



Estimating lower vocal tract features with closed-open phase spectral analyses

Elizabeth Godoy, Nicolas Malyska, Thomas F. Quatieri

MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420

{elizabeth.godoy, nmalyska, quatieri}@ll.mit.edu

Abstract

Previous studies have shown that, in addition to being speaker-dependent yet context-independent, lower vocal tract acoustics significantly impact the speech spectrum at mid-to-high frequencies (e.g. 3-6kHz). The present work automatically estimates spectral features that exhibit acoustic properties of the lower vocal tract. Specifically aiming to capture the cyclicity property of the epilarynx tube, a novel multi-resolution approach to spectral analyses is presented that exploits significant differences between the closed and open phases of a glottal cycle. A prominent null linked to the piriform fossa is also estimated. Examples of the feature estimation on natural speech of the VOICES multi-speaker corpus illustrate that a salient spectral pattern indeed emerges between 3-6kHz across all speakers. Moreover, the observed pattern is consistent with that canonically shown for the lower vocal tract in previous works. Additionally, an instance of a speaker's formant (i.e. spectral peak around 3kHz that has been well-established as a characteristic of voice projection) is quantified here for the VOICES template speaker in relation to epilarynx acoustics. The corresponding peak is shown to be double the power on average compared to the other speakers (20 vs 10 dB).

Index terms: spectral features, epilarynx, speaker's formant

1. Introduction

Though typically not the focus of speech acoustics, there are important spectral characteristics of voiced speech at mid-to-high frequencies (e.g. above 3kHz). For instance, the work in [1] showed that high-frequency energy levels of the long term average spectrum varied based on production level (soft vs loud) as well as production mode (singing vs speech). Additionally, as more data sources move from narrow to wideband channels (e.g. VoIP, video, 5G cellular, etc), questions about spectral information at higher frequencies are becoming increasingly relevant to several speech technologies. For example, in the context of speaker recognition, the work in [2] illustrated that increasing bandwidth from 4kHz (traditional telephony) to 8kHz reduced error by up to 50 percent, with a more profound impact for female speakers. However, there was no discussion in [2] of speaker-specific features in the wideband speech that helped lead to improved performance.

While significant high-frequency energy is known to be a distinguishing feature of voiceless fricatives [1], there is also speaker-dependent information in voiced speech at frequencies around and above 4kHz. Furthermore, these characteristics are not simply higher order resonances of the vocal tract proper (i.e.

pharynx, oral and nasal cavities). Specifically, resonances of the lower vocal tract (i.e. epilarynx and piriform cavities) have been shown to significantly impact the speech spectrum at mid-to-high frequencies (e.g. 3-6 kHz) [3, 4].

Previous speech science studies in [3, 4] have illustrated several notable properties of the lower vocal tract acoustics. First, lower vocal tract resonances correspond to cavities that remain largely static during voicing [4, 5]. Thus, much like subglottal resonances, those of the lower vocal tract are speaker-dependent, yet context-independent [6]. For the subglottal (Sg) case, these properties prompted development of techniques for automatic estimation of the resonances with application to speaker normalization [7]. Unfortunately, however, the Sg resonances have a subtle impact on the speech spectrum, most easily identified via a break in the frequency track of the second formant [6, 7]. On the other hand, the lower vocal tract resonances are both spectrally prominent and localized in a frequency region above lower formants (e.g. F1-F2).

In addition to being speaker-dependent, the lower vocal tract acoustics have also played a central part in studies on speaking and singing styles. Specifically, the singer's or speaker's formant has been a major focus in the voice community for decades. Several studies have clearly identified this concentration of spectral energy near 3kHz associated with projection of the speaking or singing voice [8, 9, 10, 11, 12, 4]. Explanations for this phenomenon focus on the contributions of two mechanisms: narrowing the epilarynx to enhance its resonance and clustering F3-F5 [8, 9, 4]. Though it is well established that the singer's and speaker's formant produce a concentration of spectral energy in mid-to-high frequency regions that yield a louder voice [13], analyses of the acoustics underlying this phenomenon have been limited to hypothesis and simulations [12].

The present work is the first of its kind to automatically extract spectral features from natural speech that are linked to acoustics of the lower vocal tract. Drawing from previous speech science studies, an approach is outlined that seeks to estimate features corresponding to resonance acoustics of the epilarynx and piriform cavities. In particular, to capture the cyclic closing and opening of the epilarynx tube with the vocal folds, analyses that are unique to this work exploit spectral differences between the glottal closed and open phases. Examples of the estimated spectral features are shown for several speakers, highlighting a prominent resonance pattern between 3-6kHz consistent with that shown in previous studies on the lower vocal tract. Additionally, analysis of a speaker's formant instance quantifies the observed concentration of spectral energy in relation to the other speakers in addition to relating the phenomenon to epilarynx acoustics, ultimately shedding light on one of its hypothesized production mechanisms.

This paper is structured as follows. Section 2 provides background on acoustics of the lower vocal tract. Section 3

*This work is sponsored under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

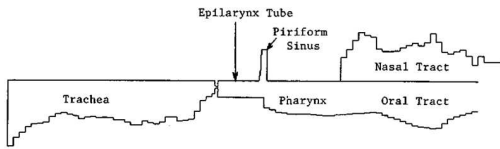


Figure 1: *Vocal tract outline with the bend removed, from [3].*

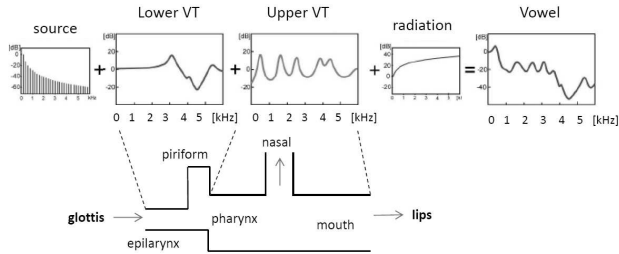


Figure 2: *Modified source-filter and transmission line views of speech production, adapted from [4] and [3] respectively.*

then outlines the proposed approach to estimate spectral features loosely representing the epilarynx and piriform cavity resonances, beginning with a description of the novel multi-resolution spectral analyses. Section 4 then shows results of the proposed feature estimation on natural speech for multiple speakers, ending with analysis of a speaker’s formant. Finally, Section 5 concludes and discusses future work.

2. Acoustics of the Lower Vocal Tract

Situated between the glottis and pharynx, the lower vocal tract (vocal tract) consists of the epilarynx tube (or laryngeal ventricle and vestibules) and piriform fossa (sinus), as shown in Fig. 1 [3]. Highlighting the role of the epilarynx and piriform cavities, the works in [3, 4] offer alternative views of speech production showing the epilarynx tube on the main transmission line contributing a peak (shown near 3.5kHz) in the lower vocal tract frequency response and the piriform fossa as a sidebranch producing a null (shown near 5kHz).

2.1. The Epilarynx Tube

Given its location in the vocal tract (Fig. 1), the epilarynx tube effectively closes and opens with the vocal folds. Thus, it alternatively acts as a quarter- and half- wavelength resonator during the respective glottal closed and open phases [5]. The resonant frequencies are determined by the cavity length, which is approximately 1/6 of the total vocal tract length [3]. Typically, the *epilarynx resonance* refers specifically to the first resonance of the epilarynx tube during the closed phase [4, 3, 5, 14]. In essence, the epilarynx resonance, i.e. spectral energy in a frequency band near 3-4kHz during the closed phase, seemingly disappears during opening [5]. This apparent *cyclicity* between the open and closed phases serves as the primary acoustic cue captured by the multi-resolution spectral analyses in Section 3.

2.2. Piriform Fossa

The Piriform fossa are two cavities (left and right) adjacent to the larynx that resonate during voicing and essentially absorb energy at their resonance frequencies, generating nulls (or zeros) in the speech spectrum [15]. Although the left and right cavities do interact acoustically, one of the spectral dips typically dominates [16]. The piriform fossa is therefore acousti-

cally modeled as a single uniform sidebranch [3], and acts as a quarter wavelength resonator (Fig. 2). Typically, the *piriform resonance* refers to the first resonance of this sidebranch, with frequency shown to be near 4-5kHz for male speakers and higher for females [4]. Consequently, as shown in Fig. 2, the piriform resonance manifests as a deep spectral null that follows the epilarynx resonance in frequency.

3. Analysis of Spectral Features

Driven by the acoustics described above in Section 2, the following section outlines a technique to estimate spectral features in natural speech corresponding to those of the lower vocal tract. A key element of the proposed analyses relies on detecting significant spectral differences between the open and closed phases of the glottal cycle, linked to epilarynx cyclicity.

3.1. Multi-Resolution Spectral Analyses

The speech analyses are pitch-synchronous, using the GLOAT package for pitch, voicing and glottal closure instant (gci) estimation via the SEDREAMS algorithm [17]. Each (full resolution) speech frame is three pitch periods long and a Hamming window is used for analysis. The frame is considered voiced if all three periods that it contains are voiced. The spectral envelope for the frame is estimated using the True Envelope [18] processing with a cepstral order of 40 (removing the zeroth coefficient to mitigate effects of energy differences between frames).

The next stage of spectral analysis isolates the glottal closed and open phases. Given that a typical closed phase for a male speaker is between 30%-45% of the glottal cycle [19], the closed phase is approximated as 1/3 of the pitch period (T_n for voiced frame n). As shown in the top panel of Fig. 3, a hamming window of length $1/3 T_n$ is applied to the closed phase region of the glottal cycle (beginning $1/9 T_n$ before the detected gci so that the left edge of the window does not overly attenuate closure). The spectral envelope for the closed phase is estimated on the resulting magnitude spectrum, again using the True Envelope of cepstral order 40 with zeroth coefficient removed. A minimum of 2.5ms is set for the window length, respecting the data requirement for spectral analyses used in [19]. In order to maintain comparable spectral resolution, the open phase is extracted using the same-size Hamming window as for the closed phase of the frame (e.g. $1/3 T_n$), centered on the remaining non-closed phase part of the pitch period (e.g. $2/3 T_n$), cf Fig. 3. The open phase spectral envelope is then estimated in the same manner as for the closed phase. An example of the spectral analyses are shown in Fig. 3. From Fig. 3 (middle), it is clear that the spectral peaks are better resolved in the full envelope while the closed-open difference (bottom) captures the most significant (over 20dB) difference in energy between the closed and open phases near 3kHz, reflecting the cyclicity property of the epilarynx resonance [4].

3.2. Cyclic Peak and Deep Valley Estimation

In order to analyze more detailed spectral features linked to the lower vocal tract, a technique for mid-to-high frequency peak (loosely epilarynx) and valley (loosely piriform) estimation and refinement is described below. This analysis technique assumes only the presence of a *cyclic peak* and *deep valley* in a mid-to-high frequency range. We do not claim to estimate the true epilarynx and piriform resonances, as they are difficult to determine (requiring MRI-based vocal tract modeling and simula-

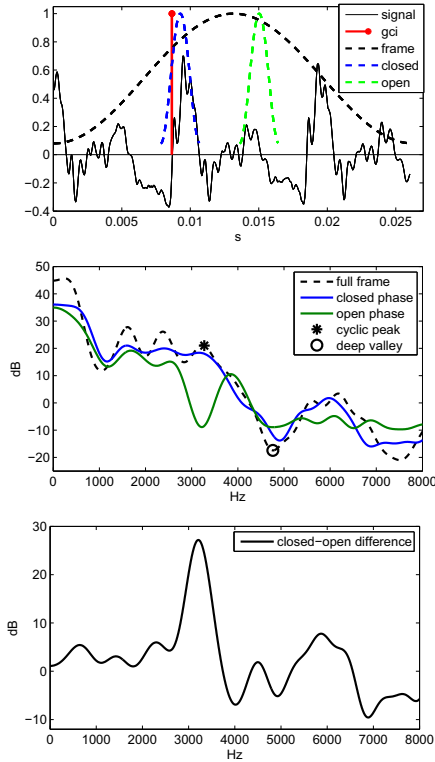


Figure 3: (Top) full-resolution frame and closed, open analysis windows; (Middle) spectral envelopes with feature estimates; (Bottom) closed-open spectral difference.

tions [4,3]), particularly for natural running speech.

For a given speaker, a single cyclic peak with frequency and amplitude $\{f_p, a_p\}$ and deep valley $\{f_v, a_v\}$ are estimated via the following process. First, for each frame n , an initial estimate of the cyclic peak $\{\hat{f}_p(n), \hat{a}_p(n)\}$ is taken to be the maxima of the (full-resolution) frame spectral envelope in the frequency range between 3-5kHz that exhibits maximal closed-open spectral difference (cf. Fig. 3). The initial estimate of the deep valley for frame n , $\{\hat{f}_v(n), \hat{a}_v(n)\}$, is then the lowest-amplitude minima of the (full-resolution) frame spectral envelope that follows $\{\hat{f}_p(n), \hat{a}_p(n)\}$ in frequency within 2.5kHz (cf. Fig. 3).

A refinement stage of this peak and valley estimation then follows, beginning with the generation of histograms (200Hz bin separation) of $\{\hat{f}_p(n), \hat{f}_v(n)\}$ for all voiced frames n . The maximum of each respective histogram, \hat{f}_p^* and \hat{f}_v^* are then used to further localize the cyclic peak and deepest valley estimates for each frame. In essence, these maxima identify the frequencies at which these phenomena occur most consistently across the voiced speech. In the refinement, the peak and valley estimates for each frame n , $\{f_p(n), a_p(n)\}$ and $\{f_v(n), a_v(n)\}$ are then taken to be the full-resolution spectral maxima and minima that are closest in frequency, respectively, to \hat{f}_p^* and \hat{f}_v^* . Finally, the overall cyclic peak $\{f_p, a_p\}$ and deep valley $\{f_v, a_v\}$ estimate for the speaker is the average of $\{f_p(n), a_p(n)\}$ and $\{f_v(n), a_v(n)\}$ for the voiced frames.

Fig. 4 plots a full-resolution (left) and closed-open difference (right) spectrogram for a segment of voiced frames from an example sentence, with the overall f_p and f_v for the speaker indicated by the solid black and red lines, respectively. The refined estimates for every frame, $f_p(n)$ (black *) and $f_v(n)$

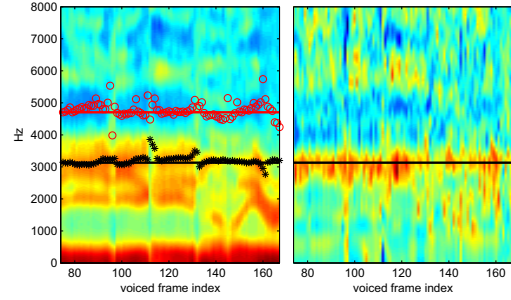


Figure 4: Full-resolution (left) and closed-open difference (right) spectrograms of voiced frames after True Envelope processing, with f_p -black ($f_p(n) -*$) & f_v -red ($f_v(n) -o$).

(red o), are also shown. First, note that the deep valley is quite prominent across the speech, indicating a clear null in the spectrum. Now further examining Fig. 3-4, the estimated cyclic peak appears to correspond to the fourth formant [14]. However, the distinction of this spectral peak from a higher-order resonance of the upper vocal tract lies in the closed-open spectral difference, clearly observed for this speaker on the bottom panel of Fig. 3 and right of Fig. 4. Comparing the left and right plots of Fig. 4, it is further evident that the full frame envelope does not capture the unique cyclicity property of the epilarynx resonance [4]. Thus, the approach outlined identifies a special or distinct formant of the vocal tract that is only observable via glottal closed-open phase spectral analyses.

4. Natural Speech Examples

4.1. Feature Estimation across VOICES Speakers

The natural speech used for analyses is from the VOICES multi-speaker corpus [20]. The corpus contains single 22kHz sampling rate microphone recordings of 12 speakers (1 template-male, 6 male and 5 female) each reading the same set of 50 phonetically rich sentences. All of the speakers were asked to mimic the template speaker's timing in their readings. The first 5 sentences are used for the analyses in this work, as extensive corpora need not be required to extract the prominent features (c.f. Section 3, where only a single sentence is used).

Statistics on the estimated spectral features for all speakers in the VOICES corpus are given in Table 1. Fig. 5 plots results for a representative selection of speakers (top-to-bottom: mwm-M-template, mam-M, zpg-M, sll-F, sas-F, zng-F). On the mean (voiced) spectral envelopes plotted on the left of Fig. 5, the estimated spectral features are shown to clearly reflect those of the lower vocal tract filter established in previous studies (cf. Sections 1-2) and canonically illustrated in Fig. 2. That is, the cyclic peak (at 3.7kHz with amplitude 9.7dB on average) clearly descends towards the deep valley (at 5.1kHz with amplitude -11.6dB on average). The frequencies and amplitudes of these peak-valley features also vary across speakers, as the lower vocal tract resonances are expected to do given their correspondence with speaker cavity sizes. Specifically, comparing the estimated features for the male and female speakers in both Fig. 5 and Table 1, it is clear that those for the female speakers are higher (specifically 500Hz for f_p and 800Hz for f_v), which is consistent with a shorter vocal tract.

Next, examining the average closed-open spectral envelope differences on the right of Fig. 5, prominent energy in a mid-to-high frequency band (3-5kHz) can be seen, reflecting the epilarynx resonance cyclicity. These localized maxima are observed in the mean closed-open spectral difference for all of the speak-

ers, indicating that the trend is consistent across voiced speech. Finally, note that for the female speakers, the closed-open spectral difference is less prominent, due to the higher pitches resulting in overlap between the closed and open phase approximations described in Section 3.1. This observation speaks to the compromise between spectral resolution and time-localization of the closed and open phases. Ultimately, there is a limitation of the proposed analyses for higher pitched speakers (e.g children). Similarly, for breathy voice or other types of non-modal phonation where the distinction between glottal closed and open phases is increasingly opaque, information from the proposed multi-resolution analyses would also be limited.

That said, the closed-open spectral difference curve offers an alternative view of the speaker acoustics and captures different information from standard analyses. In addition to epilarynx cyclicity that is observed for all of the speakers around 3-4kHz, the first (smaller) peaks under 1kHz seem to capture F1 modulation between the closed and open phases [21]. Further analysis and exploration of the information conveyed in the closed-open spectral difference is a subject for future work.

Table 1: Estimated spectral features for VOICES speakers.

	f_p (kHz)	a_p (dB)	f_v (kHz)	a_v (dB)
All	3.7	9.7	5.1	-11.6
Template	3.2	20.4	4.7	-13.4
Male	3.5	12.6	4.8	-11.7
Female	4.0	5.7	5.5	-11.4

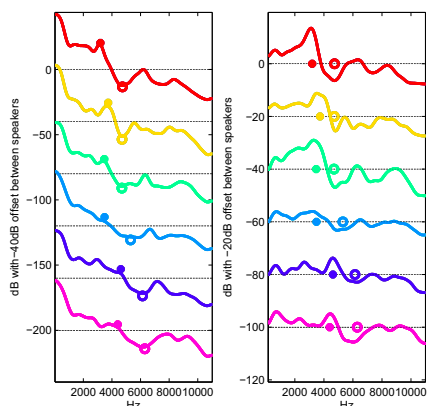


Figure 5: Estimated cyclic peaks (“*”) and deep valleys (“o”) on mean spectral envelopes (left) with feature frequencies on the closed-open differences (right) for several VOICES speakers.

4.2. Speaker’s Formant Analysis

Previous studies examining the singer’s or speaker’s formant have focused on a concentration of spectral energy (typically hand-marked) in a region near 3-3.5kHz (for males) that produces a vocal ring or resonant voice quality [8, 9, 10, 11, 12, 4]. In addition to trained voices, the speaker’s formant has also been shown to be relevant in analyzing different styles of phonation, namely loud versus soft. This particular formant has been further linked to epilarynx narrowing, which amplifies the corresponding resonance and increases energy in a frequency band where humans happen to be most sensitive to loudness [13]. Considering the present work, analyses on the acoustics of the lower vocal tract are well-equipped to highlight this paralinguistic phenomenon in speech.

Specifically, unlike clustering formants (F3-F5), epilarynx narrowing would enhance spectral differences between the glot-

tal closed and open phases located at the frequency of the observed speaker’s formant. Indeed, the above conjecture is observed for an instance of a speaker’s formant seen for the template speaker of the VOICES corpus (Fig. 6). As the template speaker set the timing for all others to mimic, he was likely attempting to speak in a clear manner and might also have been professionally trained, though this has not been confirmed. Examining Fig. 6, there is a notable concentration in spectral energy indicated on the left plot with an arrow near 3kHz for the template speaker (red) compared to the others [10]. More quantitatively in Table 1, the estimated cyclic peak is significantly higher for the template compared to the rest (20.4 dB versus 9.7 dB). At the same time, on the right panel of Fig. 6, the mean closed-open spectral difference shows a notable peak at the same frequency (also indicated by an arrow): this enhanced cyclicity thus suggesting epilarynx narrowing.

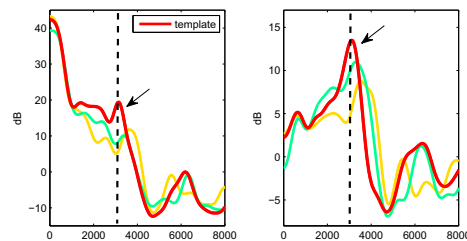


Figure 6: A speaker’s formant shown on the mean spectral envelope (left) and closed-open difference (right) for the template speaker (red) with male speakers (green, yellow) from Fig. 5.

5. Conclusions

This paper outlined an approach to automatically estimate spectral features in a mid-to-high frequency range that exhibit acoustic properties of the lower vocal tract cavities. Results of the proposed approach on natural speech are shown for multiple speakers, illustrating that the extracted features (loosely corresponding to the epilarynx and piriform resonances) form a prominent and consistently observed pattern, with individual variations based on speaker and gender. Additionally, analysis of a speaker’s formant revealed a corresponding prominence of closed-open spectral energy cyclicity, reflecting acoustics consistent with epilarynx narrowing.

As in previous studies focusing on mid-to-high frequency spectral energy in examining different production modes, voice training and disordered speech, the analyses proposed in this work could also be applied in both speech pathology and professional voice contexts. Additionally, the acoustics of the lower vocal tract, as observed with the present techniques, could also help to explain spectral differences observed for speaking styles aimed at increasing intelligibility, such as forms of clear speech [22, 23]. Furthermore, similarly to the application of subglottal resonances [7], the estimated lower vocal tract resonances will also be considered in the future as a low computation option for vocal tract length normalization (VTLN).

Finally, the current work exploits notable variations of the vocal tract within a glottal cycle that are commonly assumed to be negligible in source-filter models of speech production. Though used here for lower vocal tract feature estimation, the unique closed-open glottal phase spectral difference analyses in this work provide an alternative view of other speech features, such as F1 modulation. Ultimately, this work highlights important limitations of the standard source-filter representation, consequently motivating use of finer resolution time-variation of the vocal tract.

6. References

- [1] B. Monson, A. Lotto, and B. Story, "Analysis of high-frequency energy in long-term average spectra of singing, speech and voiceless fricatives," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1754–1764, 2012.
- [2] W. M. Campbell, D. Sturim, B. J. Borgstrom, R. Dunn, A. McCree, T. F. Quatieri, and D. A. Reynolds, "Exploring the impact of advanced front-end processing on list speaker recognition microphone tasks," in *Odyssey*, 2012, pp. 180–186.
- [3] I. Titze and B. Story, "Acoustic interactions of the voice source with the lower vocal tract," *J. Acoust. Soc. Am.*, vol. 101, no. 4, pp. 2234–2243, 1996.
- [4] K. Honda, T. Kitamura, H. Takemoto, S. Adachi, P. Mokhtari, S. Takano, Y. Nota, H. Hirata, Y. Shimada, I. Fujimoto, S. Masaki, S. Fujita, and J. Dang, "Visualisation of hypopharyngeal cavities and vocal-tract acoustic modelling," *Comp. Methods in Biomech. & Biomed. Engineering*, vol. 13, no. 4, pp. 443–453, 2010.
- [5] T. Kitamura, H. Takemoto, S. Adachi, P. Mokhtari, and K. Honda, "Cyclicality of laryngeal cavity resonance due to vocal fold vibration," *J. Acoust. Soc. Am.*, vol. 120, no. 4, pp. 2239–2249, 2006.
- [6] S. Lulich, J. R. Morton, H. Arsikere, M. S. Sommers, G. K. F. Leung, and A. Alwan, "Subglottal resonances of adult male and female native speakers of american english," *J. Acoust. Soc. Am.*, vol. 132, no. 4, pp. 2592–2602, 2012.
- [7] S. Wang, A. Alwan, and S. M. Lulich, "Automatic detection of the second subglottal resonance and its application to speaker normalization," *J. Acoust. Soc. Am.*, pp. 3268–3277, 2009.
- [8] J. Sundberg, "Articulatory interpretation of the "singing formant"," *J. Acoust. Soc. Am.*, vol. 55, no. 4, pp. 838–844, 1974.
- [9] T. Leino, A. Laukkanen, and V. Radolf, "Formation of the actor's/speaker's formant: a study applying spectrum analysis and computer modeling," *J. of Voice*, vol. 25, no. 2, pp. 150–158, 2011.
- [10] T. Nawka, L. C. Anders, M. Cebulla, and D. Zurakowski, "The speaker's formant in male voices," *J. of Voice*, vol. 11, no. 4, pp. 422–428, 1997.
- [11] K. Verdolini, D. Druker, P. Palmer, and H. Samawi, "Physiological study of "resonant voice"," National Center for Voice and Speech Status and Progress Report, Tech. Rep. 6, 1994.
- [12] I. Titze, "Acoustic interpretation of resonant voice," *J. of Voice*, vol. 15, no. 4, pp. 519–528, 2001.
- [13] I. O. for Standardization, "Equal loudness contours, ISO 226," 2003.
- [14] H. Takemoto, S. Adachi, T. Kitamura, P. Mokhtari, and K. Honda, "Acoustic roles of the laryngeal cavity in vocal tract resonance," *J. Acoust. Soc. Am.*, vol. 120, no. 4, pp. 2228–2238, 2006.
- [15] J. Dang and K. Honda, "An improved vocal tract model of vowel production implementing piriform resonance and transvelar nasal coupling," in *ICSLP*, 1996, pp. 965–968.
- [16] H. Takemoto, S. Adachi, P. Mokhtari, and T. Kitamura, "Acoustic interaction between the right and left piriform fossae in generating spectral dips," *J. Acoust. Soc. Am.*, vol. 134, no. 4, pp. 2965–2974, 2013.
- [17] T. Drugman, M. Thomas, J. Gudnason, P. Naylor, and T. Dutoit, "Detection of glottal closure instants from speech signals: a quantitative review," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 20, no. 3, pp. 994–1006, 2012.
- [18] A. Roebel and X. Rodet, "Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation," in *Digital Audio Effects (DAFx)*, 2005, pp. 30–35.
- [19] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, "Estimation of glottal closure instants in voiced speech using the dypsa algorithm," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, pp. 34–43, 2007.
- [20] A. Kain, "High resolution voice transformation," Ph.D. dissertation, Oregon Health and Science University, 2001.
- [21] M. D. Plumpe, T. F. Quatieri, and D. A. Reynolds, "Modeling of the glottal flow derivative waveform with application to speaker identification," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 7, pp. 569–586, 1999.
- [22] E. Godoy, M. Koutsogiannaki, and Y. Stylianou, "Approaching speech intelligibility enhancement with inspiration from lombard and clear speaking styles," *Computer Speech & Language*, vol. 28, no. 2, pp. 629–647, 2014.
- [23] V. Hazan and R. Baker, "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," *Journal of the Acoustical Society of America*, vol. 130, no. 4, pp. 2139–2152, 2011.