ON THE EFFECT OF STRESS ON CERTAIN MODULATION PARAMETERS OF SPEECH [1]

K. Gopalan
Department of Engineering, Purdue University Calumet, Hammond, IN 46323

ABSTRACT

This paper reports the results of correlation between demodulated amplitude and frequency variations from the AM-FM speech model and the heart rate of a fighter aircraft flight controller. It has been found that the peak frequencies in the spectrum of the amplitude envelope of the model follow F0 regardless of the center frequency of analysis. This tracking of F0 gives a qualitative estimate of stress level – as measured by the heart rate – relative to a neutral state of stress, with no a priori knowledge of F0 or formants. Additionally, the mean values of F0 estimates at low and high heart rates increased significantly from those at neutral state. Formant tracking showed an increase in F3 at both high and low heart rates while F4 generally varied directly with heart rate.

1. INTRODUCTION

Analysis and detection of physiological stress at workplaces such as the cockpit of an aircraft is important in assessing the ability of the worker and assigning tasks accordingly. Stress detection based on speech enables monitoring of stress nonintrusively and, in general, without the cooperation of the speaker. Other applications include detection of deception in speech based on the emotional state of speakers, and implementation of natural-sounding speech synthesizers.

This report presents the results of analyzing the parameters of an AM-FM model of speech as a function of the heart rate of a fighter aircraft flight controller.

2. MODULATION BEHAVIOR IN SPEECH

Teager and Teager [1, 2] showed that speech resonances have frequency modulation (FM) causing instantaneous frequency variations and also amplitude modulation (AM) resulting in time-varying amplitudes. Based on this observation the many nonlinear and time-varying phenomena during speech production have been modeled successfully by an AM-FM modulation model representing each of the resonances or formants [3, 4].

In the case of a speaker under stress, increased activation of the sympathetic or the parasympathetic nervous system is observed to occur when the speaker is angry, fearful, sad, etc [5]. This increased activation leads to change in heart rate and blood pressure, and also to tremor in muscle activity. Consequently, the articulatory and respiratory movements for speech production are affected. The spectral characteristics of the resulting stressed speech are, in general, observed to have increased fundamental frequency F0, increased amplitude, and decreased speech duration. The extent of variation in F0, however, has been shown to depend on the speaker and the type of stress. Other manifestations of stress in speech include variations in the formants and their bandwidths, increase in high frequency energy, and changes in the glottal pulse shape.

In addition, studies have shown that in stressed speech, the fundamental frequency of excitation and formants have higher variations in their instantaneous values than in neutral speech. Therefore, an analysis of the instantaneous frequency variation around F0 and the formants is expected to bring out the frequency variations due to stress.

The AM-FM model represents speech around each formant F_k as a damped AM-FM signal given by [3, 6]

$$x(t) = \sum_{k=1}^{N} a_k(t)\cos\left(w_k t + \Phi_k(t)\right) \qquad (1)$$

where N represents the number of formants in the speech signal x(t). $a_k(t)$ is the time-varying amplitude of the sinusoid at the k^{th} formant frequency ω_k with the total instantaneous frequency of $w_{ki} = w_k + \frac{d\Phi_k}{dt}$. Both the instantaneous frequency (IF) and the amplitude envelope (AE) around a formant are derived from the Teager energy operator [1, 2]. Because of the unlimited number of combinations of AM and FM that can give rise to the modulated signal at each formant, the modulation model given in (1) is analyzed at each formant.

[1] Work supported by the U.S. Air Force Office of Scientific Research under the 1999 Summer Faculty Research Extension Program, Contract No. F49620-93-C-0063.

From the discrete-time version of a speech signal s(n) band-pass filtered at a center frequency $f_c = \omega_c/2\pi$, the instantaneous frequency $\omega_i = 2\pi f_i$, which varies about the analysis frequency f_c, and the amplitude envelope |a(n)| are calculated using the nonlinear energy tracking operator ψ(.) [1, 2]. From Kaiser's algorithm, these values are given by [7, 4 and 6]

$$y[s(n)] = s^2(n) - s(n-1)s(n+1) \quad (2)$$

$$w_i(n) \approx \arcsin\sqrt{\frac{y[s(n+1) - s(n-1)]}{4y[s(n)]}} \quad (3)$$

$$|a(n)| \approx \frac{2y[s(n)]}{\sqrt{y[s(n+1)] - y[s(n-1)]}} \quad (4)$$

A preliminary study of IF and AE for utterances from a fighter aircraft pilot in an aviation emergency showed that the peak frequencies in the spectrum of AE increased with stress [8]; additionally, the spectra of both IF and AE, in general, followed the fundamental frequency F0. The present work reports on the estimates of F0 and the bandwidth of IF around the center frequency Ω_c, and their correlation with stress.

3. ANALYSIS OF AM-FM FEATURES WITH HEART RATES

Speech files from the NATO SUSC-1 CD containing utterances from male European fighter aircraft flight controllers recorded during communication with aircraft were used in this study. The recording also includes heart rates of the speakers, which were chosen as indicators of stress. The utterance *"eye"* spoken at each of three different heart rates for a speaker was analyzed using the modulation model. Table I lists the range of the fundamental frequency F0 and the formants for speaker RD, which were estimated using Entropic Xwaves package.

Table I

Range of fundamental frequency and formants for speaker RD at three different heart rates

Freq.	At Low Heart Rate (72.27) (RD41)	At Medium Heart Rate (76.49) (RD21)	At High Heart Rate (77.29) (RD55)
F0, Hz	124 – 128	106 - 110	173 – 203
F1, Hz	581 – 779	622 - 773	616 - 711
F2, Hz	1274 – 1632	1282 – 1649	1411 – 1541
F3, Hz	2234 – 2533	2198 – 2490	2421 – 2498
F4, Hz	3331 – 3560	3470 - 3655	3568 - 3607
F5, Hz	3499 – 5051	4652 - 5915	4121 - 5030

Although the high heart rate of 77.29 corresponds to high values in the fundamental frequency, medium (76.49) and low (72.27) heart rates do not have proportionally lower values of F0. This indicates that the medium heart rate may correspond to "neutral" or unstressed state of the speaker.

Utterance at each heart rate was analyzed after bandpass filtering with a bandwidth of approximately 500 Hz. Demodulation was carried out at (a) an arbitrary frequency of f_c = 3000 Hz, which falls outside of any formant for the speaker RD, (b) in the vicinity of f_c = 2300 Hz ≈ F3, and (c) in the vicinity of f_c = 3500 Hz ≈ F4. Results of the demodulated parameters are discussed in the following section.

4. RESULTS AND DISCUSSION

At the arbitrary analysis frequency of f_c = 3000 Hz, the spectrum of the amplitude envelope corresponding to each heart rate showed harmonics at F0 as seen in Fig. 1. Values of F0 from the peaks are: 130 Hz at low heart rate, 110 Hz at medium and 180 Hz at high. These are consistent with the F0 values, and their variation relative to medium heart rate, as estimated using Entrpoic Xwaves (Table I). More significantly, the analysis indicates that F0 can be estimated for speech from the spectrum of the amplitude envelope by demodulating the utterance at an arbitrary frequency. This enables comparison of relative stress levels of a speaker from different utterances without a priori knowledge of F0 or the formants. Although the spectrum of the instantaneous frequency also revealed harmonics at F0, no strong peak was detected around F0.

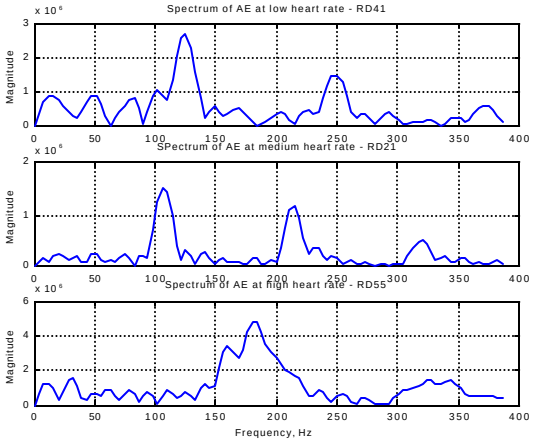


Fig. 1 Spectra of amplitude envelopes at different heart rates using arbitrary analysis frequency of f_c = 3000 Hz

Analyses at frequencies in the neighborhood of formants were used to estimate the formants. (IF from the analysis at an arbitrary frequency f_c locked in at f_c.) Unweighted and weighted estimates of the formants (F_u

and F_w) and their bandwidths (B_u and B_w) were calculated using [9]

$$F_u = \overline{f_i} = \frac{1}{T}\int_{t_0}^{t_0+T} f_i(t)dt \quad (5)$$

$$[B_u]^2 = \overline{(f_i - F_u)^2} \quad (6)$$

$$F_w = \frac{\overline{f_i a^2}}{\overline{a^2}} = \frac{\left(\frac{1}{T}\right)\int_{t_0}^{t_0+t} f_i(t)[a(t)]^2 dt}{msq(a)} \quad (7)$$

$$[B_w]^2 = \frac{\overline{[f_i - F_w]^2 a^2}}{\overline{a^2}} = \frac{\left(\frac{1}{T}\right)\int_{t_0}^{t_0+T} [f - F_w]^2 [a(t)]^2 dt}{msq(a)} \quad (8)$$

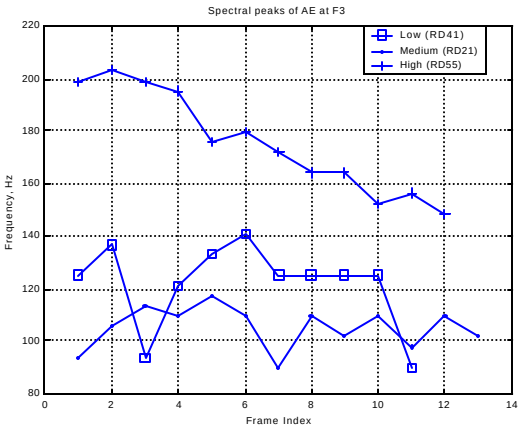
where msq(a) = mean-square value of |a|, and the small contribution of amplitude envelope |a| to weighted bandwidth is neglected.

Table II shows the average estimates of F0 (from the spectrum of AE), formants and bandwidths for the utterances at the three heart rates at the analysis frequency of f_c = 2300 Hz ≈ F3.

Table II
Average estimates of F0, formants and bandwidths at f_c = 2300 Hz ≈ F3

Heart Rate (Utterance)	F0, Hz	$F3_u$, Hz	$B3_u$, Hz	$F3_w$, Hz	$B3_w$, Hz
Low (RD41)	125	2346	96	2319	42
Medium (RD21)	105	2276	109	2269	43
High (RD55)	180	2350	106	2337	40

Clearly, all the parameters in the table show similar variations with heart rate – higher values at low and high heart rates relative to medium rate. Time variation of the modulation parameters were analyzed using 15 ms frames every 5 ms. Fig. 2 illustrates F0 variations for the utterances at the three heart rates and Fig. 3 shows the tracking of formant F3, both obtained at f_c = 2300 Hz ≈ F3. From the F0 profile the utterance at high heart rate is clearly distinguishable from those at low and medium rates. As for the lower values of F0 at medium heart rate compared to those at low, it appears that a deviation in heart rate – up or down – from "neutral" contributes to an increase in F0. The unweighted and weighted formants also indicate higher variations about the analysis frequency of 2300 Hz for utterances at heart rate below and above the "neutral" rate. Although the frame-to-frame weighted and unweighted bandwidths of the formant, shown in Fig. 4, do not appear to indicate any clear variation, the average bandwidth in each case increased with heart rate: unweighted average: 106 Hz (low), 110 Hz (medium) and 117 Hz (high); weighted: 49 Hz (low), 55 Hz (medium) and 69 Hz (high). Thus the estimate of F3 for an utterance at neutral or unstressed state at an analysis frequency in the vicinity of F3 appears to determine the relative stress level of other utterances qualitatively, while the average bandwidth in each case confirms the stress level.



.

Fig. 2 Tracking of F0 from the spectra of amplitude envelopes at f_c = 2300 Hz ≈ F3

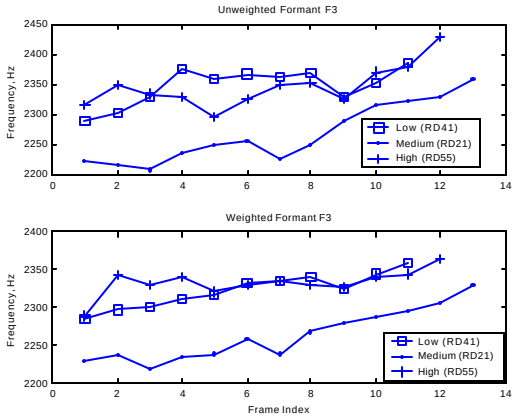


Fig.3 Unweighted and weighted formant tracking at f_c = 2300 Hz ≈ F3

At the analysis frequency of 3500 Hz, which is close to formant F4, F0 tracking (Fig. 5) again shows higher values at heart rates above and below medium rate – as observed at f_c = 2300 Hz. Average estimates of F0 were also comparable to those obtained at f_c = 3000 Hz and 2300 Hz. Formant tracking (Fig. 6), however, shows values increasing with heart rate. Average bandwidth – both weighted and unweighted – decreased with increasing heart rate.

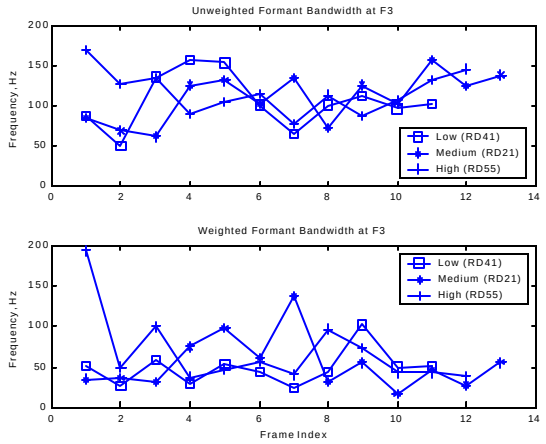


Fig. 4. Unweighted and weighted formant bandwidth at F3

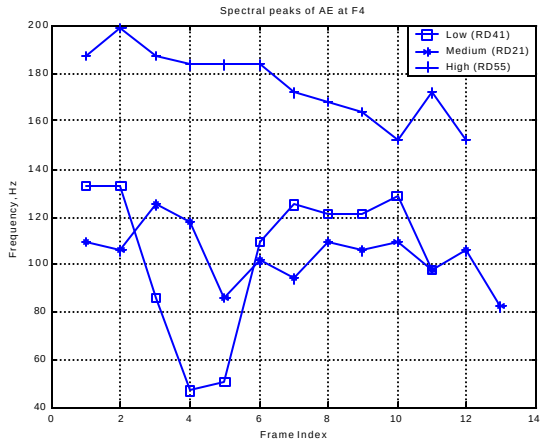


Fig. 5 Tracking of F0 from the spectra of amplitude envelopes at f_c = 3500 Hz ≈ F4

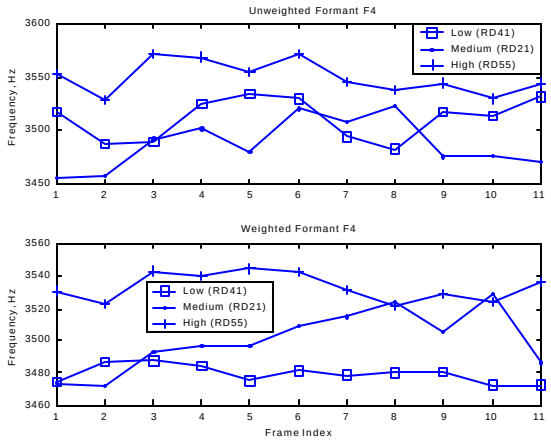


Fig.6 Unweighted and weighted formant tracking at f_c = 3500 Hz ≈ F4

5. CONCLUSION

A method of qualitative detection of stress based on the AM-FM model of speech has been presented. Although not all the model parameters changed in direct proportion to heart rate of the speaker, their variations relative to unstressed state, or medium heart rate, indicate presence of stress. While F0 as a measure of stress can be evaluated at any analysis frequency, more analyses at different formants are needed to confirm the observed correlation between heart rate and formants.

References

1. Teager, H.M., and S. Teager, "Evidence for Nonlinear Production Mechanisms in the Vocal Tract," NATO Advanced Study Inst. On Speech Production and Speech Modeling, Bonas, France, 1989, Kluwer Acad. Pub., 1990.
2. H.M. Teager, "Some Observations on Oral Air Flow during Phonation," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-28, No. 5, pp 599-601, Oct. 1980.
3. P. Maragos, T.F. Quatieri, and J.F. Kaiser, "Speech Nonlinearities, Modulations, and Energy Operators," Proc. ICASSP '91, pp. 421-424, 1991.
4. A.B. Fineberg, R.J. Mammone and J.L. Flanagan, "Application of the Modulation Model to Speech Recognition," Proc. ICASSP '92, pp. I-541-I-544, 1992.
5. C.E. Williams, and K.N. Stevens, "Vocal Correlates of Emotional Stress," in Speech Evaluation Psychiatry, J.K Darby, Jr. (Ed.), Grune & Stratton, Inc.,1981.
6. P. Maragos, J.F. Kaiser and T.F. Quatieri, "On Amplitude and Frequency Demodulations using Energy Operators," IEEE Trans. Signal Processing, Vol. 41, No. 4, pp. 1532-1550, Apr. 1993.
7. J.F. Kaiser, "On a Simple Algorithm to Calculate the 'Energy' of a Signal, Proc. ICASSP '90, pp. 381-384, 1990.
8. K. Gopalan, "Amplitude and Frequency Modulation Characteristics of Stressed Speech," Final Report, AFOSR Summer Faculty Research Program, Bolling AFB, July 1998.
9. L. Cohen and C. Lee, "Instantaneous bandwidth," in *Time-Frequency Signal Analysis – Methods and Applications* (B. Boashash, Ed.), London: Longman-Chesire, 1992.