THE ACOUSTIC STRUCTURE OF VOWELS IN MOTHERS' SPEECH TO INFANTS AND ADULTS

Jean E. Andruski
Patricia K. Kuhl

Dept. of Speech & Hearing Sciences, University of Washington, Box 357920, Seattle, WA 98195 - 7920

ABSTRACT

Research has shown that exposure to a specific language alters infants' perception of vowel sounds by the time they reach 6 months of age. This raises the important question of how infants develop language-specific patterns of vowel perception from the language-general pattern they appear to be born with. Language spoken to infants may exert an important influence in this regard. The present study compares the acoustic structure of vowels in the words "sheep" and "shoes" produced by 10 mothers in conversation with their infants, with their acoustic structure when produced by the same 10 women in conversation with an adult. Mothers were instructed to play with their infants using a toy sheep and pair of shoes. They were asked to use the same words in conversation with an adult. The infant-directed tokens exhibited higher F0, greater pitch excursions and longer duration than the adult-directed tokens, as is typical of motherese. Although F0 was significantly higher in infant-directed vowels, F1 remained at essentially the same frequency as in adult-directed /i/ and /u/. In contrast, F2 was significantly higher in infant-directed /i/ and significantly lower in infant-directed /u/ than in the adult-directed tokens. Thus, the infant-directed tokens reached more extreme acoustic targets. The formant structure also indicated an equivalent or slightly greater degree of coarticulation of infant-directed /u/ with the preceding fricative in "shoes." Overall, mothers in this study consistently hyperarticulated vowels in speech to infants. Hyperarticulation may contribute to the acquisition of native language vowel categories by increasing the degree of acoustic separation between vowel categories.

1. INTRODUCTION

A critical question for language development research is the extent to which language is innately specified as opposed to being learned as a function of input. The "innateness hypothesis" of language acquisition states that language input to human infants is too impoverished to support learning of grammatical principles, and that the essential properties of human language must therefore be innately specified.[1] Based on the complexity of the acoustic cues for phonemes, the innateness argument has been extended to the phonetic level of language.[2] In these formulations, language input plays a role, but only by maintaining certain pre-specified knowledge. Models that strongly emphasize the role of learning (e.g. Kuhl's Natural Language Magnet Theory, or NLM) hold that language input actually induces structure.[3]

Recent cross-language research on young infants has called the innateness hypothesis into question and refocused attention on learning models. Research has shown that exposure to a specific language alters infants' perception of speech sounds by 6 months of age.[4] The results suggest that language input itself provides sufficient information to instigate the learning of native-language sound structure.

During the first few months of life, language input to infants is largely from the primary caretaker, particularly the mother. Cross-language studies have shown that mothers' speech to infants undergoes predictable modifications in comparison with adult-directed speech.[5] Mothers tend to use higher vocal pitch, larger pitch excursions, longer durations and more repetition when speaking to infants. Furthermore, research shows that infants are biased to attend to speech with the pitch characteristics of "motherese."[6] Thus, motherese functions as an attention-getting device. Does it also contribute to the infant's acquisition of the sound categories of language?

The present study provides a preliminary exploration of this possibility. The acoustic structure of the vowels /i/ and /u/ in mothers' conversational speech with infants is examined and compared with the structure of the same vowels in their speech to another adult.

2. METHOD

2.1. Subjects

Twenty-five mother/infant pairs from the University of Washington subject pool were paid for their participation in this study. Mothers were native speakers of a general American English dialect, with no history of hearing or speech disorders. Mother's age was not a factor in subject selection. Infants were healthy, normally-developing babies between 2 and 4 months of age, born to monolingual families with no known history of hearing loss. Infants born 10 or more days prematurely were excluded from the study.

Data for 9 mother/infant pairs were discarded from the analyses because the baby cried, fussed, or slept throughout the recording session, because the mother had a strong regional accent or a cold, or because the infant was less than 2 months of age. Data for 6 further mother/infant pairs were discarded due to equipment failure or experimenter error which resulted in incomplete or otherwise unuseable recordings.

The final subject set consisted of 10 mother/infant pairs in which infant age ranged from 9.1 to 17.4 weeks, with a mean of 14.3 weeks.

2.2. Recording Procedure

Mothers were given a familiarization session in which the equipment and procedures for the study were explained. Mothers were then recorded in 2 conversations, the first adult-directed and the second infant-directed. In each conversation the mothers were asked to use the words "sheep" and "shoes" at least 3 times. The experimenter asked prompting questions and shared personal anecdotes in order to encourage natural conversational during the adult-directed session. For infant-directed conversations, mothers were given a toy sheep and pair of shoes, and were asked to play with their infants using the toys as if they were playing at home. Mothers were left alone with their infant for the infant-directed session. During both recording sessions mothers changed, fed and otherwise tended to their infants as necessary.

2.3. Apparatus

Recordings were made in a sound-treated room using an Electro-Voice 635-A omni-directional microphone suspended from the ceiling so that it hung 4 to 6 inches in front of the subject's face. The microphone signal was sent to a Tascam 122 MK II audio cassette tape recorder in an adjacent room.

2.4. Acoustic Analyses

All target word tokens except those which overlapped with conversation from the experimenter, infant vocalizations or toy noise, were sampled from tape to a Gateway 2000/486 33C computer. Tokens which had overdriven the tape recorder, or whose amplitude was too low for analysis were eliminated. A total of 243 tokens were retained for analysis (see Table 1).

Target Word	Addressee: Infant	Adult	Number of Tokens
Sheep	70	56	126
Shoes	57	60	117
Total:	127	116	243

Table 1: Number of target word tokens included in the acoustic analyses.

Target words were low-pass filtered at a 7.5-kHz cutoff frequency and then sampled at a rate of 15,000 12-bit samples per second using the KAY Elemetrics Computerized Speech Laboratory (CSL) Software System. CSL was also used for acoustic analysis of the tokens.

Formant and fundamental frequency (F0) measurements were made at the onset, offset and durational center of each vowel. Onset measurements were taken when F1, F2, and F3 became visible after the cessation of frication noise, and offset measurements were taken just prior to the disappearance of F2 and/or F1 preceding closure for the final stop or onset of frication noise for the final fricative. These criteria allowed formant transitions to be included as part of the vowel.

In order to estimate the location of formants across a wide range of F0, the relative intensities of harmonics were examined using narrow band spectrograms and FFT and autocorrelation LPC spectra, and formant locations were interpolated from these. Bandwidth, spectrogram preemphasis, filter order and frame length were set by the user as appropriate for the token at hand. To locate formants in areas where formant energy was weak, a series of FFTs was produced at 30- to 40-ms intervals, beginning prior to and continuing after the measurement point, and including at least one region in which the formant was clearly visible from the spectrogram. This allowed formant movement to be dissociated from harmonic movement, permitting formant frequency to be more precisely estimated.

Pitch extraction was carried out using autocorrelation analyses with all parameters set by the user as appropriate for the token at hand.

When formant and F0 measurements were complete, 10% of the tokens were selected for measurement reliability checks. Measurement error was calculated as the difference between the original and error-check values, taken as a percentage of the original measurement. Error was 5.3% for /i/ and 5.6% for /u/.

3. RESULTS AND DISCUSSION

Vowel duration, F0 and pitch excursion were averaged across speakers for each target word. Pitch excursion was calculated as the absolute change in F0 from vowel onset through vowel center to vowel offset. The results are summarized in Table 2. The infant-directed vowel tokens showed longer durations, higher F0, and larger pitch excursions than the adult-directed tokens, as is typical of motherese.

	Sheep		Shoes	
Measurement	**Infant**	**Adult**	**Infant**	**Adult**
Duration (ms)	140	113	264	152
F0 (Hz): Onset	310	208	317	202
Center	328	188	326	179
Offset	310	179	298	173
F0 Excursion (Hz)	106	42	173	48

Table 2: Average duration and F0 of infant- and adult-directed vowels.

F1, F2, and F3 frequencies for each individual vowel token were converted to Mels, and central F1/F2 values were plotted in Figure 1. As can be seen from the figure, the infant-directed tokens are more widely distributed along F1 and F2 than the adult-directed tokens. However, infant-directed /i/ and /u/ frequently show more extreme F2 values than adult-directed /i/ and /u/. F2 in infant-directed /i/ tends to shift to higher values, while in infant-directed /u/, F2 tends to shift to lower values. This results in greater acoustic separation between the infant-directed /i/ and /u/ categories than the adult-directed categories, and should also result in greater separation between /i/ and /u/ and other English vowel categories. This tendency to more extreme F2 values for infant-directed /i/ and /u/ is evident for each of the 10 mothers who participated in the study.

Figures 2 and 3 show F1, F2, and F3 values across the vowel for infant- and adult-directed "sheep" and "shoes," averaged across speakers. Average F1 values are very similar in the infant- and adult-directed vowels. F2 patterns differ in the infant- and adult-directed vowels. In /i/ (Fig 2) infant-directed F2 is higher throughout the vowel than adult-directed F2. In /u/ (Fig. 3) the onset value for infant-directed F2 is slightly higher than adult-directed F2, but moves to lower values than adult-directed F2 for the remainder of the vowel. This F2 pattern suggests that coarticulation of /u/ with the initial fricative in "shoes" is equivalent or somewhat more pronounced in infant-directed than adult-directed tokens. Informal listening confirms this impression: infant-directed tokens of "shoes" tend to be produced as /šyuz/ rather than /šuz/. In spite of this coarticulation at vowel onset, F2 again reaches acoustically more extreme values by the center of infant-directed /u/ than it reaches in adult-directed /u/.

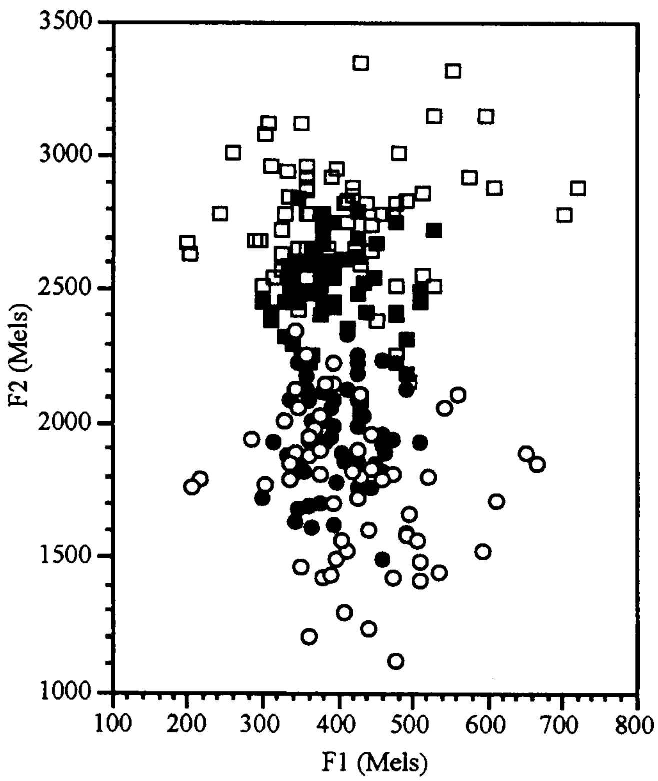

Figure 1: Central F1/F2 values for all vowel tokens. Values for /i/ in "sheep" are represented by squares, values for /u/ in "shoes" by circles. Infant-directed tokens are white, and adult-directed tokens are black.

Separate 2-way ANOVAs with Addressee (Infant & Adult) and Measurement (F0, F1, F2 & F3 at onset, center and offset) as within-subjects factors were conducted on F0 and formant values in "sheep" and "shoes." Significant main effects were found for both factors for /i/ and /u/. The interactions were also significant. (For "sheep" F and p for Addressee, Measurement, and AxM are 108.567, $\leq .0001$; 5128.241, $\leq .0001$; and 7.779, $\leq .0001$ respectively. For "shoes" F and p are 20.755, $\leq .0014$, 1686.308, $\leq .0001$ and 14.357, $\leq .0001$ respectively.)

Simple effects for the AxM interactions are summarized in Table 3. F0 is significantly higher throughout infant-directed /i/ and /u/ than the adult-directed tokens. No significant differences were found between infant- and adult-directed F1 in either vowel. In "sheep," infant-directed F2 is significantly higher throughout the vowel than in adult-directed tokens. In contrast, in "shoes," infant- and adult-directed F2 differ significantly only at the vowel center, where infant-directed F2 is significantly lower than in adult-directed tokens. Infant-directed F3 is significantly higher than adult-directed F3 throughout both /i/ and /u/.

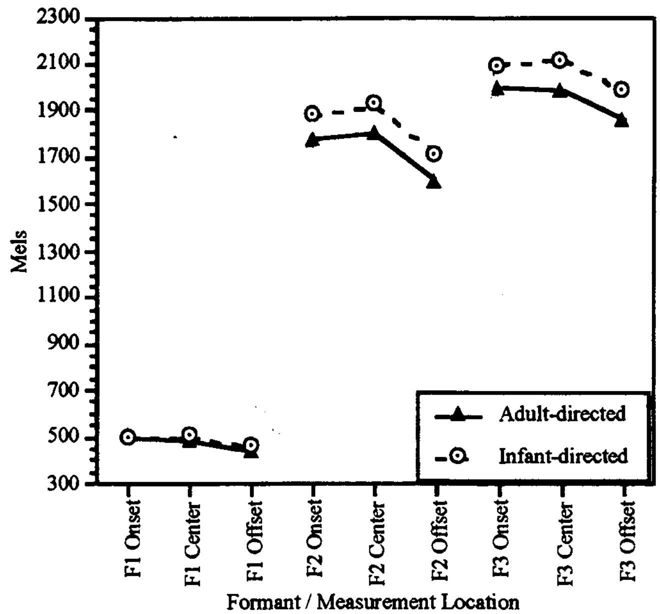

Figure 2: Average F1, F2, and F3 values in infant- and adult-directed "sheep."

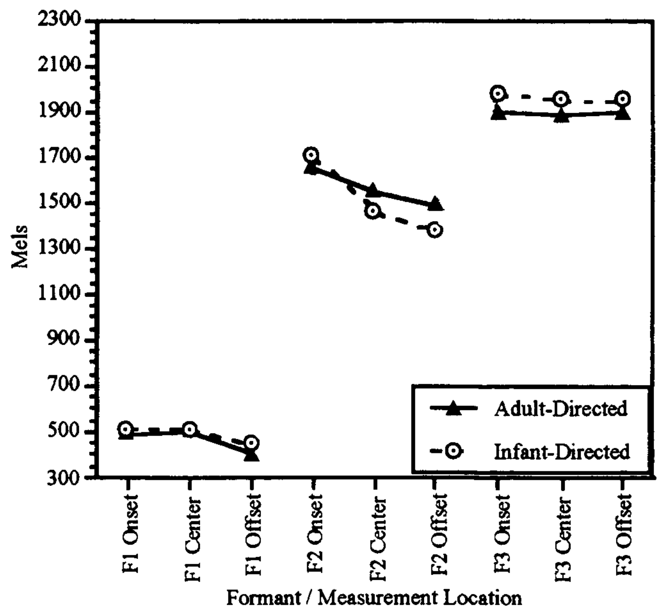

Figure 3: Average F1, F2, and F3 values in infant- and adult-directed "shoes."

Can the difference between infant- and adult-directed formant values, and particularly F2 values, be attributed to differences in vocal pitch? The answer appears to be no. Although increases in F0 are sometimes correlated with increased formant frequencies, in these data formant frequencies do not uniformly increase with F0. That is, F1 does not increase significantly in either vowel. In addition, F2 shifts in *opposite* directions in infant-directed /i/ and /u/. Importantly, in both cases F2 shifts in a direction which makes the vowel acoustically more distinct from surrounding vowels. This suggests that mothers hyperarticulate in speaking to their infants, resulting in increased acoustic separation between vowel categories in infant-directed speech.

	/i/ in Sheep		**/u/ in Shoes**	
Measurement	***F***	***p* (≤)**	***F***	***p* (≤)**
A at F0 Onset	39.908	0.001	76.091	0.001
A at F0 Center	35.195	0.001	120.858	0.001
A at F0 Offset	49.662	0.001	53.610	0.001
A at F1 Onset	0.107	0.751	2.847	0.126
A at F1 Center	0.348	0.570	0.490	0.502
A at F1 Offset	0.904	0.367	1.433	0.262
A at F2 Onset	18.517	0.002	2.429	0.154
A at F2 Center	45.898	0.001	15.928	0.003
A at F2 Offset	12.919	0.006	4.319	0.067
A at F3 Onset	16.509	0.003	27.632	0.001
A at F3 Center	32.018	0.001	17.336	0.002
A at F3 Offset	51.434	0.001	9.715	0.012
M at Adult	3197.96	0.001	2802.92	0.001
M at Infant	2302.14	0.001	730.494	0.001

Table 3: Summary of simple effect results. A=Addressee and M=Measurement.

If both /i/ and /u/ are hyperarticulated in infant-directed speech, why does F1 not shift in value as well as F2? In order to increase the acoustic distance between /i/, /u/ and other English vowels, F1 must move downward. However, F1 tends to be on or just above the first harmonic in infant-directed tokens. Thus, it is generally impossible for infant-directed F1 to shift to a lower value.

The significant upward shift in infant-directed F3 is greater in magnitude in /i/ than in /u/, and may play a role in producing a more "extreme" percept for infant-directed /i/. However, its role in infant-directed /u/ is unclear, and it may instead relate to voice quality in infant-directed speech.

4. CONCLUSION

This study indicates that the acoustic structure of the vowels /i/ and /u/ in mothers' speech to infants differs from the structure of the same vowels in mothers' speech to adults. In conversational speech, infant-directed tokens of /i/ in "sheep" and /u/ in "shoes" showed greater acoustic separation than adult-directed tokens of /i/ and /u/. This increase in acoustic distance between vowel categories was the result of shifts in F2 towards values which make /i/ and /u/ acoustically more distinct from each other and from neighboring English vowels.

Shifts in F1 value made no contribution to increasing the acoustic distance between /i/ and /u/ in these data. Due to significant increases in the value of F0, however, F1 in /i/ and /u/ could not shift downward, in the expected direction. If the tendency to increase the acoustic separation between vowel categories found here holds for the remaining English vowel categories, it should be expected that F1 will be more free to shift in value in lower vowels.

We speculate that the acoustic structure of vowels in motherese speech contributes to infants' acquisition of vowel categories prior to the infant's acquisition of word meaning. Motherese speech may support the infant's acquisition of native-language vowel categories by providing more invariant cues to vowel identity, by making these cues more salient, and by increasing between-category acoustic differences.

The formant value changes required to make vowels in the interior of English vowel space more distinct from one another are much less transparent than those required for /i/ and /u/. In the interior of English vowel space, shifts in F1 and F2 may result in the percept of a different vowel, rather than an "enhanced" percept of the intended vowel. Other acoustic characteristics may make important contributions to distinguishing these vowels from one another, however, and these characteristics could potentially be emphasized in motherese speech. In particular, vowel-inherent formant movement may provide an important cue to the identity of English vowels, and may be exaggerated in infant-directed speech.[7]

Several crucial steps remain in determining the importance of motherese speech for infant acquisition of speech sound categories. We are currently collecting data to investigate the cross-linguistic validity of the results reported here, and to examine the acoustic structure of vowels in the interior of vowel space.

5. REFERENCES

[1] Chomsky, N. *Aspects of the theory of syntax,* Cambridge, Mass: MIT Press, 1965.

[2] Liberman, A. & I.G. Mattingly. "The motor theory of speech perception revised," *Cognition*, 21(1):1-36, 1985.

[3] Kuhl, P.K. "Learning and representation in speech and language," *Current Opinion in Neurobiology*, 4:812-822, 1994; Kuhl, P.K. "Innate predispositions and the effects of experience in speech perception: The native language magnet theory" in B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (eds.) *Developmental neurocognition: speech and face processing in the first year of life*. Boston: Kluwer, 1993.

[4] Kuhl, P.K. , K.A. Williams, F. Lacerda, K.N. Stevens, & B. Lindblom. "Linguistic experience alters phonetic perception in infants by 6 months of age" *Science* 255: 606-608, 1992.

[5] Grieser, D.L. & P.K. Kuhl. "Maternal speech to infants in a tonal language: support for universal prosodic features in motherese," *Developmental Psychology*, 24(1):14-20, 1988.

[6] Fernald, A. & P.K. Kuhl. "Acoustic determinants of infant preference for motherese speech," *Infant behavior & development* 10:279-293, 1987.

[7] Andruski, J.E. & T. Nearey. "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables" *Journal of the Acoustical Society of America*, 91:390-410, 1992.