Compensatory Strategies in the Developmental Patterns of English /s/: Gender and Vowel Context Effects

Hye-Young Bang, Meghan Clayards, and Heather Goad

McGill University

Abstract

Purpose: The developmental trajectory of English /s/ was investigated to determine the extent to which children's speech productions are acoustically fine-grained. Given the hypothesis that young children have adult-like phonetic knowledge of /s/, the following were examined: (1) whether this knowledge manifests in acoustic spectra that match the gender-specific patterns of adults, (2) whether vowel context affects the spectra of /s/ in adults and children similarly, and (3) whether children adopt compensatory production strategies to match adult acoustic targets.

Method: Several acoustic variables were measured from word-initial /s/ (and /t/) and the following vowel in the productions of children aged 2-5 and adult controls using two sets of corpora from the Paidologos database.

Results: Gender-specific patterns in the spectral distribution of /s/ were found. Acoustically more canonical /s/ was produced before vowels with higher F_1 (i.e., lower vowels) in children, a context where lingual articulation is challenging. Measures of breathiness and vowel intrinsic F_0 provide evidence that children use a compensatory aerodynamic mechanism to achieve their acoustic targets in articulatorily challenging contexts.

Conclusion: Together, these results provide evidence that children's phonetic knowledge is acoustically detailed and gender-specified and that speech production goals are acoustically-oriented at early stages of speech development.

Compensatory Strategies in the Developmental Patterns of English /s/: Gender and Vowel Context Effects

A long tradition of developmental studies assume (implicitly or explicitly) that what is heard by adult transcribers accurately reflects children's production targets; in other words, what adults transcribe represents what children intended to produce (e.g., Bernhardt & Stemberger, 1998; Ingram, Christensen, Veach, & Webster, 1980; Smith, 1973). Thus, children are typically described as substituting one sound for another, for example [t] for target /s/. However, there is reason to doubt the validity of this assumption. First, individual transcribers often differ in their transcriptions of the same productions, a problem that is amplified by the fact that it is impossible to record all the gradient and fine-detailed differences in the speech signal using an exhaustive set of phonetic symbols. This assumption is also challenged by the discovery of *covert contrast* in acoustic studies of children's productions. Covert contrast refers to acoustically subtle but systematic distinctions in children's productions of a contrast, which fail to meet the perceptual criteria of the adult listener, typically resulting in perceived homophony (Li, Edwards, & Beckman, 2009; Munson, Edwards, Schellinger, Beckman, & Meyer, 2010; Scobbie, Gibbon, Hardcastle, & Fletcher, 2000). The phenomenon has been widely documented for various speech contrasts, including place of articulation for obstruents (Baum & McNutt, 1990; Li et al., 2009), voicing contrasts for stops (Macken & Barton, 1980; Maxwell & Weismer, 1982), and contrasts between consonant clusters and corresponding singletons (Scobbie, et al., 2000). For example, Baum and McNutt (1990) found that /s/ misarticulated as [θ], which results in apparent homophony between /θ/ and /s/, is acoustically discernable from target /θ/ in the productions of English speaking children aged 5 to 8.

The existence of covert contrast suggests that children's phonological and/or phonetic knowledge is not always identical to what is perceived and recorded by adults. It further raises the possibility that children are attempting to produce adult acoustic targets, but because of their articulatory limitations, they are doing so using different production strategies than adults. The use of non-adult-like strategies by children has been observed by Ménard and colleagues

(Ménard & Noiray, 2011; Ménard, Schwartz, & Boë, 2004). They suggest that children use non-adult-like articulatory strategies to achieve adult-like acoustic targets in the production of French vowels. Ménard and Noiray (2011) utilized articulatory modeling and ultrasound to compare a 4-year-old speaker and an adult control and found that both the child and adult distinguished the vowels /i, a, u/ acoustically and articulatorily despite their anatomical differences, especially in the pharyngeal region. Most importantly, they found that the range of tongue curvature gestures for the same vowel set differed between the child and adult speakers. Their findings suggest that the way sound contrasts are realized is constrained by and adjusted with morphological growth.

The studies conducted by Ménard and colleagues are concerned with children's use of non-adult-like articulatory strategies to achieve adult-like auditory targets despite their immature motor control capacities and vocal tract morphology. We further extend this interpretation and propose that child-specific compensatory mechanisms may also be evidenced in productions that are perceived as incorrect or non-target-like by adults. Our reasoning behind this assumption is that these non-target-like outputs fail to reach adult auditory targets but they do not necessarily reflect the level of children's phonetic knowledge.

To test our hypothesis that non-adult-like mechanisms may be evidenced during speech acquisition, we focus on English /s/. The literature shows that /s/ is notoriously problematic for children to acquire: it exhibits great variability during acquisition, the development of auditorily target-like production is often protracted, and it is commonly associated with pathological speech (Gruber, 1999; Shriberg et al., 2003; Smit, Hand, Feininger, Bertha, & Bird, 1990). There is wide agreement that these problems are largely attributable to the articulatory difficulty of producing /s/ (Hardcastle, 1976; Ladefoged & Maddieson, 1996): target-like production of /s/ has to fulfill specific articulatory and aerodynamic requirements. For example, an intricate aeromechanical goal—air at a high volume rate flowing through a narrow constriction—must be achieved (Howe & McGowan, 2005; Scully, Castelli, Brearley, & Shirt, 1992; Shadle & Scully, 1995), which is possible only through skilled control of lingual gestures and force (Kent, 1992). Thus, /s/ is a good candidate for exploring whether children's phonetic knowledge is underrepresented in what is perceived by adults due to its demanding articulatory requirements coupled with children's limited articulatory capabilities.

Producing an /s/+vowel (/sV/) sequence requires complex coordination of the subparts of the tongue to maintain high turbulence throughout /s/ because both /s/ and the following /V/ share the main articulator. It has been well documented that young children have limited articulatory abilities—especially with respect to lingual gestures—limitations that are magnified by sequentially developing articulators (Cheng, Murdoch, Goozée, & Dion, 2007; Green, Moore, & Reilly, 2002; Kent, 2004; Vorperian et al., 2009, 2011). Because of the immature articulators that young children have and the characteristics of producing a string of speech sounds that requires well coordinated temporal and spatial movements of speech organs, we expect that the vowel context in which /s/ is produced will impact children's productions of this consonant more than it does adults'. Examining how children respond to /sV/ co-articulatory challenges will therefore allow us to investigate potential child-specific strategies.

The existence of covert contrast discussed above suggests that children's phonetic knowledge is richer than what is perceived and is grounded in adult targets. Additional support for this comes from studies that have documented gender-specific acoustic characteristics in children that match gender-specific patterns in adults (e.g., Busby & Plant, 1995; Fox & Nissen, 2005; Nissen & Fox, 2005). These studies suggest that gender information is widely available in the speech signal, for example, in fundamental frequency (F_0) and vowel formants in adults' speech (Bachorowski & Owren, 1999; Ladefoged & Broadbent, 1957; Perry, Ohde, & Ashmead, 2001). Gender-specific differences in vowel production have also been detected in prepubescent children. For instance, Busby and Plant (1995) examined vowel production in 5-, 7-, 9-, and 11-year-olds and found that F_0 decreases as a function of age for both boys and girls. However, they found that while vowel formants remain higher for female children across all age groups, they consistently decrease with age in male children, showing gender divergence in speech cues.

One important source of gender-related acoustic variation in adults is the gender-specific morphology of the vocal tract and articulators (Fitch & Giedd, 1999; Titze, 1989). F_0 is modulated by the length, thickness, stiffness, and mass of the vocal folds and formant patterns for vowels are mainly determined by the shape and size of the vocal tract (Fant, 1960; Stevens & House, 1955; Titze, 1994). However, the observed gender differences cannot be accounted for solely by the anatomical differences between men and women. Socio-phonetic factors are also known to affect gender-related differences in speech production (Fant, 1975; Diehl, Lindblom, Hoemeke, & Fahey, 1996; Hillenbrand, Getty, Clark, & Wheeler, 1995; Simpson, 2009). That is, gender-specific patterns arise in large part due to the fact that male and female speakers learn to speak according to gender norms (Eckert, 2008; Labov, 1990). For instance, by comparing vowel dispersions between various languages or dialects of the same language, studies have found that the magnitude of the differences in vowel dispersions between men and women shows cross-linguistic and cross-dialectal variation (Henton, 1989; Johnson, 2006).

As for children, developmental sex differences have been discovered in vocal tract growth (Vorperian et al., 2009, 2011; but cf. Fitch & Geidd, 1999; Perry et al., 2001) and head circumference growth (Nellhaus, 1968), which can account for some aspect of sex-specific differences in acoustic speech outputs in prepubescent children. Interestingly, anatomical differences in the vocal tract subsequently disappear and reappear after puberty (Vorperian et al., 2011) while gender differences in speech acoustics steadily increase with age starting as young as age 4 (Perry et al., 2001), which suggests that anatomical differences alone cannot explain the consistently observed differences in sex-specific acoustics.

Gender differences in acoustics have also been reported for /s/. Studies of adult speech have shown that /s/ has a higher center of gravity and higher spectral peaks in women compared to men (Fox & Nissen, 2005; Jongman, Wayland, & Wong, 2000; Maniwa, Jongman, & Wade, 2009; Nittrouer, Studdert-Kennedy, & McGowan, 1989). Other spectral properties that can signal gender identification include spectral variance and kurtosis, which are higher for females (Jongman et al., 2000), spectral skewness, which is higher for males (Fox & Nissen, 2005; Jongman et al., 2000), and spectral slope, which is higher for females (Fox & Nissen, 2005).

Several studies have also detected gender-specific patterns in English /s/ in prepubescent children. Flipsen, Shriberg, Weismer, Karlsson, and McSweeny (1999) found gender differences in 9-15 year olds for center of gravity, standard deviation, and skewness obtained from different temporal points in /s/. Romeo, Hazan, and Pettinato (2013) observed gender differences in variability of center of gravity at age 11. Fox and Nissen (2005) reported gender differences in children as young as 6 on a variety of spectral measures including center of gravity, skewness, kurtosis, spectral peak, and spectral slope. Nissen & Fox (2005) observed gender differences in younger children (aged 3 to 5), although it was restricted to spectral slope. Gender differences in sibilants have also been observed in 5-year-old Mandarin speaking children (Li, 2011).

A recent study of various English sounds including /s/ compared boys with Gender Identity Disorder (GID), a condition where individuals are discontent with the biological sex they were born with, and boys in a control group (Munson, Crocker, Pierrehumbert, Owen-Anderson, & Zucker, 2015). The authors found that word-initial /s/ tokens with higher spectral mean, greater variation (diffuseness), and more negative skewness were associated with less boy-like ratings. This study clearly supports the view that the emergence of gender-specific patterns in children's speech cannot be reduced to biological differences but is instead at least partly due to children detecting subphonemic gender differences in the acoustic input and learning the different ways that men and women produce sounds (Avery & Liss, 1996; Fuchs & Toda, 2010).

Taken together, the literature examining acoustic markers of covert contrast and gender differences suggests that throughout the course of development, children attempt to match their productions to adult-like acoustic targets. As such, speech development involves continuous acoustic and articulatory calibration as the physiological system matures, all the while affected by socio-phonetic factors. In this paper, we explore the hypothesis that if children's phonetic knowledge is grounded in acoustically specified adult targets while their articulatory systems—and especially their lingual gestures—lack adult-like control, they may use child-specific non-lingual production strategies to handle articulatorily challenging contexts. Before testing this possibility, we discuss how canonical /s/ is achieved in different vowel contexts; that is, what the articulatory and aerodynamic requirements are to consistently produce /s/ that is resistant to the effects of the following vowel. To understand the relation between acoustics and articulation in the production of /s/, we first detail its production requirements, in the following section.

Production of Canonical /s/ in English

The production of canonical /s/ in English has several aerodynamic and gestural constraints, which likely cause challenges for children. The characteristic stridency of /s/ is formed by a jet produced from a narrow channel passing over the front part of the grooved tongue raised against the hard palate; the jet of air forcefully hits the front teeth and, as a result, the magnitude of turbulence and amplitude of noise increase (Shadle, 2012; Shadle & Scully, 1995; Stevens, 1998). Studies suggest that there are several factors that contribute to a high spectral center of gravity (CoG) in the noise portion, which is known to be one of the primary cues shaping the characteristics of /s/. First, the small constriction area during the production of /s/ prevents acoustic coupling, resulting in cancellation of back-cavity resonances, which inhibits the excitation of lower frequencies and contributes to an increase in CoG. Second, the small constriction area increases the velocity of air particles, leading to a greater generation of turbulence and excitation of high frequencies. A tight constriction made more anteriorly is also a crucial factor in producing canonical /s/ with high CoG; a shortened cavity, which acts as a filter, and/or close proximity between the teeth and the sound source will enhance the turbulent noise that excites the relevant resonances. Lastly, turbulent noise is enhanced with an increase in volume airflow (Catford, 1982; Stevens, 1998). As such, production of canonical /s/ requires fine control of complex articulatory gestures involving the tongue, teeth, jaw, larynx, and lungs. Therefore, whatever sound follows /s/ should increase this complexity to different extents.

A comprehensive study that examined direct acoustic–articulatory relations in the production of /s/ in varying contexts suggests that adults can maintain constant constriction degree and location for prevocalic /s/ until the end of the segment by adopting a compensatory tongue tip motion that moves in the opposite direction to the jaw (Iskarous, Shadle, & Proctor, 2011). However, this articulatory strategy may not be available to young children, given that their speech organs, including the lingual muscles as well as oral and pharyngeal cavities, are functionally and anatomically different from adults (Green et al., 2002; Kent, 2004; Vorperian et al., 2009, 2011). In addition, fine control of the subparts of the tongue is protracted during physiological development (Cheng et al., 2007). This may lead to children having difficulty maintaining the narrow constriction degree that characterizes /s/ in certain vowel contexts. For example, the production of /sa/ requires rapid movement of speech articulators to sequentially form a narrow constriction for /s/ and a wide opening for /a/ (Honda, 2008). In contrast, the requirements for articulatory mobility to produce the sequence from /s/ to /i/ for /si/ are less extreme. From an articulatory perspective, then, all else being equal, children are expected to produce acoustically more canonical /s/ before high vowels than low vowels. If children instead produce more canonical /s/ in articulatorily challenging contexts, this will provide evidence that their productions do not completely rely on articulatory capabilities but that children strive to find a way to utilize their speech system to overcome context-driven challenges.

We define more canonical /s/ as having the characteristic spectral distribution of excited high frequencies and well-attenuated low frequencies, which can be represented as higher CoG, lower SD, lower skewness, greater spectral slope, and higher noise amplitude. If children's productions are the result of managing articulatory constraints and thereby striving to reach adult-like acoustic targets through compensatory mechanisms, then we expect to see more limited effects of the following vowel or even more canonical acoustics of /s/ in articulatorily challenging vowel contexts, which would likely be the result of overshoot. For the case of /s/, we expect such mechanisms to be targeted at maintaining frication through effortful noise production by increasing transglottal airflow to compensate for a less tight constriction in low vowel contexts.

We expect increased airflow during /s/ to have carryover effects into the following vowel, as the increase in transglottal airflow is accompanied by an increase in glottal opening, subglottal pressure, and stiffness in the vocal folds. These effects on glottal state can be estimated from certain acoustic dimensions that are sensitive to the extent to which noise encroaches into the vowel. In the current study, we adopt three acoustic measures, H1-H2 and H1-A3 to assess breathiness and vowel intrinsic F_0 (IF_0) to assess glottal tension or stiffness (DiCanio, 2009, 2012; Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995; Honda, 1983, 2008; Kirk, Ladefoged, & Ladefoged, 1984). We discuss these in detail in the following two sections.

Breathiness (H1-H2, H1-A3)

Due to their lack of adult-like control of lingual gestures, our prediction is that children's compensatory strategy to approximate canonical /s/ will involve an aerodynamic mechanism; children will use a higher degree of airflow or more extreme glottal gestures for noise maintenance or enhancement, especially in articulatorily challenging vowel contexts. This strategy is expected to involve a wider glottal opening or higher subglottal pressure for enhanced frication noise, which will spill over into the vowel during the production of /sV/ syllables.

Studies have found that in vowels produced with relatively large glottal openings as well as vowels at fricative/vowel boundaries, higher harmonics are dampened relative to lower harmonics, when compared to vowels produced with modal phonation (Dicanio, 2009; Stevens, 1998, p. 426). There is also an increase in aspiration noise due to increased airflow passing through a wider opening (Stevens, 1998, p. 91). Thus, in breathy phonation, the noise component is higher in amplitude while the periodic component loses its strength, especially at higher frequencies, causing steeper spectral tilt. Accordingly, breathiness or noise encroachment has been computed acoustically through different measures of spectral tilt that capture the difference in the amplitudes between lower and higher harmonics (DiCanio, 2009, 2012; Kirk et al., 1984). Two measures are most commonly used. These are H1-H2: the difference between the amplitude of the first (H1) and second harmonics (H2) in the Fourier spectrum; and H1-A3: the difference between the amplitude of the first harmonic (H1) and the amplitude of the third formant (A3). H1-H2 is considered to be related to the Open Quotient (OQ), the proportion of the glottal open phase in a glottal cycle (Holmberg et al., 1995; Stevens, 1998). H1-A3 is proposed to be related to Speed Quotient (SQ)—the closing velocity of the vocal folds—and possibly to muscle tension (Keating & Esposito, 2006) and is a good measure of global changes in the slope of the spectrum (Dicanio, 2012; Hanson & Chuang, 1999; Pennington, 2005).

The degree of OQ and SQ are also affected by the degree of oral constriction during vowel production. Studies based on electroglottographic waveforms suggest that [i] has higher OQ and SQ than lower vowels, both of which indicate a longer opening phase relative to the closing phase of the vocal folds (Chen, Robb & Gilbert, 2002; Higgins, Netsell, & Schulte, 1998). These effects are proposed to result from greater vocal tract tension or elongation (Higgins et al., 1998) and/or smaller transglottal air pressure (Bickley & Stevens, 1986) for the production of [i] compared to lower vowels. Taken together, these studies predict higher H1-H2 and H1-A3 for high vowels if these vowels are produced with modal phonation. This vowel height dependent voice quality is predicted to be disturbed when vowels are produced with a different degree of breathiness caused by the control of airflow in children's productions.

Vowel Intrinsic F_0

F_0 is mainly modulated by two factors: length of the vocal folds and aerodynamics. Thinner and longer vocal folds and the resultant decrease in mass per unit length are associated with higher F_0 and length of the vocal folds can be affected by adjusting the cricoid and thyroid cartilages (Honda, 1983, 2008). F_0 also increases with an increase in transglottal airflow rate accompanied by greater subglottal air pressure and tissue stiffness in the vocal folds.

High vowels tend to be produced with higher F_0 and this is considered to be due to anatomical linkages between the tongue root and larynx (Honda, 1983, 2008). The intrinsic F_0 difference between high and low vowels (i.e., IF_0 difference) has been consistently observed across languages (e.g. Lehiste, 1976; Whalen & Levitt, 1995). For example, Whalen and Levitt (1995) performed a meta-analysis of 31 languages analyzed in 58 studies and confirmed this pattern. IF_0 has even been observed in infant babbling (Whalen, Levitt, Hsiao, & Smorodinsky, 1995) and in the speech of deaf children (Bush, 1981), both of which strengthen the view that it is the result of an interaction between articulatory and aerodynamic mechanisms (but cf. Connell, 2002; Van Hoof & Verhoeven, 2011).

If children with immature lingual gestures attempt to produce canonical /s/ in articulatorily challenging contexts, namely in low vowel contexts, they may achieve this by increasing airflow. This would involve increased subglottal pressure and higher airflow rate accompanied by a more widely open glottis and/or stiffer vocal folds during frication production. If these phonation parameters continue into the following vowel, this in turn should increase the rate at which the vocal folds vibrate, increasing the F_0 for low vowels and reducing the IF_0 gap between high and low vowels. Based on this physiological mechanism, we expect to observe that the magnitude of the IF_0 difference between high and low vowels differs between adults and children in vowels following /s/.

Research Goals

To summarize thus far, we hypothesize that children have adult-like, gender-specific phonetic knowledge of /s/ and that this knowledge is manifested in speech acoustics even when children's productions are not perceived as target-like. We further hypothesize that, even at early stages in speech development, production goals are acoustically-oriented, while at the same time, children's articulatory abilities are limited, and that to resolve this conflict, children use non-adult-like compensatory strategies to produce sounds that are closer to adult targets. In successful attempts, children's outputs should be perceived as target-like by adults. We explore these hypotheses by testing three predictions: (1) children's knowledge of /s/ should manifest in acoustic spectra that match the gender-specific patterns of adults; (2) children's compensatory mechanism adopted to handle articulatorily challenging contexts should be reflected as acoustic overshoot in low vowel contexts; and (3) children's compensatory strategy will involve an aerodynamic mechanism, evidenced as voice quality carry-over and an attenuation of the universal patterns of vowel height.

To test these predictions, we examine the effects of age, gender, and vowel context on the acoustics of /s/ in two corpora of child and adult speech. We evaluate the first two predictions by measuring multiple acoustic correlates of /s/ including spectral center of gravity, standard deviation, skewness, and slope of the spectral distribution. In addition to these, we measure intensity in /s/ and in the following vowel. We also extract the first formant frequency in the following vowel as a proxy of tongue height. To assess our third prediction, we additionally extract H1-H2 and H1-A3 to measure breathiness or the degree of noise encroachment in the vowel that follows /s/ as well as /t/. We included /t/ in this analysis as a control. Our hypothesis about articulatory limitations and subsequent compensatory mechanisms is limited to /s/ (and other sibilant fricatives) and so we do not make the same predictions for other sounds that share the same primary articulator (tongue tip/blade) and glottal abduction gesture but do not require strident airflow. Finally, we measure F_0 as an estimate of vocal fold tension in the vowels that follow both /s/ and /t/.

Method

Corpus Data

The data that we use to test our predictions are word-initial [sV] and [tV] productions from 79 monolingual English-speaking children aged 2 to 5 (39 females, 40 males; mean age 3;7) and 20 college-age adult controls (10 men, 10 women). The data were accessed from the publically available Paidologos corpus (Edwards & Beckman, 2008), available on the CHILDES website (http://childes.psy.cmu.edu/). The dialect of both child and adult participants was controlled by recruiting all participants from the Columbus, OH metropolitan area (Li, 2008; Reidy, 2015).

The Paidologos project investigated the acquisition of various word-initial lingual consonants across five languages (details available at https://www.ling.ohio-state.edu/~edwards/). The English data were collected by recording children's speech during a picture-prompted repetition task of single word productions. The experimenter controlled a "show and play" script, which presented visual images of the target words along with pre-recorded labels as prompts. The auditory prompts were produced in child-directed speech by a phonetically-trained, female native speaker of English who was from the same dialect region as the participants. Productions of single words were digitally recorded at a sampling rate of 44100 Hz at 16 bits. The consonants were coded as either 'correct' or 'incorrect' by a trained native English-speaking phonetician whose decisions were based on auditory-perceptual judgments and examination of the acoustic waveform of each sound. A second phonetician independently transcribed about 20% of the tokens and high inter-rater reliability (about 90%) was found (for further details, see Edwards and Beckman, 2008; Li, 2009; Reidy, 2015). The percentage of correct productions increased with age (2-year olds: 48%, 3-year olds: 60%, 4-year olds: 67%, and 5-year olds: 83%). For incorrect tokens, substitutions were noted. For /s/, the most common substitution errors for 2- to 3-year-olds included fronting (/s/ → [f, v, θ]), backing (/s/ → [ʃ]), and stopping (/s/ → [tʰ, t] (Li, 2008).

Initial /s/ and /t/ occurred in nine different vowel contexts, /i, u, ɪ, eɪ, oʊ, ɛ, ʌ, ɔ, ɑ/. We extracted 1,211 sound files containing /s/ initial words and excluded files with noise interruptions, inaudible or unrecognizable words, and random productions (words not in the list) before any measurements were taken (n=63). Vowel formants were automatically extracted using a script created in Praat (Boersma & Weenink, 2013) and were then hand-corrected. At this point, five more tokens were excluded due to their unusually large formant values. The corpus also includes judgments on whether the target sound is phonemically correct or incorrect. Since our hypothesis states that children's phonetic knowledge is manifested even in productions that do not meet adult perceptual criteria for target-like, we did not exclude 'incorrect' tokens. In total, 1,143 tokens from 15 /s/ initial words from 39 girls and 40 boys were analyzed. The /s/ initial words are 'seal', 'seashore', 'soup', 'suitcase', 'super', 'sister', 'safe', 'same', sodas', 'soak', 'soldier', 'seven', 'sun', 'sauce', and 'soccer'. We additionally extracted 1,205 /t/ initial tokens from the child corpus. Fourteen tokens were excluded due to noise interruption or mispronunciation, leaving a total of 1,191 tokens from 15 /t/ initial words for analysis. The /t/ initial words were 'taco', 'tail', 'tall', 'teacher', 'tent', 'tepee', 'tube', 'tuna', 'tickle', 'tongue', 'toad', 'torn', 'toast', 'taste', and 'tooth'.

The adult Paidologos data were collected using the same picture-prompted repetition task of single words that was used to elicit the children's productions. However, unlike with the children, who were recorded at day care centers or in their homes with the tester controlling the script, the adults were recorded in the sound booth in the phonetics lab at Ohio State University, and the participants were left alone to control the program at their own pace. As with the children, the adults' single word productions were recorded using a sampling rate of 44100 Hz at 16 bits. Two female speakers mispronounced 'same' as 'fame', so we excluded these productions from the data. After excluding seven more tokens that either lacked distinct frication noise visible in the spectrogram or contained substantial echoes, a total of 292 tokens from 10 female and 10 male speakers were included in the analysis. For word initial /t/ tokens, we extracted 300 tokens, but one token was excluded due to noise interruption, leaving a total of 299 tokens for analysis.

Hand Segmentation

The original adult and child corpora contained defined segment boundaries inserted by trained phoneticians. We manually checked all of these for consistency using Praat (Boersma & Weenink, 2013). The consonant-vowel boundaries were adjusted to the point where a zero-crossing of the periodic wave form is as close as possible to the onset of the first formant band in the spectrogram for both /s/ and /t/. This method was used to ensure that the acoustic output of each vowel token represents the articulatory gestures for that vowel rather than for the preceding fricative. If the first zero-crossing of a periodic portion did not coincide with the beginning of the first formant in the spectrogram, then the first subsequent one that coincided with the first formant was marked as the boundary. Boundaries between the vowel and following consonant were also carefully marked to exclude any portion of the consonant. For example, a boundary was drawn at the end of the periodic pulse when a stop followed. Auditory judgments were used when the following sound was a glide or nasal as they usually lack visually discernible boundaries. When a vowel was followed by a voiceless fricative, the boundary was carefully drawn by using auditory judgment and by examining the spectrogram in order to exclude voiceless portions resulting from early glottal opening gestures at the end of the vowel (Clayards & Knowles, 2015; Niebuhr, Clayards, Meunier, & Lancia, 2011). These breathy portions had the property of faint formants and aspiration like noise in the spectrogram and little or no voicing.

Spectral Measures of /s/

Before calculating any measures of frication noise from these adjusted intervals, frequencies below 550 Hz were filtered using the 'Stop Hann Band' filter built in Praat. This filtering was performed to remove low-frequency ambient noise in frication and to exclude the voicing portion of the following vowel from intruding into the preceding frication noise (Koenig, Shadle, Preston, & Mooshammer, 2013). From each of the filtered tokens, we obtained four measures—CoG, SD, Skewness, and Spectral Slope—that reflect spectral features of /s/.

Spectra of frication noise can be described as random probability distributions (Forrest, Weismer, Milenkovic, & Dougall, 1988) and the first three moments represent the center of gravity (CoG, also called the mean, the centroid, M1, or L1), standard deviation (SD, also called M2 or L2), and Skewness (also called M3 or L3). These measures were obtained using DFTs (Discrete Fourier Transforms) computed by averaging six spectra taken at different time points across the fricative (following Shadle, 2012). Each spectrum was created from a 15 ms long window. Windows were evenly distributed across the middle 80% of the fricative. The information present in the remaining 20% at the fricative edges was excluded to remove possible inclusion of acoustic information from the immediately adjacent sound and weak noise at the beginning of frication. For shorter fricatives (middle 80% < 90 ms), the analysis windows overlapped and for longer ones (middle 80% > 90 ms), they were evenly spaced apart.

Spectral Slope represents the logarithmic power spectral density difference between the high and low regions of the spectrum. The values were extracted from the Long-Term Average Spectrum (LTAS) of the middle 50% of each fricative token. Low and high frequency regions were set at 0-8000 Hz and 800-1500 Hz respectively following Maniwa et al. (2009). No pre-emphasis was applied to any of the measures following previous studies (e.g., Koenig et al., 2013).

Vowel Measures

All vowel measures were taken from the original sound files before filtering was performed. We extracted the first formant (F_1) values using a 5 ms spectrogram window taken at the one third point of the vowel using a Praat script with the criteria set at five formants below 5000 Hz for male adult speakers, 5500 Hz for female adult speakers, and 6000 Hz for child speakers. Each speaker's formant distribution by vowel was plotted using histograms and values at both edges were checked and hand-corrected. The F_1 values were subsequently normalized to minimize vocal tract differences between speakers. The normalization was performed based on the Nearey 2 method using the 'norm.nearey' function embedded in the R package called 'Vowels' (Kendall & Thomas, 2012). This process is important because the data contain tokens from children and adults as well as male and female speakers. The formant measures were taken at the one third point rather than at the center because some vowels were diphthongs or diphthongized in the children's data and we were interested in observing the relationship between /s/ and the following target vowel, which is the first part of the vowel in the case of diphthongs. Moreover, due to the possibility of early anticipatory coarticulation in children's speech (Katz & Bharadwaj, 2001; Nittrouer et al., 1989), the midpoint of the vowel could contain information from the following consonant or display greater variability. For consistency, the same criterion was applied to the adult data as well. All diphthongs were then manually checked to ensure that we extracted acoustic cues from the first part of these sounds only.

F_0 from each vowel token was extracted automatically in Praat using the autocorrelation-based algorithm with a 25 ms Gaussian analysis window at the midpoint of the vowel. F_0 values were hand-corrected before converting to semitones, which represent equal perceptual intervals and minimize physiologically-based pitch differences (Nolan, 2003). The conversion into semitones was made according to the following formula with a reference of 100 Hz:

$$Fi(st) = 12 * \log2 (Fi\ (Hz) / 100\ (Hz))$$

The three measures that estimate phonatory characteristics of vowels—H1, H2, and A3—were also taken at the midpoint of the vowel using the same analysis method as F_0. These values were then corrected to minimize the influence of neighboring formants on each measure using the formula in Iseli and Alwan (2004) and Iseli, Shue, and Alwan (2007), using a Praat script. For H1 and H2, the correction was done to remove the influence of the first and second formants respectively. For A3, the first three formants were corrected using the same method. As was the case for the spectral measures of /s/, no pre-emphasis was applied to any of these measures.

Intensity Measures

The root-mean-square (rms) value of the sound pressure (intensity (dB)) was taken from the fricative and the following vowel for each sound file. The intensity of /s/ was taken from the sound files after the filtering was applied while that of V was taken from the original sound files. The measure was taken from the middle 80% of the /s/ and from the middle 80% of the following vowel to minimize inclusion of information from the adjacent sounds. For example, in the case of 'seal', the very beginning of the vowel may contain information from the preceding /s/, and the end may contain information from coda /l/, which would likely decrease vowel intensity. Similar to the spectral measures, an average value was calculated from each of six 15 ms long windows, evenly distributed across the middle 80% of each interval. Then, the sum of the values from the six windows was divided by the number of windows to obtain the average intensity in dB.

As the final step, we calculated the relative intensity of /s/ to the following vowel by subtracting V-intensity from /s/-intensity in order to control any effects of the recording environment such as differences in the distance between talkers and the microphone or indexical characteristics such as the volume of a talker's voice. The value was then labeled as *IntDiff* (intensity difference).

Results

Spectral Distributions

The effects of gender and vowel context on the spectral distributions of /s/ were statistically tested with mixed-effects regression models, which take into account the dependence of observations within a group. In all models throughout the paper, *p* values were calculated based on Satterthwaite approximations to degrees of freedom using the 'lmerTest' package (Kuznetsova, Brockhoff, & Christensen, 2013). We built our models using the lmer() function from the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2014). We first built four models for each of the two datasets (child and adult) by introducing CoG, SD, Skewness, and Spectral Slope separately as a response variable. Each model was built with three predictors: F_1, IntDiff, and Gender, and tested whether and how each of these predictors contributes to the distribution of each of the response variables. F_1 was included as an acoustic proxy of tongue height to observe how each component of the spectral distribution changes as a function of tongue raising/lowering gestures. IntDiff was introduced to address how the amplitude level of frication noise modulates the spectral distribution of /s/. Gender was included to test for gender-specified information in the spectral output of both children and adults. No interaction term was included in the adult models because the likelihood ratio tests using the *anova* () function confirmed that no interaction significantly contributed to any of the four models. In the models for the child data, Age was additionally included to examine the developmental trajectory of the spectral distribution of /s/ and at what age gender divergence occurs by considering its interaction with Gender. A likelihood ratio test confirmed that the interaction term Age*Gender significantly contributes to CoG, Skewness, and Spectral Slope models in the child data with *p* values less than 0.001 while it marginally contributes to the SD model ($\chi 2 = 3.45$, $p = 0.064$).

We also assessed whether F_1 (vowel height) modulates /s/ production differently for children of different ages by adding the interaction term, Age*F_1, into the models. However, this term did not return a significant effect in any of the models and the likelihood ratio tests confirmed that the interaction did not improve the model predictions. Therefore, the interaction was excluded from the analyses. Finally, we included Correct in the child models to test whether tokens judged by the transcribers to be correct versus incorrect would differ with respect to spectral distributions of /s/. Likelihood ratio tests confirmed that Correct significantly contributed to CoG, Skewness, and Spectral Slope in the child data. Note, however, that an additional 117 tokens were excluded from the statistical analyses whenever Correct was included as a variable in the models because correctness judgments were missing for these tokens.

In preparation for the statistical analysis, Age was coded in intervals of 0.5 years (from 2 to 5.5) and was considered a continuous variable. F_1 and IntDiff were also treated as continuous variables while Gender and Correct were coded based on sum-coding (-0.5 and 0.5). To avoid collinearity and to make the effect sizes comparable, all continuous variables were centered and divided by 2 standard deviations.

In the models throughout the paper, by-speaker random intercepts and slopes were included for all within-speaker factors (i.e., all factors except Age and Gender) and by-vowel random factors for relevant within-vowel factors. To be more specific, the statistical models for adults included by-speaker random intercept and slope for F_1 to allow for speaker-specific variation in the spectral cues and speaker-specific variation for F_1 effects. By-vowel random intercepts and random slopes for Gender and IntDiff were additionally included to allow for vowel-specific variation and variation in the coefficients of the fixed effects by vowel. The inclusion of by-vowel IntDiff random effects is expected to control for vowel-inherent amplitude as low vowels have greater amplitude than high vowels (Lehiste & Peterson, 1959). In the child model, in addition to the random effects introduced in the adult models, by-speaker random intercept and slope for Correct and by-vowel random intercept and slope for Age were added. Finally, our models included all possible correlations between the random effects.

Adults.

The statistical results for the adults are summarized in Table 1. Significant main effects of Gender were found for CoG, Skewness, and Spectral Slope in adults. CoG was significantly lower for males than females ($\hat{\beta}$ = -2109.9, t = -5.56, $p < .0001$), Skewness was significantly higher for males than females ($\hat{\beta}$ = 1.092, t = 4.64, p=0.0002), and Spectral Slope was significantly lower for males than females ($\hat{\beta}$ = -7.788, t = -4.49, p = 0.0002). The directions of the spectral components in females versus males mirrored those for clear versus casual speech (Maniwa et al., 2009), supporting the finding that women tend to hyper-articulate (Labov, 1990). No significant gender difference in SD was found.

Insert Table 1 about here

Turning to the intensity measure (IntDiff), we found that /s/ with higher intensity relative to the following vowel had higher CoG ($\hat{\beta}$ = 383.4, t = 2.21, p = 0.036), lower SD ($\hat{\beta}$ = -536.4, t

= -4.41, p = 0.001), and greater Spectral Slope ($\hat{\beta}$ = 3.637, t = 5.19, p <.0001). The top row of Figure 1 shows the relationship between relative fricative intensity and the four spectral measures for adults.

Insert Figures 1 and 2 about here

The effect of F_1 in the adult data (see top panels in Figure 2) on the four spectral cues did not provide strong evidence that the spectral cues in adults' productions of /s/ are modulated by varying vowel height.

Children.

The statistical results for the children are summarized in Table 2. Older children had lower SD ($\hat{\beta}$ = -307.4, t = -2.35, p = 0.021), indicating that the spectra of /s/ produced by older children are characterized by an acutely excited small range of frequencies compared to younger children. On the other hand, older and younger children across genders did not significantly differ in CoG, Skewness, and Spectral Slope.

Insert Table 2 about here

Statistically significant gender divergences were also found in the child results. Male children produced significantly lower CoG than female children across all age groups (Gender: $\hat{\beta}$ = -764.0, t = -2.46, p = 0.016). Spectral Slope was also significantly lower for male children ($\hat{\beta}$ = -2.281, t = -2.44 , p = 0.017). These two main effects were consistent with the patterns found in adults. However, unlike adults who did not show a gender difference in SD, there was a marginal effect of Gender on SD; male children tended to have lower SD than female children across age groups ($\hat{\beta}$ = -221.0, t = -1.84, p = 0.070). However, this effect did not reach statistical significance.

Gender-specific spectral cue developments are illustrated in Figure 3. In the statistical results, significant interaction effects of Age*Gender were found for CoG ($\hat{\beta}$ = -2036, t = -3.05, p = 0.003), Skewness ($\hat{\beta}$= 0.546, t = 2.10, p = 0.038), and Spectral Slope ($\hat{\beta}$ = -6.45=6, t = -3.22, p = 0.001) while the interaction was marginal for SD ($\hat{\beta}$= -477.3, t = -1.84, p = 0.069). The interaction effects and the patterns illustrated in Figure 3 suggest that

Gender divergence in the four spectral cues emerges at age 4 or 5. Interestingly, the significant gender divergence in older children was manifested in the same spectral cues as in the adults. The two marginal effects, Age and Age*Gender for SD provide evidence (although statistically weak) that the linear decrease in SD with age was more robust for male children than for female children.

Insert Figures 3 and 4 about here

As with adults, IntDiff significantly increased CoG ($\hat{\beta}$= 439.6, t = 3.86, p = 0.004) and Spectral Slope ($\hat{\beta}$= 2.605, t = 7.32, p < .0001) and it significantly decreased SD ($\hat{\beta}$ = -405.7, t = -6.78, p=0.0002) (Figure 1 bottom panels). The effects of IntDiff in children, as was the case in adults, suggest that greater frication noise was correlated with acoustically more canonical /s/.

The effects of Correct confirm that tokens with higher CoG ($\hat{\beta}$= 1111.5, t = 8.29, p < .0001), greater Spectral Slope ($\hat{\beta}$= 4.68, t = 9.77, p < .0001), and lower Skewness ($\hat{\beta}$= -0.346, t = -4.94, p < .0001) were likely to be perceived by adults as target-like. Figure 4 illustrates how the probability of correct judgments were correlated with the three spectral cues.

Interesting patterns were observed in the effects of F_1 in the child data (Figure 2 bottom panels). In contrast to the adult data, the statistical results showed that as F_1 increased, CoG ($\hat{\beta}$ = 514.7, t = 3.89, p= 0.0001) and Spectral Slope ($\hat{\beta}$= 1.656, t = 3.78, p= 0.0002) significantly increased and Skewness significantly decreased ($\hat{\beta}$= -0.14, t = -2.57, p = 0.014) while SD did not seem to be significantly modulated by F_1. The effects of F_1 on the three spectral components suggest that children produced more canonical /s/ when F_1 of the following vowel was higher (i.e., in lower vowels), which we expected to be an articulatorily more challenging environment in which to coproduce /s/ with the following vowel.

To sum up, we found that gender-specific patterns clearly emerged in CoG, Skewness, and Spectral Slope at around 4-5 years of age. These patterns were consistent with their adult gender-matched groups. Our results also showed that /s/ produced with higher relative amplitude increased CoG and Spectral Slope but lowered SD, compatible with the spectral patterns for canonically produced /s/. We also confirmed that children's tokens that have more canonical acoustic features of /s/ were more likely to be perceived as correct tokens.

More crucially, interesting differences between adults and children were detected. The lack of significant F_1 effects in the adult data partly coincides with an earlier study in which consistent constriction formations were observed among adults in the production of /s/ in varying contexts to "maintain an invariant articulatory–aerodynamic goal" (Iskarous et al., 2011, p. 944). However, the effect of vowel height on CoG (or on the first spectral moment, M1) reported in this earlier study is not consistent with the lack of F_1 effects found in our study. CoG was higher for low vowels than high vowels in the study by Iskarous et al. but was not significantly affected by vowel height in our adult data. The conflicting results between these two studies will be discussed further below.

Our child data, in contrast with the results for adults, support our second prediction that children will produce more canonical /s/ before low vowels, a context that should create a greater production challenge. This mismatch between articulatory demand and acoustic output supports our hypothesis that children will adopt a compensatory mechanism in order to achieve their production goals. The mismatch, then, appears to be the result of overshoot in articulatorily challenging contexts.

Our third prediction was that children will use a different strategy from adults, possibly involving an aerodynamic mechanism, due to their limitations in lingual dexterity. In what follows, we test this prediction based on three acoustic parameters for vowels: H1-A3 and H1-H2, which are estimates of the level of breathiness, and IF_0, which estimates how vertical tongue body gestures and their interaction with the status of the vocal folds modulate F_0 of the vowel. As discussed earlier, we expect breathiness and F_0 to be greatest for high vowels in adults (e.g., Chen et al., 2002; Higgins et al., 1998; Honda, 1983). However, if children are increasing airflow in low vowel contexts, this predicts a different pattern—the vowel height dependent effects should be attenuated, or in a more extreme case, they should disappear altogether.

Vowel Acoustics

We compare the three measures, H1-A3, H1-H2, and IF_0, in vowels following /s/ with measures in vowels following /t/ in both adult and child datasets. The purpose of including vowels following /t/ is to control for differences in vowel production between adults and children, both overall (children may in general produce breathier vowels than adults when these vowels follow a consonant produced with glottal abduction) and as a function of F_1 (the effects of F_1 on breathiness and F_0 in vowels may differ between children and adults). We expect that comparing two consonants that require a different level of articulatory demand will guard against the possibility that our findings are due to an inherent difference between children and adults with respect to vowel production. We chose /t/ because it shares place of articulation with /s/, and like /s/, its production involves a wide glottal opening and aspiration noise before the vowel and, therefore, it should show some similarity in the results. However, unlike /s/, the production of /t/ does not require high amplitude turbulent frication noise sustained for a relatively long duration coupled with greater articulatory demand. Therefore, we do not expect children to use increased airflow or glottal gestures in low vowel contexts to the same degree for /t/ as they do for /s/. Following from this, we expect to find a greater difference between children and adults for /s/ than for /t/.

We then assess whether the /s/ tokens judged ‘correct’ or ‘incorrect’ differ with respect to the degree of compensation. Our prediction is that correct tokens will involve a higher degree of compensation for the production of /s/ in articulatorily challenging contexts. We tested this prediction from H1-H2, H1-A3, and IF_0 measures of the vowels that follow /s/ in the child data. However, significant differences between correct and incorrect tokens were found only in the IF_0 difference as a function of F_1. Therefore, we limit our discussion to the findings from the IF_0 model in this regard.

Breathiness.

The magnitude of breathiness encroachment into the following vowel is estimated using two measures: H1-A3 and H1-H2. We tested how speaker factors such as age and gender, vowel quality (F_1), and identity of the preceding consonant (/s/ vs. /t/) modulate the degree of breathiness in the vowel using mixed effects regression models.

We modeled H1-A3 and H1-H2 (each taken at 50% into the vowel) as a function of several variables introduced below in the same way that we did for the spectral models, that is, by introducing maximal by-speaker correlated random effects. Unlike the spectral models where we built separate models for the child and adult data, the breathiness models were built by combining both datasets in order to directly compare the two speaker groups (this was possible as we did not include the effect of Age). In addition to F_1 and Gender, we added two more categorical variables to the models. The first variable is Group, which has the levels 'adults' and 'children', with adults being the reference level. The other variable, Target, has two levels, /t/ and /s/, where /t/ is treated as the reference level. We also included a number of interactions. If children increase airflow for /s/ but less so for /t/, we expect more breathiness in /s/ than /t/ for children but not for adults (Group*Target). In adults we expect breathiness to decrease for vowels with higher F_1 (i.e., lower vowels), but for children we expect breathiness to be maintained or to increase for the same vowels (Group*F_1). We also test whether or not /s/ and /t/ are affected differently by vowel height (F_1*Target). Lastly, we added a three-way interaction, F_1*Group*Target, to test whether a group difference in breathiness as a function of F_1, if it exists, varies by target consonant. Again, all the categorical variables were sum-coded and all the variables were standardized before being fit into the models. By-speaker random intercept and by-speaker random slopes for F_1, Target, and the interaction between F_1 and Target were included as random effects in all the models in Table 3 and 4. These random terms would allow each speaker's H1-A3, H1-H2, and F_0 to differ from the group-level mean and the effects of the fixed effects to vary by speaker. The results of the models are summarized in Table 3 (H1-A3 & H1-H2) and Table 4 (IF_0).

Insert Table 3 about here

Table 3 shows that we found significant main effects of Group and Gender for both measures, indicating breathier vowels for children (H1-A3: $\hat{\beta}$ = 9.002, t = 7.92, p < .0001; H1-

H2: $\hat{\beta}$ = 5.367, t = 9.04, p < .0001) and for females (H1-A3: $\hat{\beta}$ = -3.097, t = -3.41, p = 0009, H1-H2: $\hat{\beta}$ = -1.822, t = -3.85, p =0.0002). The main effect of Target was significant in the H1-A3 model, indicating that /s/ is breathier than /t/ averaging across all speakers and vowels ($\hat{\beta}$ = 1.430, t = 3.32, p =0.001). We also found a significant interaction of Group*Target in the H1-A3 model ($\hat{\beta}$ = 3.993, t = 3.70, p = 0.0003), suggesting that the difference in breathiness between adults and children is greater in /s/ compared to /t/ averaging over all vowels, as illustrated in the left panel in Figure 5. However, unlike the findings in the H1-A3 model, significant differences between the target consonants and a group by target consonant interaction (Group*Target) were not observed in the H1-H2 measure.

Insert Figure 5 about here

A significant main effect of F_1 was found only in the H1-A3 model, which indicates that low vowels are breathier averaging across groups and target consonants ($\hat{\beta}$ = 1.011, t = 2.15, p = 0.033), which is in the opposite direction of previous findings in adults. The F_1*Group interaction for both measures indicates that the effect of F_1 on breathiness measures differs between groups (H1-A3: $\hat{\beta}$ = 4.778, t = 3.90, p = 0.0001, H1-H2: $\hat{\beta}$ = 1.452, t = 1.96, p = 0.058, but note that it does not reach statistical significance at the 0.05 level for H1-H2). Figure 5 (right panel) shows that for the adults, vowels with higher F_1 (i.e., lower vowels) were less breathy than vowels with lower F_1 (i.e., higher vowels), which is consistent with previous literature (Chen et al., 2002; Higgins et al., 1998). Conversely, Figure 5 also shows that for the children, vowels with higher F_1 were breathier than vowels with lower F_1. The main effect of F_1 is thus due to children having breathier vowels in lower vowel contexts. The effects of F_1 and F_1*Group support our prediction of voice quality carry-over likely with increasing airflow for children. However, we did not find evidence that the effect of F_1 differs by target consonant across groups (F_1*Target) or that group difference regarding the effect of F_1 on breathiness differs between /s/ and /t/ (F1*Group*Tartget).

Overall, these measures of breathiness support our hypothesis that children use a different mechanism from adults to maintain frication noise in low vowel contexts during the production of /s/, which may be generalizable to other sounds which, to some degree, share aerodynamic requirements and constriction location, as is the case for /t/. Our results indicate that this compensatory mechanism can be achieved through controlled glottal and subglottal activity involving the manipulation of airflow.

We seek further evidence of such differences between adults and children by looking at the effect of F_1 on F_0 (i.e., IF_0) in the following section.

Vowel intrinsic fundamental frequency.

Adults and children (/s/ and /t/).

We assume that increasing airflow to increase frication noise production will perturb the usual articulatory linkage between the glottal and lingual gestures for the production of the vowel following /s/, imposing s-to-V coarticulation at the glottis. To test our prediction, we built a mixed-effects regression model by taking F_0 (st) measured at the midpoint of the vowel as a response variable. The model was built similarly to the breathiness models. One difference between them is the introduction of an additional two-way interaction term, Gender*Group, into the IF_0 model. This term was introduced to control for the well-documented gender difference for adult speakers in F_0. The results are summarized in Table 4.

Insert Table 4 about here

The main effects of Group and Gender suggest that children had higher F_0 than adults (Group: $\hat{\beta} = 8.774$, $t = 15.98$, $p < .0001$) and that male speakers had lower F_0 than female speakers (Gender: $\hat{\beta} = -1.319$, $t = -3.03$, $p = 0.003$), as expected. We also found that differences between adults and children were greater for male than female speakers (Gender*Group: $\hat{\beta} = 9.172$, $t = 8.43$, $p < .0001$). This is not surprising considering that the developmentally associated F_0 lowering for male speakers usually emerges during puberty. The average F_0 was lower for /s/ than /t/ across children and adults (Target: $\hat{\beta} = -0.357$, $t = -4.27$, $p < .0001$).

Turning to F_1 effects, we found a significant main effect of F_1 indicating that F_0 decreased for vowels with higher F_1 (i.e., lower vowels) compared to vowels with lower F_1 (i.e., higher vowels) across group and target (F_1: $\hat{\beta} = -1.010$, $t = -11.20$, $p < .0001$), consistent with the cross-linguistically observed pattern. A between-group difference was observed; the effect of F_1 was weaker in children than in adults (F_1*Group: $\hat{\beta} = 0.924$, $t = 3.91$, $p = 0.0001$). The interaction of F_1*Target ($\hat{\beta} = -0.425$, $t = -2.56$, $p = 0.011$) showed that the size of the IF_0 difference between high and low vowels, represented as a function of F_1, was greater for /s/ than /t/ across children and adults. We also found a significant three-way interaction effect of F_1*Group*Target ($\hat{\beta} = 0.921$, $t = 2.10$, $p = 0.036$), which suggests that the difference in F_1 effect between the two groups was greater for /s/ than /t/, as both empirical and model prediction plots in Figure 6 illustrate. Note that in the empirical plot, the relationship between F_1 and F_0 is to some degree obscured by other variables such as Age, Gender, etc. In contrast, the model prediction plot more clearly shows the effect of F_1 on F_0 after controlling for all other variables. Note also that the range of F_0 values plotted for the empirical plot is larger, which makes the pattern less clear.

Insert Figure 6 about here

The results of the vowel intrinsic F_0 measure are consistent with the expected universal pattern of higher F_0 for higher vowels. However, children's vowels exhibited attenuation of the universal effects of vowel height on F_0, providing evidence that the compensatory mechanism involving airflow manipulation would have an offsetting effect on the intrinsic aspect of vowel articulation. Further, we found that attenuation of the IF_0 difference in children is greater for /s/ than for /t/, which provides evidence that the degree of compensation required to produce target-like consonants may differ according to the magnitude of the articulatory constraints involved. Built upon these results, we further tested whether the tokens that were counted as correct by the transcribers exhibited a more marked weakening of IF_0 difference than incorrect tokens. We assess this prediction in the following section.

Children (/s/).

Correct tokens are expected to exhibit a greater degree of compensation than incorrect tokens. Therefore, we predict that the effect of F_1 on IF_0 difference between high F_1 (i.e., lower vowels) and low F_1 (i.e., higher vowels) should be greater for the vowels that follow correct /s/

compared to incorrect /s/. We tested this prediction by building a mixed effect model similar to the other models with F_1, Age, Correct, and Gender as main effects and Age*Gender and F_1*Correct as interaction terms. By-speaker random slopes for F_1, Correct, and F_1*Correct as well as by-speaker random intercepts were included as random effects. The results are summarized in Table 5.

Insert Table 5 about here

The universal pattern of IF_0 was again confirmed (F_1: $\hat{\beta} = -1.046$, $t = -6.48$, $p < .0001$). We also found that F_0 decreased for older children (Age: $\hat{\beta} = -1.932$, $t = -4.33$, $p < .0001$). No gender difference was found similar to the IF_0 model that compared adults and children in the previous section. There was no difference in the overall F_0 values between correct tokens and incorrect tokens across age groups (Correct: $\hat{\beta} = -0.041$, $t = -0.20$, $p = 0.836$). The interaction effect of F_1*Age was not significant, suggesting that the effect of vowel height (F_1) on F_0 does not differ among children aged 2-5 years. Crucially, we found a significant interaction effect of F_1*Correct, which indicates that the mechanical consequence of vowel height on F_0 is significantly attenuated for correct tokens ($\hat{\beta} = 1.043$, $t = 2.68$, $p = 0.008$). Our model clearly captures this difference in IF_0 between correct and incorrect tokens, as depicted in the empirical plot (left) and, more apparently, in the model prediction plot (right) in Figure 7.

Insert Figure 7 about here

Taken together, the results of our breathiness and vowel intrinsic F_0 measures found in adults show the expected universal patterns of increasing breathiness and F_0 for higher vowels. Children's and women's vowels are breathier and have a higher F_0 overall. Children's vowels are especially breathy after /s/ compared to /t/ indicating more noise encroachment from the consonant. Crucially, children's vowels also show attenuation of the universal patterns of vowel height on breathiness and F_0, consistent with our prediction that the compensatory mechanism would have a counteracting effect on the universal pattern. For F_0, this attenuation is greater after /s/ than /t/ as well as after /s/ judged as 'correct' than /s/ judged as 'incorrect'.

Discussion and Conclusion

This study strived to determine the degree of phonetic knowledge that children have and how this knowledge is manifested in their speech outputs. To probe this issue, we examined the developmental trajectory of English /s/ using a corpus of speech from children aged 2-5 and adult controls while testing three predictions. The first prediction was that socio-phonetically driven gender-specific knowledge would be manifested in the output spectra of /s/ in the speech of prepubescent children. We found that gender-specific patterns are present in CoG, Skewness, and Spectral Slope in /s/ produced by children, which are the spectral cues in which gender-matched information is encoded in our adult data as well. Our findings replicate Flipsen et al. (1999) and Fox and Nissen (2015), who found gender differences in CoG and Skewness in 9-15 year olds and in CoG, Skewness, Spectral Slope, and other measures in 6-14 year olds, respectively. The gender differences found in younger children in our study may be partly attributable to structural differences between male and female children in head circumference (Nellhaus, 1968), which may affect the size of the front resonance cavity. However, the fact that adult male and female acoustic vowel systems cannot be modeled based on simple tube size differences calls for consideration of non-anatomical factors as well (Fant, 1975). Hence, the observed gender difference in the spectral cues for /s/ in the children in our study likely also reflects sociolinguistically accumulated acoustic knowledge. Our view is thus in line with some aspects of exemplar or episodic models of representation (Goldinger, 1996; Pierrehumbert, 2001, 2006). According to Pierrehumbert (2006), episodic memories of contrastive speech sounds consist of fine phonetic detail of both linguistic and social categories, both of which are acquired via perception and social interactions (see also Sumner, 2015).

Our second prediction was that children and adults would differ in /s/ and V coproduction across different vowel contexts. We found that the acoustic spectra of children's /s/ are more canonical in low vowel contexts (higher CoG, lower Skewness, and higher Spectral Slope) while those of adults are relatively invariant to vowel context. The directions of those spectral cues as a function of vowel height (F_1) were precisely the same as the tokens that were judged as correct compared to incorrect tokens. This difference between children and adults in the effect of vowel height was linked to our third prediction that children use a different strategy than adults to compensate for vowel context. Studies suggest that in the case of adults, canonical /s/ production, which is primarily represented through high CoG (Newman, 2003), is articulatorily achievable by shortening the length of the front cavity by, for example, forming the constriction more anteriorly (Jesus & Shadle, 2001; Shadle & Mair, 1996). Another study observed that adults can maintain relatively uniform constriction size and location during the production of /s/ in various contexts by using tongue tip gestures to compensate for differences in required jaw movements (Iskarous et al., 2011). Together, these studies suggest that adults have more than one articulatory option available for producing /s/ relatively invariantly in different vowel contexts. However, as we discussed, the use of the same articulatory strategies is not expected among children due to their limited lingual dexterity and larger tongue size relative to cavity size. An alternative way for children to increase CoG that we examined was increasing the rate of airflow through the constriction accompanied by higher subglottal pressure and forceful noise production. Specifically, given the articulatory immaturity of children and the effect of airflow on frication enhancement, we predicted that children would exploit an aerodynamic mechanism in order to compensate for the lowered tongue body gesture during the production of /s/ in low vowel contexts. The results from H1-A3, H1-H2, and IF_0 supported this prediction.

We observed greater noise encroachment for lower vowels (higher F_1) following both /s/ and /t/ in children while it was higher vowels (lower F_1) that were breathier in adults. The overall group difference in breathiness was greater for /s/ than for /t/ averaging over all vowels. Crucially, we found that the universality of IF_0 difference between high and low vowels was more attenuated after /s/ than /t/ and after auditorily correct /s/ than incorrect /s/ for children. These results, together with effects of vowel height on spectral cues, suggest that, unlike adults, children's /s/ is affected by vowel height in such a way that overshoot occurs due to the compensatory system they adopt in articulatorily challenging vowel contexts. The results of the breathiness analysis suggest that the level of breathiness as a function of F_1 differs between adults and children and that this group difference can be more or less generalized to other sounds that share the same primary articulators and glottal status in speech production. At the same time, our results from IF_0 suggest that the extent to which this compensatory strategy affects a sound may differ depending on the temporal and spatial differentiation in the coordination of multiple supraglottal and subglottal speech systems that are required for each sound. Our interpretation—that the attenuated IF_0 in low vowels in children compared to adults is due to the compensatory airflow mechanism—is supported by earlier studies that observed an increase in F_0 as a result of an increase in the rate of transglottal airflow when other muscular laryngeal activity was controlled; this has been found both in humans and in models based on an excised canine larynx (Alipour & Scherer, 2007; Baer, 1979; Lieberman, Knudson, & Mead, 1969; Titze, 1989).

However, the interpretation of our results on IF_0 may be challenged by a recent study that observed that the IF_0 difference between vowels is highly affected by lexical and phrasal accent (Jacewicz & Fox, 2015). The study found that when the target word occurs in a prosodically weak position and is unaccented, the IF_0 difference is attenuated. This finding raises the possibility that the smaller size of the IF_0 difference found in /s/ and especially in correctly produced tokens in our child data may be due to differences in the location of prominence between the productions of adults and children. However, there are two arguments against this possibility. First, the attenuation of the IF_0 difference is greater for /s/ than /t/ within children, which provides evidence that the IF_0 difference is not merely due to an inherent difference in vowel production between children and adults. Second, all the /s/- and /t/-initial tokens examined in this study come from monosyllabic words or bisyllabic words with initial stress, a word shape that is both early acquired and rarely produced with stress errors by English-speaking children (Allen & Hawkins, 1978; Gerken, 1994; Kehoe, 1998), and all tokens were produced as single word utterances. Each word thus most likely forms a single prosodic domain which carries a nuclear pitch accent on the primary stressed syllable containing /s/ or /t/ (see, e.g., Beckman & Pierrehumbert, 1986). Therefore, the effect of the presence or absence of prominence on the productions of the words in our data should be minimal compared to words of different stress profiles that are embedded in sentences with different prosodic contours. These two points support our view that the attenuation we observed in the child data is most likely due to a compensatory strategy involving an aerodynamic mechanism.

The absence of F_1 effects on CoG that we observed in the adult data, however, is in contrast with the acoustic findings reported in Iskarous et al. (2011). They observed higher CoG (M1) in low vowel contexts compared to high vowel contexts, while we found CoG to be invariable. Their explanation for the variability in CoG caused by vowel height was the raised lip and jaw positions observed relatively constantly in high vowels among their speakers. However, raised vertical motion of the jaw and lower lip cannot be solely responsible for the lower CoG without lip rounding or protrusion, which is the main trigger of a second turbulence source mechanism and also one of the causes of CoG lowering (Shadle & Scully, 1995). An alternative explanation for the discrepancy in findings may be that the pellets glued on the surface of the tongue in Iskarous et al. (2011) exerted different effects on different vowels. The effects of pellet markers on speech acoustics have been examined in several studies (Baum & McFarland, 1997; Fant, 1960; Weismer & Bunton, 1999) though, to our knowledge, no study has directly examined the effects of the markers on vowels. These studies suggest that the pellet markers increased spectral mean possibly by perturbing laminar flow (Fant, 1960; Weismer & Bunton, 1999). More related to the current study, Baum and McFarland's (1997) speech adaptation experiment involving an artificial palate suggested that speakers modify airflow turbulence to compensate for feedback they receive. This implies that airflow manipulation could be a compensatory strategy that adults can also adopt when their articulators are constrained or perturbed.

Our findings from the child data—the gender divergence present in the acoustic spectra of /s/, non-adult-like vowel context effects, and evidence of children's compensatory mechanism relying on phonation—provide evidence that, at early stages of speech development, production goals are acoustics-oriented. Our view is in line with studies that have observed that children with a phonological deficit in production also show difficulty in speech perception (Broen, Strange, Doyle, & Heller, 1983; Rvachew & Jamieson, 1989) and supports the proposal that improvements in production can be achieved through perception training by presenting participants with both more and less canonical varieties of a segment produced by multiple speakers (Guenther, Husain, Cohen, & Shinn-Cunningham, 1999; Rvachew, Nowak, & Cloutier, 2004). Our view is, however, in stark contrast with Direct Realism (Fowler, 1986, 1996), which claims that the objective of speech perception is the speaker's actual articulatory gestures as well as with Motor Theory (Liberman & Mattingly, 1985), which contends that the object of speech perception is the speaker's intended gestures. In contrast with these two views, we suggest that the goal of both speech perception and production at early stages of language development is to achieve socially-oriented and acoustically-specified adult targets. This may shift toward more articulatory goals at later stages in development. Further studies involving a wider range of age groups are needed to better understand whether and when a shift in speech goals takes place.

One may question whether the elicitation method that used pre-recorded labels as prompts produced by an adult female speaker in child-directed speech may have affected the child participants in a way that they modified their productions to mimic the acoustic patterns of the prompts. We suggest that the finding that the male children follow the speech patterns of male adults rather than the female prompter weakens this possibility. Nevertheless, a comparison between the sound records of the prompts and the child data would be necessary to definitively rule out the possibility of imitation.

As a final remark, we point out that the current study indirectly inferred kinematic and aerodynamic behavior in speech development based on acoustic information. A simultaneous recording of the acoustic signal, airflow, and articulation would be necessary to strengthen our understanding of the compensatory strategy that involves the regulation of speech aerodynamics in the face of articulatory challenges. We leave this to future research.

Acknowledgements

This research was supported by grants awarded to Meghan Clayards by the Fonds de recherche du Québec–Société et culture (FRQSC) (2011-NC-145433), to Lydia White, Heather

Goad, and colleagues by the Fonds de recherche du Québec–Société et culture (FRQSC) (2010-SE-130727), and to Heather Goad and Lydia White by the Social Sciences and Humanities Research Council of Canada (SSHRC) (435-2015-0490). We would like to thank Mary Beckman for making the adult files from the Paidologos project available and for providing information on how these data were collected.

An earlier version of this work, which reported on a subset of the data, was presented at the 39th Annual Boston University Conference on Language Development and appears in the proceedings (Bang, Clayards, & Goad, 2015).

References

Allen, G. D., & Hawkins, S. (1978). The development of phonological rhythm. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments* (pp. 173–185). Amsterdam: North-Holland.

Alipour, F., & Scherer, R. C. (2007). On pressure-frequency relations in the excised larynx. *Journal of the Acoustical Society of America*, *122*(4), 2296–2305.

Avery, J. D., & Liss, J. M. (1996). Acoustic characteristics of less-masculine-sounding male speech. *Journal of the Acoustical Society of America*, *99*(6), 3738–3748.

Bachorowski, J.-A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America*, *106*(2), 1054–1063.

Baer, T. (1979). Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. *Journal of the Acoustical Society of America*, *65*(5), 1271–1275.

Bang, H.-Y., Clayards, M., & Goad, H. (2015). A child specific compensatory mechanism in the acquisition of English /s/. In E. Grillo & J. Kyle (Eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development* (pp. 75–87). Somerville: Cascadilla Press.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2014*). lme4: Linear mixed-effects models using Eigen and S4*. R package, version 3.1.2, https://cran.r project.org/web/packages/lme4/index.html (Last accessed March 1, 2015).

Baum, S. R., & McNutt, J. C. (1990). An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics*, *18*, 51–63.

Baum, S. R., & McFarland, D. H. (1997). The development of speech adaptation to an artificial palate. *The Journal of the Acoustical Society of America*, *102*(4), 2353–2359.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*(1), 255–309.

Bernhardt, B. H., & Stemberger, J. P. (1998). *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.

Bickley, C. A., & Stevens, K. N. (1986). Effects of a vocal-tract constriction on the glottal source: Experimental and modelling studies. *Journal of Phonetics*, *14*, 373–382.

Boersma, P., Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.57). [Computer software]. Retrieved from http://www.praat.org/.

Broen, P., Strange, W., Doyle, S., & Heller, J. H. (1983). Perception and production of approximant consonants by normal and articulation delayed preschool children. *Journal of Speech and Hearing Research, 26*, 601–608.

Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *Journal of the Acoustical Society of America*, *97*(4), 2603–2606.

Bush, M. (1981). *Vowel articulation and laryngeal control in the speech of the deaf.* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Catford, J. C. (1982). *Fundamental problems in phonetics*. Bloomington: Indiana University Press.

Chen, Y., Robb, M. P., & Gilbert, H. R. (2002). Electroglottographic evaluation of gender and vowel effects during modal and vocal fry phonation. *Journal of Speech and Hearing Research*, *45*(5), 821–829.

Cheng, H. Y., Murdoch, B. E., Goozée, J. V., & Dion, S. (2007). Physiologic development of tongue-jaw coordination from childhood to adulthood. *Journal of Speech and Hearing Research*, *50*, 352–360.

Clayards, M., & Knowles, T. (2015). Prominence enhances voiceless-ness and not place distinction in English voiceless sibilants. *Proceedings of the International Congress of Phonetic Sciences*, Glasgow, UK. 4 pages.

Connell, B. (2002). Tone languages and the universality of intrinsic F0: Evidence from Africa. *Journal of Phonetics*, *30,* 101–129.

DiCanio, C. T. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the*

International Phonetic Association, *39*(2), 162–188.

DiCanio, C. T. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics*, *40*, 162–176.

Diehl, R. L., Lindblom, B., Hoemeke, K. A., & Fahey, R. P. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, *24*, 187–208.

Eckert, P. (2008). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, *1*(3), 245–267.

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics*, *22*, 937–956.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague, Netherlands: Mouton.

Fant, G. (1975). Non-uniform vowel normalization. *Speech Technology Laboratory: Quarterly Progress and Status Report, 2*(3), 1–19.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America*, *106*(3), Pt. 1, 1511–1522.

Flipsen, P., Shriberg, L., Weismer, G., Karlsson, H., & McSweeny, J. (1999). Acoustic characteristics of /s/ in adolescents. *Journal of Speech and Hearing Research*, *42*(3), 663–677.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, *84*(1), 115–123.

Fowler, C. A. (1986). An event approach to the speech of perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, *99*(3), 1730–1741.

Fox, R. A., & Nissen, S. L. (2005). Sex-related acoustic changes in voiceless English fricatives.

Journal of Speech and Hearing Research, *48*(4), 753–765.

Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? In S. Fuchs, M. Toda, & M. Żygis (Eds.), *An interdisciplinary guide to turbulent sounds* (pp. 281–302). Berlin, Germany: Mouton de Gruyter.

Gerken, L. A. (1994). A metrical template account of children's weak syllable omissions from mutisyllabic words. *Journal of Child Language, 21*(3), 565–584.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1166–1183.

Green, J. R., Moore, C. A., & Reilly, K. J. (2002). The sequential development of jaw and lip control for speech. *Journal of Speech and Hearing Research*, *45*(1), 66–79.

Gruber, F. A. (1999). Variability and sequential order of consonant normalization in children with speech delay. *Journal of Speech and Hearing Research*, *42*(2), 460–472.

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, *106*(5), 2900–2912.

Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America*, *106*(2), 1064–1077.

Hardcastle, W. J. (1976). *Physiology of speech production: An introduction for speech scientists*. New York, NY: Academic Press.

Henton, C. G. (1989). Fact and fiction in the description of female and male pitch. *Language and Cognitive Processes*, *9*(4), 299–311.

Higgins, M. B., Netsell, R., & Schulte, L. (1998). Vowel-related differences in laryngeal articulatory and phonatory function. *Journal of Speech and Hearing Research*, *41*(4), 712–724.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of

American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research*, *38*(6), 1212–1223.

Honda, K. (1983). Relation between pitch contral and vowel articulation. *Haskins Laboratories Status Report on Speech Research*, *SR 73*, 269–282.

Honda, K. (2008). Physiological Processes of Speech Production. In J. Benesty, M. M. Sondhi, & Y. Huang (Eds.), *Springer handbook of speech processing* (pp. 7–26). Berlin and Heidelberg, Germany: Springer.

Howe, M., & McGowan, R. (2005). Aeroacoustics of [s]. *Proceedings of the Royal Society of London. Series A, 461*, 1005–1028.

Ingram, D., Christensen, L., Veach, S., & Webster, B. (1980). The acquisition of word-initial fricatives and affricates in English by children between 2 and 6 years. In Yeni-Komshian, G. H., Kavanagh, J. F., & Ferguson, C. A. (Eds.), *Child phonology: Vol. 1. Production* (pp. 169–192). New York, NY: Academic Press.

Iseli, M., & Alwan, A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04): Vol. 1,* I669–I672.

Iseli, M., Shue, Y. L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America*, *121*(4), 2283–2295.

Iskarous, K., Shadle, C. H., & Proctor, M. I. (2011). Articulatory–acoustic kinematics: The production of American English /s/. *Journal of the Acoustical Society of America*, *129*(2), 944–954.

Jacewicz, E. and Fox, R. (2015). Intrinsic fundamental frequency of vowels is moderated by regional dialect. *Journal of the Acoustical Society of America*, *138*(4), EL405–EL410.

Jesus, L. M. T., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, *30*, 437–464.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, *108*(3), 1252–1263.

Katz, W. F., & Bharadwaj, S. (2001). Coarticulation in fricative-vowel syllables produced by children and adults: A preliminary report. *Clinical Linguistics & Phonetics*, *15*, 139–143.

Keating, P. A., & Esposito, C. (2006). Linguistic voice quality. *UCLA Working Papers in Phonetics*, *105*, 85–91.

Kehoe, M. (1998). Support for metrical stress theory in stress acquisition. *Clinical Linguistics & Phonetics, 12(1),* 1–23.

Kendall, T., and Thomas, E. R. (2014). *Vowels: Vowel manipulation, normalization, and plotting*, R package, version 1.2-1, https://CRAN.R-project.org/package=vowels (Last accessed May 1, 2015).

Kent, R. D. (1992). The biology of phonological maturation. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 65–90). Timonium, MD: York Press.

Kent, R. D. (2004). The uniqueness of speech among motor systems. *Clinical Linguistics & Phonetics*, *18*(6-8), 495–505.

Kirk, P. L., Ladefoged, P. & Ladefoged, J. (1984). Using a spectrograph for measures of phonation types in natural languages, *UCLA Working Papers in Phonetics, 59*, 102–113.

Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward improved spectral measures of /s/: Results from adolescents. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1175–1189.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests for random* *and fixed effects for linear mixed effect models (lmer objects of lme4 package):* R package version 2.0-29, https://CRAN.R-project.org/package=lmerTest. (Last accessed March 1, 2016).

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change, 2*, 205–254.

Ladefoged, P., & Broadbent, D. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, *29*(1), 98–104.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK: Blackwell.

Lehiste, I., & Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America*, *31*, 428-435.

Lehiste, I. (1976). Suprasegmental features of speech. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 225-239). New York, NY: Academic Press.

Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese* (Unpublished doctoral dissertation). The Ohio State University, Columbus.

Li, F., Edwards, J., & Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, *37*, 111–124.

Li, F. (2011). Speech production development of voiceless sibilant fricatives in children speaking Mandarin Chinese. *Journal of the Acoustical Society of America,* 130(4), 2523.

Lieberman, P., Knudson, R., & Mead, J. (1969). Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation. *Journal of the Acoustical Society of America*, *45*(6), 1537–1543.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: Study of voice onset time in word-initial stop consonants. *Journal of Child Language*, *7*(1), 41–74.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America*, *125*(6), 3962–3973.

Maxwell, E. M., & Weismer, G. (1982). The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops. *Applied Psycholinguistics*, *3*(1), 29–43.

Ménard, L. L., Schwartz, J.-L. J., & Boë, L.-J. L. (2004). Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood. *Journal of Speech and Hearing Research*, *47*(5), 1059–1080.

Ménard, L., & Noiray, A. (2011). The development of lingual gestures in speech: Experimental approach to language development. *Faits de langues*, *37*, 189–202.

Munson, B., Edwards, J., Schellinger, S. K., Beckman, M. E., & Meyer, M. K. (2010). Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana. *Clinical Linguistics & Phonetics*, *24*(4-5), 245–260.

Munson, B., Crocker, L., Pierrehumbert, J. B., Owen-Anderson, A., & Zucker, K. J. (2015). Gender typicality in children's speech: A comparison of boys with and without gender identity disorder. *Journal of the Acoustical Society of America*, *137*(4), 1995.

Nellhaus, G. (1968). Head circumference from birth to eighteen years practical composite international and interracial graphs. *Pediatrics*, *41*(1), 106–114.

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *Journal of the Acoustical Society of America*, *113*(5), 2850–2860.

Niebuhr, O., Clayards, M., Meunier, C., & Lancia, L. (2011). *Journal of Phonetics*, *39*, 429–451.

Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *Journal of the Acoustical Society of*

America, *118*(4), 2570–2578.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, *32*(1), 120–132.

Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. In M.-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 771–774). Barcelona, Spain.

Pennington, M. (2005). *The phonetics and phonology of glottal manner features* (Unpublished doctoral dissertation). Indiana University, Ann Arbor.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America*, *109*(6), 2988–2998.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, *34*, 516–530.

Reidy, P. F. (2015). *The spectral dynamics of voiceless sibilant fricatives in English and Japanese* (Unpublished doctoral dissertation). The Ohio State University, Columbus.

Romeo, R., Hazan, V., & Pettinato, M. (2013). Developmental and gender-related trends of intra-talker variability in consonant production. *Journal of the Acoustical Society of America*, *134*(5), 3781–3792.

Rvachew, S., & Jamieson, D. G. (1989). Perception of voiceless fricatives by children with a functional articulation disorder. *Journal of Speech and Hearing Disorders,* 54, 193–208.

Rvachew, S., Nowak, M., & Cloutier, G. (2004). Effect of phonemic perception training on speech production and phonological awareness skills of children with expressive phonological delay. *American Journal of Speech-Language Pathology, 13*, 250–263.

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, J. (2000). Covert contrast as a stage in the acquisition of phonetics and phonology. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Labaratory Phonology V: Acquisition and the lexicon* (pp. 194–207). Cambridge, UK: Cambridge University Press.

Scully, C., Castelli, Ménard E., Brearley, E. & Shirt, M. (1992). Analysis and simulation of a speaker's aero-dynamic and acoustic pattern for fricatives, *Journal of Phonetics, 20*, 39–51.

Shadle, C. H., & Scully, C. (1995). An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics*, *23*, 53–66.

Shadle, C.H. (2012). The Acoustics and Aerodynamics of Fricatives. In A. Cohn & M. K. Fougeron (Eds.), *The Oxford handbook of laboratory phonology* (pp. 511–526). Oxford, UK: Oxford University Press.

Shadle, C. H., & Mair, S. J. (1996). Quantifying spectral characteristics of fricatives. In H. T. Bunnell & W. Idsardi (Eds.), *Proceedings ICSLP 96: Fourth International Conference on Spoken Language Processing* (pp. 1521–1524). Wilmington, DE: Applied Science and Engineering Laboratories.

Shriberg, L. D., Kent, R. D., Karlsson, H. B., Mcsweeny, J. L., Nadler, C. J., & Brown, R. L. (2003). A diagnostic marker for speech delay associated with otitis media with effusion: Backing of obstruents. *Clinical Linguistics & Phonetics*, *17*(7), 529–547.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*(4), 779–798.

Smith, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge, UK: Cambridge University Press.

Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of Americ*a, *27*, 484– 493.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Sumner, M. (2015). The social weight of spoken words. *Trends in Cognitive Sciences*, *19*(5), 238–239.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, *85*(4), 1699–1707.

Titze, I. R. (1994). Toward standards in acoustic analysis of voice. *Journal of Voice*, *8*(1), 1–7.

Van Hoof, S., & Verhoeven, J. (2011). Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *Journal of Phonetics*, *39*, 168–177.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of the Acoustical Society of America*, *125*(3), 1666–1678.

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech and Hearing Research*, *54*(4), 995–1010.

Weismer, G., & Bunton, K. (1999). Influences of pellet markers on speech production behavior: Acoustical and perceptual measures. *Journal of the Acoustical Society of America*, *105*(5), 2882–2894.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*, 349–366.

Whalen, D. H., Levitt, A. G., Hsiao, P.-L., & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9-, and 12-month-old French- and English-learning infants. *Journal of the Acoustical Society of America*, *97*(4), 2533–2539.