Published in final edited form as:

J Voice. 2017 March ; 31(2): 149–156. doi:10.1016/j.jvoice.2016.04.010.

New evidence that nonlinear source-filter coupling affects harmonic intensity and f_o stability during instances of harmonics crossing formants

Lynn Maxfield¹, Anil Palaparthi¹, and Ingo Titze^{1,2}

¹ National Center for Voice and Speech, University of Utah

² Department of Communication Sciences and Disorders, University of Iowa

Abstract

The traditional source filter theory of voice production describes a linear relationship between the source (glottal flow pulse) and the filter (vocal tract). Such a linear relationship does not allow for nor explain how changes in the filter may impact the stability and regularity of the source. The objective of this experiment was to examine what effect unpredictable changes to vocal tract dimensions could have on f_o stability and individual harmonic intensities in situations in which low frequency harmonics cross formants in a fundamental frequency glide. To determine these effects, eight human subjects (five male, three female) were recorded producing f_o glides while their vocal tracts were artificially lengthened by a section of vinyl tubing inserted into the mouth. It was hypothesized that if the source and filter operated as a purely linear system, harmonic intensities would increase and decrease at nearly the same rates as they passed through a formant bandwidth, resulting in a relatively symmetric peak on an intensity-time contour. Additionally, f_o stability should not be predictably perturbed by formant/harmonic crossings in a linear system. Acoustic analysis of these recordings however revealed that harmonic intensity peaks were asymmetric in 76% of cases and 85% of f_o instabilities aligned with a crossing of one of the first four harmonics with the first three formants. These results provide further evidence that non-linear dynamics in the source-filter relationship can impact f_o stability as well as harmonic intensities as harmonics cross through formant bandwidths.

Introduction

For over 50 years, the study of speech production has been governed by the source filter theory¹, whereby the time-varying glottal airflow (source) is acoustically “shaped” as it passes through the vocal tract (filter). Much of speech production can be successfully described by this theory using a purely linear relationship between the source and the filter

Corresponding Author Lynn Maxfield, PhD, Associate Director, National Center for Voice and Speech, University of Utah, 136 S Main St, Ste. #320, Salt Lake City, UT 84101, Lynn.maxfield@utah.edu, Phone: 801-596-2012, Fax: 801-596-2013.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

* A portion of this research was presented at The Voice Foundation 44th Symposium in Philadelphia, PA, May 27, 2015

(i.e., the acoustic pressures created by the filter affect the acoustic output of the system, but do not affect the source itself). This is particularly true of male speech, where the dominant harmonics f_0 (fundamental frequency), $2f_0$ (second harmonic), and $3f_0$ (third harmonic) tend to fall below the formant frequencies of the vocal tract (F_n), resulting in only a simple time-delay interaction between the filter and the glottal airflow source.² However, in female or child speech and singing, where dominant harmonics approach and even cross the vocal tract formant frequencies, a non-linear (interactive) source-filter relationships is evident.³ This interactive relationship can cause disruption of the vocal fold vibration, resulting in unstable and unpredictable phonation.⁴

An interactive relationship between the source and filter has long been evident in computer simulations of voice.⁵⁻⁹ For example, a one-mass model failed to achieve self-sustained oscillation without a vocal tract⁶, and a two-mass model exhibited sudden jumps in f_0 as f_0 passed through F_1 (first formant).⁷ Coupling has also been theorized and observed in the Iberian red deer, both excised and in vivo.^{10, 11} Additionally, interaction has been observed in the form of f_0 jumps and the presence of subharmonics with acoustic coupling to the subglottal system¹² as well as in excised canine and porcine larynges with an added supraglottal vocal tract.^{13, 14}

In human phonation, the interactions are often not severe, but nevertheless clearly observable.^{4, 15-18} It appears that the degree to which the source and filter are coupled is variable, both between and within individuals. If the coupling is relatively weak, the source frequencies are produced independently of the acoustic pressures in the airways. Thus, the glottal flow in the larynx is produced aerodynamically without being affected by acoustic pressures¹⁹, with a quasi-steady transglottal pressure and a flow pulse that mirrors the time-varying glottal area. With stronger coupling, however, the acoustic airway pressures contribute to self-sustained vocal fold vibration, and thereby the entire sound production process at the source.²⁰ The phenomenon is well known in woodwind instrument acoustics, where the airflow through the reed is driven by acoustic pressures of the instrument bore²¹; or in brass instrument acoustics, where the lip flow is driven by the acoustic pressures in the brass tube.²²

Previous research has indicated that source filter interaction (SFI) can be subdivided into two categories: Level 1 and Level 2 (see Figure 1).³ In Level 1 interaction, the acoustic pressures within the vocal tract interact only with the glottal airflow. This can cause changes in the harmonic frequencies (e.g. changes in individual harmonic intensity and/or frequency) without perturbing vocal fold vibration. These spectral changes, influenced by reactance/resistance of the vocal tract, affect the radiated acoustic pressure wave.^{23,24}

Level 2 interactions, on the other hand, are interactions in which the vibratory pattern of the vocal folds is disturbed by the acoustic pressures within the vocal tract. Level 2 interactions occur when a harmonic with significant energy (typically f_0 to $4f_0$) is in the vicinity of a formant. These interactions can result in vocal instabilities such as frequency jumps, sub-harmonic frequencies, or chaotic vibration.^{4,25,26}

The focus of this experiment was to investigate both levels of interaction in human subjects and thereby to begin to elucidate how humans cope with unpredictable filter conditions in regions of f_0 where SFI is likely to occur. Vocal tract length was varied randomly with tube extensions at the mouth. Level 1 interactions were addressed by examining the effect of harmonics crossing formant on the intensity contours of individual harmonics during an f_0 glide. Level 2 interactions were investigated by looking at the f_0 instabilities (pitch jumps) during the same exercise.

Methods

Subjects

The subjects for this study were 5 males and 3 females, ranging in age from 22 to 45. All subjects reported no history of vocal pathologies and two males reported having received significant vocal training (singing). Prior to recording, subjects demonstrated that they could voluntarily produce a vocal fry and an f_0 glide sufficient to cover the desired range. This research was approved and overseen by the Institutional Review Board of the University of Utah (IRB number 00057403).

Procedure and Equipment

All subjects were asked to perform pitch glides (f_0 sweeps) beginning in vocal fry and extending to 500 Hz for males and 700 Hz for females, then descending to vocal fry again (Figure 2). For all repetitions, the vocal tract was extended randomly in length using 6 tubes of varying length (5 cm, 7 cm, 9 cm, 11 cm, 15 cm, and 19 cm), effectively altering the formant frequencies and ensuring that several harmonics, including the fundamental, would be forced to pass through at least one formant region. The tubes were fitted with a mouthpiece to ensure uniform jaw and lip placement and held between the subjects' teeth with a tight lip seal around the tube throughout phonation. Male subjects used tubes with a cross sectional area of 3.14 cm² while females used tubes with areas of 1.77 cm². These dimensions were chosen to roughly correspond to the average cross-sectional areas of the male and female vocal tracts.¹⁸ Tube lengths increased systematically from 5 cm to 19 cm. With a known vocal tract length, formant frequencies (F_n) could be estimated as

$$F_n = (2n - 1) (c/4L)$$

where n is the formant number, c is the speed of sound, and L is the combined length of the tube and the vocal tract in cm.²⁷ Using the average vocal tract length for males (17 cm) and females (15 cm)²⁸, table 1 shows the estimated frequencies for F_1 and F_2 for each tube length. These frequencies were low enough to ensure that f_0 would always cross F_1 .

The tubes were placed in a handle, which also served as a mount for a microphone (Countryman Isomax B3) at a constant distance of 1.5 cm from distal end of the tube (see Figure 3). The microphone signal, sampled at 44,100 Hz, was amplified using an FMC pre-amp (RNP model) before being captured along with an EGG signal (for f_0 detection, sampled at 20,000 Hz) using the ADInstruments analog-digital converter box. Signals were recorded using Labchart7 data collection software.

Subjects were instructed to listen to a computer-simulated version of the desired pitch glide. The duration of the up-down glide from fry release to fry onset was 5 seconds. The simulated glide exhibited no f_0 discontinuities, but the attention of the subjects was not directed to this fact. Subjects were instructed to mimic the task, particularly in regard to pacing of the pitch glide, and to maintain a steady, comfortable loudness to the best of their ability. Subjects were not instructed to specifically produce a smooth glide, avoiding pitch jumps, however the simulated glide they were instructed to mimic did not exhibit any such jumps. Through proper fitting of the mouthpiece, the jaw and lip configurations were controlled. Additionally, an attempt was made to minimize tongue movement by instructing subjects to maintain a neutral vowel /ə/, keeping the tip of the tongue in contact with the back of the lower teeth. Following instruction period, the subjects were allowed to practice the task using a practice tube, differing in length from any used during the experiment. During this period, researcher coached the participants to maintain a consistent f_0 glide rate, achieve the target f_0 , and, if necessary, a stable tongue and laryngeal position.

Following the practice period, subjects began data collection using a randomized order of the length of tubes to prevent subjects from anticipating the frequency location at which a harmonic might cross a formant. If a mistake was made, (e.g. inconsistent glide rate, unfinished glide, change in laryngeal or tongue position), the token was repeated added to the end of the recording session, again prevent subjects from repeating tube lengths back to back. Subjects typically needed to repeat three to four tokens for one or more reasons. Once subjects completed glides through all 6 tubes the process was repeated twice more, in new randomized orders, resulting in 18 tokens per subject.

Analysis

Formant frequencies F_1 , F_2 , and F_3 were determined from the fry segments (before and after the glide) for each recording using linear predictive coding within the PRAAT software. A 50 ms window was used. Because of the difficulty in pinpointing formants at high f_0 , it was not possible to ensure that the formant frequencies did not shift during the glide. However, for our analysis an approximate location of the formant frequencies was sufficient. Consequently, a prediction of the formant frequencies from the vocal fry segments adequately identified the formants.

Fundamental frequency (f_0) was computed using short-time Fourier transform of the audio signal. A non-overlapping window of length 50 ms was used. Frequency of the first local peak in the magnitude spectrum was considered as f_0 . Recognizing that no human would be able to produce perfectly stable f_0 glides, it was necessary to identify substantial instabilities in the rate of f_0 change during the glide that might be a result of something more than intended physiological variations. A non-overlapping window of 50 ms was determined to optimize agreement between instabilities identified by the algorithm and those identified via visual inspection of the audio spectrogram.

To do this, a rate of f_0 change was defined as

$$\dot{f}_o(n) = \frac{f_o(n) - f_o(n-1)}{f_o(n) \Delta t}$$

where Δt is the sampling rate in s, and n the sample number. This rate of change was plotted for each token and fitted with a 3rd order best fit polynomial (Figure 4a). The polynomial is meant to identify the average rate of f_o change.

Next a thresholding was applied such that $\dot{f}_o$ values smaller in magnitude than the standard deviation from the best polynomial fit were set to zero. Thus, only the deviations greater than the threshold remained as peaks on the plot (Figure 4b). These peaks were considered to significant instabilities for the purpose of this analysis. Finally, regions were identified in which any of the first four harmonics crossed any of the first three formants (Figure 4c). The threshold τ for each token was computed using the standard deviation equation, viz.

$$\tau = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\dot{f}_o(n) - \dot{f}_p(n) \right)^2}$$

Here, n is the sample number, $\dot{f}_p(n)$ is the n^{th} sample of the best fit polynomial average.

In addition to the f_o changes, the energies of the fundamental and second harmonic were extracted from the short term Fourier transform using an overlapping window of length 25 ms and step size 10 ms. In this case, a smaller overlapping window length allowed for extracting smooth harmonic energy contours, especially at the interaction regions where the frequency increases abruptly. Since the harmonic energies drift upward due to increase in lung pressure with increase in pitch. The drift was removed by subtracting a linear fit of the harmonic energy and subtracting it from the harmonic energy (Figure 5). Thus, the asymmetry in harmonic energies in the vicinity of formant frequencies was quantified after drift correction.

Results

Level 1 interactions: Non-linear effects on harmonic intensity

In a purely linear system, the result of a harmonic crossing over a formant would be predictable as a symmetric rise and fall in amplitude. The intensity of the harmonic, computed in decibels over each time window, would build as it approached the formant, reaching a peak at precisely the peak of the formant before decaying at the same rate that it grew (assuming that the formants are not closely spaced so that the formant shape is symmetric about its center frequency). Figure 6a shows an intensity-frequency contour of f_o (black) and $2f_o$ (red) for a simulated f_o glide with no source-filter nonlinearity. This was produced by passing a sawtooth waveform signal with f_o between 100 and 1000 Hz and constant harmonic intensities through a uniform filter, resulting in evenly-spaced formants of approximately 500, 1500, 2500, and 3500 Hz. The intensities of the filtered signal are shown in Figure 6. Grey bars are drawn to indicate a 50 Hz bandwidth centered on the exact formant frequency. Note that the peaks of the intensity contours of f_o and $2f_o$ align nearly

perfectly with the center of the grey bars, corresponding with the peak formant frequencies. Additionally, the ascending portions of the intensity contours are nearly symmetrical with the descending portions, again as predicted in a linear system.

However, if a nonlinear relationship exists between the source and the filter in human subjects, the effect of harmonics crossing formants on harmonic intensity becomes less predictable. Instead of harmonic intensity depending primarily on the magnitude of the acoustic impedance of the vocal tract (which is symmetric around a formant peak), the supraglottal reactance of the system plays a significant role.³ The reactance is asymmetric around a formant peak, positive below F_1 and negative above F_1 . If positive (inertive), reactance skews the glottal flow pulse and vocal fold vibration amplitude is increased. The greatest effect on both harmonic intensity and f_0 stability will occur near the peaks of the reactance curve, a positive peak that is slightly to the left (lower frequency) of the formant peak, followed by a negative peak slightly to the right (higher frequency) of the formant peak. Figure 7 shows the intensity contours for the first two harmonics during a pitch glide produced by one of the subjects plotted against frequency (a) and time (b). Vertical boxes represent the 50 Hz windows, in which each harmonic is crossing through a formant bandwidth, as estimated from the fry portions of the glide. In contrast to the simulated, non-interactive, case in Figure 6, the harmonic intensity peaks no longer align perfectly with the center of the formant bandwidth. Indeed, in some cases, a subject would raise his or her larynx, shifting formants higher and further offsetting the intensity peak alignment from the formant peak. Furthermore, there is greater dependence of one contour change on the other. Note that as $2f_0$ passes through F_1 (first bar), f_0 intensity is first increased slightly, then decreases when $2f_0$ gets its greatest boost, and finally reaches a gradual increase. Similarly, when f_0 begins to pass through F_1 (second bar), $2f_0$ intensity is diminished slightly. By calculating the slopes of the ascending and descending portions of harmonic intensities, it becomes apparent that the two sides of the peak are no longer symmetrical. This slope of harmonic intensity can be expressed as

$$S = \frac{\Delta I}{\Delta t}$$

where I is the harmonic intensity and t is time. Δt was a constant window of 150 ms preceding the contour peak for the ascending slope and following the contour peak for the descending slope.

A total of 144 tokens were produced by the 8 subjects participating in this experiment. With more than one instance of a harmonic crossing a formant taking place in each token, these 144 tokens yielded 352 crossings. Of these, 266 (76 %) exhibited a significant asymmetry as determined by a slope ratio (SR), calculated by dividing the ascending slope (S_a) by the absolute value of the descending slope (S_d) such that

$$SR = \frac{S_a}{|S_d|}$$

The slope ratio was 0.7 or 1.3 for the 266 cases between the pre- F_1 and post- F_1 portions of each harmonic intensity peak. Given that the overall spectral slope was balanced prior to analysis, the results indicate that Level 1 source-filter coupling occurred.

Interaction is by nature complex when multiple oscillators are coupled to each other and energy can flow between the oscillators. All the harmonics and all the formants may be thought of as coupled oscillators. Positive (inertive) vocal tract reactance delays the glottal flow pulse, thereby increasing the maximum flow declination and the energy into the vocal tract. However, this delay is applied differently to every harmonic because reactance is frequency-dependent. Hence, one harmonic may gain energy while another loses energy as different harmonics pass through a formant. If there were a single harmonic and a single formant only, one could categorically say that a positive reactive term (proportional to the flow derivative) in the vocal tract input impedance

$$Z = R + I \frac{du}{dt},$$

where R is resistance, I is inertance, and u is glottal airflow, would increase harmonic energy. To the contrary, a negative $I \frac{du}{dt}$ term would decrease harmonic energy. Given that this reactive term is more positive below a formant peak than above a formant peak, one would expect a simple slope change in the harmonic energy. With multiple harmonics and multiple formants, however, a more statistical measure (as described above) is needed.

Level Two Interactions: Non-linear effects on f_0 stability

In contrast to Level 1 interactions, recall that Level 2 interactions require the vibratory pattern of the vocal folds to be disturbed by the acoustic pressures within the vocal tract. These disruptions may take the form of pitch instabilities, biphonation, subharmonics, or deterministic chaos.^{16,17,9} It has already been shown that pitch jumps occur more frequently in phonations where f_0 crosses F_1 ⁹ – implying that non-linear source filter coupling may impact the stability of f_0 when it is in the vicinity of F_1 . Here the analysis is extended to include significant changes in rates of f_0 change in the glide (as identified by the process outlined in the analysis section), as well as $2f_0$, $3f_0$, and $4f_0$ interactions with F_1 , F_2 , and F_3 .

The 144 tokens collected in the experiment yielded 625 total significant instabilities, identified as peaks on the f_0 rate plot after thresholding was applied. If a non-interactive (purely source) origin were attributed to these instabilities, it would be expected that the association between an instability and a harmonic crossing a formant would be no greater than chance. If source-filter interaction did occur, given that the crossings of the first four harmonics with the first three formants covered approximately 65% of the frequency range of any given pitch glide, it would be plausible that there would be a 65% chance that any instability might fall within the region of one of those crossings. Of the 625 instabilities, 534 (85%) could clearly be identified as being associated with the crossing of one of the first four harmonics over one of the first three formants, a number significantly higher than chance. Male subjects had a slightly higher correlation between f_0 instabilities and harmonics crossing formants, with 376 of 416 (88%) of instabilities associated with a crossing, while tokens from female subjects associated only 167 of 209 (80%) instabilities

with crossings. These results provide further evidence for nonlinear dynamics within vocal fold vibration, characterized as Level 2 interaction.

Discussion

The experiments described here were designed to further reveal the extent to which nonlinear source-filter relationships may be capable of impacting human phonation. By artificially elongating the vocal tract of human subjects, the frequency regions within which harmonics cross formants were able to be predictably changed. By randomizing the order in which the tubes were presented to each subject, in addition to other measures taken to stabilize the vocal tract structures, individuals were prevented from anticipating these crossings and making small physiological adjustments to avoid significant interactions.

When the intensity contours of the first two harmonics were examined in relation to locations where harmonics crossed formants, it was clear that the intensity peak did not align perfectly with the peak of the formant, indicating that perhaps maximum reinforcement is achieved by placing a harmonic slightly below a formant frequency. Additionally, as the harmonic passed through the formant, the harmonic intensity did not grow and decay in a strictly symmetric fashion, as would be expected within a linear system. These two findings can be explained by considering that harmonic intensity changes associated with harmonics interacting with formants are likely influenced by the acoustic reactance of the vocal tract, not only the magnitude of the acoustic impedance that is symmetric within the formant bandwidths.

Because this analysis hinges on both the magnitude of intensity contour peaks and the rate at which those peaks were approached and left, it is reasonable to be concerned with whether or not the total intensity was kept steady throughout the glide, thus ensuring that the peaks were primarily an acoustic effect of the source-filter interaction, and not a result of inconsistency of the source. Prior to recording, this concern was addressed by instructing the subjects to maintain a steady loudness to the best of their ability. Still, during analysis, researchers witnessed an upward drift in total intensity throughout the glide. This drift was corrected as described earlier in this paper. Finally, the average time spent with a harmonic in or near the bandwidth of a formant was approximately 200 ms, too fast for conscious intensity control or correction when auditory feedback loop delays are considered.²⁹

This work also added to the body of evidence that nonlinear dynamics are capable of destabilizing the vibration of the vocal folds. Previously, it had been established that as long as the speaking f_0 remained well below the first formant (as it does in most speech and most vowels), a linear explanation of the source-filter relationship was sufficient, since any nonlinearities would likely not impact the stability of phonation in those ranges. However, here it was revealed that higher harmonics interacting with higher formants can still have a destabilizing effect on f_0 . With all of these interactions having the potential to cause instabilities during normal speech, the fact that humans with healthy speaking voices are able to control their f_0 and intensity with considerable accuracy indicates that they are likely constantly compensating for these interactions in order to maintain a stable speaking voice. For females speaking at higher f_0 , this may be as simple as lightening the harmonic content

by using less vocal fold adduction. This work also found that males were slightly more likely to suffer an instability as a result of a harmonic crossing over a formant. This difference could be explained by the fact that when female subjects switched from fry phonation to normal voicing, their initial f_o was often above the first formant, meaning that the strongest possibility of interaction had been avoided. Future investigations will need to examine what behaviors or strategies are being applied in these situations to maintain stability. Once these behaviors or strategies are identified, individuals whose voices suffer from non-pathological instabilities may benefit from targeted training in these areas.

This project revealed a number of questions warranting further investigation. In the current project, subjects only completed the task with a tube, never with their natural vocal tract without a tube. This elongation of the vocal tract caused the desired effect of lowering all formants in a predictable manner. However, this lowering also increased the probability of interactions between dominant harmonics and the first few formants. Future work should include a “no tube” to serve as a control. Additionally, future investigators may consider surveying participants about their perceived experiences during the glides in order to determine whether or not the acoustic affects measured can also be perceived by the subject. Finally, a more controlled investigation of how the likelihood of source-filter interaction (resulting in instability) changes in relation to changes in the vocal tract geometry. The current study did not differentiate between the lengths of tubes used by subjects, but such an analysis could provide valuable insights and should be included in future work.

ACKNOWLEDGMENT

Funding for this work was provided by the National Institutes on Deafness and Other Communication Disorders, Grant No. 5 R01 DC012045 02

References

1. Fant, G. The Acoustic Theory of Speech Production. The Hague; Mouton: 1960.
2. Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. J Acoust Soc Am. 1990; 87(2):820–857. [PubMed: 2137837]
3. Titze IR. Nonlinear source-filter coupling in phonation: Theory. J Acoust Soc Am. 2008; 123(5): 2733–2748. [PubMed: 18529191]
4. Titze IR, Riede T, Popolo P. Nonlinear source-filter coupling in phonation: Vocal exercises. J Acoust Soc Am. 2008; 123(4):1902–1915. [PubMed: 18396999]
5. Flanagan JL. Source-system interaction in the vocal tract. Ann NY Acad Sci. 1968; 155(1):9–17.
6. Flanagan JL, Landgraf LL. Self-oscillating source for vocal tract synthesizers. IEEE Trans Audio Electroacoust. 1968; 16(1):57–64.
7. Ishizaka K, Flanagan JL. Synthesis of voiced source sounds from a two-mass model of the vocal cords. Bell Syst Tech J. 1972; 51(6):1233–1268.
8. Merritt T, Raitio T, King S. Investigating source and filter contributions, and their interaction, to statistical parametric speech synthesis. Interspeech. 2014:1509–1513.
9. Koc T, Ciloglu T. Nonlinear interactive source-filter models for speech. Comp Speech & Lang. 2016; 36:365–394.
10. Titze IR, Riede T. A cervid vocal fold model suggests greater glottal efficiency in calling at high frequencies. PLoS Comput Biol. 2010; 6(8):e1000897. [PubMed: 20808882]
11. Volodin I, Volodina E, Frey R, Carranza J, Torres-Porras J. Spectrographic analysis points to source-filter coupling in rutting roars of Iberian red deer. acta ethologica. 2013; 16(1):57–63.

12. Lulich S, Zañartu M, Mehta D, Hillman R. Source filter interaction in the opposite direction: subglottal coupling and the influence of vocal fold mechanics on vowel spectra during the closed phase. *Proceedings of Meetings on Acoustics*. 2009; 6(1):1–14.
13. Smith B, Nemcek S, Swinarski K, Jiang J. Nonlinear source-filter coupling due to the addition of a simplified vocal tract model for excised larynx experiments. *J Voice*. 2013; 27(3):261–266. [PubMed: 23490131]
14. Giovanni A, Ouaknine M, Guelfucci B, Yu P, Zanaret M, Triglia JM. Nonlinear behavior of vocal fold vibration: the role of coupling between the vocal folds. *J Voice*. 1999; 13(4):465–476. [PubMed: 10622513]
15. Zañartu M, Mehta D, Ho J, Wodicka G, Hillman R. Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study. *J Acoust Soc Am*. 2011; 129(1):326–339. [PubMed: 21303014]
16. Echternach M, Traser L, Richter B. Perturbation of voice signals in register transitions on sustained frequency in professional tenors. *J Voice*. 2012; 26(5):674–e9.
17. Sundberg J, Lã FM, Gill BP. Voice Source Variation Between Vowels in Male Opera Singers. *J Voice*. 2015
18. Signorello R, Zhang Z, Gerratt B, Kreiman J. Impact of vocal tract resonance on the perception of voice quality changes caused by varying vocal fold stiffness. *Acta Acust undt with Acust*. 2016; 102(2):209–213.
19. Zhang Z, Neubauer J, Berry D. Aerodynamically and acoustically driven modes of vibration in a physical model of the vocal folds. *J Acoust Soc Am*. 2006; 120(5):2841–2849. [PubMed: 17139742]
20. Sundberg J, Lã FM, Gill BP. Formant tuning strategies in professional male opera singers. *J Voice*. 2013; 27(3):278–288. [PubMed: 23453594]
21. Fletcher N. Excitation mechanisms in woodwind and brass instruments. *Acta Acust*. 1979; 43(1):63–72.
22. Adachi S, Sato M. Trumpet sound simulation using a two dimensional lip vibration model. *J Acoust Soc Am*. 1996; 99(2):1200–1209.
23. Holmberg E, Hillman R, Perkell J. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J Acoust Soc Am*. 1988; 84(2):511–529. [PubMed: 3170944]
24. Sapienza C, Stathopoulos E. Comparison of maximum flow declination rate: Children versus adults. *J Voice*. 1994; 8(3):240–247. [PubMed: 7987426]
25. Hatzikirou H, Fitch W, Herzel H. Voice instabilities due to source-tract interactions. *Acta Acust*. 2006; 92(3):468–475.
26. Švec J, Schutte H, Miller D. On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges. *J Acoust Soc Am*. 1999; 106(3):1523–1531. [PubMed: 10489708]
27. Titze, IR. Principles of voice production. National Center for Voice and Speech; 2000.
28. Story B. A parametric model of the vocal tract area function for vowel and consonant simulation. *J Acoust Soc Am*. 2005; 117(5):3231–3254. [PubMed: 15957790]
29. Larson CR, Burnett TA, Kiran S, Hain TC. Effects of pitch-shift velocity on voice F₀ responses. *J Acoust Soc Am*. 2000; 107:559–564. [PubMed: 10641664]

Linear Source-Filter Acoustics

Level 1 and Level 2 Source-Filter Interaction

FIGURE 1.

Source-filter conceptualization: (upper) non-interaction source filter concept; (lower) two levels of source filter interaction.

FIGURE 2.

Spectrogram of a subject producing a glide through a tube 5 cm in length. Formants one, two, and three are identified on the left. Pitch instabilities are shown with vertical arrows where $4f_o \approx F_2$ and where $f_o \approx F_1$.

FIGURE 3.
Rendering of mouthpiece, tube, and microphone apparatus.

FIGURE 4.

(a) Derivative of the rate of f_o change during a f_o glide with a best fit polynomial, (b) derivative of f_o change with thresholding applied, (c) derivative with thresholding and regions where harmonics $\approx$ formants identified.

Figure 5. Harmonic intensity contours before (above) and after (below) drift correction for f_o , plotted against frequency (a).

FIGURE 6.

Harmonic intensity contours for computer simulated f_o glide plotted against frequency. Grey bars indicate 50 Hz windows, in which f_o and $2f_o$ are in the vicinity of F_1 (508 Hz) and F_2 (1,448 Hz).

FIGURE 7.

Harmonic intensity contours for human-produced f_0 glide plotted against frequency (a) and time (b) with inherent vocal tract interactions.

TABLE 1

Estimated F_1 and F_2 frequencies for artificially elongated vocal tracts.

Tube Length (cm)	5	7	9	11	15	19
Male calculated F_1 (Hz)	398	365	337	313	273	243
Female calculated F_1 (Hz)	437	398	337	313	291	257