VARIABLE PRE-EMPHASIS LPC FOR MODELING VOCAL EFFORT IN THE SINGING VOICE

Karl I. Nordstrom, Peter F. Driessen

Music Intelligence and Sound Technology Interdisciplinary Centre (MISTIC)
University of Victoria, Canada
knordstr@uvic.ca, peter@ece.uvic.ca
http://karlnordstrom.ca, http://www.ece.uvic.ca/~peter/

ABSTRACT

In speech and singing, the spectral envelope of the glottal source varies according to different voice qualities such as vocal effort, lax voice, and breathy voice. In contrast, linear prediction coding (LPC) models the glottal source in a way that is not flexible. The spectral envelope of the source estimated by LPC is fixed and determined by the pre-emphasis filter. In standard LPC, the formant filter captures variation in the spectral envelope that should be associated with the source. This paper presents variable pre-emphasis LPC (VPLPC) as a technique to allow the estimated source to vary. This results in formant filters that remain more consistent across variations in vocal effort and breathiness. VPLPC also provides a way to change the envelope of the estimated source, thereby changing the perception of vocal effort. The VPLPC algorithm is used to manipulate some voice excerpts with promising but mixed results. Possible improvements are suggested.

1. INTRODUCTION

Linear prediction coding (LPC) estimates a voice source with a fixed spectral envelope. The true voice source has a spectral envelope that varies. As a result, part of the perceived voice quality [1, 2, 3] that should be in the source ends up in the LPC filter [4]. This paper presents variable pre-emphasis LPC (VPLPC) for separating the spectral envelope into two components: formant filter and source envelope. The paper describes why the true source varies between high-effort and breathy voices, why standard LPC estimates a formant filter that captures source variation, and how VPLPC can estimate a formant filter that is more consistent across different voice qualities. An attempt is then made to manipulate the perceived vocal effort with the VPLPC algorithm.

1.1. Variation in the glottal source

The spectral envelope of the glottal source does not remain constant but varies according to changes in glottal voice quality such as tense voice, lax voice and breathiness [5]. This variation in glottal quality often happens within a single spoken or sung phrase. When a voice is tense, it requires more effort to phonate. The resulting voice has more high frequency content than the same voice in a relaxed state. When a voice is relaxed (known as lax voice), the vocal folds move freely resulting in vibrations that appear almost sinusoidal. The lower harmonics are much stronger relative to the upper harmonics. Air often leaks between the vocal folds when the voice is relaxed. When air leakage causes significant aspiration noise and the vocal folds are relaxed, it is known as a breathy voice.

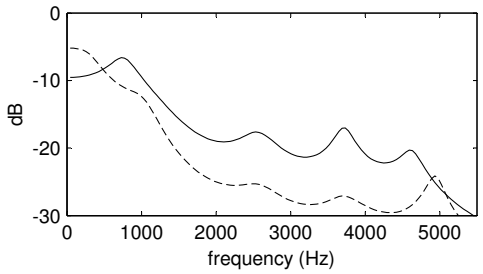

Figure 1: *Frequency spectra of LPC filters for breathy voice (dashed line) and high-effort voice (solid line). The same voice is singing the same vowel on the same pitch.*

Vocal effort is a subjective term that describes a strained or tense voice quality. The perception of vocal effort has been associated with compression of the vocal folds and a reduced open quotient [6]. When a voice exhibits vocal effort, 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. The voice source looks like a series of impulses and the spectral envelope is flatter than the envelope from a lax or breathy voice. The difference in the spectra between voices with high and low effort can be seen in Figure 1.

There are a number of indicators of vocal effort. While the overall sound level is significant, people are able to identify the effort a person is putting into speaking or singing even when the sound level has been normalized [7]. Vocal effort is also associated with higher pitches and changes in the phrasing [8]. We are looking at singing voices and therefore many of these factors are less influential. The music specifies pitch and phrasing, and the sound level is normalized to the recording. In this situation, the spectral balance between low frequencies and high frequencies is more significant. Voices with more vocal effort have more high-frequency content. Another indicator of vocal effort is the mix of harmonics and noise in the voice signal. Voices with vocal effort have little aspiration noise (as long as the source has not become aperiodic). In contrast, relaxed voices have a more aspiration noise.

1.2. Problem with LPC

Standard LPC models the voice in a way that does not appropriately capture variation in the source. This subsection describes the problem.

The source-filter concept provides a perceptual approximation of the glottal source and vocal tract and is widely used for voice

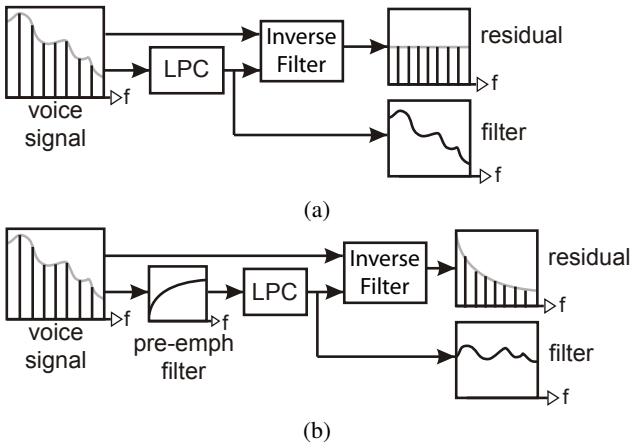

Figure 2: *(a) LPC analysis algorithm (b) LPC analysis algorithm with pre-emphasis filter. Note that the tilt of the residual spectrum is the inverse of the pre-emphasis filter.*

analysis and synthesis [9]. The most common technique for doing this is linear prediction coding (LPC). The operation of the LPC algorithm [10] and its relation to the human voice [11] have been greatly discussed 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.

LPC attempts to minimize the error between the spectrum of the signal and the frequency response of the filter. As a result, the LPC residual has a flat spectrum as seen in Figure 2(a). Most LPC algorithms compensate for lip radiation with a pre-emphasis filter. 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 tilt of the LPC residual is the inverse of the pre-emphasis filter as seen in Figure 2(b). This is closer to the expected appearance of the glottal source. However, the pre-emphasis algorithm does not estimate the spectral envelope of the source. The slope of the residual's spectrum is fixed.

The LPC algorithm does not take into account spectral changes to the glottal source. Whether the voice has much or little vocal effort, whatever the shape of the glottal spectrum, the pre-emphasis filter remains the same and the LPC residual has the same spectral envelope. This means that variation in the envelope of the glottal spectrum is captured by the LPC filter instead of the LPC residual [4, 12]. This appears to be an inherent part of LPC that has not been clearly addressed in the speech literature.

2. VARIABLE PRE-EMPHASIS LPC

This paper proposes that the pre-emphasis filter be made variable. We already know that the pre-emphasis filter determines the spectral envelope of the LPC residual. Variable pre-emphasis results in a residual that responds to broad changes to the spectral envelope. As long as this variation in the spectral envelope does not affect the perception of formants, the assumption is that variable pre-emphasis captures a glottal voice quality. We cannot verify this by physiological measurement but we can perceptually evaluate this influence.

If we compare VPLPC formant filters from high-effort and breathy voices, we should find the VPLPC formant filters to be more consistent than the corresponding formant filters from standard LPC. Variable pre-emphasis should reduce variation in the formant filters by allowing glottal voice qualities to pass through to the residual.

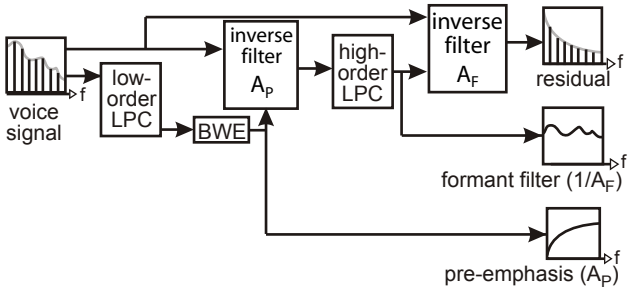

Figure 3: *Variable pre-emphasis LPC analysis. Low-order LPC plus bandwidth expansion (BWE) determines the variable pre-emphasis filter A_P. Following pre-emphasis, high-order LPC determines the formant filter $1/A_F$.*

Variable pre-emphasis is not a new idea. Some LPC algorithms use variable pre-emphasis to improve voice compression or speech recognition [13]. This paper adds to that research by suggesting that there is a physiological reason why variable pre-emphasis works and by attempting to use it to manipulate the perception of vocal effort.

2.1. Low-order LPC

One way to estimate an appropriate pre-emphasis filter is to carry out low-order LPC. The low-order LPC analysis method is presented in Figure 3. Because LPC tends to produce filters that are peaky, bandwidth expansion (BWE) is carried out on the filter coefficients using pole scaling [14].

From experimentation, an order of three appeared to work best while using a sampling rate of 22050 Hz. In standard LPC, one pole is located at 0 Hz to represent lip radiation. This can be thought of as one pole in the pre-emphasis filter. Adding another pole pair enables the algorithm to capture a broad resonance in the spectrum around 2–3 kHz that can happen in high-effort voices. A couple of example pre-emphasis filters are shown in Figure 4.

It may seem strange that there is an upper resonance in the pre-emphasis filter in Figure 4(b). Most voice analysis in linguistic research takes place at lower sampling frequencies, truncating the plot at approximately 5 kHz. This makes changes in vocal effort appear as a tilt. Plotting the spectrum up to 11 kHz reveals that many high-effort voices have a hump in the spectrum.

After pre-emphasis, the signal is fed into high-order LPC to estimate the formant filter. The order of the formant filter was informally adjusted to the order that perceptually seemed to work best. Orders between nineteen and twenty-four at a sampling rate of 22050 Hz roughly correspond to the length of a human vocal tract when LPC is interpreted as a physical model. Since LPC does not typically estimate a true vocal tract, we weren't constrained by this range. We used an order of thirty for the sound excerpts that we modified. To extract the residual, the original signal is inverse filtered by the formant filter.

2.2. Synthesis method

To modify the perceived vocal effort, the spectral envelope of the residual has to be modified and resynthesized. The process of

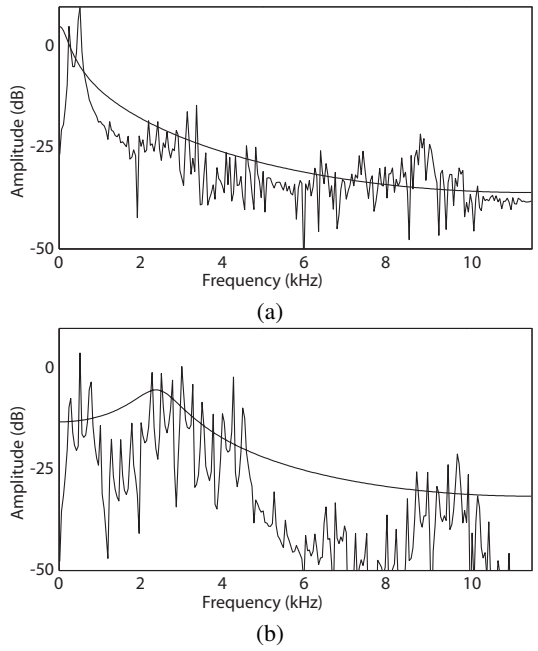

Figure 4: *Inverse of the pre-emphasis filter (A_P) estimated by low-order LPC for (a) breathy and (b) high-effort voice excerpts.*

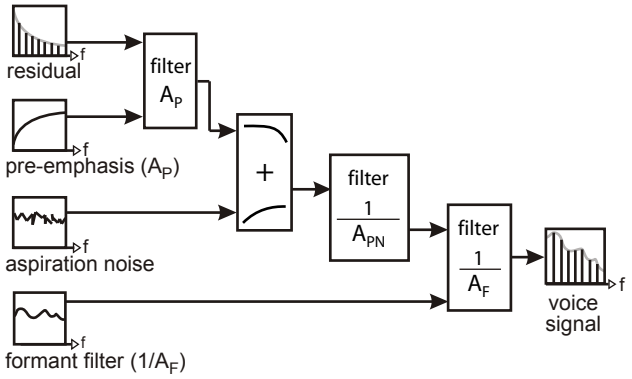

Figure 5: *VPLPC synthesis configured to modify the perception of vocal effort. The tilt of the residual is removed by the variable pre-emphasis filter (A_P), leading to a flat spectrum. Matched filters blend in aspiration noise if required. A new pre-emphasis filter (A_{PN}) applies the desired spectral envelope for the glottal source. The signal is then filtered by the formant filter ($1/A_F$) to achieve the new voice signal.*

resynthesizing the voice is illustrated in Figure 5. First, the spectral envelope of the residual is flattened by filtering with the pre-emphasis filter. The flat spectrum makes it easier to add aspiration noise if required. Two matched butterworth filters were used to blend aspiration noise, one filter applied to the flattened residual and one filter applied to the white noise. Then, a filter representing the desired spectral envelope is applied to the mix of the flattened residual and aspiration noise. The resulting signal is the modified source with the desired spectral envelope and aspiration noise when required. The modified source is fed through the formant filter to synthesize the voice.

Voices with less vocal effort have less aspiration noise. In the algorithm, the aspiration was generated as gaussian noise. This noise was pulsed in sync with the frequency of the voice using a square envelope with a 50% duty cycle. The pitch was estimated using Praat software [15]. The energy level of the noise was adjusted to be the same as the energy level of the flattened residual. Matched, first-order butterworth filters were used to low-pass the residual and high-pass the pulsed noise. This blended the noise into the flattened residual.

3. RESULTS

One of the advantages that VPLPC should provide is a formant filter (A_F) that is more consistent over varying voice qualities. We tested this aspect of VPLPC by exciting the formant filters from high-effort voices and breathy voices with the same excitation. For raw voice data, we had three pairs of voice excepts. Within each pair, the same person sang the same vowel at the same pitch varying only their voice quality between high effort and breathiness. To remove the influence of the source, we used the same LF model [16] as the excitation for all voices. The LF model is the most popular model for glottal excitation. We also carried out the same procedure using standard LPC. Due to the nature of artificial excitation, some artifacts were present in the data; however, it was still easy to hear relative differences in voice quality between samples.

When the same LF model excited the VPLPC formant filters (A_F) from the high-effort and the breathy voices, both synthesized voices took on similar voice qualities. The biggest difference between the two voices was that the formant filter for the breathy voice had more jitter and shimmer associated with it, resulting in more artifacts and slightly more perceived breathiness. The LPC formant filters sounded more different from each other. The LPC formant filter from the high-effort voice carried a significant perception of vocal effort. In comparison, the LPC formant filter from the breathy voice carried a significant perception of breathiness.

3.1. Manipulating the perception of vocal effort

In the next stage, we attempted to use VPLPC to manipulate the perception of vocal effort. The shape of the spectral envelope for the desired source was estimated from the excerpts. Using VPLPC analysis, A_P from the breathy voice provided A_{PN} for the high-effort voice. A_P from the high-effort voice provided A_{PN} for the breathy voice. Doing this gives the spectral envelope of the breathy residual to the spectral envelope of the high-effort residual and vice versa.

First, vocal effort was removed from the high-effort voice. The spectral envelope of the high-effort residual was replaced by the spectral envelope of the breathy residual. This reduced the high-frequency content of the voice. Some of the perception of vocal effort was reduced. However, the new voice did not sound as relaxed as the original breathy voice. Although the spectral envelope changed, the mix of harmonics and noise did not change. A relaxed voice should have fewer harmonics and more aspiration noise.

To further reduce the perception of vocal effort, we added noise to the residual as described in section 2.2. The addition of aspiration noise made the synthesized voice sound more natural, which improved the perception of a more relaxed voice. However, the voice that was originally perceived to have high effort was not fully transformed into a relaxed breathy voice. It was difficult to

blend much noise into the residual without creating an unnatural sounding voice. Perceptually, the noise easily separated from the source, sounding like a separate stream of noise rather than part of the voice.

Next, vocal effort was added to the breathy voice. The spectral envelope of the breathy residual was replaced by the spectral envelope of the high-effort residual. This boosted the high-frequency content of the voice. The resulting voice was perceived to have a higher degree of vocal effort but there was too much aspiration noise and not enough harmonics. The voice sounded noisy and unnatural due to the amplified noise.

Original and synthesized voice samples are available online at: http://www.ece.uvic.ca/~knordstr/dafx06

4. CONCLUSIONS

This paper presented the VPLPC algorithm as a method to estimate formant filters that are more neutral across varying voice qualities. While formant filters from standard LPC contain a significant perception of vocal effort, VPLPC appeared able to remove the perception of vocal effort from formant filters that were excited by an LF model.

The VPLPC algorithm was able to partially change the perception of vocal effort by manipulating the residual. The transformation was not complete because the mix of harmonics and noise did not change along with changes to the spectral envelope. When reducing vocal effort, it helped to add pulsed noise into the residual. Unfortunately, only a little noise could be added before there were problems with stream separation between the noise and the residual. In transformations to high effort, the residual did not have enough harmonic content, resulting in voices that sounded noisy.

There are a couple of ways that the artifacts could could be reduced. VPLPC, as presented here, involved two-stages of inverse filtering. The opportunity for artifacts increases with each stage, especially when the filters are dynamic like those from LPC. It may be possible to eliminate the pre-emphasis inverse filter by extracting key information from a standard LPC filter. Given that the variable pre-emphasis filter is of a low order, it should be possible to convert an analysis of pole locations into a single measure of vocal effort. This parametric measure of vocal effort could then be used to control the model.

Methods other than low-order LPC could be used to estimate a pre-emphasis filter. A larger number of voice excerpts could provide a better indication for an appropriate estimation method.

5. ACKNOWLEDGEMENTS

I send my thanks out to the following people. Laura Anne Bateman provided some voice excerpts from her research [17]. George Tzanetakis provided some feedback on the paper. Thanks to Glen Rutledge for useful discussions about voice modeling and John Esling for improving my understanding of phonetics. Thanks to IVL Technologies and TC Helicon for starting me in this research path and lending me some audio equipment.

6. REFERENCES

[1] F. Thibault and P. Depalle, "Adaptive processing of singing voice timbre," in *Proc. IEEE Canadian Conf. on Electrical and Computer Engineering (CCECE2004)*, Niagara Falls, Ontario, Canada, 2004, pp. 871–874.

[2] L. Fabig and J. Janer, "Transforming singing voice expression – the sweetness effect," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-04),* Naples, Italy, 2004, pp. 70–75.

[3] A. Loscos and J. Bonada, "Emulating rough and growl voice in spectral domain," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-04),* Naples, Italy, 1998, pp. 188–91.

[4] K. I. Nordstrom, P. F. Driessen, and G. A. Rutledge, "Influence of the LPC filter upon the perception of breathiness and vocal effort," in *IEEE Int. Symposium on Signal Processing and Information Technology (ISSPIT06)*, Vancouver, Canada, Aug. 2006.

[5] A. N. Chasaide and C. Gobl, *Handbook of Phonetic Sciences*. W. J. Hardcastle and J. Laver Eds. Blackwell, 1997, ch. Voice source variation, pp. 427–461.

[6] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *J. Acoust. Soc. Am.*, vol. 90, no. 5, pp. 2394–2410, 1991.

[7] J.-S. Liénard and M.-G. D. Benedetto, "Effect of vocal effort on spectral properties of vowels," *J. Acoust. Soc. Am.*, vol. 106, no. 1, pp. 411–422, July 1999.

[8] H. Traunmüller and A. Eriksson, "Acoustic effects of variation in vocal effort by men, women and children," *J. Acoust. Soc. Am.*, vol. 107, no. 6, pp. 3438–3451, June 2000.

[9] P. R. Cook, "Toward the perfect audio morph? Singing voice synthesis and processing," in *Proc. COST-G6 Workshop on Digital Audio Effects (DAFx-98),* Barcelona, Spain, 1998, pp. 223– 230.

[10] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, Apr. 1975.

[11] J. D. Markel and A. H. Gray Jr., *Linear Prediction of Speech*. New York: Springer-Verlag Berlin Heidelberg, 1976.

[12] K. I. Nordstrom, G. A. Rutledge, and P. F. Driessen, "Using voice conversion as a paradigm for analyzing breathy singing voices," in *Pacific Rim Conf. on Communications, Computers and Sig. Proc. (PACRIM05)*, Victoria, Canada, 2005, pp. 428 – 431.

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

[14] P. Kabal, "Ill-Conditioning and bandwidth expansion in linear prediction of speech," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc. (ICASSP'03),* Hong Kong, China, 2003, pp. 824–827.

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

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

[17] L. A. Bateman, "Soprano, style and voice quality: Acoustic and laryngographic correlates," Master's thesis, University of Victoria, 2004.