Cross-Language Analysis of Phonetic Units in Language Addressed to Infants

Patricia K. Kuhl,* Jean E. Andruski, Inna A. Chistovich, Ludmilla A. Chistovich, Elena V. Kozhevnikova, Viktoria L. Ryskina, Elvira I. Stolyarova, Ulla Sundberg, Francisco Lacerda

In the early months of life, infants acquire information about the phonetic properties of their native language simply by listening to adults speak. The acoustic properties of phonetic units in language input to young infants in the United States, Russia, and Sweden were examined. In all three countries, mothers addressing their infants produced acoustically more extreme vowels than they did when addressing adults, resulting in a "stretching" of vowel space. The findings show that language input to infants provides exceptionally well-specified information about the linguistic units that form the building blocks for words.

The emergence of language in a child depends on linguistic input. Socially isolated children (*1*), and profoundly deaf children who experience neither oral nor manual language (*2*), do not acquire language. Recent findings highlight the impact of natural language input on normally developing infants. For example, cross-cultural studies of speech perception show that simply listening to ambient language results in infants' acquisition of information about the phonetic, phonotactic, and prosodic regularities of their native language (*3, 4*). This learning alters infants' perceptual systems, tuning them to the properties of their native language before word learning (*5*). Moreover, in recent studies of speech production, 5-month-old infants were shown to produce specific speech sounds after short-term exposure to them in a laboratory setting, which suggests that language listening also affects speech production at an early age (*6*).

Because early linguistic experience alters speech perception, theorists' attention has focused on language input to infants. Research has established that speech directed to infants (often termed "parentese") is syntactically and semantically simpler than speech directed to adults (*7*). Moreover, cross-cultural studies have shown that infant-directed speech has a unique acoustic signature: It is produced with a higher fundamental frequency (pitch), exaggerated intonation contours, and a slower cadence (*8*). Laboratory tests

show that, when given a choice, young infants prefer infant-directed over adult-directed utterances, and this preference is governed by the intonational features of infant-directed speech (*9*).

Thus, linguistic input to infants is modified syntactically, semantically, and prosodically. A remaining question is whether the phonetic units themselves are modified in infant-directed speech in a way that might enhance learning (*10*). Phonetic units in adult-directed speech are often poorly specified. Vowel and consonant articulations undershoot their intended targets (*11*), resulting in an overlap in the acoustic cues specifying distinct categories (*12, 13*). The phonetic units of adult-directed speech may thus provide a poor signal from which to learn, contributing to the argument that language input to the child underspecifies the information needed for language acquisition (*14*).

We examined natural language input to infants in the United States, Russia, and Sweden. The results show that across all languages, there is an alteration of the phonetic units in infant-directed speech. Parents addressing their infants produce vowels that are acoustically more extreme, resulting in an expanded vowel space, one that is acoustically "stretched."

Ten native-speaking women were audiotaped in two experimental conditions in each of the three countries. In one condition, women were speaking with their 2- to 5-month-old infants (*15*). In the other, the same women spoke to an adult native speaker. Native-language words containing the vowels /i/, /a/, and /u/ were preselected for analysis in the three languages (*16*). Acoustically, the space encompassing vowels forms a "vowel triangle" whose points are determined by the vowels /i/, /a/, and /u/. These three vowels (termed "point

vowels) occur in all the world's languages (*17*).

The hypothesis was that the formant frequencies (*18*) of infant-directed (I) vowels would differ significantly from those of adult-directed (A) vowels. Target words were isolated from each tape-recorded conversation by means of computer-editing techniques. All target words except those obscured by noise or overlapping conversation were digitally sampled for spectrographic analysis (*19*). For each word, 13 acoustic measures were taken: Vowel formant frequencies (F1, F2, and F3) and fundamental frequency (pitch) measures were made at three locations (onset, center, and offset of the vowel) (*20*); vowel duration was also measured. Across all languages, 30,719 measurements were made on 2363 words (1330 I words and 1033 A words). The total number of words included 188 (I) and 141 (A) in English, 175 (I) and 135 (A) in Russian, and 967 (I) and 757 (A) in Swedish.

The results confirm the hypothesis that infant-directed speech exhibits a change in the phonetic units of language when compared with adult-directed speech. Across all three languages, mothers produced acoustically more extreme vowels when addressing their infants, resulting in an expansion of the vowel triangle during infant-directed speech (Fig. 1). Mothers did not simply raise all formant frequencies when speaking to their infants, as they might have done if they were mimicking child speech. Rather, formant frequencies were selectively increased or decreased to achieve an expansion of the acoustic space encompassing the vowel triangle.

Vowel triangle areas in the infant- and adult-directed conditions were compared for each subject. The results were highly consistent. For each of the 30 mothers, the area of the vowel triangle was greater in the I condition than in the A condition ($P < 0.0001$, by binomial test). A Friedman two-way analysis of variance (ANOVA) by ranks (*21*) on the effect of addressee (I versus A), with language (English, Russian, or Swedish) as the blocking factor, confirmed that vowel triangle areas were significantly larger in the I condition ($\chi r^2 = 39.9$, $P < 0.0001$). The degree of vowel triangle expansion was substantial. On average, mothers addressing their infants expanded the vowel triangle by 92% (English, 91%; Russian, 94%; Swedish, 90%). The ratios of mothers' area measures (I/A) across the three languages did not differ (Kruskal-Wallis = 0.38, $P = 0.83$), suggesting that across languages, mothers stretch the vowel triangle to a similar degree.

Analysis of the change in individual formant frequencies showed that they were

P. K. Kuhl and J. E. Andruski, Department of Speech and Hearing Sciences, University of Washington, Box 357920, Seattle, WA 98195, USA.
I. A. Chistovich, L. A. Chistovich, E. V. Kozhevnikova, V. L. Ryskina, E. I. Stolyarova, Early Intervention Institute, 191194, ul. Chaykovskogo, 73, St. Petersburg, Russia.
U. Sundberg and F. Lacerda, Institute of Linguistics, Stockholm University, S-106 91 Stockholm, Sweden.

*To whom correspondence should be addressed.

Fig. 1. Vowel triangles formed by the "point" vowels, /i/ (green), /a/ (red), and /u/ (blue), in infant-directed (solid circles) and adult-directed (open circles) speech in three languages—English, Russian, and Swedish. Each data point represents the coordinate of the first two formant frequencies of a vowel. A universal stretching of the vowel triangle is observed in infant-directed (solid line) relative to adult-directed (dashed line) speech.
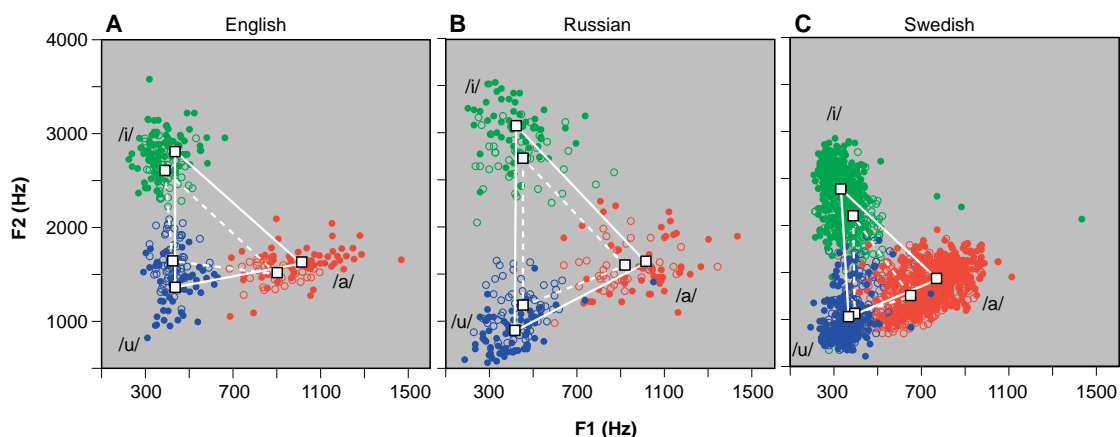
increased or decreased as necessary to achieve a stretching of the vowel triangle (*22*). The results for American English mothers showed increased F2 in /i/, decreased F2 in /u/, and increased F1 and F2 in /a/ (Fig. 1A). Russian mothers showed increased F2 in /i/, decreased F2 in /u/, and increased F1 in /a/ (Fig. 1B). Swedish mothers showed increased F2 and decreased F1 in /i/, decreased F1 in /u/, and increased F1 and F2 in /a/ (Fig. 1C). The range of formant values was greater in I speech in all languages. As expected, significant increases in fundamental frequency and vowel duration were observed in I speech in all languages.

Does a stretched vowel triangle benefit infants? We hypothesize three ways in which it could do so. First, an expanded vowel triangle increases the acoustic distance between vowels, making them more distinct from one another. In recent studies, language-delayed children showed improvements when listening to speech in which between-category phonetic differences were increased (*23*). Normally developing infants have been shown to discriminate smaller differences than those provided by an expanded vowel triangle (*24, 25*), but

may nonetheless benefit similarly from the enhanced acoustic differences provided in infant-directed speech.

Second, to achieve the stretching, mothers produce vowels that go beyond those produced in typical adult conversation. From both an acoustic and articulatory perspective, these vowels are "hyperarticulated" (*26*). Hyperarticulated vowels are perceived by adults as "better instances" of vowel categories (*27, 28*), and laboratory tests show that when listening to good instances of phonetic categories, infants show greater phonetic categorization ability (*29*). Our study shows that hyperarticulated vowels are a part of infants' linguistic experience and raises the possibility that they may play an important role in the development of infants' vowel categories.

Third, expanding the vowel triangle allows mothers to produce a greater variety of instances representing each vowel category without creating acoustic overlap between vowel categories. Greater variety may cause infants to attend to non–frequency-specific spectral features that characterize a vowel category, rather than to any particular set of frequencies the

mother uses to produce a vowel (*30*). As shown in Fig. 2, converting the formant values to spectral features (*31*) in mels (*32*) shows that infant-directed speech maximizes the featural contrast between vowels. This is especially critical for infants because they cannot duplicate the absolute frequencies of adult speech—their vocal tracts are too small (*33*). To speak, infants must reproduce the appropriate spectral features in their own frequency range (*6*). We posit that early in development, representations of speech stored in memory encode such abstract spectral dimensions. According to this view, linguistic input induces infants to attend to features that (i) allow phonetic units spoken by different talkers to be categorized and (ii) provide a non–frequency-specific metric that reveals how equivalent speech units can be produced by the infant's vocal tract.

Language development includes not only the acquisition of a complex grammar, but also the acquisition of a phonological system that allows differences in meaning to be conveyed. The acoustic forms of speech are highly variable, changing with factors

Fig. 2. Formant measures converted to spectral features (in mels) for infant-directed and adult-directed speech. Spectral features describe the acoustic components of vowels in a non–frequency-specific metric (*31*), and mels take into account the fact that at higher frequencies, larger differences are necessary to detect change (*32*). The vowel /i/ has component frequencies that are broadly distributed across the spectrum ("diffuse") and relatively high ("acute"), whereas component frequencies in the vowel /a/ are acute but more concentrated ("compact") and components of /u/ are maximally low ("grave"). The formula for calculating the compact-diffuse feature is F2 − F1; for the grave-acute feature, (F1 + F2)/2.
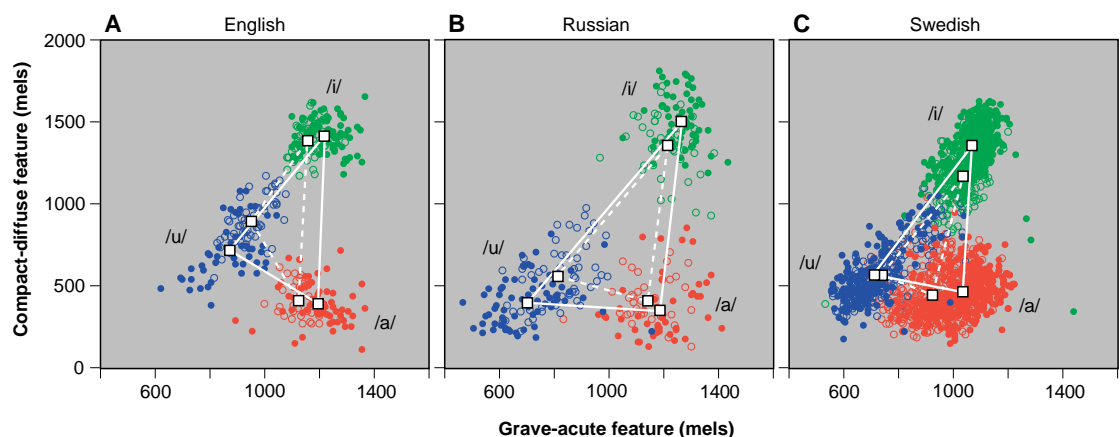
that include speaker gender and identity, speaking rate, and the phonological context of the sound (*34*), which makes sorting ambient language sounds into phonetic categories a complex task. Our results suggest that infant-directed speech assists in this process by delivering information about the sound system of the infant's native language in an exaggerated form. The exaggerated form serves two functions: It more effectively separates sounds into contrasting categories, and it highlights the parameters on which speech categories are distinguished and by which speech can be imitated by the child.

Our results contribute to an emerging view of the role of linguistic input in language development in the child and the type of learning it induces (*5*). According to this emerging view, language input is not a trigger for innately stored information. Moreover, the developmental change that ensues, given language input, is not a process that depends on Skinnerian reinforcement; infants' learning of linguistic regularities shown in recent studies (*3, 4, 35*) cannot be explained on the basis of reinforcement. Language input provides a rich and detailed source of information that instigates, before word learning, a process of species-specific mapping of information by the brain, a process that alters the infant's perceptual and perceptual-motor system to conform to a specific language.

Natural language input is a reliable feature of every typically developing child's experience. Our findings demonstrate that language input to infants has culturally universal characteristics designed to promote language learning. These characteristics are likely to be exploited by infants' developing neural systems.

REFERENCES AND NOTES

1. V. Fromkin, S. Krashen, S. Curtis, D. Rigler, M. Rigler, *Brain Lang.* **1**, 81 (1974); H. L. Lane, *The Wild Boy of Aveyron* (Harvard Univ. Press, Cambridge, MA, 1976).
2. L. A. Petitto, in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, J. Morton, Eds. (Kluwer, Dordrecht, Netherlands, 1993), pp. 365–383.
3. P. W. Jusczyk, A. Cutler, N. J. Redanz, *Child Dev.* **64**, 675 (1993); P. W. Jusczyk, A. D. Friederici, J. M. I. Wessels, V. Y. Svenkerud, A. M. Jusczyk, *J. Mem. Lang.* **32**, 402 (1993).
4. P. K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, B. Lindblom, *Science* **255**, 606 (1992).
5. P. K. Kuhl, *Curr. Opin. Neurobiol.* **4**, 812 (1994).
6. _____ and A. N. Meltzoff, *J. Acoust. Soc. Am.* **100**, 2425 (1996).
7. C. E. Snow, in *Talking to Children: Language Input and Acquisition,* C. E. Snow and C. A. Ferguson, Eds. (Cambridge Univ. Press, Cambridge, 1977), pp. 31–49.
8. A. Fernald and T. Simon, *Dev. Psychol.* **20**, 104 (1984); D. L. Grieser and P. K. Kuhl, *ibid.* **24**, 14 (1988).
9. A. Fernald, *Infant Behav. Dev.* **8**, 181 (1985); _____ and P. Kuhl, *ibid.* **10**, 279 (1987).
10. An earlier report [N. B. Ratner, *J. Child Lang.* **11**, 557 (1984)] noted phonetic modifications in linguistic input to nine older children (specific ages were not given; children ranged from "preverbal," described as older than 9 months of age, up to those with mean length of utterances = 4.0, that is, children assumed to be 2 to 4 years of age). Although the direction of the data was similar to that shown here, no statistical analyses were reported.
11. B. Lindblom, *J. Acoust. Soc. Am.* **35**, 1773 (1963).
12. G. E. Peterson and H. L. Barney, *ibid.* **24**, 175 (1952).
13. J. Hillenbrand, L. A. Getty, M. J. Clark, *ibid.* **97**, 3099 (1995).
14. N. Chomsky, *Rules and Representations* (Columbia Univ. Press, New York, 1981).
15. The mean ages of infants in the three countries were as follows: United States, 14.3 weeks (range, 9.1 to 18.1 weeks); Russia, 15.7 weeks (range, 9 to 22 weeks); Sweden, 13.3 weeks (range, 11 to 17.7 weeks).
16. In English, the words were "bead" for /i/, "pot" for /a/, and "boot" for /u/. In Russian, the words were "knizhka" (book) and "vilka" (fork) for /i/, "lapa" (paw) and "palets" (finger) for /a/, and "busy" (beads) and "ruchka" (pen) for /u/. To make it easier to incorporate preselected words in conversation, we provided mothers with small toys representing the objects. In Swedish, all content words containing the vowels /i/, /a/, or /u/ were analyzed. In all languages, women were recorded on one occasion; the conversation in each condition lasted about 20 min. Women were instructed to speak naturally and (when applicable) to use the preselected words at least three times during each conversation. All analyzed words contained stressed vowels.
17. The three languages were chosen because they represent the substantially different vowel systems that occur across languages: Russian is a 5-vowel system, American English is a 9-vowel system, and Swedish is a 16-vowel system (with the allophonic length variation) [J. Liljencrants and B. Lindblom, *Language* **48**, 839 (1972)].
18. Formants are frequency regions in which the amplitude of acoustic energy is high, reflecting natural resonances created in the vocal tract. Formants are numbered (F1, F2, and so forth, from the lowest frequency upward); the frequency value cited is the center of the band of energy. F1 and F2 are the most important formants for vowel identification and are reported here. See K. N. Stevens, in *Introduction to Communication Sciences and Disorders,* F. D. Minifie, Ed. (Singular, San Diego, CA, 1994), pp. 399–437.
19. All words were low-pass filtered at 7.5 kHz and sampled at 15 to 16 kHz with specially designed software implemented on either a 12- or 16-bit 486 computer. Formant measures on the English and Russian words were made by a trained acoustician (J.E.A.) using narrow-band spectrograms, fast Fourier transform spectra, and autocorrelation linear prediction coefficient spectra. An automatic formant tracker (16 kHz, 18 poles, six formants) was used for the Swedish words. Reliability was assessed in two ways, each using a randomly selected 10% of the words in the data set: (i) Human reliability was assessed by having English and Russian words remeasured by J.E.A. and by a second trained analyst, and (ii) human-machine reliability was assessed by having American English words remeasured by machine and by having Swedish words remeasured by J.E.A. Reliability, expressed as the mean percentage difference between measures of individual formants in each vowel, was uniformly high. For human reliability across all formants and vowels, the mean intrascorer error was 5.95% and the mean interscorer error was 5.43%. For human-machine reliability across all formants and vowels, the mean error was 6.10%.
20. The vowel center formant measure most accurately reflects the speaker's intended frequency and is the one reported here, but all three locations produced the same pattern of results.
21. The nonparametric ANOVA was most appropriate for the data [W. J. Conover, *Practical Nonparametric Statistics* (Wiley, New York, 1980)], but a parametric analysis [multivariate ANOVA (MANOVA)] yielded the same result, $F_{1,27} = 110.64$, $P < 0.0001$.
22. Each formant of each vowel in each language was assessed by MANOVA to examine the effects of addressee. All listed effects are significant at the 0.01 level.
23. P. Tallal *et al., Science* **271**, 81 (1996); M. M. Merzenich *et al., ibid.,* p. 77; M. Studdert-Kennedy and M. Mody, *Psychon. Bull. Rev.* **2**, 508 (1995).
24. P. J. Swoboda, J. Kass, P. A. Morse, L. A. Leavitt, *Child Dev.* **49**, 332 (1978).
25. D. Grieser and P. K. Kuhl, *Dev. Psychol.* **25**, 577 (1989); P. K. Kuhl, *Percept. Psychophys.* **50**, 93 (1991).
26. Lindblom argued that speakers intuitively use "hyperarticulated speech" to accommodate the needs of listeners [B. Lindblom, in *Speech Production and Speech Modeling,* W. J. Hardcastle and A. Marchal, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1990), pp. 403–439].
27. P. Iverson and P. K. Kuhl, *J. Acoust. Soc. Am.* **97**, 553 (1995).
28. K. Johnson, E. Flemming, R. Wright, *Language* **69**, 505 (1993).
29. It is of theoretical interest to compare the values of the American English "prototype" /i/ vowel used in P.K.K.'s laboratory studies (*4, 25, 27*) with the formant values for /i/ produced by mothers in this study. P.K.K.'s prototype was the male speakers' average /i/ in Peterson and Barney's (*12*) data set. When the formant values for Peterson and Barney's female average /i/ were compared with mothers' productions of /i/ in the present study, (i) the formant values fell very near the mean of American English mothers' infant-addressed /i/ (Fig. 1), and (ii) the feature values were virtually identical to the mean shown for American English mothers' infant-directed /i/ (Fig. 2). Thus, the earlier findings showing infants' superior categorization when listening to a "prototype" are consistent with our results showing that mothers hyperarticulate vowels when addressing their infants. In effect, the comparison suggests that P.K.K.'s previous tests used a vowel that would be described as "hyperarticulated." The comparison indicates that the Peterson and Barney formant values (derived from words spoken in citation speech) are not representative of formant values in adult-directed natural speech [see also (*13*)]. Also, in previous studies (*4, 25, 27, 28*), vowels judged to be "best instances" (i) maximize spectral features to a greater extent than do those judged to be poor instances, and (ii) are more extreme than those produced by adults in normal conversation.
30. Greater variety also improves phonetic category learning in foreign speakers [S. E. Lively, J. S. Logan, D. B. Pisoni, *J. Acoust. Soc. Am.* **94**, 1242 (1993)].
31. G. Fant, *Speech Sounds and Features* (MIT Press, Cambridge, MA, 1973); R. Jakobson, C. G. M. Fant, M. Halle, *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates* (MIT Press, Cambridge, MA, 1969).
32. S. S. Stevens, J. Volkmann, E. B. Newman, *J. Acoust. Soc. Am.* **8**, 185 (1937). The mel scale equates the magnitude of perceived differences in pitch at different frequencies.
33. R. D. Kent and A. D. Murray, *ibid.* **72**, 353 (1982).
34. J. S. Perkell and D. H. Klatt, Eds., *Invariance and Variability in Speech Processes* (Erlbaum, Hillsdale, NJ, 1986).
35. J. R. Saffran, R. N. Aslin, E. L. Newport, *Science* **274**, 1926 (1996).
36. Supported by the William P. and Ruth Gerberding Professorship and NIH grant DC 00520 (P.K.K.), a grant from the Social Sciences and Humanities Research Council of Canada (J.E.A.), and Bank of Sweden Tercentenary Foundation grant 94-0435 (F.L.). We thank E. Stevens for help on all aspects of the data analysis and A. N. Meltzoff for helpful comments on an earlier draft of the manuscript.

21 January 1997; accepted 17 June 1997