INFLUENCE OF THE LPC FILTER UPON THE PERCEPTION OF BREATHINESS AND VOCAL EFFORT

Karl I. Nordstrom, Peter F. Driessen

University of Victoria
Dept. of Elec. and Comp. Engineering
P.O. Box 3055 STN CSC
Victoria BC, V8W 3P6, Canada
www.ece.uvic.ca/˜knordstr, www.ece.uvic.ca/˜peter
knordstr@uvic.ca, peter@ece.uvic.ca

Glen A. Rutledge

3dB Research Ltd.
PO Box 3075 STN CSC
R Hut McKenzie Ave.
Victoria BC, V8W 3W2, Canada
3dbresearch.com
grutledge@3dBResearch.com

ABSTRACT

According to the source-filter paradigm, the perception of breathiness and vocal effort should be primarily controlled by the glottal source and be little affected by the formant filter. This experiment investigates whether the formant filter estimated by linear prediction (LPC) can influence the perception of breathiness and vocal effort. The experiment starts with a pair of voice samples. One sample exhibits high effort and the other sample exhibits breathiness. LPC estimates a filter and residual for each sample. The influence of the residual is eliminated by providing both filters with the same artificial source during resynthesis. The synthesized samples differ only according to the difference between the two filters. Three pairs of samples were evaluated by seven people in listening tests. The results demonstrate that the LPC filters do influence the perception of breathiness and vocal effort. When a voice changes between breathiness and vocal effort, the spectral envelope changes. This change is captured by the LPC filter rather than the residual. A closer look at the LPC algorithm provides an explanation for this result.

1. INTRODUCTION

Linear prediction coding (LPC) is a common technique that uses a source-filter approach to analyzing the voice (see Figure 1). LPC estimates a formant filter for the voice by fitting the spectral envelope of the voice signal. By doing this, LPC implicitly assumes that the glottal source has a fixed spectral envelope. However, the true glottal source does not have a fixed or predetermined spectral envelope. For example, the glottal source for a breathy voice has less high-frequency content than for other voice qualities [1, 2]. When singing or speaking, the spectral envelope of the source varies according to the amount of breathiness or vocal effort. Any changes to the spectral envelope of the source end up in the estimated formant filter rather than the corresponding source. The purpose of this paper is to evaluate how much the LPC formant filter influences the perception of breathiness and vocal effort for a variety of voice samples.

1.1. Breathiness and vocal effort

Voices sound like they have breathiness or high effort depending on the nature of the vibrations of the vocal folds. The difference in vibration patterns between the two voice qualities creates a corresponding difference in the voice's frequency content.

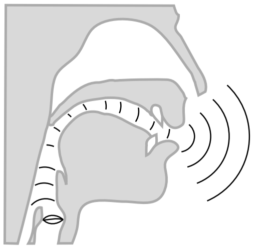
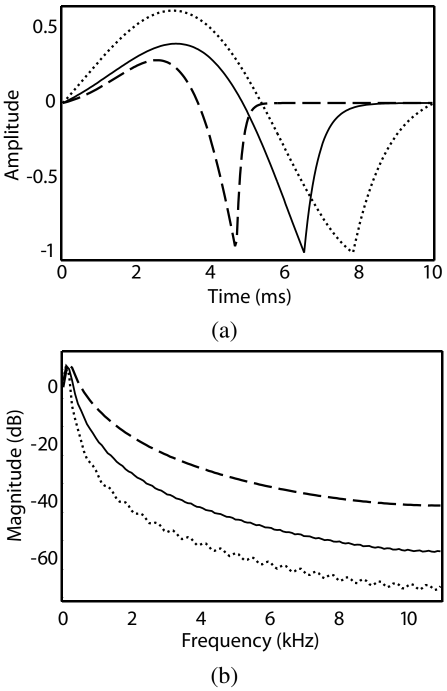


Fig. 1. The voice can be viewed as a source and a filter. The pressure waves originating at the vocal folds provide the glottal source. The vocal tract filters these pulses resulting in resonances that correspond to the vowel sounds.

Breathiness occurs when a voice is very relaxed. The vocal folds vibrate freely in a pattern that is almost sinusoidal. As a result, the lower harmonics are much stronger relative to the upper harmonics. In addition, air often leaks between the vocal folds when the voice is relaxed leading to significant aspiration noise.

Vocal effort (or tense voice [3]) is a subjective term that describes a strained or tense voice quality. The perception of vocal effort is associated with compression of the vocal folds and a reduced open quotient [1] (open quotient refers to the period that the vocal folds are open relative to the full cycle of one pulse). When a voice exhibits greater vocal effort, greater pressure builds up behind the vocal folds. When the pressure exceeds the resistance of the vocal folds, they open, releasing a short burst of air before quickly closing again. This makes the glottal source look more like a series of impulses. Given the impulsive nature of the excitation, the associated spectrum is more flat than for a breathy voice. Voices with more vocal effort have more high-frequency content [4].

The differences in the shapes of the glottal pulses can be seen by looking at some standard settings for the Liljencrant-Fant (LF) model [5]. The LF model provides time-domain pulses that represent the derivative of glottal flow. This model is widely used by the linguistic community for analyzing and synthesizing the glottal source. Figure 2 illustrates the differences in the glottal pulses between a breathy and a high-effort voice. These standard pulse shapes have been derived from careful analysis of the glottal pulse shapes [6]. The corresponding differences in the frequency spectra have also been plotted. This clearly demonstrates that a breathy source and a high effort source each have a different frequency spectrum.



(a)

(b)

Fig. 2. The LF model creates (a) a pulse representing the derivative of glottal flow in the voice source with (b) a corresponding frequency spectrum. Three voice qualities are represented here: modal voice, that is, a neutral voice (solid line); breathy voice (dotted line); and high-effort voice (dashed line). The frequency spectra have been normalized at the peak.

1.2. LPC

LPC depends on a source-filter concept of the voice [7, 8] as illustrated in Figure 1. In this concept, the vocal folds provide the source by creating air pulses. These air pulses are then acoustically transmitted through the vocal tract. The resonances in the vocal tract act as a filter, changing the shape of the acoustic spectrum (Figure 3(a)). These resonances are what we use to identify different vowels and are called formants.

LPC fits an all-pole filter to the spectrum of the signal. The all-pole filter is of the following form:

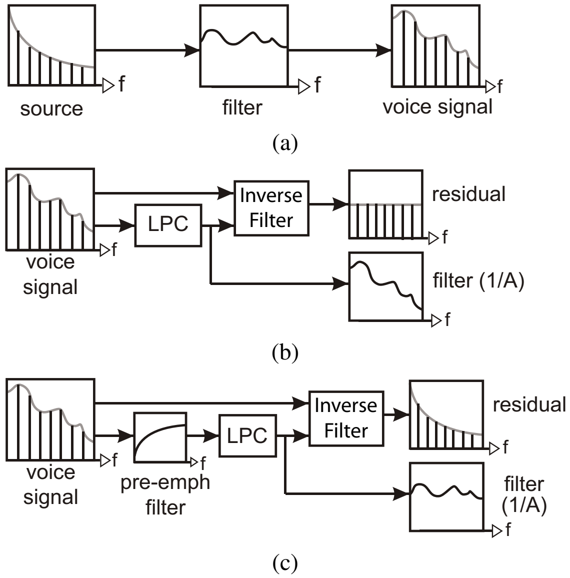
$$H(z) = \frac{G}{A(z)}, \tag{1}$$

where G is the gain and $A(z)$ is an all-zero filter, defined as follows:

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} \tag{2}$$

The order of the filter is defined by p. The operation of the LPC algorithm [7] and its relation to the human voice [8] have been thoroughly described in the literature.

LPC finds a filter to fit the spectrum of the input signal. If we apply the inverse of this filter to the original signal, we can extract the LPC residual. This residual represents the glottal source. Given that LPC attempts to minimize the error between the spectrum of the signal and the frequency response of the filter, the LPC residual has a flat spectrum as seen in Figure 3(b).



(a)

(b)

(c)

Fig. 3. (a) Source-filter model of the voice (b) LPC analysis algorithm (c) LPC analysis algorithm with pre-emphasis filter. The tilt of the residual spectrum is the inverse of the pre-emphasis filter.

However, most LPC algorithms compensate for lip radiation with a pre-emphasis filter. The voice signal goes through the pre-emphasis filter before it enters the LPC algorithm (Figure 3(c)). The pre-emphasis boosts the high frequencies, resulting in slightly better formant matching at the high frequencies and fewer scaling issues in fixed-point algorithms. The resulting LPC residual has a tilt corresponding to the inverse of the pre-emphasis filter. This is closer to the expected spectral envelope for the glottal source.

In most applications, the pre-emphasis filter is fixed. Regardless of whether the analyzed voice is breathy or whether it exhibits high effort, the pre-emphasis filter determines the spectral envelope of the residual. If the pre-emphasis filter is fixed then the spectral envelope of the residual also remains fixed. The residual does not follow changes to the voice quality such as breathiness and vocal effort.

The changes to the spectral envelope due to breathiness and vocal effort are captured in the LPC filter. This means that characteristics of the source are captured by the estimated formant filter. This can lead to problems when attempting to model the voice because the variation in the tilt of the source has not been modeled independent from the tilt of the formant filter. For example, one can attempt to make a voice sound breathy by adding aspiration noise but it becomes difficult to know how to change the spectral envelope of the source without having an estimate of the source envelope. The purpose of this experiment is to demonstrate that LPC with fixed pre-emphasis results in estimated formant filters that contain some perception of breathiness and vocal effort.

2. EXPERIMENT SETUP

One way to evaluate the influence of the formant filter is to take two different formant filters and supply them with the same glottal source. In this situation, the only difference between the resulting synthesized voices is the filter. If the vocal tract filter does not influence the perception of breathiness, then both voices should be per-

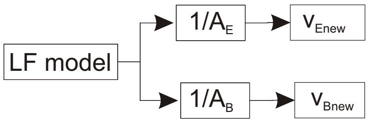


Fig. 4. The synthesized pair of voices were generated with the same artificial source using an LF model. The LPC filters were extracted from the high-effort voice ($1/A_E$) and the breathy voice ($1/A_B$). Any difference between the synthesized voices is due to differences in the LPC filters.

ceived to have the same amount of breathiness. If the formant filter does influence breathiness, then a difference will be observed. The process for creating and evaluating the samples is as follows:

1. Start with two samples in which the same person sings the same vowel at the same pitch but with differing voice qualities: high effort voice (V_E) and breathy voice(V_B).
2. Use LPC (Figure 3(c)) on each voice to estimate filters ($1/A_E$ and $1/A_B$).
3. Excite the filters ($1/A_E$ and $1/A_B$) with an LF model (Figure 2) plus noise to generate synthesized voices: V_{Enew} and V_{Bnew}. Since the source is the same for both voices, any difference between the voices will be due to the filters (see Figure 4).
4. Carry out a listening test evaluating the difference between the two filters.
   (a) Rate the relative difference in breathiness between the the original voices: V_E w.r.t. V_B.
   (b) Rate the relative difference in breathiness between the the synthesized voices: V_{Enew} w.r.t. V_{Bnew}.
   (c) A rating of zero indicates that there is no difference between V_{Enew} and V_{Bnew}, indicating that the filters ($1/A_E$ and $1/A_B$) do not influence the perception of breathiness. A non-zero rating indicates that the filters do influence the perception of breathiness. See Figure 5 for the results.
5. Repeat steps 4(a-c) for vocal effort.

2.1. Algorithm details

The voices were recorded at a sample rate of 22050 Hz, which was chosen as a compromise between having enough bandwidth to capture the breathy quality and a low enough sample rate for LPC to model the spectrum well.

A Liljencrant-Fant (LF) model [5] provides the glottal source by generating pulses in time that represent the opening and closing of the vocal folds. The LF model can be controlled by a dimensionless parameter, R_d, to provide a range of voice qualities between breathy voices with a high open-quotient ($R_d = 0.5$) to a neutral, modal voice ($R_d = 1$) to voices with a small open-quotient ($R_d = 2$) [6]. In this experiment, we found that R_d values between 0.5 and 0.8 worked well. The primary concern was for the LF model to sound natural. We can get a reasonable comparison between filters as long as the R_d parameter is kept identical for the sample pairs being compared to each other (Figure 4).

The rate at which the LF model provided time pulses was controlled by the pitch from the original voices. The pitch was extracted using Praat phonetics software [9]. The profiles of the pitch contours were similar between the breathy and high-effort voices.

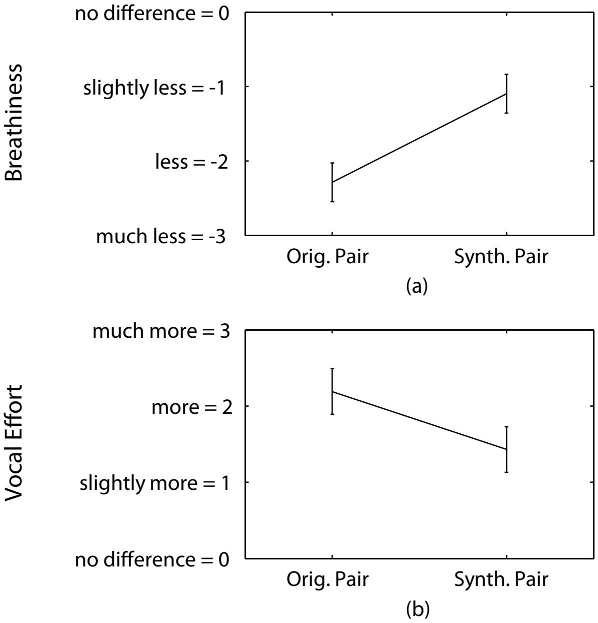


Fig. 5. Plot of the relative difference in (a) perceived breathiness and (b) perceived vocal effort within each sample pair. 95% confidence intervals have been plotted. “Orig. Pair” represents the rating of the original high-effort voice relative to the original breathy voice. “Synth. Pair” represents the rating of the synthesized high-effort voice relative to the synthesized breathy voice. The negative rating for breathiness indicates that the high-effort sample sounds less breathy than the corresponding breathy sample.

An LPC order of 22 was chosen as it approximately corresponds to a typical vocal tract length [8]. A higher LPC order, such as 50, with artificial excitation, results in a more natural-sounding voice. However, the additional spectral information from the higher order might artificially include detail about the breath quality that would otherwise remain in the LPC residual. For this reason, the LPC order was chosen to represent a physical vocal tract rather than choosing an unrealistic order to achieve better results. Bandwidth expansion was carried out using the pole-scaling method [10] to reduce peakiness in the LPC spectrum. The LPC coefficients were computed every 32 samples and linearly interpolated to reduce the influence of discontinuities between filters.

A pre-emphasis filter ($1 - .99z^{-1}$) was applied to the voice before it entered the LPC analysis algorithm (see Figure 3). With this pre-emphasis filter, The LPC residual approximately matches the spectral envelope of the LF model. This meant that no tilt adjustments were needed when replacing the residual with the LF model.

The aspiration noise consisted of white noise with a square wave envelope that was synchronized with the pulses from the LF model. Providing noise pulses to the model helped the noise to blend into the voice more easily [1].

There were three pairs of original samples resulting in three more synthesized samples. In total, there were six pairs of samples to be evaluated. After synthesis, the samples were normalized to have the same energy level.

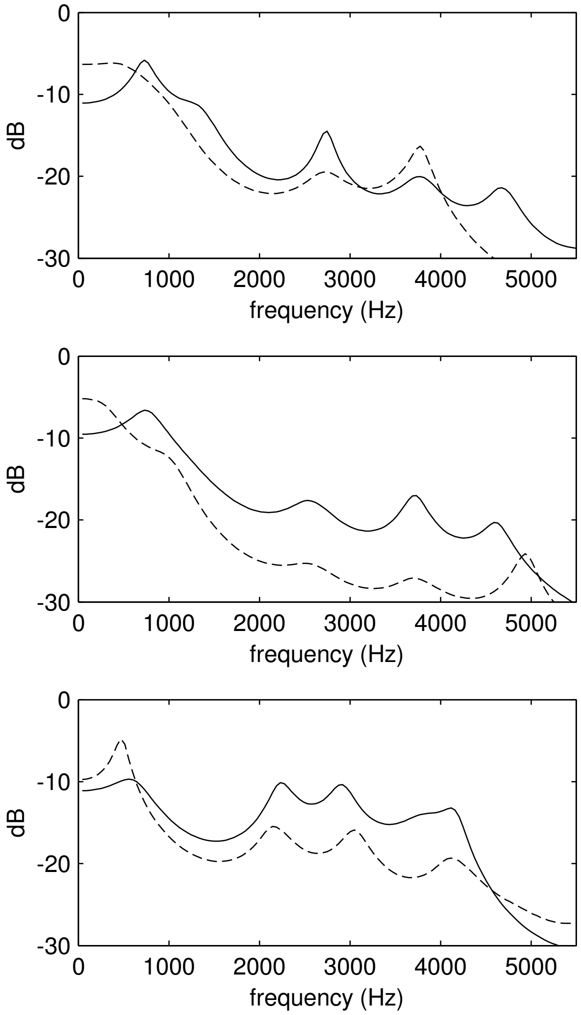


Fig. 6. Frequency spectra from a number of LPC filters for breathy voices (dashed lines) and high-effort voices (solid lines). In each plot the same voice is singing the same vowel on the same pitch. The breathy voices have a stronger first formant and less high-frequency content than the corresponding high-effort voice.

2.2. Listening Experiment

Listening experiments were carried out to evaluate the samples. There were a total of seven listeners. The perceptual criteria for this test was drawn from other studies for evaluating breathy voices [1, 2] and a prior test that we conducted [11].

The test was designed to measure relative differences between the high-effort sample and the breathy sample. This approach has been used with good results for evaluating breathy voices [2]. The questions were worded as follows:

- Listen to the two samples and rate which one sounds more breathy.
- Listen to the two samples. Rate which voice sounds like it requires more effort to sing. Vocal effort would be associated with a tense voice rather than a relaxed voice.

The difference in breathiness or vocal effort between the two samples was evaluated on a seven point scale. For example, the possible ratings for breathiness ranged from much less breathy to no difference to much more breathy. Half of the rating scale can be seen on the vertical axis of the results in Figure 5. Breathiness and vocal effort were evaluated in separate runs.

The evaluator did not know which sample pairs were being provided or the order in which they were presented. Within each sample pair, the breathy or high-effort sample was randomly chosen to be first. This order was randomized for each run. In addition, the order of the six sample pairs was randomized for each run of the test for each listener. The evaluation process was automated for the listener and did not involve intervention on part of the experimenter.

3. RESULTS

The results of the subjective evaluation are displayed in Figure 5. The ratings are presented with the high-effort voice relative to the breathy voice. An F-test was carried out to determine whether the differences between the means are significant relative to sampling noise. The results were found to be significant with an F-value of 10.2 [12].

As expected, there was a large difference in the perceived breathiness between the original sample pairs (see Figure 5(a)). The rating of -2.3 indicates that the high-effort sample sounded less breathy than the breathy sample. When the LPC filters from both of these samples were excited by the LF model, the perceived difference in breathiness was reduced to a rating of -1.1. The high-effort LPC filter sounded slightly less breathy than the breathy LPC filter. The 95% confidence interval indicates that the two filters did not sound the same. The LPC filter from the breathy voice was clearly more breathy than the LPC filter from the high-effort voice.

There was a large difference in the perceived vocal effort between the original sample pairs (see Figure 5(b)). The rating of 2.2 indicates that the high-effort sample sounded like it had more effort than the breathy sample. When the LPC filters from both of these samples were excited by the LF model, the perceived difference in vocal effort was reduced to a rating of 1.4. The high-effort LPC filter still sounded like it had more effort than the breathy LPC filter. The 95% confidence interval indicates that the two filters did not sound the same.

These results indicate that the LPC filters do have an influence on the perception of breathiness and vocal effort.

4. DISCUSSION

The LPC filters from the high-effort samples sound different than the LPC filters from the breathy samples because there is a consistent difference between their spectra. The spectra for the three pairs of filters have been plotted in Figure 6. The first formant for the breathy filters is at a lower frequency and is generally stronger than for the high-effort filters. The breathy filters typically have less energy than the high-effort filters between 1000 Hz and 4500 Hz. The breathy filters accentuate the first formant while the high-effort filters emphasize higher frequencies.

The difference in emphasis between high and low frequencies is likely due to physiological changes in the glottal source. Figure 2(b) shows how the glottal source changes between breathy and high-effort voices. This change in the spectra is being captured by the LPC filters rather than being modeled as part of the LPC residual.

The LPC algorithm does not take spectral changes to the glottal source into account. Whether the voice has a little or a lot of vocal effort, whatever the shape of the glottal spectrum, the spectral envelope of the residual does not change. The spectral envelope of the

LPC residual is fully determined by the pre-emphasis filter as seen in Figure 3(c). There is no identification of an appropriate spectral envelope for the source in the standard LPC algorithm.

5. CONCLUSIONS

This listening experiment showed that LPC filters do influence the perception of breathiness and vocal effort. The LPC filters were estimated from pairs of voice samples where one was breathy and the other had high-effort. The pairs of LPC filters were excited with the same LF excitation to ensure that the LPC filters created the only differences between the samples. Listening tests were carried out to evaluate the differences between LPC filters from the breathy and the high-effort voices. The results indicate that the LPC filters do retain some of the breathiness or vocal effort from the original voices. However, from a perceptual perspective, the formant filter should not contain information about breathiness or vocal effort. This information should be in the glottal source.

There is a reason why the LPC formant filters capture some of the perception of breathiness and vocal effort. LPC, as it is commonly implemented, assumes that the glottal source has a fixed spectral envelope. In contrast, the true glottal source varies according to different voice qualities, often within a single phrase. One way to vary the spectral envelope of the residual is to implement a variable pre-emphasis algorithm [13]. The concept of variable pre-emphasis is not new but in previous applications the purpose has typically been to improve voice compression or speech recognition [14]. Little work has been done to develop a variable pre-emphasis algorithm that matches the perceptual characteristics of the glottal source.

6. ACKNOWLEDGEMENTS

Thanks to John Esling for his enlightening discussions about voice physiology and voice quality. Thanks to Anne Bateman for sharing voice data for this experiment [15]. Thanks also to IVL Technologies and TC Helicon for their collaborative support in this research.

7. REFERENCES

[1] Donald G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *Journal of the Acoustical Society of America*, vol. 90, no. 5, pp. 2394–2410, November 1991.

[2] Dennis H. Klatt and Laura C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 820–857, February 1990.

[3] John Laver, *The Phonetic Description of Voice Quality*, Cambridge University Press, New York, 1980.

[4] Björn Granström and Lennart Nord, "Neglected dimensions in speech synthesis," *Speech Communication*, vol. 11, pp. 459–462, 1992.

[5] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPSR*, vol. 4, pp. 1–13, 1985.

[6] Gunnar Fant, "The LF-model revisited: Transformations and frequency domain analysis," *STL-QPSR*, vol. 2-3, pp. 119–156, 1995.

[7] John Makhoul, "Linear prediction: A tutorial review," in *Proceedings of the IEEE*, April 1975, vol. 63, pp. 561–580.

[8] John D. Markel and Augustine H. Gray, *Linear Prediction of Speech*, Springer-Verlag Berlin Heidelberg, New York, 1976.

[9] Paul Boersma and David Weenink, "Praat: A system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[10] Peter Kabal, "Ill-conditioning and bandwidth expansion in linear prediction of speech," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 824–827, April 2003.

[11] Karl I. Nordstrom, Glen A. Rutledge, and Peter F. Driessen, "Using voice conversion as a paradigm for analyzing breathy singing voices," in *Pacific Rim Conference (PACRIM) on Communications, Computers and Signal Processing*, Victoria, BC, Canada, August 2005, pp. 428 – 431.

[12] J. S. Milton and Jesse C. Arnold, *Introduction to Probablility and Statistics: Principles and Applications for Engineering and the Computing Sciences*, McGraw Hill, New York, 1990.

[13] Karl I. Nordstrom and Peter F. Driessen, "Variable preemphasis LPC for modeling vocal effort in the singing voice," in *Proceedings of the 9th International Conference on Digital Audio Effects*, Montreal, September 2006.

[14] Sahar E. Bou-Ghazale and John H. L. Hansen, "A comparative study of traditional and newly proposed features for recognition of speech under stress," *IEEE Transaction of Speech and Audio Processing*, vol. 8, no. 4, pp. 429–442, July 2000.

[15] Laura Anne Bateman, "Soprano, style and voice quality: Acoustic and laryngographic correlates," M.S. thesis, University of Victoria, 2004.