A HIGH-RESOLUTION GLOTTAL PULSE TRACKER

R. Rodman[*], *D. McAllister*[*], *D. Bitzer*[*], *D. Chappell*[†]

[*]Department of Computer Science
North Carolina State University
Raleigh, N.C. 27695-8206, U.S.A.

[†]LIPSinc.
110 MacKenan Dr., Suite 300-A
Cary, N.C. 27511, U.S.A.

ABSTRACT

A new method of computing the glottal pulse period of voiced speech is given. This algorithm is based on the mathematically derived fact that the amplitudes of the odd harmonics of a periodic function with period P are zero when the function is expanded in a Fourier series whose coefficients are determined by integrating over $2P$ instead of the usual P. It is shown that such a glottal pulse tracker is extremely sensitive to sudden short-lived shifts in the apparent frequency of the glottal pulse that circumscribe certain consonants in the speech stream. This method may therefore be used to segment these consonants for various analytical purposes.

1. INTRODUCTION

In many speech processing applications, it is important to determine accurately pitch period or glottal pulse positions. Knowledge of pitch period locations is very important for pitch-synchronous algorithms. Glottal pulse detection can also be used as a refined method of determining precise pitch. In addition, our lip-synchronization research is aided by the ability to identify acoustic-phonetic classes through sudden shifts in the apparent frequency of the glottal pulse [1-5]. Many glottal pulse detection algorithms are refinements of pitch estimation programs. There are a number of common methods for pitch detection, and time-domain techniques can locate specific pitch period times [7]. In this paper, we present a new method of tracking the glottal pulse period based on certain mathematical properties of Fourier analysis.

2. DERIVATION OF ODD HARMONIC METHOD OF LOCATING GLOTTAL PULSE PERIOD

Let f(t) be periodic with period P. (i.e., $f(t) = f(t+kP)$, $k = 0,1,2, \ldots$)
By Fourier analysis we have:

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n\cos(nt) + b_n\sin(nt),$$

called a Fourier series, where

$$a_0 = \frac{1}{P}\int_0^P f(t)dt$$

$$a_n = \frac{2}{P}\int_0^P f(t)\cos(\frac{2\pi n}{P})t\,dt$$

$$b_n = \frac{2}{P}\int_0^P f(t)\sin(\frac{2\pi n}{P})t\,dt$$

a_0 is the so-called “DC term.” Inside the summation, the nth term is called the nth harmonic.

We now wish to consider f(t) from the point of view of being periodic with a period of $2P$ (which, of course, it is). In that case the magnitude of the odd harmonics (those values of a_n and b_n for which n=1,3, . . in the Fourier series) is zero. That is:

$$a_n = \frac{2}{2P}\int_0^{2P} f(t)\cos(\frac{2\pi n}{2P})t\,dt = 0,\ n=1,3,..$$

$$b_n = \frac{2}{2P}\int_0^{2P} f(t)\sin(\frac{2\pi n}{2P})t\,dt = 0,\ n=1,3,..$$

Canceling the 2's gives us:

$$a_n = \frac{1}{P}\int_0^{2P} f(t)\cos(\frac{\pi n}{P})t\,dt = 0,\ n = 1,3,.. \quad \textbf{(Eq. 1)}$$

$$b_n = \frac{1}{P}\int_0^{2P} f(t)\sin(\frac{\pi n}{P})t\,dt = 0,\ n = 1,3,.. \quad \textbf{(Eq. 2)}$$

We first prove a Lemma.

Lemma: $\cos\{(\frac{\pi n}{P})(t+P)\} = -\cos(\frac{\pi n}{P}t)$

Proof of Lemma:
Use trig identity: cos(A+B)=cosAcosB-sinAsinB.

$$\cos\{(\frac{\pi n}{P})(t+P)\} = \cos(\frac{\pi n}{P}t + \pi n) =$$

$$\cos(\frac{\pi n}{P}t)\cos\pi n - \sin(\frac{\pi n}{P}t)\sin\pi n$$

The second term is zero since $\sin\pi n = 0$ for all integral values of n. As well, $\cos\pi n = -1$ for odd values of n.

Thus, $\cos\{(\frac{\pi n}{P})(t+P)\} = -\cos(\frac{\pi n}{P}t)$

Similarly, $\sin\{(\frac{\pi n}{P})(t+P)\} = -\sin(\frac{\pi n}{P}t)$

QED Lemma. (NB: The proof captures the "obvious" truth that moving half of a period along a sinusoidal gives the negative of the starting value.)

We may now prove the theorem that Eq. 1 and Eq. 2 are true.

Proof:

We can ignore the preceding coefficients ($\frac{1}{P}$) in the proof. We show that

$$\int_0^{2P} f(t)\cos(\frac{\pi n}{P}t)\,dt = 0, \quad n = 1,3, ..$$

$$\int_0^{2P} f(t)\cos(\frac{\pi n}{P}t)dt =$$

$$\int_0^{P} f(t)\cos(\frac{\pi n}{P}t)dt + \int_P^{2P} f(t)\cos(\frac{\pi n}{P}t)\,dt$$

We can rewrite the second term on the right side as follows (this amounts to a “change in variable"):

$$\int_P^{2P} f(t)\cos(\frac{\pi n}{P}t)\,dt = \int_0^{P} f(t+P)\cos\{(\frac{\pi n}{P})(t+P)\}dt$$

$f(t+P)=f(t)$ since $f(t)$ is periodic in P. By the Lemma $\cos\{(\frac{\pi n}{P})(t+P)\}=-\cos(\frac{\pi n}{P}t)$, $n = 1,3, . .$ thus, extracting the minus sign from the integral, we get

$$\int_P^{2P} f(t)\cos(\frac{\pi n}{P}t)\,dt = -\int_0^{P} f(t)\cos(\frac{\pi n}{P}t)\,dt \quad \text{and so}$$

$$\int_0^{2P} f(t)\cos(\frac{\pi n}{P}t)dt = \int_0^{P} f(t)\cos(\frac{\pi n}{P}t)dt - \int_0^{P} f(t)\cos(\frac{\pi n}{P}t)\,dt = 0, \; n = 1,3, .. \qquad \text{QED}$$

An entirely analogous proof shows that

$$\int_0^{2P} f(t)\sin(\frac{\pi n}{P}t)\,dt = 0, \; n = 1,3, ..$$

This proof embodies the "obvious" fact that if you start with a sinusoidal clipped with a rectangular window, and use a frame length of twice the period, then the discrete Fourier transform (DFT) will have a single pulse at the second non-DC term, and be zero everywhere else, including, of course, the odd harmonics.

If one were given a periodic function f(t), and an approximation to 2*P*, say X, one could compute *P* exactly by varying Δ and computing various integrals of the form

$$\int_0^{X\pm\Delta} f(t)\cos(\frac{\pi n}{P})t\,dt \qquad n=1,3, . .$$

The value of Δ for which the integrals (or the sum of some subset of them) are 0 (or very close to 0) could then be used to compute the actual period.

A practical computational method of finding the period of a periodic function f(t) whose period is nearly known is to compute DFTs on the sampled f(t). That is, given a periodic function f(t), and an approximation to the number of *samples* in *twice* its period, say S, one computes various DFTs in window sizes of $S \pm \Delta$ and examines the values of the odd harmonics. When they are closest to zero, the value of Δ can be used to determine the actual period. Note that the window length is very important, and thus the fast Fourier transform (FFT) is not generally usable.

Suppose now that we have a function that is quasi- periodic. By this we mean two things: the function repeats itself at intervals that are approximately, but not exactly, the same; and with each repetition, the function may change slightly and not be identical in all respects to the one in the previous cycle. Voiced speech is an example of a quasi-periodic function of time. We can approximate the "current" period (i.e., the period over the current window) using the following algorithm. Digitize the signal. Assuming we know the approximate period length S in number of samples (within 20%, say), we compute various DFTs in window sizes of $S \pm \Delta$. For each such computation we examine the sum of some number of odd harmonics (say the first four). The value of Δ when that sum is *minimal* is used to determine the period. The sum will be minimal when the number of samples in the window used to compute the DFT represents *twice* the sought after period.

The window size estimate must be large enough to handle the rapid excursions of the glottal pulse period (GPP) when it encounters a fricative, say, but not so large as to allow alternate extrema, which would result in the tracking of a different multiple of the period. On the other hand if the estimate is too low, the glottal pulse (GP) tracker converges to zero and crashes. In practice the size of the previous GPP is used to determine the size of the next window. We have found that $GPP_{previous}$ +/- 50 is a workable window for estimating for GPP_{next}.

Figure 1 of the Appendix is a plot of the sum of the first four harmonics versus the number of samples in the window. The voice exemplar is in the nucleus of /e/ as pronounced by an adult female. Since the window in this case is [182, 282], and the minimum is at the 52nd sample, the GPP = (182 + 52) / 2 = 117 samples. Dividing that into the sampling rate of 22.050 kHz yields a frequency (pitch) of 188 Hz. The convergence is very sharp with a resolution of one sample or approximately 0.05 milliseconds. On the other hand, an attempt to apply the tracker to an unvoiced sound correctly reveals the erratic behavior of the tracker in the absence of a glottal pulse, as seen in Figure 2 of the Appendix.

3. COMPARISON WITH STANDARD PITCH PERIOD TRACKERS

This method of glottal pulse tracking, while computationally somewhat intensive, has great sensitivity to sudden shifts in frequency induced by transitions in the vocal tract. To demonstrate this we recorded a speaker saying /amanalavazazhadha/. That is, the consonants /m/, /n/, /l/, /v/, /z/, /zh/, and /dh/ spoken with the vowel /a/ (as in *father)* separating them. The result of digitizing this speech at 22.050 kHz, 8-bit quantization, is a raw signal containing 84,992 samples. (We ignore about 1500 samples at the end of the utterance to make smoothing the transform simpler, hence the missing second half of the syllable /dha/, which is not relevant.)

If we plot the number of samples in a glottal pulse period (GPP) versus the *glottal pulse number* (GP#), we arrive at Figure 3 of the Appendix. The higher y-values are lower frequency values since the more samples in a glottal pulse, the lower the frequency. The four peaks to the right represent the drop in the apparent glottal pulse frequency that accompanies the voiced fricatives. This has been observed by acoustic phoneticians [8] and is due partly to an actual reduction in the vibration rate due to the higher supraglottal pressure caused by the resistance of the fricative, and partly by the filtering effect of the articulators.

Of more interest are the small frequency deviations that circumscribe the sonorants /m/, /n/, and /l/. These delta-frequency delimiters are found approximately at GP#s 42 and 62 for /m/; 105 (somewhat indistinctly) and 136 for /n/; and 179 and 204 for /l/. We have verified by other methods that these are indeed the loci of the transitions between the vowels and consonants.

The source of these sudden brief frequency shifts are, in the case of the nasals /m/ and /n/, the result of the rapid phase changes induced by the opening and closing of the nasal cavity port, which changes the number of acoustic paths from source to egress. In the case of the lateral consonant /l/, the phase changes result from the position of the tongue typical of the lateral approximant, which creates more than one path for the sound to propagate through the vocal tract. This behavior in the speech signals of /m/, /n/ and /l/ is anticipated in [8].

Other glottal pulse trackers are not always sensitive to the frequency shift phenomenon. For example, when we track the glottal pulse of the same utterance using the more traditional autocorrelation method we get Figure 4 of the Appendix. The resolution of this tracker is sufficient to detect the frequency shifts that accompany the voiced fricatives, but the variation in frequency that circumscribe the sonorants are either missing or barely visible except for the one at GP# 62.

Our research is ongoing and we are currently examining the Modified Simplified Inverse Filter Tracking (MSIFT) algorithm [6], among others, for comparison with the odd harmonic pitch tracker presented here.

4. CONCLUSION

The new glottal pulse tracking algorithm presented here has been shown not only to have a strong mathematical basis but also practical usefulness. A pitch tracker of such high sensitivity allows segmentation of various consonants from the speech stream for special treatment. For our work in lip synchronization [1-5], it is especially useful because it allows the use of different methods for mapping acoustic features into mouth shape parameters depending on the acoustic-phonetic class.

5. REFERENCES

[1] Koster, B., Rodman, R., and Bitzer, D. "Automated lip-sync: direct translation of speech-sound to mouth shape." *Proceedings of the 28th Annual Asilomar Conference on Signals, Systems and Computers,* (IEEE publication) 1994, pp 36-46.

[2] Koster, B. *Automatic Lip-Sync: Direct Translation of Speech-Sound to Mouth-Animation.* Ph.D. Dissertation. Department of Computer Science, North Carolina State University, Raleigh, NC 27695. 1995.

[3] McAllister, D.F., Rodman, R.D., Bitzer, D.L. and Freeman, A.S. "Lip Synchronization for Animation." *SIGGRAPH 97 Visual Proceedings,* 1997, p 225.

[4] McAllister, D.F., Rodman, R.D., Bitzer, D.L. and Freeman, A.S. "Lip Synchronization as an Aid to the Hearing Impaired," *Proc. AVIOS 97,* 1997, pp 233-248.

[5] McAllister, D.F., Rodman, R.D., Bitzer, D.L. and Freeman, A.S. "Lip synchronization of speech." *Proc. AVSP 97*, 1997, pp 133-136.

[6] Markel, J.D. "The SIFT Algorithm for Fundamental Frequency Estimation," *IEEE Trans. Audio and Electroacoust.* Vol. AU-20, No. 5. Dec. 1972.

[7] O’Shaughnessy, D. *Speech Communication: Human and Machine*. Reading, Mass.: Addison-Wesley, 1987.

[8] Stevens, K.N. *Acoustic Phonetics.* Cambridge, MA:The MIT Press, 1998.

6. APPENDIX

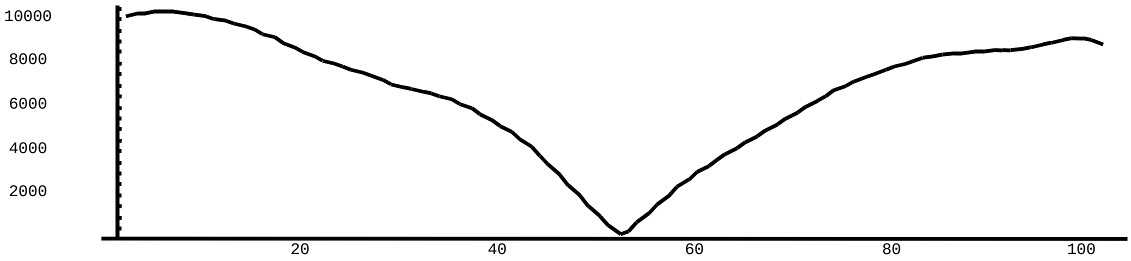

Figure 1. Sum of odd harmonics for a window of voiced speech

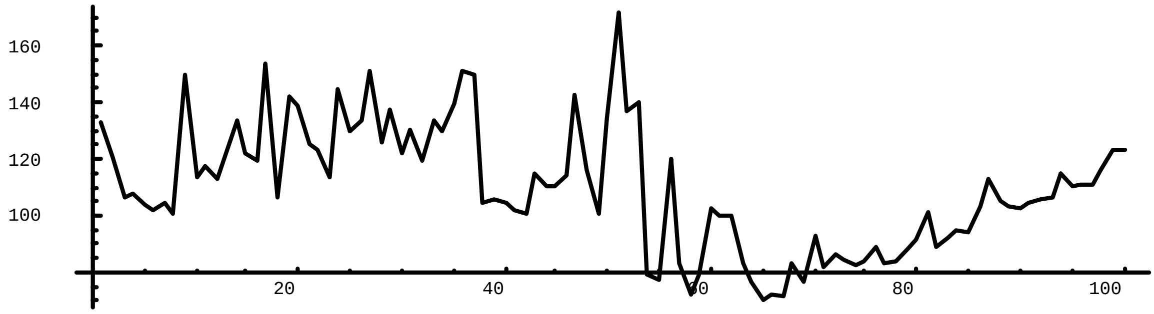

Figure 2. Sum of odd harmonics for a window of unvoiced speech

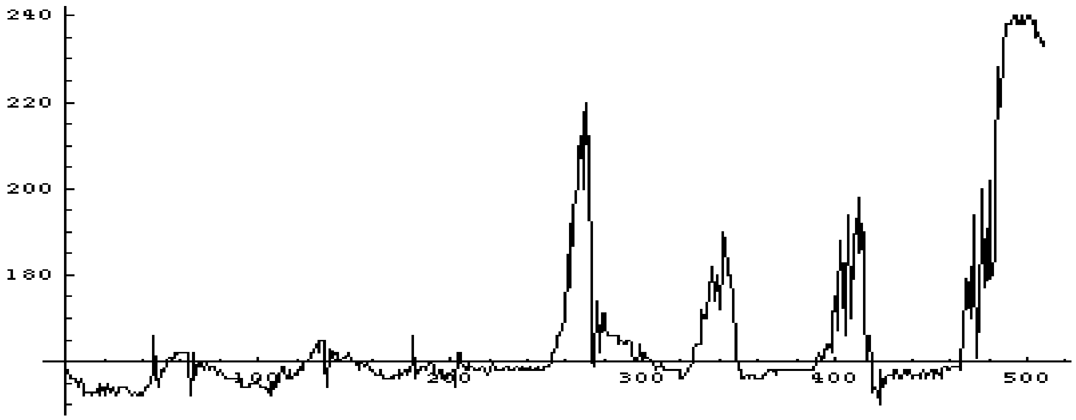

Figure 3. GPP versus GP# using the "odd harmonics" algorithm

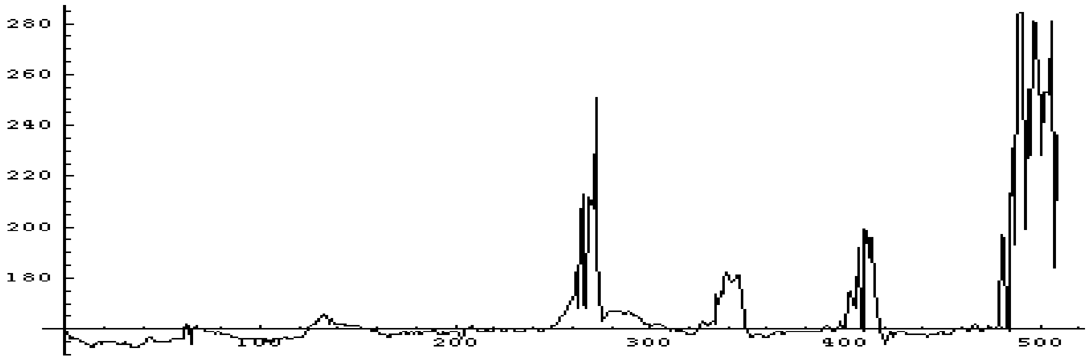

Figure 4. GPP versus GP# using a traditional autocorrelation algorithm