On certain irregularities of voiced-speech waveforms

Dolansky, L. and Tjernlund, P.

journal: STL-QPSR
volume: 8
number: 2-3
year: 1967
pages: 058-065

http://www.speech.kth.se/qpsr
I. Introduction

It is known that fast and objective quantitative evaluation of various pitch extractors present a difficult problem. While it is natural to put the responsibility for imperfect pitch extraction on equipment malfunction it is possible that other causes, for example difficulty related to pitch frequency definition, may be of importance.

II. Problems studied

This paper is concerned with two problems related to this question: (a) to study irregularities in acoustical waveforms of voiced-speech sounds and relate them to associated glottal excitation waveforms as observed by various methods, and (b) to make available a convenient means for a quantitative evaluation of the performance of various pitch extractors.

III. Problems of pitch extraction

Pitch extractors have in the past often been tested as part of an entire analysis-synthesis system, using some listening tests. Even if various extractors are incorporated into the same system in succession, the possibility exists that the extractors when used in another system would not show the same relative figure of merit.

In addition, since speech is a time-varying process, the very definition of pitch frequency is vague \(^1\). A seemingly obvious remedy is to define the instantaneous pitch frequency as the reciprocal value of the pitch period, and merely identify the time values at which the periods start, if necessary by direct visual inspection of the acoustic waveform \(^2\)(\(^3\)).

---


** Northeastern University, Boston, Mass. 02115, USA
That even such an approach will present problems can easily be understood by examination of Figs. II-D-1 and II-D-2. While in Fig. II-D-1 the individual pitch period can easily be identified certain parts of the waveform shown in Fig. II-D-2 are such that it becomes difficult to do so. Since it is assumed that the laryngeal sound source produces nearly periodic pulses the question arises why the resulting speech waveform is not also of a correspondingly periodic nature.

IV. Possible causes of irregularities

One can speculate about the causes of the occasional lack of periodicity in the voiced speech waveform. Perhaps one or more glottal pulses are missing. On the other hand, the vocal source may generate an additional pulse at times or more than one major discontinuity in the glottal time function may cause additional excitations of the vocal tract. Again, there might be destructive interference between two waveform components which are the result of two consecutive excitation pulses. Sometimes the excitation pulse might in itself be incomplete, for example only an opening or a closing may be present in an individual excitation signal (e.g., at the beginning or end of a voiced portion). Other causes of this kind of waveform irregularity might be a large rate of formant transition, a large rate of fundamental frequency variation, or the magnitude of the pitch frequency itself; sudden changes in the vocal tract (such as occur in the case of stop consonants) may also be a contributing cause.

V. Experimental approach

A. Equipment

In order to investigate the relationship between the acoustic signal and the glottal source signal and thus obtain some explanation of the above-mentioned irregularities in periodicity, a simultaneous recording of the following three signals was made, using the arrangement shown in Fig. II-D-3: (a) regular microphone signal, (b) glottograph signal, (c) larynx microphone signal. A time code generator signal is recorded on an additional channel in order to provide for a time reference for the three above-mentioned signals.
Fig. II-D-1. An example of good waveform for the purpose of visual pitch period extraction.
Fig. II-D-2. Example of a difficult waveform a) for visual pitch period extraction. Waveform b) is a glottograph signal, and c) a larynx microphone signal.
Fig. II-D-3. Recording equipment setup.
Fig. II-D-4. Equipment for simultaneous recording of acoustic, glottographic, and larynx microphone signal.
For convenient visual study the signals were recorded by means of an ink writer. In order to accommodate the entire speech frequency band within the limited frequency band of the ink writer, the reproducing speed of the FM tape recorder was reduced by a factor of 16, with respect to the recording speed.

B. Speech material

The various causes of irregularities (see p. 59, IV. Possible causes of irregularities) were investigated with the help of the utterances listed in Table II-D-I. Each of the parts 1 through 5 is intended to test periodicity irregularity with respect to a particular parameter, for example transitions between speech sounds, intonation patterns, etc.

C. Subjects

Ten persons, five males and five females, were used for the recording of the test signals according to Table II-D-I. The recording was made in an anechoic chamber. The subjects were first asked to make a trial reading of the list before the recording of the signal, and they were especially asked to try to reach the extreme values of their pitch frequencies for the signals listed in Part 2 (Table II-D-I).

VI. Results

The experimental signals which were obtained as outlined on p. 59 (V. A, Experimental approach. Equipment) were evaluated through a study of the multitrace recordings. Examples of such recordings are given in Figs. II-D-5 to II-D-9. As in Fig. II-D-2, the upper trace a) represents the ordinary microphone signal, the middle trace b) represents the glottographic signal, while the lower trace c) shows the larynx microphone signal. The horizontal line under the upper waveform corresponds to the region where irregularities occur. These can occur in the acoustic waveform alone (Fig. II-D-5) or they may be associated with corresponding irregularities in the glottographic and/or throat-microphone waveform. In Fig. II-D-6 this happens towards the end, in Fig. II-D-7 at the beginning of the utterance. The irregularity of Fig. II-D-6, consisting first of alternating complete and incomplete closures, and later of a train of regular, almost sinusoidal incomplete closures, occurs relatively often.
Table II-D-I. List of utterances

<table>
<thead>
<tr>
<th>Part</th>
<th>Utterance</th>
<th>Phonetic description</th>
<th>Tested parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>aːdi:</td>
<td>vowel: open-closed</td>
<td>VCV transitions</td>
</tr>
<tr>
<td></td>
<td>aːni:</td>
<td>consonant: manner of articulation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>aːri:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>aːji:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>aːvi:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>æː</td>
<td>sustained vowels</td>
<td>Pitch at personal extreme values</td>
</tr>
<tr>
<td></td>
<td>æː</td>
<td>rising or falling pitch</td>
<td></td>
</tr>
<tr>
<td></td>
<td>iː</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>skæ'jaː</td>
<td>different stress patterns</td>
<td>Two different intonation patterns</td>
</tr>
<tr>
<td></td>
<td>'skæjaː</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>wiwiwi ajajaj</td>
<td>alternating opening-closing gestures</td>
<td>Fast and extensive formant transitions</td>
</tr>
<tr>
<td>5</td>
<td>devaː mannenn jaaː (5 times)</td>
<td>connected speech</td>
<td></td>
</tr>
</tbody>
</table>
Fig. II-D-5. Example of an irregularity only in the acoustic waveform occurring in the beginning of an utterance.

a) Regular microphone signal.
b) Glottograph signal.
c) Larynx microphone signal.
Fig. II-D-6. Example of the alternating type of glottal irregularity in the terminal portion of the speech signal.

a) Acoustic waveform signal.
b) Glottograph signal.
c) Larynx microphone signal.
Fig. II-D-7. Example of a single glottal pulse irregularity occurring in the utterance.

a) Regular microphone signal.
b) Glottograph signal.
c) Larynx microphone signal.
Some of the most pronounced effects of the vocal tract upon the source are shown in Fig. II-D-8 and Fig. II-D-9. Certain sounds like [r] and [d] appear to lead the voice source so that the throat microphone signal is substantially reduced. Nevertheless, the periodicity of the signal persists.

Of the total number of about 30,000 pitch periods inspected, 78 were classified as irregular. Of these twelve had irregularities only in waveform a) while the remaining 66 had also a correlate in waveform c). Only one error was found in the central portion of unutterance -- all others were either at the beginning or at the end. Five of the eighteen errors in the beginning had irregularities only in waveform a) while the remaining ones were obviously caused by a single glottal pulse change.

With respect to irregularities in the final portion of utterances, six of them are confined to waveform a) only, while the remaining ones have correlates in waveforms b) and c). Irregularities in waveform c) are always of the type exemplified by Fig. II-D-6 and it should be noticed that when complete and incomplete glottal closures are alternating they tend to bunch in groups of two.

A more complete account of the experimental results obtained by means of the speech material described in Table II-D-I is given in quantitative terms in Table II-D-II.

Within the framework of the experimental material considered, it was observed that irregularities often occurred when the pitch frequency was low, and never when it was high. Only one irregularity was found in a rapid formant transition.

Quantitative information with respect to other parameters can be obtained from Table II-D-II.

VII. Excitation function tape

In order to obtain a solution to problem (b), p. 58 (II. Problems studied), a two-channel tape-recording containing the speech signal and the associated timing information for the source signal was made. While in the present investigation attention was focused on pitch extractor evaluation, the testing tape is more generally applicable, i.e., it can be used whenever exact timing of the glottal pulses is needed.
Fig. II-D-8. Example of large influence on vocal source caused by the consonant [r].
The periodicity of the vocal source is retained.

a) Regular microphone signal.
b) Glottograph signal.
c) Larynx microphone signal.
Fig. II-D-9. Example of a heavy loading caused by a stop consonant on the vocal source with prevailing regularity of the source.

a) Acoustic waveform signal.
b) Glottograph signal.
c) Larynx microphone signal.
Table II-D-II. Distribution of irregularities with respect to underlying causes or associated parameters cause or parameter. Total number of samples were about 30,000.

<table>
<thead>
<tr>
<th>Part</th>
<th>Parameter</th>
<th>Parameter value</th>
<th>No. of source irregularities</th>
<th>No. of additional speech waveform irregularities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VCV transitions</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Fundamental frequency</td>
<td>high</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>low</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rising</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>falling</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Different intonation patterns according to Table II-D-I</td>
<td>ska'ja:'</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ska:ja</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Rapid and extensive formant transitions</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Location of irregularity in utterance</td>
<td>beginning</td>
<td>13*</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>middle</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>end</td>
<td>53*</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>Distribution of irregularity according to sex</td>
<td>male</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>female</td>
<td>40</td>
<td>9</td>
</tr>
</tbody>
</table>

* Irregularities at the end of utterances are of a composite type, i.e. consisting of a group of consecutive simple errors. Irregularities in the beginning are of a simple (single period) type. A group of irregularities which appear in a periodic manner is counted as one irregularity.
The criterion to be satisfied is to have a high quality speech signal to which at least the major parts (for example the time of glottal closure) of the excitation function are correctly related in time.

The signals used for the testing tape obtained by the apparatus in Fig. II-D-3 were the same as those used in the waveform irregularity studies.

With the help of the equipment shown in Fig. II-D-10, an individual speech sample is transcribed while a finite time of clock pulses is recorded on the second channel. The next step (Fig. II-D-11) is to feed the recorded speech sample through an A/D converter to a CD-1700 computer. This conversion is made under the control of the clock signal, via an interrupt line to the computer.

The computer-stored speech sample is displayed on an oscilloscope (Fig. II-D-12). With the help of various controls, the signal can be moved along the time axis and desired points on it can be marked. These points correspond to pitch-period boundaries. (These boundaries correspond to the instant where major excitation occurs.)

Finally, (Fig. II-D-13), the stored pitch marking pulses are recorded in synchronism with the original speech sample. As in Fig. II-D-11, the clock signal is used here again to ensure synchronism. The maximum time-measurement error caused by the finite time resolution of the computer does not exceed 160 μsec. The delay of the speech signal fed into the computer by use of the sampling low-pass filter was taken care of by the program.

VIII. Conclusion

On the basis of the studies reported in this paper, the following conclusions can be drawn:

(1) Of about 30,000 pitch signals 78 were judged irregular.

(2) In about 20% of these the corresponding glottal excitation is not irregular.

(3) Irregularities in the beginning of the utterances are usually of a single period type while at the end trains of irregularities are encountered.

(4) Irregular excitation in the ending portions consists of alternating complete and incomplete glottal closures. Usually an incomplete closure is followed almost immediately by a complete closure but a larger distance is found between the complete closure and the following incomplete closure.
Fig. II-D-10. Equipment for simultaneous recording of clock signal and speech sample.
Fig. II-D-11. Equipment for A/D conversion and storage of speech sample in the computer memory under control of clock signal.
Fig. II-D-12. Equipment for displaying, time shifting, and marking of pitch period boundaries.
Fig. II-D-13. Equipment for synchronous recording of original speech signal and corresponding pitch indications.
(5) Even disregarding the multiplicity of irregularities in the terminating portion of waveforms the irregularities at the ends outnumber the ones at the beginning about four to one.

(6) Most of the irregularities occur when the pitch is low. This may be related to conclusions inasmuch as the pitch in the terminating portion is usually low.

(7) There is a considerable spread in the number of total errors among different persons: the range is from 4 to 14.

(8) Rapid rates of variation of formant frequency or fundamental frequency do not appear to cause any waveform irregularity.

This work was carried out at the Speech Transmission Laboratory, Royal Institute of Technology (KTH), Stockholm, and supported in part by a VRA Special Fellowship.

References:


(4) In essence the glottograph measures the resistance across the vocal chords. It has been shown that the glottograph signal very accurately gives the point in time where the vocal chords close.
