Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Cognition 103 (2007) 147–162
COGNITION
www.elsevier.com/locate/COGNIT

Brief article

Infant-directed speech supports phonetic category learning in English and Japanese ☆

Janet F. Werker [a,*], Ferran Pons [a], Christiane Dietrich [a], Sachiyo Kajikawa [b], Laurel Fais [a], Shigeaki Amano [b]

[a] *Department of Psychology, The University of British Columbia, 2136 West Mall, Vancouver, BC, Canada V6T 1Z4*
[b] *NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikari-dai, Seika-cho, Souraku-gun, Kyoto 619-0237, Japan*

Received 1 March 2006; accepted 30 March 2006

Abstract

Across the first year of life, infants show decreased sensitivity to phonetic differences not used in the native language [Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development, 7*, 49–63]. In an artificial language learning manipulation, Maye, Werker, and Gerken [Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82(3)*, B101–B111] found that infants change their speech sound categories as a function of the distributional properties of the input. For such a distributional learning mechanism to be functional, however, it is essential that the input speech contain distributional cues to support such perceptual learning. To test this, we recorded Japanese and English mothers teaching words to their infants. Acoustic analyses revealed language-specific differences in the distributions of the cues used by mothers (or cues present in the input) to distinguish the vowels. The robust availability of these cues in maternal speech adds support to the hypothesis that distributional learning is an important mechanism whereby infants establish native language phonetic categories.

☆ This manuscript was accepted under the editorship of Jacques Mehler.
* Corresponding author. Fax: +1 604 822 6923.
E-mail address: jwerker@psych.ubc.ca (J.F. Werker).

0010-0277/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2006.03.006

Keywords: Infant-directed speech; Distributional learning; English; Japanese

1. Introduction

One of the defining characteristics of language is its productivity. Units can be combined and recombined to allow for the creation of new words, phrases, and sentences. Languages differ in their phoneme repertoires, the sets of consonant and vowel sounds that distinguish meaning, and in the rules for how phonemes can be combined to create words and morphemes. It has been known for nearly 40 years that speakers of different languages represent and discriminate best those phonetic differences that are used phonemically (to contrast meaning) in their native language (Abramson & Lisker, 1970), but the means by which native phonetic categories are established have still not been fully explained.

Important advances were made in understanding this problem over 30 years ago when it was demonstrated that very young infants discriminate not only native, but also non-native phonetic differences (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Streeter, 1976), suggesting that sensitivity to the phonetic detail used to distinguish adult phonemic categories may be part of the initial perceptual apparatus. And, as first shown over 20 years ago (Werker & Tees, 1984), initial perceptual sensitivities change across the first year of life, resulting in diminished sensitivity to phonetic differences that are not used phonemically in the native language (see Best & McRoberts, 2003; Saffran, Werker, & Werner, 2006), and enhanced sensitivity to native distinctions (see Kuhl et al., 2006; Polka, Colontonio, & Sundara, 2001).

Several models were proposed to explain the underlying mechanisms that allow infants to tune speech sound categories so rapidly. Early models, following the structural–functional linguistics tradition (e.g., Jakobson, 1949; Trubetskoy, 1969) assumed that the tuning of native categories emerges only after the establishment of contrastive words in the lexicon (e.g., Werker & Pegg, 1992). More recent perceptual learning models posited various similarity metrics to account for the change from broad-based to language-specific phonetic perception [e.g., the ''Perceptual Assimilation Model'' (PAM) Best and McRoberts, 2003; the ''Native Language Magnet Model'' (NLM) Kuhl, 1993]. The missing piece in supporting a perceptual learning model was an explication and demonstration of an actual learning mechanism that could account for these changes.

In 2002, Maye, Werker, and Gerken provided evidence that distributional learning might underlie the rapid tuning to the categories of the native language. Using an artificial language learning manipulation, two groups of infants aged 6–8 months were exposed to all steps of an 8-step continuum of [da] to [ta][1] speech syllables.

[1] Importantly, the ''ta'' is not like the standard English /ta/ (phonetically transcribed as [t^ha]). The ''ta'' used here lacks the aspiration of an English initial position /ta/ and is more like the English ''ta'' following an ''s'' as in the word ''stop''. Although the [ta] and [da] are possible for adults and infants to distinguish, this difference is not as readily discriminated, without training, as a standard English [da]–[t^ha] difference (Pegg & Werker, 1997).

One group heard more instances of the two center points in the continuum, steps 4 and 5, corresponding to a unimodal frequency distribution (as might be experienced in a language without the [da]/[ta] distinction). The other group of infants heard more instances of steps 2 and 7, corresponding to a bimodal frequency distribution (as might be experienced by infants being raised in a language with this distinction). Both groups heard equal numbers of the remaining stimuli. Following 2.3 min of familiarization, infants in the bimodal but *not* the unimodal group showed evidence of discriminating steps 1 from 8. Maye and Weiss (2003) replicated the distributional learning finding with two new sets of speech contrasts, and more recently, Yoshida, Pons, and Werker (2006) have replicated it with a non-native distinction. Together, these results indicate that distributional learning could be a mechanism that allows for speech sound category restructuring in the first year of life, prior to the establishment of a lexicon.

While laboratory-based artificial language learning studies constitute proof in principle that a particular learning mechanism is available, infants will not be able to apply this learning mechanism unless the speech they hear comprises such defined distributional regularities. Thus, an essential step is to examine the characteristics of input speech that infants hear, to see if the frequency distribution of the relevant acoustic/phonetic cues required for this learning mechanism does indeed exist in the input.

In this study we analyze durational and spectral cues of vowels from Japanese and English maternal speech. In Canadian English, the vowels differ primarily in vowel color. The acoustic correlates of vowel color are seen in the frequency of the formants; i.e. spectral differences. In Japanese there are only five vowels that differ in color, but every vowel has two forms, a long and a short form. Although likely once a dominant cue in English as well (Lehiste, 1970; Port, 1981), the historic length difference that still exists in some tense/lax vowel pairs has been superseded by a color difference in the English vowel space. A comparison of input speech for vowel pairs that differ in length in Japanese, and primarily in vowel color in English would thus allow a test of the hypothesis that there are distributional characteristics in the input that support native category learning. Because both English infants (Cooper & Aslin, 1990; Fernald, 1985) and Japanese infants (Hayashi, Tamekawa, & Kiritani, 2001) show a preference for listening to infant-directed over adult-directed speech, the strongest evidence would be provided by a study of infant-directed speech.

Vowels are more variable than consonants. Many factors influence vowel duration, including the voicing of the surrounding consonants, emphatic stress, focus, position in an utterance, and affect (for an overview see Erickson, 2000). The spectral differences that cue vowel color distinctions are also influenced by many factors, including pitch height and degree of pitch change (see Lieberman & Blumstein, 1988; Trainor & Desjardins, 2002 for a discussion) and speaking rate (Lindblom, 1963). In infant-directed speech, both vowel duration and spectral properties are affected by high pitch and highly affective modulation in English (e.g., Fernald, 1985) and Japanese (Hayashi et al., 2001). Vowel duration is much longer in infant-directed than in adult-directed speech (e.g., Andruski & Kuhl, 1996; Fernald & Simon, 1984; Kuhl et al., 1997), raising the very real possibility that the critical

distributional cues to support distinctive categories in the domain of duration may be quite different in the input. Although it has been shown that the articulatory configurations used to distinguish vowel color differences in English are exaggerated in infant-directed over adult-directed speech (Andruski & Kuhl, 1996), the higher overall pitch could nonetheless lead to varying availability of the spectral cues distinguishing vowels.

Here we ask if the crucial distributional information is present in infant-directed speech to allow infants to modify initial sensitivities and establish native language vowel categories. We compare two languages, Japanese and English, on two very similar vowel pairs. In adult speech both vowel pairs are cued only by duration in Japanese, whereas in English both are cued spectrally, with duration as a less predictive, but most likely still available, secondary cue. If there are sufficient distributional cues in input speech to allow infants to tune their perceptual systems to the phonetic categories of the native language using distributional learning, then the following predictions must be upheld: (1) there should be two significantly distinct distributions of vowel length but not vowel color in the two members of each vowel pair as produced by the Japanese mothers, and (2) there should be two distinct distributions of vowel color in the two members of each vowel pair produced by the English mothers, and the distribution of vowel length should not be as distinct as it is for Japanese. Moreover, these differences should be apparent not only when the categories are already given, but the characteristics of maternal input – on their own – should yield language-specific categories. Specifically, (3) the input speech of Japanese mothers should better predict two categories for each vowel pair on the basis of vowel length than will the input speech of English mothers and (4) the input speech of English mothers should better predict two categories for each vowel pair on the basis of vowel color than will the input speech of Japanese mothers.

2. Method

2.1. Participants

The study was conducted at the Infant Studies Centre at the University of British Columbia, Vancouver, Canada and in the NTT Communication Science Laboratories, Keihanna, Japan. A total of 30 mothers (20 Canadian-English and 10 Japanese) and their 12-month-old infants participated in the study.

Japanese infants were able to sit through the full version of the study. The Canadian-English infants were much less compliant, and were only able to complete half of the study; thus, twice as many Canadian mothers were needed to complete the sample.

2.2. Recording apparatus

The recordings were made in a quiet and comfortable room. During the recording the mother and the baby were left alone to keep additional noise to a minimum.

In Canada, maternal speech was recorded directly onto a Macintosh G3 computer using Sound Edit 16 (Version 2-99) software. In Japan, a DAT recorder was used. The sampling rate during the recording was 44 kHz in Canada and 48 kHz in Japan. The speech data were later re-sampled at a rate of 16 kHz for consistency across labs.

2.3. Stimuli

Two vowel pairs were used in the study, /I–ii/ and /E–ee/. In adult Japanese speech, these vowels differ only in length (phonetically transcribed as /i/-/i:/ and /ɛ/-/ɛ:/). In Western Canadian English, these vowel pairs differ primarily in vowel color (phonetically transcribed as /ɪ/-/i/ and /ɛ/-/e/). These distinctions for English are exemplified in the words "bit"–"beat" and "bet"–"bait". In most dialects of English the vowel in "bait" is a diphthong, meaning that it glides from one vowel /e/, to another, /i/, transcribed as /ei/. In Western Canadian English, however (K. Russell, personal communication, December 10, 2002), the vowel in "bait" is more often a monophthong; hence it is phonetically transcribed as /e/. Despite the acoustic differences between the English and Japanese pairs, we label the pairs /I–ii/ and /E–ee/ for both languages for ease of reference.

A set of 16 nonsense CVCV words was created. All words were selected to be phonotactically possible word-forms in both Japanese and Canadian English. To control for the known effect of voicing on vowel duration in English, the nonsense words were created using all four possible combinations of voiced and voiceless consonants before and after the target vowel (see Table 1). Each Japanese mother was recorded with all possible combinations. In reminder, it was necessary to record 20 English-speaking mothers to obtain a full complement of items. Each vowel type was

Table 1
List of nonsense words in English orthography and Japanese katakana, with their phonetic realizations

/E/	/ee/	/I/	/ii/
Peckoo (/pɛku/) ペク (/pɛkɯ/)	Payku (/peku/) ペーク (/pɛ:kɯ/)	Pippa (/pIpə/) ピパ (/pipɑ/)	Peepo (/pipo/) ピーポ (/pi:po/)
Kezza (/kɛzə/) ケザ (/kɛzɑ/)	Kayghee (/kegi/) ケーギ (/kɛ:gi/)	Kibboo (/kIbu/) キブ (/kibɯ/)	Keedo (/kido/) キード (/ki:do/)
Beppy (/bɛpi/) ベピ (/bɛpi/)	Bayssa (/besə/) ベーサ (/bɛ:sɑ/)	Bicko (/bIko/) ビコ (/biko/)	Beepa (/bipə/) ビーパ (/bi:pɑ/)
Gebby (/gɛbi/) ゲビ (/gɛbi/)	Gaby (/gebi/) ゲービ (/gɛ:bi/)	Gidda (/gIdə/) ギダ (/gidɑ/)	Geeda (/gidə/) ギーダ (/gi:dɑ/)

produced by each English-speaking mother, but each voicing context was not recorded by each English mother. To counterbalance effects of sentence position, the nonsense words occurred at the beginning, middle, and end of the sentence in the reading task.

2.4. Materials and recording procedure

Mothers were asked to "teach" the 16 nonsense words to their infants while they looked at a picture-book together. This included both a picture-book reading task and a spontaneous speech task. The picture-book devoted two pages to each nonsense word. On the first page, the nonsense word was written in the three above sentence types under a colourful picture of a novel object and the mother was asked to read the sentences. On the second page, the novel object was depicted in a specific context and the mother was asked to describe the scene, using the name of the object as much as possible. This resulted in several repetitions of each word from each mother, balanced across the variety of contexts that are known to affect vowel characteristics.

2.5. Acoustic analyses

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

3. Results

The sentences were classified as "Reading" (read sentences) or "Spontaneous" speech (description of the visual scene). From the Japanese recordings, a range of 52–64 tokens from each mother from the Reading, and a range of 30–65 tokens from the Spontaneous speech were analyzed. For the English recordings, a range of 25–36 tokens from each mother were analyzed from the Reading. The Spontaneous speech was analyzed from 19 mothers (one mother did not produce any nonsense words spontaneously), with a range of 10–28 tokens each.

To test Predictions 1 and 2, the data were analyzed using ANOVA to compare the mean values for the input from each group of mothers for each acoustic dimension. To test Predictions 3 and 4, the data were then analyzed using hierarchical logistic regression to model the odds likelihood that the acoustic correlates of maternal speech would better yield two vowel categories on the basis of duration in Japanese, and on the basis of formant values in English.

Table 2
Means for duration and color for all vowels in English and Japanese Read and Spontaneous speech

		Vowel length (ms)			
		E	ee	I	ii
Read Speech	English	119	180	99	130
	Japanese	85	205	74	175
Spontaneous	English	110	157	91	114
	Japanese	86	169	72	149
		Vowel color (Hz: F2 − F1)			
		E	ee	I	ii
Read Speech	English	1341	1869	1646	2095
	Japanese	1575	1575	1839	1920
Spontaneous	English	1353	1878	1667	2039
	Japanese	1672	1565	1847	1848

3.1. Vowel length: ANOVAs

The mean durations for each vowel as pronounced by Japanese and English mothers were analyzed separately for Read and Spontaneous speech. A 2 (vowel: /I/ vs. /ii/ or /E/ vs. /ee/) × 2 (language: English vs. Japanese) mixed ANOVA was used. The results were identical for both vowels, and for both Read and Spontaneous speech. There was a significant effect of vowel for both /E–ee/, $F(1,28) = 24.78$, $p < .001$ (Read speech), $F(1,26) = 79.41$ $p < .001$ (Spontaneous speech), and /I–ii/ $F(1,28) = 93.37$, $p < .001$ (Read speech), $F(1,27) = 81.26$, $p < .001$ (Spontaneous speech), and a significant interaction between language and vowel for each vowel pair, /E–ee/, $F(1,28) = 28.63$, $p < .001$ (Read speech); $F(1,26) = 5.95$, $p = .022$ (Spontaneous speech), and /I–ii/, $F(1,28) = 27.06$, $p < .001$ (Read speech); $F(1,27) = 23.66$, $p < .001$ (Spontaneous speech). As shown in Table 2, the interaction is accounted for by a greater difference in means in Japanese than in English.

3.2. Vowel color: ANOVAs

The measure used for the analyses of spectral characteristics was the standard measure of the difference between F2 and F1[2] (the same results were also found analysing F1 and F2 separately). Using a 2 (vowel: /I/ vs. /ii/ or /E/ vs. /ee/) × 2 (language: English vs. Japanese) mixed ANOVA, the results were again identical for both vowels, and for both read and spontaneous speech. There was a significant effect of vowel for both /E–ee/, $F(1,28) = 28.41$, $p < .001$ (Read speech), $F(1,26) = 22.82$, $p < .001$ (Spontaneous speech), and $F(1,28) = 36.24$, $p < .001$ (Read speech), $F(1,27) = 14.81$, $p < .001$ (Spontaneous speech), and a significant interaction between language and vowel for each vowel pair, /E–ee/,

[2] The F2 − F1 value gives the backness of the vowel (Ladefoged, 1993).

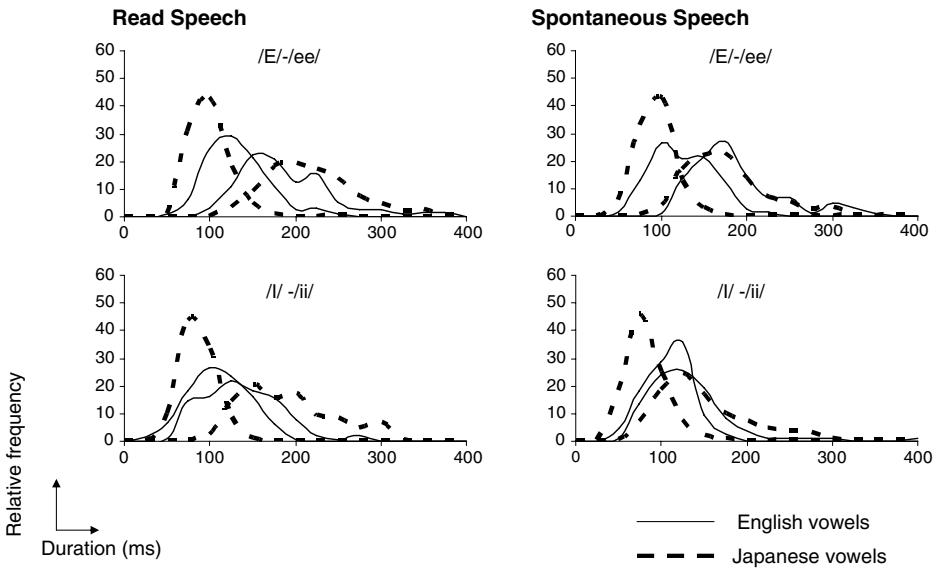

Fig. 1. Relative frequency of the vowel duration in /E–ee/ and /I–ii/ pairs in Japanese and in English, illustrated in 50 ms bins. The figure shows the /E–ee/ pair (upper left), the /I–ii/ pair (lower left) in the Reading task, and the /E–ee/ pair (upper right), the /I–ii/ pair (lower right) in the Spontaneous speech.

$F(1,28) = 28.29$, $p < .001$ (Read speech); $F(1,26) = 52.40$, $p < .001$ (Spontaneous speech), and $F(1,28) = 17.39$, $p < .001$ (Read speech); $F(1,27) = 14.65$, $p < .001$ (Spontaneous speech). As shown in Table 2, for vowel color, the interaction is accounted for by a greater difference in means in English than in Japanese. These findings are graphically illustrated in Fig. 1 (vowel length) and Fig. 2 (vowel color).

3.3. Hierarchical multi-level logistic regression

Predictions 3 and 4 required a modeling strategy in which vowel category becomes the dependent, rather than the independent, variable. Specifically, we used a multi-level hierarchical logistic regression to ask whether maternal input better predicted two vowel categories in one language vs. the other along the acoustic dimension of interest. Logistic regression, which is more suitable than linear regression for predicting nominal outcomes, tests the natural log of the odds (the "logit") that data can be classified into a predetermined number of categories (Raudenbush & Bryk, 2002). Because we began with the assumption that there are up to two categories for each vowel dimension[3], the logistic regression was set up with two nominal out-

[3] Languages with three-duration distinctions have been reported; however, they are extremely rare and the claims are not uncontroversial. It has been claimed that three-duration distinctions are unstable and tend to re-organize as two-duration distinctions. In some cases, another acoustic feature (F0, for example) marks a third distinction (Lehiste, 1997; McRobbie-Utasi, 1999).

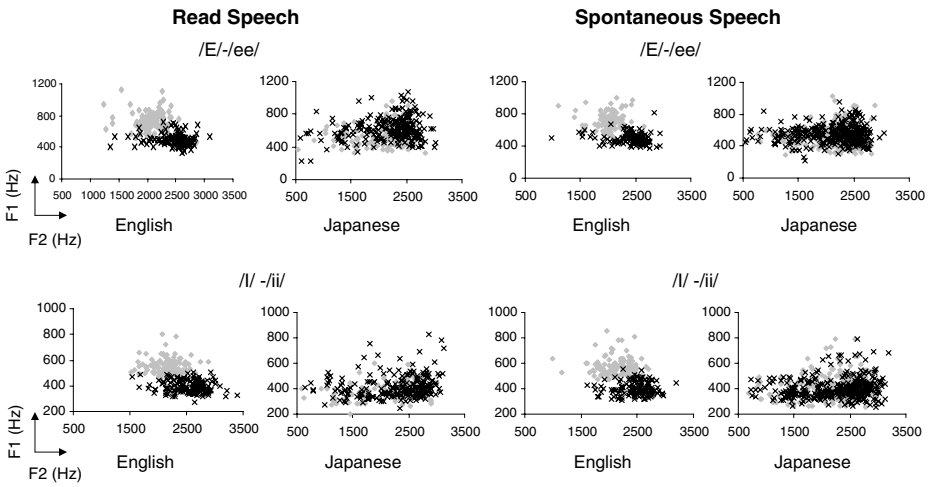

Fig. 2. F1 – F2 Scatter plots for all items for both vowel pairs in English and Japanese, in both Read and Spontaneous speech conditions.

comes: the long vowel (e.g., /ee/) was labelled 0 whereas /E/ was labelled 1. Since all tokens provided by each mother were included, it was essential to consider the heterogeneity that exists between mothers as well as the clustering and potential statistical dependencies that exist within mothers in speaking style (rate, variability, pitch, etc.). Hierarchical logistic regression controls for both by modeling the variability within mothers at the first level, Level 1, and modeling the variability between mothers at Level 2. Tokens within mothers served as Level 1 units and mothers served as Level 2 units.

To investigate whether the log-odds that acoustic characteristics of maternal speech are related to vowel category as a function of language (which was also coded as two categories; 0 for English and 1 for Japanese), we ran the following multilevel analyses for Acoustic Measure (unstandardized Vowel Length or Color). The following demonstrates the typical analysis in the present study:

$$\eta_{ij}(\mathrm{VCat}) = \ln[(\phi_{ij})/(1-\phi_{ij})] = \beta_{0j} + \beta_{1j}(\mathrm{Acoustic\ Measure_{ij}}) + e_{ij} \quad [\mathrm{Level\ 1}]$$

$$\beta_{0j} = \gamma_{00} + \gamma_{0j}(\mathrm{Language}_j) + u_{0j} \quad [\mathrm{Level\ 2}]$$

$$\beta_{1j} = \gamma_{10} + \gamma_{1j}(\mathrm{Language}_j) + u_{1j}$$

Level 1 models the log-odds (η_{ij}) of categorizing a vowel, for example as /E/ versus /ee/. While the probability of vowel category (VCat), ϕ_{ij}, is constrained to be either 0 or 1, η_{ij} can take on any real value. Acoustic Measure$_{ij}$ is the vowel duration or color for token i and participant j. The coefficients β_{0j} and β_{1j} are the intercept and slope, respectively, for participant j. That is, β_{0j} is participant j's predicted mean log-odds for VCat across the tokens when Acoustic Measure$_{ij}$ is 0, and β_{1j} is the predicted change in the log-odds as a function of Acoustic Measure of token i for participant

j. β_{0j} is a function of the mean intercepts of the VCat odds across all mothers, γ_{00}, and across language, γ_{0j}, whereas β_{1j} is a function of the mean estimated slopes for duration, γ_{10}, and language, γ_{ij}, across all participants. As such, the tests of γ_{10} and γ_{1j} represent whether, on average across mothers, log-odds of categorizing vowel type increase with Acoustic Measure and whether language category alters the relationship between vowel Acoustic Measure and VCat log-odds, respectively. Both β_{0j} and β_{1j} can vary randomly across participants as is illustrated at Level 2 of the analyses (u_{0j} and u_{1j} represent systematic unanalyzed variation across mothers).

3.4. Does vowel length better predict two categories for each vowel pair for the input speech of Japanese mothers than for English mothers?

The unstandardized duration for each vowel (i.e. vowel length) as pronounced by Japanese and English mothers was analyzed separately for both Read and Spontaneous speech and for each vowel pair, /E–ee/ or /I–ii/. For example, the model for /E–ee/, Read speech is as follows:

$$\eta_{ij}(\text{E–ee}) = \ln[(\phi_{ij})/(1-\phi_{ij})] = \beta_{0j} + \beta_{1j}(\text{Duration}_{\text{ij}}) + e_{ij} \quad \text{[Level 1]}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{0j}(\text{Language}_j) + u_{0\text{j}} \quad \text{[Level 2]}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{1j}(\text{Language}_j) + u_{1j}$$

Unit specific models with robust standard errors were specified for each analysis. Table 3 presents the estimated model results for each vowel pair for each type of input speech for all the analyses considering vowel length. The interaction between Duration at Level 1 and Language at Level 2 is the critical component of the analyses of interest here. The results for each vowel pair and for both Read and Spontaneous speech indicate that the interaction between vowel length and language of the mother was significant. This means that for /E–ee/ and /I–ii/ in both Read and Spontaneous speech, maternal Japanese input had a greater log-odds of predicting two categories than did maternal English speech.

Table 3
HGLM coefficients and standard errors for analyses of vowel length and language predicting log-odds of vowel category (*$p < .05$; **$p < .01$)

Fixed effects	/E–ee/				/I–ii/			
	Read		Spontaneous		Read		Spontaneous	
	β	*SE*	β	*SE*	β	*SE*	β	*SE*
Intercept, β_0								
Intercept, γ_{00}	3.58*	1.44	3.49	2.14	−7.90**	1.45	−0.69	1.30
Duration, γ_{01}	−17.95	11.54	−11.86	18.01	71.73**	9.87	7.25	10.22
Duration, slope, β_1								
Language, γ_{10}	4.39**	0.75	2.36	1.52	10.41**	0.94	3.35**	0.77
Lang × Dur, γ_{11}	−37.98**	6.52	−30.92*	13.69	−95.68**	5.72	−34.53**	5.87

To graphically illustrate these findings, we transformed the log-odds into probability scores. Fig. 3 illustrates the relationship among /I–ii/ vowel category probability and vowel length for English- and Japanese-speaking mothers for Read and Spontaneous speech. As can be seen in Fig. 3, the probability of categorizing a vowel as either /I/ or /ii/ based on length of vowel is stronger for Japanese-speaking mothers than for English-speaking mothers for both Read and Spontaneous speech. The curve shown for Japanese is, in both cases, close to an idealized estimated two-category function, with the majority of the predicted cases being close to 0 or 1 probability, and a steep transition, whereas the curve for English reveals a more continuous function.

3.4.1. Does vowel color better predict two categories for each vowel pair for the input speech of English mothers than for Japanese mothers?

We conducted similar analyses as above, but replaced Vowel Length with Vowel Color. We utilized unstandardized F2 − F1 values in the analyses to predict the log-odds of predicting two categories for each vowel pair, /E–ee/ or /I–ii/, separately for both Read and Spontaneous speech. Again, unit specific models with robust standard errors were specified for each analysis. Table 4 presents the estimated model

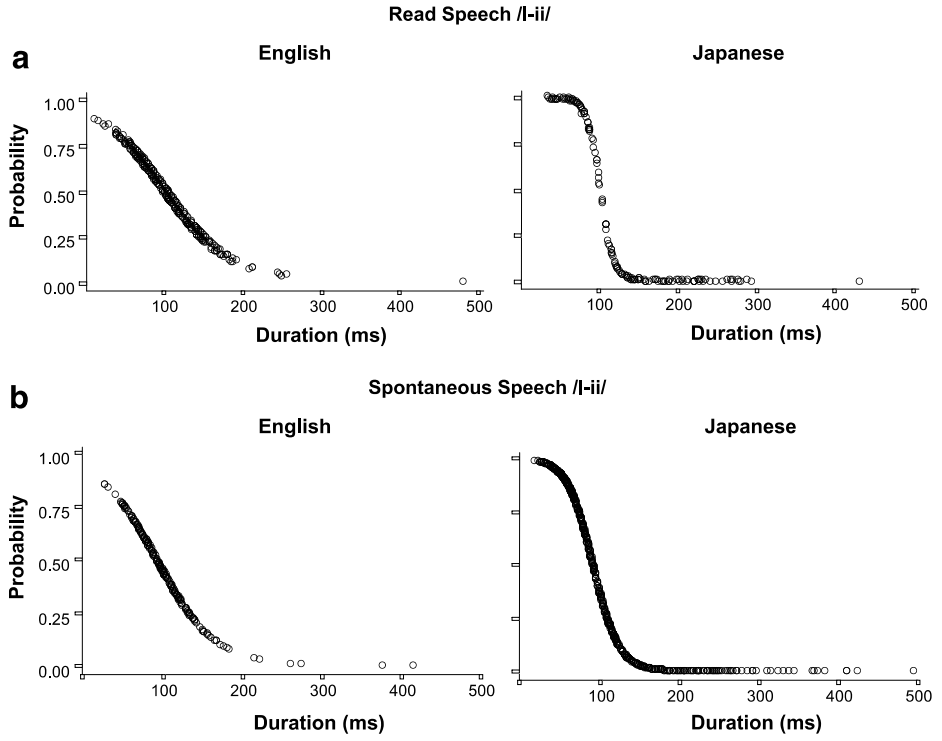

Fig. 3. Probability functions (derived from log-odds) of vowel duration differences predicting two categories for the vowel pair /I–ii/ in Read (a) and Spontaneous (b) speech in English and Japanese.

Table 4
HGLM coefficients and standard errors for analyses of vowel color and language predicting log-odds of vowel category (*$p < .05$; **$p < .01$)

Fixed effects	/E–ee/				/I–ii/			
	Read		Spontaneous		Read		Spontaneous	
	β	*SE*	β	*SE*	β	*SE*	β	*SE*
Unit specific model with Robust Standard Errors								
Intercept, β_0								
Intercept, γ_{00}	24.29**	5.16	27.05**	4.41	24.43**	4.04	22.60**	4.08
F2 − F1, γ^{01}	−0.01**	2.59	−0.02**	0.00	−0.01**	0.00	−0.01**	0.00
Duration, slope, β_1								
Language, γ_{10}	−12.19**	2.60	−13.98**	2.24	−11.94**	2.11	−11.42**	2.06
Lang × F2 − F1, γ_{11}	0.01**	0.00	0.01**	0.00	0.01**	0.00	0.01**	0.00

results for Vowel Color for each vowel pair for both Read and Spontaneous speech. The results again reveal that the interaction between vowel color and language of the mother was significant. Thus, for both Read and Spontaneous speech across both /E–ee/ and /I–ii/ vowel pairs, maternal English input had a greater log-odds of predicting two categories on the basis of vowel color than did maternal Japanese speech.

A graphic illustration of the results for /E–ee/ for both Read and Spontaneous speech, converting the log likelihood of two categories back to probabilities, is shown in Fig. 4. This figure reveals that virtually all of the estimated values for Japanese fall at or near the .50 probability line while for English input the probability of categorizing the vowel as /E/ or /ee/ falls closer to the idealized estimated two category function of 0 and 1.

4. Discussion

The goal of this study was to determine if, in the face of all the variation present in infant-directed speech, there are sufficient cues in the input to support distributional learning of native language phonetic categories. In our study of Japanese and English mothers teaching new words to their infants, we found clear and consistent language-specific cues. The vowel pairs /E–ee/ and /I–ii/ each differed more in length in the speech of Japanese mothers, whereas each differed more in color (as indicated by spectral differences) in the speech of English mothers. These results provide *prima facie* evidence that distributional cues do exist in the input, even in infant-directed speech in which both vowel duration and pitch are highly variable (Fernald et al., 1989), potentially affecting the distinctiveness of these cues for phonetic category learning. Perhaps even more convincing, when the input characteristics are modelled using hierarchical logistic regression, there is a significantly greater likelihood that duration will yield two categories in the speech of Japanese mothers, and that vowel spectral differences will yield two categories in the speech of English mothers. Hence, even when the category is the outcome rather than the predictor variable, input

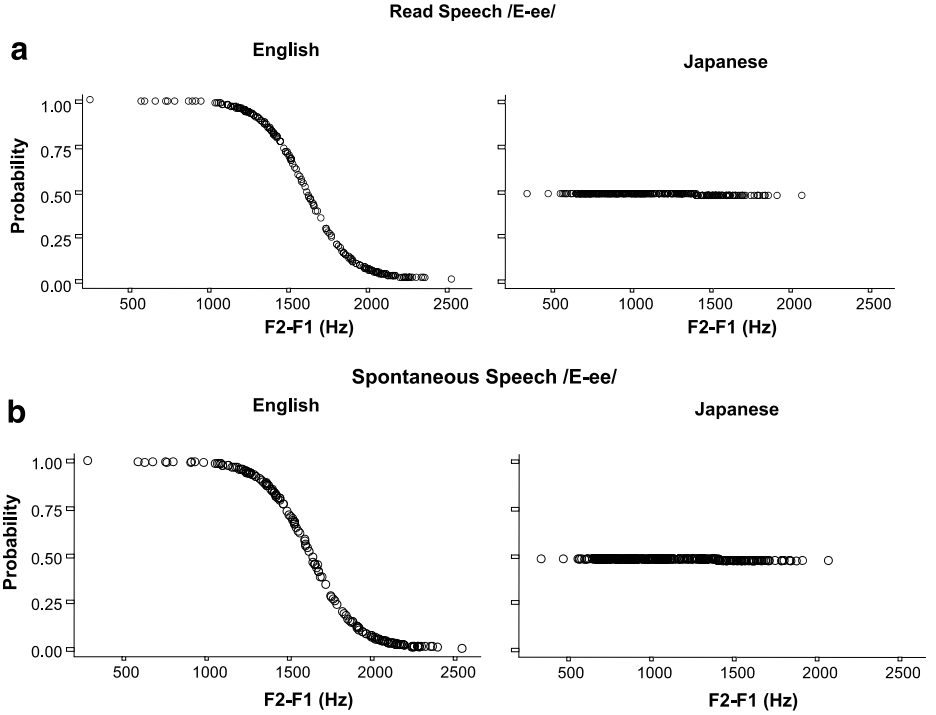

Fig. 4. Probability functions (derived from log-odds) of spectral (F2 minus F1) differences predicting two categories for the vowel pair /E–ee/ in Read (a) and Spontaneous (b) speech in English and Japanese.

speech can be seen to have the properties to lead to the weighting of the critical language-specific acoustic variables, and to the establishment of categories along those dimensions. That this is seen even when the variability within and between mothers within a language group is controlled, reveals just how predictive the statistics of the input are. The finding that the relevant cues are robustly available even in infant-directed speech adds to the likelihood that, just as they do in artificial language learning studies (Maye, Werker, & Gerken, 2002; Maye & Weiss, 2003), infants in naturalistic language-learning environments use distributional learning to establish native language phonetic categories.

It is interesting to speculate what the presence of these robust cues to phonetic categories might indicate about the function and evolutionary significance of infant-directed speech. It has been suggested that infant-directed speech serves first to attract the infant's attention to the mother (Werker & McLeod, 1989) and to communicate affect between mother and child (Trainor, Austin, & Desjardins, 2000), and only later to play a role in language acquisition (Fernald, 1992). In support, it has been shown that in American Sign Language (ASL), mothers sacrifice the grammatically necessary brow furrowing gesture to preserve positive affect in the face in their interactions with infants in the first 18 months of life (Reilly & Bellugi, 1996). The

results shown herein together with earlier research (Kuhl et al., 1997; Liu, Kuhl, & Tsao, 2003; Ratner & Luberoff, 1984), show that in addition to its well-documented role in modulating affect, infant-directed speech may also enhance the cues distinguishing the linguistic features of the basic combinatorial units of the native language from very early in infancy.

Infants prefer to listen to infant-directed speech (Cooper & Aslin, 1990; Fernald, 1985), but there is ample evidence that they listen to overheard adult speech as well (Oshima-Takane, 1988). It is thus important in future work to determine if distributional cues are equally available in adult-directed speech. Moreover, it would be informative to determine if distributional cues are more exaggerated in some pragmatic tasks, e.g., word teaching, than they are in others, even in infant-directed speech. Answers to these questions will help delimit how broadly available a distributional learning mechanism might be, and thus how well it can account for the perceptual learning of language-specific phonetic categories.

In summary, in this study we have shown that in the highly modulated style that mothers use when speaking to their infants, there are statistically regular differences in the phonetic cues necessary to support distribution-based perceptual learning. This research helps answer the 40-year-old question of just how and when it is that speakers of different language groups acquire their native language perceptual categories. The distributional information necessary to enhance some category distinctions and collapse others is evident in the input. This work, together with the previous artificial language work on distributional learning (Maye et al., 2002; Maye & Weiss, 2003), thus provides evidence of a plausible learning mechanism that could help reshape phonetic categories to conform to the phoneme repertoire of the native language. As such, the work adds to the growing evidence that perceptual learning can indeed guide, rather than necessarily trail, the establishment of a lexicon.

Acknowledgements

We thank Noshin Lalji-Samji for designing the picture books, Jeremy Biesanz for statistical advice, Eli Puterman for help performing the analyses, and Jay McClelland for feedback and comments. We are grateful to all the mothers and infants who participated in this research.

References

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum; Cross-language tests. In *Proceedings of the Sixth International Conference of Phonetic Sciences* (pp. 569–573). Prague.

Andruski, J., & Kuhl, P. K. (1996). The acoustic structure of vowels in mothers' speech to infants and children. In *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1545–1548). Philadelphia, PA.

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

Boersma, P., & Weenink, D. (2004). Praat: doing phonetics by computer (Version 4.3.02) [Computer program]. Retrieved from http://www.praat.org/.

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

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.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkwo, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 391–428). Oxford: Oxford University Press.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104–113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477–501.

Hayashi, A., Tamekawa, Y., & Kiritani, S. (2001). Developmental change in auditory preferences for speech stimuli in Japanese infants. *Journal of Speech, Language and Hearing Research, 44*, 1189–1200.

Jakobson, R. (1949). On the identification of phonemic entities. *Travaux du Cercle Linguistique de Copenhague*, *5*, 205–213. (Reprinted in R. Jakobson *Selected Writings I: Phonological Studies*. The Hague: Mouton, pp. 418–425 (1962)).

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. deBoysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Dordrecht, Netherlands: Kluwer Academic Publishers.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (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 a facilitation effect for native language perception between 6 and 12 months. *Developmental Science, 9*(2), F1–F9.

Ladefoged, P. (1993). *A course in phonetics* (3rd ed.). Harcourt Brace Jovanovich College Publishers.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35*, 1773–1781.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge: MIT Press.

Lehiste, I. (1997). Search for phonetic correlates in Estonian phonology. In I. Lehiste & J. Ross (Eds.), *Estonian prosody: papers from a symposium* (pp. 11–35). Tallinn: Institute of Estonian Language.

Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. New York, NY: Cambridge University Press.

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.

Maye, J., & Weiss, D. J. (2003). Statistical cues facilitate infants' discrimination of difficult phonetic contrasts. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 508–518). Somerville, MA: Cascadilla Press.

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

McRobbie-Utasi, Z. (1999). *Quantity in the Skolt (Lappish) Saami language: an acoustic analysis. Indiana University Uralic and Altaic Series 165*. Bloomington: Indiana University Research Institute for Inner Asian Studies.

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

Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America, 102*(6), 3742–3753.

Polka, L., Colontonio, C., & Sundara, M. (2001). A cross-language comparison of /d/-/ð/ perception: evidence for a new developmental pattern. *Journal of Acoustical Society of America, 109*, 190–220.

Port, R. (1981). Linguistics timing factors in combination. *Journal of the Acoustical Society of America, 69*, 262–274.

Ratner, N. B., & Luberoff, A. (1984). Cues to post-vocalic voicing in mother-child speech. *Journal of Phonetics, 12*, 285–289.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Reilly, J. S., & Bellugi, U. (1996). Competition on the face: affect and language in ASL motherese. *Journal of Child Language, 23*(1), 219–239.

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In D. Kuhn & M. Siegler (Eds.), *The 6th edition of the handbook of child psychology* (pp. 58–108). New York: Wiley.

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature, 259*, 39–41.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed prosody a result of the vocal expression of emotion?. *Psychological Science 11*(3), 188–195.

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

Trubetskoy, N. S. (1969). *Principles of Phonology* (C. Baltaxe, Trans.). Berkeley: University of California Press. (Original work published 1939).

Yoshida, K. A., Pons, F., & Werker, J. F. (2006, June). *Does distributional learning affect perception after phonemes are established*? Poster presented at the International Conference on Infant Studies (ICIS), Kyoto, Japan.

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., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: models, research, and implications* (pp. 285–311). York Publishing Company.

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