THE USE OF TEMPORAL FREQUENCY IN SPEECH SIGNAL ANALYSIS

DAVID A. SEGGIE

Department of Phonetics and Linguistics, University College London,
London NW1 2HE, U.K.

ABSTRACT

An analytic signal representation enables the estimation of speech signal temporal frequency. The use of this time-domain attribute in speech signal analysis is illustrated. In addition, the relationship between a signal's temporal frequency and its spectral composition is elucidated.

INTRODUCTION

In speech signal analysis, a basic goal is to extract from the signal those acoustic attributes useful in signifying phonetic contrasts. Given the fact that these attributes appear to be encoded in the speech signal in a highly complicated manner, attempts at achieving this goal often involve the transformation of the data into what is thought to be a more appropriate representation - appropriate in the sense that salient acoustic characteristics are brought to the fore. For example, models of the acoustics of speech production [1] and studies of the frequency selectivity of hearing [2], indicate that one such appropriate representation of the speech waveform is in terms of its short-term amplitude spectra. Salient acoustic features, (e.g. temporal variations in formant frequencies), can then be readily estimated from these spectra.

However, the demonstrated utility of established speech signal representations should not prohibit the assessment of novel ways of viewing speech pressure waveforms. Indeed, given that a phonetic contrast is usually signalled by many different acoustic parameters, it would seem eminently sensible to view the waveform in several different ways in order to uncover the overall acoustic pattern .

Recent work in both seismic signal processing [3,4] and ultrasonic imaging [5,6], indicates that the temporal frequency characteristics of acoustic signals encode useful information. In speech signal processing, preliminary studies [7,8] point to the possibility of extracting phonetically relevant information from this particular time-domain signal attribute. Temporal frequency, (sometimes referred to as instantaneous frequency), is defined via an analytic signal representation [9]. The analytic signal is a complex-valued function of time defined as,

$$a(t) = s(t) + j\tilde{s}(t)$$

where $j = \sqrt{-1}$, s denotes the real speech pressure waveform, and $\tilde{s}$ is the Hilbert transform of s. Manipulation of a(t) allows the unique separation of the speech signal into time-domain envelope and phase. Instantaneous envelope, e(t), is defined as,

$$e(t) = \mathrm{mod}[a(t)] = \sqrt{(s^2(t) + \tilde{s}^2(t))}$$

Instantaneous phase, $\phi(t)$, is given by,

$$\phi(t) = \arg\big[a(t)\big] = \arctan\big[s(t)/\tilde{s}(t)\big]$$

Temporal frequency (in radians) is simply the time derivative of instantaneous phase, i.e.,

$$\omega(t) = d\phi(t)/dt$$

Note that instantaneous phase as defined above is modulo 2π , and shows discontinuities whenever it extends beyond $\pm\pi$. Therefore, prior to differentiation, a standard "unwrapping" algorithm was applied in order to extract the desired continuous phase function [10].

The analytic signal and the time-domain signal attributes derived from it can be understood in the following way. a(t) can be thought of as the path traced out in complex space by a vector whose length and rate of rotation vary as a function of time. e(t) decribes the temporal variations in the length of the vector, and can be regarded as a measure of the instantaneous strength of the speech signal. $\omega(t)$ describes the temporal variations in the vector's rate of rotation. This time-domain function can be used as a measure of speech signal continuity.

The temporal frequency characteristics of speech signals are illustrated in Figs. 1 - 4. Figure 1 shows a speech pressure waveform for the simple VCV token [ɑ:dɑ:]. The waveform was low-pass filtered, (cut-off frequency = 8.4 kHz), and digitized at a sampling frequency of 20 kHz, to a maximum amplitude resolution of 12 bits. Figure 2 shows the temporal frequency function of the signal depicted in Fig. 1. Figure 3 shows a speech pressure for the utterance [tu:zɪərəʊ] ("two zero"); same speaker and data acquistion conditions as in Fig. 1. Figure 4 is the temporal frequency function for the signal shown in Fig. 3. These figures show that the quasi-periodic and noisy regions of the speech waveforms associated with sonorant and non-sonorant segments respectively, are clearly delineated by marked changes in both the structure and mean value of the temporal frequency.

MEAN TEMPORAL FREQUENCY

Although there is no one-to-one correspondence between time-domain and Fourier-domain frequencies, mean temporal frequency can be related to the spectrum of the speech signal. This relationship, outlined by Vile [11] (see also [12]), can be made more general in order to apply to speech data segments of arbitrary duration.

Without loss of generality, a speech signal segment of duration T, centred at $t = \tau$, can be modeled as,

$$s(\tau;t) = \mathrm{Re}[e(t)\exp(j\phi(t))]w(T,\tau;t) \qquad (1)$$

where e(t) is a non-negative envelope function, $\phi(t)$ is a phase function, and ,

$$w(T,\tau;t) = \begin{cases} 1 & \tau - T/2 < t > \tau + T/2 \\ 0 & \text{otherwise} \end{cases}$$

The mean Fourier-domain frequency of the data segment, f_τ , can be defined as,

$$f_\tau = \frac{\int_0^\infty f|S(\tau;f)|^2 df}{\int_0^\infty |S(\tau;f)|^2 df} \qquad (2)$$

where $S(\tau;f)$ is the Fourier transform of $s(\tau;t)$. Since $s(\tau;t)$ is real, $|S(\tau;f)|$ is an even function. Consequently, the integration in equ. (2) ranges over the positive frequencies only, to ensure a non-zero f_τ value. Given that the analytic signal can be written as,

$$a(\tau;t) = s(\tau;t) + j(1/\pi t)\bigstar s(\tau;t)$$

where $\bigstar$ denotes the convolution operator, it follows that,

$$a(\tau;t) = 2[\delta(t)/2 + j/2\pi t]\bigstar s(\tau;t)$$

i.e.,

$$A(\tau;f) = 2H(f)S(\tau;f)$$

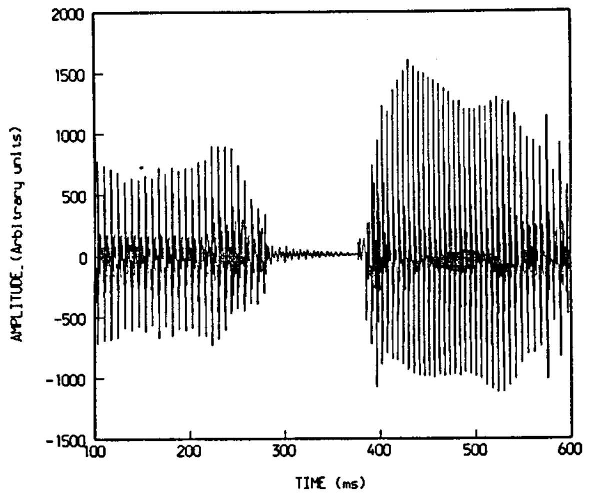
where δ is the Dirac-delta function, H is the Heaviside unit step function, and $A(\tau;f)$ is the Fourier transform of $a(\tau;t)$. Using this one-sided property of $A(\tau;f)$, equ. (2) can be re-written as,

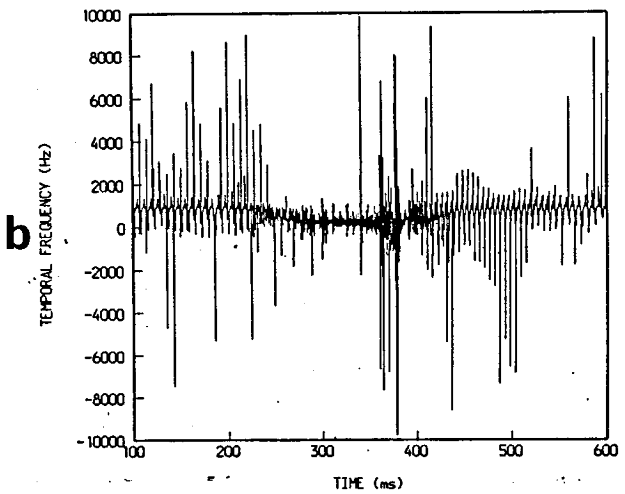
$$f_\tau = \frac{\int_{-\infty}^{\infty} f|A(\tau;f)|^2 df}{\int_{-\infty}^{\infty} |A(\tau;f)|^2 df}$$

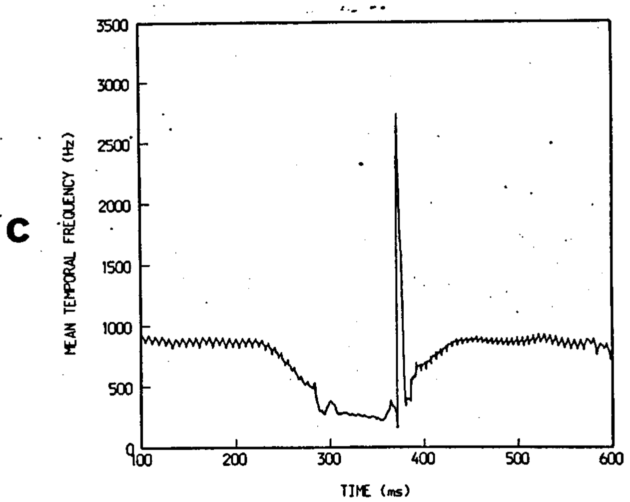
$$= \frac{\int_{-\infty}^{\infty} fA(\tau;f)A^*(\tau;f)df}{\int_{-\infty}^{\infty} |A(\tau;f)|^2 df}$$

where * denotes complex conjugate. Using the derivative theorem, and expressing $A(\tau;f)$ as a Fourier integral gives,

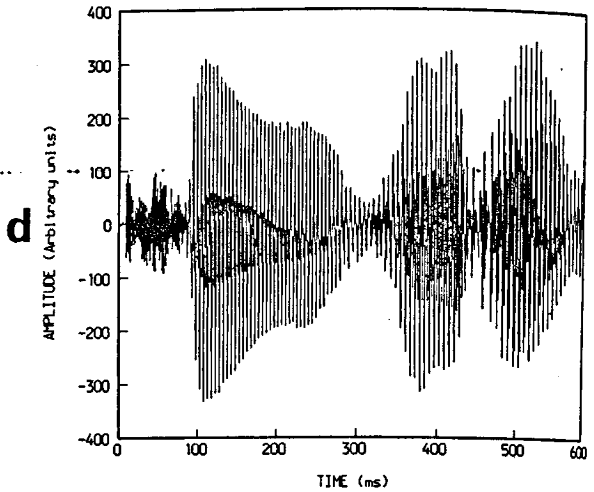
$$f_\tau = \frac{\int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' \dot{a}(\tau;t')a^*(\tau;t)\exp[2\pi jf(t-t')]}{2\pi j \int_{-\infty}^{\infty} |A(\tau;f)|^2 df}$$

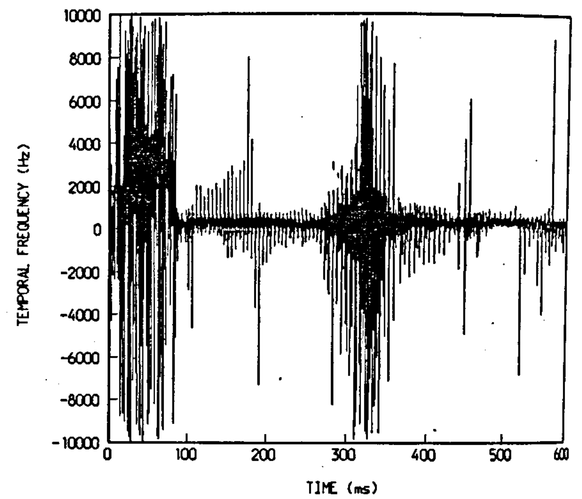
a


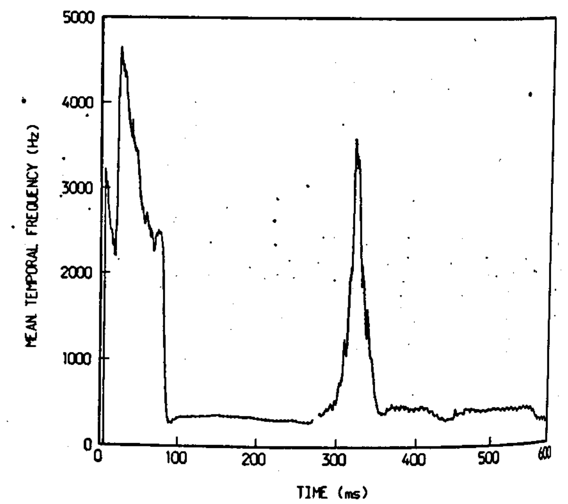
b


c


a – Fig. 1 Speech signal for [a:da:]
b – Fig. 2 Temporal frequency function for signal in Fig. 1
c – Fig. 5 Mean temporal frequency function for signal in Fig. 1

d


e


f


d – Fig. 3 Speech signal for [tu:zɪərəʊ] ("two zero")
e – Fig. 4 Temporal frequency function for signal in Fig. 3
f – Fig. 6 Mean temporal frequency function for signal in Fig. 3

where ˙ denotes time derivative. From above it follows that,

$$f_\tau = \frac{\int_{-\infty}^{\infty} dt\,\dot{a}(\tau;t')a^*(\tau;t)}{2\pi j \int_{-\infty}^{\infty} |A(\tau;f)|^2 df}$$

Using the signal representation given in equ. (1),

$$f_\tau = \frac{\int_{\tau-T/2}^{\tau+T/2} [e(t)\dot{e}(t) + e^2(t)\dot{w}(t) + j\dot{\phi}(t)e^2(t)]dt}{2\pi j \int_{\tau-T/2}^{\tau+T/2} |A(\tau;f)|^2 df}$$

Assuming $e(\tau+T/2) \simeq e(\tau-T/2)$, the above expression reduces to,

$$f_\tau = \frac{\int_{\tau-T/2}^{\tau+T/2} \dot{\phi}(t)e^2(t)dt}{2\pi \int_{\tau-T/2}^{\tau+T/2} |A(\tau;f)|^2 df}$$

Using Rayleigh's theorem,

$$\bar{f}_\tau = \frac{\int_{\tau-T/2}^{\tau+T/2} \dot{\phi}(t)e^2(t)dt}{2\pi \int_{\tau-T/2}^{\tau+T/2} |a(\tau;t)|^2 dt}$$

$$= \bar{\dot{\phi}}/2\pi = \overline{\omega_\tau}/2\pi$$

That is, for a speech signal segment of arbitrary duration, the centre of gravity of the power spectrum is equal to the envelope squared-weighted temporal frequency. Using the above expression, the time evolution of the mean temporal frequency for the signals shown in Figs. 1 and 3 was computed; the results are shown in Figs. 5 and 6 respectively. In both cases the window duration was 10 ms. Figures 5 and 6 show that plots of $\overline{\omega}(t)$ highlight the differences in the spectral composition of speech signal segments associated with sonorant and non-sonorant sounds. Note particularly the very clear delineation of the plosive release in Fig. 5.

DISCUSSION

Appropriate manipulation of speech signal temporal frequency enables the estimation of the time evolution of the centre of gravity of the signal's short-term power spectra, without the computational effort involved in moment calculation from the Fourier transforms of many short data segments. Initial results indicate that plots of $\overline{\omega}(t)$ may be useful in automatically segmenting speech waveforms; particularly in determining the presence of plosives. One other interesting aspect of this study is the presence of large, time-localized fluctuations in speech temporal frequency functions, (see Fig. 2 & 4). Work in ultrasonic signal processing has shown that an analysis of such features yields information which is of use both in imaging and in signal parameter estimation [6]. The possibility that speech signal temporal frequency structure encodes similarly useful information is being investigated.

REFERENCES

[1] L. R. Rabiner, R. W. Schafer, "Digital processing of speech signals", Prentice-Hall, 1978.
[2] B. C. J. Moore, "Frequency selectivity in hearing", Academic Press, 1986.
[3] M. T. Taner et al., "Complex trace analysis", Geophysics, vol. 44, pp. 1041-1066, 1979.
[4] R. L. Kirlin er al., "Enhancement of seismogram parameters using image processing techniques", Geoexploration, vol. 23, pp. 41-76, 1984.
[5] D. A. Seggie, S. Leeman, "Deterministic approach towards ultrasound speckle reduction", IEE Proc., vol. 134, Pt. A, no. 2, pp. 188-192, 1987.
[6] D. A. Seggie, S. Leeman, G.M. Doherty, "Time domain phase: a new tool in ultrasound imaging", Mathematics and Computer Science in Medical Imaging, Springer-Verlag, in press.
[7] C. Berthomier, "Instantaneous frequency and energy distribution of a signal Sig. Proc. vol. 5, pp. 32-45, 1983.
[8] D. A. Seggie, "The application of analytic signal analysis in speech processing", Proc. IOA, vol. 8, Pt. 7, pp. 85-92, 1986.
[9] D. Gabor, "Theory of communication", J. Inst. Elect. Eng., vol. 93, Pt. 1, pp. 429-441, 1946.
[10] A. V. Oppenheim, R. W. Schafer, "Digital signal processing", Prentice-Hall, pp. 507-509, 1975.
[11] J. Vile, "Theories et applications de la notion de signal analytique", Cables et Transmissions, vol. 1, pp. 61-74, 1948.
[12] L. Mandel, "Interpretation of instantaneous frequencies", Am. J. Phys., vol. 42, pp. 840-846, 1974.