TIME-DOMAIN TWO-DIMENSIONAL PITCH DETECTION

by

Gerard Benbassat

TECHNICAL REPORT NO. 267

December 30, 1975

PSYCHOLOGY AND EDUCATION SERIES

The work reported in this article was supported by National Science Foundation Grant NSF-EC-43997 to the Institute for Mathematical Studies in the Social Sciences, Stanford University.

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA 94305

Table of Contents

1 Introduction

A pitch detection algorithm faces two basic problems: reliability and computational efficiency. In the present case this algorithm was developed in the context of an audio response system using a large vocabulary. It was intended to be used in a pitch-synchronous encoding of the dictionary words, and for prosody experimentations. Reliability was a primary factor due to the hearability of the distortions introduced by most pitch detection errors and also to the impracticality of manual correction because of the large number of words involved (5,000 to 10,000).

The computational aspect has been considered for the convenience of a fast algorithm in experimentations on pitch and, as a result, the actual high-level language version of the program runs at about three times real time (on a PDP-10), and could, as will be shown, be optimized to run in real time.

The multiplicity of pitch detection algorithms (Markel [3], Noll [1], Miller [2], Maksim [4]), illustrates the difficulty in achieving the goals of speed and reliability. It appears that the reluctance of the speech wave to follow a simple pattern in all cases is the main source of occasional errors. A critical point is the difficulty of finding a single criterion that could separate voiced from unvoiced portions of speech in all situations. One solution could be to find a multidimensional space in which voiced and unvoiced speech are linearly separable, but this could lead to great computational

inefficiency. Another possibility is to add a continuity constraint in the voiced portions, but then occasional voicing irregularities introduce problems. The algorithm presented here uses both techniques: a continuity constraint on the pitch period in conjunction with a voiced/unvoiced separation in a two-dimensional space. In addition, various mechanisms are provided to account for known "misbehavior" of the speech wave.

2 A Two-dimensional Representation of the Speech Wave

The voiced portions of the speech wave are created by the excitation of the vocal tract by a series of pseudoperiodic high-energy pulses (pitch pulses) which, along with the resonant characteristics of the vocal tract, contribute to the creation of a high-energy peak immediately following each pitch pulse. A general damping due to the glottal excitation and the radiation of the mouth will minimize the energy of later peaks in each individual pitch period. On the other hand, in the case of unvoiced speech, the vocal tract is either excited by a white noise (fricative) or by a single burst (plosive), which results in many low-energy peaks or an isolated high-energy peak. Thus, the detection of a series of peaks of higher energy than the surrounding ones and at regular intervals is an indication of voicing, whereas the absence of such a pattern is an indication of unvoicing.

2.1 Peak Energy

The first dimension to represent speech with respect to voicing/unvoicing quality is peak energy (PKE). It is defined as the energy of a positive excursion cycle between two consecutive zero-crossings.

$$PKE(zi) = \sum_{t=zi}^{zi+1} S(t)*S(t) \qquad zi, zi+1 \text{ consecutive zero-crossings}$$

The position of each excursion cycle (or PEAK) is defined as the position of the first zero-crossing (see Figure 1).

Insert Figure 1 about here

2.2 Interperiodic Cross-correlation

Because of mechanical constraints the frequency response of the vocal tract changes slowly; when excited by a periodic train of pulses it produces a wave with a high correlation between successive pitch periods. On the other hand, successive segments of unvoiced speech, produced by random noise excitation, have a low correlation. If P1 and P2 are the respective positions of two consecutive pitch pulses, the interperiodic cross-correlation (XCORR) is defined as

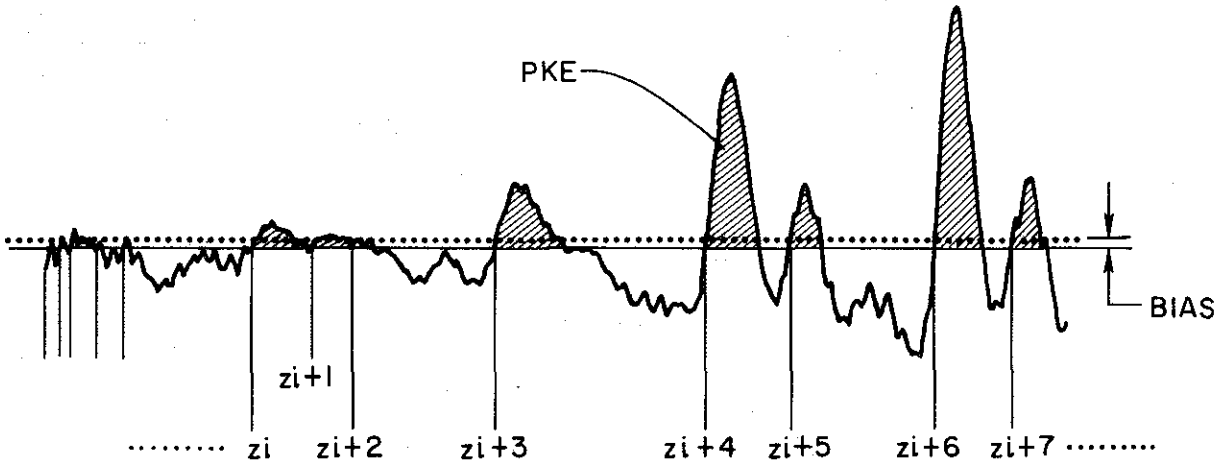


Figure 1. Peak energy

$$XCORR(p2) = \frac{\sum_{t=p1}^{p2} s(t)*s(t+T)}{\sqrt{\left(\sum_{t=p1}^{p2} s^2(t)\right) * \left(\sum_{t=p1}^{p2} s^2(t+T)\right)}}$$

with T = P2 - P1 - 1.

The nonstationary character of the speech wave introduces only a negligible error in the calculation of XCORR because of the slow variation of the vocal tract.

3 Description of the Algorithm

The pitch detection is performed in the time domain. Each segment of speech is assumed a priori to be voiced: it is attempted to extract a series of PEAKs of maximum energy spaced at regular intervals. If such PEAKs are found, XCORR is calculated for each of them and a decision is made on the position of these PEAKs in the plan (PKE,XCORR). Along with the following description, a flowchart of the algorithm is given in Appendix A.

3.1 Unbiasing and Peak Energy Calculation

Since the zero-crossing positions are important, the speech wave is first unbiased using the first 100 ms of sound to calculate the bias. Then PKE is calculated for all nonzero positive PEAKs (see Section 2.1).

3.2 Maximum PEAK Selection: "Search"

If PTR is the position of a pitch pulse and T is the value of the previous pitch period, the next pitch pulse, if it exists, should be found close to (PTR + T). In the interval (PTR + E1, PTR + SEARCHFIELD) where E1 = T/10 and SEARCHFIELD = k*T, the PEAK of larger energy (MPK) will correspond to a possible pitch pulse. If PTR is the position of an unvoicing marker, PERIOD is an a priori guess, E1 is made null, and the selected MPK is a candidate to be the first pitch pulse of a series (see Figure 2).

Insert Figure 2 about here

The value of k is chosen so that SEARCHFIELD is not larger than two times the smallest period that can satisfy the periodicity test: k = 2*(1 - v). If the investigated segment is unvoiced there is, in a truly random case, a 30 percent chance for the selected MPK to satisfy the periodicity test. Thus, this selection process makes the periodicity test alone about 70 percent efficient for the elimination of spurious PEAKs.

3.3 Selection of First PEAK: "Firstguess"

In case the last segment of speech was either unvoiced or unknown, the algorithm tries to find the beginning of a voiced portion without the use of any previous knowledge about the sound. This operation will be referred to as the "firstguess."

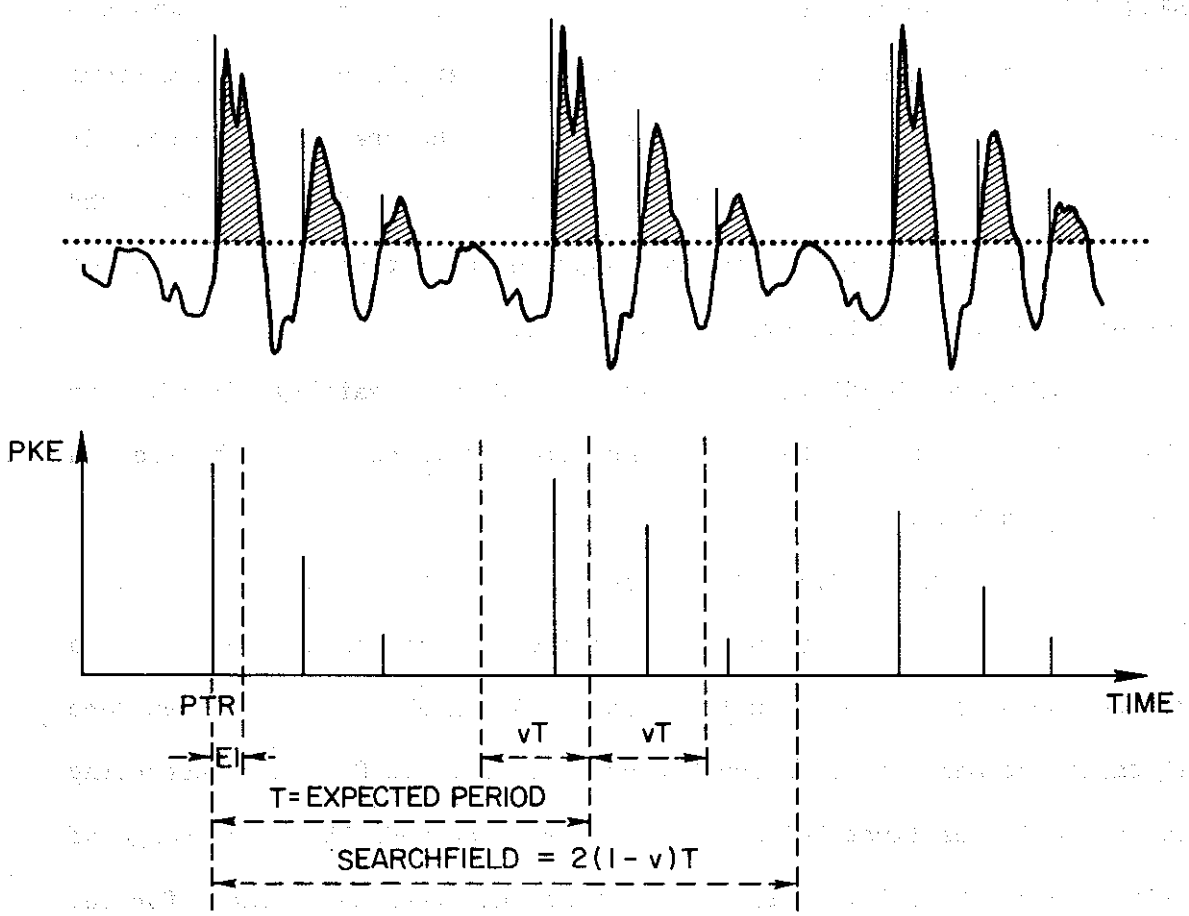


Figure 2. Searchfield

A first maximum PEAK (MPK) is selected with the "search" procedure (see Section 3.2) starting at the last valid unvoicing marker. The SEARCHFIELD is set to 1.5 times the smallest expected period (3 ms). Then two more MPKs are selected in similar SEARCHFIELDs, starting at the last selected MPK. These three MPKs are subjected to a periodicity test (PTEST) that allows for a maximum period variation of 25 percent over or under the previous period. If the three MPKs satisfy PTEST, more MPKs are selected with the "search" procedure in SEARCHFIELDs that are adjusted each time to be 1.5 times the distance between the two previous MPKs.

If up to MAXNB (actually set to 12) MPKs satisfy PTEST, then these MPKs are further tested in the voiced/unvoiced decision section (see Section 3.4).

If a nonperiodic MPK is found and if the number of periodic MPKs selected thus far is smaller than the assumed minimum length of a voiced segment (MLVS, actually set to 4 pitch periods) then more attempts are made to find another set of periodic MPKs by restarting the selection of first PEAKs with increased SEARCHFIELDs. The range of variation of SEARCHFIELD is set so that the frequency range for the first periods is 50-330 hz.

If the maximum value of SEARCHFIELD is reached without success, then the segment is declared unvoiced (see Section 3.4).

If more than MLVS but less than MAXNB periodic MPKEs are selected, and if the first nonperiodic MPK has an energy greater than a

preset voicing threshold (HIPKE), this may indicate an error in the period detection (half or double), so the following portion of speech is checked by restarting the "firstguess" after this nonperiodic MPK to search for at most MLVS MPKs. If there is a substantial difference between the period of the newly selected MPKs, if there are any, and the period of the old set of MPKs, then, to avoid a probable frequency error, the old set is rejected and the next 10 ms of sound are declared unvoiced. Otherwise, the old set of MPKs is restored and tested in the voiced/unvoiced decision section.

3.4 Voiced-Unvoiced Decision

The selected MPKs are now tested to decide whether they correspond to pitch pulses. For each MPK, XCORR is calculated and its position in the plan (PKE,XCORR) is tested with the linear functions:

(TEST1) $a1*MPKE + a2*XCORR - a3 > 0$

with $MPKE > 0$ and $-1 < XCORR < 1$ (see Figure 3).

Insert Figure 3 about here

If less than four MPKs satisfy TEST1 the segment of speech is declared unvoiced; otherwise it is accepted as voiced.

I. Unvoiced case: The next 10 ms of speech or up to the first of the selected MPKs, whichever is smaller, is declared unvoiced. Then the first peak selection is restarted from this point. All information about the MPKEs selected in this section is forgotten.

II. Voiced case: The selected MPKs are inserted in the pitch pulses list and the selection of subsequent PEAKs is started from the last inserted MPK.

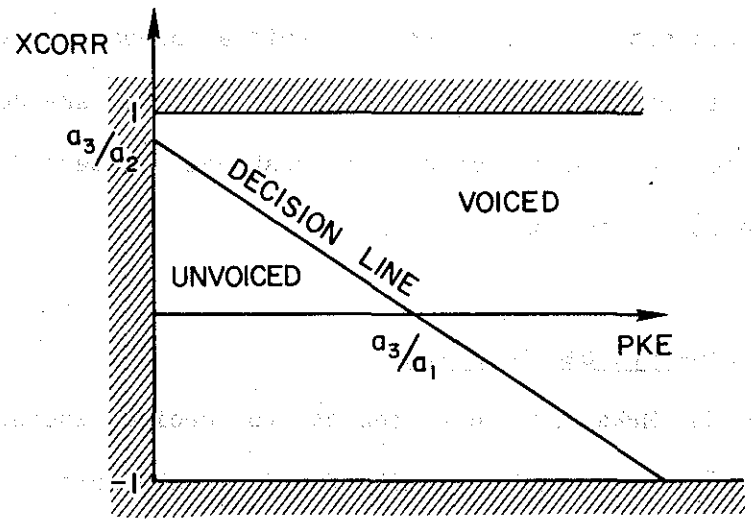


Figure 3. Decision line in the plane (PKE,XCORR)

3.5 Selection of Subsequent PEAKs

The beginning of a voiced segment of speech has been detected. It is now attempted to find the remaining pitch pulses and the end of the voiced portion of speech. One MPK at a time is selected and tested, and if the tests are positive, it is definitely accepted as a pitch pulse without waiting for more information.

The selection of an MPK is done with the "search" procedure in a SEARCHFIELD starting at the last pitch pulse and of a length equal to 1.5 times the last pitch period.

If this MPK does not satisfy the periodicity test (PTEST) but its value is large enough to indicate a probable voicing, then the second largest PEAK of the same SEARCHFIELD is selected. This takes into account the possibility of having an extra PEAK in the pitch period, due to a rapidly changing intensity.

If, again, the periodicity test is not satisfied, the selection of the subsequent PEAKs is abandoned for a "firstguess" attempt.

If the MPK selected is "periodic enough," a test similar to TEST1 of the "voiced/unvoiced decision" section is applied:

(TEST2) $b1*MPKE + b2*xcorr - b3 > 0$

with $MPKE > 0$ and $-1 < XCORR < +1$.

The coefficients b1, b2, and b3 are chosen so that TEST2 is less severe (to accept an MPK as being a pitch pulse) than TEST1. This is to account for some transition phenomena, for example, fast formant transition (low XCORR) at a low intensity level (low MPKE) (see Section

4). If TEST2 is satisfied, the selection of further PEAKs continues; otherwise, a "firstguess" is attempted starting at the last successfully selected MPK.

3.6 Unvoicing Decision Delaying

If the "firstguess" has failed to find a series of pitch pulses and the previous segment of speech is unvoiced, no further testing is applied and the current segment is declared unvoiced (see Section 3.4). But if the previous segment is voiced and the first MPK selected in the "firstguess" has an energy large enough to indicate a probable voicing, then the unvoicing decision is delayed.

The failure of the "firstguess" was probably due to a lack of of periodicity of some selected MPKs which may correspond to a voicing irregularity (see Section 4). To take such a possibility into account, a "firstguess" is again attempted but starting after the first previously selected MPK.

If the "firstguess" continues to fail, the unvoicing decision is delayed until the first selected MPK has an energy lower than the voicing threshold (HIPEAK). This allows the accepting of more than one irregular pitch pulse.

If "firstguess" is successful after such a delaying, the "hole" left between the last pitch pulse and the first selected MPK is filled with artificially inserted pulses using a linear interpolation of the period.

3.7 Smoothing of the Pitch Contour

The positions of the zero-crossings of the selected MPKs are not the exact positions of the pitch pulses, and the spacing between these two positions is essentially variable, thus introducing a noise in the pitch contour. It appeared that such a noise has a very unpleasant effect on the reproduced speech. It was possible to suppress that effect by applying a simple triangular smoothing on the originally obtained pitch contour:

$$T(i) = \left(\sum_{j=-4}^{4} T(i+j)*w(j)\right)/k$$

with $w(j) = 1 - abs(j)/5$ and $k = \sum_{j} w(j)$.

4 Hearability and Controllability of the Pitch Detection Errors

Many refinements have been introduced in the algorithm to minimize the risks of pitch detection errors and also to reduce the hearability of the errors that may be left. Control of the errors can be achieved by knowing the specific influence of the parameters on particular types of errors; it is then possible to find the best adjustments.

The method chosen as the most practical consists of making the pitch detection and then encoding (in LPC) a large set of words (500), and, by synthesizing the words from their coded form, isolating those

with "hearable" defects after a simple listening test. After this operation the behavior of the algorithm on the defective words can be traced and readjustments made. The advantage of this method is to focus the optimization on the errors that have the most perceptually distorting effect.

The errors can be classified into three categories: frequency errors (double or half the real frequency), voiced/unvoiced decision errors (on a whole segment of speech), and boundary errors (at the voiced/unvoiced or unvoiced/voiced transitions). Each category is handled by a specific section of the algorithm and the risk of occurrence can be controlled by setting the appropriate parameters.

4.1 Frequency Errors

Doubling or halving the fundamental frequency of a voiced sound, even for a short period of time, introduces a very noticeable distortion that may affect the understandability of the utterance and is, in any case, very undesirable.

If the correct frequency has been detected in the firstguess section (Section 3.3), frequency doubling or halving cannot happen in the subsequent peak detection (Section 3.5) because of PTEST. But in the firstguess section, where the frequency is guessed using no information about the past, frequency errors are possible in some situations.

Doubling of the fundamental can occur when, at the beginning of

a voiced segment, the first formant is only lightly damped and has a frequency approximately double the fundamental, thus creating a large PEAK in the middle of a pitch period, and a possible confusion of that extra PEAK for the beginning of a pitch period.

Possible halving of the fundamental generally occurs because of the presence of a high-energy PEAK preceding the beginning of the first pitch period at a distance approximately double the pitch period in conjunction with a rising intensity.

In these situations the firstguess section may detect the wrong frequency and it was not found possible to adjust TEST1, in the decision section (Section 3.4), to discriminate the extraneous MPKs without introducing unvoicing errors in other situations.

The reduction of the maximum period variation allowed in PTEST reduces the probability of such errors in the firstguess section, but not significantly.

To overcome this problem, one more hypothesis must be added to the model: the confusing situation lasts only for a limited number of periods and not for a whole voiced segment. That is, the intensity of the PEAKs will not be an increasing monotone function for more than n periods, in the case of a frequency-halving situation, or the ratio between the fundamental and the first formant frequencies will not stay stable for more than n periods, in the case of a frequency-doubling situation. If this is true, the n+1 MPK should be rejected by PTEST but accepted by TEST1. A firstguess of the segment of speech following

the n+1 MPK and a comparison between the pitch periods of these two segments can detect a frequency error (see Section 3.3). The maximum value for n (MAXNB) was determined experimentally.

Without this look-ahead mechanism, 5 percent of the words, on the test set of 500, had frequency errors (including only two cases of frequency doubling). With the look-ahead and MAXNB = 4, four words were left with frequency-halving errors; with MAXNB = 6, only one word had such an error and it was necessary to set MAXNB = 10 to eliminate it. Finally MAXNB was set to 12 to provide some more immunity.

In the case of the detection of a frequency error the simple decision of unvoicing for the beginning of the segment (see Section 3.3) may result in the devoicing of the first pitch period. Since this type of error does not introduce any significant distortion of the reproduced words, it was not felt necessary to implement a special procedure to correct it.

4.2 Voiced/Unvoiced Decision Errors

The distorting effect of a voiced/unvoiced decision error on an entire portion of speech mostly depends on the duration and the intensity level of that portion. It appeared, however, that the devoicing of a voiced segment had a less destructive effect on the intelligibility than the artificial voicing of an unvoiced segment.

The behavior of the algorithm with respect to this problem is central for the adjustment of most tests and parameters: PTEST, TEST1, TEST2, and MLVS.

When the parameters are set for a possible correct detection of the voiced portion of speech with the worst characteristics (according to the model: shortest duration, maximum period variation, and a combination of low PKE and low XCORR) it results in approximately 70 percent of the words with artificial voicing errors. Although these conditions are artificial, this test showed that some "exceptional misbehavior" should be dealt with separately and that the adjustment of the parameters can only result in the reduction of the frequency of errors.

MLVS adjustment. Although very short vowels do not occur frequently in words spoken in isolation, they tend to appear more often in continuous speech, especially at the onset; for example, in a sentence starting with "a few ...," the vowel "a" may be considerably reduced. The shortest vowel observed in the test set was four pitch periods long and we were unable to produce a shorter one, so MLVS was set to 4.

PTEST adjustment. Some voicing irregularities called "creaks" can create a sudden pitch variation of up to 100 percent. They are caused by an irregular vibration of the vocal cords. These cases must be dealt with by a special mechanism, and PTEST should only accept the maximum "normal" variation between two consecutive pitch periods. One way to find this maximum period variation is to set TEST1 and TEST2 so that they will not reject any true pitch pulse and then adjust PTEST

until no devoicing occurs except in creak situations. It was found that a 25 percent variation threshold could achieve this result on the test set.

The special mechanism provided to account for creaks consists of trying to skip the irregularity. If a high-energy irregular PEAK is detected, a firstguess is attempted starting from that PEAK. If it fails to detect more pitch pulses, the decision of unvoicing is delayed as long as the selected MPKs have an energy indicating a probable voicing (see Section 3.6) so that the irregularity can be bypassed and replaced by a continuity solution if voicing does indeed continue afterward. However, there are situations where this mechanism does not operate successfully: if the irregularity happens within the first or last four pitch periods of a voiced portion, the beginning or the end of that portion is devoiced. Only one such case was detected in the test set (see Figure 4) and it did not result in a very noticeable distortion. A more critical situation is when the irregularity occurs in the middle of a short vowel, that is, within the first and last pitch periods. The whole vowel is then devoiced. Such cases were observed on the ending vowels of six words among a set of 1,000 words (different from the test set).

Insert Figure 4 about here

TEST1 and TEST2 adjustments. Informal intelligibility tests using words containing pitch errors showed that artificial voicing is

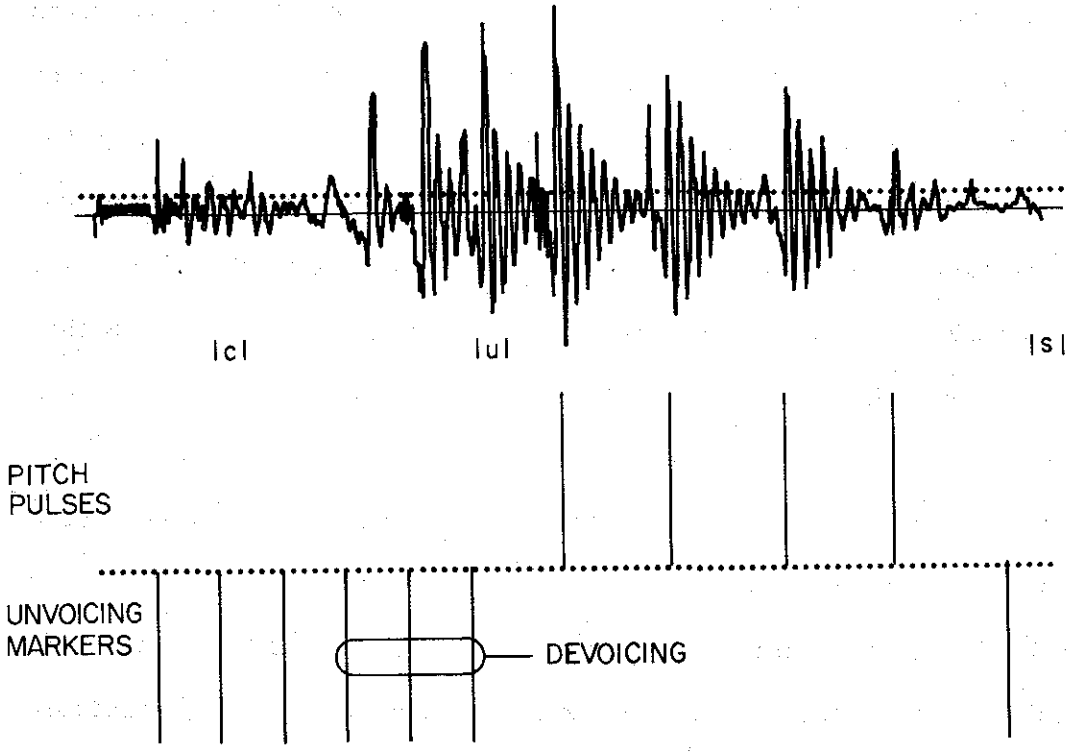


Figure 4. Pitch period irregularity at the beginning of a voiced portion

less desirable than devoicing, so TEST1 was adjusted to reject all sets of four or more unvoicing peaks that would satisfy PTEST, for the words of the test set. With TEST2 adjusted in the same way, about 10 percent of the words of the test set had devoicing errors, but since the selection of peaks is more restrictive in the selection of subsequent peaks (Section 3.5) than in the firstguess, the risk of selecting a spurious peak is also smaller and TEST2 can be made less severe than TEST1 to accept MPKs of low energy and/or low XCORR. However, it was not possible to adjust TEST2 to accept the sharpest formant transitions between consecutive pitch periods, for example, at the transition of a /b/ consonant and a vowel (see Figure 5). These cases were simply solved by trying a firstguess after a failure of TEST2 instead of immediately declaring the segment unvoiced. All noticeable voicing/unvoicing errors were then removed from the test set.

Insert Figure 5 about here

4.3 Boundary Errors

With all parameters adjusted to eliminate frequency and major voicing/unvoicing errors it was found, by graphic observation of the detected pitch pulses along with the speech wave, that in some instances voicing would start too late or stop too early, or, less frequently, start too early or stop too late. By readjusting TEST1 and TEST2, only the balance of error types can be modified. However, these

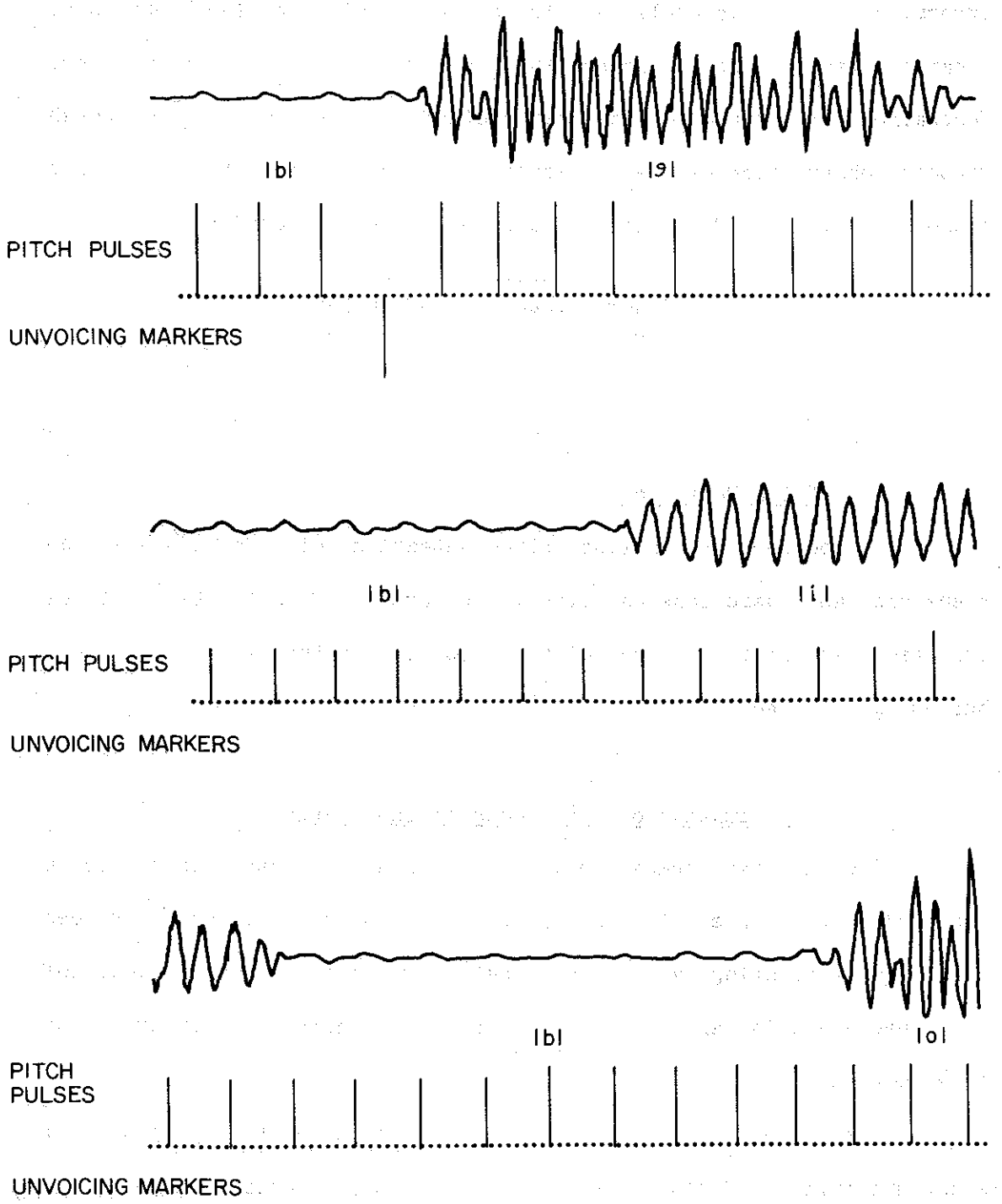


Figure 5. Sharp formant transitions

errors, happening generally at low energy levels and for very short durations (see Figure 6), do not introduce very noticeable distortions. Nontrained listeners did not hear defects in the test set although graphic observation of sample words from that same set showed that approximately 10 to 20 percent of the words contain boundary errors.

Insert Figure 6 about here

---

5 Pitch Detection On-line

To perform an on-line pitch detection of continuous speech there are two basic considerations: the number of instructions to be executed per unit of time of the sound, and the amount of sound buffering required.

5.1 Estimation of the Number of Instructions

The average number of "basic machine instructions" to be executed per sample of the speech wave being analyzed has been evaluated by counting over a large set of words the average number of times each task is executed and evaluating the number of instructions in each task.

The "basic machine instructions" selected for this purpose are: adds, multiplies, divides, program controls and data fetches. The possible optimization of data fetch instructions with respect to the

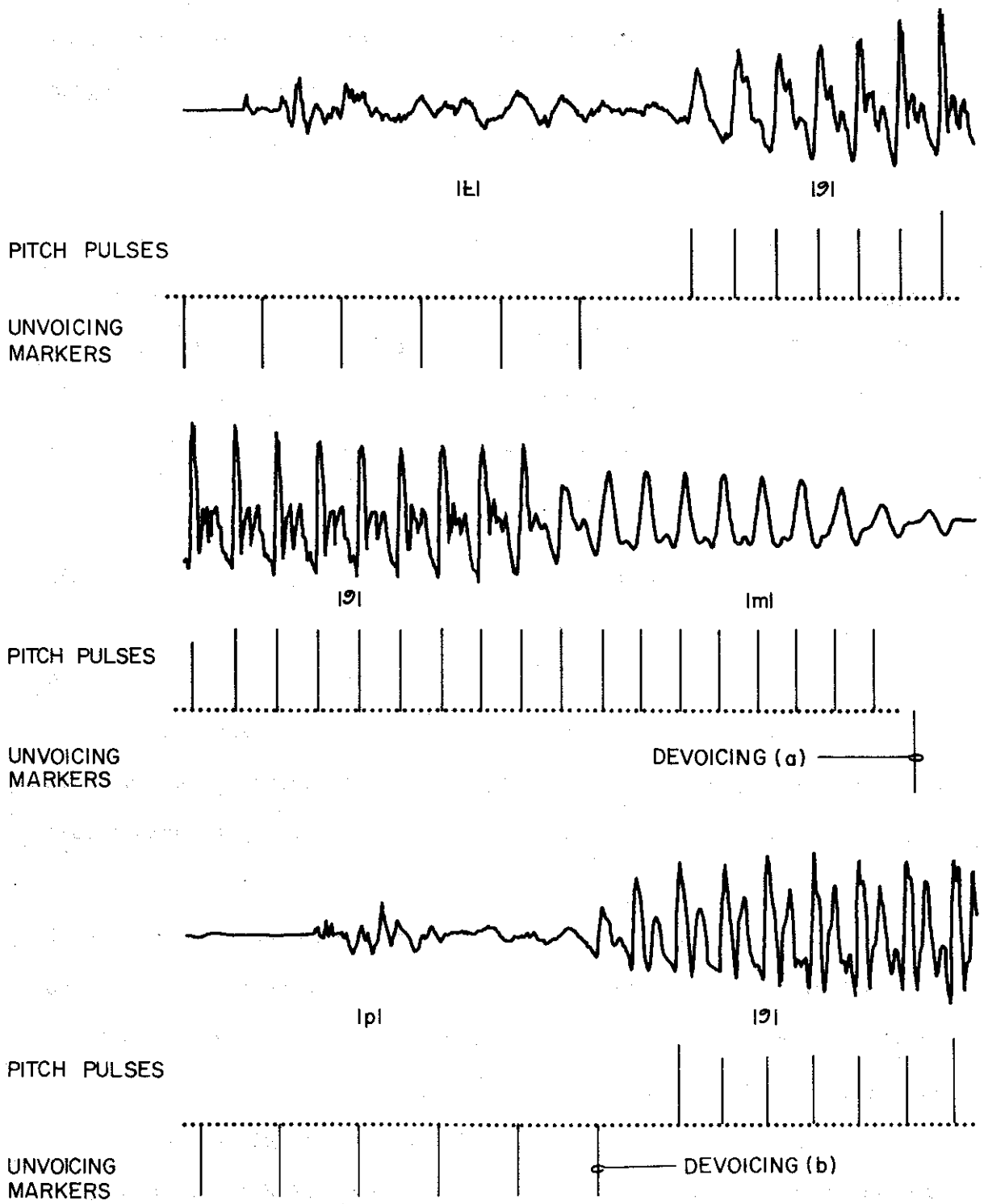


Figure 6. Boundary errors: Devoicing

other instructions is very much machine dependent. Assuming a computer having memory-register instructions and indexing, the worst case estimation used in the following count is of one data fetch per program control instruction and two data fetches for each of the other instructions.

The main tasks to be executed are the unbiasing and the peak energy calculations, the search for maximum peaks, the calculation of XCORR, PTEST, TEST1 and TEST2. A table showing the count of instructions per task is presented in Table 1.

Insert Table 1 about here

The unbiasing can be done in a loop that requires 1PC and 1A per sample.

The peak energy calculation can also be done in a loop where the sign of each sample is tested so it will require 2PC per sample. For the actual energy calculation, since the speech wave has been unbiased, on the average only .5A and .5M are executed per sample (see Section 2.1).

The search procedure used in the firstguess and the subsequent peak selection sections is used on the average once every 10 samples and there is an average of 10 peaks to be tested each time. By using a loop structure, the procedure requires 20PC per call.

XCORR is used in TEST1 and TEST2 and is calculated once every 80 samples on the average. If the average pitch period is of 60 samples

Table 1

Count of Basic Instructions

Tasks	Times/sample	Number of instructions/task[a]					Instructions/sample
		A	M	Dv	PC	D	
Unbiasing	1	1.0			1	2	1A + 1PC + 2D
Peak energy	1	.5	.5		2	6	.5A + .5M + 2PC + 3D
Search	1/10				20	20	2PC + 2D
PTEST	1/25	1.0		1	2	6	.04A + .04Dv + .08PC + .24D
XCORR	1/80	3 × 60	3 × 60	1		720	1.5A + 1.5M + .01Dv + 6D
TEST1 and TEST2	1/80	1.0	2.0		2	8	.01A + .02M + .02PC + .08D

Note. Total instructions/sample: 3A + 3M + .05Dv + 4PC + 14D.

[a]A = addition, M = multiplication, Dv = division, PC = program control, D = data fetch.

(assuming a 10 KHz sampling rate), then the calculation requires 60*3A, 60*3M and 1Dv (see Section 2.2).

As can be seen in Table 1, PTEST, TEST1 and TEST2 contribute only by a negligible amount to the total number of instructions.

There are a number of other operations performed each time a section of the algorithm is entered or exited, but each of them requires only a small number of instructions and each section is entered at most once every 100 samples, so that their contribution is at most an order of magnitude smaller than the total number of instructions.

The total number of instructions per sample amounts to:

3A + 3M + .05Dv + 4PC + 14D .

With a medium-speed computer (5 μs per Multiply and 1 μs per other instruction) it takes 36 μs per sample, which at a 10 KHz sampling rate would leave 64 μs per sample for data management and other overhead.

5.2 Sound Buffering

The "firstguess" operation requires the analysis of a substantial portion of the speech wave before a decision is made. In the present version of the program this portion is determined by the maximum number of pitch pulses that can be searched in that section: MAXNB + MLVS (actually 12 + 4). Such a method makes the amount of

buffering required essentially dependent on the pitch period, but it would be as efficient to fix a maximum duration of sound to be investigated instead of a maximum number of pitch pulses. About .15 second of sound would be sufficient to insure a good reliability. This would be a substantial delay for live transmission of encoded speech but would be suitable for on-line recording of encoded speech.

6 Conclusion

An algorithm for pitch detection was developed, using a time-domain model of the speech wave with respect to its voicing/unvoicing characteristics. Various mechanisms were developed to account for most of the cases where the speech wave does not match the model. An analysis of the correlation between the various parameters used in the algorithm and potential pitch detection errors was made in order to find an adjustment of the parameters that would satisfy the requirement of quality for the speech reproduced after linear predictive analysis and synthesis. After the algorithm had been adjusted with a test set of 500 words, it was used to encode a vocabulary of 3000 words. By listening to the synthesized version of that vocabulary, less than one percent of the words were found to have noticeable voicing/unvoicing errors, most of them due to creaks in the original recorded versions; none was found having frequency errors. It was also shown that the algorithm could be performed in real time for on-line applications.

References

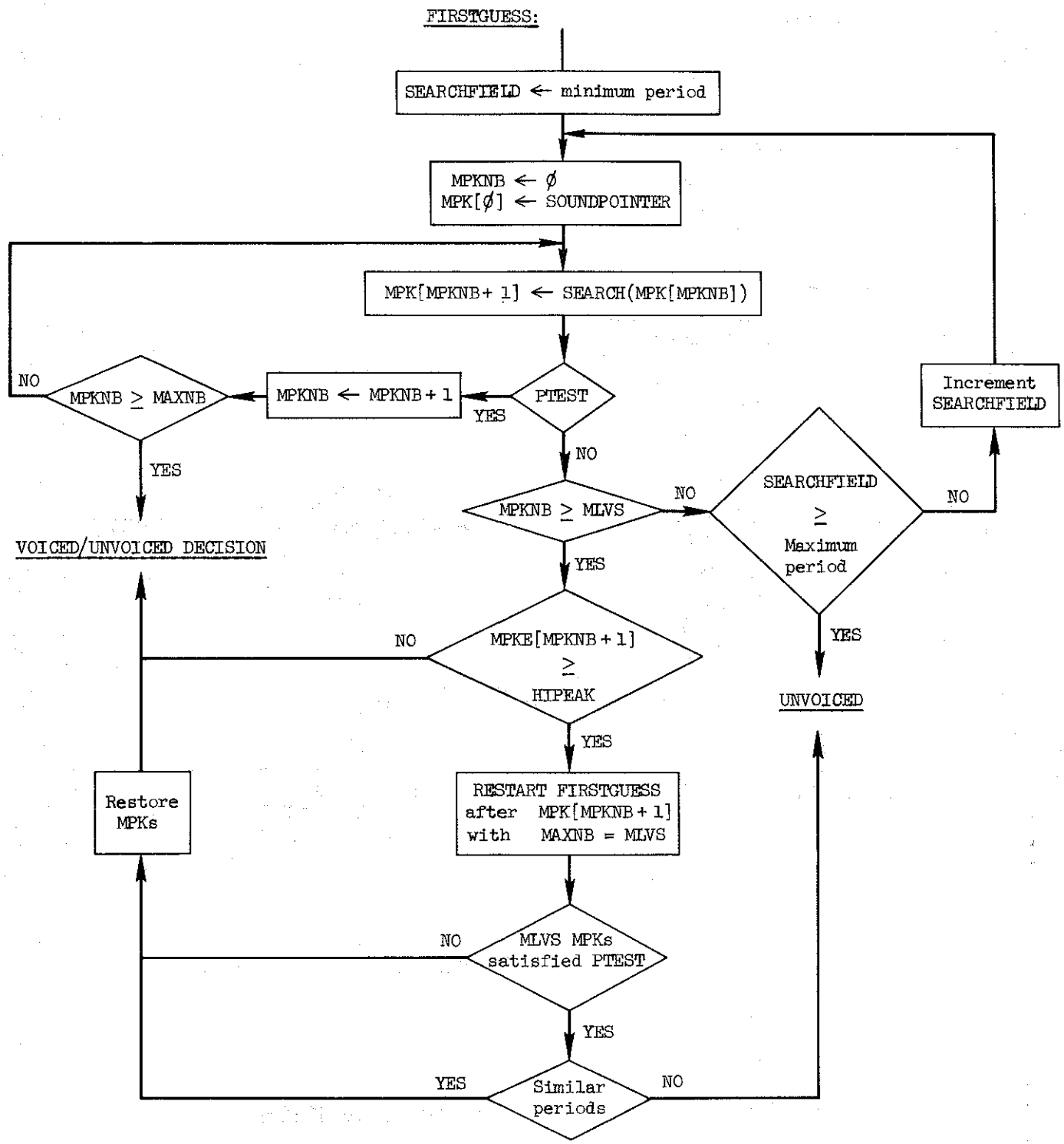
1. A. M. Noll. "Cepstrum pitch determination." J. Acoustical Society of America, Vol. 41, pp. 293-309, February 1967.

2. N. J. Miller. "Pitch detection by data reduction." IEEE Symposium on Speech Recognition, Carnegie Mellon University, Pub. IEEE, N.Y., pp. 122-130, April 1974.

3. J. D. Markel. "The sift algorithm for fundamental frequency estimation." IEEE Transactions on Audio and Electro-acoustics, Vol. AU 20, No. 5, pp. 367-377, December 1972.

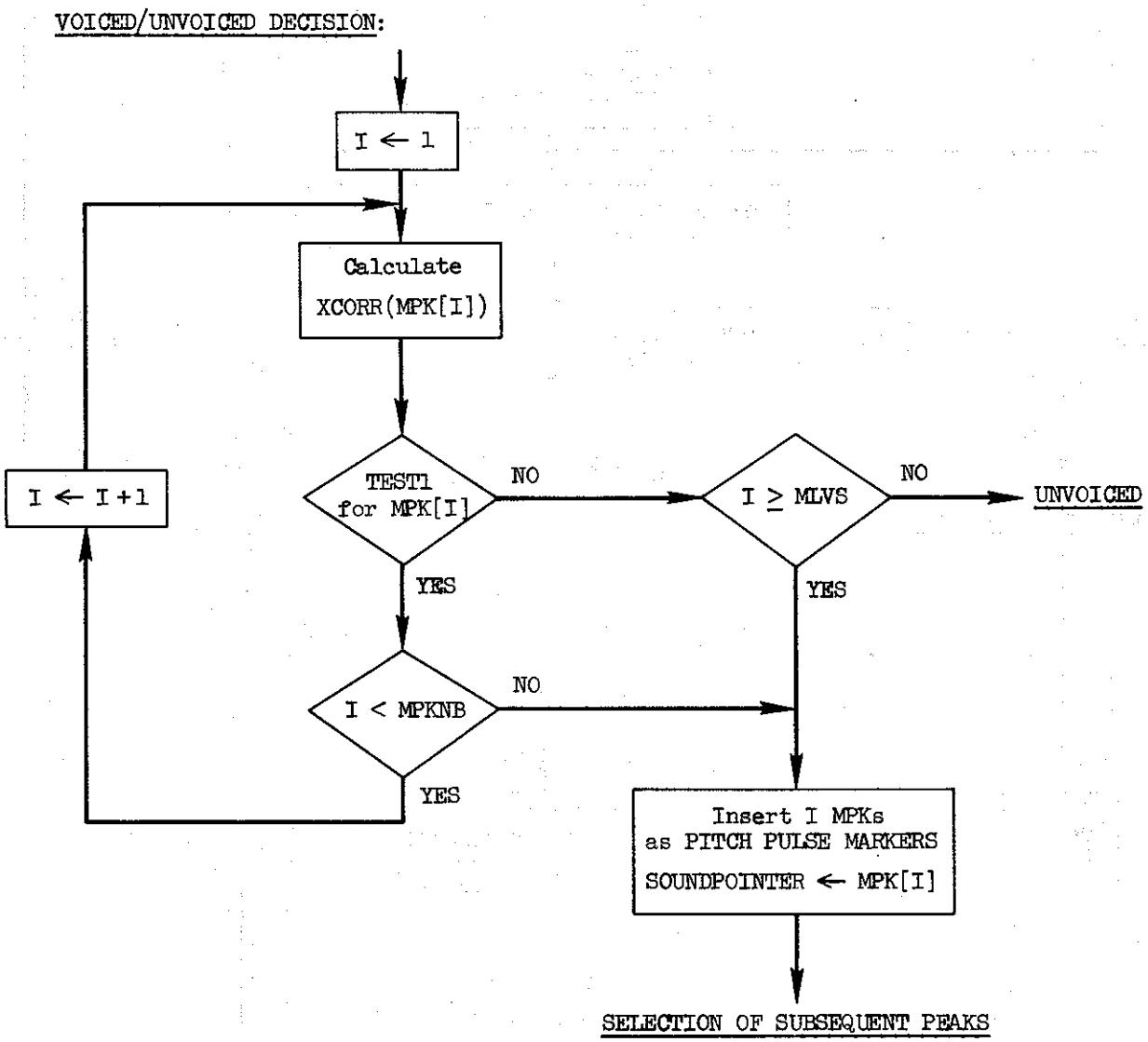
4. J. N. Maksim. "Real time pitch extraction by adaptive prediction of the speech wave." IEEE Conference on Speech Communication and Processing, Pub. IEEE, N.Y., pp. 70-73, April 1972.

APPENDIX A: Algorithm Flowchart



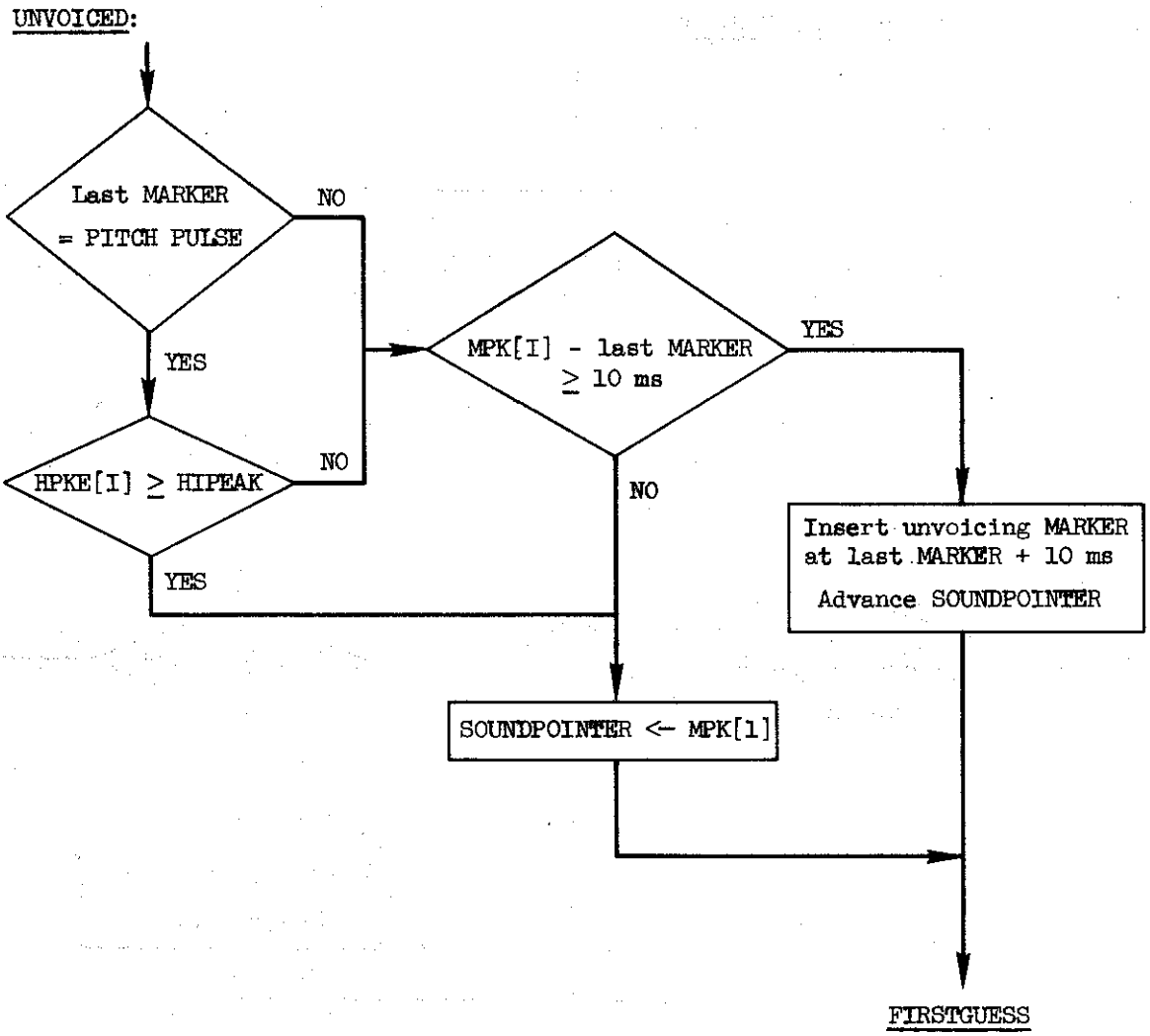
(a) Flow diagram for FIRSTGUESS

(APPENDIX A, cont.)

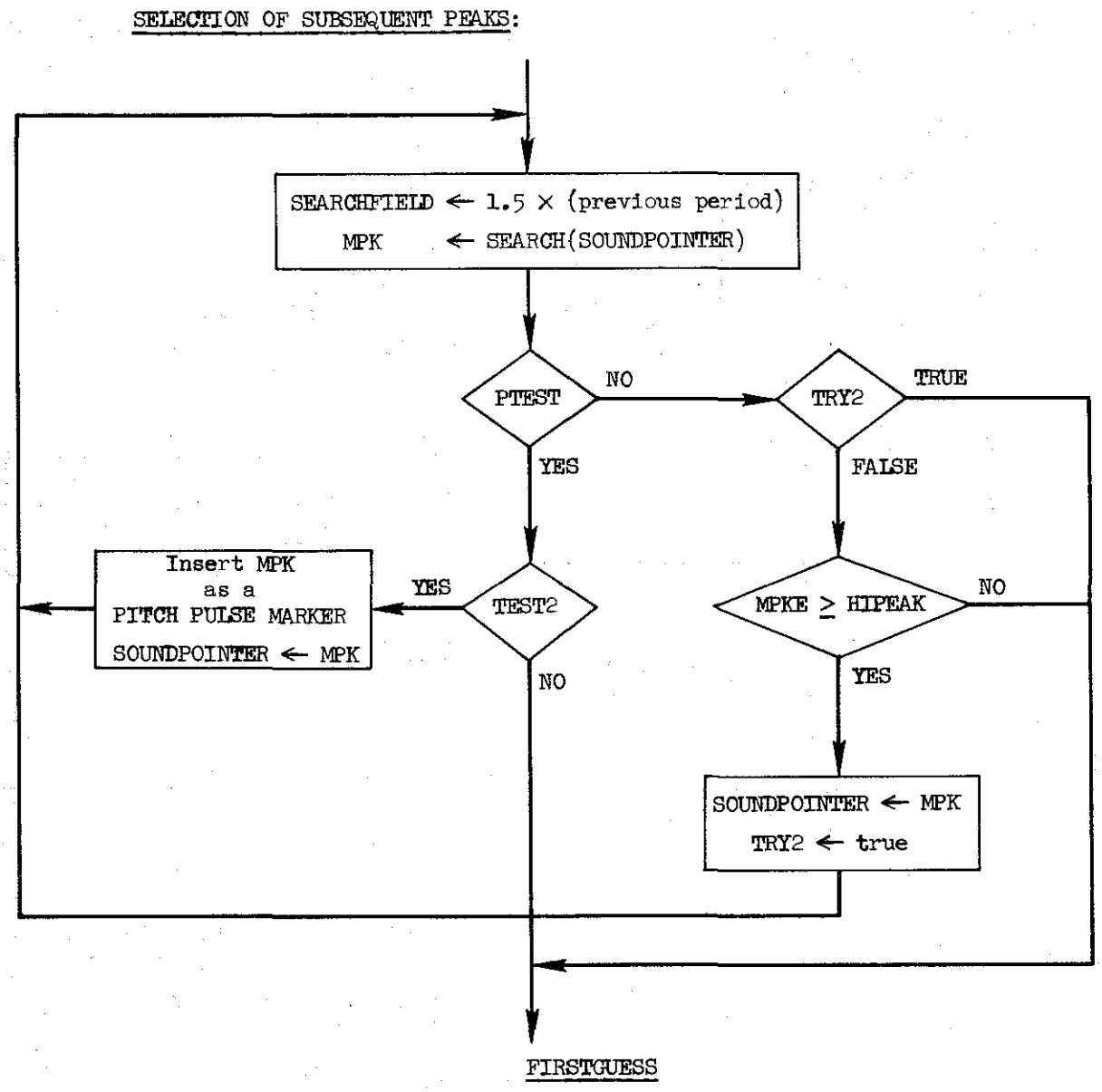


(b) Flow diagram for VOICED/UNVOICED DECISION

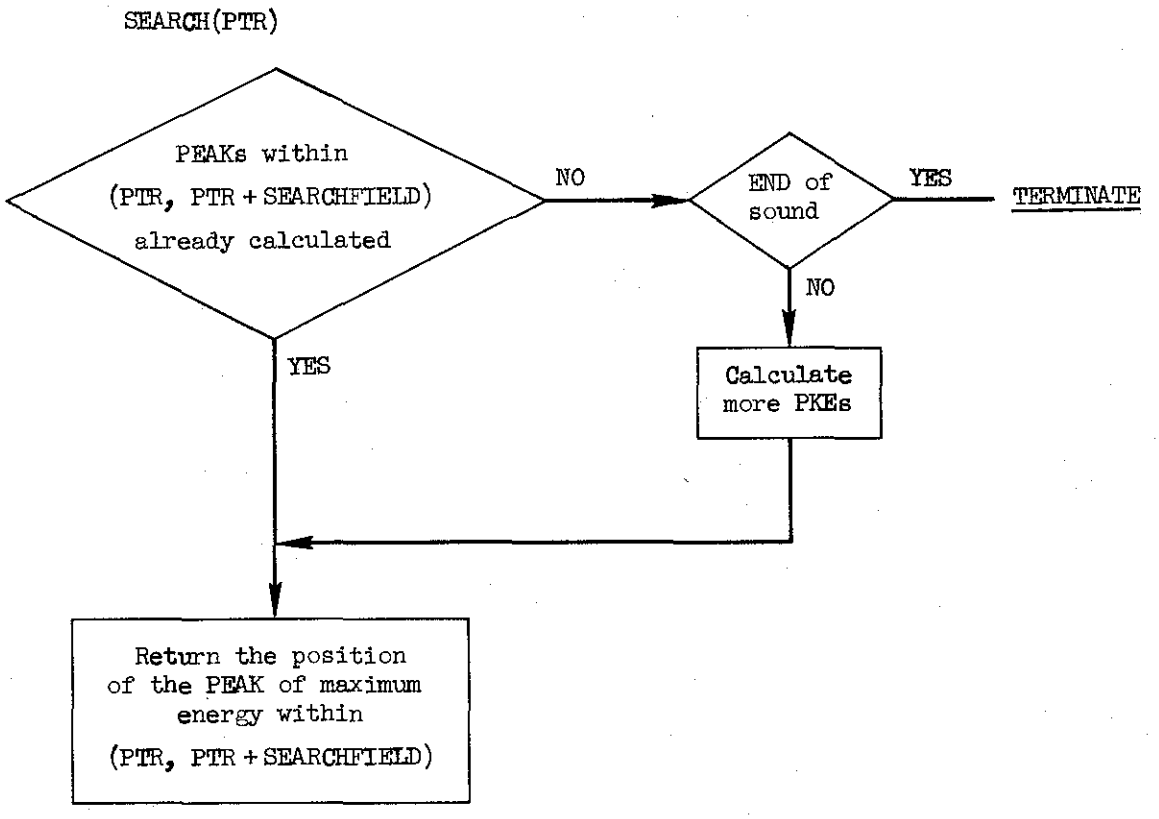
(APPENDIX A, cont.)



(c) Flow diagram for UNVOICED



(d) Flow diagram for SELECTION OF SUBSEQUENT PEAKs



(e) Flow diagram for SEARCH procedure