Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category

Alejandrina Cristià

Laboratoire de Sciences Cognitives et Psycholinguistique,

Ecole des Hautes Etudes en Sciences Sociales,

Département d'Etudes Cognitives-Ecole Normale Supérieure,

Centre National de la Recherche Scientique,

Paris 75005,

France[a)]

(Dated: February 15, 2011)

Abstract

Within the debate on the mechanisms underlying infant perceptual acquisition, one hypothesis proposes that infants' perception is directly affected by the acoustic implementation of sound categories in the speech they hear. In consonance with this view, the present study shows that individual variation in fine-grained, subphonemic aspects of the acoustic realization of /s/ in caregivers' speech predicts infants' discrimination of this sound from the highly similar /ʃ/, suggesting that learning based on acoustic cue distributions may indeed drive natural phonological acquisition.

I. INTRODUCTION

Infants can discriminate many speech sounds in the absence of prior experience at an early age, but by the beginning of the second year of life they begin to home in on those present in their ambient language(s) (see Jusczyk, 1997, for a review). This discovery raised several puzzling questions: How can infants discriminate contrasts that they had never heard before? How can they determine which contrasts to continue to attend to?

A traditional view proposes that infants' initial perception is guided by Universal Grammar, which provides them with a finite, universal set of sounds (or features) that may be contrastive in language. Therefore, this hypothesis easily accounts for discrimination of contrasts in the absence of exposure, such as what happens with non-native contrasts. As for changes later in development, early phonological acquisition is viewed as a selection process, a set of binary choices made based on the presence or absence of sounds. That is, the only important factor is which sounds, among those specified by Universal Grammar, are present in the infant's ambient language (for recent theoretical support for this view, see Hale *et al.*, 2006; Hale and Reiss, 2003; for a historical review of this and other theories of perceptual tuning, see Werker, 1994).

An alternative hypothesis postulates instead that infants start out with certain auditory-perceptual sensitivities, which allow them to easily distinguish certain pairs of sounds, and that exposure to the acoustic implementations of sounds in the input shapes these prior sensitivities (Aslin and Pisoni, 1980). By emphasizing acoustic salience and the role of the input in shaping infants' perception, this view makes somewhat different predictions. If learners' perception is partly dependent on how robustly a contrast is implemented acoustically, one would expect less acoustically salient contrasts to be harder to discriminate, and to be acquired later. Indeed, Filipino-learning 4- to 6-month-olds cannot discriminate /n/ and /ŋ/, while they succeed with the acoustically robust /n-m/ (Narayan *et al.*, 2009). Similarly, English learners struggle with the native contrast /d-ð/ even at 10-12 months (Polka

---

a)Electronic address: `alecristia@gmail.com`

et al., 2001). With regard to the input, laboratory training studies suggest that infants' discrimination is guided not by the presence of specific sounds but by the distributions of acoustic cues caused by the presence of those categories (Cristià *et al.*, 2011, in press; Maye *et al.*, 2008, 2002; Yoshida *et al.*, 2010).

In short, both the traditional and the alternative hypotheses succeed at accounting for infants' initial sensitivities and their tuning, but only the latter can also accommodate the importance of acoustic realization found in laboratory training studies. However, one may argue that the perceptual effects found in such studies are not representative of natural language learning. One way to assess the role of the input in phonological acquisition is to make use of individual variation in the way of pronouncing a given sound, such that clearer implementations should lead to better learning. That infants could be affected by subphonemic variation is credible because infants are sensitive to within-category variation (particularly in vowels: Kuhl *et al.*, 1992; Polka and Werker, 1994; but also in consonants: McMurray and Aslin, 2005; Miller and Eimas, 1996), and can use it to learn categories in laboratory training studies (Cristià *et al.*, 2011, in press; Maye *et al.*, 2008, 2002; Yoshida *et al.*, 2010). In the present study, the prediction that fine-grained, subphonemic characteristics in the input may affect infants' sound categories in natural language acquisition is tested by examining the impact on infant perception of individual variation in the realization of one specific sound in the speech of their caregivers.

Two infant age groups were tested. The younger ones were between 4 and 6 months of age (an age at which no native-language attunement has been reported for consonants), while the older ones were between 12 and 14 months (by which age they should exhibit language-specific perception; Werker and Tees, 1984). Each infants' discrimination of /s,ʃ/ (as in "sock, shock") was measured using Visual Habituation (Werker *et al.*, 1998), a procedure in which the infant is habituated to one category, and subsequently presented with novel tokens of the habituated category, or a new category. Infants' ability to discriminate between the categories is assessed through the recovery of attention for the new category as compared to that due to sheer novelty of the tokens. After this test, the infant's primary caregiver

was recorded in interaction with the child, and the acoustic characteristics of the caregiver's sibilants were measured. It was predicted that infants whose caregivers produced more extreme /s/ would be better able to discriminate /s/ from its phonemic neighbor /ʃ/, and that this effect may be stronger in the older age group, who have been exposed to their caregivers for a longer amount of time. The specifics of the design and the reasoning behind it are laid out in section II.A below.

II. EXPERIMENT

A. Design

1. Choice of procedure

A single, fixed test order is thought best in order to assess individual variation in infants, since then differences in performance are not confounded with differences in the way the test was carried out (Colombo and Fagan, 1990). Given that it was not feasible to do more than one habituation-dishabituation test, there was only one background or habituated category, and only one target or novel category. This kind of test yields two measurements: the number of trials that the infant required to meet the habituation criterion, and an index of dishabituation during test. In individual variation data, the trials to habituation measure correlates with cognitive skills (McCall and Carriger, 1993). Therefore, this measure should not be considered a clear index for the robustness of the category used during habituation. Furthermore, since by test all infants had habituated, this would level out individual differences due to the background category. On the basis of these considerations, I hypothesized that only the dishabituation index would be meaningful, and that it would depend primarily on the caregivers' implementation of the target/dishabituation category. I return to this point below.
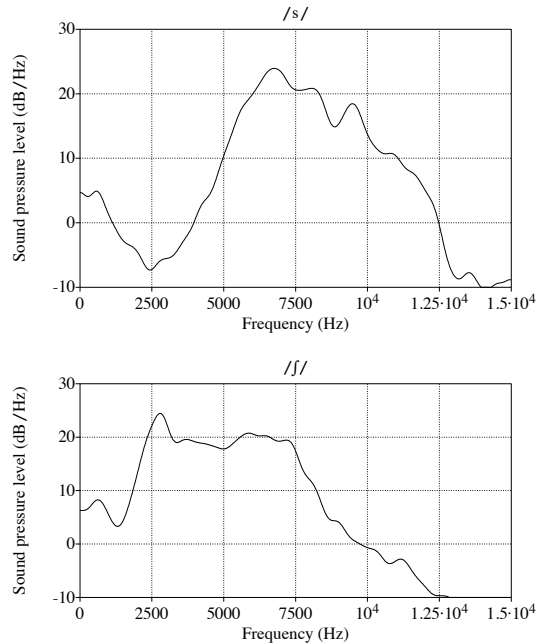
FIG. 1. The top panel is a spectrum (distribution of intensity over frequency) for /s/ and the bottom one a spectrum for /ʃ/. The peak location for /ʃ/ is usually much lower than that for /s/.

2. Choice of category

Given these constraints, it was necessary to choose carefully which category would serve as the novel one. Four reasons lead to the choice of /s/ as the target category, and its acoustic neighbor /ʃ/ as the habituated or background, detailed as follows.

There should be a great deal of individual variation. The main identifying acoustic characteristic of /s/ is that the peak of energy during the closure/frication portion is at a high frequency, as evident in the comparison between /s/ and /ʃ/ displayed in Figure 1.[1] In all languages where /s/ has been studied, a great deal of individual variation in average peak location has been recorded (Gordon *et al.*, 2002), which was confirmed to still be the case in infant-directed speech in pilot testing (see also Figure 4 in the Results section; and in clear speech, judging from the error bars in Figures 1 and 2 for /s/ and /ʃ/ in Maniwa *et al.*, 2009).

6

Phonetically clearer instantiation should not be interpretable as phonologically enhanced. In the search of individual variation, a salient option would have been a vowel such as [i] in the speech of talkers of a 5-vowel system like that of Spanish, as there is room for enhancement by pronouncing it with a higher and fronter tongue position. However, the difference between more and less extreme [i]'s could be encoded phonologically as the potentially phonemic contrast between /i/ and /ɪ/. In this case, enhanced perception in the child could still have been explained within the traditional account as the use of feature enhancement, with some Spanish-speaking parents producing the phonologically more distinct [+high][+ATR] ([i]), and the others [+high] (the less extreme version). In such a scenario, the two potential explanations, based on phonological and phonetic enhancement, would be confounded.[2] Clear pronunciations of /s/ are achieved by raising the spectral center of gravity or peak location in the aperiodic fricative energy corresponding to the closure (see e.g., Jongman *et al.*, 2000 and Maniwa *et al.*, 2009), and normal listeners' intelligibility benefits the most from increases in energy at higher frequencies (Maniwa *et al.*, 2008). Since the peak location of /s/ is already the highest among sibilants (Gordon *et al.*, 2002), this enhancement does not turn it into a different phonemic category. The phonological explanation was thus made unavailable for the category chosen here.

Phonetic implementation should be (largely) independent from general clarity of speech and infant-directed quality. Ideally, variation in the acoustic realization of the category studied should not be confounded with other acoustic parameters that may be present in infant-directed speech and that are known to affect infant performance. For instance, research on individual variation in intelligibility shows that clear speakers tend to expand their vowel spaces as well as to produce more extensive pitch ranges (Bradlow *et al.*, 2003; Ferguson and Kewley-Port, 2007; Picheny *et al.*, 1985; Smiljianic and Bradlow, 2005). This irrelevant pitch dimension greatly affects infants' performance: More extensive pitch ranges attract infants' attention (Fernald and Kuhl, 1987) as well as directly facilitate vowel discrimination (Trainor and Desjardins, 2002), and processing of non-linguistic material (Kaplan *et al.*, 1995b, 1996). Furthermore, naturally occurring individual differences

in caregivers' prosody have an important effect on infants' learning abilities (Kaplan *et al.*, 2002). Thus, if we were to find an association between caregivers' acoustic implementation of vowels and infants' performance, it could be attributed to that lurking factor (compare Liu *et al.*, 2003; see also Discussion below). Contrastingly, variation in /s/ implementation appears to be more independent from possible confounding variables, such as differences in pitch and speech rates (Baum, 2004; Maniwa *et al.*, 2009; Shadle *et al.*, 1992).

Infants should *learn* the category chosen. In Nittrouer (2001), out of 23 children aged 6-14 months who were reliably able to discriminate a vowel or a stop voicing contrast using Conditioned Headturn, only 6 (26%) were also able to tell /s/ and /ʃ/ apart. These results show that discrimination for /s,ʃ/ is not at ceiling in infancy, and, therefore, there is a strong chance that infants have to learn to discriminate these sounds. Additionally, in acoustic-phonetic terms, /s/ and /ʃ/ are highly similar between them and they stand out among other English sounds. Therefore, I hypothesized that if this surface similarity played any role, at least some learners may entertain the possibility that /s/ and /ʃ/ are two realizations of the same category, in which case three additional facts would lead them to treat sounds closer to our /s/ category as more prototypical. To begin with, /s/ is acoustically more extreme and more different from all other consonants, which could make it more salient (similarly to extreme values of vowels being more salient to infants; Polka and Bohn, 2003). Indeed, /s/ is enhanced in clear speech, whereas changes for /ʃ/ are not significant (Maniwa *et al.*, 2009). The same occurs for infant-directed speech enhancement: Comparison of infant- and adult-directed speech showed that caregivers enhance /s/, but not /ʃ/, when talking to their word-learning children (Cristià, 2010b). Furthermore, while /s/ and /ʃ/ are phonologically contrastive in pre- and post-vocalic positions, they are in complementary distribution in clusters, with /s/ being acceptable in most of them (e.g., sm, sn, sp, st, sk, sl; but [ʃr]). (In fact, in one context, /s/ and /ʃ/ are in free variation: There is a good deal of interspeaker variation as to whether the sibilant in the context "_tr" is pronounced as /s/ or /ʃ/ in the Midwest, where dyads tested in the present work reside; Durian, 2007.) Finally, /s/ is much more frequent than /ʃ/: For example, the ratio in word-
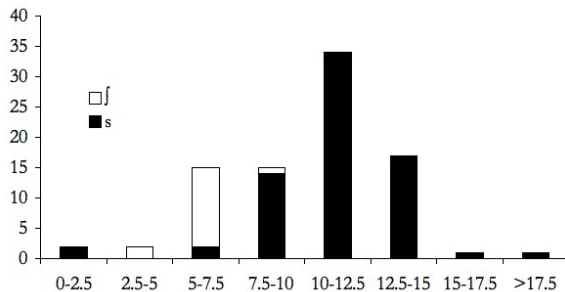
FIG. 2. Histogram of peak location values (in kHz) collapsing across /s/ and /ʃ/ tokens in a caregiver with high /s/ peak location.

initial prevocalic position in an infant-directed corpus is of 9 to 1 (in the portion of the spontaneous infant-directed speech corpus of Soderstrom *et al.*, 2008 that is addressed to infants between 6 and 10 months). As a consequence, even in talkers with a very clear /s/, histograms of the peak location in the frication that collapse across /s/ and /ʃ/ often show a single mode, around the peak location for /s/, with a long tail towards the /ʃ/ peak location frequencies. An example of this is shown on Figure 2, based on the data of a caregiver (code 3052), who has the highest average peak location for /s/ among the caregivers who participated in the present study.

3. Choice of variables

Predictor and possible confounded variables. For all the reasons given above, individual variation in the acoustic implementation of /s/ should generally be a better predictor of variation in infant performance, particularly if infants posit a single category for both sibilants. If the latter is true, the center for this 'super-category' will lie closer to our center for /s/, since it is more frequent (and probably more salient). Thus, the main correlate of interest was the acoustic implementation of /s/, rather than the acoustic implementation of /ʃ/ or the distance between /s/ and /ʃ/.[3] While it is clear that *relative* measures, such as the overlap between categories, are much more relevant for adults' labeling than absolute measures (Newman *et al.*, 2001), adults in a labeling task *know* that there are 2 underly-

9

ing categories, and they have a great deal more experience with the categories as they are instantiated by different talkers. Therefore, even for talkers who have considerable overlap between the two categories, adults may be able to recover the underlying bimodal distribution. It is unclear whether infants, for whom there could be any number of underlying categories and who have less experience with distribution-fitting, are able to do the same. Nonetheless, two indices of acoustic separation between /s/ and /ʃ/ were also calculated. One index of separation was D(a) (as in Newman *et al.*, 2001), a measure akin to d', but which does not assume the variance of the two distributions to be the same. In addition to having different variances, the two sibilants also differ in their frequency of occurrence, with /s/ being much more frequent than /ʃ/. Therefore, another measure of separation was considered, the t-value from a Welch's test (which does not assume equality in variance *or* sample size). As mentioned above, it was desirable to assess the impact of caregivers' pronunciation of /s/ beyond any wholesale effects of clarity of speech or infant-directed quality of the speech. Therefore, a number of additional measures were drawn from the corpora. Two measurements served as proxies for general clarity of speech: Rate of speech and vowel space size (the area circumscribed by /i,a,u/ in F1xF2 space), which correlate highly with intelligibility (e.g., Picheny *et al.*, 1985; Bradlow *et al.*, 1996). Measurements of average pitch and size of pitch excursions were carried out as proxies for quality of infant-directed speech or affectiveness (e.g., Grieser and Kuhl, 1988; Kaplan *et al.*, 2009). **The specific prediction was that infants whose caregivers' peak location for /s/ was higher would have better discrimination scores.** To test this prediction, infants were split on the basis of the median peak location for /s/. If the prediction is correct, the two median groups should differ in discrimination scores but not in any of the other measurements.

Outcome measure. As for the measurement to be used as an index of discrimination, previous work assessing individual variation in speech perception with visual habituation methods has found good test-retest reliability for a ratio-based novelty measure, in which variation in basic looking time is taken into account (Houston *et al.*, 2007). Such ratio-based novelty preference measures are are preferable over difference in looking times in

discrimination and dishabituation tasks (e.g., visual paired comparison in Rose *et al.*, 2009) because there are important differences in infants' basic attention (e.g., Colombo, 1993) that may obscure individual variation in other cognitive constructs. For instance, a ratio-based 'novelty score' is used in paired comparison tasks assessing speed of visual processing and memory, and it yields a reliable predictor of later educational achievement and IQ (e.g., Fagan *et al.*, 2007; Rose *et al.*, 2003). Therefore, the present study used a ratio-based novelty preference measure that was functionally equivalent to the preference quotient used by Houston et al. (2007; they divided the difference between the looking time to novel minus the looking time to old by the average between them), but which is more transparent than that one. Specifically, the looking time during the novel category trials was divided by the sum of the looking times during the novel and the habituated category trials: A performance ratio over .5 would indicate successful discrimination, due to the expected recovery of attention.

B. Methods

1. Participants

Twenty-four 12- to 14-month-olds (age: $M = 12;27$, range 12;3 to 13;26, 12 female) and 18 4- to 6-month-olds (age: $M = 4;26$, range 4;10 to 5;14, 7 female) were included in the data. The caregiver who was recorded spent at least 50% of the child's wake time with him/her (on average, 89% of the time; average for the younger group: 83%; for the older group: 93%). All caregivers were female, and all but one were the infant's mother.

An additional 21 infants participated in at least part of the experiment, but their results were not included for the following reasons: being twins or triplets (9); equipment error or data loss (4); failing to finish the discrimination task (3) or to habituate in 24 trials (1); caregiver was male (3); experimenter error (1).

2. Infants' discrimination

Infants were tested with a variant of the Visual Habituation procedure (Maye *et al.*, 2008; Werker *et al.*, 1998). The infant sat on their caregiver's lap in a small room. Images were projected onto a screen at the infant's eye level. Both sides and the rest of the front of the room were covered with black curtains, concealing the experimenter and the equipment from the infant's view, except for a small peephole through which the experimenter could monitor the infant's gaze online and videorecord the session for offline coding. All stimuli were presented using Habit (Cohen *et al.*, 2000) running on a Power Mac G5. Both the experimenter and the caregiver listened to loud masking music over headphones (Peltor Aviation).

There were two phases to the experiment, habituation and test. In both, trials started with an attention-getter (a spinning water wheel); when the infant gazed at the screen, a colorful bull's eye was displayed. Habituation trials were terminated when the infant looked away for more than 2 s, while test trials had a fixed duration of 15 s. Based on online coding, habituation was determined at the end of a trial if the average looking time for that trial and the two preceding ones dropped below 40% of the average looking times for the three trials in which that infant had looked the longest. Given the way Habit calculates this average, habituation could occur after 4 trials. The auditory stimuli during habituation consisted of 9 syllables beginning with /ʃ/, with the vocalic frame alternating between /a/ and /ɔ/ (that is, /ʃa ʃɔ ʃa ʃɔ/). When the infant habituated, or after 24 trials, she was presented with four test trials. The first two were 'change' trials, in which 9 syllables beginning with /s/ (with the vowel alternating between /a/ and /ɔ/) were presented. In the two 'same' trials, the infant heard 9 novel /ʃ/-initial syllables with those same vocalic frames. In order to be able to compare infants' performance, all infants were tested with the exact same stimuli and presentation order.

The speech stimuli used in habituation and test trials were produced by a female speaker of American English in an infant-directed register. The stimuli were recorded in a sound-

proof booth (IAC, model 403a), using a Marantz Professional Solid State Recorder (PMD 660) and a hypercardioid microphone (Audio-Technica D1000HE). The syllables used in the 'change' and 'same' test trials had been chosen so that they were matched in duration, amplitude, average pitch, and pitch excursion patterns. Acoustic measurements carried out *a posteriori* with the same methods used to analyze caregivers' speech revealed that the average peak location for /s/ was 9860 Hz with a SD of 737 Hz; the mean peak location for /ʃ/ was 4023 Hz, with a SD of 239 Hz.

Looking during test trials was coded off-line from a video that had been digitized at 30 frames per second. The response measure was a discrimination score, represented by the ratio of total looking time to 'change' trials divided by the sum of the looking time to 'change' and 'same' trials. A discrimination score of 0.5 or less indicates no recovery of attention for the new category as compared to the new tokens, suggesting lack of discrimination.

3. Caregivers' speech

After the habituation task, caregiver and infant were taken into a small, sound treated room. Caregivers were told that we were interested in finding out how infants learn categories. They were then provided with a set of objects chosen to elicit the target sounds (/s, ʃ/) in word-initial position as well as the corner vowels /i,a,u/ and some filler items, taking into account both sounds in the object label and sounds in highly related words. They were asked to show the infant the objects and explain what they are for, and how the objects could be sorted into categories. They were then left alone with the child for about 5 minutes. Objects had been selected to fit with the task proposed to caregivers, in order to maintain, as much as possible, the ecological validity of the speech sample collected. Those instructions had been given to ensure that the parent produced the target sounds without being overly conscious about the clarity in their speech, focusing instead on the sorting task. The full list of targets and one possible categorization for them are given on Table I.

The caregiver's speech was recorded with an AKG WMS40 Pro Presenter Set Flexx

UHF Diversity CK55 Lavalier mic (which has a completely flat frequency response between 50 and 20,000 Hz), into a Marantz Professional Solid State Recorder (PMD660ENG), with a sampling frequency of 44.1 kHz. Coding and analyses were done using Praat (Boersma and Weenink, 2005). Sound files were first split into prosodically-determined sentences (for the speech rate, average pitch, and pitch range calculations), and the sounds of interest were hand-tagged. All sibilants and point vowels were tagged (even those that were not part of the word target set; e.g., "strawberry shortcake" counted for /s/ and /ʃ/), unless there was noise or talker overlap in the frication. The location of the spectral peak was estimated on the basis of Fast Fourier Transforms (FFT) on a 40-ms Hamming window centered in the midpoint of the tagged frication noise (Jongman *et al.*, 2000). The peak location was considered to be the frequency of the most intense energy peak, excluding frequencies below 1000 Hz in order to avoid picking up on contextual voicing. Given the 40 ms window of analysis, frications shorter than this length were excluded. A single measure of peak location was calculated as the average over all the /s/ tokens for each caregiver. Average peak location for /ʃ/ was calculated similarly. The distance between the two sibilants for each caregiver was estimated using two measures based on the average peak location for /s/ and /ʃ/ and each category's variance within the speech of that caregiver. If, in a given caregiver's speech, the two sibilants have very different peak locations and their respective peak locations are not very variable, then the 2 categories are very distinct, and the index will be high. In contrast, caregivers for whom the distance in average peak location is small, or where the variance is large, will have lower indices, and their /s/ and /ʃ/ will be confusable because they are similar or they overlap. Both indices of the distance between /s/ and /ʃ/ had the difference in mean peak location for /s/ and /ʃ/ in the numerator, but the denominator is different in the two measures, as one takes into account unequal variances [D(a) $= \frac{\bar{x_1}-\bar{x_2}}{\sqrt{\frac{(s_1{}^2+s_2{}^2)}{2}}}$)], and the other takes into account both unequal variances and unequal frequency of occurrence [Welch's t $= \frac{\bar{x_1}-\bar{x_2}}{\sqrt{\frac{s_1{}^2}{N_1}+\frac{s_2{}^2}{N_2}}}$, where 1 and 2 stand for /s/ and /ʃ/ respectively; x is the mean, s^2 is the variance, and N the sample size]. In addition, vowel space sizes were calculated as the area circumscribed by the average F1 and F2 frequencies (in Bark) for the point vowels /i,a,u/,

14

using Heron's formula ($size = \sqrt{s(s-a)(s-b)(s-c)}$, where $s = \frac{1}{2}(a+b+c)$ and a, b, c are the lengths of the sides of the triangle defined by /i,a,u/.) The F1/F2 measurements were done using an implementation of the ceiling optimization algorithm proposed by Escudero *et al.* (2009) accessible at https://sites.google.com/site/acrsta/scripts, which is robust to interspeaker variation (Cristià, 2010a).

C. Results

As expected, there was a great deal of variability in /s/ implementation with respect to average peak location: the median average peak location was 7370 Hz, the mean 7405 Hz; and the range of average peak location across caregivers was 5599-10962 Hz. Caregivers were classified into "Higher" if their mean peak location was above the median (which included caregivers of 7 younger and 14 older children), and "Lower" otherwise (which included caregivers of 11 younger and 10 older children; no significant relationship between Median and Age, $X^2(1) = 1.56$, $p > .2$). The two groups did not differ on the number of /s/, /ʃ/, and vowel tokens tagged or excluded (all $ps > .3$), as shown on Table II. As mentioned above, effects could be mediated by other aspects of caregivers' speech which are linguistically irrelevant, but cognitively important, or by general speech clarity. Therefore, as a first measure, these two groups of caregivers were compared in the average pitch, pitch variability, speech rate, and vowel space size exhibited in the speech they addressed to their children. No significant differences emerged when collapsing across age groups [unless noted, all $ps > .1$; throughout this paper, it was not assumed that there were equal variance or equal samples, and therefore degrees of freedom are calculated using the Welch-Satterthwaite equation; average Lower versus average Higher for pitch: 251 versus 266 Hz, $t(35) = 1.5$; pitch range: 275 versus 272, $t(40) = .25$; speech rate: 4.42 versus 4.18 syllables/second, $t(39) = 1.25$; vowel space size: 4.95 versus 5.74 Bark^2, $t(39) = 0.82$]. Similarly, there were no stable significant differences when considering these comparisons within the two age groups [within the younger group, pitch: 239 versus 268 Hz, $t(14) = 2.19$, $p = .05$;[4] pitch range: 269

versus 265 Hz, $t(14) = .19$; speech rate 4.52 versus 4.27 syllables/second, $t(9) = .78$; vowel space size: 3.81 versus 5.15 Bark^2, $t(9) = 1.19$; within the older group, pitch: 265 versus 265 Hertz, $t(18) = .04$; pitch range: 282 versus 275 Hertz, $t(21) = .38$; rate of speech: 4.31 versus 4.13 syllables/second, $t(20) = .66$; vowel space size: 6.2 versus 6.03 Bark^2, $t(20) = .11$].

An Analysis of Variance with the infants' discrimination score as the dependent measure, and Age group (Younger, Older) and Median group (Higher, Lower) as independent variables, reveals no effect of Age $[F(1,38) = .66]$ and no interaction of Age and Median $[F(1,38) = .34]$, but a significant effect of Median $[F(1,38) = 4.59, p < .05]$, which was medium to large [difference in means divided by the average of the standard deviation, Cohen's $d = .67$; effect size values around .5 are considered medium, around .8 large]. The main effect of Median is due to infants' discrimination scores being significantly higher for the Higher group than the Lower group [given the predicted direction, tests are one-tailed; overall: $t(39.9) = 2.17, p = .02$; within the older age group: $t(20) = 1.79, p = .04, d = .742$; within the younger age group: $t(14) = 1.49, p = .08; d = .706$]. If the median split is done within each Age group, similar results ensue [within the older Age group $t(22) = 1.94, p = .03; d = .796$; within the younger Age group $t(15) = 1.48, p = .08; d = .687$]. Average discrimination scores and standard errors for the infants within each Median Group are shown on Figure 3; and the individual data is plotted on Figure 4.[5]

One possibility suggested during the review process was that infants' performance responded to the general clarity of sibilants (that is, both /s/ and /ʃ/) or the distance between the /s/ and /ʃ/ categories. It is true that caregivers with a clear /s/ also tend to have a higher peak location for /ʃ/; and that, since /ʃ/ is not as variable, they also have more distinct /s/ and /ʃ/, as evident in Figure 4 and Table III. Indeed, the two median groups differed significantly on all 4 measures (/s/ and /ʃ/ peak locations; D(a) and Welch's t-values for the distance between the 2 categories). Therefore, at present it is not clear whether infants in the Higher group had greater discrimination scores because they were responding based on the perceptual distance between the categories (which assumes that all infants have two
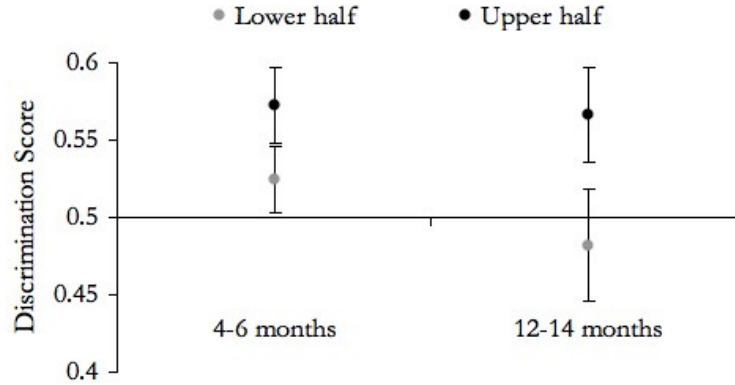
16

FIG. 3. Average discrimination scores (error bars show standard errors) by Median Group and Age Group.

underlying categories) or based on the distance from the habituated /ʃ/ to their /s/-driven prototype (which makes no assumptions with respect to the number of categories). In either case, it would still hold true that infants' perception is affected by the acoustic cue distributions that they are exposed to.

A third, less likely, explanation could also be posited: That the infants' state after the experiment determined caregivers' /s/, since the speech sample was gathered immediately after the infant test. For example, it is possible that infants who have higher discrimination scores are those that grew more tired throughout the experiment; in this case, their looking times for the 2 final trials (the "same" trials) would be lower than those in the 2 preceding ("different") trials, and their caregivers would have modulated their speech during recording also as a function of infants' tiredness. To assess this possibility, the 2 Median groups were compared in the number of habituation trials, since infants with a greater tendency to grow tired should also habituate more quickly. However, there was no significant difference in the number of trials to habituation across the two groups [9.95 versus 8.95, $t(34) = .88$, $p>.38$]. Furthermore, as mentioned above, there were no differences in pitch height and excursions, even though caregiver pitch is strongly modulated by infant state (Braarud and Stormark, 2008; Masataka, 1992; McRoberts and Best, 1997; Smith and Trainor, 2008; Stern *et al.*,
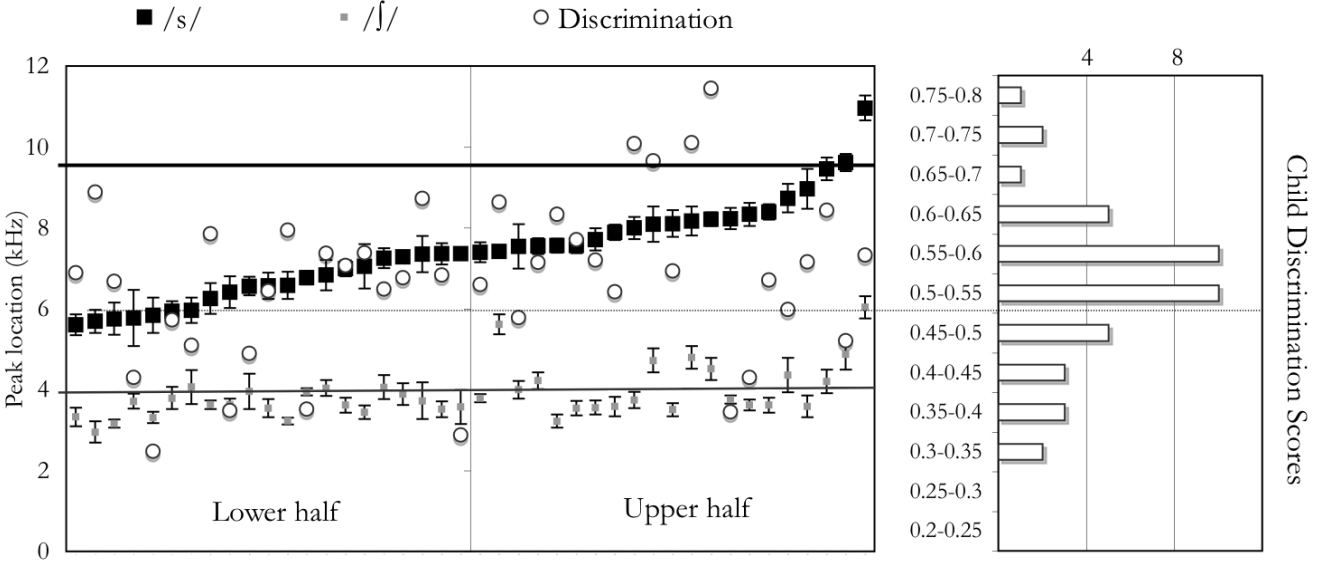
FIG. 4. Individual data are plotted on the left panel, sorted by /s/ peak location. For each dyad, peak location in the caregiver's speech is noted with squares (/s/=dark; /ʃ/=light; both plotted on the kHz scale shown on the far left; error bars indicate standard errors); and the infant's discrimination score is plotted with a white circle in the same vertical line (using the scale between the two panels). The right panel shows a histogram for infants' discrimination score (on the same scale). The dotted horizontal line crossing both panels signals the .5 proportion level for the Discrimination Scores, above which there would be evidence of successful discrimination. The other two horizontal lines, crossing only the left panel, represent the average peak location for the /s,ʃ/ stimuli used in the infant test (/s/=dark, /ʃ/=light).

1982).

III. DISCUSSION

Infants whose caregivers produced a more acoustically extreme /s/ were better able to discriminate this category from /ʃ/. This may indicate, more in general, that infants whose caregivers produce clearer /s/ categories are in a better position to learn this sound, and

18

to discriminate it from acoustically similar categories. These results are hard to explain within the traditional view of phonological acquisition, according to which infants ignore variation in acoustic realization that is irrelevant to category membership, but fit well with the hypothesis that infants' perceptual tuning is guided by acoustic cue distributions. Thus, the present study provides a crucial piece of the phonological acquisition puzzle, being the first to show that non-phonemic, fine-grained variation in implementation affects infants' perceptual development in natural language acquisition.

This effect was somewhat stronger for the older group than for the younger one, possibly due to the smaller group size, or to the fact that neither measure was as variable in the younger group; for example, overall variance in discrimination scores was less than .005 in the younger children, but over .014 in the older group. Alternatively, this difference in effect size (and variance) could be accounted for within a learning explanation, since the older infants have had several more months of exposure to their caregivers' speech than the younger ones, and are naturally more affected by it. Additionally, inspection of Figure 3 suggests that the largest difference between the younger and the older infants concerns the Lower-Median subgroups, as if changes in infants' discrimination tended to be in the direction of attenuation (i.e., infants exposed to relatively bad /s/ categories lose the ability to hear the distinction, whereas for the other children this ability is maintained.) Such a pattern of results would be expected if, based on general audition or Universal Grammar, infants are biased to divide this acoustic space into at least two regions, and /s/ with a low peak location leads to a merger.[6] Although this is an interesting possibility, there was not enough power in the present data to explore it. Future work with even younger infants and a longitudinal design may be better suited to assess this explanation.

Nonetheless, the design of the study was such that it was not necessary to assume that all infants had two underlying categories; that is, infants might either have two categories, or a single one with a mode closer to /s/. In the first case, the Higher-Median group would show higher recovery of attention because their /s/ and /ʃ/ categories are well separated. According to the second interpretation, infants in the Higher-Median group would recognize

the novel category as the prototype. As mentioned above, it is at present impossible to tease apart these explanations, both of which nonetheless attribute infants' performance to sibilant instantiation.

In contrast, it is unlikely that results are due to an affective or hyperarticulatory quality of the caregivers' speech, given that the two groups of infants did not differ on infant-directed characteristics (average pitch and range), nor on clarity of speech measurements (speech rate, vowel space size). This was important because one could imagine a causal relationship between infant-directed quality and wholesale clarity of speech, on the one hand, and infant performance, on the other, where the instantiation of sound categories is secondary: Clarity and infant-directed quality of caregiver speech would have a general positive effect on infant discrimination skills because they promote language acquisition (Kuhl *et al.*, 2008); and/or they could facilitate word segmentation, which would in turn help infants learn sound categories (Swingley, 2009). For example, Liu *et al.* (2003) report that infants' discrimination of a temporal contrast (that between the alveopalatal fricative and affricate) measured using the Conditioned Headturn procedure is predicted by the size of their mothers' vowel space. As pointed out in the Introduction, one problem with measuring vowel space is that it is significantly associated with linguistically irrelevant, but cognitively important, acoustic characteristics. For example, more distinct vowel categories have been repeatedly associated with vocally expressed positive emotion and increased affect (e.g., Schaeffler, 2006). Given that the Conditioned Headturn procedure is heavily reliant on associative learning, and that the measurement found to correlate with maternal vowel space was trials to criterion rather than the discrimination index d', the reported association could be due to any of the confounded variables, such as emotional and cognitive development. As for indirect effects through word segmentation and lexical acquisition, it is controversial that infant-directed speech or its acoustic characteristics always facilitate word recognition: Thiessen *et al.* (2005) find improved word segmentation by infants, but Singh *et al.* (2004) do not, and while Singh *et al.* (2009) documents that infants remember better words they had heard spoken in infant-directed speech, this could be an effect of heightened arousal leading

20

to better cognitive performance (Kaplan *et al.*, 1995a) since the signal itself is actually noisier (de Boer and Kuhl, 2003). In sum, the evidence in this study and elsewhere for these indirect links between caregiver speech and infant performance do not appear strong enough to warrant ruling out the simpler explanation of a link through sibilant characteristics.

There are, nonetheless, two alternative explanations for the present study that cannot be ruled out. One possibility is that infants in the Higher group were better able to recognize the /s/ tokens used in test, as these were closer to the /s/ category present in their caregivers' speech. This explanation may be explored by replicating the present experiment using /s/ tokens with a lower peak location. Nonetheless, even if this were true, it would still indicate that infants attend to fine-grained details of acoustic realization when learning categories from their caregivers' speech. Alternatively, caregivers' acoustic implementation of /s/ and infants' discrimination abilities in the present sample could be explained by inheritable individual differences in auditory sensitivity, due to the fact that most of the caregivers recorded are also the infant's mother. Thus, if individual variation in both discrimination abilities and clarity of speech were due to auditory sensitivity, and this sensitivity was inheritable, then the link between acoustic distinctness in caregivers' speech and discrimination in the infant would be due to their dependence on a third variable, rather than being causally linked. Indeed, some research shows that talkers who produce more distinct /s,ʃ/ are better able to discriminate them (Perkell *et al.*, 2004). In order to test this hypothesis, it would be necessary to test infants whose primary caregiver is *not* related to them, or to find the effect of mothers' speech to be mediated by the amount of time spent with the infant. Given that most caregivers in this sample were the infant's mother and there was little variation in time spent with the infant, this possibility remains to be explored in future work.

In summary, the results of the present study suggest that fine-grained variation in the infants' input may affect their perceptual categories, and that this process takes place early in the infant's life. This conclusion resonates with the increasing support for theories that attribute a larger role to domain-general explanations for perceptual tuning, rather than linguistically restricted, nativist accounts.

Acknowledgments

Endnotes

1. Although in English /s/ also differs significantly from /ʃ/ along other acoustic dimensions, a large production study using words elicited in isolation found that 100% accuracy in classification could be achieved with peak location/centroid alone (Li, 2008, p. 46; a less successful, but equally uni-dimensional classification appears feasible for more spontaneous data: see Cristià, 2009, pp. 110-112).

2. Naturally, hyperarticulation in English /i/ could not be explained as feature enhancement, because English requires [ATR].

3. Another option would have been to calculate the peak location over all of the tagged sibilants, not just over the /s/ tokens. However, as will be explained below, the samples elicited overestimated the frequency of occurrence of /ʃ/, since approximately the same number of target words with /s/ and /ʃ/ were selected when choosing the elicitation toys. As a result, while the corpus count mentioned above yielded a 9 to 1 ratio, the ratio in the current corpus was 3 to 1.

4. This marginal comparison (the only one among 12 t-tests) is not replicated in the older age group, where the difference in discrimination scores across median groups is stronger. Therefore, it is unlikely that differences in caregiver pitch explain the

pattern of results reported in this paper, and the marginal difference in pitch is likely spurious.

5. Inspection of the individual data suggests two side comments. First, one could initially think that any deviance from .5 is an index of discrimination, with positive deviations being novelty and negative ones familiarity preferences. A familiarity preference would indeed indicate discrimination in a design with fixed exposure; however, in a habituation design, longer looking times to the habituated category cannot be taken as a familiarity preference (because by design this has been exhausted in the first phase.) Second, although there may not be enough data points to establish the shape of the relationship between caregivers' /s/ and infant discrimination beyond any doubts, current data suggest that the relationship is not linear, that is, it is not the case that increases in /s/ peak location are always accompanied by increases in discrimination (Spearman's $\rho = .224$, and Pearson's r $= .237$, both $p > .1$).

6. Notice that even in this instantiation of the traditional hypothesis, one must give up the assumption that phonological acquisition is a selection process determined by the presence or absence of sounds: In all caregivers, both /s/ and /ʃ/ are present.

References

Aslin, R. N. and Pisoni, D. B. (**1980**). "Some developmental processes in speech perception", in *Child Phonology*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson, volume 2: Perception, 67–96 (Academic, New York).

Baum, S. R. (**2004**). "Fricative production in aphasia: Effects of speaking rate", Brain and Language **1996**, 328–341.

Boersma, P. and Weenink, D. (**2005**). "Praat: Doing phonetics by computer (version 5.0.09)", Retrieved May 26, 2007, from //www.praat.org/.

Braarud, H. C. and Stormark, K. M. (**2008**). "Prosodic modification and vocal adjustments in mothers' speech during face-to-face interaction with their two- to four-month-old infants: A double video study", Social Development **17**, 1074–1084.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for children with learning disabilities: Sentence perception in noise", Journal of Speech, Language, and Hearing Research **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 Communication **20**, 255–272.

Cohen, L. B., Atkinson, D. J., and Chaput, H. H. (**2000**). "Habit 2000: A new program for testing infant perception and cognition, version 1.0", (computer program).

Colombo, J. (**1993**). *Infant cognition: Predicting later intellectual functioning* (Sage, Newbury Park, CA; London, New Delhi, 176 pp).

Colombo, J. and Fagan, J. (**1990**). *Individual differences in infancy. Reliability, stability, prediction* (Lawrence Erlbaum Association, Hillsdale, NJ, 480 pp).

Cristià, A. (**2009**). "Individual differences in infant speech processing", Ph.D. thesis, Purdue University.

Cristià, A. (**2010**a). "A comparison of algorithms to automatically extract vowel formants and nasal poles from tagged vowels in Praat (Abstract)", Journal of the Acoustical Society of America **128**, 2291.

Cristià, A. (**2010**b). "Phonetic enhancement of sibilants in infant-directed speech", Journal of the Acoustical Society of America **128**, 424–434.

Cristià, A., McGuire, G. L., Seidl, A., and Francis, A. L. (**2011, in press**). "Effects of the distribution of acoustic cues on infants' perception of sibilants", Journal of Phonetics .

de Boer, B. and Kuhl, P. K. (**2003**). "Investigating the role of infant-directed speech with a computer model", Acoustics Research Letters Online **4**, 2238–2246.

Durian, D. (**2007**). "Getting [ʃ]tronger every day?: More on urbanization and the socio-geographic diffusion of (str) in columbus, OH", University of Pennsylvania Working Papers

in Linguistics **13**, 65–79.

Escudero, P., Boersma, P., Rauber, A. S., and Bion, R. A. H. (**2009**). "A cross-dialect acoustic description of vowels: Brazilian and European Portuguese", Journal of the Acoustical Society of America **126**, 1379–1393.

Fagan, J. F., Holland, C. R., and Wheeler, K. (**2007**). "The prediction, from infancy, of adult IQ and achievement", Intelligence **35**, 225–231.

Ferguson, S. H. and Kewley-Port, D. (**2007**). "Talker differences in clear and conversational speech: Acoustic characteristics of vowels", Journal of Speech, Language, and Hearing Research **50**, 1241–1255.

Fernald, A. and Kuhl, P. K. (**1987**). "Acoustic determinants of infant preference for motherese speech", Infant Behavior and Development **10**, 279–293.

Gordon, M., Barthmaier, P., and Sands, K. (**2002**). "A cross-linguistic acoustic study of voiceless fricatives", Journal of the International Phonetic Association **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", Developmental Psychology **24**, 14–20.

Hale, M., Kissock, M., and Reiss, C. (**2006**). "Microvariation, variation, and the features of universal grammar", Lingua **32**, 402–420.

Hale, M. and Reiss, C. (**2003**). "The subset principle in phonology: Why the tabula can't be rasa", Journal of Linguistics **39**, 219–244.

Houston, D., Horn, D. L., Qi, R., Ting, J. Y., and Gao, S. (**2007**). "Assessing speech discrimination in individual infants", Infancy **12**, 119–145.

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

Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (MIT Press, Cambridge, MA, 326 pp).

Kaplan, P. S., Bachorowski, J.-A., Smoski, M. J., and Hudenko, W. J. (**2002**). "Infants of depressed mothers, although competent learners, fail to learn in response to their own

mothers' infant-directed speech", Psychological Science **13**, 268–271.

Kaplan, P. S., Burgess, A. P., Sliter, J. K., and Moreno, A. J. (**2009**). "Maternal sensitivity and the learning-promoting effects of depressed and nondepressed mothers' infant-directed speech", Infancy **14**, 143–161.

Kaplan, P. S., Goldstein, M. H., Huckeby, E. R., and Cooper, R. P. (**1995**a). "Habituation, sensitization, and infants' responses to motherese speech", Developmental Psychobiology **28**, 45–57.

Kaplan, P. S., Goldstein, M. H., Huckeby, E. R., Owren, M. J., and Cooper, R. P. (**1995**b). "Dishabituation of visual attention by infant- versus adult-directed speech: Effects of frequency modulation and spectral composition", Infant Behavior and Development **18**, 209–223.

Kaplan, P. S., Jung, P. C., Ryther, J. S., and Strouse, P. Z. (**1996**). "Infant-directed versus adult-directed speech as signals for faces", Developmental Psychology **33**, 880–891.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (**2008**). "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e)", Philosophical Transactions of the Royal Society B **363**, 979–1000.

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.

Li, F. (**2008**). "The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese", Ph.D. thesis, Ohio State University.

Liu, H.-M., Kuhl, P. K., and Tsao, F.-M. (**2003**). "An association between mothers' speech clarity and infants' speech discrimination skills", Developmental Science **6**, F1–F10.

Maniwa, K., Jongman, A., and Wade, T. (**2008**). "Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners", Journal of the Acoustical Society of America **123**, 3962–3973.

Maniwa, K., Jongman, A., and Wade, T. (**2009**). "Acoustic characteristics of clearly spoken

English fricatives", Journal of the Acoustical Society of America **125**, 1114–1125.

Masataka, N. (**1992**). "Pitch characteristics of Japanese maternal speech to infants", Journal of Child Language **19**, 213–223.

Maye, J., Weiss, D., and Aslin, R. (**2008**). "Statistical phonetic learning in infants: Facilitation and feature generalization", Developmental Science **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.

McCall, R. B. and Carriger, M. S. (**1993**). "A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ", Child Development **64**, 57–79.

McMurray, R. and Aslin, D. (**2005**). "Infants are sensitive to within-category variation in speech perception", Cognition **95**, B15–26.

McRoberts, G. and Best, C. T. (**1997**). "Accommodation in mean f0 during mother-infant and father-infant vocal interactions: a longitudinal case study", Journal of Child Language **24**, 719–736.

Miller, J. L. and Eimas, P. D. (**1996**). "Internal structure of voicing categories in early infancy", Perception and Psychophysics **58**, 1157–1167.

Narayan, C., Werker, J. F., and Speeter Beddor, P. (**2009**). "The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination", Developmental Science **13**, 407–420.

Newman, R. S., Clouse, S. A., and Burnham, J. (**2001**). "The perceptual consequences of acoustic variability in fricative production within and across talkers", The Journal of the Acoustical Society of America **109**, 3697–3709.

Nittrouer, S. (**2001**). "Challenging the notion of innate phonetic boundaries", The Journal of the Acoustical Society of America **110**, 1598–1605.

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", Journal of Speech, Language, and Hearing Research **47**, 1259–1269.

Picheny, M. A., Durlach, N. I., and Braida, L. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech", Journal of Speech and Hearing Research **28**, 96–103.

Polka, L. and Bohn, O. (**2003**). "Asymmetries in vowel perception", Speech Communication **41**, 221–231.

Polka, L., Colantonio, C., and Sundara, M. (**2001**). "A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern", The Journal of the Acoustical Society of America **109**, 2190–2201.

Polka, L. and Werker, J. F. (**1994**). "Developmental changes in perception of nonnative vowel contrasts", Journal of Experimental Psychology: Human Perception and Performance **20**, 421–435.

Rose, S. A., Feldman, J. F., and Jankowski, J. J. (**2003**). "The building blocks of cognition", Journal of Pediatrics **143**, S54–S61.

Rose, S. A., Feldman, J. F., and Jankowski, J. J. (**2009**). "A cognitive approach to the development of early language", Child Development **80**, 134–150.

Schaeffler, S. (**2006**). "Are affective speakers effective speakers? Exploring the link between the vocal expression of positive emotions and communicative effectiveness", Ph.D. thesis, University of Stirling.

Shadle, C. H., Dobelke, C. U., and Scully, C. (**1992**). "Spectral analysis of fricatives in vowel context", Journal de Physique IV, Colloque C1, suppIement au Journal de Physique 111 **2**, 295–299.

Singh, L., Morgan, J., and White, K. S. (**2004**). "Preference and processing: The role of speech affect in early spoken word recognition", Journal of Memory and Language **51**, 173–189.

Singh, L., Nestor, S., Parikh, C., and Yull, A. (**2009**). "Influences of infant-directed speech on early word recognition", Infancy **14**, 654–666.

Smiljianic, R. and Bradlow, A. R. (**2005**). "Production and perception of clear speech in Croatian and English", Journal of the Acoustical Society of America **118**, 1677–1688.

Smith, N. A. and Trainor, L. J. (**2008**). "Infant-directed speech is modulated by infant feedback", Infancy **13**, 410–420.

Soderstrom, M., Blossum, M., Foygel, I., and Morgan, J. (**2008**). "Acoustical cues and grammatical units in speech to two preverbal infants", Journal of Child Language **35**, 869–902.

Stern, D. N., Spieker, S., and MacKain, K. (**1982**). "Intonation contours as signals in maternal speech to prelinguistic infants", Developmental Psychology **18**, 727–735.

Swingley, D. (**2009**). "Contributions of infant word learning to language development", Philiosophical Transactions of the Royal Society B: Biological Sciences **364**, 3617–3632.

Thiessen, E. D., Hill, E., and Saffran, J. R. (**2005**). "Infant-directed speech facilitates word segmentation", Infancy **7**, 53–71.

Trainor, L. J. and Desjardins, R. N. (**2002**). "Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels", Psychonomic Bulletin and Review **9**, 335–340.

Werker, J. F. (**1994**). "Cross-language speech perception: Developmental change does not involve loss", in *The development of speech perception: The transition from speech sounds to spoken words*, edited by J. Goodman and H. Nusbaum, 93–120 (MIT Press, Cambridge, MA).

Werker, J. F., Shi, R., Desjardins, R. N., Polka, L., and Patterson, M. (**1998**). "Three methods for testing infant speech perception", in *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy*, edited by A. M. Slater, 389–420 (UCL Press, London).

Werker, J. F. and Tees, R. (**1984**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life", Infant Behavior and Development **7**, 49–63.

Yoshida, K. A., Pons, F., Maye, J., and Werker, J. F. (**2010**). "Distributional phonetic learning at 10 months of age", Infancy **15**, 420–433.

Object label	s	∫	i	a	u	Category	Object label	s	∫	i	a	u	Category
shampoo		x			x	bath	shovel		x				toys
shaving gel		x				bath	bucket						toys
soap	x					bath	sifter	x					toys
bottles			x				octopus						toys
shoes		x			x	clothes	castle shaper	x	x				toys
shirt		x				clothes	*sand (toys)*	x					
shorts		x				clothes	*sea*	x		x			
socks	x		x			clothes	*beach*			x			
sandals	x					clothes	shark		x		x		toys
sunglasses	x					other	*sea*	x		x			
cellphone	x					other	blue block				x	x	toys
sunflower seeds	x					other	soft green block	x		x	x		toys
pumpkin pail						toys	ball						toys
sweets	x		x				cinderella storybook	x					toys
trick or treat			x				*read*			x			
doll						toys	bear						toys

TABLE I. Objects selected to elicit the target sibilants and vowels through the object label and through "primed" words (given in italics below the relevant items).

Median	/i/	/a/	/u/	/s/	Short /s/	/ʃ/	Short /ʃ/
Lower	18.3	13.9	13.3	69.9	4.8	21	0.1
	(9.5;5-41)	(7.6;4-30)	(8.5;3-44)	(28.2;20-162)	(3.3;1-12)	(10.3;6-49)	(0.5;0-2)
Higher	16	13.3	14.1	67.3	3.8	19.4	0.1
	(9.7;2-39)	(7.1;3-26)	(7.5;4-33)	(31;20-149)	(2.8;0-11)	(9;3-33)	(0.4;0-1)

TABLE II. Average (standard deviation; min-max) of the number of segments tagged (short /s/ and /ʃ/ are those excluded for being shorter than 40ms) in each of the Median groups.

	Median group	Peak location /s/	Peak location /ʃ/	Inter-category distance D(a)	Inter-category distance Welch's t	Discrimination scores
Overall	L	6.531 (0.624)	3.622 (0.313)	1.57 (0.51)	7.61 (3.00)	0.504 (0.094)
	H	8.28 (0.876)	4.135 (0.736)	2.61 (0.75)	11.91 (3.60)	0.568 (0.098)
t across groups		7.453 ***	2.934 **	5.29 ***	4.20 ***	2.17 *
4-6 months	L	6.422 (0.594)	3.591 (0.359)	1.51 (0.39)	7.20 (1.97)	0.524 (0.071)
	H	8.391 (1.212)	4.576 (0.962)	2.35 (0.93)	10.07 (3.12)	0.572 (0.064)
t across groups		4.003 **	2.598 *	2.28	2.18	1.491
12-14 months	L	6.651 (0.665)	3.657 (0.268)	1.63 (0.63)	8.07 (3.90)	0.482 (0.114)
	H	8.224 (0.701)	3.914 (0.499)	2.74 (0.64)	12.82 (3.56)	0.566 (0.113)
t across groups		5.583 ***	1.623	4.23 ***	3.05 **	1.792

TABLE III. Average (and standard deviation) in each of the Median groups for relevant acoustic correlates as well as the outcome measure. The first two columns show peak location for the sibilants in kHz, and the following two columns show measures of distance between the two categories within each talker's speech: D(a) (distance between the means divided by the average variance) and Welch's t (distance between the means divided by frequency-weighted variance); the fifth, rightmost column displays the outcome, Discrimination Scores. These averages are given for the whole sample (Overall), and also separating infants in their age groups; in the last line of each of these 3 groupings, the t-value of an unequal variance, two-tailed test across the 2 median groups is shown, together with its significance level (*** $p<.001$; ** $p<.01$; * $p<.05$).

List of Figures