New Evidence That Nonlinear Source-Filter Coupling Affects Harmonic Intensity and *fo* Stability During Instances of Harmonics Crossing Formants

***Lynn Maxfield, *Anil Palaparthi, and *,†Ingo Titze,** **Salt Lake City, UT, and †Iowa City, IA*

Summary: 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 *fo* 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 *fo* 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, *fo* 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 that 85% of *fo* instabilities aligned with a crossing of one of the first four harmonics with the first three formants. These results provide further evidence that nonlinear dynamics in the source-filter relationship can impact *fo* stability as well as harmonic intensities as harmonics cross through formant bandwidths.
Key Words: source-filter theory–pitch instability–formant–nonlinearity–voice registration.

INTRODUCTION

For over 50 years, the study of speech production has been governed by the source-filter theory,[1] 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 (ie, 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 *fo* (fundamental frequency), 2*fo* (second harmonic), and 3*fo* (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 nonlinear (interactive) source-filter relationship 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 tract,[6] and a two-mass model exhibited sudden jumps in *fo* as *fo* 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 *fo* jumps and the presence of subharmonics with acoustic coupling to the subglottal system[12] 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 pressures,[19] 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 bore,[21] 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 (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 (eg, 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

Accepted for publication April 19, 2016.
A portion of this research was presented at The Voice Foundation 44th Symposium in Philadelphia, PA, May 27, 2015.
From the *National Center for Voice and Speech, University of Utah, Salt Lake City, UT; and the †Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA.
Address correspondence and reprint requests to Lynn Maxfield, National Center for Voice and Speech, University of Utah, 136 S Main St, Ste. #320, Salt Lake City, UT 84101. E-mail: Lynn.maxfield@utah.edu
Journal of Voice, Vol. 31, No. 2, pp. 149–156
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.04.010

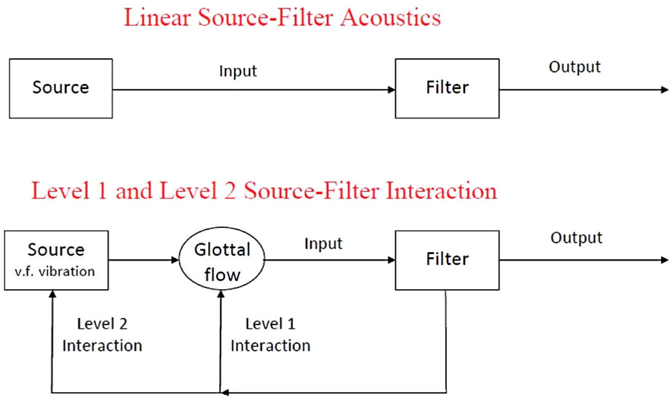

FIGURE 1. Source-filter conceptualization: (upper) linear source-filter concept, showing no interaction between the source and the filter; (lower) nonlinear source-filter concept, showing two levels of source-filter interaction.

when a harmonic with significant energy (typically *fo* to 4*fo*) is in the vicinity of a formant. These interactions can result in vocal instabilities such as frequency jumps, subharmonic 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 *fo* where SFI is likely to occur. Vocal tract length was varied randomly with tube extensions in the mouth. Level 1 interactions were addressed by examining the effect of harmonics crossing formant on the intensity contours of individual harmonics during an *fo* glide. Level 2 interactions were investigated by looking at the *fo* instabilities (pitch jumps) during the same exercise.

METHODS

Subjects

The subjects for this study were 5 men and 3 women, ranging in age from 22 to 45. All subjects reported no history of vocal pathologies and two male subjects reported having received significant vocal training (singing). Prior to recording, subjects demonstrated that they could voluntarily produce a vocal fry and an *fo* 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 (*fo* sweeps) beginning in vocal fry and extending to 500 Hz for men and 700 Hz for women, then descending to vocal fry again (Figure 2). For all repetitions, the vocal tract was extended randomly in length using six 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. The male subjects used tubes with a cross-sectional area of 3.14 cm^2, whereas the female subjects used tubes with an area of 1.77 cm^2. 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 centimeters.[27] Using the average vocal tract length for male subjects (17 cm) and female subjects (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 *fo* would always cross F_1.

The tubes were placed in a handle, which also served as a mount for a microphone (Countryman Isomax B3, Menlo Park, California, USA) at a constant distance of 1.5 cm from the distal end of the tube (Figure 3). The microphone signal, sampled at

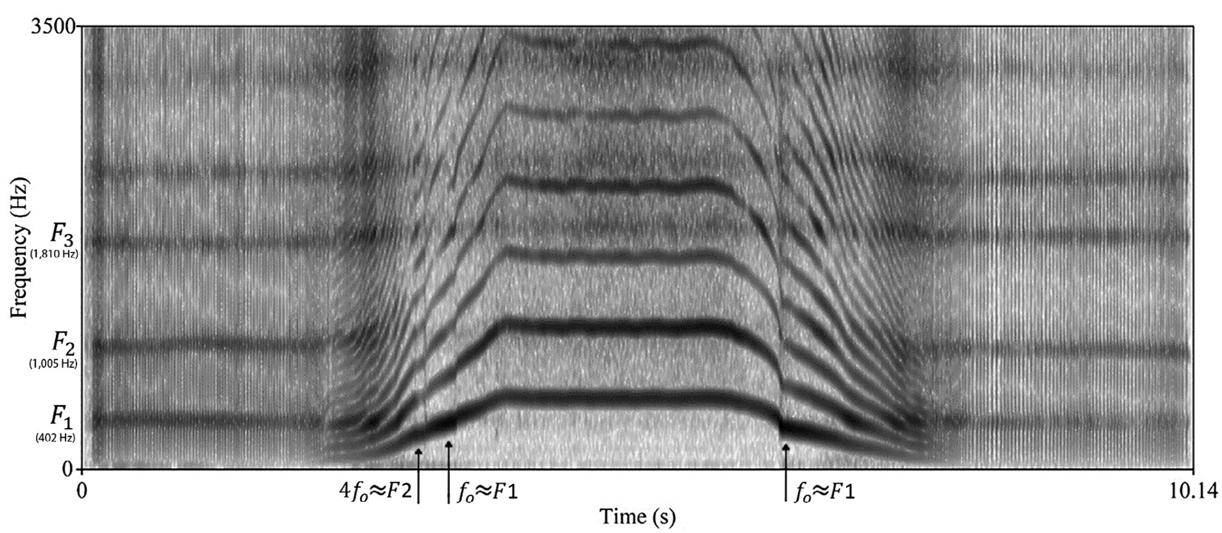

FIGURE 2. Spectrogram of a subject producing a glide through a tube 5 centimeters in length. Formants 1, 2, and 3 are identified on the left. Pitch instabilities are shown with *vertical arrows* where $4fo \approx F_2$ and where $fo \approx F_1$.