3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

LARYNX PERIOD DETECTION METHODS IN SPEECH PATTERN HEARING AIDS

Jianing WEI, David HOWELLS, Andrew FAULKNER and Adrian FOURCIN

Department of Phonetics and Linguistics, University College London, 4 Stephenson Way, London NW1 2HE, UK.

ABSTRACT

In many applications 'pitch extraction' and larynx period measurement are synonymous but in work for the profoundly hearing impaired, and also when detailed information concerning prosodic contrasts is needed, the analysis in normal speech of regularly periodic and also aperiodic laryngeal excitation are essential. The second important need relates to the requirement for period to period accuracy even in conditions of reverberation and ambient noise, and third to the ability to process both normal and pathological speech. We have concentrated on the real-time use of MLP ('neural-net') and peak-picking methods implemented on wearable prostheses, and these are compared with electro-laryngographic techniques and with cepstral processing. Our emphasis is on algorithmic efficacy for both quiet and noisy speech. The developments described come from a programme of work aimed at the real-time analysis of the voice excitation components in speech, in association with the STRIDE project in the EC TIDE programme.

Keywords: *Fundamental Period Detection, Hearing Aids, Speech Signal Processing.*

1. INTRODUCTION

For many years researchers have been investigating methods for the reliable estimation of the fundamental frequency of speech. Many algorithms have been proposed which either operate in the time or frequency domain. The choice of a particular fundamental period estimation method depends on the particular requirements of the application, since each method has its own limitations and strengths.

The present work is part of a five country project in the EC TIDE programme which is directed towards the pre-commercial development of a family of speech pattern element hearing aids for the profoundly hearing-impaired. Many of these people get little speech communication benefit from even the best available conventional hearing aids, and would often otherwise be candidates for a cochlear implant. To provide support for lip-reading, a prototype digital signal processor based hearing aid has been developed. The speech processing which is currently implemented extracts the voice fundamental period information from speech on a period-by-period basis (Tx) and presents it to the patient in the form of a sinusoidal signal, matched to residual hearing ability. Field trials are being carried out with profoundly hearing-impaired listeners in England, the Netherlands, France and Sweden.

Rosen et al. [1] compared connected discourse tracking using the fundamental frequency information derived from four different real-time hardware methods as a supplement to lipreading. These four methods were a) cepstral analysis [2], b) Gold-Rabiner method [3], c) peak-picker [8] based on the algorithm of Gruenz and Schott [4], and d) voice fundamental frequency information derived from the laryngograph [5]. The laryngograph output waveform (Lx) provides the basis for an accurate indication of fundamental period (Tx) and it is used as a reference against which other methods can be compared. The connected discourse tracking (CDT) results showed that the peak-picker was comparable to the laryngograph reference. The cepstral method gave results which were consistently poorer than the reference despite its good performance in tracking fundamenatal frequency (Fx).

In this study the real-time implementation of the MLP and the 'peak-picker' methods is described. Fundamental period information using these two methods is compared with frequency estimates from two other purely software based pitch-detection methods, cepstral analysis and a modified Gold-Rabiner method; the laryngograph signal is used as a reference.

2. REAL-TIME IMPLEMENTATIONS

A prototype wearable digital signal processor based hearing aid has been developed [6]. The aid uses a Texas Instruments TMS320C50 fixed-point digital signal processor and supports 256K bytes FLASH EPROM for non-volatile program and data storage. Analogue input/output is achieved using a 16-bit sigma-delta codec (coder/decoder) which digitizes the signal from an electret microphone at 8kHz. The output amplifier is capable of driving the receiver at levels of 140dB(SPL).

The aid can communicate with a PC via an electrical isolation unit and this link is used for two purposes. The programming of the aid is achieved by downloading the code via the isolation unit to the processor which stores it in the FLASH EPROM. During the setting-up procedure, where the aid is configured to match the needs of the patient, tone-bursts are played out via the aid and threshold, comfort and discomfort levels for a particular patient can be obtained. Once recorded the measurements can be sent to the aid where they are stored in the FLASH EPROM. These values are subsequently used to generate a sinusoidal signal at a correct level for the patient.

The first larynx period extraction algorithm to be implemented on the aid used a multi-layer-perceptron (MLP) trained to detect voice fundamental period [7].
The preliminary process, shown in Figure 1, consisted of the following:

- Filtering the speech signal, sampled at 8kHz, using a bank of six 2nd-order band-pass filters
- half-wave rectification
- low-pass filtering
- decimation to 2 KHz
- storing each resulting value in one of six 41 element delay lines

The MLP has all 246 elements as input and provides a single output which indicates when a larynx-fold closure has occurred. Training of the MLP was performed on a mini-computer using digitized speech and associated laryngograph signals as a reference. Training data for the MLP was provided by a set of recordings which included both adult male and female speech in anechoic and reverberant conditions. Noise was also optionally added to the acoustic speech data.

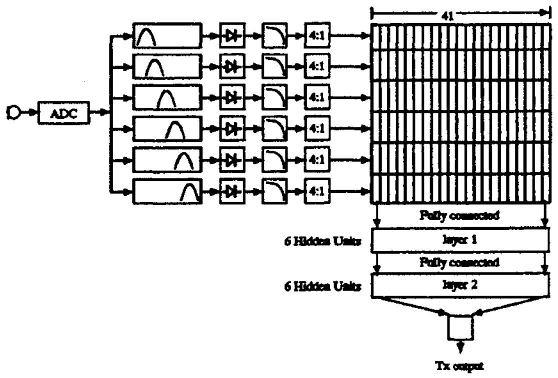

Figure 1: The MLP Process.

The second algorithm implemented on the SiVo aid was the 'peak-picker'. This algorithm works completely in the time-domain as opposed to the MLP process which combines both time and frequency domain analyses. The 'peak-picker' is an untrained algorithm.
The process as implemented on the aid first low-pass filters the digitized speech then uses a series of differentiators and 'peak-pickers' to detect the dominant leading edge of the sound pressure wave-front. Final output is produced after thresholding and pulse-stretching with a monostable multi-vibrator. This is shown in Figure 2.
The 'peak-picker' derives it name from the technique used to reject secondary peaks in the resultant speech signal.
Typically in an analogue version it is implemented as a diode fed capacitor with constant current discharge. The charge rate is controlled so that variations in the input signal after the main peak are substantially ignored - the resultant output signal being the decaying voltage across the capacitor. The capacitor is recharged as soon as the input voltage is greater than that held by the capacitor itself.
Using the digital signal processor, processes such as differentiation and 'peak-picking' are implemented easily and have the advantage of being efficient and repeatable across all devices.

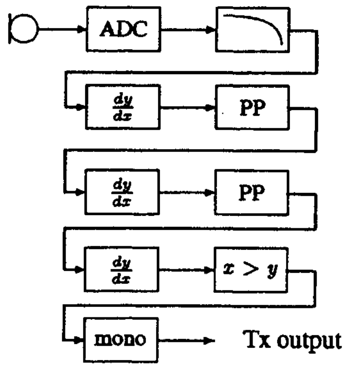

Figure 2: The Peak-picker Process.

An important factor, when comparing two algorithms such as the MLP and 'peak-picker', comes from the effect of quantization on the value of the fundamental period estimate. In the case of the MLP, period estimates are to the nearest 0.5ms. The resultant jitter is clearly visible in the fundamental frequency (Fx) contour. However, for patients with profound hearing loss this is not important. The 'peak-picker' period estimates are made to the nearest 0.125ms (the sampling period) resulting in smaller distortion of the Fx contour.

3. COMPARISON RESULTS

The MLP and the peak-picker methods are compared with other voice fundamental period extraction methods such as the cepstral analysis and a modified Gold-Rabiner method, using the laryngograph signal as a reference. The window sizes of 32ms and 64ms are used in the cepstral analyses. The modified Gold-Rabiner algorithm, written for the "Speech Pattern and Algorithmic Representations" (SPAR) project, was based on the original combined with the autocorrelation function using 32ms window. The following figures show the results of the different methods for quiet speech, speech in noise at SNR=15dB and at SNR=10dB. Here the SNR is conservatively defined as the speech maximum rms value over a 500ms window divided by the rms value of the noise. The noise signal used is "speech-spectrum" shaped noise which has greater energy at the low-frequencies than in the high-frequencies.
From Figure 3 we can see that the MLP, peak-picker, and cepstral analyses perform quite well for voice/voiceless detection on quiet speech. The voice/voiceless detection error in the modified Gold-Rabiner method is significantly higher than in the other methods. When noise is introduced at SNR=15 dB level, the cepstral analysis starts to fail in estimating the larynx period, and although the modified Gold-Rabiner method can extract the fundamental period, it failed in voice/voiceless sound classification. However, the MLP and peak-picker can still operate reasonably well at this SNR level. When SNR=10 dB, the peak-picker starts to respond to noise. The MLP can still make voice/voiceless decisions, though it starts to generate more errors in period estimation when speech is voiced.

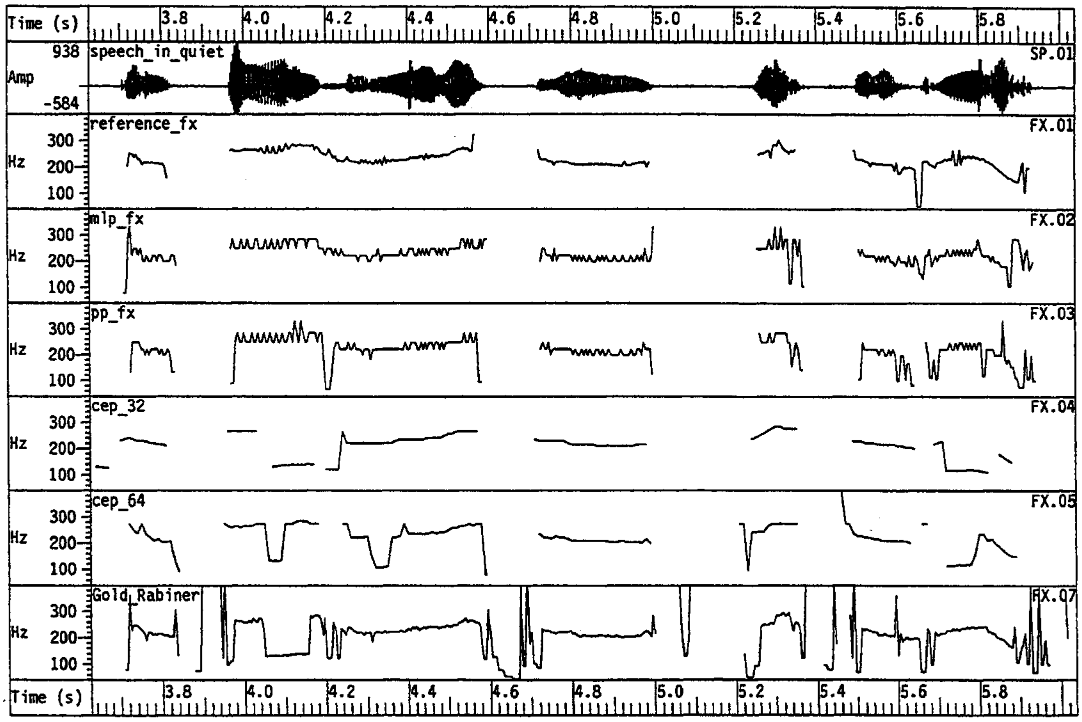

Figure 3: Different Voice Period Detection Methods Operated on Quiet Speech.

Key to Figures:
SP.01: Speech waveform by a female speaker, with added noise.
FX.01: Reference fundamental frequency contour derived from the Laryngograph signal.
FX.02: Fundamental frequency contour derived from the MLP method.
FX.03: Fundamental frequency contour derived from the peak-picker method.
FX.04: Fundamental frequency contour derived from the cepstral analysis with 32ms window.
FX.05: Fundamental frequency contour derived from the cepstral analysis with 64ms window.
FX.06: Fundamental frequency contour derived from the Gold-Rabiner method.

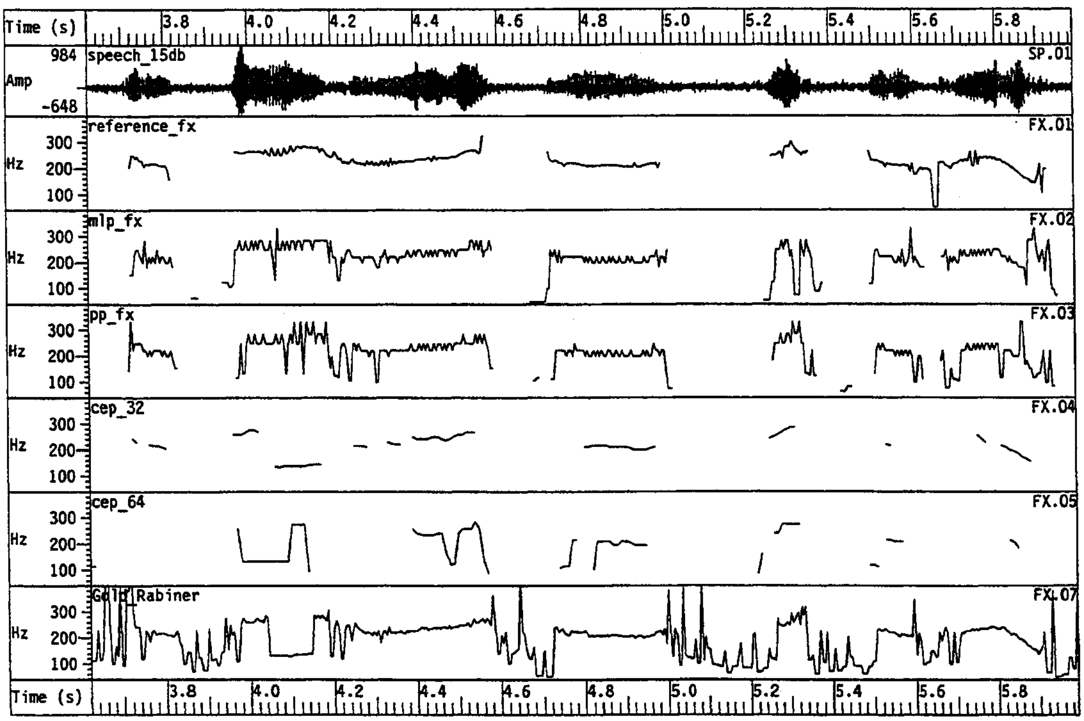

Figure 4: Different Voice Period Detection Methods Operated on Speech with SNR=15 dB.

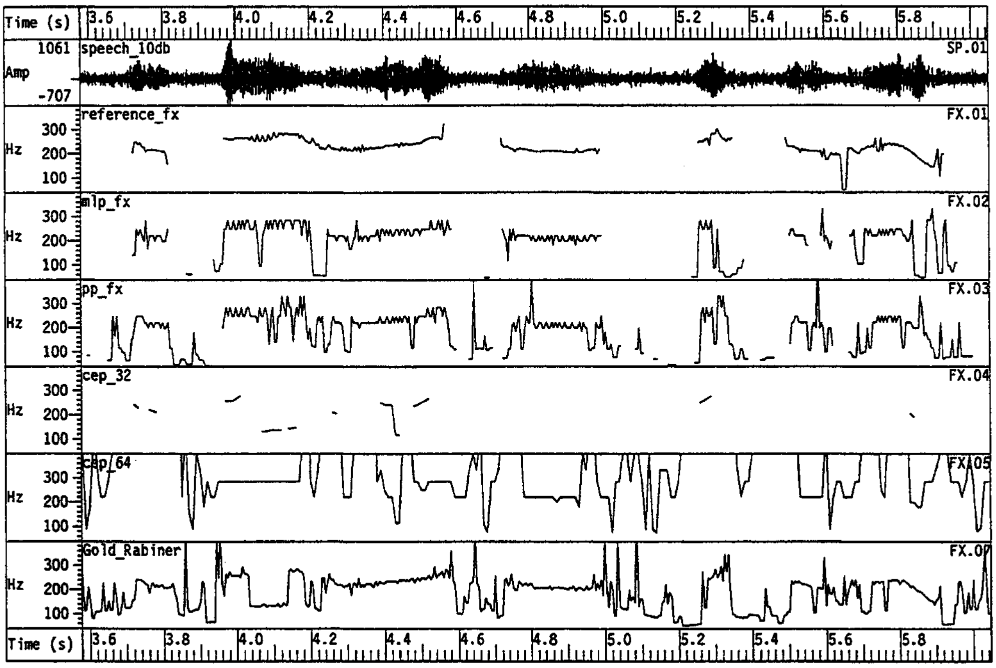

Figure 5: Different Voice Period Detection Methods Operated on Speech with SNR=10 dB.

4. CONCLUSIONS

The results presented in this study have shown that the MLP and peak-picker real-time larynx period detection methods which have been implemented on the wearable SiVo speech pattern processing hearing aid performed better in noise environment than the cepstral and modified Gold-Rabiner methods in our implementations. The MLP algorithm is more robust compared with the peak-picker, cepstral and Gold-Rabiner analysis.

ACKNOWLEDGEMENT

This work is supported by EC TIDE project 133/206, UK Department of Health, and UK MRC grant G9020214.

REFERENCES

[1] Rosen, S., Fourcin, A.J; Walliker, J.R.; Douek, E.E.; and Moore, B.C.J.: External electrical stimulation of the cochlea in the totally deaf. In J. Raviv, editor, Use of Computers in Aiding the Disabled, pages 167--183. North-Holland, 1982.

[2] Noll, A. M.: Cepstral pitch determination. J. Acoust. Soc. Amer., 41:293--309, 1967.

[3] Gold, B. and Rabiner, L. R.: Parallel processing techniques for estimating pitch periods of speech in the time-domain, J. Acoust. Soc. Amer., 46, pp.442-448.

[4] Gruenz O. O.; Schott, L.O.: Extraction and portrayal of pitch of speech sounds. J. Acoust. Soc. Amer., 21:487--495, 1949.

[5] Fourcin, A. J.; Abberton, E.: First applications of a new laryngograph. Med. and Biol., I11:172--182, 1971.

[6] Walliker, J.; Daley, J.; Smith, K.; Faulkner, A.; and Fourcin, A.: Speech analytic hearing aids for the profoundly deaf: Technical design aspects and user field trial results. In Bjorn Granstrom, Sheri Hunnicut, and Karl-Erik Spens, editors, Proceedings of an ESCA Workshop, Stockholm, Sweden, 1993.

[7] Howard, I.: Speech fundamental period estimation using pattern classification, Ph.D thesis, Dept. of Phonetics and Linguistics, University College London.

[8] Howard, D. and Fourcin, A. J.: Instantaneous voice period measurement for cochlear stimulation, Electronics Letters, 19, 776-778.