Production of child-like vowels with nonlinear interaction of glottal flow and vocal tract resonances

Brad H. Story, and Kate Bunton

Citation: Proc. Mtgs. Acoust. **19**, 060303 (2013); doi: 10.1121/1.4798416
View online: https://doi.org/10.1121/1.4798416
View Table of Contents: https://asa.scitation.org/toc/pma/19/1
Published by the Acoustical Society of America

Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 5pSC: Flow, Structure, and Acoustic Interactions During Voice Production II

5pSC2. Production of child-like vowels with nonlinear interaction of glottal flow and vocal tract resonances

Brad H. Story* and Kate Bunton

***Corresponding author's address: Speech, Language, and Hearing Sciences, University of Arizona, 1131 E. 2nd St., Rm 214, Tucson, AZ 85721, bstory@email.arizona.edu**

Acoustically, the mechanisms of vocal sound production may be considered to exist along a continuum. At one end, the glottal flow wave is weakly coupled to the resonances of the vocal tract such that the output is a linear combination of their respective acoustic characteristics, whereas at the other end there is strong nonlinear coupling of the flow source to the vocal tract resonances. To express phonetic properties in the output, such as formants, the linear case requires that the source produce sound that is rich in harmonic or broadband energy. In contrast, the nonlinear case allows for the possibility of an harmonically- rich source signal to be generated even when the glottal area variation is so simple that it may contain only one harmonic (i.e., a sinusoid) [Titze, J. Acoust. Soc. Am., 123, 2008]. The latter case is most likely to occur when the fundamental frequency is relatively high, such as in children's speech. The purpose of this study was to investigate the nonlinear end of the continuum with respect to the harmonic content of the glottal flow and pressure waveforms for vowels generated with a model of a child-like speech production system. [Supported by NIH R01-DC011275, NSF BCS-1145011]

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4798416]
Received 15 Jan 2013; published 2 Jun 2013

INTRODUCTION

Production of vowels, whether spoken by male or female, child or adult, is typically described as a linear source-filter system in which vocal fold vibration produces a periodic glottal flow signal that excites the resonances of the vocal tract filter (Fant, 1960; Stevens & House, 1961; Flanagan, 1972; Stevens, 2000), and results in an output signal that is the convolution of both components. A key assumption of the linear-source filter representation is that the filter does *not* influence the source, rather it can only enhance or suppress the amplitudes of the spectral components produced by the source. This is a simplified representation of a more complex system in which the source and filter do interact nonlinearly. Titze (2008) described a "Level 1" nonlinear interaction with regard to the dependency of the glottal flow pulse shape on subglottal and supraglottal pressures. It was shown that a sinusoidal variation of the glottal area can produce an harmonically-rich glottal flow signal, and subsequently an equally rich output pressure signal. This means that a period of glottal closure due to vocal fold collision is *not necessary* to produce a fully expressed vowel. For speech produced by adults, especially male, this nonlinearity is likely to create only secondary effects and the linear representation may be an adequate description. But when the fundamental frequency of vocal fold vibration (F0) is fairly high, and relatively close to the frequency of the first resonance of the vocal tract, as it may be in singing and children's speech, the nonlinear interaction of the glottal flow with the acoustic pressures in the subglottal and supraglottal airways may play a larger role in the production of sound.

The purpose of this study was to determine whether a speech production model based on nonlinear source-tract interaction could reasonably simulate a vowel-to-vowel transition of a 6 year-old child. Steps in the process included 1) tracking time-varying formant frequencies from recorded speech, 2) mapping the formants to a time-varying child-like vocal tract area function, and 3) simulation of the vowel transition based on sinusoidal glottal area variation combined with the acoustic pressures propagating in the trachea and time-varying vocal tract. Simulations were compared spectrographically to the original recordings of natural speech.

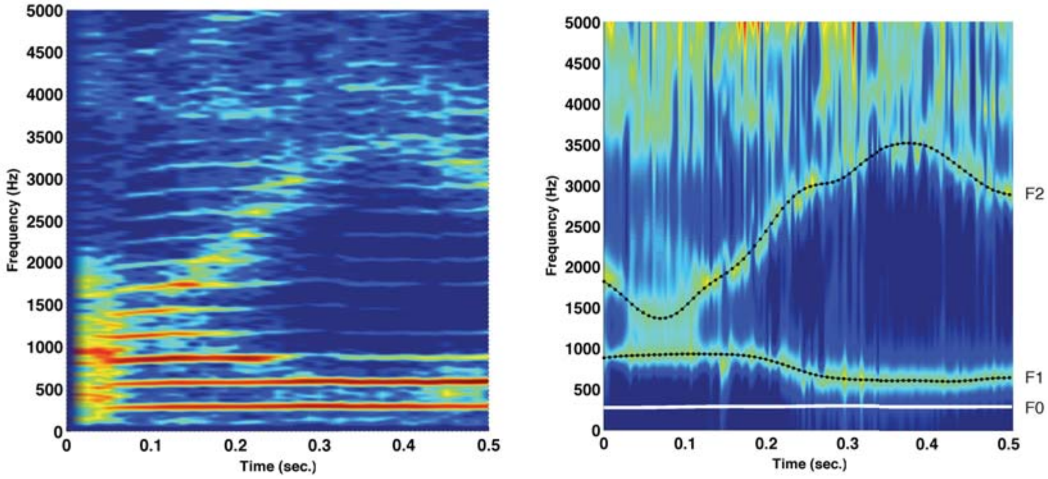
METHOD

The recording used for this study was obtained from a database previously collected by the authors. All procedures for recording were approved by the University of Arizona Institutional Review Board. The audio sample chosen for analysis and simulation was the vowel-vowel transition [ɑi] spoken by a 6 year-old female talker. A narrow-band spectrogram of the sample is shown in Fig. 1a. The mean F0 was 290 Hz and varied only slightly over the utterance. During the initial [ɑ] vowel, the first seven harmonics are fairly prominent in amplitude, whereas during the [i] only about the first three harmonics exhibit similar energy. During the transition, however, the harmonics near the second formant are clearly enhanced. In addition, there is a noise component that is apparent throughout the duration and seems to facilitate the expression of the second formant during the final [i] vowel.

Formant Tracking and Area Function Mapping

Time-varying formant frequencies were tracked with a pitch-synchronous LPC technique (Bunton & Story, 2011). A spectrographic view of the analysis is shown in Fig. 1b, where the tracked and smoothed F1 and F2 formants are shown as black dots; the F0 is indicated by the the white line.

At each time sample corresponding to the formant tracks, a vocal tract area function was estimated by an algorithm that perturbs the shape of initial vocal tract until its resonance



(a) (b)

FIGURE 1: Analysis of [ɑi] produced by a 6 year-old female talker. (a) Narrow-band spectrogram, (b) Formant tracking based on a pitch-synchronous LPC technique; black dots indicate F1 and F2, and the white line is the F0.

frequencies match the measured formants. The algorithm used in the present study is identical to that reported in Story (2006) except that the lip termination was constrained to prevent unrealistic cross-sectional areas.

The resulting time-varying area function is plotted in Fig. 2. The glottis is located at the zero point along the x-axis. The tracheal area function, shown in black and based on data scaled in length and area from Story (1995), extends from the glottis toward the bronchi in the negative x-direction. The vocal tract extends toward the lips in the positive x-direction. The red line indicates the configuration at the beginning the [ɑi] utterance, the blue line is the final shape, and the black dotted lines indicate the shape change that occurred between beginning and end. The vocal tract length was set to be 11.5 cm.

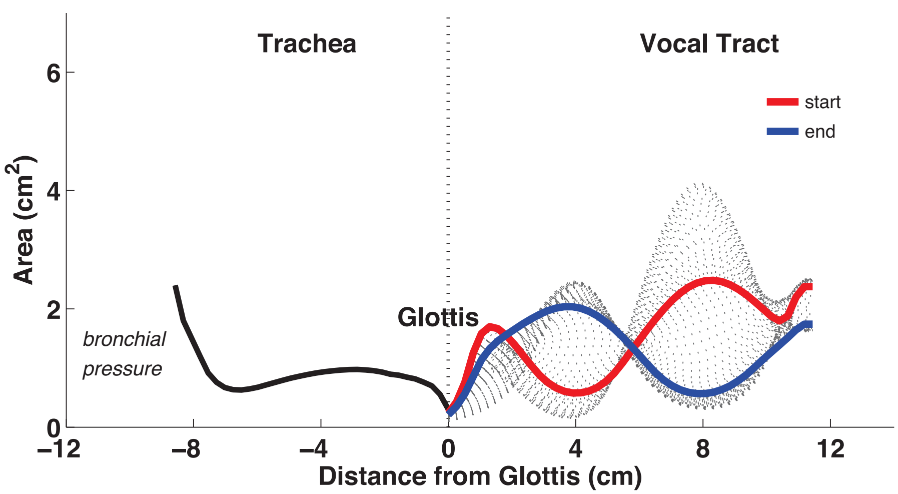


FIGURE 2: Static tracheal area function and time-varying vocal tract area function for [ɑi].

Model of Nonlinear Source-Filter Interaction

The model consists, in part, of the tubular analog of the tracheal and vocal tract airspaces as demonstrated in Fig. 2. The vocal folds would be located at the junction of the two airways marked as “Glottis.” As the vocal folds vibrate, the airspace between their medial surfaces is modulated and can be represented as a glottal area signal $a_g(t)$. The glottal flow $u_g(t)$ can be determined with the following equation (Titze, 1984),

$$u_g(t) = \frac{a_g(t)c}{k_t}\left\{-\left(\frac{a_g(t)}{A^*}\right) \pm \left[\left(\frac{a_g(t)}{A^*}\right)^2 + \frac{4k_t}{\rho c^2}(p_t^+(t) - p_e^-(t))\right]^{1/2}\right\} \quad (1)$$

where k_t is a transglottal pressure coefficient, ρ is the air density, and c is the speed of sound. Computationally, acoustic waves are propagated through the system based on a wave-reflection approach (Liljencrants, 1985; Story, 1995) that includes energy losses due to yielding walls, viscosity, heat conduction, and radiation at the lip termination. The p_t^+ and p_e^- result from the wave propagation and are the partial pressures in the subglottal and supraglottal systems, respectively, that are incident upon the glottal area at each instant of time. A^* is an equivalent vocal tract area based on the entry areas of the trachea, A_t, and the vocal tract, A_e; it is calculated as $A^* = (A_t A_e)/(A_t + A_e)$.

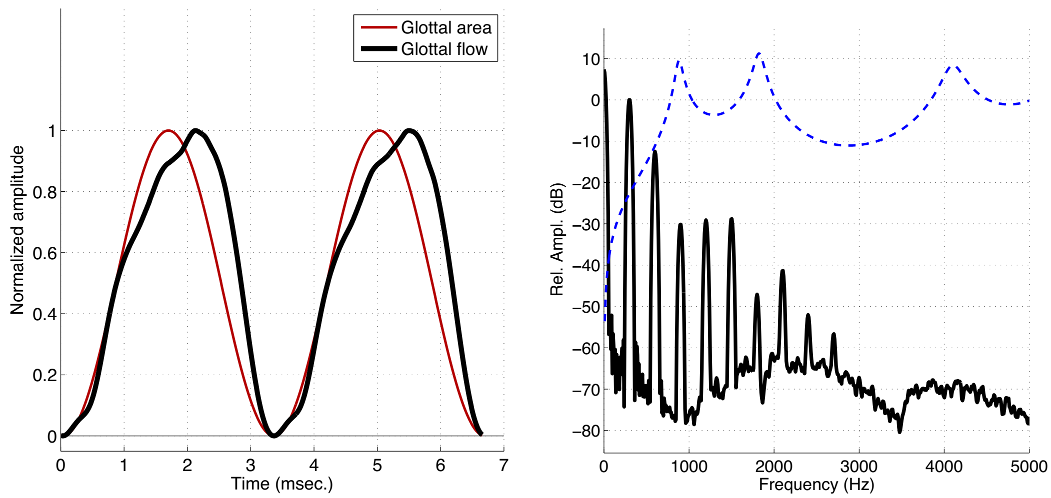
According to Eqn. 1, the glottal flow is dependent on glottal area, vocal tract shape, and the propagating acoustic pressures. As pointed out by Titze (2008), when the ratio a_g/A^* is small, due to small glottal area or large A^*, the equation reduces to one where $u_g(t)$ is proportional to $a_g(t)$, the condition assumed to exist for linear source-filter coupling. Thus, the harmonic content in the glottal flow is largely determined by characteristics of the glottal area. If either the glottal area is large or A^* is small, however, the a_g/A^* ratio becomes large and causes the glottal flow to be strongly dependent on the pressures in the subglottal and supraglottal systems.

To demonstrate the effect of tracheal and vocal tract coupling on glottal flow generation, a *sinusoidal* glottal area signal with a fundamental frequency of 300 Hz and a peak amplitude of 0.1 cm^2, shown as the red line in Fig. 3a, was set to be the output of vocal fold vibration; that is, the vocal folds were assumed to vibrate in the simplest possible pattern and do not generate a glottal closed phase. The vocal tract shape was set to the initial configuration indicated in Fig. 2a and held constant. In addition, a noise component was added to the glottal flow if the Reynolds number exceeded a preset threshold. The resulting glottal flow is shown in Fig. 3a, as the black line. Note that both the glottal area and glottal flow waveforms have been normalized to a peak amplitude of 1.0 for purposes of comparing their wave shapes. Other than the fundamental period, the shape of u_g is altered considerably relative to the glottal area signal. It is skewed rightward in time and contains ripples due to the influence of the vocal tract and tracheal resonances.

The spectrum of the glottal flow signal is shown in Fig. 3b as the black line. Nine harmonics can be observed to have been generated by the nonlinear interaction of glottal area and acoustic pressures in the trachea and vocal tract. The frequency response of the vocal tract configuration is also shown in Fig. 3b (blue dashed line), and indicates that the harmonics near a vocal tract resonance are somewhat suppressed (i.e., see 3F0 and 6F0).

Simulation of a Vowel-Vowel Transition

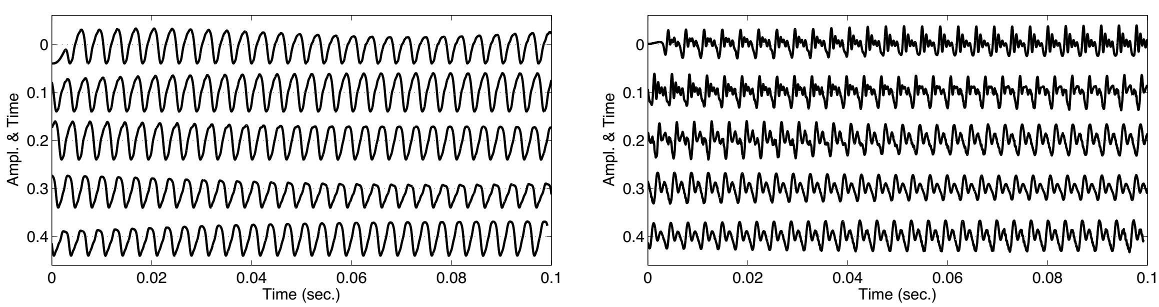
In this section, the nonlinear source-filter interaction is further demonstrated with a simulation that includes the temporal variation of both the vocal tract shape (see Fig. 2) and the F0 over a duration of 0.5 seconds. The glottal area signal was set to vary sinusoidally with the same maximum amplitude of 0.1 cm^2 as in the previous example, except that the F0 changed according to the contour shown by the white line in the lower part of Fig. 1b.



(a) (b)

FIGURE 3: Waveform and spectra of glottal flow generated by nonlinear source-filter interaction. (a) Glottal area (red) and glottal flow (black) waveforms; both have been normalized to an amplitude of 1.0 for purposes of comparison. (b) Spectrum of the glottal flow waveform (black) and frequency response the initial vocal tract configuration (blue dashed).

The glottal flow waveform generated during the simulation is plotted in five 0.1 second sections in Fig. 4a. The shape of flow pulses changes over the duration and this is due to the interaction of the glottal area with the tracheal and vocal tract pressures. The particular shape of any given flow pulse is dependent on both the F0 and resonance frequencies that are in play at that specific instant of time. Thus, with the exception of the F0, the characteristics of what is typically referred to as the voice *source* are, in this case, largely generated by the acoustic characteristics of the trachea and vocal tract. The output pressure waveform is similarly plotted in Fig. 4b and indicates typical differences in acoustic characteristics from cycle to cycle as expected for a vowel to vowel transition.



(a) Glottal flow (b) Output pressure

FIGURE 4: Waveforms resulting from [ɑi] simulation. (a) Glottal flow, (b) Output pressure

A narrow band spectrogram of the output pressure signal is shown in Fig. 5. It is similar in many ways to the spectrogram of the original recording in Fig. 1a. For example, the number of harmonics present during each portion of the utterance is comparable across both the recording and simulation. In addition, the presence of the noise component is fairly well represented by the simulation. Interestingly, the primarily difference seems to be that the harmonics in the simulation appear more prominent that in the natural production.

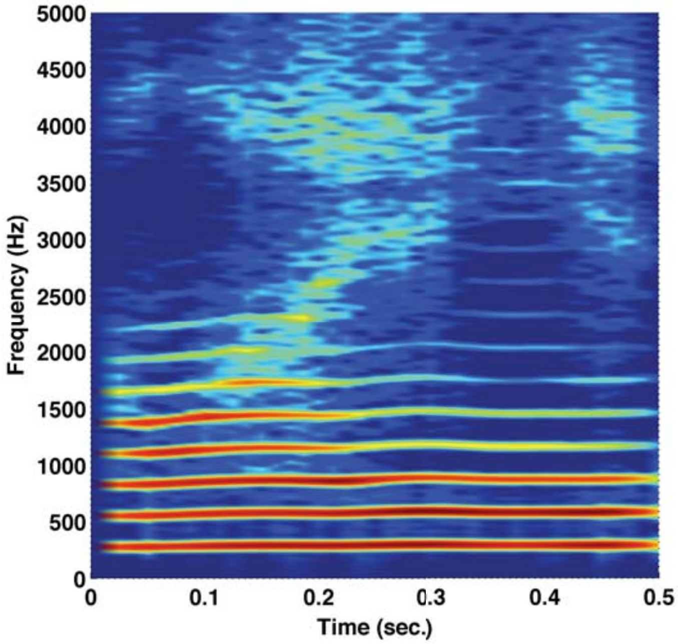


FIGURE 5: Something here.

DISCUSSION

Although a purely sinusoidal glottal area is unlikely to occur during real phonation, there is some evidence that suggests the pattern of vocal fold vibration in children may approach such a sinusoidal shape. Stathopoulos and Sapienza (1997) reported that the open quotient of the glottal airflow signal is higher for children and females than for adult males. They suggested that short durations of glottal closure (i.e., high open quotient), especially in children, may be due to the differences in laryngeal size and structure, or variations in the motor control of the intrinsic laryngeal muscles. A large open quotient may also be associated with aspiration noise which could provide an additional energy source for exciting the vocal tract resonances, especially in the vicinity of the third formant or higher (Klatt & Klatt, 1990; Hanson et al., 2001).

Based on Eqn. 1, generation of harmonics in the glottal flow with a sinusoidal (or nearly so) glottal area function is possible if 1) the mean glottal area is large, and/or 2) cross-sectional area at the vocal tract and tracheal entrances are small. The first condition is possible considering that tracheal pressure tends to be higher in children than adults (Stathopolous and Sapienza, 1997) and would likely contribute to high vibrational amplitude. In addition, the structure of the vocal ligament develops during early childhood (Titze, 1989; Hirano, 1983). Speculatively, higher vibrational amplitude could potentially be produced during the period prior to the ligament being fully developed. The effective amplitude of vibration in the simulation was a maximum glottal area of 0.1 cm^2. If this maximum area were reduced, the coupling would be tempered and the glottal flow waveform would exhibit less ripple, and consequently reduced harmonic energy. For the second condition there is little evidence, but it is known that adult vocal tracts may have entry areas on the order of 0.3 cm^2 (Story, 1995; Story et al., 1996, 1998) and tracheal areas for children have been reported to be of similar magnitude (Butz, 1969). Considering that the vocal tract structures are smaller in children that adults, entry areas at least as small as 0.3 cm^2 could be expected.

Conclusion

The results of this study contribute to a view that the mechanisms vocal sound production can be considered to exist along a continuum. At one end of the continuum the sound generated by vocal fold vibration is weakly coupled to the resonances of the vocal tract such that the output is a linear combination of their respective acoustic characteristics, whereas at the other end there is strong nonlinear coupling of the vibratory source to the vocal tract resonances. To adequately express acoustic properties in the output, such as vocal tract resonances, the *linear* case requires that the vibratory source produce sound that is rich in either harmonic or broadband energy, or both. In contrast, the *nonlinear* case allows for an harmonically-rich source signal to be generated even when the vibration pattern of the vocal folds is so simple that it may contain only one harmonic (i.e., sinusoid). The small sizes of the laryngeal and vocal tract systems of children lead to fundamental frequencies (F0) of vibration and vocal tract resonance frequencies (formants) that are higher than those of adults. These characteristics suggest that children's speech may be produced by mechanisms that lie closer to the nonlinear end of the continuum.

ACKNOWLEDGMENTS

Research supported by NIH R01-DC011275 and NSF BCS-1145011.

References

Bunton, K., Story, B.H. (2011). "A test of formant frequency analyses with simulated child-like vowels", J. Acoust. Soc. Am. 129(4) 2626.

Butz, R.O., (1968). "Length and cross-section growth patterns in the human trachea", Pediatrics 42, 336-341.

Fant, G. (1960). *Acoustic Theory of Speech Production.* The Hague: Mouton and Co.

Hanson, H.M., Stevens, K.N., Kuo, H-K. J., Chen, M.Y., an Slifka, J., (2001). "Towards models of phonation", J. Phonetics 29, 451-480.

Hirano, M., Kurita, S., and Nakashima, T., (1983). Growth, development, and aging of the human vocal folds, In Vocal Fold Physiology, DM Bless, J Abbs, Eds., San Diego, CA, College Hill Press, 22-43.

Klatt, D. and Klatt, L., (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers", J. Acoust. Soc. Am. 87, 820-857.

Liljencrants, J., (1985). *Speech Synthesis with a Reflection-Type Line Analog*, DS Dissertation, Dept. of Speech Comm. and Music Acoust., Royal Inst. of Tech., Stockholm, Sweden.

Stathopoulos, E.T., and Sapienza, C.M., (1997). "Developmental changes in laryngeal and respiratory function with variations in sound pressure level", J. Speech, Lang., Hear. Res. 40, 595-614.

Stevens, K.N., and House, A.S. (1961). "An acoustical theory of vowel production and some of its implications", J. Speech Hear. Res. 4(4), 303-319.

Stevens, K. N. (2000). *Acoustic Phonetics*, MIT Press, Cambridge, MA.

Story, B. H., (1995). *Physiologically-Based Speech Simulation Using an Enhanced Wave-Reflection Model of the Vocal Tract*, Ph. D. Dissertation, University of Iowa.

Story, B. H., (2006). “A technique for “tuning” vocal tract area functions based on acoustic sensitivity functions”, J. Acoust. Soc. Am. 119(2), 715-718.

Story, B. H., Titze, I. R., and Hoffman, E.A., (1996). “Vocal tract area functions from magnetic resonance imaging”, J. Acoust. Soc. Am. 100(1), 537-554.

Story, B.H., Titze, I.R., and Hoffman, E.A., (1998). “Vocal tract area functions for an adult female speaker based on volumetric imaging”, J. Acoust. Soc. Am. 104(1), 471-487.

Titze, I.R. (1984). Parameterization of the glottal area, glottal flow, and vocal fold contact area, J. Acoust. Soc. Am., 75, 570-580.

Titze, I.R., (1989). Physiologic and acoustic differences between male and female voices, J. Acoust. Soc. Am., 85(4), 1699-1707.

Titze, I.R., (2008). “Nonlinear source-filter coupling in phonation: Theory”, J. Acoust. Soc. Am. 123(5), 2733-2749.