Transforming Perceived Vocal Effort and Breathiness Using Adaptive Pre-Emphasis Linear Prediction

Karl I. Nordstrom, *Student Member, IEEE*, George Tzanetakis, *Member, IEEE*, and Peter F. Driessen, *Senior Member, IEEE*

***Abstract*—This paper presents a technique to transform high-effort voices into breathy voices using adaptive pre-emphasis linear prediction (APLP). The primary benefit of this technique is that it estimates a spectral emphasis filter that can be used to manipulate the perceived vocal effort. The other benefit of APLP is that it estimates a formant filter that is more consistent across varying voice qualities. This paper describes how constant pre-emphasis linear prediction (LP) estimates a voice source with a constant spectral envelope even though the spectral envelope of the true voice source varies over time. A listening experiment demonstrates how differences in vocal effort and breathiness are audible in the formant filter estimated by constant pre-emphasis LP. APLP is presented as a technique to estimate a spectral emphasis filter that captures the combined influence of the glottal source and the vocal tract upon the spectral envelope of the voice. A final listening experiment demonstrates how APLP can be used to effectively transform high-effort voices into breathy voices. The techniques presented here are relevant to researchers in voice conversion, voice quality, singing, and emotion.**

***Index Terms*—Adaptive pre-emphasis, breathiness, linear prediction (LP), pre-emphasis, spectral slope, vocal effort, voice quality.**

I. Introduction

In THIS paper, we present a technique to improve linear prediction (LP) for the transformation of high-effort singing voices into breathy voices. A common approach for this transformation is to separate the voice into a source and filter, add noise to the source to simulate aspiration noise, and to resynthesize the voice [1]. This approach works well for voices that already sound slightly breathy. However, it does not work as well for voices that exhibit high vocal effort. The noise intended to simulate aspiration noise does not sound like part of the synthesized voice and instead sounds like a segregated stream of noise. Also, the resulting voice retains the perception of high vocal effort, which is incompatible with breathy voices since breathy voices exhibit low effort. We show that synthesized aspiration noise does not blend with high-effort voices because the spectral envelopes of high-effort and breathy voices are different. We present adaptive pre-emphasis linear prediction (APLP) to model and modify the spectral envelope of the voice, thereby reducing the perceived vocal effort and improving the blending of aspiration noise.

There are many ways to describe the voice qualities that result from changes to the voice source and these terminologies often overlap. Vocal effort has been chosen in the context of this paper because increased effort describes a broad range of voice qualities where the vocal folds remain closed for a large portion of the glottal cycle. These voices have more high-frequency harmonic content due to the short length of the glottal pulses and the rapid closure of the vocal folds. We also chose the high-effort terminology because it describes something that most people can understand more easily than the standardized phonetic terminology [2]. People do not need specialized phonetic training to achieve a relatively consistent perception of vocal effort. It is more difficult to teach people the meaning of terms such as: pressed, laryngealized, creaky, or harsh voice. Vocal effort is a concept that both specialists and nonspecialists can grasp and come to agreement with less difficulty [3], [4].

Vocal effort is a subjective term that describes a strained or tense voice quality. Although the most obvious result of increased vocal effort is increased sound intensity [5], people can distinguish the quantity of effort in a voice independent of the volume at which the sample is provided [3]. Pitch can also be an indication of vocal effort [6], [7]. However, in the case of singing, the pitch has already been specified. Therefore, the dominant cue of vocal effort for the singing voice is the spectral envelope of the signal [4], [8].

When a voice involves effort, the resulting voice has more high-frequency content than the same voice in a relaxed state [9]. When a voice is relaxed (such as in lax [2] or breathy voices), the vocal folds move freely, with slow glottal closure. 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.

The spectral envelope of the voice source provides one of the most important cues for the perception of vocal effort. This envelope varies from voice to voice and can vary within the context of a single phrase [10]. Studies show that it is possible to model the spectral envelope of the voice source with a third-order low-pass filter [11], [12]. These studies modeling the spectral envelope of the voice source show that the speed of the return phase in the glottal pulses affects the spectral slope. A slow glottal return phase, such as for a breathy voice, results

Manuscript received October 5, 2007; revised April 9, 2008. Published July 16, 2008 (projected). This work was supported by IVL Technologies. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gael Richard.

K. I. Nordstrom and P. F. Driessen are with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC V8W 3P6, Canada (e-mail: knordstr@uvic.ca; northstream@gmail.com; peter@ece.uvic.ca).

G. Tzanetakis is with the Department of Computer Science, University of Victoria, Victoria, BC V8W 3P6, Canada (e-mail: gtzan@cs.uvic.ca).

Digital Object Identifier 10.1109/TASL.2008.2001105

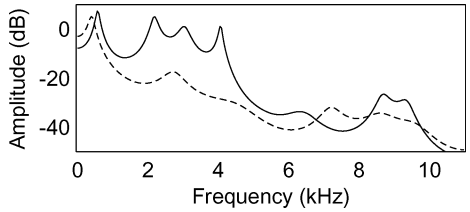

Fig. 1. Spectral envelopes estimated by linear prediction without pre-emphasis: a breathy voice (dashed line) and a high-effort voice (solid line). In each plot, the same voice is singing the same vowel on the same pitch. The breathy voice has less energy in the 1.5–4.5 kHz range than the corresponding high-effort voice.

in a steeper spectral slope starting at a lower frequency. A quick glottal return phase, such as for a high-effort voice, results in a less steep spectral slope.

The frequency response of the vocal tract also influences the spectral envelope of the voice. For wider bandwidth voice signals (above 5 kHz), the frequency response of the vocal tract is not flat throughout the entire frequency range. The difference in the spectra between voices with high and low effort can be seen in Fig. 1. The frequency response of the high-effort voice is relatively flat up to 4–5 kHz after which it drops off sharply. Physical models of the vocal tract have suggested that the cutoff frequency and the suddenness of this drop-off is due to throat constriction in the lower vocal tract [13]. This is also related to the singer's formant which results in the clustering of the third, fourth, and fifth formants [14]. Unfortunately, little voice analysis has taken place beyond 5 kHz, and further research would be useful to determine how much of the frequency response beyond 5 kHz is due to the vocal tract and how much is due to the source. The challenge with analysis beyond 5 kHz is that the acoustic waves in the vocal tract can no longer be assumed to be plane waves as the wavelengths are shorter than the width of the vocal tract. The drop-off observed in high-effort voice samples is a challenge to standard LP methods because the spectral slope of the vocal tract can no longer be considered to be consistent throughout the frequency range.

To convert a high-effort voice into a breathy voice, it is not enough to add aspiration noise. To eliminate vocal effort, we need to change the envelope of the spectrum to match the breathy envelope then add aspiration noise corresponding to breathiness. If we change one thing but not the other, then different voice qualities result. Adding aspiration noise to a high-effort voice without changing the spectral envelope results in a voice that simultaneously exhibits effort and aspiration noise. This has been described as a whispery voice [15], [16]. Alternately, transforming the spectral envelope of the high-effort voice into that of a breathy voice without adding noise results in a voice that sounds lax and unnatural. It gives the perception that the vocal folds are relaxed but there is no aspiration noise that our ears expect to hear.

To create the perception of low vocal effort and to improve the blending of aspiration noise into the voice, we propose APLP. Adaptive pre-emphasis has been used with LP, but its relationship to vocal effort and other voice qualities has not been elucidated. Adaptive pre-emphasis is often used to avoid ill-conditioning in fixed-point algorithms due to the contrast in spectral slopes between voiced and unvoiced segments [17]. Some LP algorithms use adaptive pre-emphasis to improve speech recognition [18], [19] or accent detection [20]. We present APLP as a technique to estimate the spectral emphasis of the voice due to vocal effort. This spectral emphasis, once estimated, can be manipulated to change the perceived amount of vocal effort in the voice. It is our expectation that reducing the perceived vocal effort will improve the blending of aspiration noise into the voice while transforming high effort voices into breathy voices.

The APLP algorithm does not carry out the typical ideal of separating out the vocal tract filter and the glottal source. Practically speaking, when analyzing a particular voice signal, it is usually impossible to determine how much of the change in the spectral emphasis is due to a change in the glottal source versus the vocal tract filter. What is known is that, with increases in vocal effort, both the glottal source and the vocal tract filter emphasize higher frequencies. For this reason, we created the spectral emphasis filter to capture the combined influence of both the glottal source and the vocal tract filter on the overall spectral envelope of the voice. While this differs from the typical goal of separating the glottal source and the vocal tract, the spectral emphasis filter provides an easier way to manipulate the perceived vocal effort independent of the formant filter.

There are two key differences between the APLP algorithm presented here and other implementations of adaptive pre-emphasis for voice source analysis [21]–[23]. First, voice source analysis extracts estimates of the glottal pulses whereas APLP extracts a spectrally flat excitation and a spectral emphasis filter. This spectral emphasis filter is distinct from the narrow spectral peaks associated with the perception of formants. Second, current methods of voice source analysis operate on frequencies no higher than 5 kHz while APLP works with broader bandwidth speech signals.

The APLP algorithm requires a more complex model of the spectral emphasis. In standard methods of voice source analysis, the spectrum of the pre-emphasis has a simple slope, but in APLP the pre-emphasis can have a more complex shape. This is because APLP aims to analyze musical voice signals and manipulate them in a way that is musically relevant. In doing this, frequencies above 5 kHz are important because they affect the aesthetics of the voice signal. However, when wider bandwidth signals are involved, the spectral emphasis no longer looks like a simple slope. This is likely because the frequency response of the vocal tract often drops off sharply beyond 4–5 kHz [13].

In addition, APLP does not require the ideal retention of phase information. Voice source analysis aims to extract realistic estimates of the glottal pulses [24]. This means that the original signal must retain phase information to prevent the extracted glottal pulses from looking distorted in the time domain. In contrast, the APLP algorithm presented here is not intended to extract the shapes of the glottal pulses. APLP presented here focuses on musical signals where the absolute phase information is less important.

There are other techniques for voice transformation that are capable of manipulating the perceived vocal effort. One such technique analyzes and resynthesizes the aperiodic component of the voice source [25], [26]. This technique carries out LP to estimate the formant filter before extracting the estimated source. As will be shown in this paper, constant pre-emphasis

LP results in an estimated formant filter that captures changes in the spectral emphasis of the voice. Using APLP instead of standard LP in this technique would result in a formant filter that is more consistent across varying voice qualities. In addition, APLP would provide an estimate of the spectral emphasis of the voice. This would be useful in the application of any further aperiodic manipulations to the voice quality. Lastly, the spectral emphasis filter estimated by APLP would provide a valuable starting point for any further manipulations to the spectral envelope of the source.

The aforementioned technique of aperiodic analysis and synthesis [26] is able to modify the perceived vocal effort. The type of vocal effort presented in that technique is different than the type of vocal effort manipulated by APLP in this paper. In the aperiodic synthesis, the perceived vocal effort is primarily modified by increasing variation in the aperiodic component. This results in a voice with more roughness or harshness. This roughness is associated with vocal effort. However, APLP as presented here focuses on transforming voices that do not sound rough or harsh. In the absence of these vocal aperiodicities, vocal effort is primarily manipulated by changing the spectral emphasis.

A different adaptive technique has been developed to manipulate voice timbre using harmonics plus noise analysis [27]. APLP differs from this technique in that APLP uses a source-filter approach to manipulate the voice.

The paper is organized as follows. Section II describes how LP operates and how the pre-emphasis essentially specifies the spectral envelope of the estimate glottal source. This includes an experiment that illustrates how constant pre-emphasis LP results in a formant filter that influences the perception of breathiness and vocal effort. Section III presents APLP as a method to estimate a more consistent formant filter across varying voice qualities. This includes a discussion of tradeoffs in the application of APLP to wider bandwidth voice signals. Section IV presents the APLP voice transformation algorithm and a listening experiment to validate whether APLP is able to manipulate the perception of vocal effort. This is followed by the Conclusion in Section V.

II. Influence of Pre-Emphasis on Estimated Glottal Source

The source-filter model of the voice separates the glottal source from the filtering influence of the vocal tract. However, when there are dramatic changes in the spectral envelope of the voice, constant pre-emphasis implementations of LP do not appropriately separate the source and the filter. This happens because the constant pre-emphasis specifies a constant spectral envelope for the estimated glottal source. Meanwhile, the actual glottal source varies. This section presents an explanation why the pre-emphasis controls the spectral envelope of the estimated glottal source.

The operation of LP is shown in Fig. 2(a). LP estimates a filter that approximates the spectral envelope of the signal. Inverse filtering the signal with the filter results in a spectrally flat signal. To make LP correspond more closely to the linear model of the voice [Fig. 2(b)], a pre-emphasis is typically applied as seen in Fig. 2(c). This pre-emphasis specifies that the estimated glottal signal, $\hat{G}'(z)$, will have a spectral slope that, on average, represents what would be expected according to voice physiology.

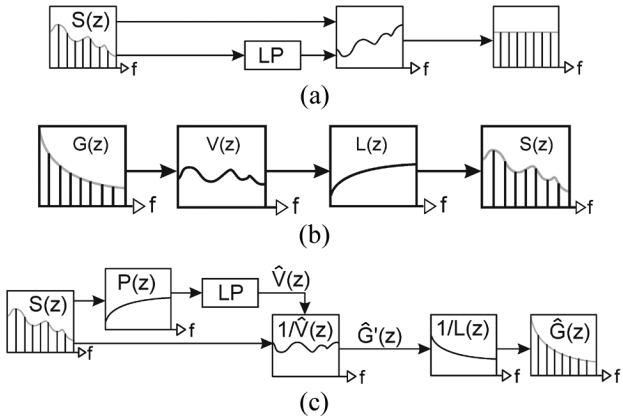

Fig. 2. (a) Linear prediction (LP) algorithm. (b) Linear model of speech production. (c) Using LP to estimate a filter representing the influence of the vocal tract $\hat{V}(z)$ and to extract the corresponding signal representing the glottal source $\hat{G}(z)$.

Mathematically, the estimated glottal signal has a spectral slope that is the inverse to the frequency response of the pre-emphasis filter. This can be demonstrated in the following way.

The linear model of the voice is as follows [Fig. 2(b)]:

$$G(z)V(z)L(z) = S(z). \tag{1}$$

$G(z)$ represents the flow signal from the glottal source (the volume-velocity wave). $V(z)$ is a filter that represents the influence of the vocal tract. $L(z)$ represents the influence of lip radiation with a filter: z transform $(1 - z^{-1})$ [28]. $S(z)$ is the acoustic pressure signal received at the ear.

Working with the derivative of the glottal airflow $G'(z)$ makes it easier to see the features of the glottal pulses. When we talk about the glottal source in this paper, we are referring to $G'(z)$. Since the filter for lip radiation $L(z)$ approximates taking the derivative [28], (1) can be simplified as

$$G'(z)V(z) = S(z). \tag{2}$$

LP analyzes $S(z)P(z)$ to estimate $\hat{V}(z)$, where $P(z)$ is the chosen pre-emphasis filter

$$S(z)P(z) = \hat{V}(z)E(z). \tag{3}$$

LP estimates an all-pole vocal tract filter $\hat{V}(z)$ that matches the spectral envelope of $S(z)P(z)$. The excitation signal $E(z)$ represents the spectrally flat excitation signal that LP would estimate through inverse filtering. Note that for the excitation signal $E(z)$ to be spectrally flat, the estimated vocal tract filter $\hat{V}(z)$ and pre-emphasis $P(z)$ must appropriately fit the voice signal $S(z)$. Typically, LP is of sufficient order in estimating the formant filter $\hat{V}(z)$ that the resulting excitation $E(z)$ appears spectrally flat.

In the application of LP to voice analysis, the original signal $S(z)$ is inverse filtered instead of $S(z)P(z)$. This filtering

process can be seen in Fig. 2(c). By manipulating (3), the result of inverse filtering can be shown to be

$$S(z)/\hat{V}(z) = E(z)/P(z). \tag{4}$$

Since the excitation signal $E(z)$ is spectrally flat, the extracted residual $S(z)/\hat{V}(z)$ has a spectral slope that is inverse to the spectral slope of the pre-emphasis $P(z)$.

From (2), we can also see how to estimate the glottal source $\hat{G}'(z)$

$$S(z)/\hat{V}(z) = \hat{G}'(z). \tag{5}$$

Combining (4) and (5) shows that the estimated glottal source $\hat{G}'(z)$ has a spectral slope that is inverse to the spectrum of the pre-emphasis $P(z)$

$$\hat{G}'(z) = E(z)/P(z). \tag{6}$$

Constant pre-emphasis is commonly used in voice analysis. This enforces a constant spectral envelope for the estimated glottal source, as seen in (6). In contrast, the spectral envelope of the actual glottal source varies through time. This means that the estimated source does not capture variation in the spectral envelope of the actual glottal source. Instead, this variation is captured by the LP filter, $\hat{V}(z)$. In constant pre-emphasis LP, the influence of varying vocal effort is entangled in the estimated vocal tract filter $\hat{V}(z)$ making it difficult to separate out and control the influence of effort.

In fixed-rate LP analysis, the pre-emphasis fully specifies the spectral envelope of the estimated glottal source. However, in closed-phase analysis, some of the variation in the spectral envelope does affect the estimated glottal source. This is because closed-phase analysis excludes large spikes in the time domain that occur at the instances of glottal closure. The amplitude of these spikes vary with the rate of glottal closure, and this is closely related to the perception of vocal effort. With higher levels of vocal effort, the spikes are larger, and this tends to flatten the voice spectrum. Since these spikes are excluded from analysis, their influence is passed on to the estimated glottal source $\hat{G(z)}$ during inverse filtering. This results in an estimated glottal source that somewhat follows variation in the spectral slope of the voice signal. With careful preparation, closed-phase analysis can be used to make good estimates of the spectral slope of the glottal source [29]. That said, in at least one case, adaptive pre-emphasis has been used to improve closed-phase analysis [23].

We are using fixed-rate analysis because we cannot depend upon the source data being accurate enough to estimate reasonable glottal pulses. Therefore, it is safe to say that the pre-emphasis fully specifies the spectral slope of the estimated glottal source.

If we want to control the perceived vocal effort, it is necessary to estimate formant filters that remain consistent across varying voice qualities. The theoretical analysis above suggests that constant pre-emphasis LP results in estimated formant filters that capture changes in the spectral emphasis of the voice. We carried out a listening experiment to determine whether the formant filters estimated by constant pre-emphasis LP affect the perceived breathiness and vocal effort [30]. The experiment compares constant pre-emphasis LP filters extracted from breathy and high effort voices.

TABLE I
ORIGINAL VOICE SAMPLES FOR CONSTANT PRE-EMPHASIS LP EXPERIMENT

Singer	Female A		Male A		Male B	
Tessitura	mezzo-soprano		baritone		baritone	
Vowel	[e]		[a]		[a]	
Note	A#3		G#3		A2	
Phonation	breathy	h.e.*	breathy	h.e.	breathy	h.e.
Pitch (Hz)	237	237	210	208	111	112
F1 (Hz)	475	475	630	830	610	670
F2 (Hz)	1900	1900	1270	1240	1230	1230

* h.e. = high effort

The experiment used pairs of samples where the same voice sang the same vowel at the same pitch while varying between breathiness and high vocal effort. To remove the influence of the estimated voice source, LP filters were estimated from the two samples using constant pre-emphasis LP. These LP filters were then excited with the same artificial source. This resulted in two synthesized samples where the only difference between them was their associated constant pre-emphasis LP filters. One of the LP filters was extracted from a high-effort voice and the other LP filter was extracted from a breathy voice.

In the experiment, listeners rated the difference in perceived vocal effort and breathiness between the LP filters. If the LP filters from the high-effort and breathy voices were perceived to be the same, then it should be possible to transform the voices by manipulating only the estimated voice source. However, if the LP filters were perceived to be different, then it would be difficult to change the perceived breathiness and vocal effort by modifying the estimated voice source alone.

Three pairs of voice samples were used in the experiment. The characteristics of the extracted vowels are summarized in Table I. LP was carried out on the voice samples with a constant pre-emphasis filter with z-transform $(1 - 0.99z^{-1})$. This resulted in two LP filters, one representing a high-effort voice and one representing a breathy voice. To compare the LP filters, both LP filters were excited with the same Liljencrants–Fant (LF) model for the voice source [31].

The perceptual criteria for this test was drawn from other studies for evaluating breathy voices [29], [32] and a prior test that we conducted [33]. The listener was given the following question: "Listen to the two samples and rate which one sounds more breathy." After they evaluated all of the sample pairs, they were given the next question: "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 samples were rated on a seven-point scale with -3 being much less breathy and $+3$ being much more breathy.

The listener did not know which sample pairs were being provided or the order in which they were presented. Within each

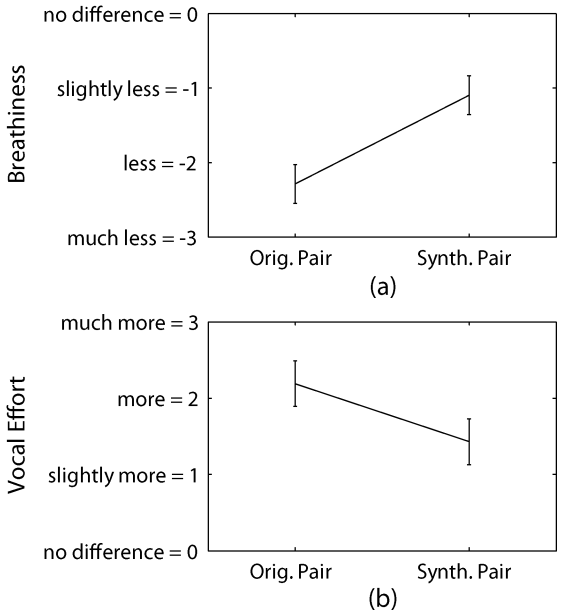

Fig. 3. Plot of the relative difference in (a) perceived breathiness and (b) perceived vocal effort within each sample pair. 95% confidence intervals have been plotted. "Orig. Pair" represents the rating of the original high-effort voice relative to the original breathy voice. "Synth. Pair" represents 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.

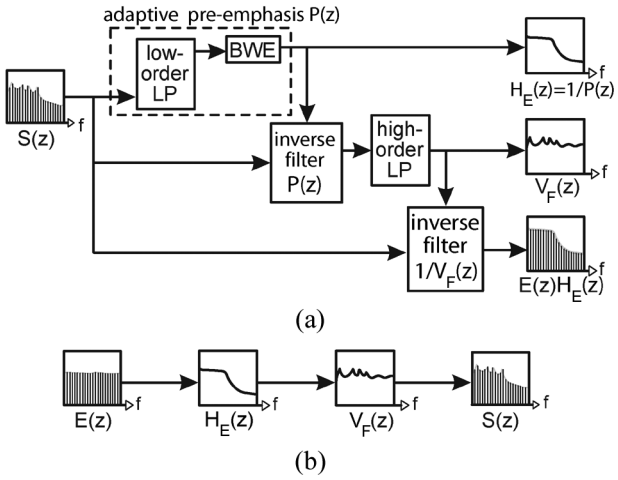

Fig. 4. (a) APLP analysis. (BWE refers to bandwidth expansion.) (b) Linear model of the voice resulting from APLP.

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. There were a total of seven listeners.

A test for statistical significance was carried out. The F-test on the breathiness ratings resulted in an F-value of 22.0, indicating that there is less than 0.01% chance that the observed differences occurred due to noise. The F-test on the vocal effort ratings resulted in an F-value of 6.8, indicating that there is only a 1.4% chance that the observed differences occurred due to noise.

In the presentation of the results, the high effort voice is rated with respect to the breathy voice. The results show that the original pair of samples (unmodified) exhibit a large contrast in breathiness as shown in Fig. 3(a). On average, the high-effort sample sounds between less breathy and much less breathy than the corresponding breathy sample. The synthesized pair provides the comparison between the estimated formant filters while using the same artificial excitation to eliminate the influence of the estimated source. When the filters from the high-effort and breathy samples are compared, the contrast in breathiness is reduced but not eliminated. On average, the high-effort filter sounds slightly less breathy than the corresponding breathy filter. Even when the high-effort filter and breathy filter are provided with the same excitation, some of the perception of breathiness remains.

The results show that the original pair of samples (unmodified) exhibit a large contrast in vocal effort as shown in Fig. 3(b). On average, the high-effort sample is perceived to have between more effort and much more effort than the corresponding breathy voice sample. When the high effort filters and the breathy filters are compared, the contrast in vocal effort is reduced but not eliminated. On average, the high-effort filter sounds like it has between more and slightly more effort than the corresponding breathy filter. Even when the high-effort filter and breathy filter are provided with the same excitation, much of the perception of vocal effort remains.

This experiment demonstrated that constant pre-emphasis LP does not fully separate the perception of breathiness and vocal effort from the LP filter. Instead, the LP filter contains some of the perception of breathiness and vocal effort. Further details about the experiment can be found in the original paper [30]. To appropriately separate variations in vocal effort and breathiness from the LP filter, we propose adaptive pre-emphasis, which is presented in the following section.

III. Adaptive Pre-Emphasis LP Algorithm

As we described above, constant pre-emphasis LP does not appropriately model the glottal source, especially in voices with varying levels of vocal effort. In this section, we present APLP as a way to address this problem. The APLP algorithm presented here separates the voice signal into a spectrally flat excitation, a spectral emphasis filter and formant filter. The purpose of this separation is to make it easier to manipulate the spectral emphasis filter and thereby change the perceived voice quality.

The algorithm for APLP analysis is presented in Fig. 4(a). The difference between APLP and the constant pre-emphasis LP can be seen by comparing it to Fig. 2(c). In APLP, the pre-emphasis filter $P(z)$ is estimated using low-order LP. This enables the pre-emphasis filter to track variations in the spectral envelope of the voice signal. The voice signal $S(z)$ is inverse filtered with the pre-emphasis $P(z)$ to spectrally flatten the signal before the second stage of LP. This second stage of LP captures the formant information in a filter $V_F(z)$ using a higher order for LP. Because the spectral emphasis is removed before the second stage, the estimated formant filter $V_F(z)$ from APLP is more spectrally flat than the estimated formant filter $\hat{V}(z)$ from constant pre-emphasis LP. In contrast, constant pre-emphasis LP causes variation in the spectral emphasis to be included in the estimated vocal tract filter $\hat{V}(z)$.

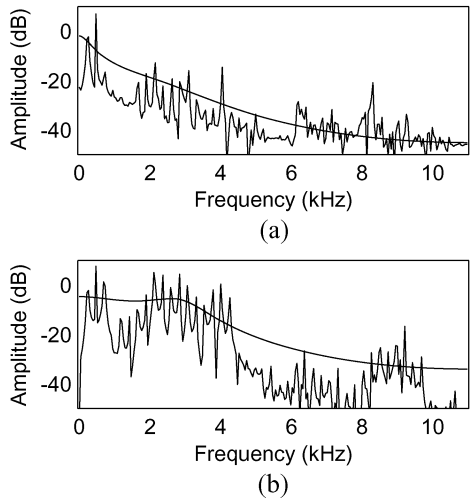

Fig. 5. Voice spectra from (a) a breathy voice and (b) a high-effort voice. In each plot, the same voice is singing the same vowel on the same pitch. The spectral emphasis filter ($H_E(z)$) estimated by low-order LP has also been plotted.

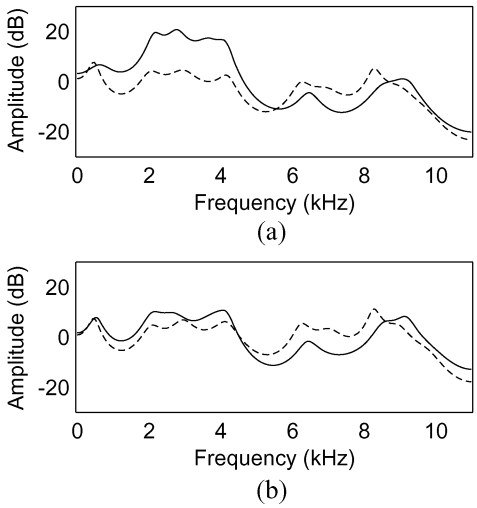

Fig. 6. Formant filters estimated using (a) constant pre-emphasis LP and (b) APLP. The same voice sang the same vowel at the same pitch while varying the quantity of vocal effort. The dashed lines are estimated formant filters from the breathy voice while the solid lines are estimated formant filters from the high-effort voice. Note that the formant filters for breathy and high-effort voices are more similar for APLP than for constant pre-emphasis LP. The sampling rate is 22 050 Hz, the formant filter has an order of 30, and the spectral emphasis filter has an order of 3.

The algorithm for APLP analysis, presented in Fig. 4, produces the following model of the voice:

$$E(z)H_E(z)V_F(z) = S(z) \tag{7}$$

which appears quite different from the typical linear model of the voice as presented in (1) and (2). A spectrally flat excitation $E(z)$ is shaped by the spectral emphasis filter $H_E(z)$ which is further shaped by the formant filter $V_F(z)$ to produce the sound received by the ear $S(z)$. Since the spectral emphasis filter $H_E(z)$ models the overall spectral emphasis, the formant filter $V_F(z)$ becomes more spectrally flat. The formant filter $V_F(z)$ captures the narrow peaks in the spectrum while the spectral emphasis filter $H_E(z)$ captures the overall shape of the spectrum. The perception of high vocal effort is caused by changes to both the glottal source and the vocal tract filter. In the APLP algorithm presented here, we lump both of these changes into one spectral emphasis filter $H_E(z)$.

When working with voices sampled at or below 10 kHz, the spectral emphasis filter has a spectrum that looks like a simple slope. However, voice signals sampled at higher frequencies often exhibit a trend that does not look like a spectral slope. Instead, for high effort voices, the spectrum can look relatively flat up to 4.5 kHz and then drop off suddenly beyond that range. If we estimate a pre-emphasis filter using LP with an order larger than one, there is no longer a guarantee that the pole of the filter will be at 0 Hz. For breathy voices, the regime of the voice appears relatively consistent throughout the frequency range and the spectral emphasis filter $H_E(z)$ has a spectrum that looks like a slope as seen in Fig. 5(a). However, for high-effort voices, a pole of the filter ends up being at an intermediate frequency as seen in Fig. 5(b). In this situation, $H_E(z)$ no longer looks like the envelope of the glottal source.

In estimating the spectral emphasis filter with LP, it is important to use bandwidth expansion (BWE) [34], [35] as seen in Fig. 4. This ensures that the estimated spectral emphasis filter does not become too peaky. If the spectral emphasis filter $H_E(z)$ becomes too sharply peaked, then it may capture formant information that should, instead, be captured by the formant filter $V_F(z)$. In the algorithm presented here, we used radial scaling on the filter coefficients [34].

When voices exhibit high vocal effort, both the vocal tract and the glottal source change in a way that affects the spectral envelope of the voice. The spectral slope of the glottal source steepens in a negative direction and the vocal tract exhibits a drop-off in the spectrum above 4–5 kHz. The spectral emphasis filter $H_E(z)$ with a resonance [Fig. 5(b)] models this drop-off in the spectrum. This means that the spectral emphasis filter is now capturing more than just the spectral slope of the glottal source. This is why the excitation of the formant filter is no longer referred to as the glottal source but is now described as a spectrally flat excitation $E(z)$ shaped by a spectral emphasis filter $H_E(z)$. In addition, $V_F(z)$ has been given the subscript "F" to indicate that this filter models the perceptual influence of the formants and not all of the frequency response of the vocal tract.

In the current application, we are interested in manipulating the perceived vocal effort, and it is convenient that the influence of vocal effort upon the vocal tract is captured in the spectral emphasis filter $H_E(z)$. In this way, the influence of vocal effort is captured in one place rather than being partially represented by the estimated glottal source and partially by the estimated vocal tract filter.

In order to manipulate vocal effort in singing voices, we want to separate the phonetic information (the formant filter) from the information about the perceived breathiness and vocal effort. APLP is useful because it is able to capture the spectral emphasis of the voice $H_E(z)$ independent of the formant filter V_F. As a result, APLP provides a more consistent estimate of the formant filter than constant pre-emphasis LP. This can be observed in Fig. 6, which compares the formant filters from constant pre-emphasis LP and APLP. Because APLP better separates the phonetic information from the spectral emphasis, it is

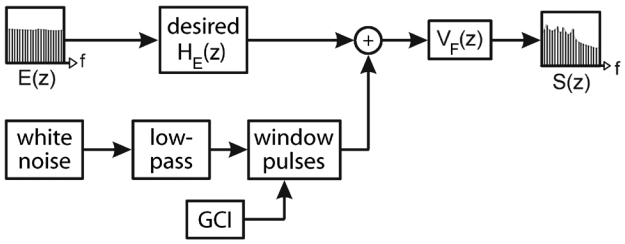

Fig. 7. APLP synthesis configured to modify the perception of vocal effort. The spectrally flat excitation $E(z)$ is shaped by the desired spectral emphasis filter $H_E(z)$. White noise is low-passed to remove the uppermost frequencies and pulsed according to the glottal closure instants (GCI). The noise is added to the spectrally emphasized excitation. The formant filter $V_F(z)$ is then applied to generate the newly synthesized voice signal $S(z)$.

easier to manipulate the spectral emphasis independently of the formant filter. In contrast, constant pre-emphasis LP entangles the spectral emphasis information in the formant filter for signals sampled above 10 kHz [see Fig. 6(b)].

The following section explains how the APLP model of the voice is used to convert high-effort voices into breathy voices that exhibit low effort.

IV. Voice Transformation Algorithm

To transform high-effort singing voices into breathy voices, we need to manipulate the spectral emphasis to change the perceived effort and add noise to simulate breathiness. This section describes the algorithm that carries out this transformation.

To modify the perceived vocal effort, the spectral envelope of the residual has to be modified and resynthesized. The process of resynthesizing the voice is illustrated in Fig. 7. First, the spectrally flat excitation $E(z)$ from the high-effort voice is filtered with the desired spectral emphasis filter $H_E(z)$. The values for this spectral emphasis filter are from a breathy voice. Pulsed white noise is added to the signal to simulate aspiration noise. The resulting signal is a voice source with the desired spectral emphasis plus noise. This signal is then fed through the formant filter $V_F(z)$ from the high-effort voice to synthesize the modified voice.

In the current experiment, we manipulated voice samples at 22 050 Hz using an LP order of 30 for the formant filter and an LP order of 1 or 3 for the spectral emphasis filter. The window size for autocorrelation LP was 416 samples with LP computing filters every 32 samples. Bandwidth expansion was carried out on the LP filters using radial scaling on the filter coefficients [34] with $\alpha = 0.9$ for the spectral emphasis filter $H_E(z)$ and $\alpha = .975$ for the formant filter $V_F(z)$. The perceived influence of pole scaling on the spectral emphasis filter is large while the influence of pole scaling on the formant filter is subtle.

The greatest challenge in synthesizing breathy voices is that the noise does not blend easily into the voice. Instead, listeners typically perceive the noise to be a separate sound distinct from the voice signal. Blending in the noise after the pre-emphasis gives it approximately the correct shape. The noise floor representing breathiness in the voice source $(E(z)H_E(z))$ has an approximately flat spectrum. A first-order, low-pass, all-zero filter $(1 - 0.98z^{-1})$ was also added to reduce some of the uppermost noise frequencies that do not blend easily. Synthesized noise blends more easily when it is pulsed in sync with the glottal closure instants [36]. We used Hann windows to provide the shape of the pulses and a duty cycle of approximately 50%.

Now that the synthesis algorithm has been described, the following section will present a listening experiment that was carried out to demonstrate that APLP is an improvement over constant pre-emphasis LP in transforming high-effort voices into breathy voices.

TABLE II
Original Voice Samples for Voice Transformation Experiment

Singer	Female A		Male A		Female B	
Tessitura	mezzo-soprano		baritone		soprano	
Vowel	[e]		[a]		[a]	
Note	A#3		G#3		G#4	
Phonation	breathy	h.e.*	breathy	h.e.	breathy	h.e.
Pitch (Hz)	237	237	210	208	415	421
F1 (Hz)	475	475	630	830	830	860
F2 (Hz)	1900	1900	1270	1240	1250	1270

* h.e. = high effort

A. Listening Experiment

We have described how the APLP algorithm can transform the spectrum of the voice. This section describes a perceptual experiment to verify whether these changes are subjectively perceivable.

As source data, we had pairs of voice samples where the same person phonated the same vowel at the same pitch but with two different voice qualities: one breathy and one high-effort. The goal was to transform the high-effort sample into the breathy sample. We had three different sample pairs as source data as described in Table II.

We compared the capabilities of three different algorithms to carry out the transformation of high-effort voices into breathy voices: APLP with first-order pre-emphasis, APLP with third-order pre-emphasis, and constant pre-emphasis LP. A high-effort voice without processing was also included in the comparison. The listener rated the effectiveness of the four methods (including no processing) with respect to the target breathy sample for the voice being transformed. The various comparisons made for each sample pair are listed in Table III.

We applied APLP synthesis as described in Fig. 7 to transform high-effort voices into breathy voices. The spectral emphasis $H_E(z)$ used during synthesis was extracted from the target breathy voice in the sample pair. Constant pre-emphasis LP does not estimate a spectral emphasis filter. This is equivalent to using the same spectral emphasis for analysis and synthesis, meaning that the spectral emphasis is not modified. The same quantity of aspiration noise was added to both the APLP algorithm and the constant pre-emphasis LP algorithm.

TABLE III
COMPARISON OF VOICE SAMPLES IN VOICE TRANSFORMATION LISTENING EXPERIMENT

orig. breathy voice	orig. high effort voice (o.h.e.v.)
orig. breathy voice	o.h.e.v. through constant pre-emphasis LP
orig. breathy voice	o.h.e.v. through first order pre-emphasis APLP
orig. breathy voice	o.h.e.v. through third order APLP

The onset for high-effort voices is typically much faster than the onset for breathy voices. A steady-state section was extracted from the center of each voice sample to eliminate the influence of voice onsets upon the perceptual rating.

There were 16 listeners in total. The listeners for the experiment were audio engineers with experience in voice processing (ten listeners), trained linguists (three listeners), and experienced singers (three listeners). The processed voice samples were rated relative to benchmark breathy samples. The unprocessed high-effort voice samples were also rated relative to benchmark breathy samples. The variously processed sample pairs were presented in a random order. In addition, the samples were randomly presented before or after the benchmark breathy samples without specifying the order to the listener.

The listeners went through the experiment three times to make three different ratings.

- BREATHINESS: Please listen to the two samples and rate how much more BREATHY one sample sounds than the other sample. BREATHINESS corresponds to a soft, relaxed voice.
- VOCAL EFFORT: Please listen to the two samples and rate how much more EFFORT is required on the part of the singer to generate one sample rather than the other sample. VOCAL EFFORT corresponds to a strained or tense voice.
- ARTIFICIALNESS: All of the samples have been digitally modified in some way. Please listen to the two samples and rate how ARTIFICIAL one sample sounds than the other.

In each iteration of the experiment, the order of the samples was re-randomized. The relative rating was on a seven-point scale as per ITU Standard 1284 [37].

The results of the experiment are presented in Fig. 8. A test for statistical significance was carried out. The F-test on the breathiness, vocal effort, and artificialness ratings resulting in an F-values of 16.7, 16.1, and 10.0 respectively. This indicates that there is less than a 0.01% chance that the observed differences could occur due to noise in each of the sets of ratings.

What we found was that all of the processing techniques provided a increase in perceived breathiness [Fig. 8(a)]. The original high-effort voice sounded less breathy than the corresponding breathy voice. After the transformation, all of the voice samples sounded only slightly less breathy than the breathy voice. The third-order APLP algorithm performed better than the first-order APLP and constant pre-emphasis algorithms. This is likely because the third-order APLP algorithm models the drop-off in the high-effort voice spectrum, whereas the first-order APLP algorithm estimates a spectral emphasis that is similar in shape to the constant pre-emphasis filter.

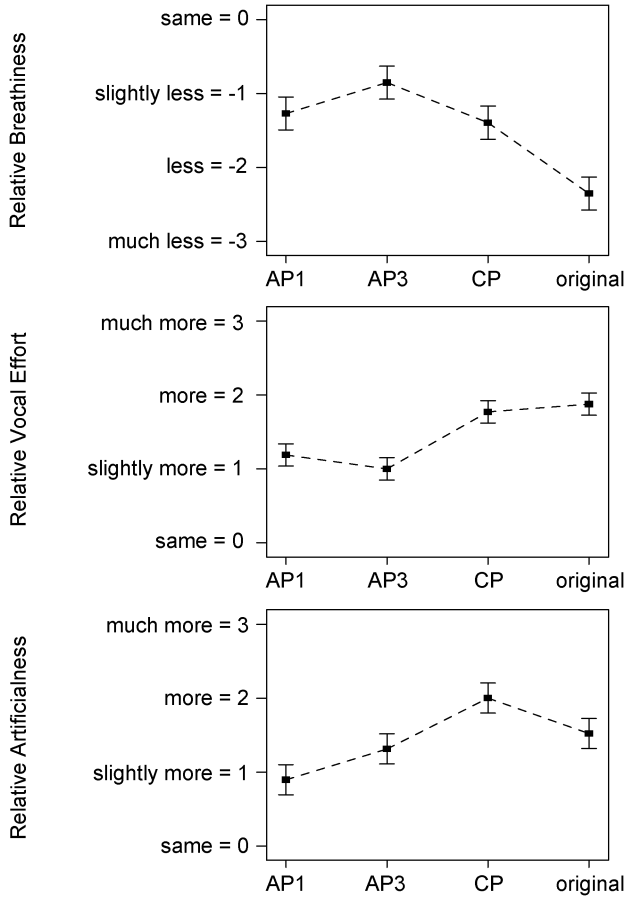

Fig. 8. Statistical results from relative ratings of breathiness (top), vocal effort (middle), and artificialness (bottom). The horizontal axis represents the processing applied to the high-effort voice: AP1 = first-order adaptive pre-emphasis, AP3 = third-order adaptive pre-emphasis, CP = constant pre-emphasis, and original = no processing. All samples were rated relative to a corresponding breathy voice. Each point represents three data sets rated by 16 listeners. 95% confidence intervals have been plotted.

The similar ratings for breathiness between the APLP and the constant pre-emphasis LP algorithms might give the impression that APLP is not much more effective than constant pre-emphasis LP. However, there are other factors to consider. A breathy voice should exhibit low effort and the transformation should, ideally, be free of unnatural artifacts.

The APLP algorithm was able to reduce the perceived effort of the voice more effectively than constant pre-emphasis LP [Fig. 8(b)]. The constant pre-emphasis LP algorithm exhibited nearly the same amount of vocal effort as the original high-effort voice. In contrast, the first- and third-order APLP algorithms reduced the perceived vocal effort almost all the way down to the breathy benchmark voices.

The best performance of the APLP algorithm was in the perceptual rating of artificialness [Fig. 8(b)]. Constant pre-emphasis LP sounded more artificial than the original high-effort voice. In contrast, APLP sounded less artificial. This is because APLP was able to transform the spectral emphasis of the voice to more closely match the target breathy voice.

The APLP algorithm actually sounded less artificial than the original high-effort voice. This is likely because the original high-effort voice exhibited so much effort that it sounded

slightly unnatural. Using APLP on the high-effort voices brought them into a range that people are more used to hearing. This may be the reason why the listeners rated the re-synthesized APLP voices as sounding less artificial than the original high-effort voice.

In conclusion, all of the voice transformation algorithms increased the perceived breathiness with third-order APLP performing the best. However, our goal was not just to make a voice that sounds breathy, but to make a voice that exhibits low-effort and sounds free of artifacts. In the ratings for vocal effort and artificialness, APLP outperformed the constant pre-emphasis algorithm. APLP even sounded less artificial than the original high-effort voice. This indicates that APLP is able to synthesize natural-sounding voices when transforming high-effort voices into breathy voices.

V. Conclusion

This paper presented APLP to estimate a spectral emphasis filter for manipulating the perceived vocal effort of singing voices. APLP also results in an estimated formant filter that is more consistent across varying voice qualities. A listening experiment was carried out to demonstrate that APLP's spectral emphasis filter can be used to transform high-effort voices into breathy voices. This resulted in breathy voices that sounded more relaxed and exhibited fewer artifacts than the corresponding transformation using constant pre-emphasis LP.

The APLP algorithm can be used during voice analysis as an indication of the perceived vocal effort in the voice [38]. Since vocal effort is influenced by the person's emotional state, this technique can be used to analyze the stress in a person's voice, which is a useful application in its own right. The filters extracted with APLP could be manipulated to synthesize new voices with different levels of vocal effort and correspondingly different emotional states.

In this paper, we focused on the contrast between high effort and breathy voices. There are many other voice qualities that exhibit range of spectral envelopes. APLP could be used to manipulate the spectral envelope to match any of these voice qualities. APLP could also be enhanced with a complementary system to control voice aperiodicities [26]. This could be useful in generating other types of voices such as creaky voice and harsh voice.

There are a few ways that the APLP algorithm could be improved. We used low-order LP to estimate the spectral emphasis filter. However, for high-effort voices, low-order LP does not model the steepness of the drop-off around 4–5 kHz [see Fig. 5(b)]. A spectral emphasis filter that more closely matches the drop-off might result in a more effective filter for controlling vocal effort. Second, we were able to apply APLP for signals sampled up to 22 kHz. Implementing a warped [39] version of APLP could result in an algorithm that works for even wider bandwidth signals.

The transformation algorithm presented here focused on isolated vowels. In order to transform phrases, it would be necessary to carry out voiced-unvoiced detection and to apply APLP only to the voiced sections. Since voices with high vocal effort exhibit a stronger attack, it is likely that this attack would have to be controlled. In addition, the influence of the APLP algorithm may not be consistent across different vowels. When a voice changes from one vowel to another, some of the spectral changes are similar to changes that occur between high and low effort voices. For example, in changing between higher and lower vowel qualities, there is an increase in the first formant frequency that also increases the amplitude of the higher formants [40]. In implementing APLP for phrases, it may be necessary to adjust for these changes between different vowel qualities.

Acknowledgment

The authors would like to thank TC-Helicon for providing equipment to use in the experiments. They would also like to thank J. Esling for providing valuable background information about phonetics, L. A. Bateman for providing voice samples from her research [41] and advice regarding the analysis of singing voices, G. Rutledge, who provided continuing mentorship through the early stages of this project, and the reviewers who provided advice that significantly improved this paper.

References

[1] M. W. Macon, L. Jensen-Link, J. Oliverio, M. A. Clements, and E. B. George, "A singing voice synthesis system based on sinusoidal modeling," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'97)*, Munich, Germany, Apr. 1997, vol. 1, pp. 435–438.

[2] J. Laver, *The Phonetic Description of Voice Quality*. New York: Cambridge Univ. Press, 1980.

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

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

[5] G. D. Allen, "Acoustic level and vocal effort as cues for the loudness of speech," *J. Acoust. Soc. Amer.*, vol. 49, no. 6, pp. 1831–1841, Jun. 1971.

[6] G. Fairbanks and M. S. Miron, "Effects of vocal effort upon the consonant-vowel ratio within the syllable," *J. Acoust. Soc. Amer.*, vol. 29, no. 5, pp. 621–626, May 1957.

[7] A. Andersson, A. Eriksson, and H. Traunmüller, "Cries and whispers: Acoustic effects of variations of vocal effort," KTH, Stockholm, Sweden, Speech, Music and Hearing Quarterly Progress and Status Report (TMH-QPSR), 1996, pp. 127–130.

[8] B. Granström and L. Nord, "Neglected dimensions in speech synthesis," *Speech Commun.*, vol. 11, pp. 459–462, 1992.

[9] S. Ternström, M. Bohman, and M. Södersten, "Loud speech over noise: Some spectral attributes, with gender differences," *J. Acoust. Soc. Amer.*, vol. 119, no. 3, pp. 1648–1665, Mar. 2006.

[10] A. N. Chasaide and C. Gobl, "Voice source variation," in *Hand of Phonetic Sciences*, W. J. Hardcastle and J. Laver, Eds. Oxford, U.K.: Blackwell, 1997, pp. 427–461.

[11] B. Doval, C. d'Alessandro, and N. Henrich, "The voice source as a causal/anticausal linear filter," in *Proc. Voice Quality: Functions, Anal. Synth., ISCA Workshop (VOQUAL'03)*, Geneva, Switzerland, Aug. 2003, pp. 15–20.

[12] B. Doval, C. d'Alessandro, and N. Henrich, "The spectrum of glottal flow models," *Acustica United With Acta Acustica*, vol. 92, pp. 1026–1046, 2006.

[13] H. Imagawa, K.-I. Sakakibara, N. Tayama, and S. Niimi, "The effect of the hypopharyngeal and supra-glottic shapes on the singing voice," in *Proc. Stockholm Music Acoust. Conf. (SMAC'03)*, Stockholm, Sweden, Aug. 2003, pp. 471–474.

[14] J. Sundberg, *The Science of the Singing Voice*. Dekalb, IL: Northern Illinois Univ. Press, 1987.

[15] J. H. Esling and J. G. Harris, "Expanded taxonomy of states of the glottis," in *Proc. 15th Int. Congr. Phonetic Sci.*, 2003, vol. 1, pp. 1049–1052.

[16] J. H. Esling, "The laryngeal sphincter as an articulator: How register and phonation interact with vowel quality and tone," in *Proc. Western Conf. Linguist.*, Nov. 2002, pp. 68–86, UBC.

[17] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*. New York: Springer-Verlag, 1976.
[18] J. Picone, "Signal modeling techniques in speech recognition," *Proc. IEEE*, vol. 81, no. 9, pp. 1215–1247, Sep. 1993.
[19] 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 Audio Process.*, vol. 8, no. 4, pp. 429–442, Jul. 2000.
[20] M. Heldner, "Spectral emphasis as an additional source of information in accent detection," in *Proc. Prosody 2001: ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*, Red Bank, NJ, Oct. 2001, pp. 57–60.
[21] P. Alku, "An automatic method to estimate the time-based parameters of the glottal pulseform," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'92)*, San Francisco, CA, Mar. 1992, pp. 29–32.
[22] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Commun.*, vol. 11, pp. 109–118, 1992.
[23] H. R. Pfitzinger, "Influence of differences between inverse filtering techniques on the residual signal of speech," in *Proc. DAGA'05*, Munich, Germany, Mar. 2005, vol. 1, pp. 223–224.
[24] D. Y. Wong, J. D. Markel, and A. H. Gray, "Least squares glottal inverse filtering from the acoustic speech waveform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 4, pp. 350–355, Aug. 1979.
[25] C. d'Alessandro, V. Darsinos, and B. Yegnanarayana, "Effectiveness of a periodic and aperiodic decomposition method for analysis of voice source," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 1, pp. 12–23, Jan. 1998.
[26] G. Richard and C. d'Alessandro, "Analysis/synthesis and modification of the speech aperiodic component," *Speech Commun.*, vol. 19, pp. 221–224, 1996.
[27] F. Thibault and P. Depalle, "Adaptive processing of singing voice timbre," in *Proc. Can. Conf. Elect. Comput. Eng. (CCECE 2004)*, Niagara Falls, ON, Canada, May 2004, pp. 871–874.
[28] J. L. Flanagan, *Speech Analysis Synthesis and Perception*. New York: Springer-Verlag, 1972.
[29] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *J. Acoust. Soc. Amer.*, vol. 90, no. 5, pp. 2394–2410, 1991.
[30] K. I. Nordstrom, P. F. Driessen, and G. A. Rutledge, "Influence of the LPC filter upon the perception of breathiness and vocal effort," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT06)*, Vancouver, BC, Canada, Aug. 2006, pp. 23–27.
[31] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," Speech, Music, and Hearing Quarterly Progress and Status Rep. (STL-QPSR), 1985, vol. 4, pp. 1–13.
[32] D. H. Klatt and L. C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Amer.*, vol. 87, no. 2, pp. 820–857, Feb. 1990.
[33] K. I. Nordstrom, G. A. Rutledge, and P. F. Driessen, "Using voice conversion as a paradigm for analyzing breathy singing voices," in *Proc. Pacific Rim Conf. Commun., Comput. Signal Process. (PACRIM05)*, Victoria, BC, Canada, 2005, pp. 428–431.
[34] P. Kabal, "Ill-conditioning and bandwidth expansion in linear prediction of speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'03)*, Hong Kong, 2003, pp. 824–827.
[35] P. Kabal, "Ill-Conditioning and Bandwidth Expansion in Linear Prediction of Speech," TSP Lab, Dept. Elect. Eng., McGill Univ., Montreal, QC, Canada, Oct. 2003, Tech. Rep..
[36] D. J. Hermes, "Synthesis of breathy vowels: Some research methods," *Speech Commun.*, vol. 10, pp. 497–502, 1991.
[37] *General methods for the subjective assessment of sound quality*, Standard 1284, ITU, Geneva, Switzerland, 1997.
[38] G. Zhou, J. H. L. Hansen, and J. F. Kaiser, "Methods for stress classification: Nonlinear TEO and linear speech based features," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'99)*, Phoenix, AZ, Mar. 1999, vol. 4, pp. 2091–2094.
[39] A. Harma and U. K. Laine, "A comparison of warped and conventional linear predictive coding," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 579–588, Jul. 2001.
[40] K. N. Stevens, *Acoustic Phonetics*. Cambridge, MA: MIT Press, 2000.
[41] L. A. Bateman, "Soprano, style and voice quality: Acoustic and laryngographic correlates," M.S. thesis, Univ. of Victoria, Victoria, BC, Canada, 2004.

Karl I. Nordstrom (SM'05) received the B.Eng. and M.A.Sc. degrees in mechanical engineering from the University of Victoria, Victoria, BC, Canada, in 1995 and 2000, respectively. His M.A.Sc. work was in evaluating the performance of full-suspension mountain bikes in collaboration with Rocky Mountain Bicycles in Vancouver, BC. He is currently pursuing the Ph.D. degree in electrical engineering, also at the University of Victoria. His Ph.D. work is in the transformation of musical voice signals in collaboration with IVL Technologies and TC-Helicon, two voice-related companies in Victoria, BC.

Prior to the M.A.Sc., he built a destructive testing system for evaluating bicycle components for Rocky Mountain Bicycles. Prior to the Ph.D. degree, he carried out noise and vibration testing as well as other analysis work at Western Star Trucks, then located in Kelowna, BC, Canada. In parallel with the Ph.D. degree, he has been working on commercial voice transformation algorithms at TC-Helicon.

George Tzanetakis (S'98–M'02) received the B.Sc. degree in computer science from the University of Crete, Heraklion, Greece, and the M.A and Ph.D. degrees in computer science from Princeton University, Princeton, NJ. His Ph.D. work involved the automatic content analysis of audio signals with specific emphasis on processing large music collections.

In 2003, he was a Postdoctoral Fellow at Carnegie Mellon University, Pittsburgh, PA, working on query-by-humming systems, polyphonic audio-score alignment, and video retrieval. Since 2004, he has been an Assistant Professor of Computer Science (also cross-listed in Music and Electrical and Computer Engineering) at the University of Victoria, Victoria, BC, Canada. His research deals with all stages of audio content analysis such as analysis, feature extraction, segmentation, and classification, with specific focus on music information retrieval (MIR). He is the Principal Designer and Developer of the open source Marsyas audio processing software framework.

Dr. Tzanetakis is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and the *Computer Music Journal*. His work on musical genre classification is frequently cited, and he received an IEEE Signal Processing Society Young Author Award in 2004.

Peter F. Driessen (SM'93) received the B.Sc. degree in physics and the Ph.D. degree in electrical engineering from the University of British Columbia, Vancouver, BC, Canada, in 1976 and 1981, respectively.

He was with various companies in Vancouver designing modems for five years before joining the University of Victoria, Victoria, BC, where he is currently a Professor at the Department of Electrical and Computer Engineering and cross-appointed to the Department of Computer Science and the School of Music. He spent two sabbaticals and several summers at AT&T Bell Laboratories, Holmdel, NJ, working on various aspects of wireless communications systems. He has more than 100 technical publications and 11 patents. His research interests include audio and video signal processing, computer music, sound recording, wireless communications, and radio propagation.

Dr. Driessen was an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 1999 to 2004.