Psicológica (2012), 33, 175-207.

Phonetic category cues in adult-directed speech: Evidence from three languages with distinct vowel characteristics

Ferran Pons[*1], Jeremy C. Biesanz[2], Sachiyo Kajikawa[3], Laurel Fais[2], Chandan R. Narayan[4], Shigeaki Amano[5], & Janet F. Werker[2]

[1]Universitat de Barcelona, Spain; [2]University of British Columbia, Canada; [3] Tamagawa University, Japan; [4] University of Toronto, Canada; [5]Aichi Shukutoku University, Japan.

Using an artificial language learning manipulation, Maye, Werker, and Gerken (2002) demonstrated that infants' speech sound categories change as a function of the distributional properties of the input. In a recent study, Werker et al. (2007) showed that Infant-directed Speech (IDS) input contains reliable acoustic cues that support distributional learning of language-specific vowel categories: English cues are spectral and durational; Japanese cues are exclusively durational. In the present study we extend these results in two ways. 1) we examine a language, Catalan, which distinguishes vowels solely on the basis of spectral differences, and 2) because infants learn from overheard adult speech as well as IDS (Oshima-Takane, 1988), we analyze Adult-directed Speech (ADS) in all three languages. Analyses revealed robust differences in the cues of each language, and demonstrated that these cues alone are sufficient to yield language-specific vowel categories. This demonstration of language-specific differences in the distribution of cues to phonetic category structure found in ADS provides additional evidence for the types of cues available to infants to guide their establishment of native phonetic categories.

In the early months of life, infants can discriminate both native and non-native contrasts with equal ease (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). However, by the end of their first year, these perceptual abilities change, involving decreasing sensitivity to non-native speech contrasts and increasing sensitivity to native ones (Anderson, Morgan, & White, 2003; Bosch & Sebastián-Gallés, 2003; Burns, Yoshida, Hill, &

[*] Corresponding author: Ferran Pons. Departament de Psicologia Bàsica, Facultat de Psicologia, Universitat de Barcelona. Pg. Vall d'Hebrón, 171 Barcelona, 08035, Spain. E-mail: ferran.pons@ub.edu

Werker, 2007; Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Narayan, Werker, & Beddor, 2010; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés; Werker & Tees, 1984). The dominant accounts of these changes in speech perception during the first year of life are based on the assumption that listening experience in infancy leads to the full or partial collapse of category distinctions that are not used, and the strengthening of those that are (Best & McRoberts, 2003; Kuhl, 2004; Werker & Curtin, 2005).

One potential source of information that could lead to the strengthening of native, and the collapse of non-native, categories, is the distributional characteristics of the input (see Kuhl, 1993 for an early discussion of this possibility). Maye, Werker and Gerken (2002) tested a perceptual learning mechanism based on distributional learning, exploiting the characteristics that might underlie the rapid tuning to the categories of the native language during the first year of life. Using an artificial language learning manipulation, they found that infants could change their speech sound categories as a function of the distributional properties of the input. In their study, 6- and 8-month-old infants were familiarized with eight tokens along the phonetic continuum /ta/-/da/. One group was familiarized with a unimodal frequency distribution; that is, they were exposed more to the central tokens (steps 4 and 5) of the continuum than to its endpoints. The other group was familiarized with a bimodal frequency distribution; they were exposed more to the peripheral tokens (steps 2 and 7) of the continuum than to the central ones. Infants in both groups had equal exposure to steps 1, 3, 6, and 8. After this familiarization phase, infants were tested on a phonetic discrimination task. There were two types of test trials; on half of the test trials (non-alternating trials) a single stimulus from the continuum (token 3 or 6) was repeated, while on the other test trials (alternating trials) infants heard an alternation between two different stimuli (tokens 1 and 8).

The results showed that infants in the unimodal condition responded to the non-alternating and the alternating trials equivalently, suggesting that they had perceived a single category. However, infants in the bimodal condition showed a significant difference in their perception of the alternating and the non-alternating trials; that is, they discriminated the endpoints of the continuum. The frequency distribution of the familiarization stimuli significantly affected subsequent discrimination in the test phase. Maye, Weiss and Aslin (2008) replicated this finding with new sets of speech contrasts. In this experiment, 8-month-old infants were exposed to a voicing contrast of pre-voiced versus short-lag stops, which is poorly discriminated at this age. Again, only infants exposed to the bimodal distribution later discriminated the contrast; in this case, the bimodal distribution facilitated discrimination. Thus, they showed that distributional learning can not only collapse an existing distinction but can facilitate learning of difficult distinctions as well (Maye et al., 2008). More recently, Yoshida, Pons, Maye and Werker (2010) have found that distributional phonetic learning remains effective at older ages (10 month-olds), but it is more difficult than before perceptual reorganization. Together, these results indicate that distributional learning may be a mechanism that contributes to speech sound category restructuring in the first year of life, prior to the establishment of a lexicon.

It is important to note that there are studies that challenge the interpretation of the rapid tuning to the native language categories merely based on a distributional account. Sebastián-Gallés and Bosch (2009) examined the ability of monolingual and bilingual infants to discriminate two common vowel contrasts. The results of their study revealed that other variables would be needed to be considered on top of the distributional ones to account for bilingual infants' processes in establishing certain language-specific phonetic categories.

Laboratory-based artificial language learning studies, such as the studies presented above, provide evidence in principle that a particular learning mechanism is available to infants. But regardless of whether or not infants actually learn phonetic categories using distributional learning (Maye et al., 2002; 2008, but see Sebastián-Gallés & Bosch, 2009), it is clear that infants will not be able to engage this learning mechanism unless the speech they listen to actually contains these distributional regularities. Consequently, Werker and collaborators (2007) examined Infant-directed Speech (IDS) input to explore whether this type of speech contains sufficient cues to support distributional learning of phonetic categories by infants. Analyzing two vowel pairs /E-ee/ and /I-ii/ from English and Japanese IDS input, Werker and colleagues found consistent language-specific cues. In English, the vowel pairs were distinguished primarily by a bimodal distribution of spectral cues and secondarily by a bimodal distribution of duration cues. In Japanese, the vowels in each pair were identical in spectral properties, but exhibited a pronounced bimodal distribution of vowel quantity, significantly more pronounced than that seen in English. Of importance, analyses showed not only that the vowels in maternal speech differed in these properties, but also that it is possible to begin with the exemplars from infant-directed speech and mathematically derive the respective categories in each language. Moreover, using these same data, Vallabha, McClelland, Pons, Werker, and Amano (2007) showed that the language-specific categories are learnable by a variant of an

Expectation-Maximization algorithm on the basis of the acoustic cues present. Taken together, these findings confirm that there are sufficient distributional cues in infant-directed speech to support distributional learning of phonetic categories.

Previous work has shown that the acoustic characteristics of language-specific phonetic distinctions are exaggerated in infant-directed speech (Bernstein-Ratner, 1984; Kuhl et al., 1997) and that a computer model may be better able to derive language-specific categories from infant-directed speech than from adult-directed speech (deBoer & Kuhl, 2003).

In the present paper, we build, in two ways, upon the question of whether the cues in adult-directed speech input also have the distributional profiles necessary to support distributional learning of the native speech sound categories. First, we record and analyze adult-directed speech (ADS) to determine whether the cues that could direct distributional learning of language-specific vowel categories are equally available in speech overheard by, rather than directed to, infants. Second, to better understand how the specific devices utilized in each particular language (in this case, the kind and number of cues) influence the information load of those cues for determining linguistic contrast, we examine three languages, Japanese, English, and Catalan, that together, provide a sample of three of the most important ways in which vowels can be distinguished across the world's languages. In Catalan vowels are classically described as differing only in quality (spectral characteristics); in Japanese the vowels we test are classically described as differing in only quantity (duration), and in English they are described as differing primarily in quality, but with a secondary quantity dimension. There is one pair of vowels, /E/-/ee/, that is used contrastively in each of these three languages. That vowel pair, which will be described in more detail in the Stimulus section, was examined in this study.

Adult-directed speech

Most of the speech that infants hear is not directly addressed to them, but instead, occurs in conversations among the adults around them; the amount of speech directed to infants is estimated to comprise only about 14% of the total speech infants hear (van de Weijer, 2002) with considerable variation across families (Aslin, Woodward, LaMendola, & Bever, 1996). In fact, there are some cultures in which infants are not addressed directly by adults at all (Pye, 1986; Schieffelin & Ochs, 1983). Although infants show a preference for IDS (Cooper & Aslin, 1990; Fernald, 1985) and may be better able to learn some properties of language in infant-directed speech (e.g. Karzon, 1985; Liu, Kuhl, & Tsao, 2003; Thiessen, Hill, & Saffran, 2005), they also listen to and learn from overheard adult speech (Oshima-Takane, 1988). Indeed, the success children show in acquiring the personal pronoun system is due entirely to learning from overheard speech (Oshima-Takane, 1988). In this study, imitative behaviors of 18-month-olds were analyzed under two modeling conditions. The non-addressee condition provided the child with systematic opportunities to observe the parents saying *me/you* with pointing actions directed towards each other as well as the parents saying *me/you* with pointing action directed towards the child. The addressee condition provided the child only with systematic opportunities to observe the parents saying *me/you* with pointing actions directed towards the child. The results revealed that only children in the non-addressee condition imitated their parents' pointing actions and use of *me/you* without errors, suggesting that even children under two years old can attend to and can learn from speech not addressed to them.

Children also learn other important aspects of language and social-cognitive skills from overheard speech (Forrester, 1993; for an overview see Rogoff, Paradise, Mejia-Arauz, Correa-Chavez, & Angelillo, 2003). Moreover, infants of 24 months (Akhtar, 2005; Akhtar, Jipson, & Callanan, 2001) and even 18 months of age (Floor & Akhtar, 2006) can learn novel words used in a third-party conversation. That infants listen to and learn from overheard speech is supported by the recent report that infants around 2 years of age are able to distinguish fluent, well-formed utterances from disfluent utterances in adult-directed speech (Soderstrom & Morgan, 2007). Similarly the phonetic system is influenced by overheard speech. Au, Knightly, Jun, and Oh (2002) examined Spanish productions of American adult learners of Spanish who had overhead Spanish during their childhood. They found that their production of the voiceless stop consonants was more native-like than typical late L2 learners.

Although previous studies have revealed that the acoustic characteristics of language-specific phonetic distinctions are exaggerated in infant-directed speech, there is mounting evidence that some phonetically relevant acoustics aspects of infant directed speech are more complex than would be predicted by facilitation. For example, a recent study of voice-onset time, the primary acoustic cue for voicing perception, showed that American English mothers speaking to 9-month-old infants exhibit more overlap between voiced and voiceless consonants than women speaking in an adult-directed register (Narayan, in press). These results are consistent with older studies examining voice-onset time in English (Baran, Laufer, &

Daniloff, 1977; Malsheen, 1980) and Swedish (Sundberg & Lacerda, 1999), suggesting that the acoustic cues may be less specified in IDS than ADS for some contrasts. The blurring of segmental contrasts in IDS is supported by more general intelligibility effects. Bard and Anderson (1994) showed that adult listeners performed only slightly better than chance when asked to identify individual words extracted from IDS. The implementation of phonetic contrast in IDS has also been shown to not differ significantly from its implementation in ADS. In her examination of singleton and geminate consonants in Lebanese Arabic child- and adult-directed speech, Khattab (2006) found no difference in the contrast of acoustic duration, the primary perceptual cue for consonant length, in the two registers. The level of acoustic complexity of IDS is further qualified by the age and gender of the child being addressed. For example, Baran and colleagues (1977) found that mothers' acoustic specification of consonant voicing became more adult when the infant was 12 months old, suggesting that the speech directed towards infants develops as the infant ages. Foulkes, Docherty, and Watt (2005) found that mothers were more likely to provide clear phonetic evidence for the standard variety of word-medial and word-final alveolar consonants when speaking to boys rather than girls, who received more instances of vernacular varieties of the contrast. van de Weijer (2001) observed that ID speech exhibits an enlarged vowel space in content words but a reduced vowel space in function words (which form the larger part of everyday speech - Cutler, 1993 -). That is, function words had a more expanded vowel space in adult-directed speech than in IDS. Bard and Anderson (1994) reported that word intelligibility was inversely related to word predictability: predictable words without their context were actually less intelligible in child-directed than in adult-directed speech. Finally, Kirchhoff and Schimmel (2005) explored the statistical properties of infant-directed versus adult-directed speech and found that class separability is actually poorer in IDS than in ADS.

Taken together, the emerging picture of the phonological significance of the infant-directed speech is one of complexity, with certain phonological contrasts being amenable to categorization by the learner and others being non-different from, or less consistent than the adult-directed register. Given the acoustic complexity of speech directed towards infants, the present study asks whether the acoustic clarity of phonetic contrasts in ADS is itself sufficient for the formation of phonological categories in the learner. While the relationship between acoustic-phonetic clarity and learning has inspired research on the differences between "clear" versus "conversational" adult-directed speech in L2 learning (e.g., Bradlow & Bent, 2001; Smiljanic & Bradlow, 2009), there have been almost no studies to date analyzing how the characteristics of adult-directed speech might match the learning propensities of the child during first-language acquisition. Our study begins to address that gap by measuring the acoustic properties of ADS to determine if the cues are available in ADS, as they are in IDS, to support distributional learning of phonetic categories.

The role of number and importance of cues to vowel distinctiveness

The cues that distinguish some Japanese and English vowel categories, namely duration and spectral properties, are not symmetrically weighted. In Japanese there are five short vowels [a] [i] [u] [e] [o], and each of them has a long vowel counterpart, [a:] [i:] [u:] [e:] [o:], with no accompanying spectral differences. In illustration, the words *kado* and *ka:do* are distinguished solely by a durational difference in the first vowel, but have different meanings (*corner* and *card* respectively). On the other hand, vowels in English are distinguished on the basis of their differences in vowel color, that is, primarily by their spectral properties. Length is a secondary cue that often accompanies the spectral differences in English vowels. For example, the vowels [eɪ] [iː] [aɪ] [əʊ] [juː] (e.g., the vowels in *bait, beet, bite*, *boat*, and *beauty*) differ from the vowels [æ] [ɛ] [I] [ɔː] [ʊ] (*bat*, *bet*, *bit*, *bought*, and *put*) not only in spectral properties, but often also in duration. In some instances, this is because vowels in the former set are diphthongs; in others, a difference in duration is the remnant of a historic length difference that has been replaced by a color difference.

Catalan differs from both Japanese and English. The Catalan inventory consists of seven vowels [i] [e] [ɛ] [a] [ɔ] [o] [u] plus [ə], a reduced vowel in unstressed position. Unlike English, there are thought to be no temporal differences accompanying vowel color contrasts in Catalan; thus, vowel duration is assumed to play no role in the distinction of vowel categories (Cebrian, 2006; Harrison, 1997). Hence, in the comparison of Japanese, English, and Catalan we have three languages that distinguish vowels in three different ways. In Japanese, duration alone can cue a vowel distinction; in Catalan color is the crucial cue, and finally in English, vowel color is the primary cue, but color categories can have duration as a secondary cue as well.

There are many factors that influence vowel length and color. Differences in vowel duration covary with vowel height such that low vowels (vowels that are produced with the tongue low in the mouth) have intrinsically longer durations than high vowels, due to the longer traveling time for tongue (-plus-jaw) to and from consonantal positions. The voicing of the surrounding consonants also influences vowel length: vowels tend to be longer before voiced consonants such as /b/, /d/, and /g/ than they are before voiceless consonants such as /p/, /t/, and /k/. Other factors that influence vowel length include emphatic stress, focus, position in an utterance, and affect (for an overview see Erickson, 2000). Similarly, pitch height and degree of pitch change affect vowel color (Trainor & Desjardins, 2002). All of these factors increase the acoustic variability of each individual vowel, and hence have a direct influence on the task of learning native vowel categories.

Summary of goals

The present study has two goals. The first goal is to examine ADS to determine whether overheard speech has the information that would be needed to support distributional learning of phonetic categories. If so, Japanese ADS should better predict two categories of vowel length than either English or Catalan ADS, and English and Catalan ADS should better predict two categories of vowel color than Japanese ADS. Because duration is a secondary cue to vowel color in English, the predictive power of duration in English ADS should fall between that in Japanese and Catalan. It is an empirical question whether Catalan or English ADS will better predict two categories of vowel color. Establishing the existence of identifiable, language-appropriate categories in adult-directed speech will contribute to the argument that distributional learning is a potential, viable mechanism by which infants can acquire these categories, and will extend our understanding of the types of information infants might retrieve from overheard speech.

The second goal is to explore the generalizability of the hypothesis that the cues in input speech are available to support distributional learning, irrespective of how a particular language chooses to contrast a particular pair of phones. As such, we examine the relative contribution of the phonetic cues for a specific vowel pair, *E-ee*, in Japanese, English and Catalan, three languages that use different combinations of cues to distinguish these vowels. In two of these languages, only a single cue (either duration or color) is used, whereas in the other language, both cues exist. Thus the results of this analysis will be informative for furthering our understanding of the interplay between number and weighting of cues for determining vowel category.

METHOD

Participants. The study was conducted in Vancouver, Canada; Keihanna, Japan; and Barcelona, Spain. A total of 40 participants (10 Japanese, 20 Canadian-English, and 10 Catalan) took part in the study. The Canadian and the Japanese participants were mothers who had previously been recorded in a different task interacting with their 12-month-old infants for the research reported in Werker et al. (2007). The reason for using twenty Canadian participants in the current study instead of ten was for consistency with this previous study, where it was necessary to record twice as many Canadian as Japanese mothers because the Canadian infants had difficulty sitting through a full recording session. The Catalan participants were Catalan-Spanish bilingual adult women, with Catalan as a dominant language. It is important to note that although participants were bilingual, they had no problem producing the two Catalan vowel contrasts since these vowels were from their native and dominant language.

The Japanese participants were recorded in NTT Communication Science Laboratories, the Canadian-English participants were recorded at the Infant Studies Center of the University of British Columbia, and the Catalan participants were recorded at GRNC (Grup de Recerca en Neurociència Cognitiva) at the University of Barcelona.

Apparatus. The participants were recorded in a quiet and comfortable room with an omni-directional lapel microphone. Utterances were recorded directly onto a Macintosh G3 computer (Japan and Canada) and a Pentium-III PC (Spain) using Sound Edit (version 2-99) software. The operating system was Mac OS9.2.2 in Japan and Canada, and Windows XP in Spain. Online monitoring of the sound quality and amplitude allowed the adjustments needed to ensure optimal sound quality.

Stimuli. We used a similar vowel pair across all three languages, which we will denote here as *E-ee*. In Japanese, these vowels are described as differing by only a length distinction (phonetically transcribed as [e]-[e:]). In English, these vowels differ phonologically by vowel color (phonetically transcribed as [ɛ]-[e]), but may also have an accompanying durational difference. The distinction for English is exemplified in the words *bet* and *bait*. In most dialects of English the vowel in *bait* is a diphthong, that is, it glides from one vowel /e/, to another, /i/, transcribed as [eI]; however, in Western Canadian English, the vowel in *bait* is often a monophthong, phonetically transcribed as [e] (Hagiwara, 2005; K. Russell, personal communication, December 10, 2002). In Catalan these vowels differ by only a color distinction (transcribed as [ɛ]-[e]).

A set of eight nonsense words was used, four words for each vowel. All words were phonotactically possible word-forms in Japanese, English and Catalan, and were created using a balanced combination of voiced and voiceless consonants before and after the target vowel to control for the known effect of voicing on vowel duration. Thus, all nonsense words were created using the four possible combinations of voiced and voiceless consonants before and after the target vowel. The eight nonsense words were the same in English and Japanese, but a new set of eight nonsense words was created for Catalan in order to make them phonotactically possible (see Appendix A).

Materials and recording procedure. Two tasks were employed. The first task consisted of sentence reading; the eight nonsense words were placed in three different sentences with the nonsense word occurring at the beginning, middle or end of the sentence, e.g., *The bayssa looks like a cloud, The politicians are showing the bayssa the results,* and *The electrician is fixing the bayssa* (the complete list of sentences appears in Appendix A). Participants were given a sheet of paper that contained different sentences and they were asked to read all of them. No training or previous instructions were given to the participants. Some participants first produced the target word in isolation. Other participants decided to repeat (without any instruction from the experimenter) some of the sentences. For the second task, participants played a puzzle board game with the researcher, in which the researcher could not see the board of the participant and each game piece was labeled with one of the nonsense words. The participant's task was to help the researcher solve a puzzle formed by the pieces, by providing verbal instructions as to how to move each piece to the correct location on the board. The purpose of the game-task was to elicit pronunciations of the nonsense words (and thus the target vowels) in spontaneous speech. Following the same conditions as those in Werker et al.'s (2007) study, 10 Japanese and 10 Catalan participants were recorded using all the nonsense words, whereas half of the English participants (10) were recorded using one set of words, and the other half were recorded using the other set, ensuring that each vowel type was produced by each subject.

Acoustic analyses. Acoustic analyses were performed using the software Praat version 4.2 (Boersma & Weenink, 2004). For each session, the sentences containing the target words, the target words themselves, and the target vowels were labeled by trained phoneticians. Target vowels that were potentially problematic for subsequent analyses (complicated by noise, breathiness, etc.) were not labeled. We used Praat scripting language routines to obtain the duration (in ms) of each labeled vowel, and the values of the first two formants (F1 and F2) in the first quarter portion of the labeled vowels.

In both the read speech and the spontaneous speech that we collected for this study, adults produced the target words primarily in sentence contexts (as reported in van de Weijer, 2002), but they also produced them in isolation. Prior to analyzing the entire data set, we compared the acoustic values of the vowels of each category in words spoken in isolation to those in ongoing utterances. There were no significant differences in spectral measurements. All vowels were longer in words spoken in isolation, but the relative differences across the vowel pairs remained. Thus the analyses below contain vowels both from words spoken in isolation and from those spoken in ongoing speech.

RESULTS

From the Japanese recordings, a range of 30-36 tokens from each subject in the reading condition, and a range of 15-20 tokens from each subject in the spontaneous condition were analyzed. From the English recordings, a range of 13-17 tokens from each subject in the reading condition, and a range of 11-20 tokens from each subject in the spontaneous condition were analyzed. From the Catalan recordings, a range of 32-34 tokens from each subject in the reading, and a range of 23-27 tokens from each subject in the spontaneous condition were analyzed. Table 1 gives the means for spectral and duration measures for all three languages in both the read and spontaneous speech conditions; Figures 1 and 3 show scatter plots of F1 and F2 for all three languages, in read and spontaneous speech, respectively; and Figures 2 and 4 plot the distributions of the tokens in each language with regard to duration in read and spontaneous speech, respectively.

Table 1. Spectral and duration values for Japanese, English, and Catalan /E/ - /ee/, in read and spontaneous speech.

/E/	Duration (ms)		Color (F2 – F1) Hz	
	Read	Spontaneous	Read	Spontaneous
Japanese	76	83	1699	1702
English	92	94	1315	1297
Catalan	97	109	1565	1521
/ee/	Duration (ms)		Color (F2 – F1) Hz	
	Read	Spontaneous	Read	Spontaneous
Japanese	177	203	1749	1561
English	143	139	1963	1930
Catalan	98	99	1801	1744

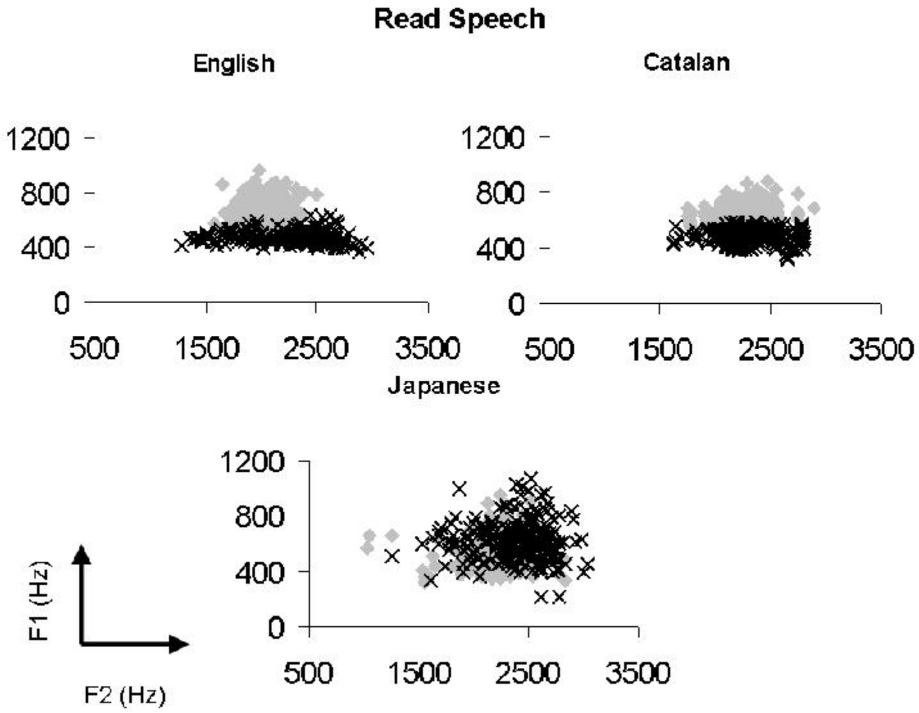

Figure 1. F1 – F2 Scatter plots for all items for both /E/ and /ee/ in Japanese, English and Catalan, read speech.

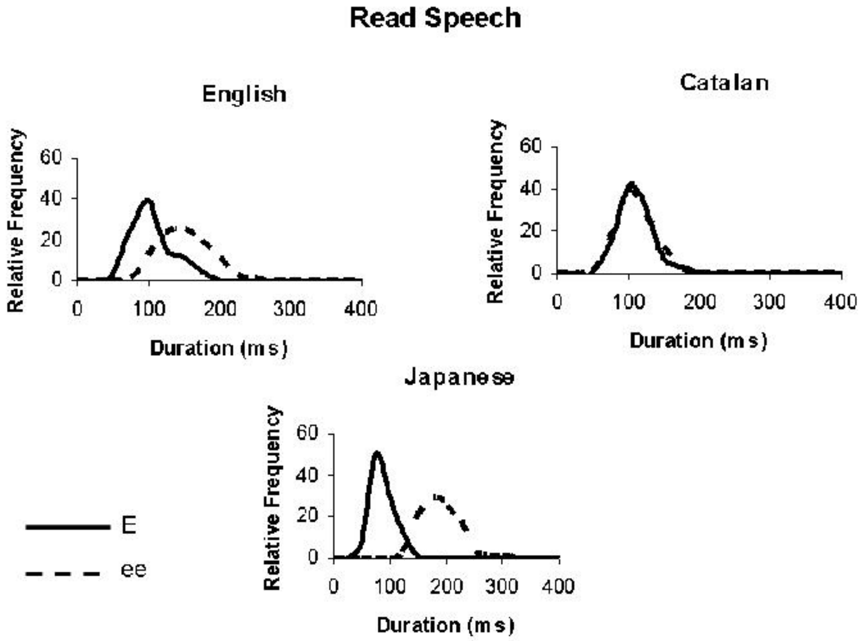

Figure 2. Relative frequencies of the vowel durations for /E/ - /ee/ in Japanese, English, and Catalan, read speech, illustrated in 50 ms bins.

To determine whether the distributions were significantly different, we compared the means of the two vowels for two measures. The first set of analyses focused on vowel length by comparing vowel duration. The second set of analyses focused on vowel color by comparing the joint contribution of F1 and F2. In reminder, it was predicted that vowel length would be more bimodal in the speech of Japanese speakers than in the speech of English speakers or Catalan speakers, and that vowel color would be bimodal in the speech of English and Catalan subjects, but unimodal in the speech of Japanese subjects.

Analysis of variance –ANOVA–

Results for read speech - Duration. The mean duration for each vowel was analyzed in a (vowel: /E/ vs. /ee/) X (language: Japanese, English, Catalan) mixed ANOVA. There was a significant effect of vowel, *F*(1,37) = 303.013 p < .001, meaning that overall /ee/ was longer that /E/, and a significant interaction between language and vowel, *F*(1,37) = 81.716

p < .001, The interaction is accounted for by a greater difference in means when comparing each of the three languages. In order to specifically explore these durational differences between each language a 2x2 mixed ANOVA using vowel /E/ vs /ee/ and two languages was performed. The results demonstrated that, comparing English and Japanese, there was a significant effect of vowel, *F*(1,28) = 382.169 p < .001, and a significant interaction between language and vowel, *F*(1,28) = 41.845 p < .001. The interaction is accounted for by a greater difference in means in Japanese than in English. For Japanese and Catalan a significant effect of vowel was also found, *F*(1,18) = 418.427 p < .001, as well as a significant interaction between language and vowel, *F*(1,18) = 405.608 p < .001. The interaction is accounted for by a greater difference in means in Japanese than in Catalan. Finally in English vs. Catalan, we observed a significant effect of vowel, *F*(1,28) = 53.616 p < .001, and also a significant interaction between language and vowel, *F*(1,28) = 50.398 p < .001. This last interaction is accounted for by a greater difference in means in English than in Catalan.

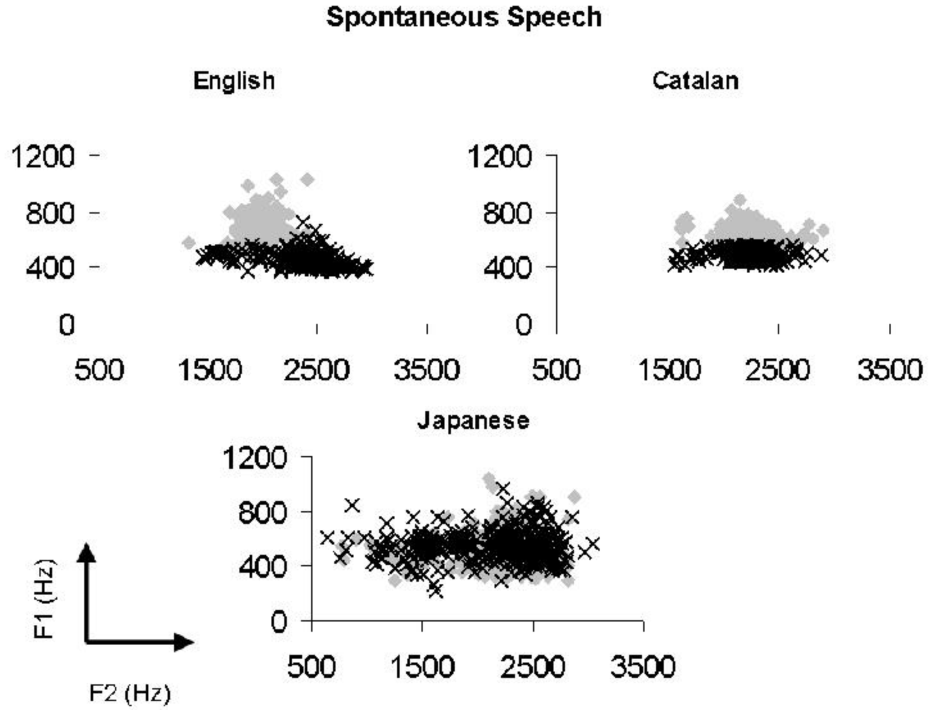

Figure 3. F1 – F2 Scatter plots for all items for both for /E/ and /ee/ in Japanese, English and Catalan, spontaneous speech.

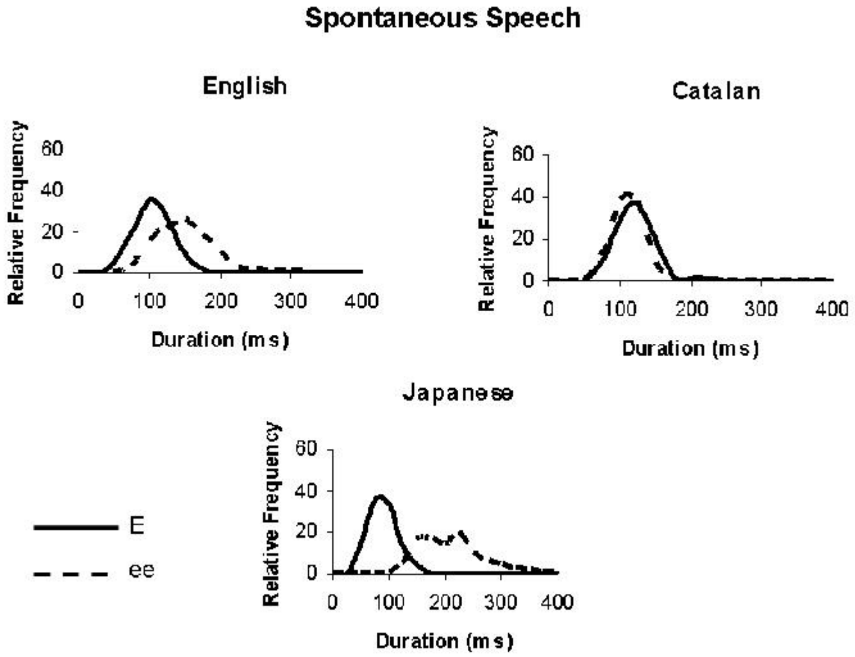

Figure 4. Relative frequencies of the vowel durations for /E/ - /ee/ in Japanese, English, and Catalan, spontaneous speech, illustrated in 50 ms bins.

Results for read speech - Color. To examine the prediction that there would be a bimodal distribution of vowel color in English and a unimodal distribution in Japanese, an analysis comparing the combined effects of the first (F1) and second (F2) formants was done. The difference between F2 and F1 was used as a measure for the analyses (F2-F1 difference is a standard measure of vowel color; Ladefoged, 1993).

There was a significant effect for vowel, $F(1,37) = 114.897$ $p < .001$, and a significant interaction between language and vowel, $F(1,37) = 47.195$ $p < .001$. The interaction is accounted for by a greater difference in F2-F1 when comparing each of the three languages. As before, a 2x2 mixed ANOVA using vowel /E/ vs /ee/ and two languages was run in order to specifically explore these differences between each language. The results demonstrated that, comparing English and Japanese, there was a significant effect of vowel, $F(1,28) = 86.813$ $p < .001$, and a significant interaction between language and vowel, $F(1,28) = 68.644$ $p < .001$, meaning that vowels in English differed more in color than in Japanese. For Japanese and

Catalan, the same pattern of results was observed: a significant effect of vowel was also found, $F(1,18) = 12.443$ $p < .005$, as well as a significant interaction between language and vowel, $F(1,18) = 6.525$ $p < .05$, that is, the vowels in Catalan differed more in color than in Japanese. Finally for English and Catalan, we observed a significant effect of vowel, $F(1,28) = 230.044$ $p < .001$, and a significant interaction between language and vowel, $F(1,28) = 49.436$ $p < .001$. This interaction revealed that in English, vowels differed more in color, than they did in Catalan.

Results for spontaneous speech. Similar results were obtained for the spontaneous speech condition.

Results for spontaneous speech - Duration. There was a significant main effect for vowel $F(1,37) = 147.531$ $p < .001$, and a significant interaction between language and vowel $F(1,37) = 61.693$ $p < .001$. A 2x2 mixed ANOVA using vowel /E/ vs /ee/ and two languages demonstrated that, comparing English and Japanese, there was a significant effect of vowel, $F(1,28) = 220.403$ $p < .001$, and a significant interaction between language and vowel, $F(1,28) = 41.473$ $p < .001$. For Japanese and Catalan, a significant effect of vowel was also found, $F(1,18) = 89.812$ $p < .001$, and a significant interaction between language and vowel, $F(1,18) = 121.415$ $p < .001$. Finally in English *vs.* Catalan we observed a significant effect of vowel, $F(1,28) = 19.284$ $p < .001$, and a significant interaction between language and vowel, $F(1,28) = 41.708$ $p < .001$.

Results for spontaneous speech - Color. There was a significant effect for vowel $F(1,37) = 80.747$ $p < .001$, and a significant interaction between language and vowel $F(1,37) = 77.934$ $p < .001$. A 2x2 mixed ANOVA using vowel /E/ vs. /ee/ and two languages demonstrated that, comparing English and Japanese, there was a significant effect of vowel, $F(1,28) = 51.383$ $p < .001$, and a significant interaction between language and vowel, $F(1,28) = 118.951$ $p < .001$. For Japanese and Catalan, there was no effect of vowel, but there was a significant interaction between language and vowel, $F(1,18) = 23.896$ $p < .001$. Finally in English and Catalan we observed a significant effect of vowel, $F(1,28) = 252.856$ $p < .001$, and a significant interaction between language and vowel, $F(1,28) = 57.634$ $p < .001$.

Analytic strategy

The ANOVA provides a descriptive portrayal of how the cues differ when the vowel category labels are provided. But in order to determine whether the cues in adult-directed speech are sufficient to allow category learning, and in particular to determine how different types of cues can serve to simultaneously discriminate among vowel category membership, it is essential to conduct an analysis wherein the categories are the dependent variable rather than the independent variable. This is more akin to the infant learner who does not know, when she encounters a word, whether the vowel should be categorized as an /ee/ or an /E/, and which cues – duration and/or color – they should pay attention to. Thus a second set of analyses was conducted using the data to predict categories.

In previous studies, a discriminant function analysis has been used to examine the characteristics of linguistic input (Assmann & Katz, 2005; Hillenbrand, & Nearey, 1999; Hillenbrand, Clark, & Nearey, 2001; Morrison, 2008). Such an analysis considers the entire data set and determines the optimal number of categories given the best fit from the data set. In our case a discriminant function analysis was not appropriate as we needed to consider the input from multiple speakers in each language. That is, to examine the relationship between color and duration in predicting vowel category as a function of language, it was necessary to estimate the variation at both the level of language and the level of the individual speaker. A discriminant function analysis is not able to simultaneously include the variance at both of these levels. Logistic regression provides the same essential information as a discriminant function analysis with less restrictive assumptions on the nature of the predictors (e.g., see Meyers, Gamst, & Guarino, 2006). Most importantly, logistic regression is extendable, as in the present analysis, to multilevel designs where multiple observations are available for each participant and consequently the between- and within-person relationships can be simultaneously estimated. In reminder, the theoretical question this analysis was designed to address was whether the acoustic cues of color and duration were better at predicting the categories in one language vs. the others. Thus the logistic regression was designed to compare the relative strength of each cue in predicting category among pairs of languages.

The specific model examined was:

$$Y_{ij} = \beta_{0i} + \beta_{1i}Duration + \beta_{2i}Color + \varepsilon_{ij} \qquad \text{Level 1}$$

$$\begin{aligned} \beta_{0i} &= \beta_{00} + \beta_{01}L1 + \beta_{02}L2 + u_{0i} \\ \beta_{1i} &= \beta_{10} + \beta_{11}L1 + \beta_{12}L2 + u_{1i} \quad . \qquad\qquad \text{Level 2} \\ \beta_{2i} &= \beta_{20} + \beta_{21}L1 + \beta_{22}L2 + u_{2i} \end{aligned}$$

Here Y_{ij} is the log odds of correct vowel classification for /ee/ for person *i* on trial *j*; *Duration* is length in milliseconds, and *Color* is the F2-F1 formant difference. Because both *Color* and *Duration* are expected, theoretically, to be present in English (albeit with *Color* as the expected dominant predictor), English was used as the reference group. To compare the other two languages to English, the dummy codes L1 and L2 were assigned to the Catalan and Japanese languages, respectively. Consequently β_{10} and β_{20} are the partial regression coefficients respectively for *Duration* and *Color* in predicting vowel category for English. The tests of the coefficients for the language dummy codes represent tests of the difference in the relationship between *Duration* and *Color* in predicting vowel classification when compared to either L1 (Catalan) or L2 (Japanese). For example, when compared to L1, β_{11} and β_{21} thus represent tests of the difference in the relationship between *Duration* and *Color* in predicting vowel classification between English and Catalan.

In logistic regression the magnitude of the odds ratio effect size is a function of the scaling of the continuous predictors. We present odds ratios where the units are 50ms for duration and the F2-F1 formant difference is 500Hz.

Results for read speech - Duration. Controlling for color, on average across participants, a 50ms increase in duration was significantly associated with vowel classification in English (odds ratio = 98.26, z = 4.50, p < .0001). However, the relationship between *Duration* and vowel classification was significantly weaker for Catalan (odds ratio = 1.51, test of interaction z = -3.93, p = .00008) and stronger in Japanese (odds ratio = 51844, test of interaction z = 2.17, p = .03). Figure 5 shows the predicted probability of vowel classification as a function of duration, holding color constant. At the bottom of the graph is a non-parametric kernel density estimate of the probability density function (essentially a smoothed histogram of the distribution of duration values), provided separately for each language.

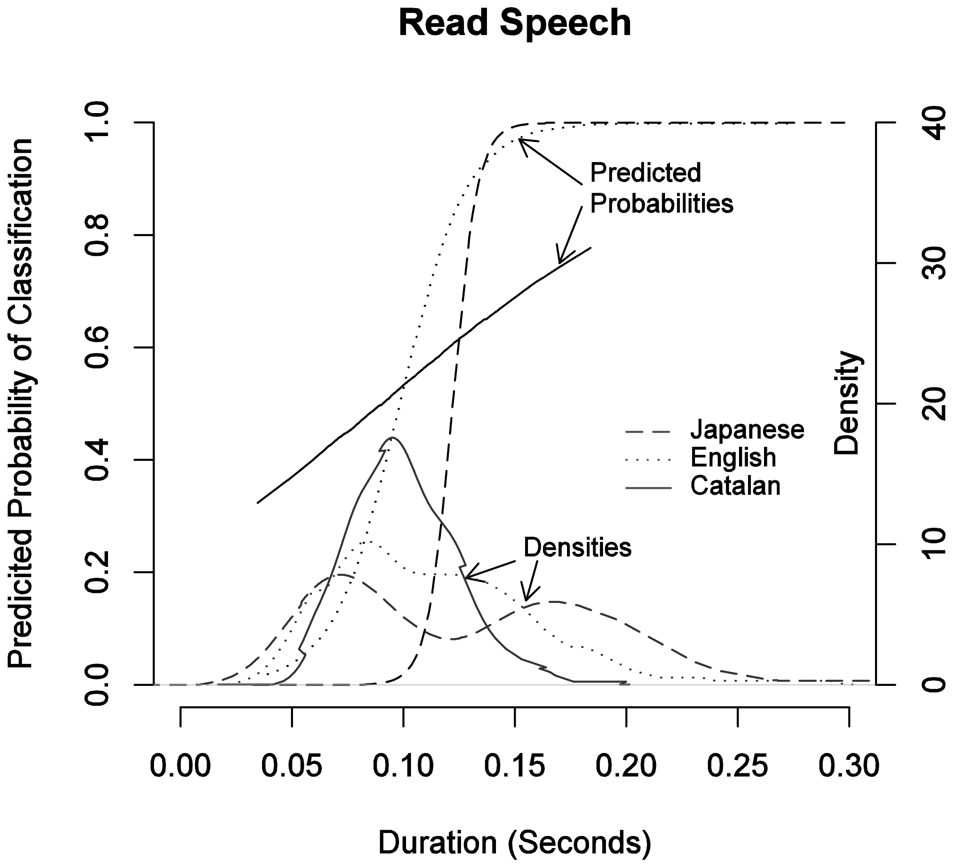

Figure 5. Predicted probability of correct vowel classification based on the multilevel logistic regression for duration after partialling color in the Reading condition (averaged across mothers). Steeper slopes indicate better discrimination between the vowels. The non-parametric kernel density estimate of the probability density function of duration within each language is also shown.

Results for read speech - Color. Partialling duration, on average across participants, color was significantly associated with vowel category in English (odds ratio = 6940, z = 5.75, p < .00001). The relationship between color and vowel category in Catalan did not differ from that in English (odds ratio = 11556, z = 0.30, p = .76). In contrast, the relationship between color and vowel category was significant weaker for Japanese relative to English (odds ratio = 5.48, test of interaction z = -3.57, p = .0004). Figure 6 shows the predicted probability of vowel classification as a function of color, holding duration constant. At the bottom of the graph is a non-parametric kernel density estimate of the probability density function (essentially a smoothed histogram of the distribution of color values), provided separately for each language.

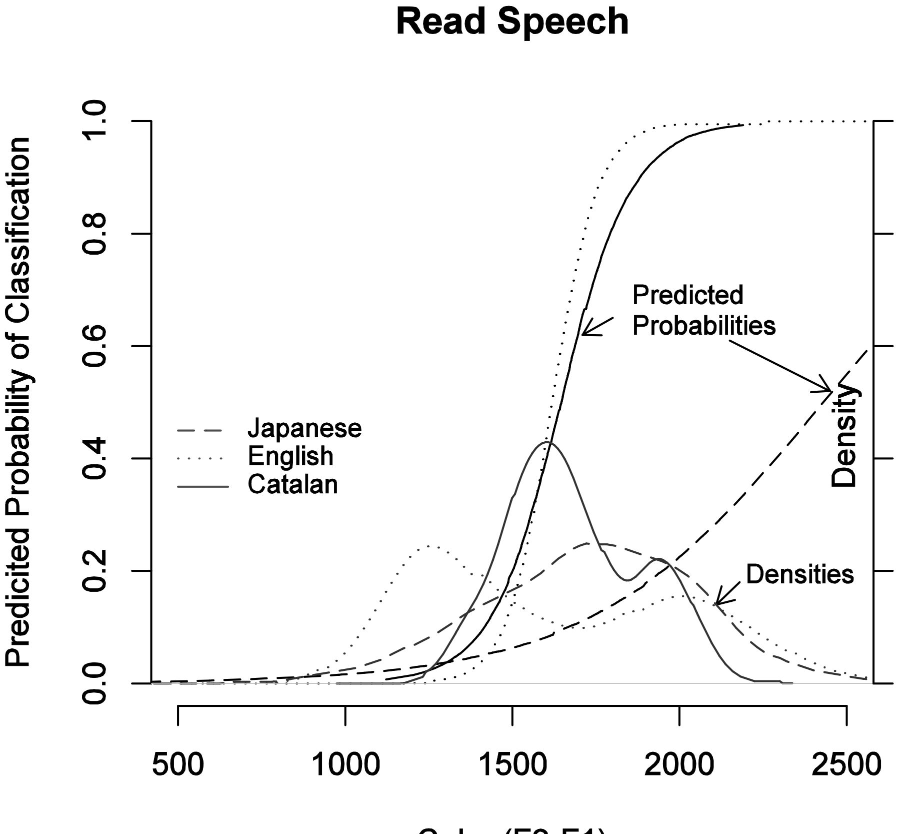

Figure 6. Predicted probability of correct vowel classification based on the multilevel logistic regression for color after partialling duration in the Reading condition (averaged across mothers). Steeper slopes indicate better discrimination between the vowels. The non-parametric kernel density estimate of the probability density function of color within each language is also shown.

Results for spontaneous speech. Overall, the pattern of results for spontaneous speech fully mirrored that for read speech.

Results for spontaneous speech - Duration. Controlling for color, on average across participants, a 50ms increase in duration was significantly associated with vowel classification in English (odds ratio = 37.05, $z = 5.43$, $p < .0001$). However, the relationship between Duration and vowel classification was significantly weaker for Catalan (odds ratio = 0.35, test of interaction $z = -6.30$, $p < .0001$) and stronger in Japanese (odds ratio = 328.3, test of interaction $z = 2.14$, $p = .03$). Figure 7 shows the predicted probability of vowel classification as a function of duration, holding color constant. At the bottom of the graph is a non-parametric kernel density estimate of the probability density function (essentially a smoothed histogram of the distribution of duration values), provided separately for each language.

Results for spontaneous speech - Color. Partialling duration, on average across participants, color was significantly associated with vowel category in English (odds ratio = 814.85, $z = 7.49$, $p < .00001$). The association between color and vowel category in Catalan did not differ from that in English (odds ratio = 159.25, $z = -1.59$, $p = .11$). In contrast, the relationship between color and vowel category was significant weaker for Japanese relative to English (odds ratio = 0.62, test of interaction $z = -7.59$, $p < .0001$). Figure 8 shows the predicted probability of vowel classification as a function of color, holding duration constant. At the bottom of the graph is a non-parametric kernel density estimate of the probability density function (essentially a smoothed histogram of the distribution of color values), provided separately for each language.

Which is more important, duration or color?

To examine this question we first standardized both color and duration based on all responses (both between and within mothers) separately for each language. This expresses the units of each of these predictors in terms of standard deviations and allows comparison of the regression coefficients for these different predictors. Constraining these two standardized predictors to have equal regression coefficients significantly reduced model fit for all three languages. We report here the results for spontaneous language noting that results for read language were equivalent. For both English and Catalan, color had a stronger relationship than duration, $\chi^2(1) = 12.37$, $p=.0004$ and $\chi^2(1) = 153.8$, $p<.00001$, respectively.

In contrast, for Japanese duration was a stronger predictor than color, $\chi^2(1) = 383.4, p < .00001$.

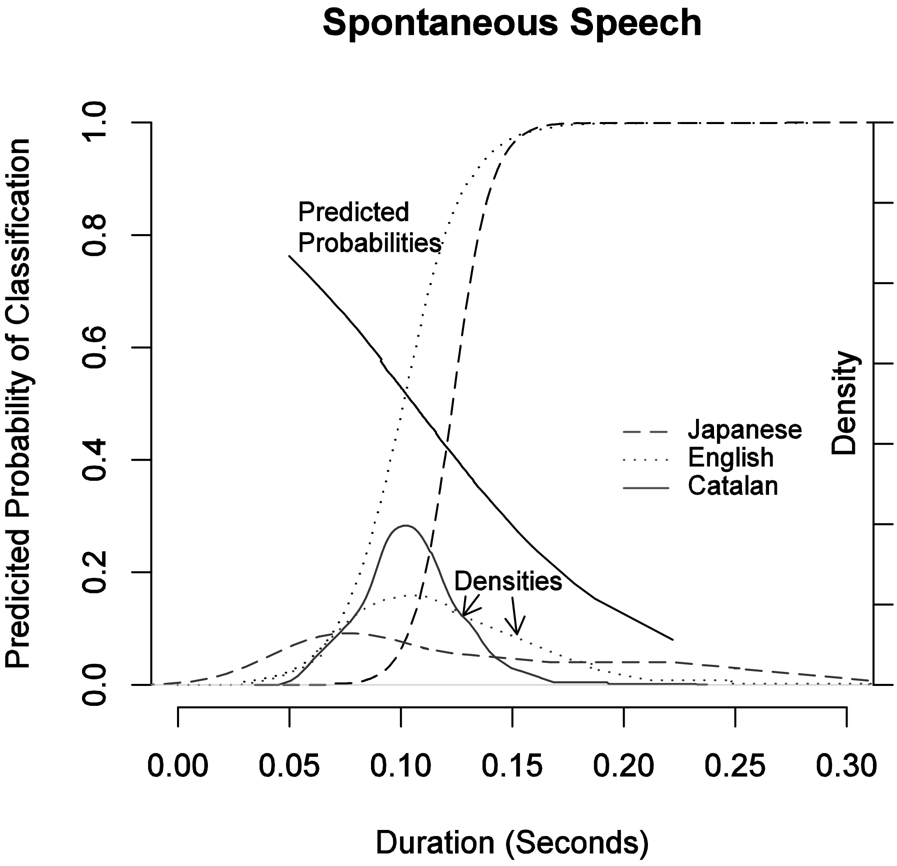

Figure 7. Predicted probability of correct vowel classification based on the multilevel logistic regression for duration after partialling color in the Spontaneous condition (averaged across mothers). Steeper slopes indicate better discrimination between the vowels. The non-parametric kernel density estimate of the probability density function of duration within each language is also shown.

Are there differences in the predictive relationships for duration and color between read and spontaneous language?

We examined whether the type of language, read or spontaneous, interacted with either color or duration in predicting vowel category. The only significant interactions that emerged were that duration was a weaker predictor for spontaneous language in Japanese, *z*=-1.96, *p*=.05, as well as Catalan, *z*=-4.39, *p*<.001, but not for English, *z*=0.30, *ns*. Type of speech did not moderate the ability of color to discriminate vowel category for any of the three languages, all *z*'s <1.59, *ns*.

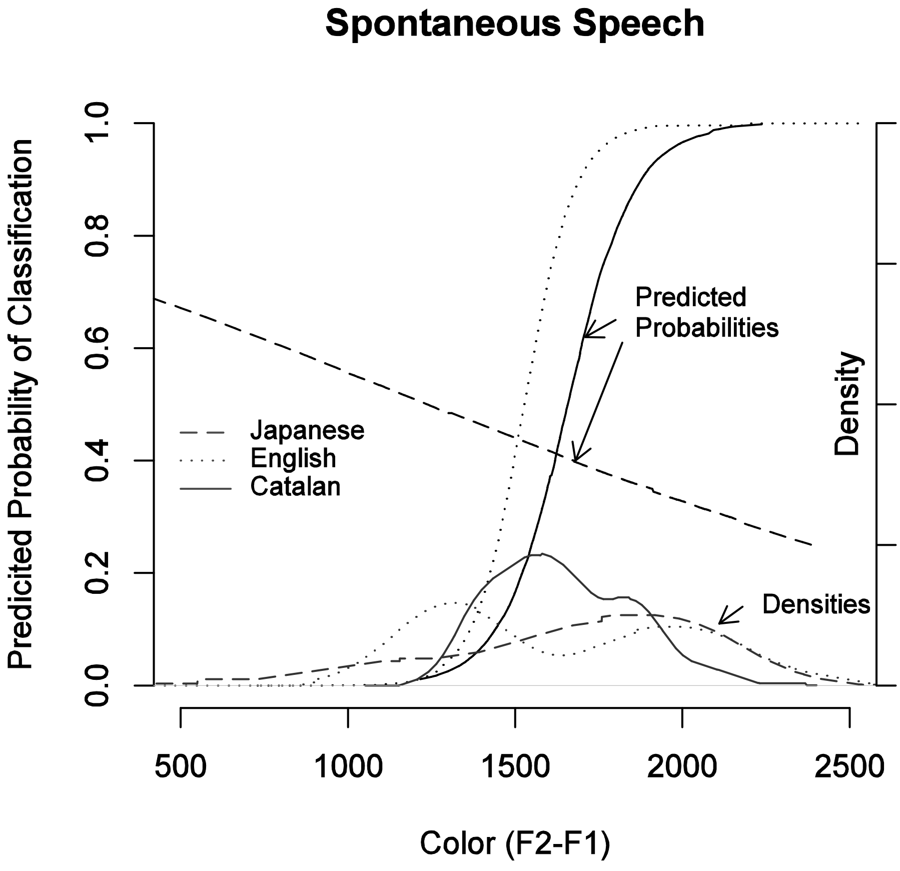

Figure 8. Predicted probability of correct vowel classification based on the multilevel logistic regression for color after partialling duration in the Spontaneous condition (averaged across mothers). Steeper slopes indicate better discrimination between the vowels. The non-parametric kernel density estimate of the probability density function of color within each language is also shown.

DISCUSSION

One of the most intractable problems in language acquisition is how listeners, particularly prelinguistic infants, can induce the phonetic category structure of their native language without a priori knowledge of the acoustic-phonetic dimensions on which one meaningful word, such as *bait,* differs from another, such as *bet*, and without knowledge of which words contrast in meaning. It has been suggested that a distributional learning mechanism may provide an entry into this otherwise unmanageable perceptual learning problem. However, for distributional learning to be successful, the information has to be available in the highly variable input speech. The research reported here advances our knowledge of the information available in two important ways. First, it provides a test of whether or not there are cues in adult-directed speech that could, in principle, support the distributional learning of language-specific categories by infants hearing such speech. Second, it provides a between-language comparison of the relative importance of vowel duration and color to vowel category assignment by comparing three languages that differ in the use of these cues. Our findings that the information in input speech is sufficient to describe the appropriate categories in each language increase the plausibility of the claim that infants may indeed be able to use a distributional-based approach to learn the phonetic categories of the native language. Thus although our results do not necessarily rule out a contribution from possible initial biases (Polka & Bohn, 2003; 2011), or eliminate other types of learning mechanisms, they add strong support to the feasibility of frequency-based models of phonetic learning (Kuhl, 1993; Maye et al., 2002; Werker & Curtin, 2005). Finally, although the current study has only focused on vowels, the specificity of these results cannot be accounted for by the different function of vowels and consonants in language acquisition (Nespor, Peña, & Mehler, 2003; Pons & Toro, 2010) as distributional learning has also been observed with consonants (Maye et al., 2002; 2008).

Cues in adult-directed speech

The results of this study provide clear evidence that, as they are in IDS, the spectral and durational cues available in ADS have the properties that are required to support phonetic category learning. This adds to our appreciation of the richness of the language input potentially available for young infants who are in the process of establishing their native language phonetic categories. The finding that relevant cues are robustly available not only in IDS but also in overheard speech strengthens the proof in principle that distributional learning is one plausible mechanism for category formation.

It is interesting to note that, although the relevant cues for language-specific vowel categories are present in both IDS and ADS, they are not necessarily manifested in the same way. A comparison of absolute durations and F2 – F1 values for /E/ and /ee/ in English and Japanese IDS and ADS reveals that, in English, the values for color in IDS and ADS tend not to differ, while those for duration do (Table 2). This is consistent with robust findings of elongated vowel production in English IDS (e.g., Kuhl et al., 1997). In Japanese, on the other hand, the values for color do differ, while duration is the same for the short vowel, but not for the long vowel. The limited nature of the data examined restricts the conclusions that can be drawn; however, the Japanese data suggest that short vowels may be produced within a consistent durational range, while long vowels may be subject to elongation processes similar to those at work in English IDS. Color differences in Japanese may reflect the sort of exaggeration of vowel forms reported by Kuhl and collaborators (1997). Variations in values for duration and color parameters in read and spontaneous speech in IDS and ADS need to be examined with a much broader data set. Despite these differences (and similarities) in IDS and ADS for both Japanese and English, we have shown that Japanese, English and Catalan ADS nonetheless contain cues to language-appropriate vowel categories, just as English and Japanese IDS do (Werker et al., 2007).

Table 2. Average values for duration and color measures for Japanese and English, read and spontaneous /E/ - /ee/. The asterisk denotes *p* < 0.05 for the comparison of IDS and ADS values in each category (IDS values are taken from Werker et al., 2007).

		English read		English spont.		Japanese read		Japanese spont.	
		IDS	ADS	IDS	ADS	IDS	ADS	IDS	ADS
/E/	Duration (ms)	119	91 *	110	94 *	85	76	86	83
	Color (F2-F1)	1341	1315	1353	1297	1575	1699 *	1672	1701 *
/ee/	Duration (ms)	180	143 *	157	139 *	205	177 *	169	203 *
	Color (F2-F1)	1869	1963	1878	1929	1575	1749 *	15	1560 *

Although there is considerable evidence that infants pay more attention to IDS (Cooper & Aslin, 1990; Fernald, 1985; Werker & McLeod, 1989), and some evidence that they may be better able to extract statistical regularities from IDS over ADS (Thiessen et al., 2005), there is increasing evidence that infants can also learn from overheard ADS (Ahktar, 2005; Oshima-Takane, 1988). Laboratory studies of distributional learning have used single syllable, monotone tokens, presentations that more closely resemble ADS than IDS, and have successfully induced category change (Maye et al., 2002, 2008; Yoshida et al., 2010), suggesting that the vowel category cues in ADS can in principle drive distributional learning. What remains unexplored is whether infants are able, in fact, to use the distributional information in overheard adult speech to support phonetic category learning. There is some evidence that phonetic learning requires engagement and contingent vocal interactions (Kuhl, Tsao & Liu, 2003). Yet, it is unknown if distributional learning requires explicit attention. It has been demonstrated that implicit learning can be facilitated by attention, and failing to attend to a stimulus severely impairs implicit learning (Stadler, 1995; Toro, Sinnett & Soto-Faraco, 2005). Thus in future work it will be of interest to compare the impact of overheard vs. attended speech on distributional learning.

Language-specific cues to vowel category

In reminder, duration is the primary cue for distinguishing /ee/ and /E/ in Japanese; color is the primary cue in Catalan, and English distinguishes the vowels first by color but secondarily by duration.

The logistic regression revealed that even without prelabelling and across variations among speakers, and sentence and voicing contexts, duration is sufficient to derive two categories in English. Consistent with classic linguistic descriptions, our analyses revealed that duration alone is significantly better at predicting two categories in Japanese than it is in English, and also significantly better at predicting two categories in English than it is in Catalan. Similarly, without prelabelling, spectral cues alone also predict two categories in English. Moreover, in comparison to English, spectral cues were equivalent at predicting two categories in Catalan and significantly worse at predicting two categories in Japanese.

These comparisons provide a first step toward illuminating the relative roles played by primary and secondary cues to phonetic categories. Duration is only a secondary cue to vowel category in English, so it is not surprising that it is computationally less predictive of vowel category in English than in Japanese, and that color is more predictive than duration in

English. Where this analysis yields new insight is in the relative informativeness of vowel color in English and Catalan. Vowel color as carried by spectral properties is the primary cue to vowel category in both English and Catalan. The two languages differ, however, in that color is the *only* distinguishing property in Catalan, whereas English has duration available as a secondary cue. This raises the interesting question of whether color is equally informative in English and Catalan. Our results show that, in a logistic regression, when the categories are not pre-specified, spectral cues are equally good at generating the two appropriate categories in both English and Catalan.

We can conclude from this result that primary cues are indeed primary. Vowel color is equally good at yielding the two appropriate categories in both Catalan where it is the only cue, and English where it is accompanied by a secondary duration cue. There was no difference between Catalan and English in the predictive value of color in the logistic regression. This reveals that even when secondary cues are available, primary cues alone are sufficient to predict language-specific phonetic categories. On the other hand, secondary cues are indeed secondary. As a secondary cue in English, duration is neither as distinct between the two given categories nor as predictive of categories as is the primary cue of vowel color in the logistic regression. Furthermore, duration is not as predictive when it is a secondary cue in English as it is when it is a primary cue in Japanese.

New directions

We have been focusing on the availability of information in overheard speech that could support distribution-based learning of phonetic categories by infants, but it is of interest to step back and consider the role that distributional cues might play in maintaining and sharpening phonetic categories in adults. In principle we might expect that experienced adult native speakers could tolerate variability in the cues that define phonetic categories, but in fact our results show that the speech between two fluent adults is highly regular. Perhaps this regularity helps to maintain parity in the precise characteristics of phonetic categories within language communities. We know that when adult Portuguese-English bilinguals move between Brazil and the United States, their voicing categories shift (Sancier & Fowler, 1997) in accord with the values of the current language community. Our results show that in their speech to one another, two speakers from the same language and dialect provide constant affirmation of shared phonetic features. Hence distributional information may not only be used to drive the initial acquisition of native phonetic categories in infancy, but may contribute as well to maintaining those categories in adulthood.

RESUMEN

Claves para la categorización fonética en el habla dirigida a adultos: Evidencia de tres lenguas con características vocálicas distintas. Utilizando un lenguaje artificial Maye, Werker, y Gerken (2002) demostraron que las categorías fonéticas de los bebés cambian en función de la distribución de los sonidos del habla. En un estudio reciente, Werker y cols. (2007) observaron que el habla dirigida a bebés (habla maternal) contiene claves acústicas fiables que sustentan el aprendizaje de las categorías vocálicas propias de la lengua: las pistas en inglés eran espectrales y de duración; las pistas en japonés eran exclusivamente de duración. En el presente estudio se amplían estos resultados de dos formas, 1) examinamos una nueva lengua, el catalán, que distingue las vocales únicamente a partir de las diferencias espectrales, y 2) ya que los bebés aprenden también del habla dirigida a los adultos (Oshima-Takane, 1988), analizamos este tipo de habla en las tres lenguas. Los análisis revelaron diferencias considerables en las pistas de cada lengua, e indicaron que, por sí solas, son suficientes para establecer las categorías vocálicas específicas de cada lengua. Esta demostración de las diferencias propias de cada lengua en la distribución de las pistas fonéticas presentes en el habla dirigida al adulto, proporciona evidencia adicional sobre el tipo de pistas que pueden estar usando los bebés cuando establecen sus categorías fonéticas maternas.

REFERENCES

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Sci*ence, *8 (2)*, 199–209.

Akhtar, N., Jipson, J., & Callanan, A. (2001). Learning words through wverhearing. *Child Development*, *72 (2)*, 416–430.

Anderson, J. L, Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language & Speech*, *46(2)*, 155-182.

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants, in *Signal to Syntax,* edited by J. L. Morgan and K. Demuth (pp. 117-134). Mahwah, NJ: LEA

Assmann, P. F., & Katz, W. F. (2005). Synthesis fidelity and time-varying spectral change in vowels. *Journal of the Acoustical Society of America*, *117*, 886-895.

Au, T. K., Knightly, L. M., Jun, S. A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, *13(3)*, 238-243.

Baran, J. A., Laufer, M. Z., & Daniloff, R. (1977). Phonological contrastivity in conversation: a comparative study of voice onset time. *Journal of Phonetics, 5*, 339-350.

Bard, E. G., & Anderson, A. H. (1994). The unintelligibility of speech to children: effects of referent availability. *Journal of Child Language, 21*, 623-648.

Bernstein Ratner, N. (1984). Patterns of vowel modification in motherese. *Journal of Child Language*, *11*, 557-578.

Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language & Speech*, *46(2–3)*, 183–216.

Boersma, P., & Weenink, D. (2004). *Praat: Doing phonetics by computer (Version 4.3.02)*. [Computer program] (University of Amsterdam, Amsterdam, The Netherlands).

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language & Speech, 46*, 217-244.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. Journal of the Acoustical Society of America, *112 (1),* 272-284.

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28(3)*, 455-474.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*, 372-387.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech within the first month after birth. *Child Development*, *61*, 1584–1595.

Cutler, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences, *Journal of Psychological Research, 22*, 109-131.

de Boer, B., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustic Research Letters Online (ARLO)*, *4*, 129–134.

Eimas, P. D., Siqueland, E. D., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.

Erickson, M. L. (2000). Simultaneous effects on vowel duration in American English: A covariance structure modeling approach. *Journal of the Acoustical Society of America, 108(6)*, 2980–2995.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behaviour and Development*, *8*, 181–195.

Floor, P., & Akhtar, N. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy*, *9(3)*, 327-339.

Forrester, M. A. (1993). Affording social–cognitive skills in young children: the overhearing context, in *Critical influences on child language acquisition and development Edited by* D. J. Messer and G. J. Turner (pp. 40–61). New York: St Martin's Press.

Foulkes, P., Docherty, G. J. & Watt, D. J. L. (2005). Phonological variation in child directed speech. *Language, 81*, 177-206.

Hagiwara, R. (2005). Visualizing the Canadian English vowels. Paper presented *at the Acoustical Society of America/ Canadian Acoustical Association meeting*. Vancouver, Canada.

Harrison, P. (1997). The relative complexity of Catalan vowels and their perceptual correlates. *UCL Working Papers in Linguistics The Internet Edition, 9*.

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America, 109*, 748-763.

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hvd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America, 105*, 3509-3523.

Karzon, R. G. (1985). Discrimination of polysyllabic sequences by one-to-four-month-old infants. *Perception & Psychophysics*, *232*, 105-109.

Khattab, G. (2006). Does child-directed speech really facilitate the emergence of phonological structure? The case of gemination in Arabic CDS. Presented at *the 10th Laboratory Phonology Conference*, Paris.

Kirchhoff, K., & Schimmel, S. (2005). Statistical properties of infant-directed versus adult-directed speech: Insights from speech recognition. *Journal of Acoustical Society of America*, *117 (4)*, 2238-2246.

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*, *21(1)*, 125-139.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831-843.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolvarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*, 684-686.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show facilitation for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*, 13-21.

Kuhl, P. K., Tsao F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceeding of the National Academy of Sciences*, *100*, 9096–9101.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants six months of age. *Science*, *255*, 606-608.

Ladefoged, P. (1993). *A course in Phonetics* (New York: Harcourt Brace Jovanovich College Publishers).

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

Malsheen, B. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In: G. Yeni-Komishan, J. Kavanaugh, & C. Ferguson (Eds). *Child Phonology, volume 2: Perception.* Academic Press, San Diego.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*, 122-134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82(3)*, B101–B111.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied Multivariate Research: Design and Interpretation* (Thousand Oaks: Sage Publications.)

Morrison, G. S. (2008). Complexity of acoustic-production-based models of speech perception. *Proceedings of Acoustics'08* (pp. 2369–2374). Paris: Société Française d'Acoustique.

Narayan, C. R. (in press). Developmental perspectives on phonological typology and sound change. In: A. C. L. Yu (Ed). *Origins of Sound Patterns: Approaches to Phonologization*. Oxford University Press.

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science, 13(3),* 407-420.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio, 2*, 203-230.

Oshima-Takane, Y. (1988). Children learn from speech not addressed to them: the case of personal pronouns. *Journal of Child Language*, *19*, 111–131.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences, 106(26)*, 10598-10602.

Pons, F., & Toro, J. M. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition, 116*, 361-367.

Pye, C. (1986). Quiche Mayan speech to children. *Journal of Child Language 13(1)*, 85-100.

Rogoff, B., Paradise, R., Mejía Arauz, R., Correa-Chávez, M., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology*, *54*, 175–203.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25(4)*, 421-436.

Schieffelin, B. B., and Ochs, E. (1983). A cultural perspective on the transition from prelinguistic to linguistic communication, in R. M. Golinkoff (Ed), *The Transition from Prelinguistic to Linguistic Communication* (pp. 115–131). Hillsdale, NJ: Lawrence Erlbaum Associates.

Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistics and Language Compass, 3(1),* 236–264.

Soderstrom, M., & Morgan, J. L. (2007). Twenty-two-month-olds discriminate fluent from disfluent adult-directed speech. *Developmental Science*, *10*, 641-653.

Stadler, M. A. (1995). The role of attention in implicit learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *21*, 674–685.

Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica, 56*, 186–199.

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

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*, B25-B34.

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

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104(33)*, 13273-13278.

van de Weijer, J. (2001). Vowels in infant- and adult-directed speech. *Lund University Department of Linguistics Working Papers, 49*, 172-175.

van de Weijer, J. (2002). How much does an infant hear in a day? Paper presented at *the GALA 2001 Conference on Language Acquisition*, Lisboa.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1(2)*, 197-234.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, *43(2)*, 230–246.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L, & Amanos, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, *103*, 147-162.

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

Yoshida, K., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy, 15(4)*, 420-433.

APPENDIX

Japanese

1. *Kezza* wa kore da yo. (This is a *kezza*) / Kono *kezza* o yubisashi te mi te (Try to point at this *kezza*) / Omoshiroi katachi o shi te iru no wa *kezza* (The *kezza* has a funny shape)

2. *Gaby* wa sora ga daisuki da yo. (The *gaby* loves the sky) / Kono *gaby* o tsukame tai ne (We want to catch this *gaby*) / Ton de iru no wa *gaby* (The *gaby is flying*)

3. *Bayssa* wa okuchi o ake te iru yo (The *bayssa* is opening its mouth) / Kono *bayssa* o mi ta koto wa aru ka na (I wonder if we have seen this *bayssa*) / Koko ni iru no wa *bayss*. (Here is a *bayssa*)

4. *Payku* wa asonde tte itte iru yo. (The *payku* is saying “let’s play”) / Kono *payku* o nadenade shi te age you ka (Let’s pat this *payku*) / Poyon tte yure te iru no wa *payku* (This payku is swinging)

5. *Kayghee* wa banzai o shite iru mitai (The *kayghee* seems to be cheering) / Kono *kayghee* o mi te (Look at this *kayghee*) / Koko ni iru aoi no wa *kayghee* (This blue one is a *kayghee*)

6. *Beppy* wa ironna iro o shi te iru ne (The *beppy* is colorful) / Kono *beppy* o mama mo mi ta koto nai yo (Mom has not seen this *beppy* before) / Korokoro dekiru no wa *beppy* (The *beppy* can roll over)

7. *Gebby* wa midoriiro da ne (The *gebby* is green) / Kono *gebby* o otomodachi ni mo mise te age tai ne (We want to show this *gebby* to our friends) / Tanoshii na tte i tte iru no wa *gebby* (The *gebby* is saying “I’m happy”)

8. *Peckoo* wa soto ga daisuki da yo (The *peckoo* loves the outdoors) / Kono *peckoo* o ouchi ni tsure te kaeri tai ne (We want to bring this *peckoo* home) / Nikoniko shi te iru no wa *peckoo* (The *peckoo* is smiling.)

English

1. Look at the *kezza /* The *kezza* is smiling / Katie wants to give the *kezza* a hug.

2. Here is a *gaby /* The *gaby* is green / Mary has to give the *gaby* a bath.

3. Max has a *bayssa /* The *bayssa* is tall and bumpy / Max needs to paint the *bayssa* a new color.

4. Here is the *payku /* The *payku* is very heavy / Ed hopes he can give the *payku* a ride in his wheelbarrow.

5. Look at the *kayghee /* The *kayghee* is very flexible / Sue has to knit the *kayghee* a sweater.

6. Lets watch the *beppy /* The *beppy* is yellow / Isabelle would like to sing the *beppy* a song.

7. Let's watch the *gebby /* The *gebby* is jumping / Ryan can bake the *gebby* a cake.

8. Janet has made a *peckoo /* The *peckoo* is pie-shaped / Janet will make the *peckoo* a star.

Catalan

1. El nen està buscant el *bèpi* (the kid is looking for *bèpi) /* El *bèpi* és de color groc (The *bèpi* is yellow)/ Els nens portaran el *bèpi* a l'altra habitació (The kids will bring the *bèpi* to the other room)

2. Aquest és l'últim *kèdu* (This is the last *kèdu*)/ El *kèdu* sembla brillant (The *kèdu* seems shiny)/ Els científics observen el *kèdu* amb el microscopi (The scientists observe the *kèdu* with the microscope)

3. Sempre m'ha agradat el *dèbi* (I have always like the *dèbi*)/ El *dèbi* ha pujat de preu (The *dèbi* has gotten more expensive)/ Aquell noi ven el *dèbi* a un preu raonable (This boy sells the *dèbi* for a reasonable price)

4. La noia està vigilant el *pèku* (The girl is watching the *pèku*)/ El *pèku* és molt delicat (The *pèku* is very delicate)/ La noia ha deixat el *pèku* a l'escola (The girl has left the *pèku* at school)

5. L'electricista està arreglant el *béga* (The electrician is fixing the béga)/ El *béga* està totalment trencat (The béga is completely broken)/ El noi està observant el *béga* desde les vuit del matí (The guy is watching the *béga* since eight o'clock)

6. El venedor s'ha quedat sense *téko* (The seller has no more *téko*)/ El *téko* és molt dolç (The *téko* is very sweet)/ El venedor ha trobat una bossa de *téko* al magatzem (The seller has found a bag of *téko* at the warehouse)

7. El gos té un *kéba* (The dog has a *kéba*)/ El *kéba* és una joguina increible (The *kéba* is an incredible toy)/ El gos ha estat jugant amb el *kéba* durant més de dues hores (The dog has been playinh with the *kéba* for more than two hours)
8. El noi s'ha comprat un *bédo* (The boy has bought a *bédo*)/ El *bédo* funciona perfectament (The *bédo* works perfectly/ El noi provarà el *bédo* abans de decidir-se (The boy will try the *bédo* before making any decision).

(Manuscript received: 15 December 2010; accepted: 26 June 2011)