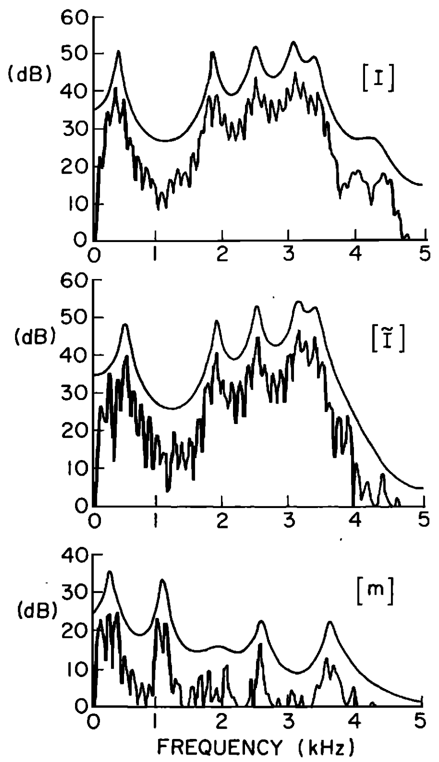


FIG. 10. Spectra are compared of a the vowel [I], the same vowel when nasalized, and a nasal murmur [m], all obtained from the recorded syllable "dim". The nasal murmur and the nasalized [Ĩ] have an extra transfer function pole pair and zero pair near $F1$. The extra peak and valley are not appearant in the linear prediction spectrum, but can be discerned in the pattern of harmonic amplitudes near $F1$ in the discrete Fourier transform spectrum.

presence of a side-branch resonator. The oral cavity forms the side-branch resonator in the case of a nasal murmur, while the nose should be considered a side-branch resonator in a nasalized vowel (because the amount of sound radiated through the nostrils is insignificant compared to the effect of the lowered velum on the formant structure of the sound output from the lips).

Nasalization of adjacent vowels is an important element in the synthesis of nasal consonants. The perceptually most important change associated with nasalization of a vowel is the reduction in amplitude of the first formant, brought on by the presence of a nearby low-frequency pole pair and zero pair. The first formant frequency also tends to shift slightly higher in most nasalized vowels.

Nasal murmurs and vowel nasalization are approximated by the insertion of an additional resonator RNP and anti-resonator RNZ into the cascade vocal tract model. The nasal pole frequency FNP can be set to a fixed value of about 270 Hz for all time. The nasal zero frequency FNZ should also be set to a value of about 270 Hz during non-nasalized sounds, but the frequency of the nasal zero must be increased during the production of nasals and nasalization. The RNP–RNZ pair is effectively removed from the cascade circuit during the synthesis of non-nasalized speech sounds if FNP = FNZ. Stragegies for controlling FNZ are given in Sec. V.

F. Parallel vocal tract model for frication sources

During frication excitation, the vocal tract transfer function contains both poles and zeros. The pole frequencies are temporally continuous with formant locations of adjacent phonetic segments because, by definition, the poles are the natural resonant frequencies of the entire vocal tract configuration, no matter where the source is located. Thus the use of vocalic formant frequency parameters to control the locations of frication maxima is theoretically well-motivated (and helpful in preventing the fricative noises from "dissociating" from the rest of the speech signal).

The zeros in the transfer function for fricatives are the frequencies for which the impedance looking back toward the larynx from the position of the frication source is infinite, since the series-connected pressure source of turbulence noise cannot produce any output volume velocity under these conditions. The effect of transfer-function zeros is twofold; they introduce notches in the spectrum and they modify the amplitudes of the formants. The perceptual importance of spectral notches is not great because masking effects of adjacent energy in format peaks limit the detectability of a spectral notch (Gauffin and Sundberg, 1974; Carlson, Granstrom, and Klatt, 1979). We have found that a satisfactory approximation to the vocal tract transfer function for frication excitation can be achieved with a parallel set of digital formant resonators having amplitude controls, and no antiresonators.

Formant amplitudes are set to provide frication excitation for selected formants, usually those associated with the cavity in front of the constriction (Stevens, 1972). The presence of any transfer function zeros is accounted for by appropriate settings of the formant amplitude controls. Relatively simple rules for determination of the formant amplitude settings (and bypass path amplitude values) as a function of place of articulation can be derived from a quantal theory of speech production (Stevens, 1972). The theory states that only formants associated with the cavity in front of the oral constriction are strongly excited. The theory is supported by the formant amplitude specifications for fricatives and plosive bursts in Sec. V. These amplitude control data were derived from trial-and-error attempts to match natural frication spectra.

There are six formant resonators in the parallel configuration of Fig. 6. A sixth formant has been added to the parallel branch specifically for the synthesis of very high frequency noise in [s, z]. The main energy concentration in these alveolar fricatives is centered on a frequency of about 6 kHz. This is above the highest frequency (5 kHz) that can be synthesized in a 10 000 sample/second simulation. However, in an [s], there is gradually increasing frication noise in the frequencies immediately below 5 kHz due to the low-frequency skirt of the 6-kHz formant resonance, and this noise spectrum can be approximated quite well by a resonator positioned at about 4900 Hz. We have found it better to include an extra resonator to simulate high-frequency noise than to move $F5$ up in frequency whenever a sibilant is to be synthesized because clicks and moving energy concentrations are thereby avoided.

Also included in the parallel vocal tract model is a bypass path. The bypass path with amplitude control AB is present because the transfer functions for [f, v, θ, ð, p, b] contain no prominent resonant peaks, and the synthesizer should include a means of bypassing all of the resonators to produce a flat transfer function.

During the production of a voiced fricative, there are two active sources of sound, one located at the glottis (voicing) and one at a constriction in the vocal tract (frication). The output of the quasi-sinusoidal voicing source is sent through the cascade vocal tract model, while the frication source excites the parallel branch to generate a voiced fricative.

G. Simualtion of the cascade configuration by the parallel configuration

The transfer function of the laryngeally excited vocal tract can also be approximated by five digital formant resonators connected in parallel. The same resonators that form the parallel branch for frication excitation can be used to synthesize any sonorant if suitable values are chosen for the formant amplitude controls.

The following rules summarize what happens to formant amplitudes in the transfer function $T(f)$ of a cascade model as the lowest five formant frequencies and bandwidths are changed. These relations follow directly from Eq. (6) under the assumption that each formant frequency $F(n)$ is at least 5 to 10 times as large as the formant bandwidth $BW(n)$:

1. The formant peaks in the transfer function are equal for the case that formant frequencies are set to 500, 1500, 2500, 3500, and 4500 Hz and formant bandwidths are set to be equal at 100 Hz. This corresponds to a vocal tract having a uniform cross-sectional area, a closed glottis, open lips (and a nonrealistic set of bandwidth values), as shown in part (a) of Fig. 11.

2. The amplitude of a formant peak is inversely proportional to its bandwidth. If a formant bandwidth is doubled, that formant peak is reduced in amplitude by 6 dB. If the bandwidth is halved, the peak is increased by 6 dB, as shown in part (b) of Fig. 11.

3. The amplitude of a formant peak is proportional to formant frequency. If a formant frequency is doubled, that formant peak is increased by 6 dB, as shown in part (c) of Fig. 11. [This is true of $T(f)$ but not of the resulting speech output spectrum since the glottal source spectrum falls off at about −12 dB/octave of frequency increase and the radiation characteristic imposes a + 6 dB octave spectral tilt, resulting in a net change in formant amplitude of + 6 −12 + 6 = 0 dB.]

4. Changes to a formant frequency also affect the amplitudes of higher formant peaks by a factor proportional to frequency squared. For example, if a formant frequency is halved, amplitudes of all higher formants are decreased by 12 dB, i.e., one half squared, as shown in part (c) of Fig. 11.

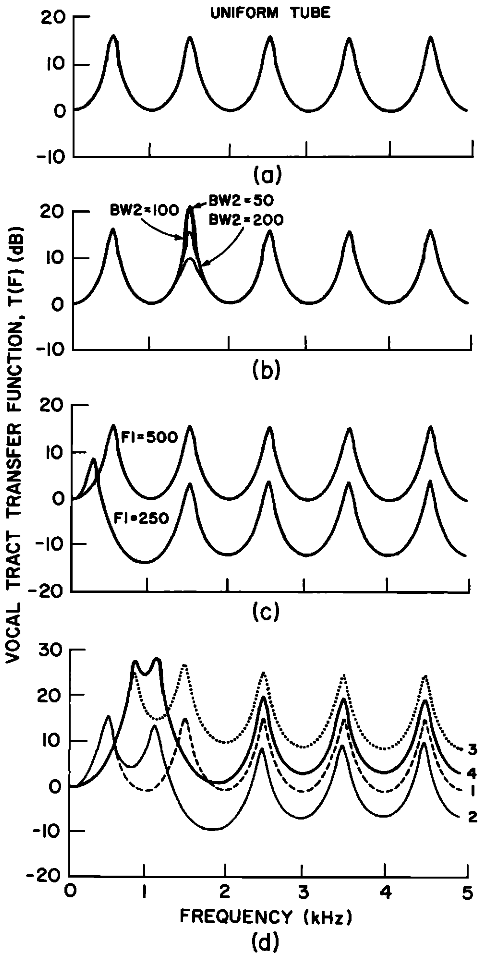


FIG. 11. Vocal tract transfer function Illustrating (a) the transfer function of a uniform vocal tract, (b) the influence of a change to a formant bandwidth, (c) changes to formant amplitudes caused by a shift in the frequency of a lower formant, and (d) the increase in formant amplitudes when two formants get close in frequency, see text.

5. The frequencies of two adjacent formants cannot come any closer than about 200 Hz because of coupling between the natural modes of the vocal tract. However, if two formants approach each other by about this amount, both formant peaks are increased by an additional 3 to 6 dB, as shown in part (d) of Fig. 11.

The amplitudes of the formant peaks generated by the parallel vocal tract model have been constrained such that, if A1 to A5 are all set to 60 dB, the transfer function will approximate that found in the cascade model. This is accomplished (1) by adjusting the gain of the higher frequency formants to take into account frequency changes in lower formants [since a higher formant rides on the skirts of the transfer function of all lower formants in a cascade model (Fant, 1960)], (2) by incorporation rules to cause formant amplitudes to increase whenever two formant frequencies come into proximity, and (3) by using a first difference calculation to remove low-frequency energy from the higher formants; this energy would otherwise distort the spectrum in the region if $F1$ during the synthesis of some vowels (Holmes, 1973).

The magnitude of the vocal tract transfer functions of the cascade and parallel vocal tract models are compared in Fig. 12 for several vowels. The match is quite good in the vicinity of formant peaks, but the parallel model introduces transfer function zeros (notches) in the spectrum between formant peaks. The notches are of relatively little perceptual importance because energy in the formant peak adjacent to the notch on the low frequency side tends to make the detectability of a spectral notch (Gauffin and Sundberg, 1974).

Many early parallel synthesizers were programmed to add together formant outputs without filtering out the energy at low frequencies from resonators other than $F1$. In other cases, formant outputs were combined in alternating sign. The deleterious effects of these choices are illustrated in Fig. 13. As can be seen, some vowel spectra are poorly modeled in both of these parallel methods of synthesis. The perceptual degradation is less in the alternating-sign case because spectral notches are less perceptible than energy fill in a spectral valley between two formants. Comparison of Fig. 12 and Fig. 13 indicates that our parallel configuration is better than either of those shown in Fig. 13.

A nasal formant resonator RNP appears in the parallel branch to assist in the approximation of nasal murmurs and vowel nasalization during parallel formant synthesis of vowels. Neither the parallel nasal formant nor the parallel first formant resonator are needed in the normal cascade/parallel synthesizer configuration (SW= 0), but they are required for the simulation of nasalization in the special-purpose all-parallel configuration (SW= 0).

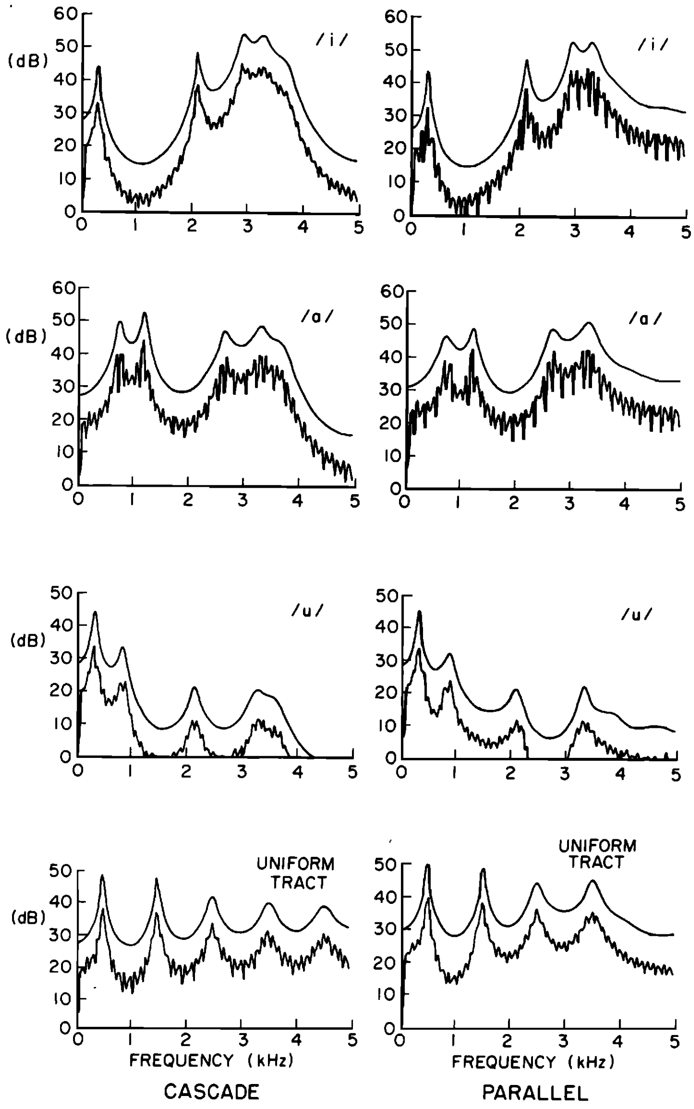


FIG. 12. Preemphasized output spectra (the DFT spectrum and a 14-pole LPC approximation) are shown for the vowels [i], [a], [u], and a uniform vocal tract when simulated by the theoretically correct cascade model and by the parallel approximation to the cascade model.

III. RADIATION CHARACTERISTIC

The box labeled radiation characteristic in Fig. 6 models the effect of directivity patterns of sound radiating from the head as function of frequency. The sound pressure measured directly in front of and about a meter from the lips is proportional to the temporal derivative of the lip-plus-nose volume velocity, and inversely proportional to r, the distance from the lips (Fant, 1960). The transformation is simulated in the synthesizer by taking the first difference of lip–nose volume velocity:

$$p(nT)=u(nT)-u(nT-T) \; . \tag{7}$$

The radiation characteristic adds a gradual rise in the overall spectrum, as shown in Fig. 14.

IV. HOST COMPUTER

The computer on which the software of Appendix B is installed must have digital-to-analog and analog-to-digital converters capable of transferring 10 000 12-bit waveform samples per second with precise control over the time between each sample. The computer should also have the appropriate audio equipment such as amplifiers, speaker, earphones, and tape recorder, as well as an external low-pass filter. The analog low-pass filter must have a sharp frequency cutoff near 5000 Hz, no appreciable ripple in the passband, and a nearly linear phase response in the passband. A filter that we constructed for this purpose is described in Appendix A.

A. Execution time

The synthesizer program is written in FORTRAN using floating point variables, so it is rather slow-running on some general-purpose digital computers (about 200 times real time on a PDP-11/40, but only about six times real time on a PDP-11/45 which has a faster floating-point multiply instruction). Even on a slower computer, the computational delay is not a serious handicap for most perceptual studies because stimuli can be generated and stored on disk for later presentation to subjects. Ultimately, it is hoped that the program will be implemented as a real-time digital device (Allen, 1977; Caldwell, 1979).

B. Graphical specification of variable control parameter data

A user could specify parametric data for an utterance to by synthesized in one of two general ways. One is to type in a sequence of ⟨time-value⟩ points for each variable control parameter and have the computer draw straight lines between them. However this is fairly time consuming and subject to error if no visual feedback in the form of a time plot of parameter values is provided.

In our laboratory, parameter specification can also be accomplished by selecting an individual parameter track for display. The computer is programmed to plot parameter values versus time as points on a refresh display, as shown in Fig. 15. Also displayed is the current ⟨time-value⟩ position of a cursor that is controlled by a hand-held pen. The pen can be moved over the surface of a graphical tablet (e.g., a Summagraphics model HW-1-11 data tablet digitizer or a Tektronix model 4953) to specify the parameter contour. A three-character symbol and a maximum y-axis value for the control parameter are also displayed.

It is relatively easy to watch the screen while moving the pen over the tablet with one hand. The other hand is positioned over three toggle switches which, when raised, have the following functions.

1. *Continuous input.* If the pen is swept horizontally past a 5 ms time point while toggle 1 is up, the stored value corresponding to this time is changed to the vertical value indicated by the pen at that moment.

2. *Straight line segment input.* A straight line is

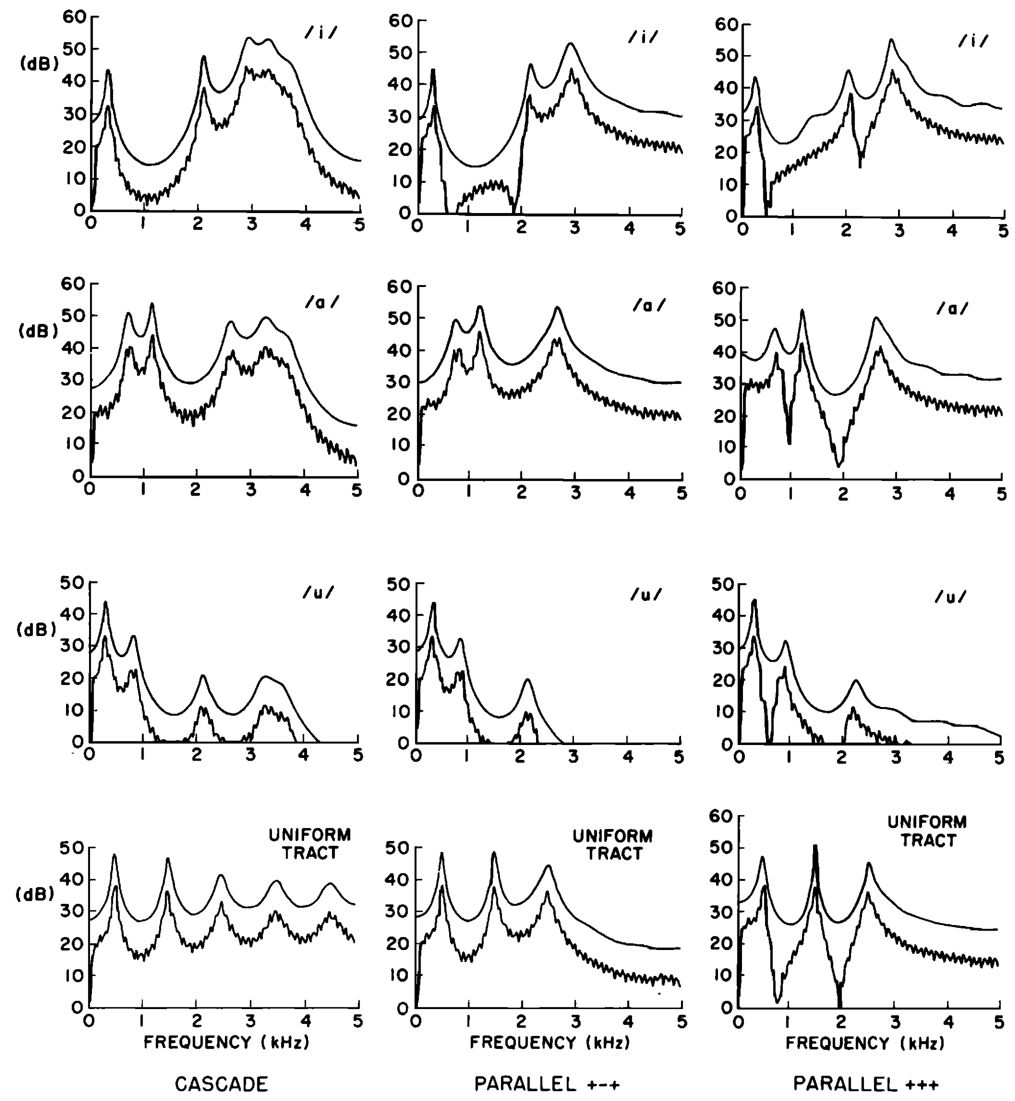


FIG. 13. Outputs from two common parallel synthesis configurations (in which resonator outputs are added with the same sign, or with alternating (+−+) signs, but without the first-difference preemphasis for $R2$ and $R3$, as in our model) are compared with the theoretically correct cascade output.

drawn from the previous ⟨time-value⟩ point to the point nearest the cursor at the moment the toggle is raised. The display is frozen until the toggle is returned to a down position. If no previous ⟨time-value⟩ point had been established, the action taken is to simply set the nearest point to the value of the cursor at the moment the toggle was raised.

3. *Set parameter track to a constant.* All time points are set to the value of the cursor at the moment the toggle is raised, and the displayed cursor position is frozen until the toggle is lowered.

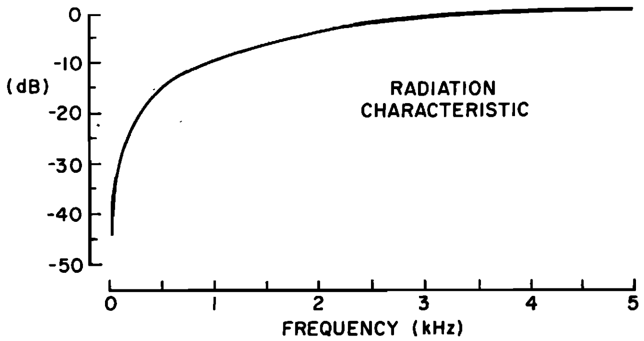


FIG. 14. Transfer function of the radiation characteristic.

In most experimental situations, we have found it advantageous to specify parameter contours in term of straight line segments, using toggle 2.

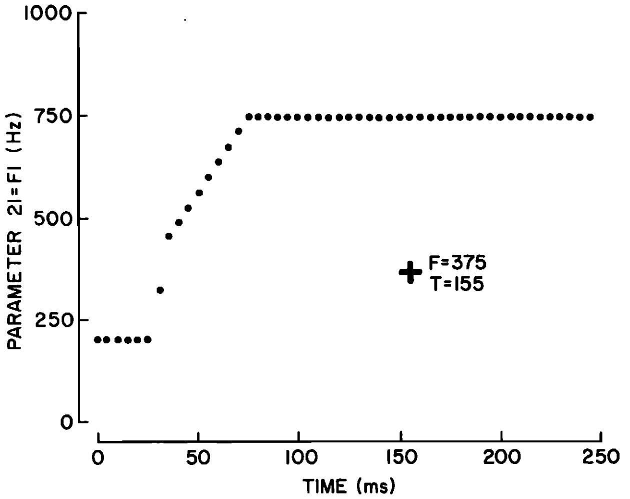


FIG. 15. A display of the first formant control parameter for the synthesis of the syllable (ba) is shown as it would appear on the computer scope. Formant frequency data are specified every 5 ms. The pen-controlled cursor position is indicated by a cross near the center of the plot.

C. Other software

The following executive commands are available for the manipulation of synthetic and natural waveforms:

1. *Select a waveform buffer*. At the beginning of a session, the user specifies the number and length of a set of disk waveform buffers. At any point in the session, one of these buffers is "current" and may be viewed or otherwise manipulated.

2. *Synthesize a waveform*. Take the control parameter files that have been drawn, and compute a synthetic waveform which is placed in the current waveform buffer.

3. *Display the waveform*. A 50-ms segment of the current waveform buffer is displayed. A variable knob is used to select the desired portion of the waveform for viewing.

4. *Listen*. The contents of the current waveform buffer are played out through the digital-to-analog converter.

5. *Listen to all*. The contents of each waveform buffer are played out in turn through the A/D converter with pauses of NPAUSE ms between waveforms.

6. *Digitize*. Begin to digitize a waveform obtained from a tape recorder or microphone and place it into the current waveform buffer. Stop digitizing when the duration of the buffer is exceeded.

7. *Edit waveform beginning time*. Chop off that portion of the current waveform before a centrally displayed waveform cursor, i.e., redefine the beginning of the waveform buffer. Use the knob to position the waveform relative to the fixed central waveform cursor before executing this command.

8. *Edit waveform ending time*. Chop off that portion of the current waveform after a centrally displayed cursor.

9. *Generate DFT spectrum*. Compute and display the magnitude of the discrete Fourier transform of a segment of waveform centered on the waveform display (see below for details).

10. *Generate LP spectrum*. Compute and display the magnitude of the linear prediction spectrum of a segment of waveform centered about the waveform display cursor (see below for details). Also list estimates of the five lowest formant frequencies.

11. *Plot estimated fundamental frequency contour*. Use the autocorrelation method to plot fundamental frequency versus time for the contents of the current waveform buffer.

12. *Plot intensity contour*. Compute rms intensity every 10 ms using a smooth time weighting window, and plot in decibels as a function of time for the contents of the current waveform buffer.

13. *Generate an identification test*. Randomize the set of waveform buffers to generate an identification test consisting of NTRIAL trials of each waveform buffer, with pauses of NPAUSE ms between trials.

14. *Generate an* AX *discrimination test*. Compare adjacent waveform buffers in a randomized order of NTRIAL trials with NPAUSE ms between trials and MPAUSE ms between members of an AX stimulus pair.

15. *Generate an* AXB *discrimination test*. Same as above, except the task is to detect whether $X = A$ or $X = B$.

Although not listed above, it is also desirable to have some means of saving parametric data and synthetic waveforms in digital form between computer sessions.

D. Spectral analysis

A spectral analysis capability is needed to verify that the output synthetic waveform has the desired spectral characteristics. Spectral analysis is particularly useful when attempting to mimic a digitized natural utterance. In our experience, if synthesis and natural spectral peaks are matched to within a couple of dB throughout the utterance, and the $F0$ contour and overall intensity contour are accurately duplicated, the synthetic utterance will be virtually indistinguishable from the original in both intelligibility and naturalness. Based on experience with several alternative spectral displays, we have found the following strategy to result in maximally useful spectral information.

A 25.6-ms segment of waveform is selected, the first difference of waveform samples is computed (to remove dc components and tilt the spectrum up slightly, somewhat analogous to the processing that takes place in the human peripheral auditory system), the differenced waveform is multiplied by a 25.6-ms Kaiser window with BETA = 7.0 (Kaiser, 1966) (so as to minimize the deleterious effects of having a nonintegral number of periods within the waveform segment), the discrete Fourier transform is computed, the magnitude of each spectral sample is converted to decibels and plotted as a set of lines connecting the ⟨dB, frequency⟩ samples, as shown in Fig. 16.

More useful in most cases is the linear prediction spectrum, which is obtained in the same way, except the discrete Fourier transform step is preceeded by a linear prediction analysis. The algorithm that we use is called the autocorrelation method (Makhoul, 1975) and uses 14 poles. The linear prediction coefficients are placed in the waveform buffer and padded with zeros before computing the discrete Fourier transform. Both DFT and linear prediction spectra can be superimposed, as in the figure, if there is some question as to the interpretation of the idealized spectra produced by the linear prediction algorithm.

Formant frequencies can often be estimated from the peaks observed in the linear prediction spectrum, such as is shown in Fig. 16, or by a root-solving technique (Markel and Gray, 1976). Plots of formant frequency trajectories can be obtained in this way, although the computations required make formant tracking of natural digitized waveforms a fairly lengthy process on most computers. One such plot is shown in Fig. 18 below. Extra points and missing points in this plot indicate that these quasi-formant tracks must be interpreted with care.

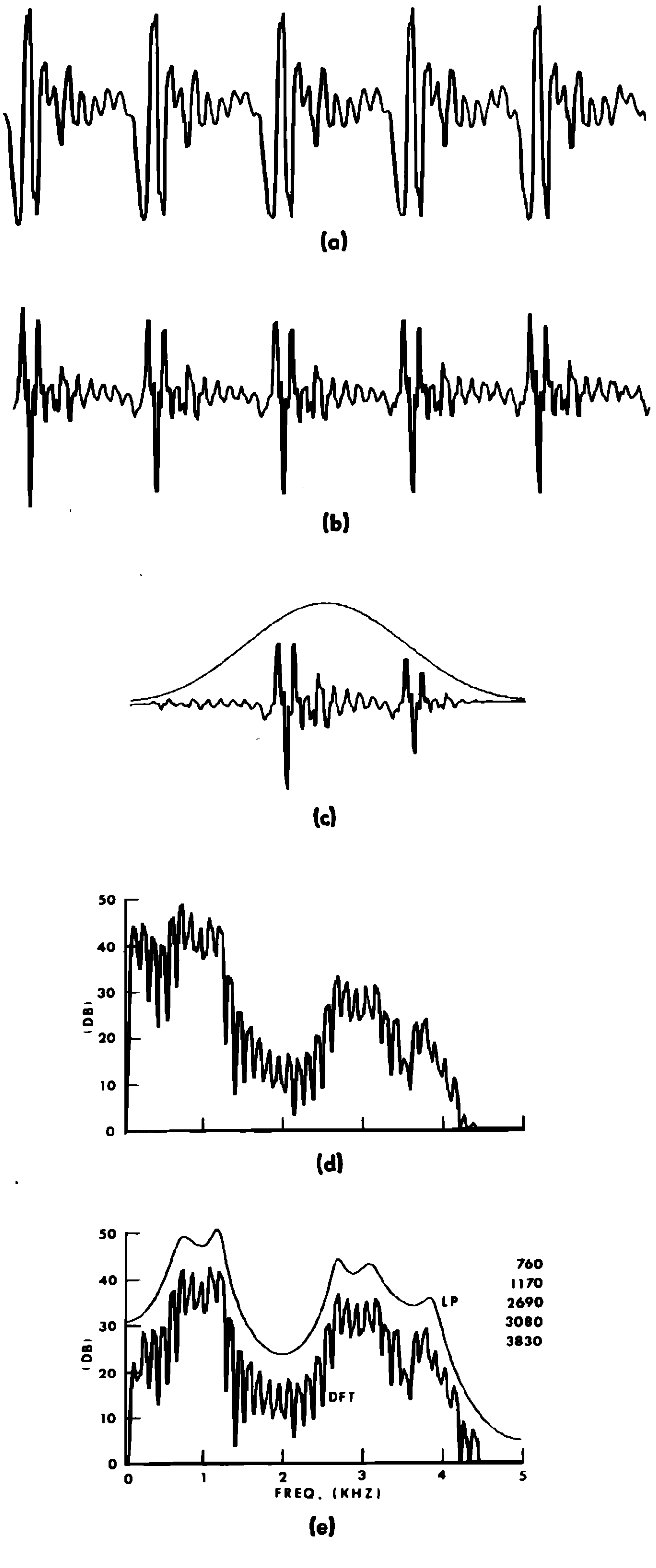


FIG. 16. The 25.6-ms (256 point) waveform segment extracted from a natural [ɑ] with a fundamental frequency of 122 Hz, shown in (a) has been first differenced in (b), and multiplied by a Kaiser window in (c). The magnitude of the discrete Fourier transform of the non-preemphasized windowed waveform is shown in (d) and the DFT and magnitude of a linear prediction spectrum of the preemphasized windowed waveform is plotted in (e). Also listed are the frequencies of local maxima in the linear prediction spectrum; these maxima are usually good estimates of formant frequencies.

V. SYNTHESIS STRATEGY

General strategies for the synthesis of English syllables are beyond the scope of this paper, but the following paragraphs are intended to provide typical parameter values for a number of static speech sounds. The values presented below in Tables II and III should constitute a good starting point for the synthesis of an utterance if the procedures outlined below are adopted. The steps used to synthesize an utterance in our laboratory are described and a simple example is presented.

A. Synthesis of vowels

The control parameters that are usually varied to generate an isolated vowel are the amplitude of voicing AV, the fundamental frequency of vocal fold vibrations $F0$, the lowest three formant frequencies $F1$, $F2$, and $F3$, and bandwidths $B1$, $B2$, and $B3$. The fourth and fifth formant frequencies may be varied to simulate spectral details, but this is not essential for high intelligibility. To create a natural breathy vowel termination, the amplitude of aspiration AH and the amplitude of quasi-sinusoidal voicing AVS can be activated.

Table II includes suggested target values for variable

TABLE II. Parameter values for the synthesis of selected vowels. If two values are given, the vowel is diphthongized or has a schwa-like offglide in the speech of the author. The amplitude of voicing, AV, and fundamental frequency, $F0$, must also be given contours appropriate for an isolated vowel.

Vowel	$F1$	$F2$	$F3$	$B1$	$B2$	$B3$
[i^y]	310	2020	2960	45	200	400
	290	2070	2960	60	200	400
[ɪ^ə]	400	1800	2570	50	100	140
	470	1600	2600	50	100	140
[e^y]	480	1720	2520	70	100	200
	330	2020	2600	55	100	200
[ɛ^ə]	530	1680	2500	60	90	200
	620	1530	2530	60	90	200
[æ^ə]	620	1660	2430	70	150	320
	650	1490	2470	70	100	320
[ɑ]	700	1220	2600	130	70	160
[ɔ^ə]	600	990	2570	90	100	80
	630	1040	2600	90	100	80
[ʌ]	620	1220	2550	80	50	140
[o^ω]	540	1100	2300	80	70	70
	450	900	2300	80	70	70
[ʊ^ə]	450	1100	2350	80	100	80
	500	1180	2390	80	100	80
[u^ω]	350	1250	2200	65	110	140
	320	900	2200	65	110	140
[ɚ]	470	1270	1540	100	60	110
	420	1310	1540	100	60	110
[a^y]	660	1200	2550	100	70	200
	400	1880	2500	70	100	200
[a^ω]	640	1230	2550	80	70	140
	420	940	2350	80	70	80
[o^y]	550	960	2400	80	50	130
	360	1820	2450	60	50	160

control parameters that are used to differentiate among English vowels (Klatt, in preparation). Formant frequency and bandwidth targets were obtained by trial-and-error spectral matching to a large set of CV syllables spoken by the author. Bandwidth values are often larger than closed-glottis values obtained by Fujimura and Lindqvist (1971) because the bandwidths of Table II have been adjusted to take into account changes to observed formant amplitudes caused by factors such as glottal losses and irregularities in the voicing source spectrum.

The amplitude of the voicing source, AV, is set to about 60 dB for a stressed vowel, and falls gradually by a few dB near the end of the syllable. The fundamental frequency contour for an isolated vowel can be approximated by a linear falling $F0$, e.g., from 130 to 100 Hz.

B. Synthesis of consonants

If the vowel is to be preceded by a consonant, additional control parameters may have to be varied. Table III includes target values for variable control parameters that are used to synthesize portions of English consonants (frication spectra of fricatives, burst spectra of plosives, nasal murmurs for nasals, and steady portions of sonorants).

The sonorant consonants [w], [y], [r], and [l] are similar to vowels and require the same set of control parameters to be varied in order to differentiate among them. Formant values given in Table III for the prevocalic sonorants [r] and [l] depend somewhat on the following vowel. The source amplitude, AV, for a prevocalic sonorant should be about 10 dB less than in the vowel. The sonorant [h] (not shown in Table III) can be synthesized by taking formant frequency and bandwidth parameters from the following vowel, increasing the first formant bandwidth to about 300 Hz, and replacing voicing by aspiration.

The fricatives characterized in Table III include both voiceless fricatives (AF = 60, AV = 0, AVS = 0) and voiced fricatives (AF = 50, AV = 47, AVS = 47). Formants to be excited by the frication noise source are determined by the amplitude controls $A2$, $A3$, $A4$, $A5$, $A6$, and AB. Values presented in the table are appropriate only for consonants before front vowels.

The amplitude of the parallel second formant, $A2$, is zero for all of these consonants before front vowels, but the second formant is a front cavity resonance for velars before nonfront vowels and $A2$ is should be set to about 60 dB. The values given for $F2$ and $F3$ are not only valid during the fricative, but also can serve as "loci" for the characterization of the consonant–vowel formant transitions before front vowels (Klatt, in preparation). These are virtual loci in that formant frequency values observed at the onset of voicing are somewhere between the locus and the vowel target fre-

TABLE III. Parameter values for the synthesis of selected components of English consonants before front vowels (see text for source amplitude values).

Sonor	$F1$	$F2$	$F3$	$B1$	$B2$	$B3$						
[w]	290	610	2150	50	80	60						
[y]	260	2070	3020	40	250	500						
[r]	310	1060	1380	70	100	120						
[l]	310	1050	2880	50	100	280						
Fric.	$F1$	$F2$	$F3$	$B1$	$B2$	$B3$	$A2$	$A3$	$A4$	$A5$	$A6$	AB
[f]	340	1100	2080	200	120	150	0	0	0	0	0	57
[v]	220	1100	2080	60	90	120	0	0	0	0	0	57
[θ]	320	1290	2540	200	90	200	0	0	0		28	48
[ð]	270	1290	2540	60	80	170	0	0	0	0	28	48
[s]	320	1390	2530	200	80	200	0	0	0	0	52	0
[z]	240	1390	2530	70	60	180	0	0	0	0	52	0
[š]	300	1840	2750	200	100	300	0	57	48	48	46	0
Affricate												
[č]	350	1800	2820	200	90	300	0	44	60	53	53	0
[ǰ]	260	1800	2820	60	80	270	0	44	60	53	53	0
Plosive												
[p]	400	1100	2150	300	150	220	0	0	0	0	0	63
[b]	200	1100	2150	60	110	130	0	0	0	0	0	63
[t]	400	1600	2600	300	120	250	0	30	45	57	63	0
[d]	200	1600	2600	60	100	170	0	47	60	62	60	0
[k]	300	1990	2850	250	160	330	0	53	43	45	45	0
[g]	200	1990	2850	60	150	280	0	53	43	45	45	0
Nasal	FNP	FNZ	$F1$	$F2$	$F3$	$B1$	$B2$	$B3$				
[m]	270	450	480	1270	2130	40	200	200				
[n]	270	450	480	1340	2470	40	300	300				

quency—the amount of virtual transition being dependent on formant-cavity affiliations. The specification of frication spectra in the table is accurate only before front vowels in the speech of the author. Before back and rounded vowels, systematic changes are observed to the fricative spectra because of anticipatory coarticulation. Specification of control parameter values for consonants in any phonetic environment is beyond the scope of the present paper, but appropriate values are easily found by trial-and-error matching to natural speech.

The affricate parameters in Table III refer to the fricative portion of the affricate. Similarly, the plosive parameters in Table III refer to the brief burst of frication noise generated at plosive release. Formant frequency values again serve as loci for predicting formant positions at voicing onset. In addition to differences in source amplitudes, voiced and voiceless consonants differ in that $F1$ is higher and $B1$ is larger when the glottis is open.

The parameters that are used to generate a nasal murmur include the nasal pole and zero frequencies FNP and FNZ. The nasal pole and zero are used primarily to approximate vowel nasalization at nasal release by splitting $F1$ into a pole-zero-pole complex. The details of nasal murmurs that have been described by Fujimura (1962) are approximated by formant bandwidth adjustments rather than by the theoretically correct method of pole-zero insertion. The reason is that it is not possible to simulate both the higher-frequency pole-zero details of nasal murmurs and vowel nasalization simultaneously without moving the frequency of the nasal pole and zero very fast at release, which would generate an objectionable click in the output, and vowel nasalization has been found to be perceptually more important. A nasalized vowel is generated by increasing $F1$ by about 100 Hz, and by setting the frequency of the nasal zero to be the average of this new $F1$ value and 270 Hz (the frequency of the fixed nasal pole).

Not included in Tables II and III are unstressed allophones, postvocalic allophones, flaps, glottal stops, voicebars, and consonant clusters. Characterization of even the static properties of these phonetic segments is beyond the scope of the present paper, but it it hoped that the information contained in the tables can be combined with the synthesis strategy described below for the rapid synthesis of an arbitrary utterance.

C. Synthesis of a novel utterance

The first step in the preparation of a new utterance is to obtain a natural model. The availability of a naturally spoken utterance is important because experience has shown that not all of the synthesis control parameter values can be deduced from theoretical considerations, and an unnatural, marginally intelligible synthetic utterance often results if one relies entirely on available theory.

A broadband spectrogram of the spoken word is then produced in order to visualize general acoustic characteristics of the utterance and determine the approximate duration of its component acoustic events. Computer analyses described below can provide much of the same information, but it is easier to visualize the time–frequency–intensity relations in the recording if a spectrogram is available.

The utterance is then 5-kHz low-pass filtered, digitized at 10 000 sample/s, and saved as a disk file for subsequent direct comparisons with the synthetic imitation that is to be created. The utterance "string", as spoken by a female talker, will be used as an illustration of the steps required to achieve a close acoustic match to a natural model of a word. A spectrogram of the recorded word is shown at the left in Fig. 17.

The intensity-versus-time plot shown in Fig. 18(a) was obtained by computing rms energy in dB every 10 ms, using a 25.6-ms Kaiser weighting window centered on the time of each displayed point. Ultimately, the synthetic utterance should have a matching intensity pattern, although it is not eady to deduce values for the various amplitude controls that will result in a close match to this contour on the first try.

The voiced portion of the word "string" was further analyzed by computer routines that extract an estimate of voicing fundamental frequency, as shown in Fig. 18(b), and formant frequency trajectories, as shown in Fig. 18(c). Observed formant motions can be used directly to specify formant frequency control parameters $F1$ through $F5$ during voiced portions of the utterance (only four formants are seen below 5 kHz for many female speakers, in which case the control parameter NFC that controls the number of formants in the cascade branch of the synthesizer would be set to 4). The third formant is invisible in the formant track of Fig. 18, but its position can be deduced by examination of the spectra in Fig. 19.

Linear prediction spectra sampled at various times in the natural utterance are plotted in Fig. 19. The spectra were all obtained in the manner described previously, except that the average spectrum of the [s] frication noise that is shown in Fig. 19(a) was obtained using a time weighting window having a longer effective duration of 40 ms. The longer duration window provides a better estimate of stationary spectra.

The general procedure for synthesis of voiced sonorants (the [rɪŋ] of "string") is to use information such as appears in Figs. 17–19 to (1) set the number of formants in the cascade vocal tract model, (2) adjust the fundamental frequency contour to within 2 Hz and formant frequencies to within about 5% by matching dig-

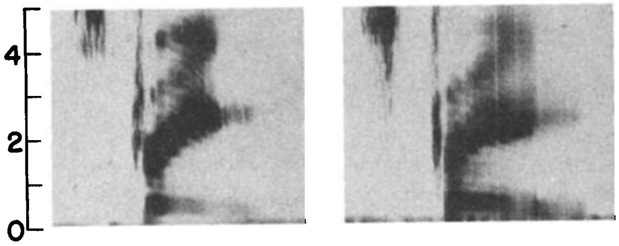


FIG. 17. Broadband spectrograms are compared of a natural and synthetic word, "string," spoken by a female talker.

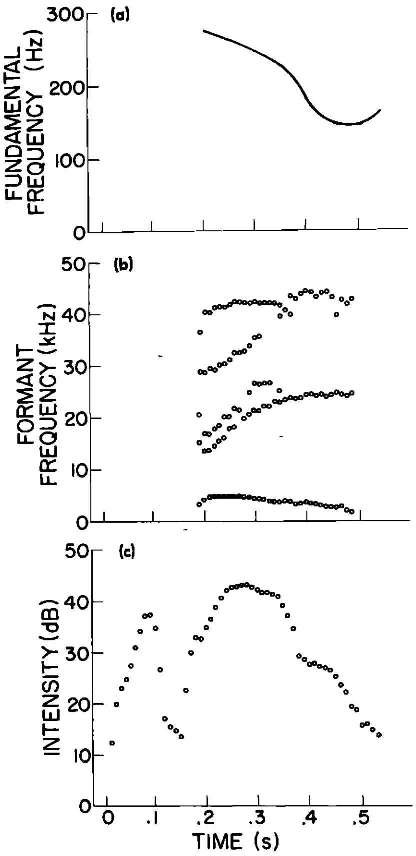


FIG. 18. The natural utterance "string" was analyzed by computer programs that produce plots of (a) fundamental frequency, (b) formant frequencies, and (c) intensity versus time.

itized waveform and spectra every 10 ms during transitions and every 50 ms during quasi-stationary intervals, (3) modify the source spectrum and/or formant bandwidths in order to set relative formant amplitudes to within about 2 dB, and (4) use the voicing source amplitude control, AV, to set the overall intensity contour to within about 2 dB. An intelligible utterance and a convincing imitation of the speaker are likely to result from satisfying these criteria (Holmes, 1961).

Detailed examination of the Fourier spectra of sonorant intervals indicates the presence of some aspiration noise during voicing for this speaker, i.e., the spectrum is less than perfectly harmonic at higher frequencies. Therefore, AH was adjusted to follow the same contour as AV, but with a value about 3 dB less. This produces a somewhat breathy voice quality that is typical of many female talkers. It also (fortunately) alleviates a fundamental problem of digital synthesis having to do with the perfect periodicity of portions of the synthesis when control parameter values do not change very much. Voiced sounds are often observed to become unpleasantly mechanical sounding as fundamental frequency is increased for female and children's voices, but a little aspiration noise breaks up the regularity of the harmonic structure.

The general procedure for synthesis of fricatives, affricates, and plosive bursts is to use the information in Figs. 17–19 to (1) select which formants are excited on the basis of continuity between the spectral peaks of the noise and formant peaks in adjacent voiced intervals, (2) set formant frequencies, (3) use formant bandwidth controls and parallel-branch formant amplitude controls to adjust the amplitude and width of peaks in the frication spectra, and (4) use AF, AV, and AVS to adjust the overall intensity contour and the mixture of periodic voicing to aperiodic noise (the discrete Fourier transform is useful for this purpose since it

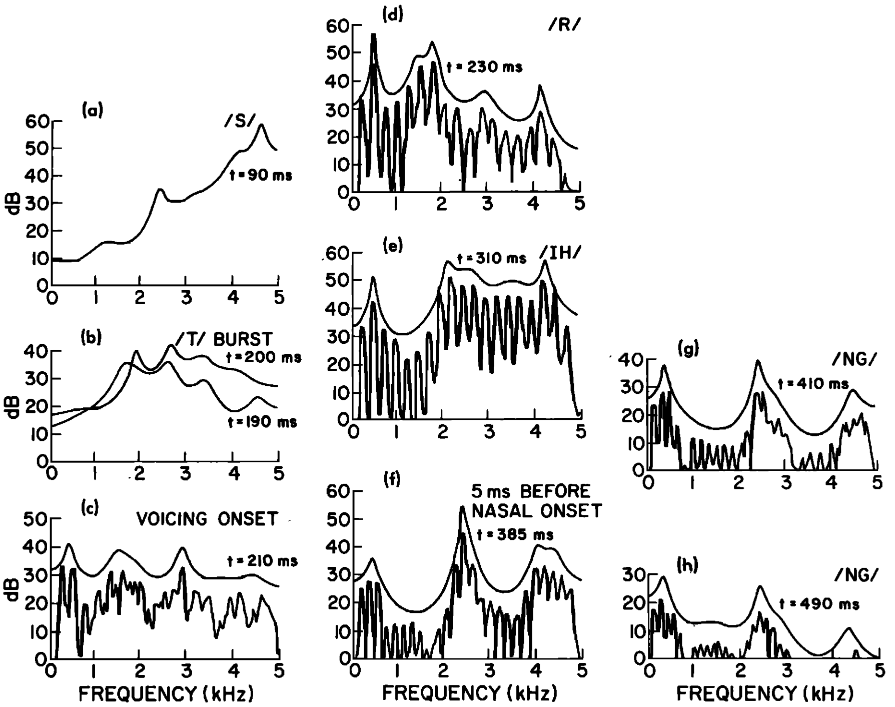


FIG. 19. Linear prediction spectra are plotted (a) during [s] noise, (b) centered on the [t] burst, (c) at voicing onset for the [r], (d) during the sonorant-vowel transition, (e) at the midpoint of the vowel, (f) just prior to closure, and (g)–(h) during the nasal murmur.

indicates individual harmonic amplitudes). For those formants that are invisible during frication generation, formant frequencies are filled in from continuity constraints and theoretical considerations.

For example, in order to synthesize the observed [s] spectrum, it is necessary to produce a strong spectral peak near 5 kHz. To do this, $F6$ can be positioned to 4.9 kHz and turned on at the appropriate time. Default values are chosen for the other formant frequencies and formant bandwidths are adjusted on a trial-and-error basis in order to match the [s] spectrum of Fig. 19.

VI. CONCLUSIONS

We have described a flexible software synthesizer that can run on any laboratory computer having sufficient core, peripheral equipment, and a Fortran compiler. The software is included as an appendix. The synthesizer and an associated control program can be used by a novice with minimal training and thus can serve as a research tool for those whose primary interest is not speech synthesis per se, but, e.g., the perception of speech and the relative perceptual importance of different acoustic cues to phonetic contrasts. It should be easier to replicate the results of perceptual experiments performed using this synthesizer because the synthesizer is fully documented here, and hopefully the control parameter values for the stimuli of an experiment will be carefully specified in the publication. The synthesizer may also find application in programs for speech synthesis by rule (Klatt, 1976a), for computer audio response, and in a reading machine for the blind (Allen, 1977).

Experience over the past few years suggest that the synthesizer is sufficiently flexible to generate good imitations to most if not all male and female voices. It also appears possible to synthesize any phonetic sequence of English with excellent intelligibility if the steps outlined in the previous section are followed. A consonant–vowel synthesis cookbook that is based on this synthesizer is in preparation (Klatt, in preparation). Intelligibility tests of 337 different CV syllables produced by the rules contained in the cookbook indicate that better than 98% of the vowels and 95% of the consonants are identified correctly by trained phoneticians who are unfamiliar with synthetic speech.

ACKNOWLEDGMENT

This research was supported by an NIH grant.

APPENDIX A: EXTERNAL 5000 Hz LOW-PASS FILTER

An external low-pass filter is required to convert the staircase waveform that comes from the D/A converter into an analog signal containing energy only below 5000 Hz. A suitable filter for this purpose has been designed and built by Dr. Joe Teirney of the M. I. T. Lincoln Laboratories. It is a seventh-order passive eliptic low-pass filter with component values that are indicated in Fig. A1.

The steps required to fabricate this filter are (1) wind the coils on something like Ferroxcube Corp. 3622A-600387 to plus-or-minus 0.5% (coils should have Q's of over 100 at the frequencies to which they are tuned), (2) use trim capacitors to tune the three L-C combinations along the top of the ladder network to have resonant frequencies in isolation of 7765, 5030, and 5427 Hz, respectively, (3) trim the remaining capacitors of the circuit to 0.5%, and (4) use 1% resistors.

The magnitude of the filter transfer function is shown at the bottom of Fig. A1. All frequency components in the input signal above 5 kHz are attenuated by at least 40 dB, while frequency components below 4780 Hz are within 0.6 dB of no attenuation at all.

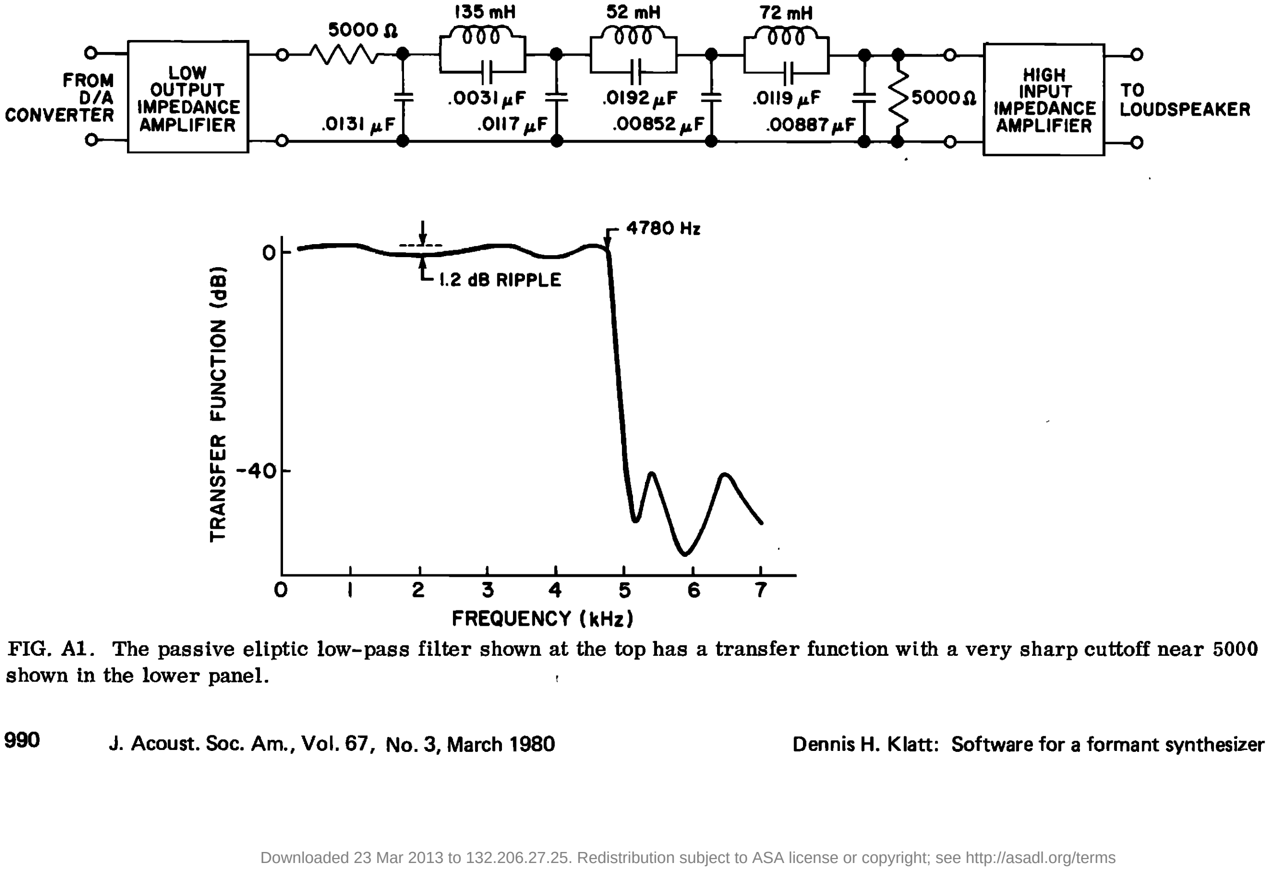


FIG. A1. The passive eliptic low-pass filter shown at the top has a transfer function with a very sharp cuttoff near 5000 Hz, as shown in the lower panel.

APPENDIX B. FORTRAN LISTINGS

This appendix contains listings of (1) the synthesizer control program, HANDSY.FOR, (2) the subroutine for converting user-oriented control parameter data into difference equation coefficients, PARCOE.FOR, and (3) the subroutine for converting these coefficients into a synthetic waveform, COEWAV.FOR . Also included are two small subroutines for converting from decibels to linear amplitudes, GETAMP.FOR, and for converting from formant frequency and bandwidth to difference-equation coefficients, SETABC.FOR.

As listed below, the programs should compile and run on, e.g., any Digital Equipment Corporation PDP-11 having sufficient core. Some of the Fortran input–output instructions may have to be changed for other computer environments. For a machine with insufficient core, it may be possible to rewrite the routine HANDSY.FOR so as to use the disk for storage of parameter/waveform data instead of the 10 050-word core array IWAVE. The arrays MAXVAL and MINVAL in HANDSY.FOR are included primarily to detect accidental typing errors and conceptual errors on the part of naive users; these values may have to be changed in order to synthesize unusual stimuli.

```
C       HANDSY.FOR              D. KLATT            6/1/79
C
C           SPECIFY AN ARRAY OF CONTROL PARAMETER DATA
C           AND SYNTHESIZE A SPEECH WAVEFORM
C
C       LOAD WITH PARCOE.FOR, COEWAV.FOR, SETABC.FOR, GETAMP.FOR
C
C   IF THIS PROGRAM DOES NOT FIT INTO CORE, DECREASE D(10050),
C   IWAVE(10050), AND WSIZE ALL BY THE SAME FRACTION
C
        IMPLICIT INTEGER (A-Z)
        REAL DB,DBLPNT,EPSLON,XMAXWA
C   EACH OF THE FOLLOWING VARIABLES HOLDS UP TO 5 ASCII CHARACTERS
        REAL QUIT,NAMEV,NAMES(39),NAMEX(39)
        DIMENSION MAXVAL(39),MINVAL(39),VALUES(39),IPAR(39)
        DIMENSION VARPAR(39),LOC(39),LOCSAV(39),D(10050),IWAVE(10050)
        DIMENSION COEFIC(50)
        COMMON /PARS/IPAR
        COMMON /COEFS/ COEFIC
        EQUIVALENCE (D(1),IWAVE(1))
C
C   3-CHARACTER SYMBOL FOR EACH OF 39 CONTROL PARAMETER VALUES
        DATA NAMES /'AV','AF','AH','AVS','FO','F1','F2','F3','F4','FNZ'
        1,'AN','A1','A2','A3','A4','A5','A6','AB','B1','B2'
        1,'B3','SW','FGP','BGP','FGZ','BGZ','B4','F5','B5','F6'
        1,'B6','FNP','BNP','BNZ','BGS','SR','NWS','GO','NFC'/
C
C   MAXIMUM POSSIBLE VALUE FOR EACH OF 39 CONTROL PARAMETERS
        DATA MAXVAL/80,80,80,80,500,900,2500,3500,4500,700
        1,80,80,80,80,80,80,80,80,1000,1500
        1,2000,1,600,2000,5000,10000,3000,4900,4000,4999
        1,2000,500,500,500,1000,20000,200,80,6/
C
C   MINIMUM POSSIBLE VALUE FOR EACH OF 39 CONTROL PARAMETERS
        DATA MINVAL/0,0,0,0,0,150,500,1300,2500,200
        1,0,0,0,0,0,0,0,0,40,40
        1,40,0,0,100,0,100,100,3500,150,4000
        1,200,200,50,50,100,5000,1,0,4/
C
C   DETERMINATION OF VARIABLE (=1 OR =2) OR CONSTANT (=0) PRAMETERS
C   (PROGRAM SETS =2 IF ACTUALLY VARIED)
        DATA VARPAR/1,1,1,1,1,1,1,1,1,1
        1,0,0,1,1,1,1,1,1,1,1
        1,1,0,0,0,0,0,0,1,0,0
        1,0,0,0,0,0,0,0,0,0/
C
C   DEFAULT VALUES FOR EACH OF 39 CONTROL PARAMETERS
        DATA VALUES/0,0,0,0,0,450,1450,2450,3300,250
        1,0,0,0,0,0,0,0,0,50,70
        1,110,0,0,100,1500,6000,250,3750,200,4900
        1,1000,250,100,100,200,10000,50,47,5/
C
C   SIZE OF PARAMETER AND WAVEFORM ARRAYS THAT RESIDE IN CORE


        DATA WSIZE/10050/
C
C   NAMES OF SOME RESPONSE CHARACTERS
        DATA QUIT,QUIT1,YES,NO,VAR,CON/'Q','Q','Y','N','V','C'/
C
1000    WRITE (5,1010)
1010    FORMAT (/' KLATT CASCADE/PARALLEL FORMANT SYNTHESIZER  6/1/79
        1'//)
C
C     SEE IF FILE PARAM.DOC EXISTS; IF SO, READ CONFIGURATION
        OPEN(UNIT=1,NAME='PARAM.DOC',ACCESS='SEQUENTIAL',ERR=1140)
        OPENPA=1
        WRITE (5,1020)
1020    FORMAT (' READING INITIAL SYNTHESIZER CONFIGURATION FROM FILE
        1  "PARAM.DOC"'/)
        DO 1060 M=1,13
        N=M+13
        N1=M+26
        READ (1,2617) DUMMY,NAMES(M),VARPAR(M),VALUES(M),DUMMY,NAMES(N)
        1,VARPAR(N),VALUES(N),DUMMY,NAMES(N1),VARPAR(N1),VALUES(N1)
1060    CONTINUE
C
C     CHANGE CONFIGURATION, CHANGE WHICH PARS ARE VARIABLE
1140    WRITE (5,1160)
1160    FORMAT (' PRINT AND/OR CHANGE CONFIGURATION (Y,Q):'$)
1170    READ (5,1180,ERR=1140) ANSWER
1180    FORMAT (A1)
1185    IF (ANSWER.EQ.QUIT1) GO TO 1740
        GO TO 1685
1190    WRITE (5,1220)
1220    FORMAT (/' NAME OF PARAMETER TO BECOME VAR OR CON (QUIT="Q"):'$)
1240    READ (5,1260,ERR=1190) NAMEV
1260    FORMAT (A3)
1270    IF (NAMEV.EQ.QUIT) GO TO 1500
        DO 1280 N=1,39
        IF (NAMEV.EQ.NAMES(N)) GO TO 1320
1280    CONTINUE
        WRITE (5,1300) NAMEV
1300    FORMAT (' ',A5,', TYPING ERROR, TRY AGAIN')
        WRITE (5,1555) (NAMES(M),M=1,39)
        GO TO 1190
1320    MODPAR=YES
        IF (N.LT.35) GO TO 1330
        WRITE (5,1325) NAMES(N)
1325    FORMAT (' PARAMETER ',A3,' CANNOT BE MADE VARIABLE')
        GO TO 1190
1330    IF (VARPAR(N).NE.0) GO TO 1380
1340    VARPAR(N)=1
        WRITE (5,1360)NAMEV
1360    FORMAT (' ',A3,' IS NOW A VARIABLE')
        GO TO 1190
1380    IF (VARPAR(N).NE.2) GO TO 1390
C     IF VARIED IN PREVIOUS SYNTH ATTEMPT, CAN'T MAKE INTO A CONSTANT

        WRITE (5,1385) NAMEV
1385    FORMAT (' ',A3,' CAN NO LONGER BE MADE A CONSTANT')
        GO TO 1190
1390    VARPAR(N)=0
1400    WRITE (5,1420) NAMEV
1420    FORMAT (' ',A3,' IS NOW A CONSTANT')
1440    FORMAT (' DONE')
        GO TO 1190
C
C     CHANGE DEFAULT VALUE FOR A PARAMETER
1500    WRITE (5,1520)
1520    FORMAT (' NAME OF PARAMETER WHOSE
        1 DEFAULT VALUE TO BE CHANGED (QUIT="Q"):'$)
        READ (5,1260,ERR=1550) NAMEV
1530    IF (NAMEV.EQ.QUIT) GO TO 1140
        DO 1540 N=1,39
        IF (NAMEV.EQ.NAMES(N)) GO TO 1560
1540    CONTINUE
1550    WRITE (5,1300) NAMEV
        WRITE (5,1555) (NAMES(M), M=1,39)
1555    FORMAT (' PARS= ',13A4)
        GO TO 1500
1560    IF ((N.NE.36).AND.(N.NE.37)) GO TO 1570
C     DON'T CHANGE NWS OR SR IF READING FROM PARAMETER FILE
        IF (OPENPA.EQ.0) GO TO 1570
        WRITE (5,1565) NAMEV
1565    FORMAT (' CANNOT CHANGE THE VALUE OF ',A3,' ANYMORE')
        GO TO 1500
1570    WRITE (5,1580) NAMEV
1580    FORMAT (' NEW DEFAULT VALUE FOR ',A3,'='$)
        READ (5,1900,ERR=1560) VALUE
1590    IF (VALUE.LE.MAXVAL(N)) GO TO 1620
        WRITE (5,1600)VALUE,MAXVAL(N)
1600    FORMAT (' ',I6,' EXCEEDS MAXIMUM OF ',I5,' TRY AGAIN')
        GO TO 1560
1620    IF (VALUE.GE.MINVAL(N)) GO TO 1660
        WRITE (5,1640)VALUE,MINVAL(N)
1640    FORMAT (' ',I5,' IS LESS THAN MINIMUM=',I5,' TRY AGAIN')
        GO TO 1560
1660    MODPAR=YES
        VALUES(N)=VALUE
        WRITE (5,1440)
        GO TO 1500
C
C     PRINT CONFIGURATION
1680    IF (MODPAR.EQ.NO) GO TO 1740
1685    WRITE (5,1690)
1690    FORMAT (' CURRENT CONFIGURATION (NAME,VAR/CON,DEFAULT-VALUE):')
        DO 1720 M=1,13
        N=M+13
        N1=M+26
        WRITE (5,1700) M,NAMES(M),VARPAR(M),VALUES(M),N,NAMES(N)
        1,VARPAR(N),VALUES(N),N1,NAMES(N1),VARPAR(N1),VALUES(N1)


1700    FORMAT(' ',I2,'  ',A4,I2,I6,2('        ',I2,'  ',A4,I2,I6))
1720    CONTINUE
        GO TO 1190
C
C     COUNT NUMBER OF VARIABLE PARAMETERS, NVAR,
C     AND PLACE NAMES IN NAMEX(NVAR)
1740    NSAMP=VALUES(37)
        DENOM=VALUES(36)/10
        DELTAT=(NSAMP*100)/DENOM
        NVAR=0
        DO 1760 N=1,39
        IF (VARPAR(N).EQ.0) GO TO 1760
        NVAR=NVAR+1
        LOC(NVAR)=N
        NAMEX(NVAR)=NAMES(N)
1760    CONTINUE
        IF (NVAR.GT.0) GO TO 1800
        WRITE(5,1780)
1780    FORMAT (' ILLEGAL CONFIG, NO VARIABLE PARAMS, TRY AGAIN')
        GO TO 1680
1800    MAXDUR=((WSIZE/NSAMP)*DELTAT)-20
        WRITE (5,1820) NVAR
1820    FORMAT (/' THERE ARE ',I2,' VARIABLE PARAMETERS')
        WRITE (5,1840) DELTAT
1840    FORMAT (' PARAMETERS ARE TO BE SPECIFIED EVERY ',I2,' MSEC')
1860    IF (OPENPA.EQ.0) GO TO 1870
        READ (1,2625) VALUE
        WRITE (5,1867) VALUE
1867    FORMAT (' LENGTH OF UTTERANCE IN MSEC = ',I5)
        GO TO 1910
1870    WRITE (5,1880) MAXDUR
1880    FORMAT (' DESIRED LENGTH OF UTTERANCE IN MSEC (MAX=',I4,'):'$)
1885    READ (5,1900,ERR=1860) VALUE
1900    FORMAT (I5)
1910    IF (VALUE.GE.DELTAT) GO TO 1920
        WRITE (5,1300) NAMEV
        GO TO 1860
1920    IF (VALUE.LE.MAXDUR) GO TO 1960
        WRITE (5,1940) VALUE,MAXDUR
1940    FORMAT (' ',I4,' ILLEGAL, MAXIMUM DURATION=',I4,', TRY AGAIN')
        GO TO 1860
1960    UTTDUR=VALUE
C
```

```
C      INSERT DEFAULT VALUES INTO PARAMETER TRACKS
         NSAMT1=((UTTDUR+20)/DELTAT)-1
         DO 2000 M=0,NSAMT1
         M1=M*NSAMP
         M2=0
         DO 1980 N=1,39
         IF (VARPAR(N).EQ.0) GO TO 1980
         M2=M2+1
         D(M1+M2)=VALUES(N)
1980     CONTINUE

2000     CONTINUE
         WRITE (5,2020)
2020     FORMAT (/' DEFAULT VALUES INSERTED IN PARAMETER TRACKS')
C
C      PUT VARIABLE DATA FROM FILE PARAM.DOC INTO PARAMETER TRACKS
2040     IF (OPENPA.EQ.0) GO TO 2050
         WRITE (5,2041)
2041     FORMAT (' READING VARIABLE PARAMETRIC DATA FROM FILE
         1  "PARAM.DOC"')
         READ (1,2043) DUMMY,(DUMMY,M=1,NVAR1)
2043     FORMAT (27A5)
         NVAR1=0
         DO 2045 N=1,NVAR
         IF (VARPAR(LOC(N)).NE.2) GO TO 2045
         NVAR1=NVAR1+1
         LOCSAV(NVAR1)=N
2045     CONTINUE
         IF (NVAR1.GT.0) GO TO 2047
         WRITE (5,1780)
         STOP
2047     IF (NVAR1.GT.26) NVAR1=26
         NSAMT1=(UTTDUR/DELTAT)-1
         DO 2048 M=0,NSAMT1
         M1=M*NSAMP
         READ (1,2660) TIME,(D(LOCSAV(N)+M1),N=1,NVAR1)
2048     CONTINUE
         CLOSE(UNIT=1)
C
C      ACCEPT MODIFICATIONS TO PARAMETER TRACKS
2050     OLDTIM=0
         SETPNT=NO
         MAXD1=UTTDUR-DELTAT
         WRITE (5,2060)
2060     FORMAT (/' NAME OF PARAMETER TRACK TO BE MODIFIED (QUIT="Q"):'$)
2065     READ (5,1260,ERR=2090) NAMEV
2075     IF (NAMEV.EQ.QUIT) GO TO 2600
         DO 2080 N=1,NVAR
         IF (NAMEV.EQ.NAMEX(N)) GO TO 2120
2080     CONTINUE
         WRITE (5,1300) NAMEV
2090     WRITE (5,2100) (NAMEX(M), M=1,NVAR)
2100     FORMAT (' VARIABLE PARS= ',10A4)
         GO TO 2050
2120     CONTINUE
         VARPAR(LOC(N))=2
         MAXV=MAXVAL(LOC(N))
         MINV=MINVAL(LOC(N))
2180     WRITE (5,2200)
2200     FORMAT ('  T='$)
2220     READ (5,2240,ERR=2550) TIME
2240     FORMAT (I3)
C
C      QUIT DRAWING CURRENT PARAMETER CONTOUR?

         IF ((TIME.EQ.0).AND.(SETPNT.EQ.YES)) GO TO 2050
         IF (TIME.LT.0) GO TO 2050
C      MAKE SURE LEGAL TIME
         IF (TIME.GE.OLDTIM) GO TO 2280
2255     WRITE (5,2260) TIME,OLDTIM
2260     FORMAT (' ILLEGAL TIME=',I3,', LESS THAN OLDTIM=',I3)
         GO TO 2180
2280     IF (TIME.LE.MAXD1) GO TO 2320
         WRITE (5,2300) TIME,MAXD1
2300     FORMAT (' ILLEGAL TIME=',I3,', GREATER THAN MAX=',I3)
         GO TO 2180
2320     NPTS=TIME/DELTAT
         TIME=NPTS*DELTAT
         POINTR=((NPTS)*NSAMP)+N
2330     WRITE (5,2340)
2340     FORMAT ('  V='$)
2345     READ (5,1900,ERR=2550) VALUE
C
C      SEE IF LEGAL VALUE
2369     IF (VALUE.LE.MAXV) GO TO 2400
2370     WRITE (5,2371) MINV,MAXV
2371     FORMAT (' VMIN=',I5,',  VMAX=',I5)
         GO TO 2330
2400     IF (VALUE.GE.MINV) GO TO 2420
         GO TO 2370
2420     IF ((SETPNT.EQ.YES).AND.(TIME.GE.(OLDTIM+DELTAT))) GO TO 2480
C
C      SET A POINT
         D(POINTR)=VALUE
2460     OLDTIM=TIME
         OLDVAL=VALUE
         SETPNT=YES
         GO TO 2180
C
C      DRAW A LINE
2480     NPTS=(TIME-OLDTIM)/DELTAT
         DVALUE=VALUE-OLDVAL
         EPSLON=0.1
         IF (DVALUE.LT.0) EPSLON=-EPSLON
         TIME1=OLDTIM/DELTAT
         DO 2500 M=1,NPTS
         DBLPNT=FLOAT(M)*FLOAT(DVALUE)
         DBLPNT=DBLPNT/FLOAT(NPTS)
         VALUE2=OLDVAL+IFIX(DBLPNT+EPSLON)
         POINTR=((TIME1+M)*NSAMP)+N
2500     D(POINTR)=VALUE2
         GO TO 2460
C
C      UNRECOVERABLE I/O ERROR, SAVE PARAMETERS AND QUIT
2550     WRITE (5,2560)
2560     FORMAT (' UNRECOVERABLE TYPING ERROR, SAVE PARAMETERS')
C
C      MAKE FILE OF PARAMETER VALUES VS TIME THAT CAN BE LISTED


C      ON LINE PRINTER
2600     CONTINUE
         OPEN(UNIT=1,NAME='PARAM.DOC',ACCESS='SEQUENTIAL',ERR=2600)
         DO 2620 M=1,13
         N=M+13
         N1=M+26
         DUMMY='     '
         WRITE (1,2617) DUMMY,NAMES(M),VARPAR(M),VALUES(M)
         1,DUMMY,NAMES(N),VARPAR(N),VALUES(N)
         1,DUMMY,NAMES(N1),VARPAR(N1),VALUES(N1)
2617     FORMAT (' ',3(A5,A3,I2,I5))
2620     CONTINUE
         WRITE (1,2625) UTTDUR
2625     FORMAT (' ',I5)
```

```
         NVAR1=0
         DO 2630 N=1,NVAR
         IF (VARPAR(LOC(N)).NE.2) GO TO 2630
         NVAR1=NVAR1+1
         LOCSAV(NVAR1)=N
2630     CONTINUE
         IF (NVAR1.GT.0) GO TO 2640
         WRITE (5,1780)
         GO TO 2900
2640     IF (NVAR1.GT.26) NVAR1=26
         WRITE (1,2650) (NAMEX(LOCSAV(M)),M=1,NVAR1)
2650     FORMAT ('     ',26A5)
         NSAMT1=(UTTDUR/DELTAT)-1
         DO 2665 M=0,NSAMT1
         TIME=M*DELTAT
         M1=M*NSAMP
         WRITE (1,2660) TIME,(D(LOCSAV(N)+M1),N=1,NVAR1)
2660     FORMAT (I3,26I5)
2665     CONTINUE
         CLOSE(UNIT=1)
         WRITE (5,2667)
2667     FORMAT (' PARAMETER FILE  "PARAM.DOC"  SAVED')
C
C      SET ALL PARAMETERS IN ARRAY IPAR TO DEFAULT VALUES
2670     IF (PPSW.EQ.1) GO TO 2676
         WRITE (5,2675)
2675     FORMAT (/' BEGIN WAVEFORM GENERATION')
2676     DO 2680 N=1,39
2680     IPAR(N)=VALUES(N)
C
C      INITIALIZE SYNTHESIZER
         MAXWA=-1
         XMAXWA=-1.
C
C      MAIN SYNTHESIZER LOOP, PUT WAVEFORM IN IWAVE(WSIZE1)
C      ADD 20 MSEC TO DURATION TO ENSURE SIGNAL WILL DECAY TO ZERO
         NPTS=(UTTDUR+20)/DELTAT
         TIME1=0
         WSIZE1=1

         DO 2740 M=1,NPTS
         POINTR=(M-1)*NSAMP
         DO 2700 N=1,NVAR
2700     IPAR(LOC(N))=D(POINTR+N)
         CALL PARCOE(MAXWA)
         CALL COEWAV(IWAVE(WSIZE1),XMAXWA)
2740     WSIZE1=WSIZE1+NSAMP
C
C      MAKE SURE SIGNAL IS LESS THAN OR EQUAL TO 0.0 DB
         DB=20.*ALOG10(XMAXWA/32767.)
         WRITE (5,2760) DB
2760     FORMAT (' PEAK SIGNAL LEVEL
         1 IN SYNTHETIC WAVEFORM =',F6.1,' DB')
C
C      SAVE WAVEFORM FILE IWAVE(WSIZE1) ON DISK
         OPEN(UNIT=1,NAME='WAVE.I6',ACCESS='SEQUENTIAL',ERR=2800)
         WRITE (1,2775) WSIZE1
2775     FORMAT (I5)
         WRITE (1,2785) (IWAVE(M),M=1,WSIZE1)
2785     FORMAT (50I6)
         CLOSE (UNIT=1)
         WRITE (5,2795)
2795     FORMAT (/' WAVEFORM FILE  "WAVE.I6"  SAVED'//)
         GO TO 2900
2800     WRITE (5,2805)
2805     FORMAT (' DISK ACCESS ERROR DURING ATTEMPT TO SAVE WAVEFORM')
2900     STOP
         END


C        PARCOE.FOR             D.H. KLATT              8/1/78
C
C        "PARAM-TO-COEF" TRANSFORMATION SUBROUTINE
C
C      THIS PROGRAM CONVERTS SYNTHESIZER CONTROL PARAMETERS FROM ARRAY I(39)
C      INTO DIFFERENCE EQUATION CONSTANTS FOR SYNTHESIZER HARDWARE
C      STORED IN ARRAY C(50)
C
         SUBROUTINE PARCOE(INITPC)
C                INITPC INITIALIZES THIS ROUTINE IF =-1
C
         REAL IMPULS
         DIMENSION I(39),NDBSCA(12),NDBCOR(10),C(50)
C      INPUT PARAMETER VALUES (CONSTANT AND VARIABLE) PASSED THROUGH I
         COMMON /PARS/ I
         COMMON /COEFS/ C
         COMMON /PIXX/ PIT,TWOPIT
C      COEFICIENT VALUES IN C(50) ARE REAL
         EQUIVALENCE (C(1),IMPULS),(C(2),SINAMP),(C(3),AFF)
         EQUIVALENCE (C(4),AHH),(C(5),A1P),(C(6),A2P)
         EQUIVALENCE (C(7),A3P),(C(8),A4P),(C(9),A5P)
         EQUIVALENCE (C(10),A6P),(C(11),ABP),(C(12),ANPP)
         EQUIVALENCE (C(13),AGP),(C(14),BGP),(C(15),CGP)
         EQUIVALENCE (C(16),AGZ),(C(17),BGZ),(C(18),CGZ)
         EQUIVALENCE (C(19),AGS),(C(20),BGS),(C(21),CGS)
         EQUIVALENCE (C(22),A1),(C(23),B1),(C(24),C1)
         EQUIVALENCE (C(25),A2),(C(26),B2),(C(27),C2)
         EQUIVALENCE (C(28),A3),(C(29),B3),(C(30),C3)
         EQUIVALENCE (C(31),A4),(C(32),B4),(C(33),C4)
         EQUIVALENCE (C(34),A5),(C(35),B5),(C(36),C5)
         EQUIVALENCE (C(37),A6),(C(38),B6),(C(39),C6)
         EQUIVALENCE (C(40),ANP),(C(41),BNP),(C(42),CNP)
         EQUIVALENCE (C(43),ANZ),(C(44),BNZ),(C(45),CNZ)
         EQUIVALENCE (C(46),PLSTEP)
C      NAMES OF INPUT CONTROL PARAMETERS
         EQUIVALENCE (I(1),NNAV),(I(2),NNAF),(I(3),NNAH),(I(4),NNAVS)
         1,(I(5),NNF0),(I(6),NNF1),(I(7),NNF2),(I(8),NNF3),(I(9),NNF4)
         1,(I(10),NNFNZ),(I(11),NNAN),(I(12),NNA1),(I(13),NNA2)
         1,(I(14),NNA3),(I(15),NNA4),(I(16),NNA5),(I(17),NNA6)
         1,(I(18),NNAB),(I(19),NNB1),(I(20),NNB2),(I(21),NNB3)
         1,(I(22),NNSW),(I(23),NNFGP),(I(24),NNBGP),(I(25),NNFGZ)
         EQUIVALENCE (I(26),NNBGZ),(I(27),NNB4),(I(28),NNF5),(I(29),NNB5)
         1,(I(30),NNF6),(I(31),NNB6),(I(32),NNFNP),(I(33),NNBNP)
         1,(I(34),NNBNZ),(I(35),NNBGS),(I(36),NNSR),(I(37),NNNWS)
         1,(I(38),NNG0),(I(39),NNNFC)
C      CONSTANTS NEEDED BY SUBROUTINE SETABC
         DATA PI/3.14159265/
C
C      SCALE FACTORS IN DB FOR GENERAL ADJUSTMENT TO:
C                      A1  A2  A3  A4  A5  A6  AN  AB  AV   AH  AF AVS
         DATA NDBSCA/-58,-65,-73,-78,-79,-80,-58,-84,-72,-102,-72,-44/
C      INCREMENT IN DB TO FORMANT AMPLITUDES OF PARALLEL BRANCH IF
C      FORMANT FREQUENCY DIFFERENCE 50, 100, 150, ... HZ

         DATA NDBCOR/10,9,8,7,6,5,4,3,2,1/
C      PRINT INPUT PAR VALUES AT T=NTIMPR, OR AT ALL TIMES IF NTIMPR=0
         DATA NTIMPR,NPPBEG,NPPEND/-1,1,39/
C
C
C    INITIALIZE SYNTHESIZER BEFORE COMPUTING WAVEFORM CHUNK IF ARG.LT.0
100      IF (INITPC.GE.0) GO TO 130
         INITPC=0
C      SET CUMULATIVE TIME COUNTER TO ZERO
         NTIMEP=0
         NAFLAS=0
```

```
C     COMPUTE SAMPLING PERIOD T (ALL CONSTANT CONTROL PARAMETERS
C     MUST BE SET BEFORE CALLING INIT)
        SAMRAT=NNSR
        T=1./SAMRAT
        PIT=PI*T
        TWOPIT=2.*PIT
        NTIMED=(NNNWS*1000)/NNSR
C     CONVERT INHERENTLY INTEGER PARAMS TO REAL COEFICIENTS
        C(48)=NNNWS
        C(49)=NNSW
        C(50)=NNNFC
110     CONTINUE
C
C
C
C   UPDATE ALL COEFICIENTS OF HARDWARE SYNTHESIZER
C
C     COMPUTE PARALLEL BRANCH AMPLITUDE CORRECION TO F2 DUE TO F1
130     DELF1=FLOAT(NNF1)/500.
        A2COR=DELF1*DELF1
C     COMPUTE AMPLITUDE CORRECTION TO F3-5 DUE TO F1 AND F2
        DELF2=FLOAT(NNF2)/1500.
        A2SKRT=DELF2*DELF2
        A3COR=A2COR*A2SKRT
C     TAKE INTO ACCOUNT FIRST DIFF OF GLOTTAL WAVE FOR F2
        A2COR=A2COR/DELF2
C     COMPUTE AMPLITUDE CORRECTIONS DUE TO PROXIMITY OF 2 FORMANTS
        N12COR=0
        N23COR=0
        N34COR=0
        NF21=NNF2-NNF1
        IF (NF21.LT.50) GO TO 135
        IF (NF21.LT.550) N12COR=NDBCOR(NF21/50)
        NF32=NNF3-NNF2-50
        IF (NF32.LT.50) GO TO 135
        IF (NF32.LT.550) N23COR=NDBCOR(NF32/50)
        NF43=NNF4-NNF3-150
        IF (NF43.LT.50) GO TO 135
        IF (NF43.LT.550) N34COR=NDBCOR(NF43/50)
C     PRINT INPUT PARAMETERS IF NTIMPR SET TO ZERO OR TO A SPECIFIC TIME
        IF (NTIMPR.EQ.0) GO TO 135
        IF (NTIMPR.NE.NTIMEP) GO TO 146

135     WRITE (5,140) NTIMEP
140     FORMAT (' INPUT PARS AT T = ',I4,' MS')
        WRITE (5,141) (I(NPP),NPP=NPPBEG,NPPEND)
141     FORMAT (' ',13I5)
        WRITE (5,142)
142     FORMAT (' ')
145     NPAR=1
146     NTIMEP=NTIMEP+NTIMED
C     SET AMPLITUDE OF VOICING
        NDBAV=NNGO+NNAV+NDBSCA(9)
        IMPULS=GETAMP(NDBAV)
C     AMPLITUDE OF ASPIRATION
150     NDBAH=NNGO+NNAH+NDBSCA(10)
        AHH=GETAMP(NDBAH)
C     AMPLITUDE OF FRICATION
C     (IN AN ALL-PARALLEL CONFIGURATION, AF=MAX[AF,AH])
        IF ((NNAH.GT.NNAF).AND.(NNSW.EQ.1)) NNAF=NNAH
        NDBAF=NNGO+NNAF+NDBSCA(11)
        AFF=GETAMP(NDBAF)
C     ADD A STEP TO WAVEFORM AT A PLOSIVE RELEASE
        PLSTEP=0.
        IF (NNAF-NAFLAS.LT.49) GO TO 151
        PLSTEP=GETAMP(NNGO+NDBSCA(11)+44)
151     NAFLAS=NNAF
C     AMPLITUDE OF QUASI-SINUSOIDAL VOICING SOURCE
        NDBAVS=NNGO+NNAVS+NDBSCA(12)
        SINAMP=10.*GETAMP(NDBAVS)
C     SET AMPLITUDES OF PARALLEL FORMANTS A1 THRU A6
        NDB=NNA1+N12COR+NDBSCA(1)
        A1P=GETAMP(NDB)
        NDB=NNA2+N12COR+N12COR+N23COR+NDBSCA(2)
        A2P=A2COR*GETAMP(NDB)
        NDB=NNA3+N23COR+N23COR+N34COR+NDBSCA(3)
        A3P=A3COR*GETAMP(NDB)
        NDB=NNA4+N34COR+N34COR+NDBSCA(4)
        A4P=A3COR*GETAMP(NDB)
        NDB=NNA5+NDBSCA(5)
        A5P=A3COR*GETAMP(NDB)
        NDB=NNA6+NDBSCA(6)
        A6P=A3COR*GETAMP(NDB)
C     SET AMPLITUDE OF PARALLEL NASAL FORMANT
        NDB=NNAN+NDBSCA(7)
        ANPP=GETAMP(NDB)
C     SET AMPLITUDE OF BYPASS PATH OF FRICATION TRACT
        NDB=NNAB+NDBSCA(8)
        ABP=GETAMP(NDB)
C     RESET DIFFERENCE EQUATION CONSTANTS FOR RESONATORS
230     CALL SETABC(NNF1,NNB1,A1,B1,C1)
        CALL SETABC(NNF2,NNB2,A2,B2,C2)
        CALL SETABC(NNF3,NNB3,A3,B3,C3)
        CALL SETABC(NNF4,NNB4,A4,B4,C4)
        CALL SETABC(NNF5,NNB5,A5,B5,C5)
        CALL SETABC(NNF6,NNB6,A6,B6,C6)

        CALL SETABC(NNFNP,NNBNP,ANP,BNP,CNP)
C     AND FOR NASAL ANTIRESONATOR
        MNFNZ=-NNFNZ
        IF (MNFNZ.GE.O) MNFNZ=-1
        CALL SETABC(MNFNZ,NNBNZ,ANZ,BNZ,CNZ)
C     AND FOR GLOTTAL RESONATORS AND ANTIRESONATOR
        NPULSN=1
        IF (NNFO.LE.O) GO TO 245
C     ISSUE NO PULSE IF NNAV AND NNAVS BOTH .LE.0
        IF ((NNAV.LE.O).AND.(NNAVS.LE.O)) GO TO 245
C     WAVEFORM MORE SINUSOIDAL AT HIGH FUNDAMENTAL FREQUENCY
        NXBGP=(NNBGP*100)/NNFO
        CALL SETABC(NNFGP,NXBGP,AGP,BGP,CGP)
        CALL SETABC(O,NNBGS,AGS,BGS,CGS)
        MNFGZ=-NNFGZ
        IF (MNFGZ.GE.O) MNFGZ=-1
        CALL SETABC(MNFGZ,NNBGZ,AGZ,BGZ,CGZ)
C     SET GAIN TO CONSTANT IN MID-FREQUENCY REGION FOR RGP
        AGP=.007
C     DO NOT LET FO DROP BELOW 40 HZ
        IF (NNFO.LT.40) NNFO=40
C     MAKE AMPLITUDE OF IMPULSE INCREASE WITH INCREASING FO
        IMPULS=IMPULS*NNFO
C     NUMBER OF SAMPLES BEFORE A NEW GLOTTAL PULSE MAY BE GENERATED
        NPULSN=NNSR/NNFO
245     CONTINUE
C     CONVERT INHERENTLY INTEGER PARAMS TO REAL COEFICIENTS
        C(47)=NPULSN
        RETURN
        END


C       COEWAV.FOR              D.H. KLATT              8/1/78
C
C       "COEF-TO-WAVE" TRANSFORMATION SUBROUTINE
C               (FOR A 16-BIT PDP-11 COMPUTER)
C
C       SIMULATION OF THE HARDWARE KLATT SYNTHESIZER
C       TAKE 50 COEFIECIENTS FROM COMMON ARRAY C, AND
```

```
C       SYNTHESIZE NEXT NNXWS SAMPLES OF THE OUTPUT WAVEFORM
C
        SUBROUTINE COEWAV(IWAVE,OUTMA)
C               IWAVE IS AN ARRAY IN WHICH WAVEFORM SAMPLES ARE PLACED
C                 LEFT-JUSTIFIED IN A 36-BIT WORD
C               OUTMA IS RETURN ARG INDICATING MAX ABSOL. VALUE OF WAVE
C                 IF CALLING PROGRAM SETS TO -1., COEWAV IS INITIALIZED
C
        REAL NOISE,INPUTS,INPUT,IMPULS
        DIMENSION IWAVE(1),C(50)
        COMMON /COEFS/ C
C     COEFICIENT VALUES IN C(50) ARE REAL
        EQUIVALENCE (C(1),IMPULS),(C(2),SINAMP),(C(3),AFRICI)
        EQUIVALENCE (C(4),AASPI),(C(5),A1PAR),(C(6),A2PAR)
        EQUIVALENCE (C(7),A3PAR),(C(8),A4PAR),(C(9),A5PAR)
        EQUIVALENCE (C(10),A6PAR),(C(11),ABPAR),(C(12),ANPAR)
        EQUIVALENCE (C(13),AGP),(C(14),BGP),(C(15),CGP)
        EQUIVALENCE (C(16),AGZ),(C(17),BGZ),(C(18),CGZ)
        EQUIVALENCE (C(19),AGS),(C(20),BGS),(C(21),CGS)
        EQUIVALENCE (C(22),A1),(C(23),B1),(C(24),C1)
        EQUIVALENCE (C(25),A2),(C(26),B2),(C(27),C2)
        EQUIVALENCE (C(28),A3),(C(29),B3),(C(30),C3)
        EQUIVALENCE (C(31),A4),(C(32),B4),(C(33),C4)
        EQUIVALENCE (C(34),A5),(C(35),B5),(C(36),C5)
        EQUIVALENCE (C(37),A6),(C(38),B6),(C(39),C6)
        EQUIVALENCE (C(40),ANP),(C(41),BNP),(C(42),CNP)
        EQUIVALENCE (C(43),ANZ),(C(44),BNZ),(C(45),CNZ)
        EQUIVALENCE (C(46),PLSTEP)
C     MAXIMUM VALUE FOR A WAVEFORM SAMPLE (LEFT-JUSTIFY IN 36-BIT WORD)
        DATA WAVMA,WAVMAX/32767,-32767/
C
C     INITIALIZE COEWAV IF OUTMA=-1.
C     ZERO MEMORY REGISTERS IN ALL RESONATORS
        IF (OUTMA.GE.O.) GO TO 250
249     YL11P=0.
        YL12P=0.
        YL21P=0.
        YL22P=0.
        YL31P=0.
        YL32P=0.
        YL41P=0
        YL42P=0.
        YL51P=0.
        YL52P=0.
        YL61P=0.
        YL62P=0.
        YLNP1=0.
        YLNP2=0.
        YL11C=0.
        YL12C=0.
        YL21C=0.
        YL22C=0.
        YL31C=0.
        YL32C=0.
        YL41C=0.
        YL42C=0.
        YL51C=0.
        YL52C=0.
        YL61C=0.
        YL62C=0.
        YLNP1C=0.
        YLNP2C=0.
        YLNZ1C=0.
        YLNZ2C=0.
        YLGP1=0.
        YLGP2=0.
        YLGS1=0.
        YLGS2=0.
        YLGS3=0.
        YLGS4=0.
        YLGZ1=0.
        YLGZ2=0.
C     ZERO ALL OTHER MEMORY REGISTERS
        NPULSE=1
        MPULSE=0
        UGLOTX=0.
        UGLOTL=0.
        OUTMA=0.
        AFRIC=0.
        STEP=0.
        AASPIR=0.
C
C     GENERATE NNXWS SAMPLES OF OUTPUT WAVEFORM
250     CONTINUE
C     TRANSLATE SOME COEFICIENTS TO INTEGER
        NPULSN=C(47)
        NNXWS=C(48)
        NXSW=C(49)
        NNXFC=C(50)
        XNSAMI=1.0/FLOAT(NNXWS)
C     DELTA AMPLITUDE OF ASPIRATION
        DAHH=(AASPI-AASPIR)*XNSAMI
C     DELTA AMPLITUDE OF FRICATION
        DAFF=(AFRICI-AFRIC)*XNSAMI
C
C   MAIN LOOP
        DO 530 NTIME=1,NNXWS
C     GENERATE NEW GLOTTAL PULSE IF PERIOD COUNTER EXCEEDED
        NPULSE=NPULSE-1


        IF (NPULSE.GT.0) GO TO 260
C     AND IF NPULSN.GT.1 (I.E. IF FO>0 AND AV+AVS>0)
        IF (NPULSN.LE.1) GO TO 260
C     RESET PULSE COUNTER
        NPULSE=NPULSN
C     PULSE COUNTER FOR MODULATED NOISE
        MPULSE=NPULSE/2
C     SET AMPLITUDE OF NORMAL VOICING IMPULSE
        INPUT=IMPULS
C     AMPLITUDE OF QUASI-SINUSIODAL VOICING
        INPUTS=SINAMP
        GO TO 275
C     SET INPUT TO ZERO BETWEEN GLOTTAL IMPULSES
260     INPUT=0.
        INPUTS=0.
C     RESONATOR RGP:
275     YGP=AGP*INPUT + BGP*YLGP1 + CGP*YLGP2
        YLGP2=YLGP1
        YLGP1=YGP
C     GLOTTAL ZERO PAIR RGZ:
290     YGZ=AGZ*YGP + BGZ*YLGZ1 + CGZ*YLGZ2
        YLGZ2=YLGZ1
        YLGZ1=YGP
C     QUASI-SINUSOIDAL VOICING PRODUCED BY IMPULSE INTO RGP AND RGS:
        YGS=INPUTS*AGS + BGS*YLGS1 + CGS*YLGS2
        YLGS2=YLGS1
        YLGS1=YGS
        YGS=AGP*YGS + BGP*YLGS3 + CGP*YLGS4
        YLGS4=YLGS3
        YLGS3=YGS
C     GLOTTAL VOLUME VELOCITY IS THE SUM OF NORMAL AND
C     QUASI-SINUSOIDAL VOICING
        UGLOT2=YGZ + YGS
```

```
C     RADIATION CHARACTERISTIC IS A ZERO AT THE ORIGIN
        UGLOT=UGLOT2-UGLOTX
        UGLOTX=UGLOT2
C
C   TURBULENCE NOISE OF ASPIRATION AND FRICATION
C     GENERATE RANDOM NOISE, RANDOM PRODUCES UNIFORM DIST. (0. TO 1.)
370     NOISE=0.
C     MAKE PSEUDO-GAUSSIAN
        DO 371 NRANDX=1,16
371     NOISE=NOISE+RAN(IRAN1,IRAN2)
C     SUBTRACT OFF DC
        NOISE=NOISE-8.
C     MODULATE NOISE DURING SECOND HALF OF A GLOTTAL PERIOD
375     IF (MPULSE.LE.0) NOISE=NOISE/2.
        MPULSE=MPULSE-1
C     LOW-PASS NOISE AT -6 DB/OCTAVE TO SIMULATE SOURCE IMPEDANCE
C     HIGH-PASS NOISE AT +6 DB/OCTAVE FOR RADIATION CHARACTERISTIC
C          (TWO EFFECTS CANCEL ONE ANOTHER)
C     GLOTTAL SOURCE VOLUME VELOCITY = VOICING+ASPIRATION
        AASPIR=AASPIR+DAHH

        UASP=AASPIR*NOISE
380     UGLOT=UGLOT+UASP
C     SET FRICATION SOURCE VOLUME VELOCITY
390     AFRIC=AFRIC+DAFF
C     PREPARE TO ADD IN A STEP EXCITATION OF VOCAL TRACT
C     IF PLOSIVE RELEASE (I.E. IF PLSTEP.GT.0.)
        IF (PLSTEP.LE.0.) GO TO 391
        STEP=-PLSTEP
        PLSTEP=0.
391     UFRIC=AFRIC*NOISE
C
C   SEND GLOTTAL SOURCE THRU CASCADE VOCAL TRACT RESONATORS
C     DO FORMANT EQUATIONS FOR NNXFC FORMANTS IN DESCENDING ORDER
C     TO MINIMIZE TRANSCIENTS
        IF (NXSW.EQ.1) GO TO 430
C     BYPASS R6 IF NNXFC LESS THAN 6
        Y6C=UGLOT
        IF (NNXFC.LT.6) GO TO 415
        Y6C=A6*UGLOT + B6*YL61C + C6*YL62C
        YL62C=YL61C
        YL61C=Y6C
C     BYPASS R5 IF NNXFC LESS THAN 5
415     Y5C=Y6C
        IF (NNXFC.LT.5) GO TO 416
        Y5C=A5*Y6C + B5*YL51C + C5*YL52C
        YL52C=YL51C
        YL51C=Y5C
416     Y4C=A4*Y5C + B4*YL41C + C4*YL42C
        YL42C=YL41C
        YL41C=Y4C
        Y3C=A3*Y4C + B3*YL31C + C3*YL32C
        YL32C=YL31C
        YL31C=Y3C
        Y2C=A2*Y3C + B2*YL21C + C2*YL22C
        YL22C=YL21C
        YL21C=Y2C
        Y1C=A1*Y2C + B1*YL11C + C1*YL12C
        YL12C=YL11C
        YL11C=Y1C
C     NASAL ZERO-PAIR RNZ:
420     YZC=ANZ*Y1C + BNZ*YLNZ1C + CNZ*YLNZ2C
        YLNZ2C=YLNZ1C
        YLNZ1C=Y1C
C     NASAL RESONATOR RNP:
        YPC=ANP*YZC + BNP*YLNP1C + CNP*YLNP2C
        YLNP2C=YLNP1C
        YLNP1C=YPC
        ULIPSV=YPC
C     ZERO OUT VOICING INPUT TO PARALLEL BRANCH
C     IF CASCADE BRANCH HAS BEEN USED
425     UGLOT=0.
        UGLOTL=0.
C
C   SEND VOICING AND FRICATION NOISE THRU PARALLEL RESONATORS
C     INCREMENT RESONATOR AMPLITUDES GRADUALLY
430     CONTINUE
C     FIRST PARALLEL FORMANT R1' (EXCITED BY VOICING ONLY)
        Y1P=A1*A1PAR*UGLOT + B1*YL11P + C1*YL12P
        YL12P=YL11P
        YL11P=Y1P
C     NASAL POLE RN' (EXCITED BY FIRST DIFF. OF VOICING SOURCE)
        UGLOT1=UGLOT-UGLOTL
        UGLOTL=UGLOT
        YN=ANP*ANPAR*UGLOT1 + BNP*YLNP1 + CNP*YLNP2
        YLNP2=YLNP1
        YLNP1=YN
C     EXCITE FORMANTS R2'-R4' WITH FRIC NOISE PLUS FIRST-DIFF. VOICING
```

```
        Y2P=A2*A2PAR*(UFRIC+UGLOT1) + B2*YL21P +C2*YL22P
        YL22P=YL21P
        YL21P=Y2P
        Y3P=A3*A3PAR*(UFRIC+UGLOT1) + B3*YL31P +C3*YL32P
        YL32P=YL31P
        YL31P=Y3P
        Y4P=A4*A4PAR*(UFRIC+UGLOT1) + B4*YL41P +C4*YL42P
        YL42P=YL41P
        YL41P=Y4P
C     EXCITE FORMANT RESONATORS R5'-R6' WITH FRIC NOISE
        Y5P=A5*A5PAR*UFRIC + B5*YL51P +C5*YL52P
        YL52P=YL51P
        YL51P=Y5P
        Y6P=A6*A6PAR*UFRIC + B6*YL61P +C6*YL62P
        YL62P=YL61P
        YL61P=Y6P
C     ADD UP OUTPUTS FROM RN', R1' - R6' AND BYPASS PATH
        ULIPSF=Y1P-Y2P+Y3P-Y4P+Y5P-Y6P+YN-ABPAR*UFRIC
440     CONTINUE
C     ADD CASCADE AND PARALLEL VOCAL TRACT OUTPUTS
C     (SCALE BY 170 TO LEFT JUSTIFY IN 16-BIT WORD)
450     ULIPS=(ULIPSV+ULIPSF+STEP)*(170.)
        STEP=.995*STEP
C     FIND CUMULATIVE ABSOL. MAX. OF WAVEFORM SINCE BEGINNING OF UTT.
500     IF (ULIPS.GT.OUTMA) OUTMA=ULIPS
        IF (-ULIPS.GT.OUTMA) OUTMA=-ULIPS
C     TRUNCATE WAVEFORM SAMPLES TO ABS[WAVMA]
        IF (ULIPS.LE.WAVMA) GO TO 510
        ULIPS=WAVMA
510     IF (ULIPS.GE.WAVMAX) GO TO 520
        ULIPS=WAVMAX
520     IWAVE(NTIME)=ULIPS
530     CONTINUE
540     RETURN
        END
C       SETABC.FOR             D.H. KLATT               8/1/78
C
C       CONVERT FORMANT FREQENCIES AND BANDWIDTH TO RESONATOR
C       DIFFERENCE EQUATION CONSTANTS
C
        SUBROUTINE SETABC(F,FB,A,B,C)
C
        INTEGER F,FB
        COMMON /PIXX/ PIT,TWOPIT
C
C---REPLACE BY R=EXPTAB(FB) FOR FASTER EXECUTION
        R=EXP(-PIT*FLOAT(FB))
        C=-R*R
C---REPLACE BY B=COSTAB(F) FOR FASTER EXECUTION
        B=2.*R*COS(TWOPIT*FLOAT(F))
        A=1.-B-C
620     CONTINUE
C     IF F IS MINUS, COMPUTE A,B,C FOR A ZERO PAIR
        IF (F.GE.0) RETURN
630     A=1./A
        B=-A*B
        C=-A*C
        RETURN
        END
C       GETAMP.FOR             D.H. KLATT               8/1/78
C
C       CONVERT DB ATTEN. (FROM 96 TO -72) TO A LINEAR SCALE FACTOR.
C               (TRUNCATE NDB IF OUTSIDE RANGE)
C
        FUNCTION GETAMP(NDB)
C
        DIMENSION DTABLE(11),STABLE(28)
        DATA DTABLE/1.8,1.6,1.43,1.26,1.12
        1,1.0,0.89,0.792,0.702,0.623,0.555/
        DATA STABLE/65536.,32768.,16384.,8192.
        1,4096.,2048.,1024.,512.,256.,128.
        1,64.,32.,16.,8.,4.,2.
        1,1.,.5,.25,.125,.0625,.0312,.0156,.0078,.0039,.00195
        1,.000975,.000487/
C
650     NDB1=NDB
        GETAMP=0.
        IF (NDB1.LE.-72) RETURN
        IF (NDB1.GT.96) NDB1=96
        NDB2=NDB1/6
        NDB3=NDB1-(6*NDB2)
        XX1=STABLE(17-NDB2)
        XX2=DTABLE(6-NDB3)
        GETAMP=XX1*XX2
660     CONTINUE
        RETURN
        END
```

Allen, J. (1977). "A Modular Audio Response System for Computer Output," 1977 IEEE Int. Conf. on Acoust. Speech and Signal Proc., Hartford, IEEE Catalog No. 77CH1197-3 ASSP, 579-581.

Caldwell, J. (1979). "Flexible, High-Performance Speech Synthesizer Using Custon NMOS Circuitry," J. Acoust. Soc. Am. Suppl. 1 **64**, S72(A).

Carlson, R., Granstrom, B., and Klatt, D. H. (1979). "Vowel Perception: The Relative Perceptual Salience of Selected Acoustic Manipulations," in *Speech Transmission Laboratories QPSR* 2-3/1979 (Royal Institute of Technology, Stockholm, Sweden).

Cooper, F. S., Liberman, A. M. and Borst, J. M. (1951). "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," Proc. Natl. Acad. Sci. (U.S.) **37**, 318–325.

Dudley, H., Riesz, R. R., and Watkins, S. A. (1939). "A Synthetic Speaker," J. Franklin Inst. **227**, 739–764.

Dunn, H. K, and White, S. D. (1940). "Statistical Measurements on Conversational Speech," J. Acoust. Soc. Am. **11**, 278–288.

Epstein, R. (1965). "A Transistorized Formant-Type Synthesizer," Status Report on Speech Research SR-1, part 7, Haskins Labs.

Fant, C. G. M. (1956). "On the Predictability of Formant Levels and Spectrum Envelopes from Formant Frequencies," in *For Roman Jakobson* (Mouton, The Hague), pp. 109–120.

Fant, C. G. M. (1959). "Acoustic Analysis and Synthesis of Speech with Applications to Swedish," Ericsson Technics **1**, 1–106.

Fant, C. G. M. (1960). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fant, C. G. M. and Martony, J. (1962). "The Instrumentation for Parametric Synthesis (OVE II)" Speech Trans. Labs. QPSR 18–24, Royal Inst. of Tech., Stockholm.

Flanagan, J. L. (1957). "Note on the design of Terminal Analog Speech Synthesizers," J. Acoust. Soc. Am. **29**, 306–310.

Flanagan, J. L. (1958). "Some Properties of the Glottal Sound Source," J. Speech Hear. Res. **1**, 99–116.

Flanagan, J. L., Coker, C. H., and Bird, C. M. (1962). "Computer Simulation of a Formant Vocoder Synthesizer," J. Acoust. Soc. Am. **35**, 2003(A).

Flanagan, J. L., Ishizaka, K. and Shipley, K. L. (1975). "Synthesis of Speech from a Dynamic Model of the Vocal Cords and Vocal Tract," Bell System Tech. J. **54**, 485–506.

French, N. R., and Steinberg, J. C. (1947). "Factors Governing the Intelligibility of Speech Sounds," J. Acoust. Soc. Am. **19**, 90–119.

Fujimura, O., and Lindqvist, J. (1971). "Sweep-Tone Measurements of Vocal Tract Characteristics," J. Acoust. Soc. Am. **49**, 541–558.

Fujimura, O. (1961). "Analysis of Nasilized Vowels", M. I. T. Res. Lab. Electron. QPR **62**, 191–192.

Fujimura, O., (1962). "Analysis of Nasal Consonants," J. Acoust. Soc. Am. **34**, 1865–1875.

Gauffin, J., and Sundberg, J. (1974). "An Attempt to Predict the Masking Effect of Vowel Spectra," Speech Trans. Labs. QPSR 4/74, Royal Inst. of Tech., 57–62.

Gold, B., and Rabiner, L. R., "Analysis of Digital and Analog Formant Synthesizers," IEEE Trans. Audio Electroacoust. **AU-16**, 81–94.

Holmes, J. N. (1961). "Research on Speech Synthesis Carried out during a Visit to the Royal Institute of Technology, Stockholm," Research Report JU11-4, Joint Speech Research Unit, Eastcote, England.

Holmes, J. N. (1973). "The Influence of the Glottal Waveform on the Naturalness of Speech From a Parallel Formant Synthesizer," IEEE Trans. Audio Electroacoust. 298–305.

Holmes, J. N., Mattingly, I., and Shearme, J. (1964). "Speech Synthesis by Rule," Language Speech **7**, 127–143.

Kaiser, J. F. (1966). "Digital Filters," Chap. 7 in *System Analysis by Digital Computer*, edited by F. F. Kuo and J. F. Kaiser (Wiley, New York).

Klatt, D. H. (1972). "Acoustic Theory of Terminal Analog Speech Synthesis," Proc. 1972 Int. Conf. on Speech Communication and Processing, IEEE Catalog No. 72 CH0567-7 AE, 131–135.

Klatt, D. H. (1976a). "Structure of a Phonological Rule Component for a Synthesis-by-Rule Program," IEEE Trans. Acoust. Speech Signal Processes **ASSP-24**, 391–398.

Klatt, D. H. (1976b). "A Digital Filter Bank for Spectral Matching," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, IEEE Catalog No. 76 CH1067-8 ASSP, 537–540.

Klatt, D. H. (in preparation). *Analysis and Synthesis of Consonant-Vowel Syllables in English*.

Lawrence, W. (1953). "The Synthesis of Speech from Signals which have a Low Information Rate," in *Communication Theory*, edited by W. Jackson (Butterworths, London), pp. 460–469.

Liljencrants, J. (1968). "The OVE-III Speech Synthesizer," IEEE Trans. Audio Electroacoust. **AU-16**, 137–140.

Makhoul, J. (1975). "Spectral Linear Prediction: Properties and Applications," IEEE Trans. Acoust. Speech Signal Processes **ASSP-23**, 283–296.

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

Rabiner, L. R., Jackson, L. B., Schafer, R. W., and Coker, C. H. (1971). "A Hardware Realization of a Digital Formant Speech Synthesizer," IEEE Trans. Comm. Tech. **COM-19**, 1016–1070.

Scott, R. J., Glace, D. M., and Mattingly, I. G. (1966). "A Computer-Controlled On-Line Speech Synthesizer System," 1966 IEEE Int. Commun. Conf., Digest of Tech. Papers, Philadelphia, 104–105.

Stevens, K. N. (1971). "Airflow and Turbulence Noise for Fricative and Stop Consonants: Static Considerations," J. Acoust. Soc. Am. **50**, 1180–1192.

Stevens, K. N. (1972). "The Quantal Nature of Speech: Evidence form Articulatory-Acoustic Data," in *Human Communication: A Unified View*, edited by E. E. David and P. B. Denes (McGraw-Hill, New York).

Stevens, K. N., and Klatt, D. H. (1972). "Current Models of Sound Sources for Speech," in *Ventilatory and Phonatory Control Systems: An International Symposium*, edited by B. D. Wyke (Oxford U.P., London), pp. 279–291.

Stevens, K. N., Bastide, R. P., and Smith, C. P. (1955). "Electrical Synthesizer of Continuous Speech," J. Acoust. Soc. Am. **27**, 207(A).

Tomlinson, R. S. (1974). "SPASS-An improved Terminal Analog Speech Synthesizer," M.I.T. Research Lab. of Electronics QPR 80, 198–205.