COMPUTER TECHNOLOGY

A computer assisted system for the extraction and visual display of pitch

A. B. ROSTRON and C. P. WELBOURN
Hull University, Hull, HU6 7RX, England

The fundamental frequency of the vocal output is obtained from an accelerometer output by an adaptive filtering technique. This is converted to a voltage analogue of frequency via software on a standard PDP-8/E computer. Pitch values are either output directly or stored for subsequent display. Software is available for permanent storage and access of recorded data from backing store.

A wide variety of methods have been explored for examining the manner and rate of vibration of the vocal cords. Previous authors have tended to regard their systems as capable of producing good pitch analyses of both speech and singing, but in some respects the requirements for recording these two types of vocalizations are contradictory. In speech the pitch extractor must respond rapidly to sudden bursts of voicing and must be able to follow large and rapid frequency transitions. On the other hand, the accuracy with which the pitch measures are plotted is not crucial, since little if any speech information in English is conveyed by frequency as such. In contrast, with singing, voicing is continuous and the rate of frequency change is relatively slow, but the measured pitch values need to be accurate, since only small errors can cause large perceptual effects.

Of the methods available for examining vocal fold movement directly, the obvious approach is to record the sound pressure signal from a microphone. The disadvantage of this method is that anechoic conditions are needed to guarantee that the actual output is recorded, rather than a signal made up of the acoustic output plus reflections from surrounding objects. Also, the recording is sensitive to environmental noise.

Alternatively, the vocal folds can be illuminated by conventional optical techniques or by fiber optics through the nostril and nasopharynx (Coker, Sondi, Ishizaka, & Flanagan, 1974). The main disadvantage of such techniques is that they make normal phonation very difficult, and some subjects find their use objectionable. Furthermore, tedious setting-up procedures are required for each subject.

Fourcin and Abberton (1971) measured impedance from two electrodes strapped to the neck; Stevens, Kalikow, and Willemain (Note 1) used a small accelerometer taped to the neck. Both techniques provided simple methods for examining vocal fold vibration uninfluenced by environmental noise. (The latter technique was adopted here because of its smaller size.)

Once a low harmonic glottal waveform is obtained, it must be processed so that the fundamental frequency can be extracted. A number of complex frequency

analysis software algorithms, such as those of Noll (1967) and Lukatela (Note 2), can be applied to achieve this end. One advantage of such algorithms is that they can be readily applied to a complex signal. The general procedure is to digitize the signal and employ time series analysis over a temporal window which is moved along the signal. Some procedures employ an adaptive component, e.g., Lukatela (Note 2). The main problem with digital filtering techniques is that they are too slow for real-time use, and versions which simplify the approximation procedures to speed calculation are often inaccurate.

The alternative approach is to filter the vocal signal with an analogue filter and use off-the-shelf modules for the frequency to voltage conversion. This is not a particularly satisfactory solution, since there is a basic incompatibility between low output ripple (good sensitivity) and a fast settling time (Read, 1975). The present software solution involving a PDP-8/E computer and DK8-EP real-time clock attempts to overcome this last problem. The accelerometer output is amplified and passed through a variable bandpass filter (Cawkill Type FU4). The filtered signal is then fed to a Schmitt trigger. Every firing of the Schmitt generates an interrupt, and intervals between successive interrupts are timed by the clock. Thus, the period of each cycle of the input waveform is measured and assigned a digital value. This pitch value is sampled by the background software at the specified time interval and either output or stored. The machine code for this process is given in Table 1.

Software is available for displaying the recorded frequency/time pattern on a DEC VT8E video terminal. The frequency range of the display and the length of input sampled are under software control, and either logarithmic or linear scales can be used. Two modes of display exist, one in which the output is displayed while the input is being recorded, the second in which the display is updated only after a section of input has been recorded.

The program is organized with a control section which carries out operations given by the operator. Once

the command is completed, the control section is re-entered to await new commands. A list of commands is as follows:

- A Record a frequency sample from the external world.
- B Display contents of current data buffer on the video terminal.
- C Clear video terminal.
- D Output frequency sample to magnetic tape.
- E Input a frequency sample from magnetic tape to the data buffer.

The system provides an acceptable plot of speech intonation patterns (see Figure 1) when used with a filter of fixed bandwidth. Figure 1 presents a frequency contour very similar to that obtained by Fromkin and Ohala (1968) for the same utterance. As it stands, it is usable as a training aid in helping deaf people to acquire appropriate intonation patterns. Any lack of following the exact pitch contours, which is liable to occur in aperiodic bursts of voicing, may be of advantage since it is the overall general pitch which has to be matched rather than any small and sudden changes.

The system can be used also to provide a single some visual indication of intonation. Because of the high harmonic content of notes, a filter of fixed bandwidth restricts the notes that can be sung at any one filter setting to a narrow range. To help with this problem, a voltage controlled filter module (Fairchild semiconductors, 1972) is introduced into the input circuit, so that the center frequency of the passband can be con-

Table 1
Machine Code Software for the Pitch Extractor

/PERIOD VALUES BETWEEN SUCCESSIVE INTERRUPTS ARE
/MEASURED WITH A DAP-EP REAL-TIME CLOCK
/AND STORED IN A DATA BUFFER STARTING AT 2000.

```

/CLOCK INSTRUCTIONS
CLZE=6130
CLSK=6131
CLOE=6132
CLAB=6133
CLEN=6134
CLSA=6135
CLBA=6136
CLCA=6137
/GENERAL
GTF=6004
RTF=6005
CAF=6007

M0
0
JMP I 2
SERVE

ABUFF, 2000
//

MFREQ, CAF /CLEAR FLAGS
CLA
TAD ZEROC
CLZE /CLEAR CLOCK ENABLE

```

Table 1
Continued

	CLA		
	TAD ENABLE		
	CLZE	/SET CLOCK ENABLE	
	KSF	/START MEASURING FREQUENCY,	
	JMP -1	/WHEN TELETYPE IS STRUCK.	
	KCC		
	CLA		
	TAD MMAXL		
	CLAB	/START CLOCK COUNTING	
	ION		
SIT,	ISZ CNTA	/THIS SAMPLES PITCH ONCE	
	JMP -1	/EVERY TIME THROUGH THE	
	ISZ CNTB	/DOUBLE LOOP AND STORES IT	
	JMP -3	/IN THE DATA BUFFER, (2000 ON).	
	CLA		
	TAD SCNTB		
	DCA CNTB		
	TAD PITCH	/GET CURRENT PITCH VALUE	
	DCA I ABUFF		
	ISZ CNTC		
	JMP SIT		
	IDF	/INTERRUPT OFF	
	CLA		
	TAD SCNTC	/RESET COUNTERS	
	DCA CNTC		
	TAD SABUFF		
	DCA ABUFF		
	TAD ONE	/SETS FIRST TWO DATA POINTS	
	OCA I <2000	/TO SHOW RANGE IN DISPLAY	
	TAD MAXL		
	OCA I <2001		
	JMS CLEAN	/SETS ALL VALUES TO ADMISSABLE	
		/VALUES.	
	NOP		
	HLT		
SABUFF,	1777	/ADDRESS OF DATA BUFFER	
MAXL,	3770	/LONGEST FREQUENCY	
MMAXL,	4010	/MINUS LONGEST FREQUENCY	
CNTA,	0000	/COUNTERS DETERMINING HOW	
CNTB,	-1	/OFTEN PITCH VALUE IS	
SCNTB,	-1	/SAMPLED	
ONE,	0300	/SHORTEST PERIOD VALUE	
ZERDC,	7777	/CONSTANT TO CLEAR CLOCK ENABLE	
ENABLE,	6577	/VALUE OF CLOCK ENABLE	
PITCH,	0000	/PITCH VALUE STORED HERE	
CNTC,	-275	/COUNTERS DETERMINING NUMBER	
SCNTC,	-275	/OF PITCH SAMPLES TAKEN	
SERVE,	OCA AC	/INTERRUPT SERVICE ROUTINE	
	GTF		
	OCA FLAGS		
	CLSK	/SKP ON CLOCK FLAG	
	SKP		
	JMP RECORD		
	CAF		
	JMP EXIT		
RECORD,	CLCA	/CLOCK TO AC	
	DCA DUMMY	/STORE CLOCK VALUE	
	TAD MMAXL	/RESET CLOCK	
	CLAB	/START CLOCK COUNTING	
	CLA	/AC=MAXL	
	TAD DUMMY	/ADD MINUS LAST PERIOD	
	DCA PITCH	/UPDATE PITCH	
	CLSA	/CLEARS CLOCK STATUS FLAGS	
EXIT,	CLA		
	TAD FLAGS		
	RTF		
	CLA		
	TAD AC		
	JMP I 0	/END OF INTERRUPT SERVICE ROUTINE	
AC,	0000		
DUMMY,	0000	/TEMPORARY STORE FOR PITCH	
FLAGS,	0000		
CLEAN,	0000	/THIS SETS ANY VALUE IN THE	
START,	CLA CLL	/DATA BUFFER WHICH IS LESS	
	TAD MTHREEH	/THAN 300 TO 300.	
	TAD I ABUFF	/IS STORED DATUM	
	SMA	/LESS THAN 300?	
	JMP CONT	/IF NO, LEAVE IT ALONE	
	CLA	/IF YES, SET DATUM TO 300	
	TAD ONE		
	DCA I ABUFF		
CONT,	ISZ CNTC		
	JMP START		
	CLA		
	TAD SCNTC		
	DCA CNTC		
	TAD SABUFF		
	OCA ABUFF		
	JMP I CLEAN		
MTHREE,	-300	/MINUS HIGHEST FREQUENCY	
	5

Figure 1. Illustrates the falling fundamental frequency due to decrease in subglottal pressure as a subject says "Bev bombed Bob." The filter bandwidth was 100-300 Hz.

trolled by the computer's D/A converter. The control voltage, and thus the center frequency of the passband, is dependent on the frequency recorded. The main advantage of such a hybrid hardware/software pitch extraction system is that the adaptive algorithms can be tailored to fit any particular application. A display of a singer attempting to match a frequency of approximately 220 Hz is shown in Figure 2. (The note is articulated about halfway through the time period sampled.) The large and rapid initial frequency oscillations are typical of when an attempt is made to match a note. Experiments are in progress to examine whether such oscillations are the corrections of a control system dependent on feedback information.

In general, it is clear that success of the pitch extraction algorithm is dependent on the efficiency with which the various harmonics from the accelerometer output can be filtered. The filter bandwidth is the critical parameter determining the purity of the processed signal. To some extent, the selected passband must be a compromise between adequate filtering, which demands the smallest possible bandwidth, and an acceptable frequency range over which signals can be recorded, which requires the largest possible passband. (These opposing requirements are due to the fact that the frequency of the first harmonic of a person speaking at the bottom of his frequency range may be very similar to the fundamental frequency at the top of the same person's speaking range. The problem is more acute in singing, where the frequency of the fundamental can vary over a much wider range than in speech.)

For recording, the passband is set individually for each speaker, since the frequency range of vocal fold vibration depends on the speaker. A bandwidth of about 1.5 octaves gives good results. A typical value of

the passband for male speakers is from about 100 to 300 Hz; for female speakers the range is from 200 to 500 Hz. If further filtering is needed to provide a clean trace of the fundamental frequency, the adaptive filter can be switched in. This narrows the bandwidth to about 30 Hz over 100-800 Hz and provides an additional 26 dB of gain for the passed frequencies.

Two main factors affect the accuracy of the system: the frequency of the clock and the time elapsing between receipt of an interrupt and the clock value being recorded. A clock frequency of 1 MHz produces an error of 1 μ sec, which is considerably smaller than the time needed for the several instructions needed to have the clock read. In absolute terms, periods are underestimated by a constant error of about 11 μ sec, the interval between the clock being read and restarted. But the relative accuracy, of greatest importance for the present applications, is two program instructions (5.2 μ sec). This estimate is based on the assumption that the start and finish interrupts defining a particular period can each be in error by one machine instruction. To check the accuracy of these theoretical calculations, a fixed frequency of 125 Hz from a standard oscillator was input to the pitch extractor. A straight line was produced. The sensitivity of the display was then increased until some ripple was observed in the straight line. This ripple was measured as a time difference of about 10 μ sec, so that the observed errors correspond roughly with the theoretical estimates.

The software system described is available, together with user instructions, from the senior author (Department of Psychology, The University, Hull, HU6 7RX, England).

Figure 2. Shows the first second of a singer attempting to produce a fixed frequency on the vowel /a/. (The subject is not attempting to match a given frequency merely to produce any straight line.) The filter bandwidth was 120-300 Hz.

REFERENCE NOTES

1. Stevens, K. N., Kalikow, D. N., & Willemain, T. R. *The use of a miniature accelerometer for detecting glottal waveforms and nasality*, Bolt, Beranek & Newman Research Report No. 2907, 1974.

2. Lukatela, G. *Pitch determination by an adaptive autocorrelation method*. Haskins Laboratories Status Report on Speech Research, SR-33, 1973, 185-193.

REFERENCES

COKER, C. H., SONDHI, M., ISHIZAKA, K., & FLANAGAN, J. L. Measurement of articulatory dynamics. *Proceedings 8th International Congress of Acoustics*, 1974, 1, 223.

FAIRCHILD SEMICONDUCTORS. A voltage-tuned active filter. *Electronic Engineering*, 1972, 44 (528), 56.

FOURCIN, A. J., & ABBERTON, E. First applications of a new laryngograph. *Medical and Biological Illustration*, 1971, 21, 172-182.

FROMKIN, V., & OHALA, J. Laryngeal control and a model of speech production. *Working papers in phonetics*. Los Angeles: University of California Press, 1968. Pp. 98-110.

NOLL, M. A. Cepstrum Pitch Determination. *Journal of the Acoustical Society of America*, 1967, 41, 293-309.

READ, C. W. W. Analogue function modules. *Electron*, 1975, 78, 17-18.

(Received for publication October 28, 1975;
revision accepted March 11, 1976.)