Spectral dynamics of sibilant fricatives are contrastive and language specific

Patrick F. Reidy[a)]
Department of Linguistics, The Ohio State University, 108A Ohio Stadium East, 1961 Tuttle Park Place, Columbus, Ohio 43210, USA

(Received 2 June 2016; accepted 14 September 2016; published online 13 October 2016)

Previous research has extensively investigated the spectral properties of sibilant fricatives with little consideration to how these properties vary over time. To investigate such spectro-temporal variation, productions of English /s/ and /ʃ/ and of Japanese /s/ and /ɕ/ in word-initial, prevocalic position were elicited from adult native speakers. The spectral dynamics of these productions were analyzed in terms of a psychoacoustic measure of peak frequency: "peak ERB_N number." Peak ERB_N number was computed at 17 evenly spaced points across each fricative production. The resulting peak ERB_N number trajectories were analyzed with orthogonal polynomial growth-curve models, to determine how peak frequency varied temporally within each fricative. Three analyses compared (1) the English sibilants to each other, (2) the Japanese sibilants to each other, and (3) English /s/ to Japanese /s/. The results indicated that, in both English and Japanese, the sibilant fricatives differ acoustically in terms of both static (i.e., overall level) and dynamic (i.e., shape) aspects of the peak ERB_N number trajectories. Furthermore, English /s/ and Japanese /s/ exhibited language-specific differences in the shape, but not overall level, of peak ERB_N number trajectories.
 [http://dx.doi.org/10.1121/1.4964510]

[JFL] Pages: 2518–2529

I. INTRODUCTION

A. Articulatory kinematics, but spectral statics

During the articulation of a sibilant fricative, such as English /s/ or /ʃ/, the tongue is raised toward the palate, shaping a narrow constriction within the oral cavity. As air travels through this narrow linguapalatal constriction, its flow becomes turbulent, generating a noise source at the anterior end of the constriction (Shadle, 1991; Stevens, 1971). Additional noise sources are generated when the turbulent airflow impinges on the incisors downstream from the constriction Narayanan and Alwan (2000). The generation of noise sources, therefore, involves the tongue and jaw since these articulators, respectively, form the constriction and position the lower incisors. Furthermore, articulatory studies of sibilant fricatives have found that the tongue and jaw both move continuously during their production (e.g., Iskarous *et al.*, 2011; Mooshammer *et al.*, 2006; Zharkova *et al.*, 2014).

The continuous movement of the tongue and jaw suggests that the generated turbulence noise may not be stationary; however, many acoustic studies of sibilant fricatives have characterized them in terms of static spectral properties, computed from either a single window of the fricative noise, or by averaging multiple windows placed across the fricative (e.g., Brunner *et al.*, 2011; Ghosh *et al.*, 2010; Holliday *et al.*, 2015; Li, 2012; Li *et al.*, 2009; Newman *et al.*, 2001; Perkell *et al.*, 2004; Romeo *et al.*, 2013). Relatively little effort has been given to characterizing the temporal variation in spectral properties of sibilant fricatives across their duration.

In an early study, Behrens and Blumstein (1988) analyzed temporal changes in spectral properties across English /s/ and /ʃ/. Peak frequency was measured at the beginning, middle, and final 15 ms of each fricative, revealing that peak frequency was higher for /s/ (3.8–8.5 kHz) than for /ʃ/ (2.3–7 kHz). The authors further noted that the spectral "patterns appeared to be maintained across the three time windows," which led them to conclude that any characterization of /s/ and /ʃ/ "based on spectral properties can probably be derived from…a static configuration of the frication noise itself…irrespective of where the frication noise is measured" (pp. 297–298). One interpretation of this conclusion is that the relationship between the peak frequencies of /s/ and /ʃ/ is the same regardless of where in the time course of the sibilant the spectrum is estimated. Under this interpretation, a single measure of peak frequency is sufficient to characterize the difference between /s/ and /ʃ/ because the relationship between the two remains constant—i.e., /s/ has a higher peak frequency than /ʃ/. A second, stronger interpretation is that the peak frequency of each sibilant is reasonably constant across time. Under this second interpretation, a single measure of peak frequency would be sufficient for characterizing both /s/ and /ʃ/, in their own right, since this spectral property would not be sensitive to where in the fricative it is measured.

While Behrens and Blumstein did not report any statistical tests that would have indicated the extent to which peak frequency varies across the time course of either sibilant fricative, they did report that "high frequency peaks tended to appear more often at the midpoint" of frication (p. 297), which suggests some amount of variation with time,

[a)]Current address: Callier Center for Communication Disorders, The University of Texas at Dallas, 1966 Inwood Road, Dallas, TX 75235, USA. Electronic mail: patrick.francis.reidy@gmail.com

supporting the weaker interpretation. However, it is the stronger interpretation that seems to have persisted. For example, Behrens and Blumstein (1988) is cited as the basis of the following claims: spectral properties of sibilants "are relatively stable throughout the noise portion" (Jongman *et al.*, 2000, p. 1255); "previous research has not found that the spectral [peak] varies greatly throughout the course of the fricative" (Munson, 2001, p. 1203); spectral "peak measures [remain] relatively constant across time" (Newman *et al.*, 2001, p. 1184).

A recent study of English /s/, however, provides strong evidence that only the weak interpretation of Behrens and Blumstein (1988) should be followed. Iskarous *et al.* (2011) found that, in adults' productions of /s/, centroid frequency followed an increasing, concave trajectory across the course of the fricative, rising until reaching a global maximum around 80% of the fricative's duration, before falling off. Moreover, this temporal variation in centroid frequency corresponded to articulatory movements such as the raising of the jaw across the first half of frication, and the release of the linguapalatal constriction near the end of frication. These results suggest that static spectral features are insufficient to characterize /s/.

B. Research questions and hypotheses

Because Iskarous *et al.* (2011) studied only English /s/, it is unknown whether the temporal variation in spectral properties that they observed is specific to this sibilant or whether it is a general property of sibilant fricatives, one that holds across sibilants within a given language, or cross-linguistically across sibilants that are articulated with comparable sustained postures. This paper extends the results of Iskarous *et al.* (2011) in three ways.

First, instead of investigating temporal variation in centroid frequency, the analyses here investigate the dynamics of a psychoacoustic measure of peak frequency: peak ERB_N number. This measure denotes the ERB_N number—a psychoacoustic frequency scale—of the auditory filter that is most activated by an incoming acoustic signal, and is fully described in Sec. III E.

Second, the peak ERB_N number trajectories of contrastive sibilant fricatives from two languages will be compared. In particular, English /s/ and /ʃ/ will be compared, and Japanese /s/ and /ɕ/ will be compared. The purpose of these within-language analyses is to ask whether contrastive sibilants in a given language differ not just in terms of static spectral properties measured at a given point in the fricative (e.g., at midpoint), but also in terms of the pattern of temporal variation in one such property. Specifically of interest here is whether the peak ERB_N number trajectories of English /s/ and /ʃ/, or of Japanese /s/ and /ɕ/, differ in terms of their shape, such that one is not simply the translation of the other along the peak ERB_N number scale. Since, within either language, the two sibilant fricatives are articulated with different postures, it is hypothesized that differences in how the articulators must move to form and release these postures will lead to differences in the shapes of the two sibilants' peak ERB_N number trajectories in both English and Japanese.

Third, since any observed differences across sibilants within a language may be due to kinematic requirements on the articulators as they move to form, maintain, and release the linguapalatal constriction, English /s/ and Japanese /s/—two sibilants that have comparable articulatory postures and comparable spectral properties at fricative midpoint—will be compared cross-linguistically. This comparison asks whether two sibilants that are comparable in terms of their "steady state" properties are also comparable in terms of their dynamic spectral properties. No hypothesis is made regarding this research question.

The remainder of the paper is organized as follows. Section II reviews previous findings for the articulatory postures and static spectral properties of English /s/ and /ʃ/ and of Japanese /s/ and /ɕ/. Section III describes a picture-prompted word-repetition task used to elicit productions of sibilant fricatives in word-initial position, and the method for computing trajectories of peak ERB_N number from these productions. Section IV reports the results from multiple growth-curve analyses, designed to test differences in temporal variation in peak ERB_N number across pairs of sibilant fricatives. Section V discusses the implications and limitations of the findings.

II. BACKGROUND

A. Articulatory postures and static spectral properties of English /s/ and /ʃ/

In articulatory terms, English /s/ and /ʃ/ differ foremost in place of articulation and front cavity size. For /s/, the tongue tip or blade is raised to form the constriction at the upper incisors or on the alveolar ridge, whereas, for /ʃ/, the constriction involves the tongue blade and is posterior to the alveolar ridge (see Fletcher and Newman, 1991; McLeod *et al.*, 2006; Narayanan *et al.*, 1995; Stone *et al.*, 1992). Since the constriction for /ʃ/ lies posterior to that of /s/, the front cavity is larger for /ʃ/ (Toda and Honda, 2003). Furthermore, this quantitative difference in volume is accompanied by a qualitative difference in front cavity shape: for /ʃ/, the tongue is postured such that a sublingual cavity forms posterior to the lower incisors, whereas /s/ is more often articulated with the underside of the tongue tip contacting the lower incisors, eliminating this sublingual cavity (Perkell *et al.*, 2004).

Because of their being articulated with different front cavity sizes, English /s/ and /ʃ/ differ spectrally in terms of the distribution of their resonant frequencies, which are higher for /s/. A number of studies have indexed this difference in resonant frequencies with centroid frequency (i.e., first spectral moment) or peak frequency (i.e., the frequency of the most prominent peak in the spectrum). When estimated from either the beginning, middle, or end of the fricative noise, centroid frequency has been found to be greater in /s/ than in /ʃ/ (Jongman *et al.*, 2000), and similar results have been found when centroid frequency measurements from multiple locations across the fricative are pooled together (Fox and Nissen, 2005; Maniwa *et al.*, 2009). In

each of these studies, the difference in centroid frequency was evaluated across a group of talkers, and when the perspective shifts to the individual talker, the same pattern obtains (Haley *et al.*, 2010; Li *et al.*, 2009). As is the case with centroid frequency, peak frequency has been found to be higher in /s/ than in /ʃ/, regardless of whether it is measured at the beginning, middle, or end of frication (Behrens and Blumstein, 1988; Heinz and Stevens, 1961; Hughes and Halle, 1956). Finally, when used in classification tasks, centroid frequency and peak frequency have been found to yield high, sometimes even perfect, accuracy (Forrest *et al.*, 1988; Li *et al.*, 2009; McMurray and Jongman, 2011).

B. Articulatory postures and static spectral properties of Japanese /s/ and /ɕ/

The difference in articulatory posture between Japanese /s/ and /ɕ/ is traditionally described as one of the constriction's degree of palatalization, the constriction being more palatalized for /ɕ/ than for /s/ (Akamatsu, 1997). Analyzing magnetic resonance images (MRI) of sustained articulatory postures, Toda and Honda (2003) quantified a posture's "palatalization index" as the ratio of the area between the tongue and hard palate to the length from the anterior end of the constriction to the median between the anterior and posterior nasal splines. They found that each of their nine participants differentiated the two sibilants in terms of palatalization index, and that there was almost perfect talker-independent separation in terms of this articulatory feature. Toda and Honda (2003) also found, for each of their participants, that the area of the front cavity in the mid-sagittal plane was greater for /ɕ/ than for /s/; however, across talkers, two-thirds of the front cavity areas for /ɕ/ fell within the observed range for /s/. Thus, the consistent talker-internal difference in front cavity size is qualified by a significant amount of inter-talker overlap along this articulatory dimension.

The differences in front cavity size and palatalization engender spectral differences in both the fricative noise and the fricative-vowel transition of /s/ and /ɕ/. In particular, because /s/ is articulated with a relatively smaller front cavity, both centroid frequency (i.e., first moment) and peak frequency of the frication noise are higher for /s/ than for /ɕ/ (Li *et al.*, 2009; Toda, 2007). Additionally, because /ɕ/ is articulated with greater palatalization it is also produced with a smaller back cavity, which is revealed spectrally as a relatively higher *F*2 frequency at vowel onset (Li *et al.*, 2009; Toda, 2007).

Depending on which acoustic features are used, it is possible to automatically classify /s/ and /ɕ/ using either features from the fricative noise and the fricative-vowel transition, or those from the fricative noise alone. Li *et al.* (2009) built a classifier using, as predictor variables, the first four spectral moments of the middle 40 ms of the fricative noise and the *F*2 frequency at the fricative-vowel boundary. They found that centroid frequency and *F*2 frequency were necessary to perfectly classify adults' productions of /s/ and /ɕ/. In an earlier study, though, Fujisaki and Kunisaki (1978) used some number of poles and zeroes to model the fricative noise in natural productions of /s/ and /ɕ/ by an adult male talker. They found that when two poles and one zero was used to characterize each production, an automatic classifier was 100% accurate; however, when using only one pole (and no zeroes), the classifier was 92% accurate. Fujisaki and Kunisaki then validated their pole-and-zero models with analysis-by-synthesis. First, they found that when a continuum of tokens were synthesized from either the model with two poles and one zero, or the model with one pole and no zeroes, the gradient of the equiprobability contour closely matched the decision surface of the classifier. Second, they found that listeners rated tokens that were synthesized with either model to be almost as natural sounding as the natural productions. In sum, the automatic classification experiments of Li *et al.* (2009) and Fujisaki and Kunisaki (1978) suggest that, while transitional information does facilitate identification, adults' productions of /s/ and /ɕ/ can still be identified with reasonably high accuracy from a single measure of the fricative noise's peak frequency.

C. Cross-linguistic similarities between English /s/ and Japanese /s/

While the language-internal contrasts between English /s/ and /ʃ/ and between Japanese /s/ and /ɕ/ are instantiated with different articulatory parameters, the MRI data reported by Toda and Honda (2003) suggest that the articulatory posture for /s/ is comparable in these two languages. First, within either language, talkers varied as to whether they articulated /s/ at the teeth or the alveolar ridge, and the amount of variation in place of articulation was similar between the two languages. Additionally, the English and Japanese talkers exhibited similar ranges of front cavity size ($\approx$5–50 mm^2) and of palatalization index ($\approx$5–10 mm). This cross-linguistic similarity in the articulatory posture of sustained /s/ is reflected in similar values of centroid frequency. Previous studies of English /s/ and Japanese /s/ have reported values ranging from 6 to 12 kHz, with means between 7 and 8 kHz when averaged across talkers (e.g., for English: Jongman *et al.*, 2000; Li *et al.*, 2009; Maniwa *et al.*, 2009; for Japanese: Li *et al.*, 2009; Toda, 2007).

III. METHOD

A. Participants

Adults' productions of sibilant fricatives were drawn from the English and Japanese portions of the Paidologos corpus, which resulted from a large-scale cross-linguistic study of obstruent contrasts (see Edwards and Beckman, 2008a,b). For the Paidologos study, 20 adult native speakers of each language completed a picture-prompted word-repetition task. Each group of twenty speakers was balanced across gender. The English-speaking participants were recruited from the Columbus, OH, metropolitan area, while the Japanese-speaking participants were recruited from Tokyo, Japan. All participants passed a hearing screening of otoacoustic emissions at 2, 3, 4, and 5 kHz. Furthermore, none of the participants reported any history of speech, language, or hearing disorders.

B. Materials

The target words for the repetition task were sibilant-initial real words, in which the sibilant occurred before a vowel (see Tables I and II). Since the vowel inventory of English is larger than that of Japanese, the English monophthongs were grouped into classes that correspond roughly to the five Japanese monophthongs /i, e, ɑ, o, u/. These English vowel classes elided certain features, like the tense-lax distinction. Specifically, the English /i/ category comprised /i, ɪ/; /e/, /e, ɛ/; /ɑ/, /ʌ, ɑ, ɔ/; and /u/, /u, ʊ/. For English, three words for each sibilant in each vowel context were chosen, which yielded a total of 15 target words for /s/ and /ʃ/, respectively. In Japanese, /si/ is not a phonotactically legal sequence, and /ɕe/ occurs rarely and only in loan words (e.g., /ɕeri/ "sherry"). Consequently, no target words were included for these two sibilant-vowel sequences, which left 12 target words for each sibilant in Japanese. In addition to these target words, the word list for each language included stop-initial filler words.

For each language, an adult female native speaker, who had received phonetic training, produced multiple repetitions of each word. These productions were recorded digitally at 22.5 kHz, and from these recordings, three repetitions of each word were chosen to combine with other words to create six lists of auditory stimuli (i.e., two ordered-lists for each of the three sets of audio recordings of the words). The order within each list was pseudo-randomized, so that the words for each target sibilant-vowel pair were distributed evenly across the list. Finally, the auditory stimuli were paired with digital images of the referent of the target word, and these audiovisual pairs were used as prompts in the repetition task.

C. Elicitation and recording procedure

The English speakers completed the repetition task inside a sound-attenuated room on the campus of The Ohio State University, and the Japanese speakers were tested in a quiet room in Tokyo, Japan. Prior to the task, the participants were instructed that they would be completing a task that would require them to repeat real words of their native language after being prompted by paired images and audio recordings of those words. The participants completed the task at their own pace, using a custom software program that allowed them to initiate each trial and to track their progress through the task. On a given trial, the software program first displayed the associated image on a computer screen, and then, after a 300 ms delay, played the audio recording over speakers. The adults' repetitions of the test words were spoken into an AKG C5900M (AKG Acoustics, Nashville, TN) condenser microphone with a cardioid response and recorded using a Marantz PMD660 (Marantz, Kanagawa, Japan) flash card recorder with 44.1 kHz sampling frequency. The full duration of the session was recorded for subsequent annotation and acoustic analysis.

TABLE I. The /s/- and /ʃ/-initial target words used in the English word-repetition task.

Vowel context	/s/-initial words	/ʃ/-initial words
/i/	sister	sheep
	seal	shield
	seashore	ship
/e/	safe	shape
	same	shell
	seven	shepherd
/ɑ/	sauce	shark
	soccer	shop
	Sun	shovel
/o/	soak	shore
	sodas	shoulder
	soldier	show
/u/	soup	chute
	suitcase	shoe
	super	sugar

TABLE II. The /s/- and /ɕ/-initial target words used in the Japanese word-repetition task.

	/s/-initial words		/ɕ/-initial words	
Vowel context	Gloss	Transcription	Gloss	Transcription
/i/			"bullet train"	/ɕiɴkɑɴseɴ/
			"seesaw"	/ɕiːsoː/
			"zebra"	/ɕimauma/
/e/	"back"	/senɑkɑ/		
	"cicada"	/semi/		
	"teacher"	/seɴseː/		
/ɑ/	"cherry blossom"	/sɑkuɾɑ/	"rice paddle"	/ɕɑmodʒi/
	"fish"	/sɑkɑnɑ/	"shampoo"	/ɕɑwɑː/
	"monkey"	/sɑɾu/	"shower"	/ɕɑmpuː/
/o/	"sausage"	/soːseːdʒi/	"bread"	/ɕokupɑɴ/
	"sky"	/sokːusu/	"fire engine"	/ɕoːboːɕɑ/
	"socks"	/soɾɑ/	"soy sauce"	/ɕoːju/
/u/	"sand"	/sunɑ/	"dumpling"	/ɕuːmɑi/
	"sparrow"	/sudzume/	"creme puff"	/ɕuːkuriːmu/
	"watermelon"	/suikɑ/	"shoes"	/ɕuːdzu/

D. Annotation of frication landmarks

Trained phoneticians marked the frication onsets and the fricative-vowel boundaries of the target sibilants using a custom Praat script that displayed the recording's waveform and spectrogram simultaneously and that allowed the audio signal to be played at will. Frication onset was marked at the earliest point at which an increase in the waveform's amplitude coincided with the presence of high-frequency energy in the spectrogram. For the fricative-vowel boundary, the onset of periodicity in the vocalic portion was first determined. The fricative-vowel boundary was then marked at the zero-crossing of the waveform's first upswing following the onset of periodicity.

In the English data, two productions of /se/ were not annotated, and were omitted from subsequent analysis, because participants e9gt03fw and e9gt10fw produced "fame" instead of the target "same." This seemed to be due to an ambiguous initial fricative in the audio prompt for "same" that was used in wordlist enrw111. This left 298 productions of /s/ and 300 of /ʃ/ for acoustic analysis.

In the Japanese data, six productions were not annotated because the participant's repetition overlapped the audio prompt that was still playing in the background. These six omitted productions included one token each of /se/, /sɑ/, /so/, and /ɕo/ and two tokens of /ɕi/. This left 237 productions of both /s/ and /ɕ/ for acoustic analysis.

E. Computation of peak ERB_N number trajectories

For each annotated sibilant production, the times of frication onset and fricative-vowel boundary were used to define seventeen 20-ms analysis intervals that were spaced evenly across its duration, such that the first interval was centered at the frication onset and the last, at the fricative-vowel boundary. The amount of overlap or separation between consecutive intervals depended on the duration of the production, which ranged from 75.088 to 382.197 ms in English, and from 62.663 to 264.955 in Japanese. For the English data, interval spacing ranged from 16.76 ms overlap to 1.31 ms separation [mean (M) = 10.04 ms overlap, standard error (SE) = 2.61 ms], and for the Japanese data, from 17.33 ms to 4.69 ms overlap (M = 12.42 ms overlap, SE = 1.86 ms).

A psychoacoustic spectrum was computed for the waveform in each analysis interval by first estimating that waveform's spectrum, and then passing that spectrum through a model of the auditory periphery. A waveform's spectrum was estimated with the multitaper method (Thomson, 1982), using parameter values $K = 8$ and $nW = 4$. The multitaper spectrum (MTS) is equivalent to the pointwise average of K uncorrelated spectra estimated with the discrete Fourier transform (DFT); hence, the asymptotic distributions of the ordinates of an MTS-estimate have $1/K$th the variance of those of a DFT-estimate (Percival and Walden, 1993). A number of studies have estimated sibilant spectra with the MTS, rather than the DFT (e.g., Blacklock, 2004; Koenig *et al.*, 2013; Todd *et al.*, 2011); however, the results presented here are likely not dependent on this methodological choice since the linguistic effects on acoustic measures, such as peak frequency, have been found to be comparable when either the DFT or the multitaper method is used to estimate the spectrum (Reidy, 2015).

To transform an acoustic spectrum into a psychoacoustic spectrum, it was passed through a filter bank that modeled how the auditory periphery logarithmically compresses the frequency scale and how it differentially resolves frequency components across the audible range (see Fig. 1, top). This filter bank comprised 361 fourth-order gammatone filters (see Glasberg and Moore, 1990; Patterson, 1976). The center frequencies of the filters were equally spaced along the ERB_N number scale, every 0.1 from 3 to 39. This scale logarithmically transforms the hertz frequency scale similarly to the tonotopic organization of the basilar membrane (see Greenwood, 1990). The bandwidth of each filter was set to 1.019 times the equivalent rectangular bandwidth of that filter's center frequency in hertz (see Moore *et al.*, 1997; Patterson, 2000); thus, on the hertz scale, the width of each filter increased with its center frequency, similarly to how listening experiments suggest the widths of auditory filters increase with their center frequencies (Moore and Glasberg, 1983). Each gammatone filter in the auditory model acted on an input spectrum as a bandpass filter, parsing the energy content within a limited frequency band of the spectrum. Finally, the psychoacoustic spectrum was constructed by summing the total energy (or "auditory excitation") at the output of each filter and then plotting these excitation levels against the ERB_N numbers of the filters' center frequencies (see Fig. 1, middle).

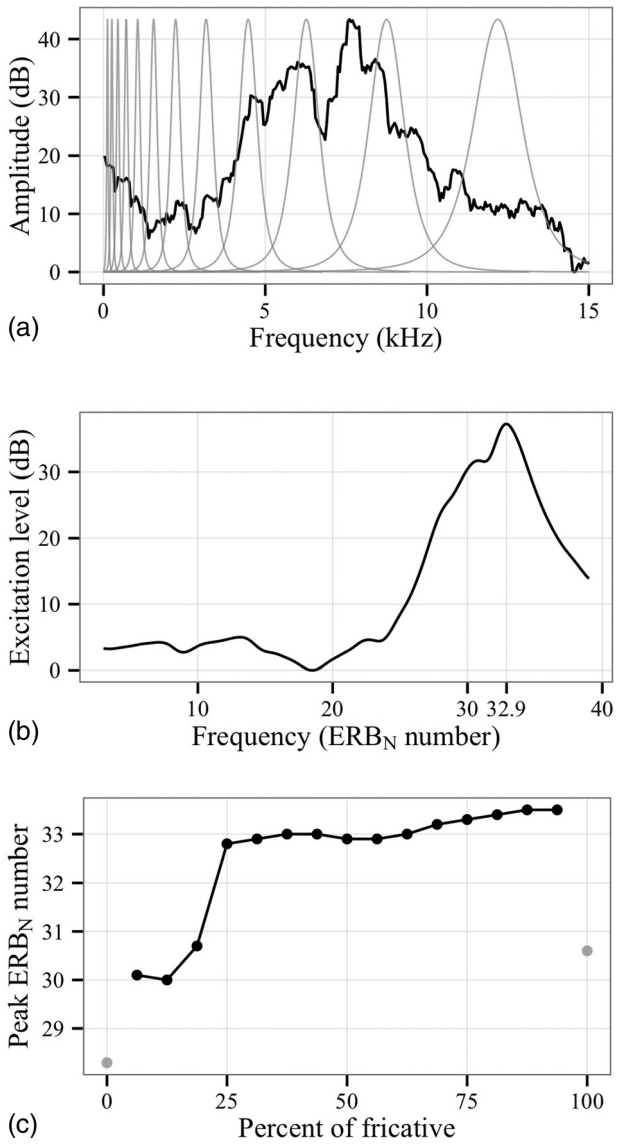

FIG. 1. (Top) MTS estimated from the middle 20 ms of an adult female's production of English /s/. Frequency responses of 12 gammatone filters of different center frequencies are shown overlaid on the spectrum. (Middle) The psychoacoustic spectrum resulting from passing the spectrum in the top panel through a 361-channel gammatone filter bank model of the auditory periphery. Peak frequency in this psychoacoustic spectrum is 32.9 ERB_N numbers. (Bottom) Peak ERB_N number trajectory computed from seventeen 20-ms windows spaced evenly across the duration of /s/. The endpoints of this trajectory were excluded from analysis.

From each psychoacoustic spectrum, the peak psychoacoustic frequency—referred to as "peak ERB_N number"—was computed. Peak ERB_N number denotes the center frequency, in ERB_N numbers, of the channel in the filter bank model that had the greatest amount of excitation at output. Peak ERB_N number corresponds to the image of the acoustic spectrum's peak frequency under the auditory model, and

can be thought of as the psychoacoustic analog of peak frequency. The peak ERB_N number trajectory was represented as a function of percent of the fricative duration (see Fig. 1, bottom).

IV. RESULTS

A. English /s/ and /ʃ/

To analyze how peak frequency varied across the duration of English /s/ and /ʃ/, a growth curve model was fitted to the peak ERB_N number trajectories of these two sibilants. The endpoints of the trajectories were discarded before fitting the growth-curve model because when these points were included, the residuals of the fitted model showed heteroskedasticity at the endpoints; thus, the growth-curve model was fitted to the middle 15 points, which ranged from 6.25% to 93.75% of the fricative duration. The models were fitted using orthogonal polynomial powers of time, up to the fifth power. Previous growth curve analyses of acoustic features of sibilants (e.g., Iskarous *et al.*, 2011) have included time polynomials up to only the second power; however, the present analysis included higher powers of time in order to be able to better capture differences between the tails of the trajectories—i.e., differences between the onset and offset of frication. Since there was no reason to suppose that the onset and offset portions of the trajectories would mirror each other, it was desirable for the model to be able to capture any asymmetries between them.

The fixed- and random-effects structures of the model were built up using a stepwise forward selection protocol: the base model included only a fixed effect intercept and random effects of intercept by talker and by consonant-within-talker. The first step considered a fixed effect of consonant, with /ʃ/ as the reference level. Subsequent steps considered a simple fixed effect of a nonzero power of time, and then an interaction between that power of time and consonant. The powers of time were considered serially in increasing order—e.g., a simple effect of linear time and its interaction with consonant were considered before any effects of quadratic or higher powers of time. At each step, the fixed effects structure was augmented only if a likelihood ratio test revealed that model fit was significantly improved at the $\alpha = 0.05$ level. At the step where a given power of time was added as a simple fixed effect, random effects by talker and by consonant-within-talker were added for that power of time as well.

Table III shows the results of fitting a growth curve model to the peak ERB_N number trajectories. The rows of each table are sorted according to the order in which the fixed effects listed in the first column were added to the model; hence, a fixed effect on a particular row was added after all fixed effects above it, and before all fixed effects below it. The second column in each table displays Akaike information criterion (AIC); each row on this column corresponds to the AIC of the model that includes all fixed effects listed on and above that row. The third, fourth, and fifth columns display the results of likelihood ratio tests performed during model fitting; for a given row r, the values in these columns report the likelihood ratio test that compared (*a*) the model comprising all fixed effects on rows 1 through r to (*b*) the model comprising all fixed effects on rows one through $r - 1$.

The fitted growth-curve model was checked by visually inspecting various diagnostic plots of its residuals. To check that the error terms were independent, the residuals were plotted against time window. A loess smoothing of this scatterplot indicated that the residuals were distributed around zero at each time window, suggesting that there was no temporal correlation between the error terms. To check that the error terms had equal variance within each level of the random effects, a scatterplot of the residuals against the fitted values was made for each consonant within each participant. Only 2 of these 40 scatterplots showed any evidence of heteroskedasticity. A Q-Q plot of the standardized residuals indicated a unimodal distribution with a thin positive tail. Between the 2.5 and 97.5 percentiles, only minor deviations from normality were observed.

The fitted model included simple effects of time up to the fourth power, a simple effect of consonant, and interactions between consonant and linear and quadratic time. The rightmost three columns of Table III report the estimates, standard errors, and confidence intervals for the coefficients in the fitted model—i.e., the model that included all fixed effects listed in the table. The confidence intervals denote percentile bootstrap confidence intervals that were constructed from 1000 replicates of the data. The simple effects of powers of time indicated that peak ERB_N number was not static across the duration of /s/. The positive coefficient for linear time [$\hat{\beta} = 1.722$, SE = 0.240, 95% CI = (1.242,

TABLE III. Results of fitting a growth curve model to the peak ERB_N trajectories of English /s/ and /ʃ/.

		Likelihood ratio test			Fitted model coefficients		
Fixed effect	AIC	df	χ^2 Statistic	p-value	$\hat{\beta}$	Standard Error	Confidence Interval
(Intercept)	36575				32.163	0.376	[31.384, 32.874]
Consonant	36506	1	71.07	<0.001	−6.383	0.343	[−7.043, −5.663]
Time	36134	3	378.60	<0.001	1.722	0.240	[1.242, 2.226]
Time × Consonant	36124	1	11.67	<0.001	−1.198	0.318	[−1.836, −0.574]
$Time^2$	35575	3	555.50	<0.001	−1.678	0.285	[−2.231, −1.126]
$Time^2$ × Consonant	35572	1	4.12	<0.05	0.647	0.311	[0.037, 1.274]
$Time^3$	35499	3	79.43	<0.001	−0.412	0.129	[−0.676, −0.170]
$Time^4$	35496	3	8.99	<0.05	−0.209	0.070	[−0.350, −0.073]

2.226)] indicated that peak ERB_N number increased across the midpoint of the fricative. The negative coefficient for quadratic time [$\hat{\beta} = -1.678$, SE = 0.285, 95% CI = (−2.231, −1.126)] indicated that peak ERB_N number followed a concave downward trajectory. Last, the coefficients for cubic [$\hat{\beta} = -0.412$, SE = 0.129, 95% CI = (−0.676, −0.170)] and quartic time [$\hat{\beta} = -0.209$, SE = 0.070, 95% CI = (−0.350, −0.073)] indicated an asymmetry between the onset and offset of the peak ERB_N number trajectory.

The negative coefficient for the simple effect of consonant [$\hat{\beta} = -6.383$, SE = 0.343, 95% CI = (−7.043, −5.663)] indicated that, at fricative midpoint, the peak ERB_N number of /ʃ/ is lower than that of /s/, as expected. The interactions between consonant and powers of time indicated that, across the duration of /ʃ/, peak ERB_N number varied less than it did for /s/. Relative to the simple effect of linear time, the coefficient for the interaction between consonant and linear time [$\hat{\beta} = -1.198$, SE = 0.318, 95% CI = (−1.836, −0.574)] was opposite in sign and smaller in magnitude; thus, this interaction indicated that, across fricative midpoint, the rate of increase in peak ERB_N number was less for /ʃ/ than for /s/. Similarly, the coefficient for the interaction between consonant and quadratic time [$\hat{\beta} = 0.647$, SE = 0.311, 95% CI = (0.037, 1.274)] was opposite in sign and smaller in magnitude, when compared to the simple effect of quadratic time; thus, this interaction indicated that the curvature of the peak ERB_N number trajectory was less for /ʃ/ than for /s/.

Figure 2 shows the means and standard errors of the observed peak ERB_N numbers of /s/ (darker curve) and /ʃ/ (lighter curve) at each analysis window. Also shown in this figure are 95% percentile bootstrap prediction intervals for the fitted model described in Table III. These prediction intervals were computed from 1000 bootstrap replicates of the observed data; for each replicate the random effects and the errors of the model were resampled. The median bootstrap predictions for each analysis window are shown as the solid lines within the prediction intervals. For both /s/ and /ʃ/, the model predictions indicate that peak ERB_N number follows an increasing concave downward trajectory. Both predicted trajectories reach their maximum at 68.75% of fricative duration. The predicted trajectory for /s/ begins at 30.807 ERB_N numbers, rises to a maximum of 32.914 ERB_N numbers, and then falls to 31.865 ERB_N numbers. After converting these ERB_N numbers to the hertz scale, these excursions correspond to a rise of 1624.919 Hz and a fall of 855.140 Hz. The predicted trajectory for /ʃ/ begins at 25.206 ERB_N numbers, rises to a maximum of 26.206 ERB_N numbers, and then falls to 25.261 ERB_N numbers. These excursions correspond to a rise of 396.220 Hz and a fall of 375.521 Hz.

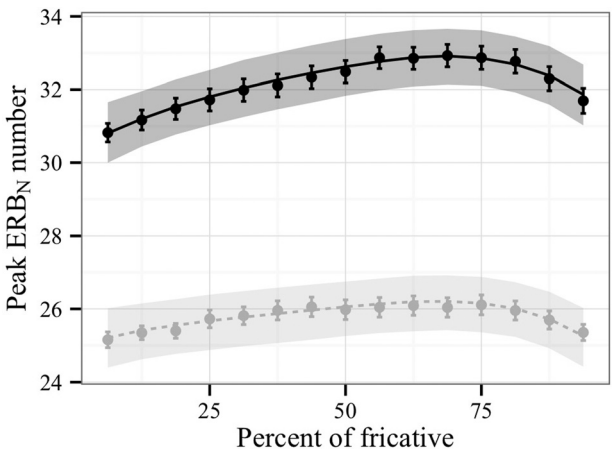

FIG. 2. Observed and predicted peak ERB_N number trajectories for English /s/ (solid, darker curve) and /ʃ/ (dashed, lighter curve). Means and ±2 standard errors of observed trajectories are shown as points with error bars. The 95% bootstrap prediction intervals are shown as ribbons. The line within each prediction interval denotes the median bootstrap prediction for each analysis window.

B. Japanese /s/ and /ɕ/

To analyze temporal variation in peak ERB_N number across Japanese /s/ and /ɕ/, a growth curve model was fitted to the peak ERB_N trajectories computed from the adults' productions of these sibilants. The structure of the model was constructed using a forward stepwise selection protocol, just as was done for the English sibilants. The results of the model-fitting procedure are shown in the third, fourth, and fifth columns of Table IV. The fitted model was again checked by visually inspecting diagnostic plots of its residuals. A loess-smoothed scatterplot of the residuals against the analysis window locations suggested that there was no temporal correlation between the error terms. When the residuals were plotted against the fitted values for each participant and each target sibilant, loess-smoothed curves suggested that the error terms were heteroskedastic in only three of these 40 levels of the random effects. A Q-Q plot of the standardized residuals indicated a unimodal distribution centered at zero, with a long thin negative tail. For all percentiles above the 2.5 percentile, only minor deviations from normality were observed.

The fitted model comprised simple effects of time up to the fourth power, a simple effect of consonant (with /ɕ/ as the reference level), and an interaction between consonant and quadratic time. Estimates, standard errors, and 95% percentile bootstrap confidence intervals of the coefficients in the fitted model are shown in the rightmost three columns of Table IV. The simple effects of powers of time indicated that peak ERB_N number varied across the duration of Japanese /s/. The negative coefficient for linear time [$\hat{\beta} = -2.531$, SE = 0.420, 95% CI = (−3.419, −1.774)] indicated that peak ERB_N number decreased across the midpoint of the fricative. As with English /s/, the negative coefficient for quadratic time [$\hat{\beta} = -4.695$, SE = 0.347, 95% CI = (−5.363, −4.007)] indicated that peak ERB_N number followed a concave downward trajectory. Last, the coefficients for cubic [$\hat{\beta} = -0.965$, SE = 0.151, 95% CI = (−1.257, −0.639)] and quartic time [$\hat{\beta} = -0.509$, SE = 0.150, 95% CI = (−0.816, –0.208)] indicated that the rise of peak ERB_N number near fricative onset was not symmetric to the fall of peak ERB_N number near fricative offset.

The effects involving consonant in the fitted model indicated that peak ERB_N number was lower overall and varied less across the duration of /ɕ/. The negative coefficient for the simple effect of consonant [$\hat{\beta} = -2.442$, SE = 0.360,

TABLE IV. Results of fitting a growth curve model to the peak ERB_N trajectories of Japanese /s/ and /ɕ/.

		Likelihood ratio test			Fitted model coefficients		
Fixed effect	AIC	df	χ^2 Statistic	p-value	$\hat{\beta}$	Standard Error	Confidence Interval
(Intercept)	36546				31.276	0.375	[30.524, 32.015]
Consonant	36524	1	24.64	<0.001	−2.442	0.360	[−3.184, −1.731]
Time	36084	3	445.42	<0.001	−2.531	0.420	[−3.419, −1.774]
$Time^2$	35298	3	791.76	<0.001	−4.695	0.347	[−5.363, −4.007]
$Time^2 \times$ Consonant	35279	1	21.68	<0.001	2.313	0.384	[1.536, 3.050]
$Time^3$	35228	3	57.04	<0.001	−0.965	0.151	[−1.257, −0.639]
$Time^4$	35217	3	17.13	<0.001	−0.509	0.150	[−0.816, −0.208]

95% CI = (−3.184, −1.731)] indicated that, at fricative midpoint, the peak ERB_N number of /ɕ/ was lower than that of /s/. Compared to the simple effect of quadratic time, the coefficient for the interaction between consonant and quadratic time [$\hat{\beta} = 2.313$, SE = 0.384, 95% CI = (1.536, 3.050)] was opposite in sign and smaller in magnitude; thus, this interaction indicated that the curvature of the peak ERB_N number trajectory was less for /ɕ/ than for /s/.

Figure 3 shows the means and standard errors of the observed peak ERB_N number trajectories, as well as the fitted model's predictions at each analysis window. For Japanese /s/ (darker curve), the model predictions indicate that peak ERB_N number follows a concave downward trajectory that falls more across its second half than it does rise across its first half. In particular, the predicted trajectory for /s/ begins at 30.366 ERB_N numbers, rises to a maximum of 32.484 ERB_N numbers at fricative midpoint, and then falls to 27.345 ERB_N numbers. On the hertz scale, these excursions correspond to a rise of 1558.432 Hz and a fall of 3246.688 Hz. The model predictions for /ɕ/ (lighter curve) indicate that peak ERB_N number begins at 29.002 ERB_N numbers and rises slightly to a maximum of 29.413 ERB_N numbers at 31.25% of fricative duration—a rise equivalent to one of just 237.544 Hz. Peak ERB_N number then decreases only slightly until 62.50% of fricative duration, after which point it decreases from 29.270 to 25.985 ERB_N numbers—i.e., a drop of 1693.040 Hz.

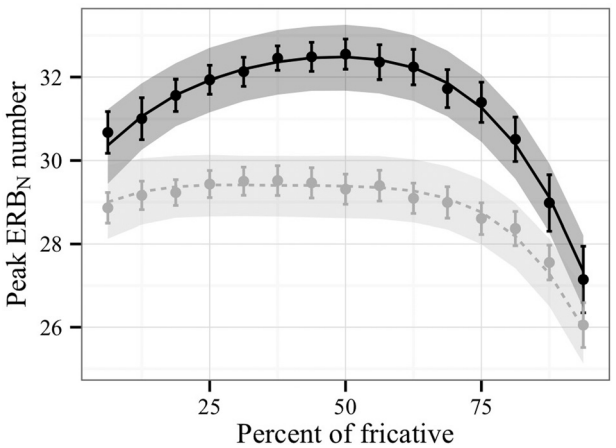

FIG. 3. Observed and predicted peak ERB_N number trajectories for Japanese /s/ (solid, darker curve) and /ɕ/ (dashed, lighter curve). Means and ±2 standard errors of the observed trajectories are shown as points with error bars. 95% bootstrap prediction intervals are shown as ribbons. The line within each prediction interval denotes the median bootstrap prediction for each analysis window.

C. Cross-linguistic comparison of /s/

To analyze cross-linguistic differences in how peak ERB_N varies temporally in sibilants, a growth curve model was fitted to the peak ERB_N trajectories of the English- and Japanese-speaking adults' productions of /s/. For this analysis, the model was fitted to just the middle 50% of the fricative in order to compare the "steady-state" interval of the sibilants, and to attempt to minimize the language-specific effects of resting articulator posture or of vowel-on-sibilant coarticulation. The model structure was determined through a stepwise forward selection protocol, the results of which are shown in Table V. Different from the language-specific models, the cross-linguistic model included random effects of intercept and powers of time by language and by participant-within-language.

The fitted model was checked through diagnostic plots of the residuals. A loess-smoothed scatterplot of the residuals against the analysis window locations indicated no temporal correlation between the error terms. Plotting the residuals against fitted values for each participant suggested that the error terms may be heteroskedastic for three of the Japanese talkers. A Q-Q plot of the standardized residuals indicated a unimodal distribution centered at zero, with a long thin negative tail. For all percentiles above the 2.5 percentile, only minor deviations from normality were observed.

The fitted model included an intercept, simple effects of linear and quadratic time, and interactions between language and the nonzero powers of time. The rightmost three columns of Table V report the estimates, standard errors, and 95% percentile bootstrap confidence intervals for the fitted coefficients. The simple effects of linear [$\hat{\beta} = 1.217$, SE = 0.238, 95% CI = (0.737, 1.679)] and quadratic time [$\hat{\beta} = -0.286$, SE = 0.151, 95% CI = (−0.588, 0.015)] indicated that peak ERB_N number followed an increasing linear trajectory across the middle half of English /s/. The interactions between language and linear [$\hat{\beta} = -1.707$, SE = 0.342, 95% CI = (−2.367, −1.039)] or quadratic time [$\hat{\beta} = -0.675$, SE = 0.222, 95% CI = (−1.112, −0.235)] indicated that, across the middle half of Japanese /s/, peak frequency followed a trajectory that increased less but was more curved than the analogous trajectory for English /s/.

TABLE V. Results of fitting a growth curve model to the peak ERB_N trajectories of English /s/ and Japanese /s/.

		Likelihood ratio test			Fitted model coefficients		
Fixed effect	AIC	df	χ^2 Statistic	p-value	$\hat{\beta}$	Standard Error	Confidence Interval
(Intercept)	21268				32.305	0.293	[31.759, 32.896]
Time	21134	2	129.32	<0.001	1.217	0.238	[0.737, 1.679]
Time × Language	21125	1	20.06	<0.001	−1.707	0.342	[−2.367, −1.039]
$Time^2$	21080	2	48.10	<0.001	−0.286	0.151	[−0.588, 0.015]
$Time^2$ × Language	21074	1	8.68	<0.005	−0.675	0.222	[−1.112, −0.235]

The means and standard errors of the observed peak ERB_N number trajectories are shown in Fig. 4, along with the fitted model's predictions at each analysis window. The model predictions indicate that peak frequency rises monotonically across the middle half of English /s/ (darker curve), from 31.532 ERB_N numbers, at 25% fricative duration, to 32.786 ERB_N numbers, at 75% fricative duration—a rise in peak frequency that corresponds to an increase of 997.052 Hz. For Japanese /s/ (lighter curve), the model predictions indicate that the peak frequency trajectory reaches a maximum of 32.671 ERB_N numbers at fricative midpoint. Before fricative midpoint, peak frequency rises 0.620 ERB_N numbers—or, equivalently 503.553 Hz. After midpoint, peak frequency falls 1.127 ERB_N numbers—or, equivalently 891.350 Hz.

V. DISCUSSION

A. Language-internal differences in sibilants' spectral dynamics

In both the growth-curve model fitted to the English sibilants and the one fitted to the Japanese sibilants, there was a significant interaction between consonant and at least one nonzero power of time, which indicated that within each language the shape of the peak ERB_N number trajectory differs across the two sibilant fricatives. One noticeable commonality between the results for English and Japanese is that in both languages, peak ERB_N number exhibited less temporal variation for the sibilant that had the larger front cavity. Using a computational model of the vocal tract, Toda and Maeda (2006) simulated its transfer function in response to turbulence noise sources, while varying the size of the front cavity and the constriction. They used the resonant frequencies of these simulated transfer functions to develop a map of how changes in front cavity size relate to changes in resonant frequency. Their simulations suggest that changes in front cavity size—which might arise from the upward movement of the jaw during articulation of the sibilant—perturb the resonant frequency when either the front cavity is relatively large (or when the constriction is relatively long. Thus, the large front cavity of English /ʃ/ may explain the relative stability of peak ERB_N number across its duration. Similarly, the long palatal constriction of Japanese /ɕ/ may contribute to it having a flatter peak ERB_N number trajectory, at least across the middle half of the fricative.

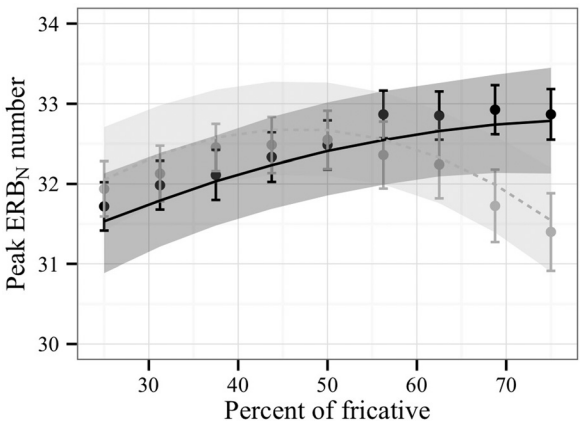

FIG. 4. Observed and predicted peak ERB_N number trajectories for English /s/ (solid, darker curve) and Japanese /s/ (dashed, lighter curve). Means and ±2 standard errors of the observed trajectories are shown as points with error bars. 95% bootstrap prediction intervals are shown as ribbons. The line within each prediction interval denotes the median bootstrap prediction for each analysis window.

B. Language-specificity in sibilants' spectral dynamics

In the model that compared English /s/ and Japanese /s/, the significant interactions of consonant with linear and quadratic time indicated that these two sounds differ in terms of the dynamic aspects of their peak ERB_N number trajectories (i.e., their slope and curvature), rather than the static properties of these trajectories (i.e., their level). Indeed, in Fig. 4, the prediction intervals for these two sounds overlap at every time-window analyzed; however, the difference in shape between the two predicted trajectories is apparent.

Under the task-dynamic model speech production (Fowler and Saltzman, 1993; Saltzman and Munhall, 1989), sequences of sounds, such as a fricative-vowel syllable, are composed through gestural scores that determine how the gestures of the individual sounds are coproduced. Thus, one interpretation of these cross-linguistic differences in the peak ERB_N number trajectories of English /s/ and Japanese /s/ is that they are due to language-specific differences in gestural coproduction: In Japanese, the fricative and vowel gestures are coproduced with greater temporal overlap such that the jaw lowers and the linguapalatal constriction releases earlier in the course of the fricative, in anticipation of the upcoming vowel. Under this view, then, temporal variation in peak ERB_N number is epiphenomenal, falling out from the coordination of gestures.

Even if such a view is adopted, though, the importance of a sibilant's specific spectral dynamics is maintained from the perspective of language acquisition. If the coproduction of gestures is governed by language-specific principles, then these principles must be learned during acquisition. Furthermore, since an infant may have, at best, only partial access to the articulators of the adult caretakers—i.e., through the infant's vision of the adult's lip and jaw movement—the spectral dynamics and other time-varying properties of speech must be leveraged to learn fluent, language-specific gestural coordination.

C. Potential artifacts of time normalization

In the analyses presented here, the peak ERB_N number trajectory of each production was represented by a 15-point sequence that ranged across the productions' proportional duration, rather than its absolute duration. The decision to analyze temporal variation in peak frequency relative to proportional duration was made in order to maintain methodological consistency with previous studies that also analyzed the acoustic and articulatory dynamics of sibilants relative to proportional duration (e.g., Iskarous *et al.*, 2011; Koenig *et al.*, 2013; Zharkova *et al.*, 2014); however, there is the worry that some of the apparent differences between the peak ERB_N number trajectories of two sibilants may have arisen as artifacts of the normalization to proportional time. For example, one consequence of normalizing each production to proportional time is that as the duration of a production decreased, the overlap between adjacent analysis windows increased. Assuming that peak ERB_N number varies smoothly across a sibilant, then, in the limit, where the duration of the production is similar to the duration of the analysis window, there would be significant overlap between adjacent analysis windows and little variation in peak ERB_N number between adjacent points in the trajectory.

In English, the effects of consonant on linear and quadratic time (see Table III) indicated that the peak ERB_N number trajectory for /ʃ/ was more horizontal and had shallower curvature than the trajectory for /s/, indicating that peak frequency varied less across /ʃ/. If the productions of English /ʃ/ were significantly shorter than those of English /s/, then the relative stability of peak ERB_N number across the former sibilant may simply have been due to multiple points in its peak ERB_N number trajectory having been computed from overlapping data. To assess this possibility, a linear mixed effects model of consonant duration, with random effects for participant and for consonant-within-participant was built for the English sibilants. In this model, the effect of consonant was estimated with /ʃ/ as the reference level. A significant positive effect of consonant [$\hat{\beta} = 10.835$, SE = 3.147, 95% CI = (4.666, 17.003)] indicated that the productions of /ʃ/ ($\hat{\mu} = 196.072$, SE = 43.466 ms) were longer than the /s/ productions ($\hat{\mu} = 185.155$, SE = 44.658 ms). Thus, in English, differences in the temporal variation in peak ERB_N number between /s/ and /ʃ/ do not seem to be artifacts of normalizing the productions to proportional time, since the differences are in the opposite direction from what would be expected given the observed duration differences between the two sibilants.

A similar argument may be made for the Japanese sibilants. In particular, the effect of consonant on quadratic time (see Table IV) indicated that the peak ERB_N number trajectory for /ɕ/ had shallower curvature than the trajectory for /s/, indicating that the variation in peak frequency was less for /ɕ/. Comparing the durations of these two sibilants, a linear mixed effects model revealed a significant positive effect of consonant [$\hat{\beta} = 9.312$, SE = 2.497, 95% CI = (4.419, 14.205)], indicating that /ɕ/ ($\hat{\mu} = 146.044$, SE = 27.355 ms) was longer in duration than /s/ ($\hat{\mu} = 136.620$, SE = 31.377 ms). Consequently, the lower variation in peak ERB_N number observed across /ɕ/ does not seem to be due to the time-normalization method employed prior to the growth-curve analysis.

D. Implications of the current study

The results presented here suggest that a static characterization of sibilant fricatives elides language- and consonant-specific acoustic information. Yet, a number of studies, using only static spectral measures, have sought to characterize the development of a child's productive knowledge of how a phonological sibilant contrast is implemented phonetically (e.g., Fox and Nissen, 2005; Li, 2012; Li *et al.*, 2009; McGowan and Nittrouer, 1988; Nissen and Fox, 2005; Nittrouer *et al.*, 1989; Romeo *et al.*, 2013). Since the spectral patterns that must be acquired and produced by a language learner are dynamic in nature, the view of acquisition provided by static measures is likely incomplete. A more fine-grained perspective of children's development toward adult-like sibilant fricative categories is likely to be had once the spectral dynamics of children's productions are considered.

E. Limitations of the current study

The present analyses have demonstrated temporal variation in peak ERB_N number of word-initial sibilant fricatives; thus, the conclusions are limited in so far as they do not inform how spectral features other than peak frequency varies temporally, or how the effect of prosodic position affects the spectral dynamics of sibilant fricatives. In previous work, Nossair and Zahorian (1991) found that dynamic aspects of global spectral properties (i.e., DCTC coefficients) better characterized initial stop consonants than did dynamic aspects of local spectral properties (i.e., formants). The current work examined only a single local property; hence, an even better understanding of sibilants' spectral dynamics may be achieved through parameters such as DCTC coefficients that index global shape properties of the spectrum. Other research has found that prosodic position may affect the spectral properties of fricatives (Silbert and de Jong, 2008). While the authors considered only static (time-averaged) spectral properties of the fricatives, it is possible that their findings may extend to dynamic spectral properties of fricatives.

ACKNOWLEDGMENTS

This research was supported by National Institutes of Health Grant No. R01DC02932 to Jan Edwards and Mary Beckman, and by a Raymond H. Stetson Scholarship from

the Speech Communication Technical Committee of the Acoustical Society of America to the author. Many thanks are due to Mary Beckman for helpful comments.

Akamatsu, T. (**1997**). *Japanese Phonetics: Theory and Practice* (Lincom Europa, Newcastle, UK), Vol. 3, 429 pp.

Behrens, S. J., and Blumstein, S. E. (**1988**). "Acoustic characteristics of English voiceless fricatives: A descriptive analysis," J. Phonetics **16**, 295–298.

Blacklock, O. S. (**2004**). "Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods," Ph.D. thesis, University of Southampton, Southampton, UK.

Brunner, J., Ghosh, S., Hoole, P., Matthies, M., Tiede, M., and Perkell, J. (**2011**). "The influence of auditory acuity on acoustic variability and the use of a motor equivalence during adaptation to a perturbation," J. Speech Lang. Hear. Res. **54**, 727–739.

Edwards, J. R., and Beckman, M. E. (**2008a**). "Methodological questions in studying consonant acquisition," Clin. Linguist. Phonet. **22**(12), 937–956.

Edwards, J. R., and Beckman, M. E. (**2008b**). "Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development," Lang. Learn. Dev. **4**(2), 122–156.

Fletcher, S. G., and Newman, D. G. (**1991**). "[s] and [ʃ] as a function of linguapalatal contact place and sibilant groove width," J. Acoust. Soc. Am. **89**(2), 850–858.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**(1), 115–124.

Fowler, C. A., and Saltzman, E. (**1993**). "Coordination and coarticulation in speech production," Lang. Speech **36**(2,3), 171–195.

Fox, R. A., and Nissen, S. L. (**2005**). "Sex-related acoustic changes in voiceless English fricatives," J. Speech Lang. Hear. Res. **48**(4), 753–765.

Fujisaki, H., and Kunisaki, O. (**1978**). "Analysis, recognition, and perception of voiceless fricative consonants in Japanese," IEEE Trans. Acoust. Speech ASSP **26**(1), 21–27.

Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A., Tiede, M., Ménard, L., Guenther, F. H., Lane, H., and Perkell, J. S. (**2010**). "An investigation of the relation between sibilant production and somatosensory and auditory acuity," J. Acoust. Soc. Am. **128**(5), 3079–3087.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hearing Res. **47**, 103–138.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species–29 years later," J. Acoust. Soc. Am. **87**(6), 2592–2605.

Haley, K. L., Seelinger, E., Mandulak, K. C., and Zajac, D. J. (**2010**). "Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach," J. Phonetics **38**, 548–554.

Heinz, J. M., and Stevens, K. N. (**1961**). "On the properties of voiceless fricatives," J. Acoust. Soc. Am. **33**(5), 589–596.

Holliday, J. J., Reidy, P. F., Beckman, M. E., and Edwards, J. (**2015**). "Quantifying the robustness of the English sibilant fricative contrast in children," J. Speech Lang. Hear. Res. **58**, 622–637.

Hughes, G. W., and Halle, M. (**1956**). "Spectral properties of fricative consonants," J. Acoust. Soc. Am. **28**(2), 303–310.

Iskarous, K., Shadle, C. H., and Proctor, M. I. (**2011**). "Articulatory-acoustic kinematics: The production of American English /s/," J. Acoust. Soc. Am. **129**(2), 944–954.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**(3), 1252–1263.

Koenig, L. L., Shadle, C. H., Preston, J. L., and Mooshammer, C. R. (**2013**). "Toward improved spectral measures of /s/: Results from adolescents," J. Speech Lang. Hear. Res. **56**(4), 1175–1189.

Li, F. (**2012**). "Language-specific developmental differences in speech production: A cross-language acoustic study," Child Dev. **83**(4), 1303–1315.

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

Maniwa, K., Jongman, A., and Wade, T. (**2009**). "Acoustic characteristics of clearly spoken English fricatives," J. Acoust. Soc. Am. **125**(6), 3962–3973.

McGowan, R. S., and Nittrouer, S. (**1988**). "Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/," J. Acoust. Soc. Am. **83**(1), 229–236.

McLeod, S., Roberts, A., and Sita, J. (**2006**). "Tongue/palate contact for the production of /s/ and /z/," Clin. Linguist. Phonet. **20**(1), 51–66.

McMurray, B., and Jongman, A. (**2011**). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," Psychol. Rev. **118**(2), 219–246.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**(3), 750–753.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**(4), 224–240, available at http://www.aes.org/e-lib/browse.cfm?elib=10272.

Mooshammer, C., Hoole, P., and Geumann, A. (**2006**). "Interarticulator cohesion within coronal consonant production," J. Acoust. Soc. Am. **120**(2), 1028–1039.

Munson, B. (**2001**). "A method for studying variability in fricatives using dynamic measures of spectral mean," J. Acoust. Soc. Am. **110**(2), 1203–1206.

Narayanan, S. S., and Alwan, A. A. (**2000**). "Noise source models for fricative consonants," IEEE Trans. Speech Audio Processing **8**(2), 328–344.

Narayanan, S. S., Alwan, A. A., and Haker, K. (**1995**). "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am. **98**(3), 1325–1347.

Newman, R. S., Clouse, S. A., and Burnham, J. L. (**2001**). "The perceptual consequences of within-talker variability in fricative production," J. Acoust. Soc. Am. **109**(3), 1181–1196.

Nissen, S. L., and Fox, R. A. (**2005**). "Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective," J. Acoust. Soc. Am. **118**(4), 2570–2578.

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

Nossair, Z. B., and Zahorian, S. A. (**1991**). "Dynamic spectral shape features as acoustic correlates for initial stop consonants," J. Acoust. Soc. Am. **89**(6), 2978–2991.

Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**(3), 640–654.

Patterson, R. D. (**2000**). "Auditory images: How complex sounds are represented in the auditory system," J. Acoust. Soc. Jpn. **21**(4), 183–190.

Percival, D. B., and Walden, A. T. (**1993**). *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques* (Cambridge University Press, Cambridge, UK), pp. 331–374.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F. H. (**2004**). "The distinctness of speakers' /s/–/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect," J. Speech Lang. Hear. Res. **47**(6), 1259–1269.

Reidy, P. F. (**2015**). "A comparison of spectral estimation methods for the analysis of sibilant fricatives," J. Acoust. Soc. Am. **137**(4), EL248–EL254.

Romeo, R., Hazan, V., and Pettinato, M. (**2013**). "Developmental and gender-related trends of intra-talker variability in consonant production," J. Acoust. Soc. Am. **134**(5), 3781–3792.

Saltzman, E., and Munhall, K. (**1989**). "A dynamical approach to gestural patterning in speech production," Ecol. Psychol. **1**, 333–382.

Shadle, C. H. (**1991**). "The effect of geometry on source mechanisms of fricative consonants," J. Phonetics **19**, 409–424.

Silbert, N., and de Jong, K. (**2008**). "Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production," J. Acoust. Soc. Am. **123**(5), 2769–2779.

Stevens, K. N. (**1971**). "Airflow and turbulence noise for fricative and stop consonants: Static considerations," J. Acoust. Soc. Am. **50**(4B), 1180–1192.

Stone, M., Faber, A., Raphael, L. J., and Shawker, T. H. (**1992**). "Cross-sectional tongue shape and linguapalatal contact patterns in [s], [ʃ], and [l]," J. Phonetics **20**(2), 253–270.

Thomson, D. J. (**1982**). "Spectrum estimation and harmonic analysis," Proc. IEEE **70**, 1055–1096.

Toda, M. (**2007**). "Speaker normalization of fricative noise: Considerations on language-specific contrast," in *Proceedings of the 16th International Congress of Phonetic Sciences*, edited by J. Trouvain and W. J. Barry (Saarland University, Germany), pp. 825–828.

Toda, M., and Honda, K. (**2003**). "An MRI-based cross-linguistic study of sibilant fricatives," in *Proceedings of the 6th International Seminar on Speech Production* (Macquarie University, Sydney, Australia), pp. 1–6.
Toda, M., and Maeda, S. (**2006**). "Quantal aspects of non-anterior sibilant fricatives: A simulation study," in *Proceedings of the 7th International Seminar on Speech Production* (Cefala, Ubatuba, Brazil), pp. 573–580.
Todd, A. E., Edwards, J. R., and Litovsky, R. Y. (**2011**). "Production of contrast between sibilant fricatives by children with cochlear implants," J. Acoust. Soc. Am. **130**(6), 3969–3979.
Zharkova, N., Hewlett, N., Hardcastle, W. J., and Lickley, R. J. (**2014**). "Spatial and temporal lingual coarticulation and motor control in preadolescents," J. Speech Lang. Hear. Res. **57**, 374–388.