CORRELATES OF TEMPORAL HIGH-RESOLUTION FIRST FORMANT ANALYSIS AND GLOTTAL EXCITATION

Manfred Pützer and Wolfgang Wokurek†*

*Institut für Phonetik, Universität des Saarlandes, Saarbrücken, Deutschland
†Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Deutschland
puetzer@coli.uni-sb.de, wolfgang.wokurek@ims.uni-stuttgart.de

ABSTRACT

This preliminary study visualizes the glottal excitation in a temporally highly resolved estimate of the first formant. Instantaneous estimates of the frequency and bandwidth of the first formant closely follow the electroglottographic contour. This is demonstrated for modal, breathy, and hoarse phonation of an [aː] produced by one male and one female speaker. The temporally highly resolved formant contours show glottal features such as the different durations of the open phase and fundamental frequency and/or amplitude perturbations of the vocal fold vibration.

Keywords: linear prediction, electroglottography

1. INTRODUCTION

In voiced speech the larynx produces a complex sound that excites the resonances of the vocal tract. Interpreting this situation in terms of narrowband spectral analysis, the source signal consists of the fundamental oscillation and higher harmonics. The harmonics which are close to the resonances of the vocal tract get amplified. They convey information on the place and manner of articulation to the listener whereas the fine structure of the harmonic spectrum carries the voice quality. In this analysis the long window of the narrowband spectrum looks at the formants as slowly varying characteristics of the vocal tract. Spectral gradient measurements may be associated with voice quality to a certain extent [3], but the minimum vowel duration of about 50 milliseconds together with the necessary stationarity excludes many short vowels and diphthongs from analysis with this technique.

On the other hand it is evident that the vocal tract changes with every movement of the vocal folds in the degree of coupling to the subglottal cavity. Widely opened glottal folds couple the subglottal cavity and lower at least the lowest - the first - formant. The increased cavity surface introduces additional damping and causes an increased bandwidth of the first formant. In contrast the closed glottis acoustically decouples the subglottal cavity and leads to the highest frequency and the smallest bandwidth of the first formant. Immediately after the acoustic excitation by the completed glottal closure, the acoustic energy in the vocal tract is at its peak and this high-frequency low-bandwidth formant state is radiated most prominently.

The instantaneous frequency and bandwidth of the first formant is estimated by the high temporal resolution linear prediction described in [5].

This study presents contours of the first formant, which is determined most strongly by the supraglottal oral cavity together with the height of the dorsum.

2. SIGNAL ANALYSIS

The recorded speech signal was lowpass filtered at 5kHz since this study focuses on the first formant. A standard first-order difference filter was applied for preemphasis of the formants over the low frequency components of voice excitation. Frames of this signal were extracted with a rectangular analysis window of 2.5 to 5 ms duration and a stepsize of 0.2 ms. At least the short analysis window duration is even shorter than female pitch periods and justifies the term instantaneous formant analysis. On the other hand such short windows lead to strong noise in the formant estimates.

Depending on the location of the analysed frame in the pitch cycle, a certain part of this - now slowly varying - contour is present. Compensation of this trend is attempted with constant, linear, and quadratic regression.

An estimate of the autocorrelation sequence was transformed to the linear prediction polynomial by the Levinson-Durbin algorithm. The roots of the polynomial were extracted as the eigenvalues of the companion matrix.

To reduce the strong residual noise of the formant frequency estimates, a two step smoothing procedure was applied. First, a frequency range of the expected first formant was defined $[400, 800]$ kHz for the male and $[500, 1000]$ kHz for the female speech samples. Second, the mean and standard deviation of all 5 frequency estimates within the expectation

range were taken. This smoothing frame was shifted afterwards to the next frequency estimates within the expectation range.

3. EVALUATION

3.1. Material and Method

One male and one female speaker were asked to produce the vowel [aː] at a normal pitch with (a) modal, (b) breathy and (c) hoarse voice quality. Electroglottogram (EGG) and microphone signals were recorded simultaneously, and both were digitized with a sampling rate of 48kHz and with 16bit amplitude resolution. The microphone signal was recorded using a headset condenser microphone AKG C420. By using a head set microphone, the distance to the lips remains constant during speech independent of head movements [4]. The EGG signal was measured with a laryngograph model Lx Proc type PCLX from Laryngograph LTD. The signal was fed directly into a digital mixing console Yamaha 03D and stored in a computer via the digital input of the Midiman/M-Audio Delta DIO2496 soundcard.

The perceptual evaluation of each stimulus using the RBH-system ([1], [2]) provides the following (expected) classification: (a) Modal voice quality samples had a score of 0 (not present) for breathiness and hoarseness. (b) Breathy voice quality samples had a score of 3 (very) for breathiness. (c) Hoarse voice quality samples also had a score of 3 (very) for breathiness and hoarseness.

Initial tests rapidly settled most of the analysis parameters: the preemphasis zero at 0.99, the order of linear prediction at 49 and 5 point smoothing, which corresponds to 1 millisecond if every analysis frame yields a formant estimate in the range where we look for the first formant. The analysis window duration and the regression order {none, 0, 1, 2} remained flexible and were adjusted to the recorded vowel samples. The window duration was varied in the range of {125, 150, 200, 250} points corresponding to {2.5, 3, 4, 5} milliseconds. The regression was omitted or applied with the orders 0, 1, or 2. The figures in the paper were selected in order to display the electroglottographic contour most clearly and to have a minimum of defects like outliers, jumps or extra peaks.

3.2. Results

Three contours are shown for every speaker and voice quality: the electroglottographic measurement as a phonation reference, the instantaneous frequency estimate of the first formant (F1), and the bandwidth estimate of the first formant (B1). Note that the duration is adjusted to display about seven pitch cycles of both the male and the female recordings.

Since the glottal excitation travels to the microphone acoustically with the speed of sound and to the laryngograph electrically, all acoustic contours show a delay of about 2 milliseconds.

3.2.1. Modal voice quality

Figure 1: EGG Signal (male modal)

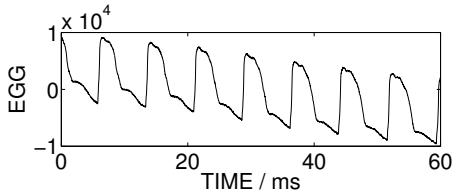
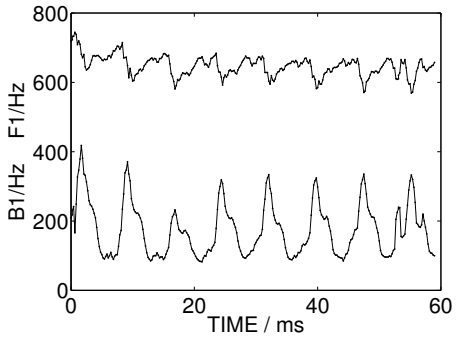


Figure 2: Instantaneous F1 and B1 (male modal)



Closing and contact phase: The beginning closing phase of each pitch cycle is displayed as an ascent of the EGG contour in figures 1 and 3. The ascent ends during the contact phase. The locally maximal contact is marked by the upper peak. The same upward and peak course is visible in the first formant in figures 2 and 4.

The bandwidth of the first formant is minimal during the closed phase, i.e. the peaks in EGG and F1 are aligned with a B1 valley in figures 2 and 4. The low bandwidth indicates a low loss of acoustic energy in this phase of the pitch cycle when the subglottal cavity is minimally coupled to the supraglottal vocal tract.

Opening and open phase: The beginning opening phase of the glottal cycle is characterised by a decreasing contact of the vocal fold tissue. The EGG contour, displaying the electrical conductivity across the larynx, falls and reaches its valley when the vocal folds are open (figures 1 and 3). Again, after the acoustic delay the frequency of the first formant

Figure 3: EGG Signal (female modal)

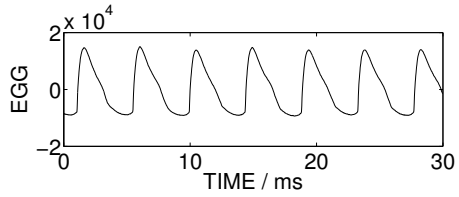
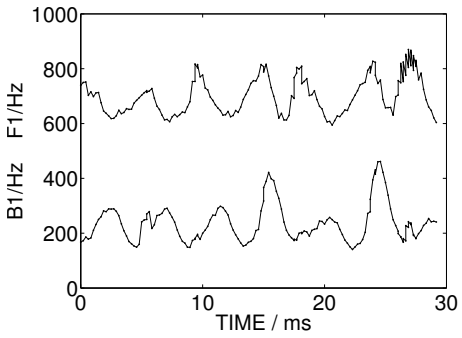


Figure 4: Inst. F1 and B1 (female modal)



in figures 2 and 4 decreases and reaches a valley. This is interpreted as a decreasing cavity resonance frequency due to an increasing acoustic coupling of the subglottal and supraglottal cavities. The first formant's frequency is minimal and its bandwidth is maximal during the open phase. The increasingly large bandwidth corresponds to an increasingly large loss of acoustic energy in the subglottal cavity.

3.2.2. Breathy voice quality

The trend of the F1 contour to follow the EGG contour is still visible in figures 5 and 6 for male and in figures 7 and 8 for female breathy voice recordings. Due to breathiness the open phase is longer in comparison to modal phonation. The longer open phase is displayed by EGG and F1 contours in figures 5-8 for both, the male and female voice samples. The average bandwidth is higher for female compared to male voices and for breathy compared to modal phonation.

Figure 5: EGG Signal (male breathy)

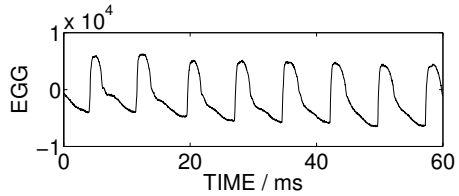


Figure 6: Instantaneous F1 and B1 (male breathy)

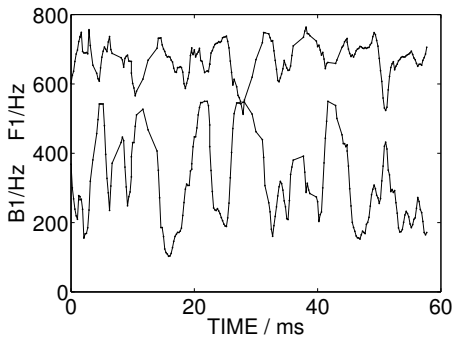


Figure 7: EGG Signal (female breathy)

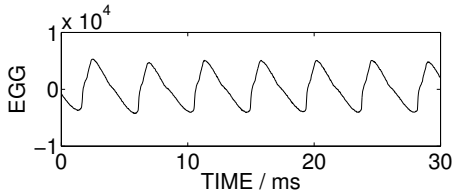
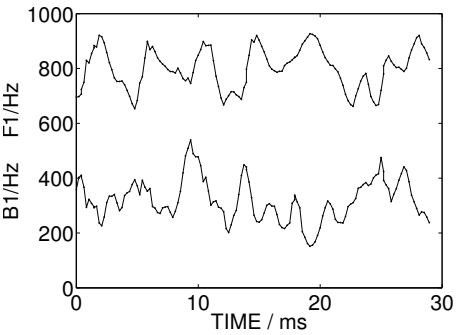


Figure 8: Inst. F1 and B1 (female breathy)



3.2.3. Hoarse voice quality

EGG and instantaneous formant measurements of hoarse voice quality are shown in figures 9 and 10 for the male and in figures 11 and 12 for the female recordings.

The pitch cycles of both speakers show strong fundamental frequency and amplitude perturbations. The EGG contours in figures 9 and 11 show a closed phase intermediate between modal and breathiness. The open phases show more variation from cycle to cycle. The F1 and B1 contours follow the EGG course less closely. In figure 10 the bandwidth contour skips some cycles which may be due to the too long analysis window of 3 milliseconds compared to

the pitch cycle of 4.5 milliseconds.

Figure 9: EGG Signal (male hoarse)

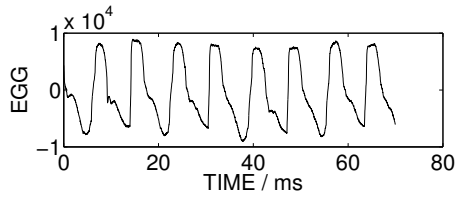


Figure 10: Inst. F1 and B1 (male hoarse)

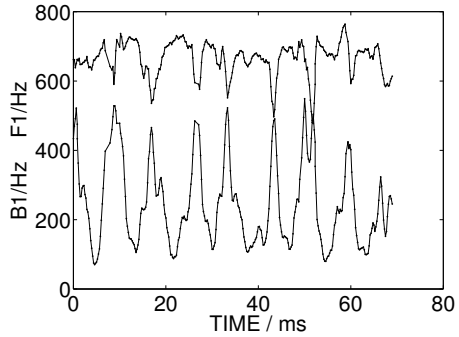


Figure 11: EGG Signal (female hoarse)

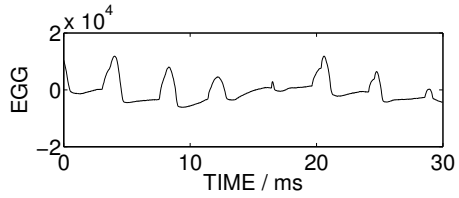
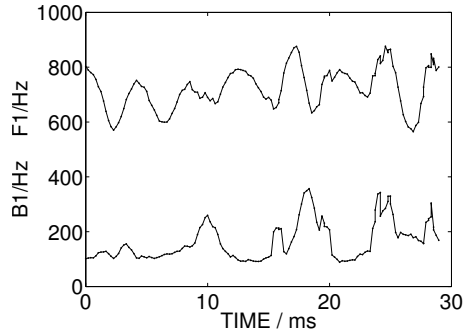


Figure 12: Inst. F1 and B1 (female hoarse)



4. DISCUSSION

This pilot study demonstrates that the temporally highly resolved first formant analysis is a candidate tool for the acoustic observation of different voice qualities (e.g., modal voice, breathy voice, hoarse voice). To sum up the observations, some similarities between the form of the electroglottographic contour and the frequency contour of the first formant are demonstrated: For modal voice quality, a regular process of the glottal cycle (closing and contact phase; opening and open phase) can be seen in the oscillation of the first formant frequency and in its narrow bandwidth. For breathy voice quality, however, a longer open phase and a larger bandwidth is observed. Finally, for hoarse voice quality, fundamental frequency and amplitude perturbations of the vocal fold vibration are apparent. Furthermore, the bandwidth of the first formant indicates a lower or higher loss of acoustic energy according to whether the subglottal cavity is minimally or more strongly coupled to the supraglottal vocal tract.

5. CONCLUSION

The observations in the present study may be important for speech perception for the following reason: The measurements are made in the frequency range of the formants. They are known to be linked very strongly to vowel quality and to be located in the frequency range of very high sensitivity for the auditory system. Furthermore, the measurements are instantaneous (short term within the pitch cycle) instead of long term over many pitch cycles and therefore they avoid the accompanying averaging effects.

6. REFERENCES

[1] Nawka, T., Anders, L., Wendler, J. 1996. Die auditive Bewertung heiserer Stimmen nach dem RBH-System. *Sprache Stimme Gehör* 18, 130–133.

[2] Pützer, M., Barry, W. 2004. Methodische Aspekte der auditiven Beurteilung von Stimmqualität. *Sprache Stimme Gehör* 28, 188–197.

[3] Pützer, M., Wokurek, W. 2006. Multiparametrische Stimmprofildifferenzierung zu männlichen und weiblichen Normalstimmen auf der Grundlage akustischer Analysen. *Laryngol Rhino Otol* 85, 1–8.

[4] Titze, I., Winholts, W. 1993. Effect of microphone type and placement on voice perturbation measurements. *J. Speech. Hear. Res.* 36, 1177–1190.

[5] Wokurek, W. March 2007. Erfassung des glottalen Öffnungsgrades durch Formantveränderungen während der Sprachgrundperiode. *Fortschritte der Akustik, Deutsche Arbeitsgemeinschaft für Akustik DAGA '07* Stuttgart. –.