

# Phonetic enhancement of sibilants in infant-directed speech

Alejandrina Cristià<sup>a)</sup>

Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS-DEC-ENS-CNRS, Paris 75005, France

(Received 25 October 2009; revised 28 April 2010; accepted 28 April 2010)

The hypothesis that vocalic categories are enhanced in infant-directed speech (IDS) has received a great deal of attention and support. In contrast, work focusing on the acoustic implementation of consonantal categories has been scarce, and positive, negative, and null results have been reported. However, interpreting this mixed evidence is complicated by the facts that the definition of phonetic enhancement varies across articles, that small and heterogeneous groups have been studied across experiments, and further that the categories chosen are likely affected by other characteristics of IDS. Here, an analysis of the English sibilants /s/ and /ʃ/ in a large corpus of caregivers' speech to another adult and to their infant suggests that consonantal categories are indeed enhanced, even after controlling for typical IDS prosodic characteristics. © 2010 Acoustical Society of America.

[DOI: 10.1121/1.3436529]

PACS number(s): 43.71.Ft, 43.71.Es [RSN]

Pages: 424–434

## I. INTRODUCTION

In many (if not all) cultures, people speak differently when addressing infants and children than when talking to other adults (Lieven, 1994). The driving forces behind the acoustic characteristics of this special register, variably referred to as motherese, baby talk, or infant-directed speech (IDS), have long been the object of speculation. Most researchers agree that caregivers adapt their speech in order to engage their children's attention and maintain positive rapport by enhancing the expression of emotion and affect (Fernald, 1989, 1992; Fernald and Simon, 1984; Kaplan *et al.*, 1991; Stern *et al.*, 1982; Uther *et al.*, 2007). In addition, some researchers further hypothesize that IDS may also serve to facilitate language acquisition (Englund, 2005; Englund and Behne, 2006; Fernald, 2000; Liu *et al.*, 2007; Sundberg, 2001; Werker *et al.*, 2007).

While the idea that caregivers may enhance linguistic categories when addressing their children is as old as the concept of IDS itself (see e.g., Ferguson, 1964), it was only recently that researchers turned to the acoustic signal as a possible level at which this enhancement may apply. This recent work indeed suggests that across-category acoustic distance is augmented in speech to infants when vowels and tones are measured (on vowels: Burnham *et al.*, 2002; Kirchoff and Schimmel, 2005; Kuhl *et al.*, 1997; but see Englund and Behne, 2006; on tones: Liu *et al.*, 2007, 2009; but see Kitamura *et al.*, 2001; Papoušek and Hwang, 1991).

In contrast, consonantal contrasts have received less attention, and the little work on them has reported conflicting results: some report enhancement (e.g., Malsheen, 1980; Sundberg, 2001), and others no difference or even deterioration (e.g., Baran *et al.*, 1977; Sundberg and Lacerda, 1999). Given that consonants play a fundamental role in word-learning and lexical access, both in childhood and later on (to give just a few examples, Cutler *et al.*, 2000; Nazzi, 2005;

New *et al.*, 2008), it is important to assess the possibility that consonantal categories may be enhanced in speech addressed to young children.

Here, this question is revisited. Previous mixed results are interpreted as the unfortunate outcome of four limiting factors. First, there have been two different definitions of phonetic enhancement, one of which did not take into account the identifying acoustic features of the particular contrasts under study. Second, enhancement may be modulated by the child's age, in which case even negative results could be meaningful. Third, side effects of the prosody of IDS, when not controlled for, may introduce additional variability that complicates documenting phonetic enhancement. Finally, small sample sizes may have limited the power to detect trends. In addition to this interpretation of existing literature, new evidence for phonetic enhancement of consonants is provided bearing in mind all these factors. Specifically, the contrast between /s/ and /ʃ/ was studied within a perceptually-based definition of phonetic enhancement, using a large corpus of speech to two very different infant age groups, and possible confounding factors were carefully considered.

### A. Interpreting previous mixed results on consonantal phonetic enhancement

To date, only a few studies have looked at consonantal enhancement (Baran *et al.*, 1977; Englund, 2005; Englund and Behne, 2006; Malsheen, 1980; Sundberg, 2001; Sundberg and Lacerda, 1999). Except for one study on /s/ duration (Englund and Behne, 2006), all these have investigated Voice Onset Time (the lag between the constriction release and the onset of voicing; VOT), an important acoustic cue to stop voicing.

#### 1. Conceptualizing phonetic enhancement

The first complication in the interpretation of previous evidence on phonetic enhancement relates to the fact that two very different concepts of enhancement have been used. In some of that research, enhancement is conceived as an

<sup>a)</sup>Electronic mail: alecristia@gmail.com

exaggeration along the temporal dimension (Englund, 2005; Englund and Behne, 2006; Sundberg, 2001; Sundberg and Lacerda, 1999), possibly based on the assumption that this temporal expansion draws the listener's attention toward the sounds thus emphasized. Along these lines, overall duration for /s/ was reported to be stably longer in speech to 1- to 6-month-old Norwegian children (Englund and Behne, 2006). However, while some have reported longer VOTs in IDS (Englund, 2005; Sundberg, 2001), others have found VOTs to be overall shorter in IDS (Sundberg and Lacerda, 1999).

Furthermore, this view does not take into consideration whether the temporal dimension involved is perceptually relevant either for the category under study, or for how this category enters into contrastive relationships with other sounds in the language. For instance, lengthening of VOT has been considered enhancement, regardless of whether it results in an expansion or a compression of the voiced/voiceless contrast, and longer /s/s would be viewed as enhanced, with no consideration that length is not a relevant dimension for any contrast involving /s/ or for the definition of /s/ as a category (Gordon *et al.*, 2002). Therefore, it is harder to explain how this enhancement would affect learners' perception to facilitate language acquisition.

A second definition of phonetic enhancement arises from the work on sound implementation in clear speech (Bond and Moore, 1994). Clear speech is that oriented to listeners who have a difficulty understanding (normal listeners in noisy environments, hearing-impaired individuals, non-native interlocutors; Biersack *et al.*, 2005; Lindblom, 1990; Payton *et al.*, 1994; Picheny *et al.*, 1985) as well as the speech of talkers who are highly intelligible (Bradlow *et al.*, 1996). This literature suggests that phonetic enhancement is heavily dependent on the sound inventory, such that categories become more different from sounds that are neighboring in phonetic space. Thus, enhancement of consonantal categories is specific to the contrasts in which they participate, and occurs along perceptually relevant dimensions. An excellent example of how important this is comes from the results reported by Kang and Guion (2008), who show that speakers expand consonantal contrasts along the acoustic dimensions that *they themselves* rely on to discriminate the contrastive sounds. Talkers who distinguish voiced and voiceless stops on the basis of VOT tend to expand the distance along the VOT dimension. In contrast, speakers who rely on the pitch at the onset of a following vowel increase the distance between voicing categories along this pitch dimension.

Returning to the literature on IDS, for English stops, one would predict that mothers would increase the VOT difference between voiced and voiceless stops, rather than making VOTs overall longer regardless of the category. Unfortunately, not even within this perceptually-based definition do results line up neatly. Thus, Malsheen (1980) reports that only some mothers increased the VOT difference between voiced and voiceless stops to differ in IDS, while Baran *et al.* (1977) report that 1 out of 3 mothers in their sample did so.

## 2. Phonetic enhancement and age

It is conceivable that enhancement of consonants may vary with the child's age and development, since such a variation has already been documented in vowels Ratner (1984) and tones (Liu *et al.*, 2009). Three hypotheses would predict age-dependent phonetic enhancement. Malsheen (1980) has proposed that caregivers emphasize phonetic categories in response to the child's first comprehensible productions. In support of her hypothesis, she reports voicing enhancement in the speech of both caregivers whose child was at the first-words or holophrastic stage at the time of recording, while results were more variable in the other 2 groups (whose children were in the babbling and phrase stages). On the other hand, this hypothesis cannot account for the vowel space expansion repeatedly documented in speech to prelinguistic infants, as documented by Kuhl *et al.* (1997) among others.

A second possibility has been put forward by Sundberg (1998), who suggested that consonantal categories could be underspecified in speech directed to younger infants, and overspecified when addressing older infants, and the opposite would be found for vowels. This timeline responds in part to developmental changes in infants' perception of native vowel categories, which occurs earlier than those for consonants (e.g., compare 6 months in Kuhl *et al.*, 1992 and 10–12 months in Werker and Tees, 1984). In consonance with this hypothesis, she reports shorter VOTs in speech to 3-month-olds (Sundberg and Lacerda, 1999) but longer ones in speech to 11- to 14-month-olds (Sundberg, 2001). However, Englund and Behne (2006) found no support for Sundberg's model of overspecification of phonetic contrasts dependent on the infant's age. They measured duration of /s/ and vowel space size longitudinally in speech to infants who were between 1 and 6 months of age. According to their interpretation of the model, the vowel space size should decrease with infants' age, and to be larger in IDS than ADS. Additionally, the model predicts that duration of /s/ ought to increase with age, as consonantal categories would be emphasized in speech to older infants. Neither prediction was supported.

A third possibility arises from the literature documenting prosodic characteristics of IDS. This work shows that the prosody of IDS changes markedly over the first year (e.g., Kitamura and Burnham, 2003). If consonantal implementation is at least partly modulated by variation in prosody (e.g., changes in speech rate), then we can expect it to vary over the first year. This hypothesis has not received as much attention as the others, but it will be considered as a possible explanation in the corpus analysis reported below. In order to consider this possibility, it is necessary to bear in mind segmental side-effects of IDS prosody, as noted in the next subsection.

## 3. Segmental side effects of IDS prosody

The study of the acoustic implementation of consonants in IDS should take into account the complex articulatory bases of these and other IDS acoustic characteristics, and the possibility of cross-talk between the prosodic and the seg-

mental levels. The typical prosodic profile of IDS includes higher pitch, larger pitch excursions, and lower speech rate. A higher average pitch is found in almost all studies (Grieser and Kuhl, 1988; Stern *et al.*, 1983; but see Pye, 1986). Expanded pitch ranges or greater pitch variability is frequently reported (e.g., Fernald and Simon, 1984; Golinkoff and Ames, 1979; Grieser and Kuhl, 1988; Niwano and Sugai, 2003; see also Shute and Wheldall, 1999, who suggest this depends on the talker's gender; Fernald *et al.*, 1989 for evidence of cross-linguistic variation in the use of pitch range in IDS; and Kitamura *et al.*, 2001, and Stern *et al.*, 1983, for evidence that the modulation of pitch depends on the child's age and sex). Finally, lower speech rates are also commonly reported (Grieser and Kuhl, 1988).

Each of these prosodic characteristics could affect the acoustic characteristics of speech sounds in complex ways. Let us take speech rate as an example, since this timing parameter is a consideration for the VOT research summarized above. To begin with, there is individual variation in how speech rate is reduced, by lengthening pauses or words, according to findings from the clear speech literature (Picheny *et al.*, 1985). Furthermore, speech rate may affect some categories more than others; for instance, Kessinger and Blumstein (1997) reports that rate affects long-lag VOT to a larger extent than short-lag VOT. If speakers lengthen long-lag VOT items when speaking slowly more than they lengthen short-lag VOT items, slowed speech would result in apparent phonetic enhancement (greater difference in VOT in the slower, IDS speech). However, such differences may not be the result of any intention to clarify the phonetic distinction on the part of the speaker, and may instead be a side effect of the desire to speak more slowly. In short, as a result of these 2 complicating factors, it is unclear what kinds of controls would be ideal to "clean" the effects of speech rate on VOT. For example, Englund (2005) normalizes on the basis of vowels immediately following the stop, which would solve the problem of individual variability, but not that of category-specific rate effects. Therefore, a safer strategy is to choose contrasts for which we have a clear idea of the segmental consequences of the prosodic profile and that are minimally affected by them. In this quest, VOT may not be a very good candidate.

#### 4. Small sample sizes and individual variation

A final limitation to the interpretation of those mixed results is that they have used relatively small talker samples. Thus, Malsheen (1980) includes 2 speakers in each of 3 age groups; 3 caregivers' speech was studied in Baran *et al.* (1977) and Sundberg (2001); and the sample size was 6 in Sundberg and Lacerda (1999) and Englund (2005)/Englund and Behne (2006). While it is understandable that small sample sets arise due to the fact that acoustic analyses are time- and resource-consuming, this does entail a loss of power, particularly if there is some individual variation.

#### B. Present study: Goals and predictions

In sum, the evidence in favor of consonantal categories in IDS being enhanced is unclear, but these conflicting re-

sults could be explained by the empirical and conceptual considerations raised above. In this paper, evidence was sought by trying to incorporate those considerations. First, the perceptually-based definition of phonetic enhancement was adopted, which proposes that enhancement occurs along a perceptually relevant dimension which is used to discriminate neighboring sounds. This definition was preferred in order to ensure that the dimensions to be considered were important to the speakers, so that speakers are more likely to enhance the contrasts if that is their goal.

Second, the corpus contained speech to infants at two age ranges (4–6 months, and 12–14 months), in order to detect age-related changes. These 2 ages were chosen because infants are at very different stages of linguistic development. While the younger group is just beginning to babble and has an almost non-existent receptive vocabulary (likely limited to their own names and those of their family members), the older group is already at a stage where they can understand a larger number of words and even produce a few. Each group was represented by at least 20 speakers, in order to have the statistical power to control for a number of dimensions within the same sample even in the presence of individual variation. Finally, a fricative sibilant place of articulation was chosen (/s/ and /ʃ/, as in "sock" vs. "shock"), since these are more impervious to the prosodic characteristics of IDS, as detailed below.

Sibilants are produced by forming a narrow constriction that allows a rapid flow of air through the teeth, where the noise is generated. The 'vertical' size of this constriction has a rather limited variance, such that it is not the extension of the tongue movement which affects the acoustic implementation of the contrast (Shadle, 1991). Instead, the filtering of the noise source responds to the size of the semi-independent cavities formed by that constriction. In English, the contrast depends primarily on the place at which the constriction is made, which in /s/ is at the alveolar ridge and in /ʃ/ further back toward the palate. In addition, the acoustic profile of both sibilants can be affected by the lip shape (through rounding and/or protrusion) and the tongue shape, which affects the size of the front cavity (the space between the constriction and the lips, Shadle, 1991; Toda *et al.*, 2008), and that of the sublingual cavity (the space below the tongue and behind the lower teeth, Toda *et al.*, 2008; but see also Shadle *et al.*, 2009). Therefore, more distinct /s,ʃ/ contrasts are achieved by changing the place of constriction (Toda *et al.*, 2008), by eliminating the sublingual cavity when producing /s/ (letting the tongue make contact with the lower teeth, Perkell *et al.*, 2004b), by shaping the dorsum of the tongue into a 'dome' for /ʃ/ (Narayanan and Alwan, 2000; Shadle *et al.*, 2008; Toda *et al.*, 2008), and by protruding and/or rounding the lips in /ʃ/ (Shadle *et al.*, 2009).

With these articulatory considerations in mind, specific predictions can be made regarding the acoustic realization of /s,ʃ/ and how the contrast may be enhanced. In acoustic terms, this sibilant place of articulation contrast is cued by different distributions of energy in the friction portion (Jongman *et al.*, 2000). If the energy distribution is treated as a probability distribution (Forrest *et al.*, 1988), the prevalence of lower-frequency energy in /ʃ/ and higher-frequency

energy in /s/ can be detected as higher location of the centroid (first moment or M1) and more negative skewness (third moment or M3) in /s/ as compared to /ʃ/. Both of these acoustic correlates reflect mainly the size of the front and the sublingual cavities. In addition, it has been proposed that the formant transitions into the vowels can provide a cue into the specific tongue shape, with an increase of doming being reflected in a higher onset F2 (Toda *et al.*, 2008). As for enhancement, Maniwa *et al.* (2009) show that, in clear speech, M1 tends to be slightly higher for fricatives across all places of articulation as pitch increases, with both changes likely triggered by a general increase in articulatory tension. Alongside these across-the-board changes, the distribution of energy for /ʃ/ tends to remain the same in clear and conversational speech, whereas M1 increases and M3 decreases for /s/, suggesting a more forward constriction for the latter sound during clear speech. These changes entail a significant increase in the acoustic distance between /s/ and /ʃ/, as /s/ effectively moves away, in acoustic space, from the neighboring /ʃ/. On the other hand, at least in their sample, onset F2 was significantly higher for clear speech without interacting with place of articulation. This may suggest that talkers do not use distinct tongue dorsum shapes to enhance contrasts, and thus contrastive enhancement can be reliably captured by analyzing the frication portion alone. Furthermore, Newman *et al.* (2001) show that adult listeners are better able to identify /s,ʃ/ in the speech of talkers whose sibilants are very different along this dimension, the spectral characteristics of the frication, and less so for talkers whose sibilants overlap in that dimension. In sum, production and perception data converge to suggest that M1 is likely to be relevant to sibilant enhancement.

With respect to interference from IDS prosody, speech rate is not likely to be a problem. Fricatives in general tend to be relatively stable, such that spectral characteristics remain more or less constant throughout the frication period (Shadle *et al.*, 1992). Furthermore, phonetic enhancement is not achieved by more extreme gestures in place-of-articulation contrasts in fricatives, making fricative place enhancement more independent from speech rate (which plays a limiting role in the extremeness of articulations, Lindblom, 1963). Indeed, variation in duration as a function of speech rate appears not to affect greatly the spectral characteristics of fricatives (Baum, 2004). As for pitch, if higher pitch is the result of tenser vocal configurations or an increase in airflow, then we expect higher M1 for all sibilants, but no category-specific effects. Finally, there is little cause for concern regarding the larger pitch ranges typical of IDS, as no clear articulatory link arises between sibilant implementation and this prosodic characteristic.

In short, based on this previous research on fricative production, we can be fairly certain that phonetic enhancement should be independent from prosodic IDS characteristics, since pitch, pitch excursions, and the general slowing down of speech rate are unlikely to interact with tongue shaping and the position of the constriction. It should be noted that the only potential problem relates to lip protrusion or rounding, which would be prevented by smiling. In this case, there would be an increase in M1 for both /s/ and /ʃ/, or

possibly more for /ʃ/, as the normal lip protrusion for this consonant would be reduced. Most importantly, the specific phonetic enhancement prediction is that, if caregivers enhance phonetic contrasts in speech to their child, the acoustic implementations of /s/ and /ʃ/ will diverge, most likely by raising M1 in /s/. This increase should occur beyond any increases that are correlated with higher pitch, and should be independent from the size of the pitch excursions.

## II. EXPERIMENT: SIBILANT ENHANCEMENT IN IDS

The infant's caregiver was invited to describe and classify some toys and objects of everyday use in conversation with the infant, first, and with an experimenter later. This recording was carried out in a laboratory setting because, being essentially noise, fricatives require carefully controlled recording conditions. The resulting corpus was hand-tagged by a highly trained phonetician and submitted to detailed acoustic analyses in order to gather measures of pitch, rate of speech, and sibilant characteristics.

### A. Methods

#### 1. Participants

The final data included a homogeneous sample of 55 female caregivers, all of them native American English speakers and of comparable education levels ( $M=16.3$ , range 12–22 years of education). Thirty-two of them had infants of between 12 and 14 months of age ( $M=12.29$ , range 12.3 to 13.29 in months.days, 14 female); 2 of these caregivers had multiple infants (one set of triplets and one set of twins, respectively). The infants of the remaining 23 caregivers were between 4 and 6 months of age ( $M=5$ , range 4.10 to 5.27, 13 female); 2 of these caregivers had multiple infants (each had a set of twins). The person being recorded reported being the child's primary caregiver and spending at least 50% of the child's wake time with him/her (on average, 90% of the time in the older age group; 86% in the younger set). All caregivers in both age groups were female, and all but one were the infant's mother (one was the grandmother). All of the infants had been born at full term and were healthy.

An additional 13 dyads participated in at least part of the experiment, but their results were not included for the following reasons: equipment error during recording (4), data loss (4), infant was premature (10 weeks; 1); caregiver was male (4). Male caregivers were excluded for a number of reasons: to preserve the homogeneity of the sample, because most of them were not the infant's primary caregiver; caregiving styles may vary with caregiver's sex in ways affecting their speech and language (Fernald *et al.*, 1989; Golinkoff and Ames, 1979; Tamis-LeMonda *et al.*, 2004); and some research suggests that male and female talkers differ in the extent to which they enhance the clarity of their speech (e.g., Bradlow *et al.*, 1996, 2003; see also Maniwa *et al.*, 2009). No infants were excluded for fussing, although data for some caregivers is scarce as a result of overlap with the child's vocalization (see Tables I–III).

TABLE I. Number of analyzable tokens in each category and dyad (subset).

Dyad	Age	Sex	ADS			IDS		
			Sentences	s	f	Sentences	s	f
2582	13.32	F	39	33	34	232	162	21
2849	13.45	M	54	23	2	241	64	14
2850	12.96	F	52	23	27	229	125	29
2857	12.93	M	48	25	7	191	99	41
2868	13.13	F	30	12	2	167	73	16
2870	13.8	F	80	57	25	122	63	12
2874	12.57	M	82	47	18	146	70	33
2877	13.42	M	80	46	14	109	44	18
2883	13.39	M	78	54	18	173	88	49
2887	12.4	M	48	28	3	234	75	40
2890	12.3	F	77	24	40	215	55	22
2891	12.76	F	57	32	29	116	64	17
2893	12.5	F	83	44	20	125	71	23
2895	12.99	M	61	44	8	219	87	12
2900	13.13	M	118	52	14	155	65	21
2901	12.2	M	88	61	15	145	67	18
2906	12.66	F	89	78	62	104	26	15
2907	12.14	M	119	43	17	115	56	24
2921	12.11	M	59	40	4	128	63	7
2922	13.85	M	35	16	7	241	149	29
2925	12.89	M	55	20	11	146	49	19
2934	13.95	F	127	55	25	182	71	29
2939	13.13	M	36	24	7	149	50	25

## 2. Caregiver's speech recordings: Procedure, equipment, and acoustic analyses

The caregiver and infant were taken into a small, sound treated room. In one corner of the room, there was a table with three tubs on it. Each tub contained objects chosen to

TABLE II. Number of analyzable tokens in each category and dyad (continued).

Dyad	Age	Sex	ADS			IDS		
			Sentences	s	f	Sentences	s	f
2947	13.18	M	72	46	13	107	62	29
2951	12.7	M	167	116	30	216	96	36
2953	12.99	F	79	32	25	69	20	10
2981	13.09	M	30	26	8	97	48	11
2983	12.43	F	127	53	50	132	51	22
3018	13.32	F	88	38	30	101	44	6
3020	13.52	F	66	18	19	196	74	28
3049	12.96	F	63	30	10	66	20	6
3088	12.66	M	90	43	6	181	56	26
3044	5.03	M	102	60	9	167	57	23
3050	4.97	F	186	111	45	140	86	33
3052	4.93	M	96	44	10	137	71	17
3064	5.1	M	64	40	7	187	82	29
3065	5.1	M	77	62	21	118	47	14
3068	4.7	M	58	36	8	89	36	24
3069	4.97	F	71	24	9	157	54	11
3073	4.8	F	91	56	20	134	70	29
3082	4.8	F	59	28	16	74	29	12
3083	4.9	F	111	69	18	119	56	16
3084	4.54	M	91	56	13	185	85	26
3086	5.43	M	56	38	9	136	121	32

TABLE III. Number of analyzable tokens in each category and dyad (continued).

Dyad	Age	Sex	ADS			IDS		
			Sentences	s	f	Sentences	s	f
3087	4.87	M	26	14	1	126	47	3
3089	5.3	M	87	49	7	173	99	28
3091	4.74	M	132	89	31	146	88	30
3092	5.46	M	106	65	17	138	46	18
3093	4.61	M	62	44	7	137	89	13
3095	4.47	M	89	64	8	116	61	14
3098	5.46	F	84	44	29	133	61	28
3100	4.34	F	56	33	6	121	79	13
3101	4.61	F	23	18	4	113	48	16
3108	5.86	F	167	106	9	160	77	16
3112	5.89	F	79	63	11	112	47	17
Totals			4350	2496	915	8167	3743	1170

elicit the target sounds (/s,f/, e.g., shoes, socks, sand toys) as well as some filler items. Most of the filler items were selected to elicit /p/- or /b/-initial words and to contain the point vowels /i,a,u/, to avoid biasing the caregiver toward the sounds of interest.

The caregiver was fitted with a lavalier mic and was asked to "tell [your child] what these objects are, what they are for, and sort them into groups or categories." It was emphasized that the purpose of the experiment was to investigate how infants learn categories from their caregivers. These instructions were given to ensure that the parent produced the target sounds but without being overly conscious about the clarity in their speech, and instead focused on object categories (such as clothes, toys, etc.) Caregiver and child were then left alone for about 5 min. Then the experimenter returned and engaged the parent in conversation to ensure they produced the target sounds in an adult-directed register. Although the infant was still in the room, sentences addressed to the child were not included in the ADS sample. Extreme care was taken to make sure that caregivers were comfortable and that the conversation with the adult was as natural as possible.

The caregiver's speech was recorded with an AKG WMS40 Pro Presenter Set Flexx UHF Diversity CK55 Lavalier mic and receiver, into a Marantz Professional Solid State Recorder (PMD660ENG). It should be noted that this microphone has a completely flat frequency response between 50 and 20,000 Hz, which should allow a faithful representation of sibilants. A wireless system was chosen to maximize the caregiver's comfort, allowing him or her to move around.

All of the coding and analyses were done using Praat (Boersma and Weenik, 2005). All instances (regardless of syllable position) of /s,f/ that were free of noise and talker overlap were marked (/i,a,u/ were also tagged for a separate study). For the pitch and speech rate calculation, sentences were determined prosodically (not syntactically) and the number of syllables pronounced as well as the register were noted. Sentences where the pitch tracker was not reliable (because of talker overlap, part of the sentence being produced in glottal fry, or due to mistracking by pitch halving or

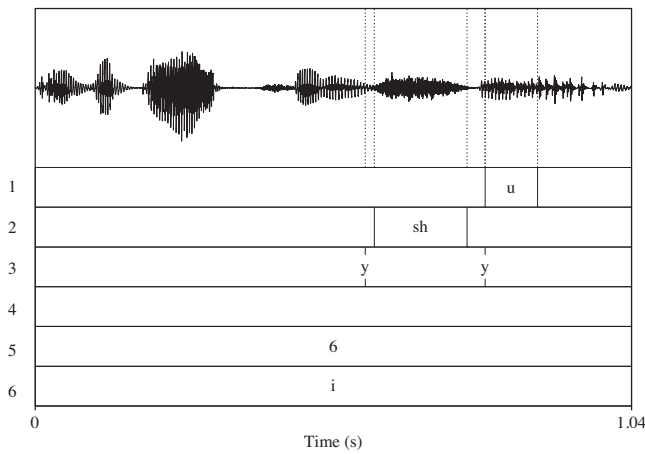


FIG. 1. Example of the coding system. The top tier was used to mark the point vowels. The edges of the frications of the sibilants were marked in the second tier, while on the third tier it was noted whether there was a vowel preceding and/or following the sibilant, in order to extract formants only in the case when there was a vowel. The fourth tier was used for comments. In the fifth tier, the number of syllables in the sentence was noted, while the sixth encoded the register (in this case, “i,” for IDS).

doubling) were excluded from the pitch calculations, and those where syllable counts were uncertain, due to overlap between different speakers or the caregiver mumbling, were excluded from the rate of speech calculation. Sibilants shorter than 40 ms were also excluded, as the window of analysis used would then include neighboring sounds. A sample of this coding is given in Fig. 1.

The tagged sound files were then analyzed using a custom-written script, yielding a number of measurements. With respect to place of articulation contrast in the sibilants, the dimension of interest here was the first spectral moment (M1) of the energy distribution from a 40 ms Hamming window taken from the midpoint of the frication, following the methods in Jongman *et al.* (2000). This dimension provides a robust separation of /s/ and /ʃ/ in English, as evident in Fig. 2. In addition, M1 has been found to reflect changes in clear speech and appeared more reliable with the analyses methods used here (over e.g., peak location). At the sentence level, the following correlates were considered: average pitch (in Hertz and ERB; both showed the same pattern; linear measures are reported here for ease of comparison with previous research); pitch excursions (maximum-minimum within the sentence; converted into semitones, as in other research, e.g., Fernald *et al.*, 1989), and rate of speech (number of syllables/duration in seconds).

## B. Results and analyses

The number of sentences and sibilants tagged in each caregiver’s samples is given in Tables I–III, split into three groups for ease of inspection. In order to describe the global acoustics of this sample, measurements for all tagged sentences were submitted to Repeated Measures Analyses of Variance with Register (IDS, ADS) as within- and Age Group as between-subjects factors. As shown in Fig. 3, when talking to their child, caregivers spoke more slowly [4.3 compared to 4.7 syllables/second; main effect of Register,  $F(1, 53)=21.27$ ,  $p<0.001$ ; no interaction with Group], with

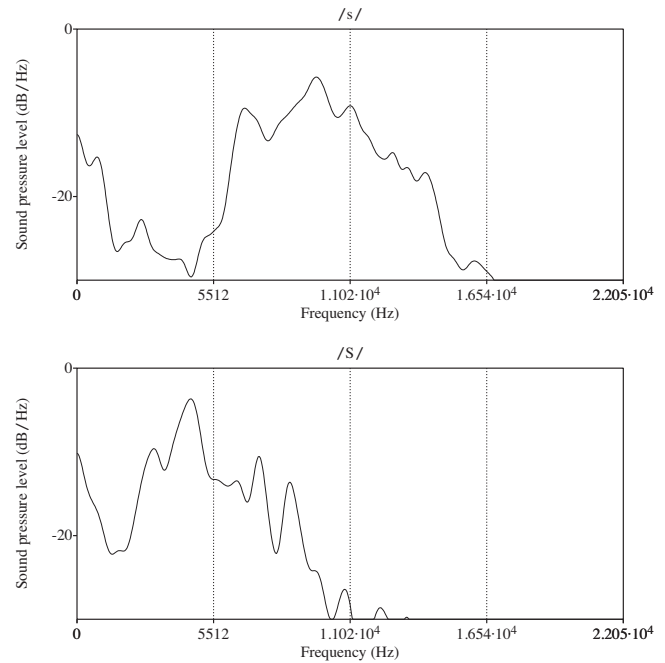


FIG. 2. The top panel is a spectrum (distribution of intensity over frequency) for /s/ and the bottom one a spectrum for /ʃ/. The peak location for /ʃ/ is usually much lower than that for /s/, and /ʃ/ has more evenly distributed energy. As a result, M1 is higher in /s/ than /ʃ/.

a higher average pitch [259 compared to 210 Hz; main effect of Register,  $F(1, 53)=189.36$ ,  $p<0.001$ ; interaction with Group,  $F(1, 53)=5.05$ ,  $p<0.03$ , due to higher average pitch in speech to older infants, 265 compared to 249 Hz] and produced larger pitch excursions within the sentences [13.55 compared to 10.8 semitones; main effect of Register,  $F(1, 53)=23.89$ ,  $p<0.001$ ; no interaction with Group].

In order to test the predictions for the sibilants, the measurements for all the sibilants were submitted to a Repeated Measures Analysis of Variance with M1 as the dependent measure, Label and Register as within-measures and Group as between-measure. Results suggested that the phonetic implementation of /s,ʃ/ differs across the two infant groups and registers [Label\*Register\*Group  $F(1, 53)=6.89$ ,  $p<0.02$ ]. Separate ANOVAs within each age group were used to investigate this pattern. In the analyses for the speech to 4–6-month-olds, there was no significant interaction Label\*Register [ $F_s<1$ ]. In contrast, in the speech to the older infants, a Label\*Register interaction [ $F(1, 31)=4.3$ ,  $p<0.05$ ] revealed that M1 was higher in IDS for /s/, but not for /ʃ/ [7903 compared to 7626 Hz Welch Two Sample  $t(2673)=5.33$   $p<0.001$ ; for /ʃ/ 5102 compared to 5035 Hz, Welch Two Sample  $t(1295)=0.06$ ].

As reported on Tables I–III, very few tokens were usable for some of the caregivers, usually due to overlap of the child vocalizing or fussing, a problem that affected the ADS portion and the older age group to a greater extent. Analyses excluding caregivers for whom fewer than 5 tokens of either sibilant could be used in either register (namely, 2849, 2868, 2887, 2921, 3087, and 3101) shows the same pattern of results, although weaker due to the reduced degrees of freedom

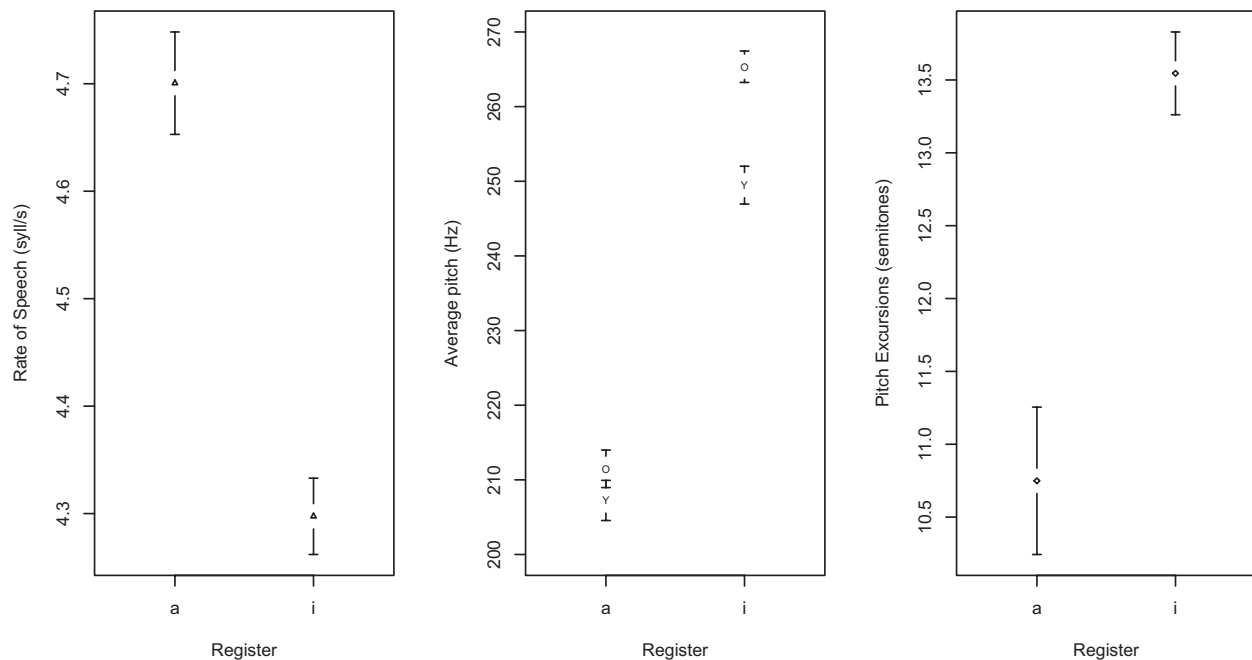


FIG. 3. Averages and 95% confidence intervals for Rate of speech, Average pitch, and Pitch excursions by register. Given that for Rate and Excursions there is no interaction with the child's age, data are collapsed across the groups. In contrast, caregivers of the 12–14 month-olds (“o”) addressed their child with an average pitch that was significantly higher to that used by caregivers of the younger infants (“y”).

[Label\*Register\*Group,  $F(1,47)=7.21$ ,  $p<0.01$ ; Label\*Register within the 12- to 14-month-olds,  $F(1,27)=3.87$ ,  $p=0.06$ ].

This higher M1 only for /s/ fits the predictions based on the clear speech literature. Nonetheless, in order to rule out that this difference in the acoustic realization of /s/ is a side effect of the lower speech rate or pitch characteristics of IDS, several linear regression models predicting M1 were fitted to the data. In the first one, the variable Label (/s,/f/) was entered, as well as the measured acoustic characteristics of the sentences in which each /s,/f/ token occurred: average pitch, pitch excursion, and rate of speech. Notice that these data should capture local effects, as they take into account the pitch of the same sentence from which the sibilant measures were taken, rather than an average over the whole sentence set for that talker. Additionally, notice that the degrees of freedom for these models do not equal the total number of sibilants, as data points where either pitch or rate of speech is missing were excluded. This model was highly significant, and so were the estimates for Label, Rate of speech, and Average pitch, although Excursion was not.

Several factors were added in a second model, including Register and Age Group as well as interactions involving every combination of Label, Group, and Register. This also resulted in a significant model, where all the predictors (except for Excursions), and most importantly the interaction Label\*Group\*Register, were significant. This second model constituted a significant improvement over the first, as confirmed through an ANOVA [ $F(6,7549)=13.49$ ,  $p<0.001$ ], which lends support to the hypothesis that not all of the changes in /s,/f/ are explained by changes in local acoustic characteristics, but rather that there were some register-specific effects, and that these were found in speech to older infants. The full results for these two models are reported in

Table IV. In a third model, an attempt was made to control for the fact that some caregivers contributed more data than others, and could thus affect results to a larger extent. To this end, Dyad was added as well as its interactions with Label, Group, and Register. In accordance with the previous conclusions, the Label\*Group\*Register interaction remained significant [ $t=2.2$ ,  $p<0.05$ ].

As in previous work (Baran *et al.*, 1977), there was some variation in the fit of this general pattern to individual

TABLE IV. Summary of linear regression models with M1 as the outcome, and Average Pitch, pitch Excursion, Rate of Speech, phonetic Label, Register, and age Group as predictors. Significance at the 0.05 level is signaled with \*; 0.01 with \*\*; 0.001 with \*\*\*.

	Estimate	SE	<i>t</i>
Model 1: $R^2=.518$ , $F(4,7555)=2026$ , $p<0.001$			
Intercept	7538.367	85.01	88.674***
Label (f)	-3285.74	36.632	-89.695***
Rate of speech	-66.153	11.120	-5.949***
Average pitch	2.173	0.26	8.359***
Excursion	-1.697	1.32	-1.289
Model 2: $R^2=.523$ , $F(10,7549)=826.5$ , $p<0.001$			
Intercept	7494.5884	89.442	83.79***
Label (f)	-3120.836	73.036	-42.73***
Group (4–6 months)	123.252	59.006	2.09*
Register (IDS)	125.376	53.392	2.35*
Rate of speech	-69.114	11.169	-6.19***
Average pitch	2.398	0.278	8.62***
Excursion	-2.011	1.311	-1.53
Label*Group	-242.279	116.393	-2.08*
Label*Register	-324.913	95.586	-3.4***
Group*Register	-492.31	75.924	-6.48***
Label*Group*Register	504.924	151.817	3.33***

data. Among the caregivers of older infants, 19 showed an increase in mean M1 for /s/, 6 a decrease, and the remainder little change; in contrast, only 6 showed a marked change with respect to the /ʃ/ category (3 in each direction). Among the younger cohort, 10 showed some decrease in mean M1 for /s/, but for most the distribution of each consonant overlapped across registers.

### III. DISCUSSION

The idea that IDS may be shaped by the implicit goal of facilitating language acquisition is as old as the concept of ‘motherese’. In his classical paper, Charles Ferguson hypothesized that the apparent universality of certain IDS features arose from the fact that this register served the same functions in all languages and cultures, including simplifying and clarifying the linguistic material (Ferguson, 1964). More recently, training studies have shown that infants’ perception of sound categories can be shaped by the distribution of acoustic cues that they are exposed to (Maye *et al.*, 2008, 2002). These findings revolutionized the field, insofar as they lent support to the possibility that phonological learning could be driven by bottom-up learning of categories from the acoustic distributions in the input (McMurray *et al.*, 2009). As a result, the question of whether the acoustic implementation of sound categories is enhanced in IDS is not just of descriptive interest, but becomes a key element in the puzzle of how natural language acquisition comes about.

Extant evidence on consonantal enhancement in IDS was not conclusive, but this may have been due to factors such as interference from prosodic characteristics and reduced power due to small sample sizes. The present study has attempted to deal with these roadblocks, and is thus able to provide more conclusive evidence for phonetic enhancement in consonantal implementation in IDS. The speech of 32 caregivers of 12- to 14-month-olds, and 23 4- to 6-month-olds was analyzed for the acoustic implementation of the sibilant place of articulation contrast between /s/ and /ʃ/. This contrast was chosen to minimize the effects of prosodic characteristics, but prosodic IDS characteristics (speech rate and pitch) were also measured, in order to control for their possible effect on the sibilants’ acoustic characteristics. The prediction, based on independent studies of clear speech, was that M1 would be higher for /s/ in IDS than in ADS. This was the case in speech to 12- to 14-month-olds. Variation in M1 was not completely accounted for by an increase in average pitch (which would be the case if they are both due to a third factor, such as increased vocal effort or airflow, or a reduction of the length of the vocal tract, e.g., due to smiling), or by a reduction in the speech rate (in which case more extreme articulations are invited). The following subsections detail the theoretical and empirical implications of these results.

#### A. Factors affecting phonetic enhancement

The fact that enhancement occurs only in speech to the older infants is compatible with both hypotheses that propose that phonetic enhancement is mediated by the child’s age or linguistic development. In other words, this effect could arise

because caregivers enhance phonetic contrasts in speech to children who are just beginning to talk (Malsheen, 1980), or because they are following the more general timeline and adjusting only the consonants (Sundberg, 1998). In fact, a third possibility is that caregivers enhance consonants when they become aware of the fact that their child is beginning to learn words, as consonants appear to be particularly important to word learners. As the linguistic development of the participants in this study was not measured, the present data cannot adjudicate between these two interpretations. Future research could focus on this question by attempting to relate consonantal enhancement with the child’s productive vocabulary (along Malsheen’s hypothesis), her comprehensive vocabulary (following the word-learning hypothesis), and the child’s age (in line with Sundberg’s model).

On the other hand, these data do suggest that variation in pitch or speech rate as a function of the child’s age is the *not* the sole factor driving enhancement. Since the main prosodic characteristics of IDS were controlled for, the present results provide stronger support for the independent existence of phonetic enhancement, rather than this occurring solely as a side effect of the IDS characteristics that vary with the child’s age.

Present results also provide indirect support for the perceptually-based view of phonetic enhancement proposed in the Introduction. That is, results fit well with the prediction that the direction of change may be predicted from the existence of highly confusable sounds in the inventory. Specifically, the direction of enhancement and its limitation to the realization of /s/ is consistent with previous results from clear speech, since both in Maniwa *et al.* (2009) and here the M1 for /s/ was significantly higher, which could be interpreted as an attempt to distinguish /s/ from its acoustic neighbor, /ʃ/.

Furthermore, it appears that the consideration that prosodic IDS characteristics may affect consonantal realization was not overly cautious. As expected from their codependence on lip shape and vocal tension, M1 and average pitch covaried to some extent, a fact that was predicted given previous research on the articulatory bases of prosodic and consonantal characteristics. Notice, furthermore, that rate of speech did bear some relationship with M1, in spite of the fact that there is no simple articulatory link between these 2 variables and there is no reported relationship between them in the phonetic literature. Such a result should serve as a reminder that the articulatory configurations that give rise to speech are peculiarly complex, and future studies of IDS characteristics should continue to incorporate direct controls for possible confounding effects.

#### B. Individual variation and the input to acquisition

The large pool of talkers recruited for the present study also afforded an invaluable perspective on individual variation in caregivers’ speech, showing a great deal of individual variation. These data could help explain previous null results as owing to limited sample sizes. Of the 32 caregivers, only 19 clearly followed the general pattern, and 6 displayed the opposite one. If these frequencies are representative of the



behavior of the population, it is unsurprising that about half of the studies in voicing vouch for enhancement and the other half do not. However, it is possible that the extent of individual variation was greater for the category under study, as /s/ appears to be implemented in diverse ways across speakers (Gordon *et al.*, 2002; Newman *et al.*, 2001).

Aside from this methodological consideration, a deeper conceptual lesson is to be drawn. As Baran *et al.* (1977) pointed out over 30 years ago, general patterns should always be checked at the individual level. If we are to uphold the view that language development is affected by the input infants hear, then assessing individual variation is not only a means to understand IDS as a register, but also responds to the theoretical need of a window into the “real” input infants hear. That is, when the phonetic results of specific behaviors are averaged, the resulting picture may not be an accurate depiction of the input specific infants are provided with. Imagine the following scenario: a subset of the speakers enhances a contrast, while others do not modify them at all. As a result, averaged results would show effects of phonetic enhancement, leading to the (incorrect) conclusion that *all* infants are presented with outstanding phonetic categories. As is evident from this and other research, this is not the case, and the impact of this individual variation on infants’ language acquisition only begins to be investigated. For example, Liu *et al.* (2003) report that the size of a mother’s vowel space predicts, to some extent, the child’s performance in sound discrimination. Although this finding is open to alternative interpretations (for example, talkers with finer discrimination abilities tend to produce sounds more distinctly, e.g., Perkell *et al.*, 2004b, 2004a and hearing abilities could be inheritable), it also suggests that maternal variation in phonetic implementation could contribute to the well-documented inter-individual variation in children’s language acquisition.

### C. Limitations of current research on phonetic enhancement

It is equally important to bear in mind an important limitation of current research on phonetic enhancement in IDS, including the present study. Measuring phonetic enhancement on the basis of separation in acoustic/auditory space makes two unwarranted assumptions: first, that this enhancement is perceivable; and second, that it is conducive to learning. While we rely on findings from the clear speech literature to assume that larger vowel spaces present learners with more discriminable categories, this should be directly tested with IDS samples, as many other acoustic dimensions differ between IDS and clear speech. To take a simple example, Trainor and Desjardins (2002) find that vowels produced with high average pitch are harder for infants to discriminate. Therefore, it could be the case that the disadvantageous spectral blurring caused by high pitch outweighs the benefits of a larger vowel space. Similarly, a wider separation between voiced and voiceless stops on the basis of VOT could have a negligible effect on perception, as the discrimination between short-and long-lag VOT is categorical and independent from experience (Eimas *et al.*, 1971; Kuhl and Miller, 1978). Regarding the positive impact on learning, there is

some research using automatic speech recognition systems. For instance, de Boer and Kuhl (2003) report that an automatic learner showed improved performance when trained on IDS. In contrast, Kirchoff and Schimmel (2005) argue that IDS provides a much noisier sample for the learner, and thus the learning curve is shallower. Moreover, even if better or worse performance were found with these automatic learners, this would not entail an advantage or a disadvantage for human learners, as there are other factors to consider. Thus, an automatic learner may quickly pick up categories with small variances, whereas research on both infant and adult learners suggests that, contrary to expectation, *increased* variability may play an important role (Jamieson and Morosan, 1989; Rost and McMurray, 2009).

### D. Conclusion

In this study, phonetic enhancement in IDS was documented, even after controlling for other acoustic characteristics, such as speech rate and average pitch. Moreover, this enhancement was specific to speech addressed to 12–14-month-olds, but not 4–6-month-olds, suggesting that this phenomenon is modulated by characteristics of the addressee. Results also lend indirect support to the hypothesis that individual variation could underlie previous conflicting results, although meta-analyses of currently published data may provide further confirmation. Finally, this paper demonstrates the importance of taking into account perception and production in the hypotheses made regarding the enhancement of specific categories. In particular, enhancement is likely inventory-dependent, with the dimensions along which category implementation varies being affected by the presence of sounds that are close in perceptual space. Taken together, these results provide a moderate degree of support to phonetic enhancement in IDS consonants, but the impact of this enhancement on perception and learning awaits confirmation.

### ACKNOWLEDGMENTS

Data collection was carried out at Purdue University, for which I thank the families and staff associated to the Purdue Baby Labs, the funding from a Purdue Research Foundation Fellowship and the Linguistics Program, and from NICHD Grant No. R03 HD046463-0 to Amanda Seidl. I also gratefully acknowledge the support of the Ecole de Neurosciences de Paris and the Fondation Fyssen during the writing stage. This work has been greatly improved by helpful comments from 2 reviewers, Mary Beckman, Alex Francis, Lisa Goffman, Amanda Seidl, and Martine Toda, and the audiences at ASHA November 2008 (Brown-Forman Corporation grant), ASA November 2008 (Purdue Graduate Student Government travel grant), and ICIS 2010.

- Baran, J. A., Laufer, M. Z., and Daniloff, R. (1977). “Phonological contrastivity in conversation: A comparative study of voice onset time,” *J. Phonetics* 5, 339–350.
- Baum, S. R. (1996). “Fricative production in aphasia: Effects of speaking rate,” *Brain Lang* 52, 328–341.
- Biersack, S., Kempe, V., and Knapton, L. (2005). “Fine-tuning speech registers: A comparison of the prosodic features of child-directed and foreigner-directed speech,” in *Interspeech-2005*, Lisbon, Portugal, pp.

- Boersma, P., and Weenik, D. (2005). "Praat: Doing phonetics by computer (version 5.0.09)," <http://www.praat.org/> (Last viewed 05/26/2007).
- Bond, Z. S., and Moore, T. J. (1994). "A note on the acoustic-phonetic characteristics of inadvertently clear speech," *Speech Commun.* **14**, 325–337.
- Bradlow, A. R., Kraus, N., and Hayes, E. (2003). "Speaking clearly for children with learning disabilities: Sentence perception in noise," *J. Speech Lang. Hear. Res.* **46**, 80–97.
- Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (1996). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," *Speech Commun.* **20**, 255–272.
- Burnham, D., Kitamura, C., and Vollmer-Conna, U. (2002). "What's new, pussycat? On talking to babies and animals," *Science* **296**, 1435.
- Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., and van Ooijen, B. (2000). "Constraints of vowels and consonants on lexical selection," *Mem. Cognit.* **28**, 746–755.
- de Boer, B., and Kuhl, P. K. (2003). "Investigating the role of infant-directed speech with a computer model," *ARLO* **4**, 129–134.
- Eimas, P., Siqueland, E., Jusczyk, P. W., and Vigorito, J. (1971). "Speech perception in infants," *Science* **171**, 303–306.
- Englund, K. (2005). "Voice onset time in infant directed speech over the first six months," *First Lang.* **25**, 219–234.
- Englund, K., and Behne, D. (2006). "Changes in infant directed speech in the first six months," *Infant Child Dev.* **15**, 139–160.
- Ferguson, C. A. (1964). "Baby talk in six languages," *Am. Anthropol.* **66**, 103–114.
- Fernald, A. (1989). "Intonation and communicative intent in mothers speech to infants: Is the melody the message?," *Child Dev.* **60**, 1497–1510.
- Fernald, A. (1992). "Human maternal vocalizations as biologically relevant signals: An evolutionary perspective," in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, edited by J. H. Barkow, L. Cosmides, and J. Tooby (Oxford University Press, New York, pp. 391–428).
- Fernald, A. (2000). "Speech to infants as hyperspeech: Knowledge-driven processes in early word recognitions," *Phonetica* **57**, 242–254.
- Fernald, A., and Simon, T. (1984). "Expanded intonation contours in mothers' speech to newborns," *Dev. Psychol.* **20**, 104–113.
- Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., Boysson-Bardies, B., and Fukui, I. (1989). "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," *J. Child Lang.* **16**, 977–1001.
- 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**, 115–123.
- Golinkoff, R. M., and Ames, G. J. (1979). "A comparison of fathers' and mothers' speech with their young children," *Child Dev.* **50**, 28–32.
- Gordon, M., Barthmaier, P., and Sands, K. (2002). "A cross-linguistic acoustic study of voiceless fricatives," *J. Int. Phonetic Assoc.* **32**, 141–174.
- Grieser, D. L., and Kuhl, P. K. (1988). "Maternal speech to infants in a tonal language: Support for universal prosodic features of motherese," *Dev. Psychol.* **24**, 14–20.
- Jamieson, D. G., and Morosan, D. E. (1989). "Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques," *Can. J. Psychol.* **43**, 88–96.
- Jongman, A., Wayland, R., and Wong, S. (2000). "Acoustic characteristics of English fricatives," *J. Acoust. Soc. Am.* **108**, 1252–1263.
- Kang, K., and Guion, S. G. (2008). "Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies," *J. Acoust. Soc. Am.* **124**, 3909–3917.
- Kaplan, P. S., Fox, K., Scheuneman, D., and Jenkins, L. (1991). "Cross-modal facilitation of infant visual fixation: Temporal and intensity effects," *Infant Behav. Dev.* **14**, 83–109.
- Kessinger, R. H., and Blumstein, S. E. (1997). "Effects of speaking rate on voice-onset time in Thai, French, and English," *J. Phonetics* **25**, 143–168.
- Kirchhoff, K., and Schimmel, S. (2005). "Statistical properties of infant-directed versus adult-directed speech: Insights from speech recognition," *J. Acoust. Soc. Am.* **117**, 2238–2246.
- Kitamura, C., and Burnham, D. (2003). "Pitch and communicative intent in mothers speech: Adjustments for age and sex in the first year," *Infancy* **4**, 85–110.
- Kitamura, C., Thanavishuth, C., Burnham, D., and Luksaneeyanawin, S. (2001). "Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language," *Infant Behav. Dev.* **24**, 372–392.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., and Lacerda, F. (1997). "Cross-language analysis of phonetic units in language addressed to infants," *Science* **277**, 684–686.
- Kuhl, P. K., and Miller, J. (1978). "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," *J. Acoust. Soc. Am.* **63**, 905–917.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science* **255**, 606–608.
- Lieven, E. V. M. (1994). "Crosslinguistic and crosscultural aspects of language addressed to children," in *Input and Interaction in Language Acquisition*, edited by C. Gallaway and B. J. Richards (Cambridge University Press, Cambridge), pp. 56–73.
- Lindblom, B. (1963). "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* **35**, 1773–1781.
- Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H and H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hard-Castle and A. Marchal (Kluwer, Dordrecht), pp. 403–439.
- Liu, H.-M., Kuhl, P. K., and Tsao, F.-M. (2003). "An association between mothers' speech clarity and infants' speech discrimination skills," *Dev. Sci.* **6**, F1–F10.
- Liu, H.-M., Tsao, F.-M., and Kuhl, P. K. (2007). "Acoustic analysis of lexical tone in Mandarin infant-directed speech," *Dev. Psychol.* **43**, 912–917.
- Liu, H.-M., Tsao, F.-M., and Kuhl, P. K. (2009). "Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: A longitudinal study," *J. Child Lang.* **36**, 909–922.
- Malsheen, B. J. (1980). "Two hypotheses for phonetic clarification in the speech of mothers to children," in *Child Phonology*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, San Diego, CA), Vol. **2**, pp. 173–184.
- Maniwa, K., Jongman, A., and Wade, T. (2009). "Acoustic characteristics of clearly spoken English fricatives," *J. Acoust. Soc. Am.* **125**, 3962–3973.
- Maye, J., Weiss, D., and Aslin, R. (2008). "Statistical phonetic learning in infants: Facilitation and feature generalization," *Dev. Sci.* **11**, 122–134.
- Maye, J., Werker, J. F., and Gerken, L. (2002). "Infant sensitivity to distributional information can effect phonetic discrimination," *Cognition* **82**, B101–B111.
- McMurray, B., Aslin, R. N., and Toscano, J. C. (2009). "Statistical learning of phonetic categories: Insights from a computational approach," *Dev. Sci.* **12**, 369–378.
- Narayanan, S., and Alwan, A. (2000). "Noise source models for fricative consonants," *IEEE Trans. Speech Audio Process.* **8**, 328–344.
- Nazzi, T. (2005). "Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels," *Cognition* **98**, 13–30.
- New, B., Araujo, V., and Nazzi, T. (2008). "Differential processing of consonants and vowels in lexical access through reading," *Psychol. Sci.* **19**, 1223–1227.
- Newman, R. S., Clouse, S. A., and Burnham, J. (2001). "The perceptual consequences of acoustic variability in fricative production within and across talkers," *J. Acoust. Soc. Am.* **109**, 1181–1196.
- Niwano, K., and Sugai, K. (2003). "Pitch characteristics of speech during mother-infant and father-infant vocal interactions," *Jpn. J. Spec. Educ.* **40**, 663–674.
- Papoušek, M., and Hwang, S.-F. C. (1991). "Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction," *Appl. Psycholinguist.* **12**, 481–504.
- Payton, K. L., Uchanski, R. M., and Braid, L. D. (1994). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **95**, 1581–1592.
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (2004a). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," *J. Acoust. Soc. Am.* **116**, 2338–2344.
- Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F. H. (2004b). "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**, 1259–1269.
- Picheny, M. A., Durlach, N. I., and Braid, L. (1985). "Speaking clearly for

- the hard of hearing I: Intelligibility differences between clear and conversational speech," *J. Speech Hear. Res.* **28**, 96–103.
- Pye, C. (1986). "Quiché Mayan speech to children," *J. Child Lang.* **13**, 85–100.
- Ratner, N.B. (1984). "Patterns of vowel modification in mother-child speech," *J. Child Lang.* **11**, 557–578.
- Rost, G. C., and McMurray, B. (2009). "Speaker variability augments phonological processing in early word learning," *Dev. Sci.* **12**, 339–349.
- Shadle, C. H. (1991). "The effect of geometry on source mechanisms of fricative consonants," *J. Phonetics* **19**, 409–424.
- Shadle, C. H., Dobelke, C. U., and Scully, C. (1992). "Spectral analysis of fricatives in vowel context," *J. Phys. IV* **2**, C1-295–C1-298.
- Shadle, C. H., Proctor, M. I., Berezina, M., and Iskarous, K. (2009). "Revisiting the role of the sublingual cavity in the /s/-/ʃ/ distinction," in 157th Meeting of the ASA, Portland, OR.
- Shadle, C. H., Proctor, M. I., and Iskarous, K. (2008). "An MRI study of the effect of vowel context on English fricatives," *Proc. Meet. Acoust.* **19**, 5099–5104.
- Shute, B., and Wheldall, K. (1999). "Fundamental frequency and temporal modifications in the speech of British fathers to their children," *Educ. Psychol.* **19**, 221–233.
- Stern, D. N., Spieker, S., Barnett, R. K., and MacKain, K. (1983). "The prosody of maternal speech: Infant age and context related changes," *J. Child Lang.* **10**, 1–15.
- Stern, D. N., Spieker, S., and MacKain, K. (1982). "Intonation contours as signals in maternal speech to prelinguistic infants," *Dev. Psychol.* **18**, 727–735.
- Sundberg, U. (1998). "Mother tongue-phonetic aspects of infant-directed speech," Ph.D. thesis, Stockholm University, Stockholm.
- Sundberg, U. (2001). "Consonant specification in infant-directed speech. Some preliminary results from a study of voice onset time in speech to one-year-olds," *Linguistics Working Papers, Lund University* **49**, 148–151.
- Sundberg, U., and Lacerda, F. (1999). "Voice onset time in speech to infants and adults," *Phonetica* **56**, 186–199.
- Tamis-LeMonda, C. S., Shannon, J. D., Cabrera, N. J., and Lamb, M. E. (2004). "Fathers and mothers at play with their 2- and 3-year-olds: Contributions to language and cognitive development," *Child Dev.* **75**, 1806–1820.
- Toda, M., Maeda, S., Aron, M., and Berger, M.-O. (2008). "Modeling subject-specific formant transition patterns in /aʃa/ sequences," in 8th International Seminar on Speech Production–ISSP'08, Strasbourg, France, pp. 357–360.
- Trainor, L. J., and Desjardins, R. N. (2002). "Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels," *Psychon. Bull. Rev.* **9**, 335–340.
- Uther, M., Knoll, M. A., and Burnham, D. (2007). "Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech," *Speech Commun.* **49**, 2–7.
- Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (2007). "Infant-directed speech supports phonetic category learning in English and Japanese," *Cognition* **103**, 147–162.
- Werker, J. F., and Tees, R. (1984). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," *Infant Behav. Dev.* **7**, 49–63.