AN ACOUSTIC COMPARISON OF VOWEL SYSTEMS IN ADULT-DIRECTED-SPEECH AND CHILD-DIRECTED SPEECH: EVIDENCE FROM FRENCH, ENGLISH & JAPANESE

**Christelle Dodane and **Jalaleddin Al-Tamimi*

*Laboratoire Dipralang, Université Paul Valery, Montpellier III, Montpellier, France
**Laboratoire Dynamique du Langage UMR CNRS 5596, Université Lyon II, Lyon, France
christelle.dodane@univ-montp3.fr ; Jalal-Eddin.Al-Tamimi@univ-lyon2.fr

ABSTRACT

This research investigated the role of child-directed speech in the acquisition of vowel systems in a cross-linguistic perspective. In order to determine if vocalic systems are expanded in child-directed speech and if this extension varies cross-linguistically, child-directed speech (CDS) was compared to adult-directed speech (ADS) in three rhythmically different languages: French, English and Japanese. The same short story was successively read by mothers to their infant and to an adult (5 mother-infant dyads per language). In the three languages, mothers tended to produce more opened vowels in CDS than in ADS, resulting in a downward shift of the vowel space on the high-low dimension (F1).

Keywords: Child-directed speech, Adult-directed speech, point vowels, Hypo- & Hyper-articulation.

1. INTRODUCTION

When adults are in interaction with infants, they adopt a special way to talk, child-directed speech (CDS) or motherese. This speech modality is characterized by an extreme simplification of the syntaxic, semantic and phonological levels and a dramatic exaggeration of the prosodic level [4, 11, 13]. In CDS, adults use a 3 or 4 semi-tones higher pitch, a greatly expanded pitch range and a high degree of pitch continuity [4]. With respect to other prosodic parameters, CDS utterances are produced with much more regularity and rhythmicity. In particular, speech rate dramatically slows down and varies in function of the attentional state of the baby [11]: The more attentive the baby is, the slower the rate is. For Lindblom [8], speech articulation can be located on a continuum from hypo-articulation to hyper-articulation (H&H theory). According to this theory, speakers adapt their speech rate to listeners, and vice-versa. These rate and rhythmic characteristics of CDS classify this speech modality as hyper-articulated (greater amplitude and duration of articulatory gestures). At the segmental level, Kuhl & al. [7] observe the same effect in comparing adult-directed speech (ADS) vs. (CDS) in the production of the three point vowels /i/, /a/ and /u/. In English, Swedish and Russian, vowels are lengthened and made salient by prosodic contours. In particular, mothers produce acoustically more extreme vowels (Hyper-articulation). The result of this study is the observation of an expansion of the vowel triangle from ADS to CDS on two axes, the high-low dimension (first formant, F1) and the front-back dimension (second formant, F2). These results are consistent with the Hyper-articulation~Hyper-space effect [6]. Kuhl & al. [7] consider these vowels as good instances of vowel categories [6]. In this study, the three languages belong to the same rhythmic class, stress-timed languages [1]. In the light of the H&H theory, it is important to compare languages with different rhythmic classes like syllable-timed languages and moraïc-timed languages. Moreover, words used in Kuhl's study were highly constrained by the consonant context (three specific monosyllabic CVC words) and the situation of production (mothers had to produce the words at least three times during a 20 minute conversation). If these kinds of constraints facilitated the acoustic comparison of vowel formants, they didn't provide a very natural speech condition. In fact, mothers had to focalize on the production of the specific words used in the study. It would be very interesting to compare Kuhl's results with another speech conditions, like telling a story to a child. This represents a much more natural speech condition, even if the speech is not spontaneous and if it would certainly provide much more variability in the production of vowels. In this paper, we propose to compare cross-linguistically the CDS version of a story to the ADS version, exactly the same story being read to a child and to an adult. In order to compare languages from different rhythmic classes and with different vowel systems, we chose English, French & Japanese. Traditionally, English is classified as a stress-timed language, French, a syllable-timed language and Japanese, a moraic-timed language. Considering the size of their vocalic repertory (oral and monophtongs only), British English is a 12-vowel system, French, a 12-vowel system and Japanese, a 5-vowel system. In our corpus, we will compare the production of the three point vowels /i/, /a/ and /u/ in the two modalities and in the three languages.

2. METHODOLOGY

2.1. Speakers

15 volunteer mothers took part in the experiment: 5 native speakers of British English, 5 native-speakers of French and 5 native-speakers of Japanese. All mothers were recorded in Lyon (France). English & Japanese mothers were recruited only if they spoke to their baby in their native language since birth. The babies' ages ranged between 6 months and 22 months (average age: 13 months +/- 5 months).

2.2. Recordings

Mothers were recorded at home by the father or a familiar person in order to obtain the most natural speech conditions. They had to read the same story to a familiar adult (ADS modality) and to their child (CDS modality). They were recorded over two sessions, in order for modalities to be clearly differentiated, beginning with either ADS or CDS. The story was taken from a French children's book. It was translated into English and into Japanese by native speakers (for some audio examples: ADS then CDS, English <audio_file_1.wav>, French <audio_file_2.wav>, and Japanese <audio_file_3.wav>). Recordings were made with a minidisk recorder (Sony MZN-910S) and an unidirectional Philips SBC-MD695 microphone. Each word was labeled for its generic grammatical class. For Japanese, coding was verified by a linguist and a native-speaker of Japanese.

2.3. Acoustic analyses

Recordings were sampled at 22 kHz, 16 bits, mono. Data were segmented manually, and acoustic measurements of the first 3 formant frequencies were carried out with Praat, using "Burg" algorithm (equivalent to a 24 LPC coefficients (autocorrelation)) with a 12.5 ms Gaussian window, and a 5 ms step, at the temporal mid-point for /i/, /a/ and /u/. Formant frequencies were manually verified to prevent automatic error extraction values, and then converted to Barks [9]. In addition, fundamental frequency ($f0$ in Hz) was extracted automatically for all segments. Across the three languages, 20360 measurements were made on 965 vowels in ADS and on 1071 vowel in CDS (see Table 1). The number of vowels is different in ADS and in CDS because some measurements were removed (errors in automatic formant extraction).

Table 1: Number of vowels per language and modality.

	ADS			CDS		
	i	a	u	i	a	u
English	119	40	25	120	40	23
French	119	125	60	133	144	80
Japanese	151	233	93	175	249	107

2.4. Statistical analyses

A multivariate analysis (MANOVA) was conducted to evaluate the effect of speech modality (ADS, CDS), language (English, French, and Japanese) and speaker (15 mothers). Additional Student T-Test was applied on F1, F2, and the distance between $f0$ and F1 (F1-$f0$) in each vowel. A correlation test was conducted to assess the effect of child age on the size of vowel dispersion areas. In order to assess the validity of the expansion of vowel space in CDS in comparison with ADS (Hyper-Articulation in CDS), vowel dispersion areas were calculated (Convex Hull).

3. RESULTS

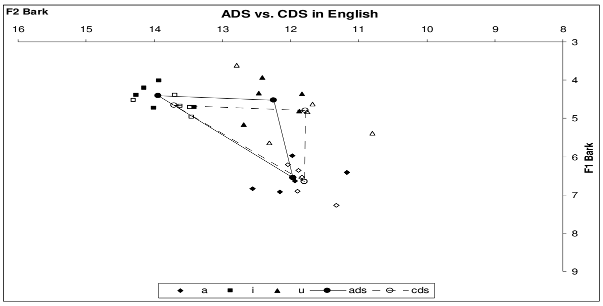
3.1. General results

Previous analysis carried out on these corpora show a significant prosodic exaggeration from ADS to CDS across languages, in particular a slower speech rate in CDS across languages (4.14 syll./sec. in ADS and 3.26 syll./sec. in CDS in English, 5.39 syll./sec. in ADS and 4.26 syll./sec. in CDS in French, 5.78 syll./sec. in ADS and 4.73 syll./sec. in Japanese) and a higher pitch (+ 4 semi-tones in English, + 3 semi-tones in French and + 1 semi-tone in Japanese). The increase of $f0$ values in CDS may affect F1 values [5]. Whereas F1 is significantly affected by modality ($F(1,1962)=40{,}42$; $p<.001$), F2 and F3 are not. Thus, we have calculated the F1-$f0$ difference for each language. This distance is decisive for the perception of openness [5].

3.2. English ADS vs. CDS

In English, results show that the vowel triangle area does not expand from ADS to CDS, but on the contrary, is slightly reduced by 3.49 % with a mean ADS area of 1.72 Barks^2 and a mean CDS area of 1.66 Bark^2 (non significant, ns), (for individual results, see <image_file_1.jpg>). Nevertheless, the most noticeable result is a downward shift of the vowel space on the high-low dimension (F1) from ADS to CDS (see Figure 1).

Figure 1: Mean vowel triangles formed by the point vowels /i/, /a/ and /u/ in ADS (solid line) and CDS (dotted line) for English, and individual values.

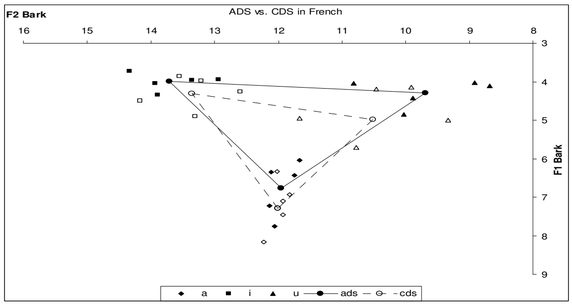


In particular, the high vowel /i/ is more open in CDS (higher F1 value) ($p<.05$). For /a/ and /u/, there is no such difference between modalities (ns). On the front-back dimension (F2), there is a shift of the vowel triangle to the right of the vocalic space: /i/ is more backed (lower F2 value) ($p<.05$), and no significant effects for /a/ and /u/ are noticeable. Nevertheless, we observe for /u/ some backing effect on the chart, which can be caused by inter-speaker differences. We observe a high correlation degree between the children age and vowel areas ($r=0.94$) in CDS: Vowel area size increases as the child age increases. Compared to French and Japanese, the F1-$f0$ difference (in Barks) is more reduced in CDS than in ADS for the three vowels /i/, /a/ and /u/. This difference could be explained by an important increase of $f0$ from ADS to CDS in English (+ 4 semi-tones) compared to French and Japanese. For both modalities, if the F1-$f0$ distance is greater for the low vowel /a/, this distance is much more reduced for the closed vowels /i/ and /u/ because of their lower F1 (see <image_file_2.jpg>).

3.3. French ADS vs. CDS

As can be seen on Figure 2, the vowel triangle is much larger in French than in English for both speech modalities. The three vowels /i/, /a/ and /u/ are more extreme in French. Nevertheless, the vowel triangle area is reduced by 24.35% in CDS with a mean ADS area of 5.34 Barks^2 and a mean CDS area of 4.05 Barks^2 (ns), (see <image_file_1.jpg> for individual results). This reduction is greater than in English. We observed, as in English, a downward shift of the vowel triangle on the high-low dimension (F1) from ADS to CDS (see Figure 2) for the three vowels ($p<.01$ for /i/, $p<.001$ for /a/, and $p<.0001$ for /u/).

Figure 2: Vowel triangles formed by the point vowels /i/, /a/ and /u/ in ADS (solid line) and CDS (dotted line) for French, and individual values.

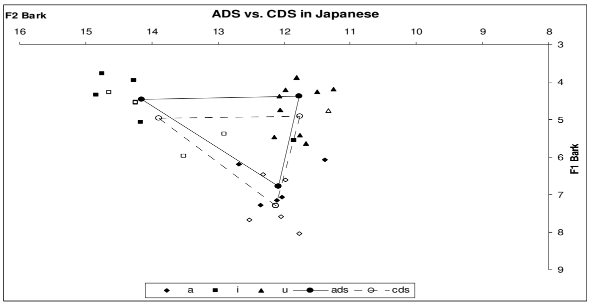


On the front-back dimension (F2), /u/ is fronted ($p<.0001$) whereas the vowel /i/ is backed in CDS ($p<.01$) and there is no significant difference for /a/ between both modalities. We observed a high correlation degree between the children age and vowel areas ($r = 0.72$) in CDS: Vowel area size increases as the child age increases. Contrary to English, the F1-$f0$ difference (in Barks) is not as reduced from ADS to CDS for the vowels /a/ and /u/ (see <image_file_3.jpg>), even if it is the case for /i/ ($p<.01$).

3.4. Japanese ADS vs. CDS

In Japanese, we obtained a similar phenomenon as in English and French (see Figure 3). The vowel triangle area is slightly reduced by 4.6 % in CDS (mean area size of 2.83 Barks^2 in ADS and 2.70 Barks^2 in CDS (ns)), (see <image_file_1.jpg> for individual results). Here again, there is a downward shift of the vowel triangle on the high-low dimension for the three vowels /i/ ($p<.001$), /a/ ($p<.0001$) and /u/ ($p<.001$) (higher F1 values) across modalities. This effect is more noticeable than in English and French because it concerns all the vowels. In CDS, on the front-back dimension, /i/ is slightly backed ($p<.05$), and no significant differences for /a/ and /u/ are noticeable. We observed a negative correlation between children's age and vowel areas ($r = -0.49$) in CDS: Vowel area size decreases as the child age increases. In comparison with English and French, F1-$f0$ difference (in Barks) in Japanese is not reduced from ADS to CDS (see <image_file_4.jpg>).

Figure 3: Vowel triangles formed by the point vowels /i/, /a/ and /u/ in ADS (solid line) and CDS (dotted line) for Japanese, and individual values.



4. DISCUSSION AND CONCLUSIONS

If our study reveals some differences in the production of the three vowels /i/, /a/ and /u/ across speech modalities and across languages, our results are different from Kuhl's [7]. We didn't notice a "stretching" in both dimensions of vowel spaces in CDS, but rather a shift of the vowel triangle on the high-low dimension (F1): Closed vowels are more opened (higher F1) in CDS than in ADS in English, French and Japanese. Nevertheless, F1 values could be affected by increase of $f0$ values from ADS to CDS [5]. These effects are particularly strong in English, where the F1-$f0$ difference is more reduced compared to French and Japanese. On the front-back dimension (F2), the observed variability can be explained by consonantal influences [14]. Our results are not consistent with the hyper-articulation hypothesis, rather a mixture of Hypo- and Hyper-articulation. Thus, the back vowel /u/ was less extreme on the front-back dimension. Vowel dispersion area differences could be explained by the different positions of "point" vowels across the 3 languages. Even if some features could be explained by some differences between the two corpora (situation of production, speech style and greater variability in consonantal contexts), the downward shift of the triangle is quite striking. Why do mothers produce more opened vowels when they address to their child? One explanation involves the importance for the baby to match sounds and articulatory gestures produced by his mother. The gestures carried out to produce more opened sounds, much more visually salient and thus, easier to extract. Another explanation could be that mothers try to imitate the productions of their babies: With a smaller vocal tract, babies produce more opened (higher F1) and more fronted (higher F2) vowels [10]. The "age" variable cannot validly been discussed, as the babies' ages ranged between 6 and 22 months; although, we observed in English and in French an effect of the babies' age on vowel area size. These results are consistent with the fact that CDS characteristics evolve with the age and abilities of the child [9]. Vowels and consonant could be pronounced more distinctively when the child begins to speak than when he is younger or older. These questions are quite fascinating and further investigation is needed to explore the effect of age on the acoustic characteristics of vowels in CDS.

5. REFERENCES

[1] Abercrombie, D. 1967. *Elements of General Phonetics*. Chicago: Aldine.

[2] Broen, P.A. 1972. The verbal environment of the language-learning child. *American Speech and Hearing Association Monographs*, 17.

[3] Fernald, A. 1989. Intonation and communication intent in mother's speech to infants: Is the melody the message? *Child Development*, 60, 1497-1510.

[4] Fernald, A., Simon, T. 1984. Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20/1, 104-113.

[5] Fujisaki, H., Kawashima, T. 1968. The roles of pitch and higher formants in the perception of vowels. *IEEE Trans. Audio. Electroacoust*, 1, 73-77.

[6] Johnson, K., Flemming, E., Wright, R. 1993. The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 69, 505-528.

[7] Kuhl, P., Andruski, J., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ryskina, V., Stolyarova, E., Sundberg, U., Lacerda, F. 1997. Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277, 684-686.

[8] Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds) *Speech Production and Speech Modelling*, Dordrecht: Kluwer Academic, 403-439.

[9] Malsheen, B. 1980. Two hypotheses for phonetic clarification in the speech of mothers-to-childen. In G. Yeni-Komshian, J.F. Kavanagh and C.A. Ferguson (Eds) *Child Phonology: Perception*, New-York: Academic Press, II, 173-184.

[10] Ménard, L., Schwartz, J.L., Boë, L.J. 2002. Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *J. Acoust. Soc. Amer.* 111(4), 1892-1905.

[11] Papousek, M., Papousek, H. 1981. Musical elements in the infant's vocalizations: Their significance for communication, cognition and creativity". In *Advances in Infancy Research*, 1, XXII, Norwood Lipsitt-Collier, Norwood Ablex, 163-224.

[12] Schroeder, M.R., Atal, B.S., Hall, J.L. 1979. Optimizing digital speech coders by exploiting masking properties of the human ear. *J. Acoust. Soc. Amer.* 66, 1647-1652.

[13] Stern, D., Spiker, S., Mac Kain. 1983. The prosody of maternal speech: Infant age and context related changes. *J. Child Lang*, 10, 1-15.

[14] Stevens, K., House, A. 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. *J. Speech and Hear. Res.* 6, 111-128.