# A Comparison of F0 Extraction Algorithms for Sustained Vowels*

Vijay Parsa and Donald G. Jamieson,
Hearing Health Care Research Unit, University of Western Ontario, London, Ontario.

## I Introduction

Measures of vocal perturbation are routinely used in clinical evaluation of a patient's voice and also in monitoring the patient's progress over the treatment period [1]. These measures can be broadly classified into: a) jitter, which measures the cycle-to-cycle variation of the fundamental frequency, b) shimmer, which calculates cycle-to-cycle amplitude variations, and c) Signal-to-Noise Ratio (SNR) which represents the amount of vocal noise. These measures are calculated pitch-synchronously, hence accurate estimation of the fundamental frequency is crucial.

## II Fundamental frequency (F0) estimation

Several algorithms are proposed in the literature for F0 extraction [2]. In a recent study, Titze *et al.*[3] compared the performance of waveform matching, peak picking and zero crossing based pitch algorithms and concluded that the waveform matching (WM) algorithm, one which adjusts the pitch period such that consecutive cycles are matched in a mean-squared sense, offers the best performance under a variety of conditions. This waveform matching technique is also the heart of a high resolution pitch algorithm proposed by Medan *et al.* [4].

The WM algorithm requires an initial F0 estimate. Titze *et al* [3] accomplished this by making an initial pass through the entire waveform and judiciously marking negative going zero-crossings. This could be a time consuming process, especially for longer signals. This study is an extension to [3] where the performance of six additional algorithms based on the Average Magnitude Difference Function (AMDF), Autocorrelation Function (AF), Autocorrelation function with Center Clipping (ACC), Inverse Filter Autocorrelation (IFAC), Cepstrum (CEP), and Harmonic Product Spectrum (HPS) respectively, is compared . In addition, the possibility of using these algorithms to provide an initial F0 estimate to the waveform matching algorithm is investigated.

## III Results

To evaluate these algorithms, synthetic vowel waveforms were used. The vowel waveform was generated following the procedure described in [4] with a sampling rate of 20 kHz. The performance of the algorithms was tested in the presence of background noise with a fixed F0 = 150 Hz and no amplitude variations. The Signal-to-Noise Ratio (SNR, also termed Harmonics-to-Noise Ratio in [2]) was varied from -2 to 20 dB, a range that covers both normal and pathological voices. Note that the F0 is not an integer multiple of the sampling frequency, hence interpolation techniques are required. We used the interpolation technique proposed in [4] as we found this to be better than the second order interpolation used in [3]. Algorithm performance was quantified by mean relative deviation, α, between the true pitch contour, $P_{N,}$ and the estimated pitch contour, $Q_N$, where $N$ is the number of pitch values.

Figure 1 shows the effect of background noise on the performance of these algorithms. The salient points from this graph are: a) the performance of all the algorithms deteriorates with an increase in the background noise, b) the AMDF method is more sensitive to background noise compared to others, b) the AC and ACC methods are more robust to background noise than IFAC, and the ACC method breaks down at low SNRs, as both the center-clipping and inverse filtering operations are affected at larger noise levels, c) time-domain methods (AC, ACC, IFAC) are more accurate than frequency domain (HPS and CEP), due to lower resolution in the frequency domain. A window size of 2048 samples was used resulting in a frequency resolution of around 9 Hz. Larger window sizes will no doubt improve the pitch estimates but also increase the computational complexity. Figure 2 depicts the performance of these algorithms when they were used to provide the F0 estimate to the WM algorithm. The F0 was estimated using the first 50 ms of the signal, and this estimate was refined using the WM algorithm over the entire waveform. Notice that almost all algorithms perform equally well with increased robustness to background noise. The HPS method appears to break down towards the lower end of the SNR spectrum due to the loss of distinct harmonic structure.

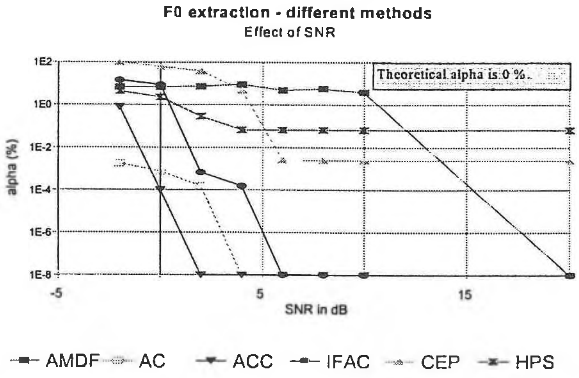

**Figure 1:** Effect of background noise on F0 estimation.

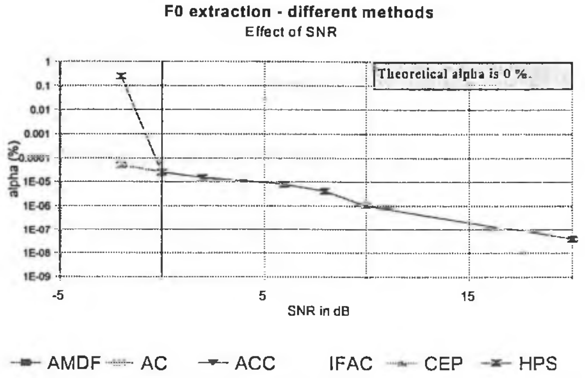

**Figure 2:** Effect of SNR on WM algorithm with initial F0 provided by different methods.

## IV Conclusions

The WM algorithm provides a high resolution F0 estimate. Any of the AC, ACC, IFAC and CEP algorithms can be chosen to provide an initial estimate of the F0 based on the initial few cycles of waveform. This averts the need for marking the whole waveform based on zero-crossings, resulting in a computationally more efficient pitch estimation.

## V References

[1] Baken, R.J., *Clinical Measurement of Speech and Voice*, College-Hill Press, Boston, 1987.

[2] Hess, W., *Pitch Determination of Speech Signals*, Springer-Verlag, Germany, 1983.

[3] Titze, I.R., and Liang, H., "Comparison of F0 Extraction Methods for High-Precision Voice Perturbation Measures", *JSHR*, V 36, 1120-1133, 1993.

[4] Medan,Y., Yair, E., and Chazan, D., "Super Resolution Pitch Determination of Speech Signals", *IEEE trans. SP*, V 39, 40-48, 1991.

* supported by NSERC and the CAA Shaw postdoctoral fellowship.