DRAFT COPY ONLY: DO NOT QUOTE

Measuring vocal emotion using electroglottography.

Tom Johnstone, Tanja Banziger, & Klaus Scherer

Abstract

This experiment was designed to test the feasibility of using electroglottography (EGG) to measure changes in vocal fold function during emotional speech production. Speakers were recorded pronouncing standard phrases while expressing a number of different emotions. In addition to acoustic recordings, EGG recordings were made. The temporal characteristics of the EGG signal across different emotions were compared to f0 and spectral measures derived from the acoustic recordings. The correlations between the EGG parameters and f0 parameters indicated that EGG measurements might provide useful information on the glottal and laryngeal mechanisms responsible for emotional changes to speech, although more precision in measuring EGG would be required to draw conclusions about spectral acoustic characteristics.

Introduction

This experiment was motivated largely by previous findings that average spectral energy distribution, as indicated by the relative proportion of energy under 1000 Hz, differs significantly as a function of experimental manipulations of emotion (Johnstone, van Reekum and Scherer, in prep.). This dependence of spectral energy distribution on emotional response is consistent with similar findings in studies of acted emotional speech (Banse and Scherer, 1996). The effects of emotional response on spectral energy distribution might be due in part to differences in the configuration of the vocal tract, as suggested by Scherer (1986). For example, changes to precision of articulation and therefore the amplitude and bandwidth of formants, restriction or contraction of the faucal arches and changes to the amount of mucous and saliva in the vocal tract will all have an impact on the resonance of vocal harmonics.

It is also possible, however, that emotional changes to spectral energy distribution are caused by changes to vocal fold vibration patterns. Much work on voice source modeling and estimation has pointed to the relationship between the manner in which the vocal folds open and close and the dynamics of glottal air flow (e.g. Titze, 1994; Sundberg, 1994). The glottal air flow in turn strongly determines the relative strength of vocal harmonics feeding into the vocal tract (e.g. see Fant, 1993). For example, when vocal folds open and close abruptly, the result is proportionately more power in higher harmonics, and thus proportionately more high frequency energy in the speech signal (all other resonant characteristics being equal) than when the vocal folds open and close slowly and smoothly. Such variations in vocal fold function, categorised into a range of laryngeal settings, have been described phonetically by Laver (1980, 1991). They make up part of a broader system for describing and categorising laryngeal and articulatory voice quality settings, although such settings have more often been the focus of research on pathological changes to voice quality rather than more subtle vocal changes. Scherer

(1986) also makes reference to appraisal-mediated changes from “tense” to “relaxed” vocal settings, although the link between vocal settings, vocal fold dynamics and spectral aspects of the resulting acoustic signal are not made explicit. The two possible sources – vocal folds and vocal tract - of spectral changes to emotional speech, make the results from experiments one and two difficult to interpret. A method of independently measuring the emotion-provoked changes to vocal tract resonance and vocal fold dynamics is necessary to resolve such ambiguity.

Electroglottography (EGG) is a technique that can be applied to better understand the link between vocal fold dynamics and the spectral characteristics of emotional speech. With EGG, a small, high frequency electric current is passed between two surface electrodes placed on either side of the speaker’s neck, at the level of the larynx. Since the electrical impedance of the path between the two electrodes will change as the glottis opens and closes, a measurement of the impedance can be used as an indicator of glottal opening and closing.

Figure 5.1 shows the acoustic and EGG signals from recording of an extended [a] vowel (the “a” sound in the word “father”). As can be clearly seen, the EGG signal is free of resonance from the vocal tract, which is apparent in the acoustic signal as rapid fluctuations occurring in between successive vocal excitations. In fact, aside from its application to measuring vocal fold dynamics, the EGG signal allows for an extremely precise measurement of F0 and F0 perturbation.

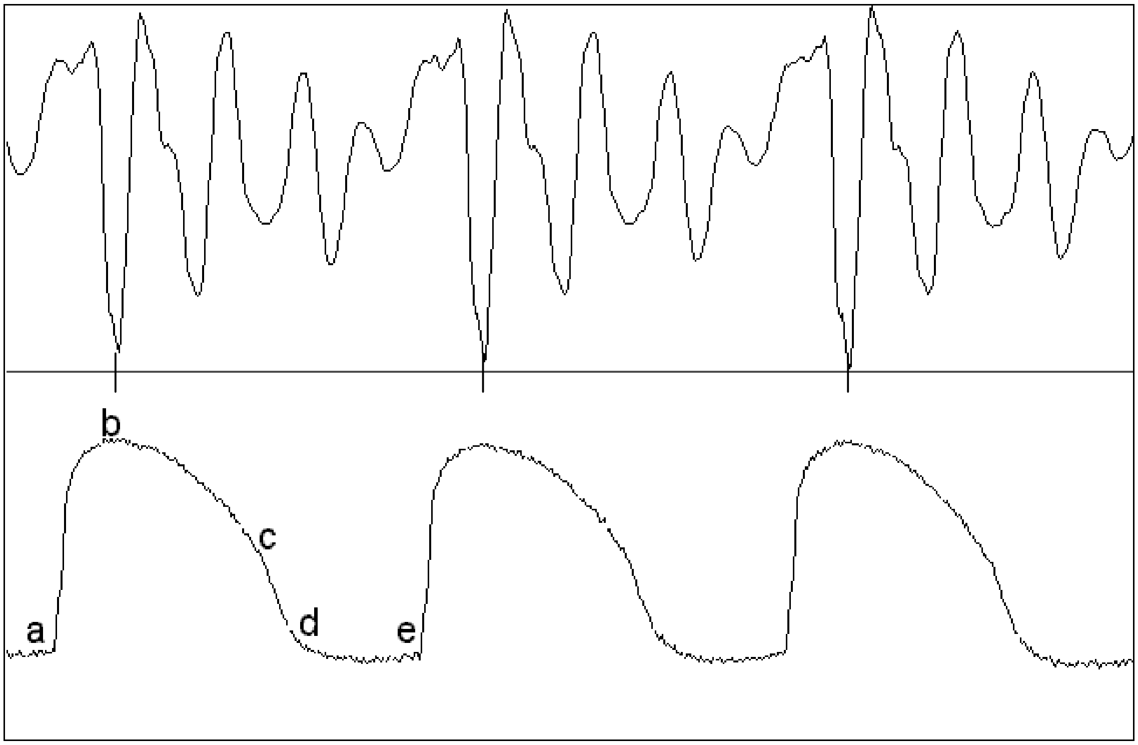

Figure 5.1. Acoustic (top) and EGG (bottom) signals from recording of an extended [a] vowel. Marks on the horizontal axis indicate fundamental periods. Symbols on the EGG signal indicate the instant of glottal closure (a), instant of maximum vocal fold contact (b), instant of glottal opening (c), instant of maximum minimum vocal fold contact (d) and instant of glottal closure for subsequent glottal period (e).

Although the exact relationship between the size of the glottal opening and the electrical impedance is not linear, the EGG signal is inversely proportional to the contact area of the vocal folds (Childers & Krishnamurthy, 1985), and as such serves as a useful indicator of vocal function. It is at least reasonable to assume that the temporal characteristics of the measured EGG signal are good indicators of the temporal aspects of glottal opening and closing. The magnitude of the EGG is a less valid measure of the relative area of the glottal opening, although in the absence of alternative estimates (other than invasive measurements such as photoglottography) it remains the best approximate indicator.

Glottal analysis of imagined emotions.

This experiment was carried out to develop a set of analysis procedures for extracting the main features from the EGG signal, as well as a test of the feasibility and applicability of EGG analysis to the study of emotional voice production. In particular, it was intended to apply such analysis to the final experiment of this thesis, in which computer tasks and games would be used to induce emotional speech. To be assured of high quality speech recordings, which were known to vary across different emotional states, an imagination/acting procedure was used. Speakers were asked to imagine themselves in specific emotional states and then to pronounce aloud two short phrases and the sustained [a] vowel. Rather than recording expressions of extreme emotions such as rage, elation and fear, seven non-extreme emotions (tense, neutral, happy, irritated, depressed, bored, anxious) corresponding to those that could realistically be induced with computer tasks and games, were selected for study.

Method

Participants

Speakers were eight research workers and postgraduate students (two males and six females) studying emotion psychology at the University of Geneva. The familiarity of the speakers with emotion psychology was seen as an advantage in this pilot study, as they were more readily able to follow the emotion imagination procedure and produce emotional (albeit possibly stereotypical) vocal samples.

Equipment

Speech was recorded with a Casio DAT recorder using a Sony AD-38 clip-on microphone to one channel of the DAT. EGG recordings were made with a Portable Laryngograph onto the second DAT channel.

Procedure

Speakers were seated in front of a computer display at a comfortable viewing distance. They were told that the experiment consisted of being presented with a number of emotion words on the computer screen. When each emotion word was presented, they were to try to imagine themselves in a situation that would provoke that emotion. They were asked to imagine the situation as vividly as possible and try to actually feel the emotion. When they felt that they were involved in the imagined situation as much as possible, they were to click the mouse, whereupon they would be presented with a list of phrases on the screen. Speakers were instructed to read aloud all the phrases on the screen, while still trying to feel the imagined emotion. The order of the seven emotions (tense, neutral, happy, irritated, depressed, bored, and anxious) was randomised across speakers. For each emotion, each speaker was asked to pronounce the phrases "Je ne peux pas le croire!" ("I can't believe it!"), "En ce moment, je me sens…<emotion>" ("at the moment, I feel…<emotion>"; speaker completed the phrase with the appropriate emotion term), and the extended [a] vowel. Each phrase was pronounced twice by each speaker in each emotion condition.

Results

Acoustic and glottal analyses

The samples from each subject were acoustically analysed using LabSpeech. In contrast to the F0 measures made in the previous experiments, which were based upon an autocorrelation analysis of the acoustic speech signal, in this experiment F0 was calculated on the basis of the low-pass filtered EGG signal. First the EGG waveform was high-pass filtered to remove low frequency fluctuations due primarily to larynx movement. Next, a peak-picking algorithm was applied to the differentiated EGG waveform to locate the instants of glottal closure, which are typically the points of maximum positive gradient in each glottal period. Those sections of the EGG signal displaying successive glottal closures corresponding to a frequency within the preset permitted F0 range were designated voiced, and the F0 values saved. All other segments were labeled non-voiced. RMS energy of voiced segments and the mean proportion of energy under 1000 Hertz for voiced segments were calculated from the acoustic speech signal as in experiment two.

Glottal opening, closing, open and closed times were calculated from the EGG signal following guidelines given by Marusek (1997). Each glottal period, as demarcated by successive instances of glottal closure, was analysed to identify closing, closed, opening, and open times, as depicted in figure 5.2. An amplitude criterion of 90% of total period amplitude was used to determine the onset and offset of closed phase. The onset of glottal open phase was located using the equal level criterion as described by Marasek (1997, section 12), which is the point at which the glottal signal drops to a level equal to the level at the instant of glottal closure. These values were then converted to quotients by dividing each by the total glottal period. The use of quotients, rather than absolute times, was preferred since quotients give a better measure of the shape, rather than the length, of the glottal EGG waveform. Such shape parameters were hypothesised to correspond to spectral acoustic measures such as the proportion of energy under 1000 Hertz. In addition, because quotients are normalised to the length of each glottal period, they can meaningfully be averaged across glottal cycles and compared between conditions.

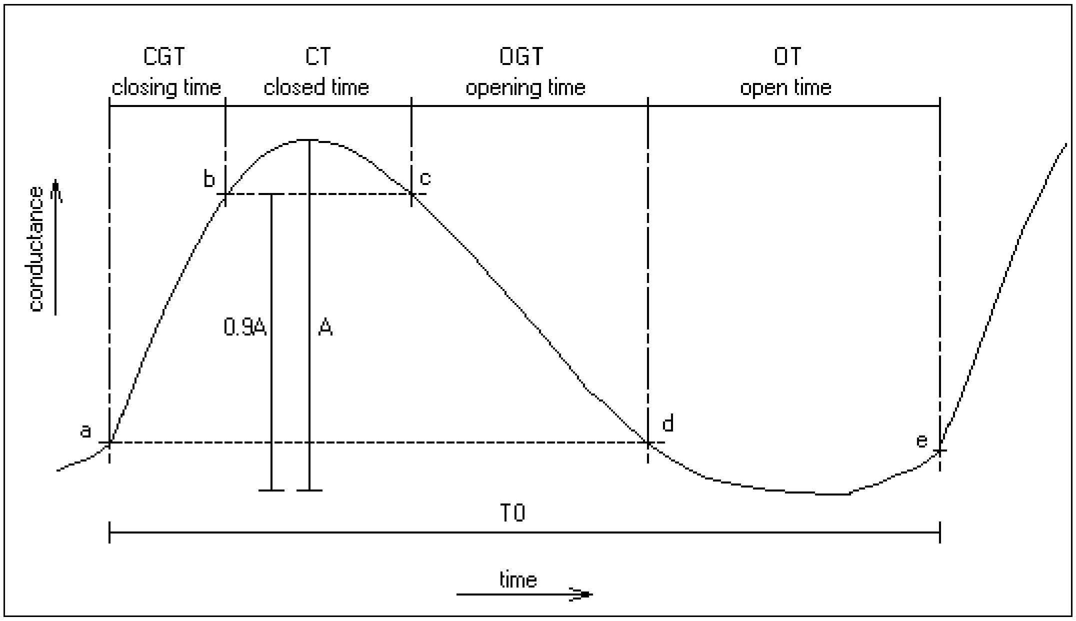

Figure 5.2. Stylised example of a glottal cycle measured with EGG, showing how the cycle is divided into four segments, based upon the point of maximum curvature (a), points of positive-going (b) and negative-going (c) 90% amplitude threshold, point of equal value to preceding point of maximum curvature (d) and point of maximum curvature of following glottal cycle (e). (A = amplitude, T0 = glottal period).

Very low frequency EGG energy was also calculated as a potential measure of larynx movement. The low frequency EGG energy is manifest as a slow waveform upon which the faster fluctuations of glottal cycles are superimposed. In this experiment, low frequency EGG energy was isolated by using a low pass filter with a frequency cut-off of 10 Hz. The RMS power of the resulting waveform was then calculated.

A measure of jitter, the short-term, random period to period variation in F0, was also calculated. For the jitter calculation, a quadratic curve was fitted to a running window of five successive F0 values on the F0 contour using a least mean squares curve-fitting algorithm. The quadratic curve was then subtracted from that section of the F0 contour. This served to remove long term, prosodic F0 movements, which would otherwise contaminate jitter measurements. Jitter was then calculated as the mean of the magnitude of period to period variation in the residual F0 values.

Statistical analysis

All acoustic and glottal parameters were tested with univariate mixed effect ANOVA's, with emotion and phrase as fixed factors and speaker as a random factor. The results are presented below organised by vocal parameter.

Median F0. Median F0 differed significantly across emotions (F(6,78)=17.5, p<0.000), and was categorised by high values for happy speech and low values for depressed and bored speech. The middle curve in Figure 5.2 shows the emotion profile for median F0.

F0 ceiling. The emotion profile for F0 ceiling is shown in Figure 5.3, top curve. The F0 ceiling varied significantly across emotions (F(6,78)=11.3, p<0.000), and was characterised by a higher value for happy speech than for the other expressed emotions.

F0 Floor. F0 floor varied significantly with emotion (F(6,78)=7.3, p<0.000) in much the same way as median F0, with relatively high values of F0 floor for happy speech, and low values for depressed and bored speech (see Figure 5.3, bottom curve).

Voiced RMS energy. The RMS energy of voiced segments of speech was significantly different across emotions (F(6,78)=16.5, p<0.000), as shown in Figure 5.4. In particular, voiced energy was high for happy and irritated speech and low for depressed speech.

Jitter. Jitter values varied significantly with emotion (F(6,54)=2.9, p=0.013), with jitter values being highest for bored and depressed speech and lowest for tense speech (see Figure 5.5).

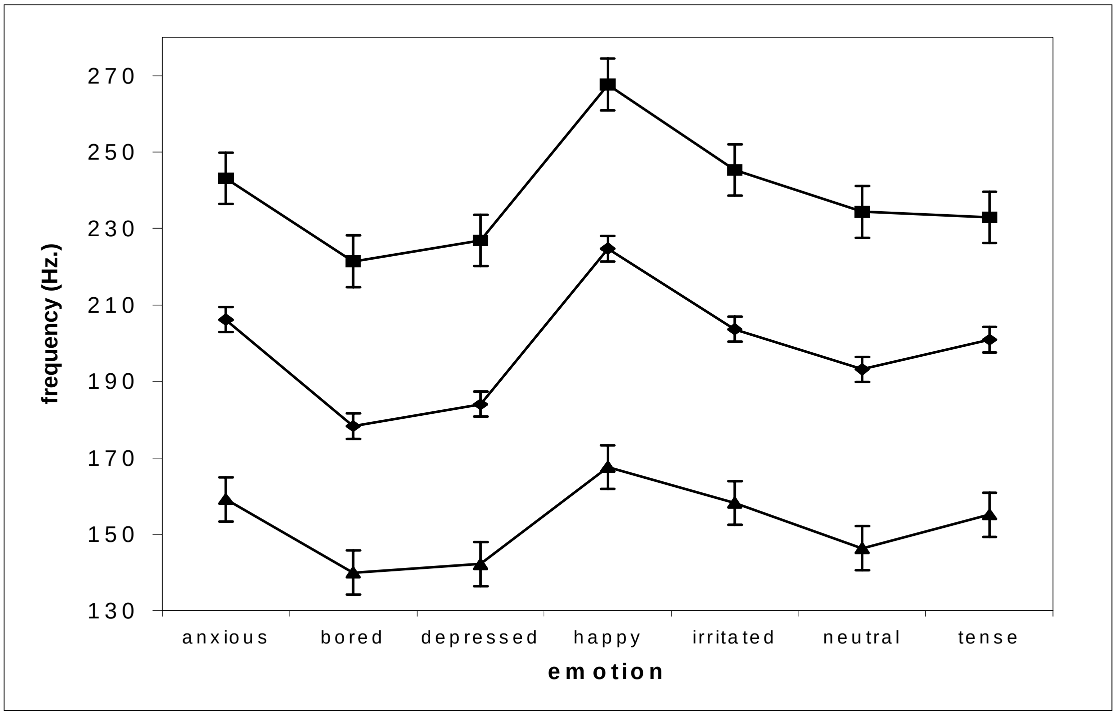

Figure 5.3. Mean values for F0 ceiling (top), median F0 (centre) and F0 floor (bottom), shown as a function of emotion. Bars represent 95% within-subjects confidence intervals.

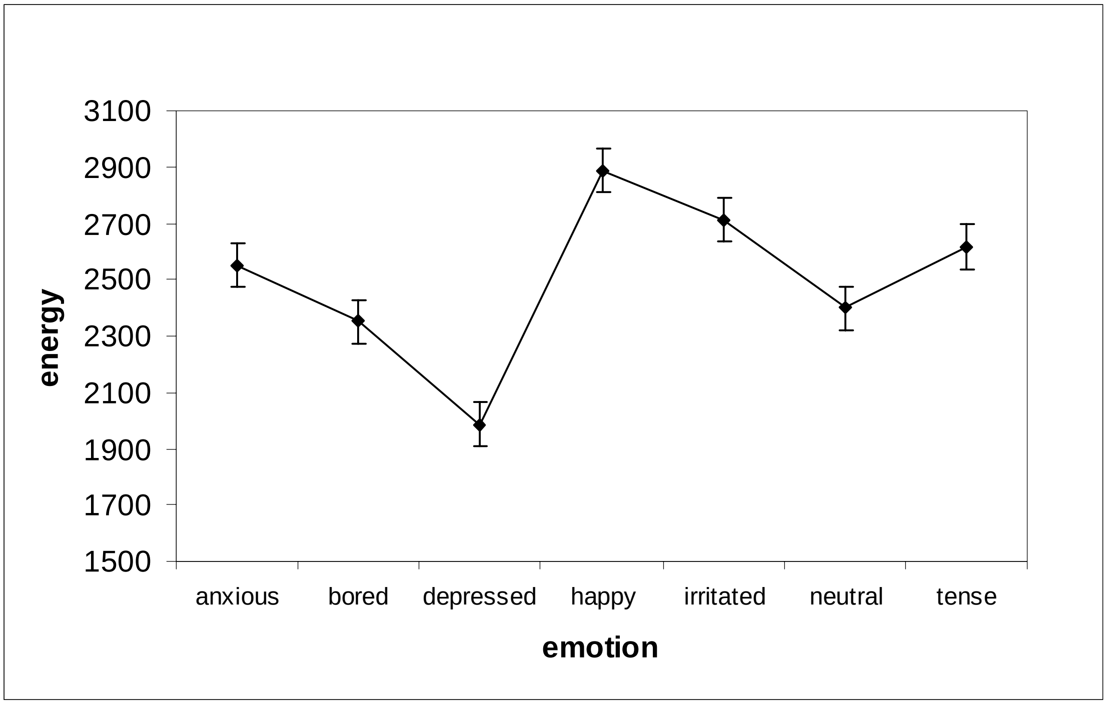

Figure 5.4. Mean values for RMS Energy, shown as a function of emotion. Bars represent 95% within-subjects confidence intervals.

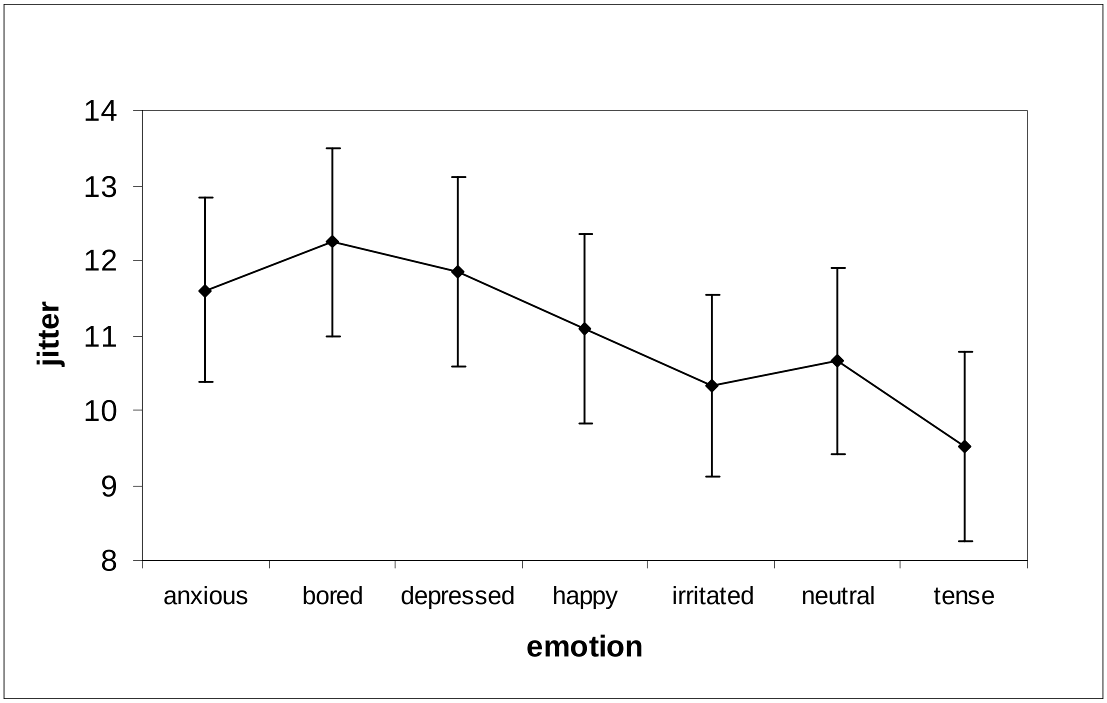

Figure 5.5. Mean values for jitter, shown as a function of emotion. Bars represent 95% within-subjects confidence intervals.

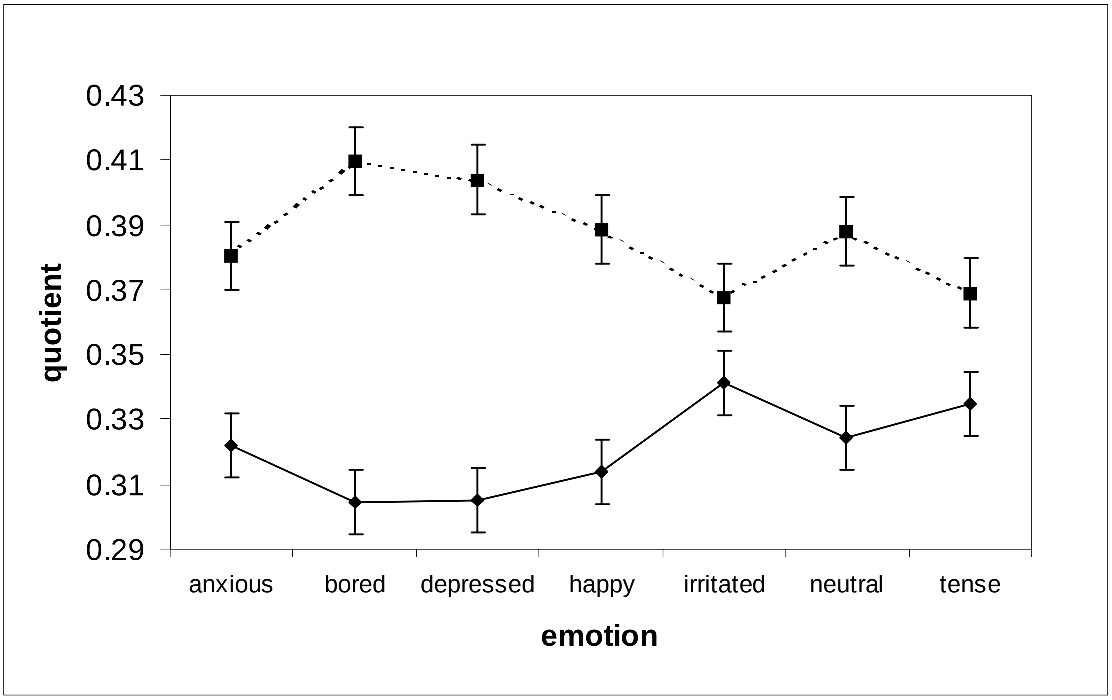

Figure 5.6. Mean values for open quotient (top) and opening quotient (bottom), shown as a function of emotion. Bars represent 95% within-subjects confidence intervals.

Opening quotient. The glottal opening quotient varied significantly as a function of emotion (F(6,78)=3.3, p=0.007), with high opening quotients for irritated and tense speech, and low opening quotients for bored and depressed speech (figure 5.6).

Open Quotient. Glottal open quotient varied significantly across emotions (F(6,78)=3.2, p=0.008), showing the inverse pattern of results from glottal opening quotient, with high values for bored and depressed speech, and low values for irritated and tense speech (figure 5.6).

Closing quotient. The glottal closing quotient did not vary significantly across expressed emotions (F(6,78)=1.2, p=0.31).

Closed Quotient. Glottal closed quotient did not vary significantly across emotions (F(6,78)<1).

Low frequency EGG power. Low frequency EGG power varied across emotions (F(6,78)=11.9, p<0.000), with high values for happy speech and low values for bored and depressed speech (figure 5.7).

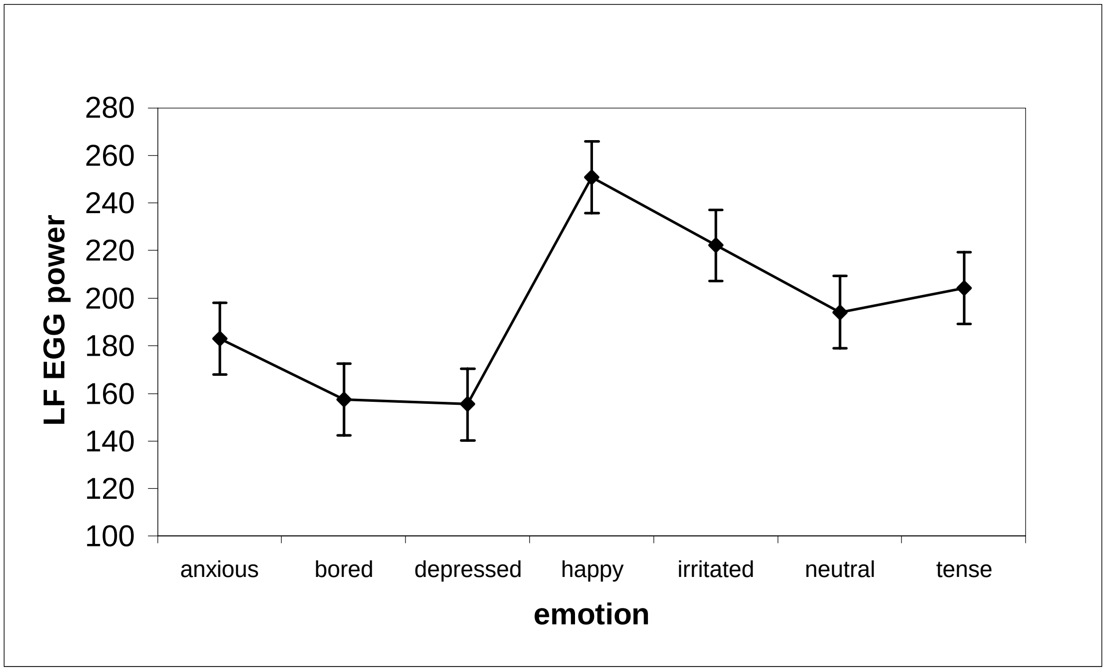

Figure 5.7. Mean values for low frequency EGG power, shown as a function of emotion. Bars represent 95% within-subjects confidence intervals.

Voiced low frequency energy. The proportion of total energy below 1000 Hertz for voiced segments of speech varied significantly according to the emotion expressed (F(6,78)=9.9, p<0.000), as did the proportion of energy under 500 Hz (F(6,78)=28.4, p<0.000). Figure 5.8 indicates that there was relatively more low frequency energy for expressed depression and boredom than for irritation and happiness.

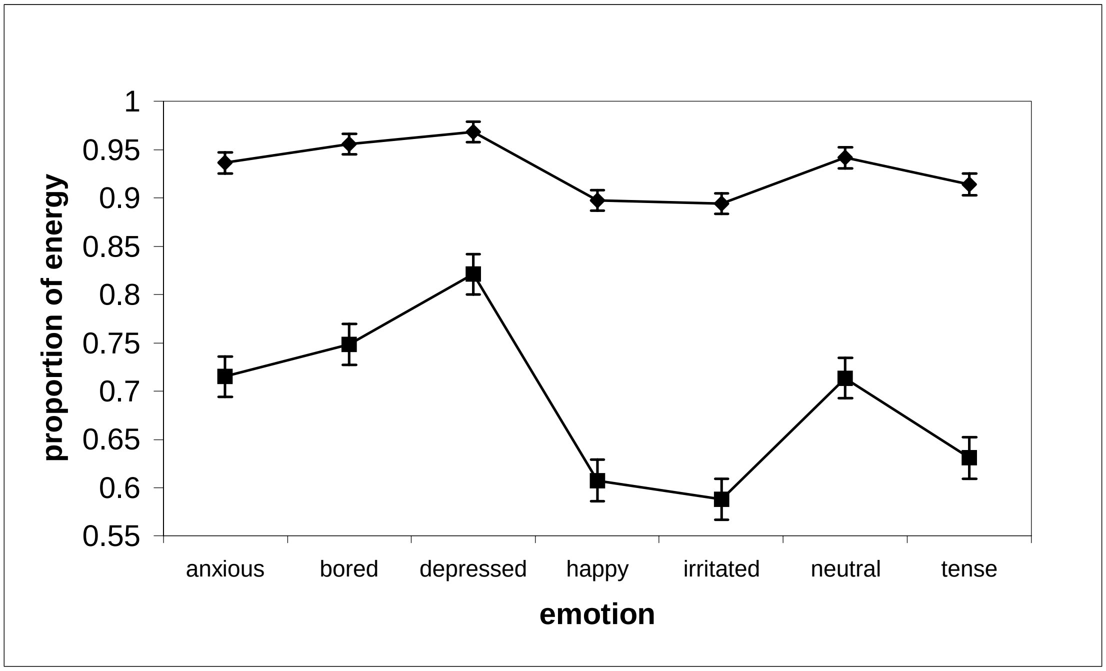

Figure 5.8. Mean values for the proportion of energy under 1000 Hz (top line) and proportion of energy under 500 Hz (bottom line), shown as a function of emotion. Bars represent 95% within-subjects confidence intervals.

Correlations between parameters. Table 5.1 provides the correlations between the different vocal parameters, after the effects of speaker, stimulus and repetition have been factored out. Many of the correlations are moderate, indicating that although the parameters might share some underlying mechanism, or to a small extent measure some common vocal feature, they capture different aspects of vocal production. Of particular note are the correlations between those glottal parameters and acoustic parameters that differed significantly across emotions.

Table 5.1. Correlations between vocal parameter residuals after effects of speaker, phrase and repetition have been factored out.

	F0 ceiling	F0 floor	Voiced energy	jitter	Opening quotient	Closing quotient	Open quotient	Closed quotient	LF EGG	Energy < 1 Khz	Energy < 500 Hz
F0 median	0.63	0.57	0.49	-0.15	0.24	0.10	-0.37	0.37	0.32	-0.29	-0.39
F0 ceiling		0.28	0.32	0.15	0.00	0.15	-0.11	0.17	0.30	-0.23	-0.31
F0 floor			0.47	-0.36	0.34	-0.08	-0.40	0.35	0.10	-0.28	-0.31
Voiced energy				-0.16	0.25	-0.01	-0.29	0.18	0.16	-0.18	-0.35
jitter					-0.30	0.22	0.24	-0.17	0.01	0.13	0.23
Opening quotient						-0.36	-0.90	0.37	0.04	-0.19	-0.29
Closing quotient							0.01	-0.18	0.20	0.04	0.04
Open quotient								-0.57	-0.13	0.20	0.31
Closed quotient									0.06	-0.10	-0.18
LF EGG										-0.16	-0.21
Energy < 1 KHz											0.65

Opening quotient correlated positively with F0 floor, F0 median and energy, but negatively with jitter and low frequency spectral energy. Open quotient was highly negatively correlated with opening quotient and thus showed the opposite pattern of correlations. Low frequency EGG energy correlated positively with median F0 and F0 ceiling but not with F0 floor. Low frequency spectral energy correlated negatively with the F0 measures.

Discussion

Variation of acoustic measures across emotions.

As with most prior research on acted expressions of emotional speech, F0 floor and median F0 were found to be lowest for the emotions bored and depressed, and highest for happy speech. A similar profile across emotions was found for F0 ceiling. Although these results would seem consistent with the idea that arousal causes changes to muscle tension, which in turn affects F0, the results for energy only partly support such an explanation. Energy was high for happy speech and low for depressed speech, thus corresponding to an arousal explanation for the F0 values, but bored speech did not have low energy. The correlation of 0.47 between energy and F0 floor indicates a high level of correspondence between the two measures, but given the result for bored speech, it is evident that these two parameters do not necessarily covary.

Low frequency spectral energy also varied across expressed emotions, indicating that there was more low frequency energy in bored and depressed speech, and less in happy and irritated speech.

Variation of EGG measures across emotions and with acoustic measures.

Of the EGG parameters, opening quotient correlated positively and open quotient negatively with F0 floor and F0 median. The pattern of these two EGG values across emotions was also similar to those of F0 floor and median F0, indicating that emotional changes to F0 level were principally mediated by changes to the time during which the glottis was open. This can be concluded because as the vocal cycle becomes shorter (i.e. as F0 rises) the open quotient falls but the opening quotient rises. Thus in absolute terms (i.e. in milliseconds rather than as a proportion of the vocal period), the open time has fallen but the opening time has remained relatively constant. An exception was the vocal expression of happiness, in which neither opening quotient nor open quotient were particularly low or high. For happiness then, it seems that an elevated F0 level was caused by relatively equal reductions in both absolute opening and open times.

F0 ceiling (but not F0 floor) was also found to correlate positively with low frequency EGG power, which was measured in this experiment as an approximate indicator of larynx movement. As has been reported previously (Iwarsson and Sundberg, 1998), movement of the larynx is one mechanism that can be used in combination with respiratory changes to vary F0 (indeed, beginner singers have to be trained in how to vary their singing pitch *without* moving their larynx). It is thus possible in this case that larynx movement was a mediator of increased F0 variation in happy speech.

The values for jitter, which correlated negatively with F0 floor and opening quotient, were highest for bored and depressed speech and lowest for tense speech. This result is in agreement with previous findings of a reduction of jitter for speakers under stress (Smith, 1977), and hints that emotional effects on jitter might be due to variation in the opening phase of the glottal cycle.

Low frequency energy correlated positively with open quotient and negatively with opening quotient, although the correlations were only moderate, indicating that spectral energy depends partly on glottal dynamics, but might also be affected by other factors including vocal tract resonance.

Conclusions

In this experiment, EGG and acoustic recordings of expressed emotional speech were analysed in an attempt to test the feasibility of using EGG techniques in studies of emotional speech. The EGG parameters open quotient, opening quotient and low frequency EGG energy were found to vary significantly across expressed emotions in ways consistent with the acoustic parameters measured as well as past results and theory of emotional speech production. Correlations between the EGG parameters and acoustic parameters indicated that EGG measurements might provide useful information on the glottal and laryngeal mechanisms responsible for emotional changes to speech. On the basis of these results it is possible to make tentative hypotheses relating laryngeal movement, as measured by low frequency EGG components, to F0 variability, and the length of glottal open phase to F0 level. Such measurements thus provide us with the potential to separate the mechanisms responsible for emotional changes to F0 level and to F0 variation. Although the links between spectral aspects of the speech signal and EGG parameters were not clear from this experiment, it is probable that techniques that better quantify the shape of the EGG pulses will lead to clearer results.