Electroglottogram Analysis of Emotionally Styled Phonation

Peter J. Murphy[1] and Anne-Maria Laukkanen[2]

[1] Department of Electronic and Computer Engineering,
University of Limerick, Limerick, Ireland
peter.murphy@ul.ie
[2] Department of Speech Communication and Voice Research,
University of Tampere, Tampere, Finland
anne-maria.laukkanen@uta.fi

Abstract. Acoustic descriptors of emotional voice quality have traditionally been provided in terms of fundamental frequency (f0), amplitude and duration. These cues have formed the basis for numerous studies of the emotional content of voice signals. The present work examines additional factors that arise as a result of the underlying production mechanism. Specifically glottal characteristics of different emotionally styled voice types are examined using the electroglottogram signal. This preliminary study, using a single speaker, investigates a number of electroglottogram measures relating to closed quotient (closed time of the glottis/cycle length) and rate of contact at closure and opening for the sustained vowel /a/, produced when simulating the emotional states sad, tender, neutral, joy and anger. The results suggest that differences exist in terms of glottal attributes for a particular emotion.

1 Introduction

The study of emotion as expressed in speech has been examined in terms of fundamental frequency (f0), syllable duration and waveform intensity in a number of studies; a review of the literature is provided in [1]. Speaker-independent automatic recognition of emotion, using four or five emotions, has been reported to be in the range of 50%-60% accuracy [2]. A more fine-grained feature extraction, such as the analysis of glottal characteristics, may allow for an improved recognition rate. Relatively few studies have focussed on glottal attributes of emotional expression [3-8]. In the present study electroglottography is used to examine emotional content in the sustained vowel /a/.

2 Electroglottography

The electroglottogram (EGG) consists of two electrodes placed external to the larynx. A high frequency current is passed between the electrodes and the output signal varies depending on the impedance of the material between the electrodes. As the vocal folds vibrate they move through high impedance (glottis open) to low impedance (glottis closed) values. As the impedance decreases with contact the electroglottogram signal provides a measure of vocal fold contact [9,10] (Fig.1 – top row). The electroglottogram provides complimentary information to the glottal flow; the maximum

A. Esposito et al. (Eds.): Multimodal Signals, LNAI 5398, pp. 264–270, 2009.

in the electroglottogram occurs when contact is maximum while the maximum in the glottal flow occurs during the open phase. An important aspect of the EGG signal is that it is essentially free of supra-glottal acoustic influences, which can produce source–filter interaction, making glottal flow determination more difficult.

2.1 Measures Taken from the EGG Waveform

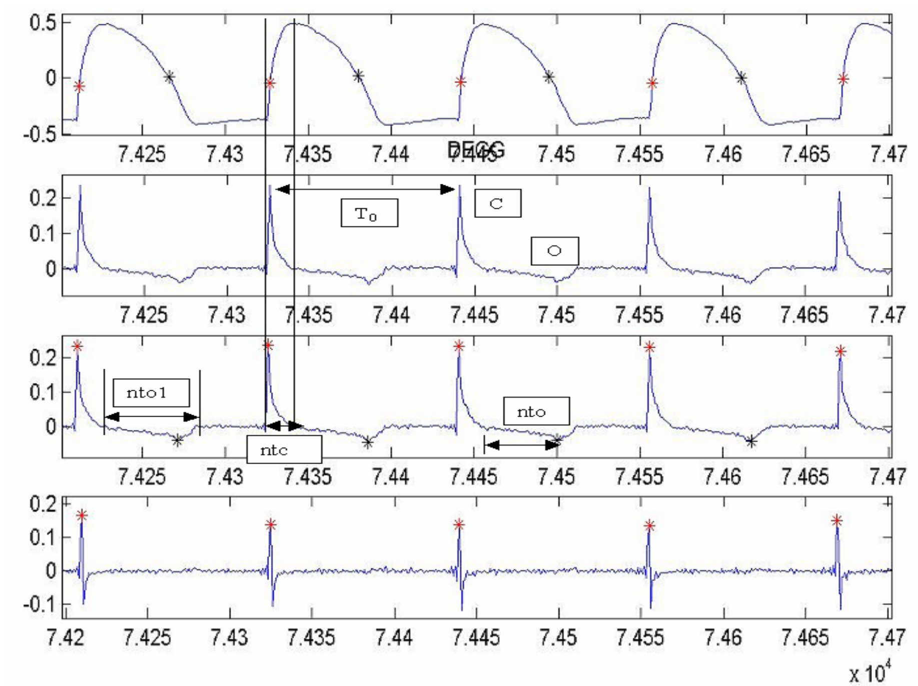

Fig. 1. EGG analysis (neutral phonation) (x-axis – sample number, y-axis amplitude - arbitrary units) top row EGG with (nearest sample to) zero crossing points indicated (negative to positive – red asterisk, positive to negative – black asterisk), 2nd row DEGG, 3rd row DEGG peaks at closure (red asterisk) and opening (black asterisk), 4th row D2EGG peak (red asterisk). Measure symbols (nto1 etc. are described in Table 1).

The following measures were extracted from the EGG signal.

Table 1. Measures estimated from the electroglottogram (EGG), first derivative of the electroglottogram (DEGG) and the second derivative of the electroglottogram (D2EGG) signals

Measure Symbol	**Description**	**Method of Measurement**
T_0	glottal cycle length	measured between points of glottal closure as determined from the positive peaks in the DEGG signal
f0	fundamental frequency	$1/T_0$
CQ	closed quotient – closed time of the glottis divided by cycle duration (T)	the closed time is measured from the positive peak to the negative peak in a glottal cycle of the DEGG signal

Table 1. (*Continued*)

CtQ	contacting quotient – contact time of the vocal folds divided by cycle duration	the contact time is measured from the beginning to end of contact
Vc	max. velocity of contact	max. positive peak amplitude occurring in a glottal cycle of the DEGG signal
Vo	max. negative velocity at opening	max. negative peak amplitude occurring in a glottal cycle of the DEGG signal
D2	acceleration of contact	max. peak in a glottal cycle cycle of the second derivative of EGG
SQ	speed quotient	Vc/Vo - velocity of closure divided by velocity of opening
ntc	normalised closing time - closing time divided by cycle length	closing time is measured from the beginning of contact – one sample point before the peak in D2EGG to the max point in the EGG signal
nto	normalised opening time – approximate opening time divided by cycle length	measured from peak in EGG to minimum in DEGG
nto1	normalised opening time one	opening time (including the final phase of opening after the minimum in the DEGG signal) divided by cycle length

3 Analysis

The EGG and speech signals were recorded in a sound treated room in the Department of Speech Communication and Voice Research, University of Tampere, Tampere, Finland. The sustained vowel /a/ was phonated while simulating the emotions angry, joy, neutral, sad and tender by a single speaker who was experienced in phonation types and emotional portrayals in voice. One hundred and thirty eight cycles of a

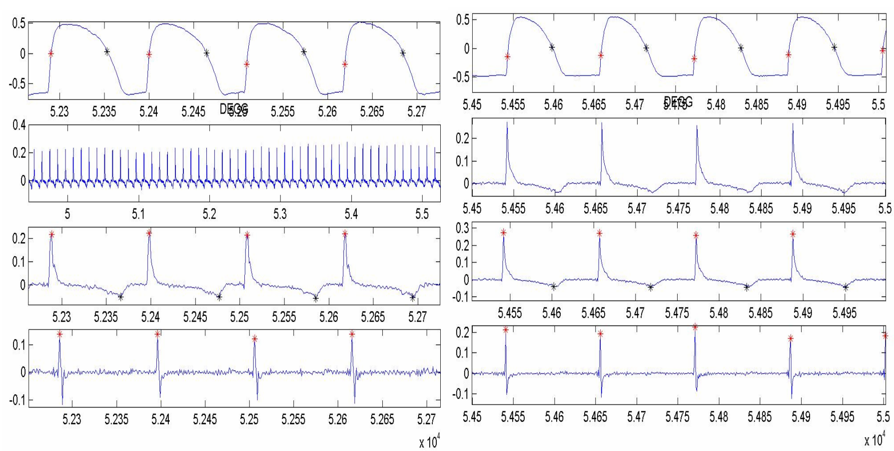

Fig. 2. EGG segment of emotions anger (left) and joy (right) (x-axis – sample number, y-axis amplitude - arbitrary units) top row EGG with zero crossing points indicated (negative to positive – red asterisk, positive to negative – black asterisk), 2nd row – first derivative of EGG (DEGG), 3rd row DEGG peaks at closure (red asterisk) and opening (black asterisk), 4th row – second derivative of EGG (D2EGG) peak (red asterisk)

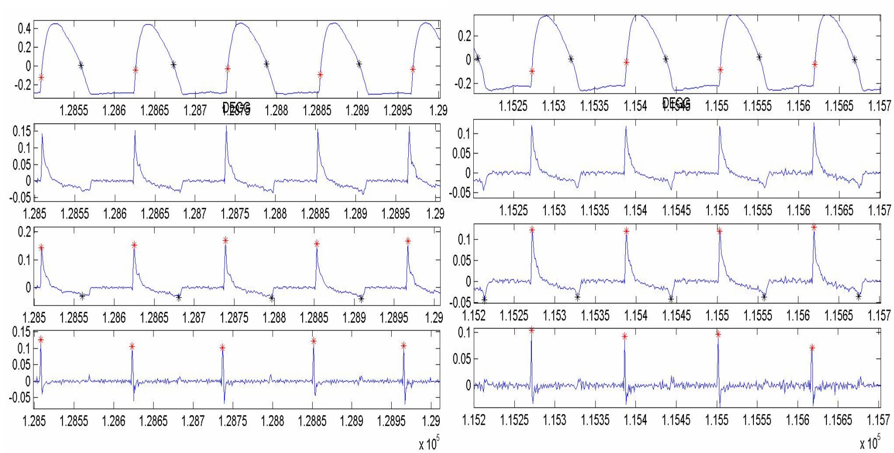

Fig. 3. EGG segment of emotions sad (left) and tender (right) (x-axis – sample number, y-axis amplitude - arbitrary units) top row EGG with zero crossing points indicated (negative to positive – red asterisk, positive to negative – black asterisk), 2nd row DEGG, 3rd row DEGG peaks at closure (red asterisk) and opening (black asterisk), 4th row D2EGG peak (red asterisk)

steady portion of the vowel were selected for analysis. The five emotions chosen have been analysed previously [3,4] and reflect positive and negative valence, and high and low psycho-physiological activity levels. Sadness and tenderness have low activity levels, while joy and anger have a high level of activity. Sadness and anger have negative valence while joy and tenderness represent positive valence. Fig.2 shows the electroglottogram and it's first and second derivatives for anger (left) and joy (right) while Fig.3 shows the same information for sad (left) and tender (right).

4 Results

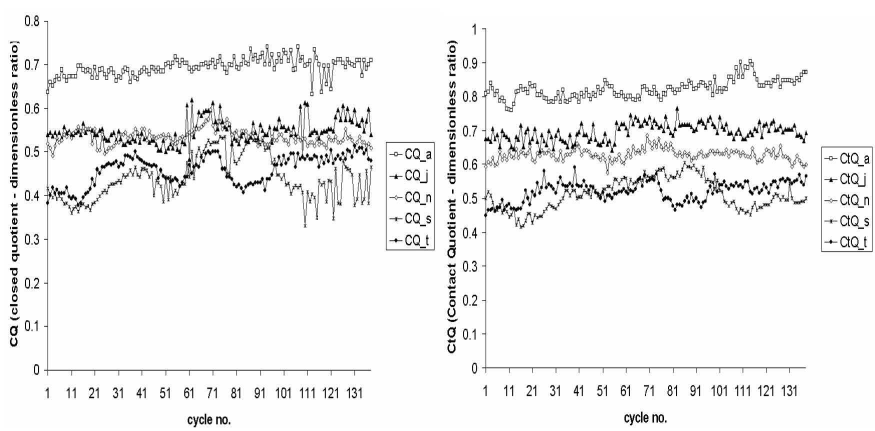

Fig. 4. (left) CQ (closed quotient) versus cycle number and right CtQ (contacting quotient) versus cycle number for the emotions anger (a), joy (j), neural (n), sad (s) and tender (t)

Fig.4 (left) shows closed quotient (CQ) and (right) contacting quotient (CtQ), versus cycle number. CtQ shows greater separability of the emotions. CQ is noticeably higher for the emotion anger.

Fig.5 (left) shows Vc (velocity of contact at closure) and (right) Vo (velocity of contact at opening) versus cycle number. Both indices show good data separability. Vc is highest for joy and lowest for tender, while Vo is highest for anger and lowest for tender.

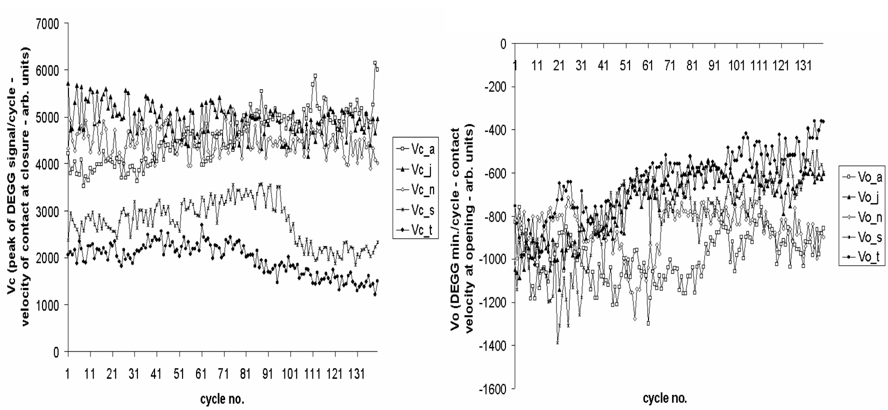

Fig. 5. (left) Vc (velocity of contact at closure) versus cycle number and right Vo (velocity of contact at opening) versus cycle number for the emotions anger (a), joy (j), neural (n), sad (s) and tender (t)

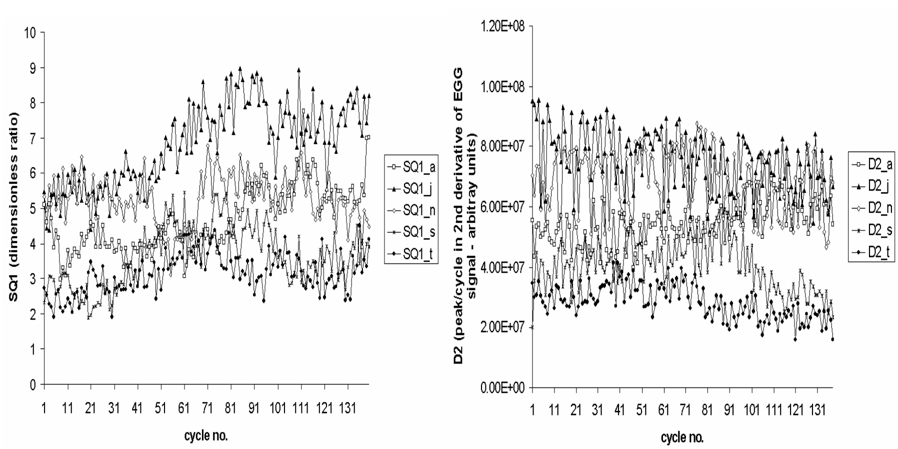

Fig. 6. (left) SQ (speed quotient-) versus cycle number and right D2 (acceleration of contact at closure) versus cycle number for the emotions anger (a), joy (j), neural (n), sad (s) and tender (t)

Fig.6 shows SQ (speed quotient) versus cycle number on the left and D2 (acceleration of contact at closure) versus cycle number on the right. Both indices show good discrimination between emotion simulations. SQ and D2 are highest for joy and lowest for tender. Each index is higher in neutral phonation than for angry phonation.

Table 2. Electroglottogram based measures averaged over 138 cycles

EGG Measure/ Emotion	f0	CQ	CtQ	Vc	Vo	SQ	D2	ntc	nto	nto1
Anger	180	**0.69**	0.82	**4547.6**	995.0	**4.68**	5.49E+07	**0.20**	0.51	**0.62**
Joy	178	**0.55**	0.70	**4940.6**	747.0	**6.84**	7.63E+07	**0.17**	0.40	**0.53**
Neutral	177	**0.53**	0.63	**4483.4**	860.0	**5.28**	6.88E+07	**0.14**	0.40	**0.48**
Sad	175	**0.44**	0.51	**2772.3**	796.0	**3.64**	4.04E+07	**0.17**	0.28	**0.34**
Tender	176	**0.46**	0.52	**1974.5**	648.0	**3.14**	2.85E+07	**0.15**	0.32	**0.37**

Table 2 summarised the average data for all measures. Normalised closing time and the normalised opening times are greatest for anger.

5 Discussion

Several electroglottogram measures show significant discriminatory ability. The results are discussed in terms of the underlying oscillatory mechanism. Contacting quotient (CtQ) is ordered as angry > joy > neutral > sad > tender. This suggests a more pressed phonatory mode for anger and a more lax configuration for sad and tender as expected. Joy is better separated from neutral using contacting quotient as opposed to contact quotient. This is because the final phase of opening (nto1-nto) is greater for joy. Speed quotient and acceleration of contact at closure are ordered as joy > neutral > angry > sad > tender. It is postulated that these values are less for anger because the more adducted glottal configuration does not allow for the vocal folds to develop to their full velocity when compared to a state of zero adduction. Normalised closing quotient did not provide clear separation of the emotions and was highest for anger. A previous study [3] examining the same emotions as those reported presently has shown that a measure of normalised glottal flow closing quotient did not provide statistically significant separation of neutral and anger simulations. It should be noted that EGG and glottal flow measures of closing quotient are not directly comparable.

6 Conclusion

Several features extracted from the EGG signal show promise for the automatic recognition of emotion e.g. contacting quotient, speed quotient and acceleration at closure distinguish between the different emotions. Future work will involve a more detailed statistical analysis with more speakers and will include a recognition task.

Acknowledgements

The recording and initial analysis development was performed during a COST 2103 (Advanced Voice Function Assessment) supported short-term scientific mission to the Department of Speech Communication and Voice Research, University of Tampere, Tampere, Finland in December, 2007.

References

1. Murray, I.R., Arnott, J.L.: Toward a simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. J. Acoust. Soc. Am. 93, 1097–1108 (1993)
2. McGilloway, S., Cowie, R., Douglas-Cowie, E., Gielen, S., Westerdijk, M., Stroeve, S.: Approaching automatic recognition of emotion from voice: a rough benchmark. In: Proceedings of the ISCA workshop on Speech and Emotion (Belfast), pp. 207–212 (2000)
3. Airas, M., Alku, P.: Emotions in vowel segments of continuous speech: analysis of the glottal flow using the normalized amplitude quotient. Phonetica 63, 26–46 (2006)
4. Toivanen, J., Waaramaa, T., Alku, P., Laukkanen, A.-M., Seppänen, T., Väyrynen, E., Airas, M.: Emotions in [a]: A perceptual and acoustic study. Logopedics Phoniatrics Vocology 31, 43–48 (2006)
5. Gobl, C., Ní Chasaide, A.: The role of voice quality in communicating emotion, mood and attitude. Speech Communication 40, 189–212 (2003)
6. Johnstone, T., Scherer, K.: The effects of emotion on voice quality. In: Intl. Conf. Phonetic Sci. (1998)
7. Laukkanen, A.-M., Vilkman, E., Alku, P., Oksanen, H.: Physical variations related to stress and emotional state: a preliminary study. J. Phonetics 24, 313–335 (1996)
8. Cummings, K.E., Clements, M.A.: Analysis of the glottal excitation of emotionally styled and stressed speech. J. Acoust. Soc. Am. 98, 88–98 (1995)
9. Baken, R.J.: Electroglottography. J. Voice 6, 98–110 (1992)
10. Fourcin, A., Abberton, E., Miller, D., Howells, D.: Laryngograph: Speech pattern element tools for therapy, training and assessment. Eur. J. Disord. Commun. 30, 101–115 (1995)