Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Learning the sound system of Japanese: What does it tell us about language acquisition?

Reiko Mazuka (1, 2)

(1) Laboratory for Language Development, RIKEN Brain Science Institute, Wako-shi, Japan
(2) Department of Psychology and Neuroscience, Duke University, Durham, USA

PACS: 43.71.Es, 43.71.Ft, 43.71.Hw.

ABSTRACT

Infants learn much about the phonology of their own language during the first year of their lives. To date, however, the vast majority of the research on infant speech perception has been carried out with infants learning English and other European languages, and we know very little about how infants learning other languages learn the sound system of their languages. The phonological characteristics of Japanese differ from English and other European languages in important ways, and investigation of its acquisition has a potential of shedding important light onto our understanding of phonological acquisition. In this paper, we present data from Japanese are presented to exemplify this point; acquisition of mora-timed rhythm, edge-prominent prosody, lexical pitch-accent and segmental distribution.

INTRODUCTION

Learning the sound system of a language is one of the first task infants are faced with in language acquisition. Research in infant speech perception during the past several decades have revealed many aspects of how infants go about this task. For segmental side, infants begin their life with an ability to discriminate most (but not all) of phonemic contrasts employed in human languages. But by the second half of their first year, infants' sensitivity to non-native contrasts begins to decline while their sensitivity to the native ones becomes more robust. This shift from a general sensitivity to all languages to a selective-sensitivity to the language they are acquiring is sometimes called "reorganization"[1, 2]. Similar developmental changes have also been reported for prosodic [3] as well as phonotactic properties [4, 5] of a language's phonological system.

To date, however, the vast majority of the research on infant speech perception has been carried out with infants learning English and a few other European languages, and little data has been gathered with infants learning other languages. For some phonological characteristics, however, Indo-European languages represent only part of the range of the variations, and the lack of empirical evidence from non-European languages could jeopardize our view of how phonological system is acquired by human infants. In this paper, we present data from Japanese that exemplify this point.

Japanese is a head-final, right-branching language with a Subject-Object-Verb word order. Syntactictically, Japanese shares many of its properties with other languages of the same type. Phonologically, however, some of the characteristics of Japanese are not widely shared with other languages.

MORA-TIMED RHYTHM

Japanese is said to have a mora-timed rhythm [6, 7]. Mora-timed rhythm is one of three types of linguistic rhythm, others being a stress-timed rhythm of languages such as English, German, and Dutch, and a syllable-timed rhythm of languagees such as French, Spanish and Italian. Other languages such as Luganda (a Bantu language spoken in Uganda) and Gilbertese (an Austronesian language spoken in Kiribati and other South Pacific islands) have been argued to have mora-timed rhythm. Still, Japanese remains to be the only language that is spoken by a large enough number of speakers in an industrialized contry, such that it is realistic to investigate its acquisition empirically.

Research during the past several decates has demonstrated that the rhythm of a language plays important roles in speech perception and production as well as lexical processing [8, 9]. In language acquisition, neonates are able to discriminate languages on the basis of their rhythms[10], and linguistic rhythm may bootstrap infants' speech segmentation and lexical acquisition [11]. Such proposal presupposes that infants learning any language should be able to use the rhythm of their language for segmentation. Yet, the proposal was built solely on the basis of infants learning stress-timed and syllable-timed languages[12].

In a recent study, Japanese infants' ability to discriminate long versus short vowels in word pairs such as /maana/ vs. /mana/ was investigated [13]. As shown in Figure 1, Japanese infants were unable to discriminate the contrast until they became 9.5-months of age. A heavy syllable /maa/ contains two morae, while a light syllable /ma/ one mora. If Japanese infants are able to access "mora," which is the unit that defines the mora-timed rhythm, they should be able to discriminate the contrast. The fact that the Japanese infants cannot access “mora” until 9.5-months of age means that they are unable to use “mora” to segment speech until then.

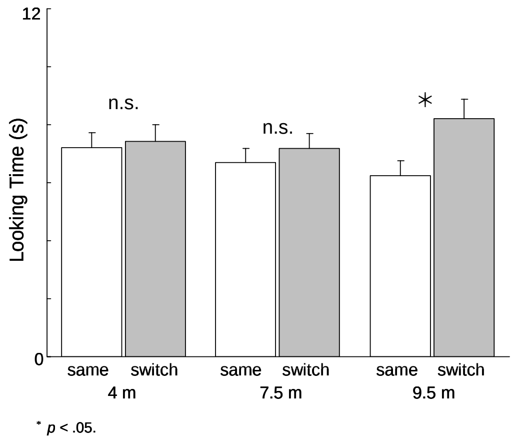

Figure 1. Japanese infants were able to discriminate long-short vowel distinction only at 9.5 months of age. (Figure 2, Sato, Sogabe, & Mazuka, 2010)

English learning infants were found to show a robust ability to segment bisyllabic or trysyllabic words from continuous speech by 9 months of age[14, 15]. If the rhythm of a language is to play facilitating role for speech segmentation, it should be available to infants younger than when they begin to show an ability to segment the speech. Thus, Japanese infants’ inability to access “mora” until they are 9.5 months of age makes it unlikely that the rhythm of their language could bootstrap them into speech segmentation.

If it does not work for one of the three types of of rhythm, the usefulness of the rhytym-based prosodic bootstrapping hypothesis would be seriously challenged.

EDGE-PROMINENT PROSODY

Recent works on prosodic typology suggest that languages can be dichotomized in the light of prominence realization. One is the “head-prominence language” such as English, in which the prominence of a word is signaled by positing (intonational) pitch accent on the stressed syllable of the word. Thus, the prominent word can be located not only at the beginning or end, but also in the middle of the prosodic phrase. The other is “edge-prominence language” such as Japanese and Korean, in which prominence of a word is achieved by placing the word at the edge of a phrase, and thus there is no choice in the location of prominent word within a phrase: it is always at the phrasal edge[16]. Analysis of Japanese IDS revealed that the “edge-priminent” characteristics of Japanese intonation can account for the so-called lack of pitch exaggeration in Japanese [17].

Adults modify their speech when they talk to infants. This style of speech is called infant-directed speech (IDS) as opposed to adult-directed speech (ADS). The modifications of prosody include higher pitch level and more expanded pitch range in IDS. The latter modifications are called “exaggeration” and often claimed to be universal characteristics of IDS.

At the same time, some of the past works suggest that Japanese may differ from English with regard to “exaggerated intonation” in IDS. For example, Japanese IDS does not show pitch range expansion, while English, French, and Italian do [18]. It was also reported that although English-learning infants responded appropriately to positive and negative IDS in English, German and Italian, they failed to do so in Japanese [19].

Using a large corpus of Japanese infant-directed speech [20], we first measured the pitch expansion of Japanese IDS using the conventional method; viz., by measuring the maximum and minimum of F0 for each utterance. As shown in Figure 2, the results replicated the previous study exactly – pitch range for both IDS and ADS were approximately 8 semitones, and no pitch expansion was found for IDS. The average pitch for IDS was higher than that of ADS, which is also the same as the previous study [18].

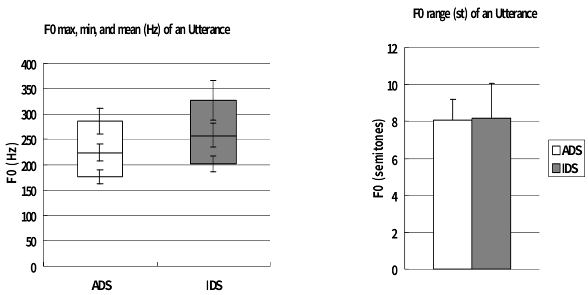

Figure 2. When measured in the conventional manner, pitch range of Japanese IDS is not larger than that of ADS, although the average pitch for IDS is higher than ADS. (From Igarashi & Mazuka, 2008)

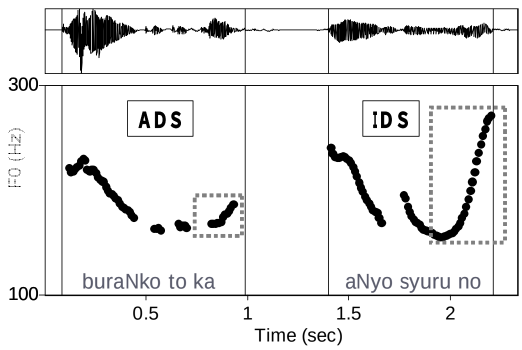

Figure 3. Pitch expansion in Japanese IDS was found in phrase-final syllables with BPM. This is a property of a language with edge-prominent prosody. (From Igarashi & Mazuka, 2008)

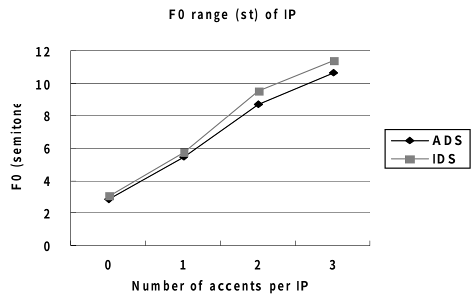

Figure 4. When syllables with BPM are removed, pitch ranges of IDS and ADS utterances are largely determined by the number of accented words in the utterance. (From Igarashi & Mazuka, 2008)

However, when we examined the pitch expansion in Japanese IDS, taking into account the intonational system of Japanese, it was found that the pitch expansion in Japanese IDS was concentrated in the phrase final syllables with Boundary Pitch Movement (BPM), as shown in Figure 3. This is a property of a language with Edge-prominent prosody. Once the syllables with BPM were removed, the pitch ranges of IDS and ADS utterences are determined largely by the number of Accentual Phrase (AP) that are included in an utterance, as shown in Figure 4. As discussed below, Japanese lexical-pitch accent is characterized by the sharp fall of pitch after an accented syllable, and the pitch of an utterance with multiple accented words shows a characteristic fall after each accented syllable (down-step.) Consequently, as the number of accented word increases in an utterance, the pitch range becomes larger. Importantly, the rate of pitch fall per accented words is identical for IDS and ADS. Since ADS utterences are significantly longer, and contain more accented words per utterance than IDS, the pitch range of an average ADS utterance (exclusing BPM) are larger than that of IDS.

Overall, the larger pitch range for BPMs in IDS and the larger pitch range of longer ADS utterences cancel out each other, and the average pitch range for IDS and ADS cames out the same. Yet, the pitch range expansion that is a natural consequence of longer utterences, as in Figure 4, is not an "exaggerated" intonation. In contrast, the larger pitch movement in a syllable with a BPM in IDS, as in Figure 3, is an "exaggered" intonation. Thus, for Japense listeners, Japanese IDS does have an exaggerated intonation. These results demonstrate that although the "exaggeration" of intonation in IDS may occur in any language, how it is realized in an actual utterance is constrained by the intonational structure of the language.

LEXICAL PITCH-ACCENT

English learning infants show a strong bias towards the strong-weak, trochaic pattern of syllables in wide variety of situations such as speech segmentation, word learning, and speech production [21]. Since the predominant trochaic pattern is the product of English lexical level prosody, it is considered that the lexical level prosody plays important roles in early stages of language acquisition.

Lexical level prosody is often categorized into three types. Lexical stress as in languages like English and Dutch, tones in languages like Chinese and Thai, and lexical pitch-accent in a language like Japanese [22]. Substantial body of research in linguistics have revealed many interesting commonalities and differences among the three types of lexical level prosody. Yet, other than the lexical stress in infants learning languages such as English, German, and Dutch, very little is known as to how infants learning other types of lexical level prosody master their languages' lexical prosody.

Unlike lexical stress, which involves multiple cues including amplitude, duration as well as vowel quality, the primary cue for tone and lexical pitch-accent is the movement of pitch contour. Such acoustic differences could impact on how native speakers of tone and lexical pitch accent languages process them, and how they are acquired by infants learning those languages.

In addition, lexical stress in stress-timed languages, such as English and German, is also involved in the rhtym of the language. Thus, the contribution of lexical level prosody and the rhtym are confounded in stress-timed languages. Investigation into how tones or lexical pitch-accent are acquired by infants and how they are represented/processed by adults could shed important light onto our understanding of how lexical level prosody contributes to language acquisition.

In Tokyo Japanese, words are either accented or unaccented. An accented word exhibits a sharp pitch fall from the accented mora (a subsyllabic unit of phonological weight) to the following mora. For example, TO'YOTA has a lexical accent on TO (as indicated by a following apostrophe), and pitch falls from TO to YO and keeps declining gradually toward the end of the word. Unaccented words do not possess such a local pitch fall (e.g., HONDA is produced with no drastic pitch fall). When spoken in isolation or after a prosodic juncture, both accented and unaccented words form a prosodic unit called accentual phrase (AP). The beginning of an AP is tonally described as LH-, a rise from the initial to the second mora, unless the word has an accent on the initial mora. Therefore, the word-initial tonal specification follows either one of two patterns in Japanese – HL for words with an initial accent, or LH for unaccented words and accented words with a non-initial accent.

Electrophysiological responses to lexical pitch-accent

In adult psycholinguistic literature, how lexical level prosody contributes to lexical access has been the topic of many studies. In particular, whether lexical level prosody is accessed during the initial stages of lexical access or it is used in later stages of processing has not been clear. We investingated this question in an ERP study with Japanese adults[23, 24].

Native speakers of Tokyo Japanese were presented with a photograph of an object, such as an egg, and asked whether or not the word they heard matched the picture. Speech stimuli were (1) correct name of the object, i.e., /tama'go/ (2) a wrong word for the picture (semantic mismatch) e.g., /hiyoko/ (chick), (3) word with segmental mispronunciation, e.g., /tana'go/ (consonant mismatch) or /tamo'go/ (vowel mismatch) or (5) words with accent mispronunciation, e.g., /ta'mago/ (early accent shift) or /tamago'/ (late accent shift).

When there was a semantic mismatch, a classic N400 was elicited as shown in Figure 5. In contrast, when there was a segmental mismatch either in consonant (Figure 6) or in vowel (Figure 7), a biphasic response pattern with an early negativity and a late positivity was observed.

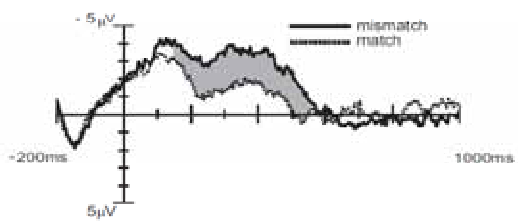

Figure 5. Grand average ERP waveforms across the medial posterior region for semantic mismatch condition. Shaded area shows a significant difference from the match condition. (From Horie et al., 2008)

When the accent was shifted to an earlier position, the biphasic pattern with an early negativity and a late positivity was found as shown in Figure 8. This is similar to the pattern found with consonant (Figure 6) or vowel mismatch conditions (Figure 7). When the accent was shifted to a later position, in contrast, a sustained positivity was found (Figure 9). Neither the segmental mismatch nor the accent mismatch condition elicited N400.

The results of the ERP study revealed that the lexical pitch-accent information is processed immediately and incrementally during lexical access. As mentioned above, there was a debate as to when the lexical level prosody is accessed ruing lexical processing. The results of the present study shwed that they are involved both in the initial and late stages of lexical processing.

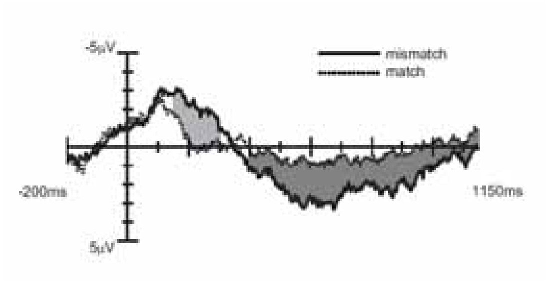

Figure 6. Grand average ERP waveforms for consonant mismatch condition. Light gray colored area shows a significant negative component, and dark gray colored area shows a significant positive component (From Horie et al, 2008).

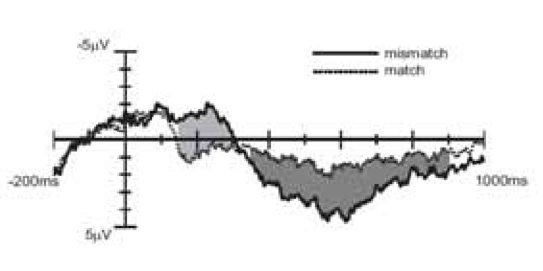

Figure 7. Grand average ERP waveforms for vowel mismatch condition (From Horie et al, 2008).

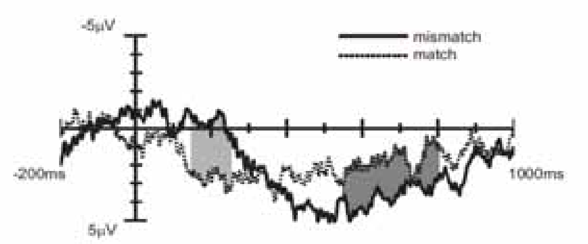

Figure 8. Grand average ERP waveforms across the medial-posterior region for early accent mismatch condition. Light gray colored area shows a significant negative component, and dark gray colored area shows a significant positive component (From Horie et al, 2008).

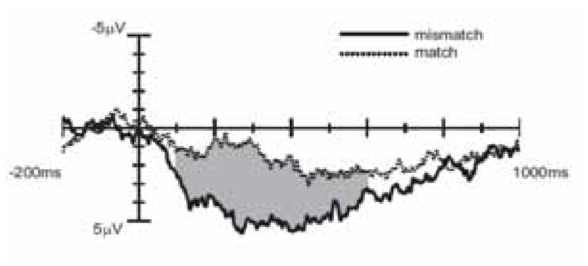

Figure 9. Grand average ERP waveforms for late accent mismatch condition. Gray colored area shows a significant positive component (From Horie et al, 2008).

Hemodynamic responses for lexical pitch-accent processing

Brain imaging techniques, such as fMRI, PET, or NIRS (Near Infrared Spectoscopy) that measures the hemodynamic responses could also provide useful information. In adults, it is well known that the left and right cerebral hemispheres work differently for speech processing: the left hemisphere is more heavily involved in processing segmental contrasts in one's native language, and the right hemisphere typically processes prosodic cues including affective prosody. Bilateral activation is seen in the processing of non-speech or non-native contrasts. The differential involvement of two hemispheres for lexical level prosody seems to be functionally determined. When the pitch cues for the lexical prosody are processed as linguistically relevant, left-lateralized activations are found, and bilateral or no left dominance activation is seen when the same cue is processed non-linguistically [25].

In Sato, Sogabe and Mazuka [25], Japanese adults presented were presented with two syllable hompphones that has either high-to-low (HL) pitch pattern or low-to-high (LH) pitch pattern such as ha'shi (HL: "chop stick") versus hashi' (LH: "bridge"). The contrast was embedded either in variable words (e.g., a'me (rain)/ame(candy), ki'ri (paulownia)/kiri (fog), ka'me (turtle)/kame (jar) etc.), single word pair (a'me (rain)/ame(candy), or in pure tones. As a control, the participants were also presented with phoneme contrast as in a'me (rain) vs. ka'me (turtle).

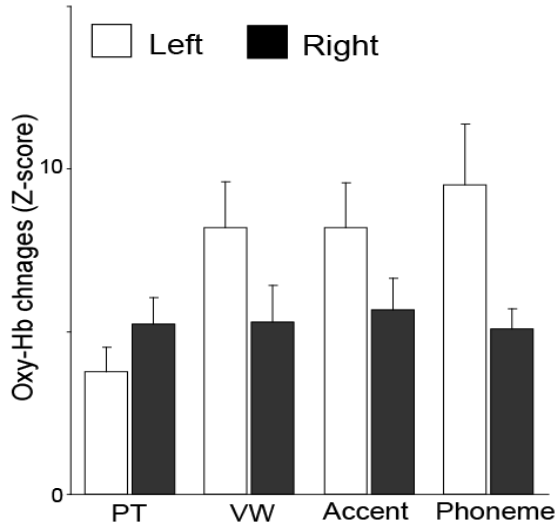

Figure 10. Hb responses in the left and right temporoparietal regions. PT=pure tone, VW=variable words, Accent=single word pair (From Stao, Sogabe, & Mazuka, 2007).

The participants' hemodynamic activation was measured using NIRS. As shown in Figure 10, left-hemisphere showed significantly higher activation for lexical pitch accent contrast (either in single word or multiple word condition), which was similar to the phoneme distinction. In contrast, the same pitch change that was presented in pure tone did not elicit higher activation in the left-hemisphere.

The results showed that the processing of lexical pitch-accent was functionally lateralized to the heft hemisphere in the same way as phonemic contrasts. When the same pitch changes were presented in non-linguistic context in pure tones, no left lateralization was found. Thus, the left lateralization was due not to the acoustic nature of the stimuli, but to the linguistic function it plays. This is consistent with the findings of the ERP study discussed above.

Development of lexical pitch-accent processing

Adults' processing of lexical level prosody is left-laterarized. But the cross-linguistic research revealed that such laterilization is specific to the lexical prosody of their own languages.

When Chinese and Thai speakers were presented with tones in Chinese and Thai, Chinese speakers showed a left-lateralized activation for Chinese tones, but not for Thai tones, while Thai speakers showed a left-lateralized activation for Thai tones but not for Chinese tones [26] . Consistent with the NIRS study with Japanese adults above[25], these results suggest that the emergence of left-lateralized activation for lexical-level prosody must be linked with acquisition of the native language.

With English and French learning infants, it has been reported that they were able to discriminate tones in Thai at 6 months of age, while they became unable to discriminate them by 9 months of age[27, 28]. These results suggest that infants' perception of lexical level prosody may go through "reorganization" similar to the perception of segmental categories. It has been proposed that the "reorganization" involves a shift of infants' speech sound processing from general auditory processing to language specific one. If true, we may observe the emergence of left-lateralized brain activation for lexical level prosody before and after the "reorganization."

In Sato, Sogabe and Mazuka [29], we tested the emergence of hemispheric latelirization for Japanese lexical pitch-accent processing. Using visual habituation-dishabituation paradigm, we first confirmed that 4 and 10 month old Japanese infants are capable of discriminating pairs of lexical items that minimally differ in LH versus HL pitch-accent, as shown in Figure 11.

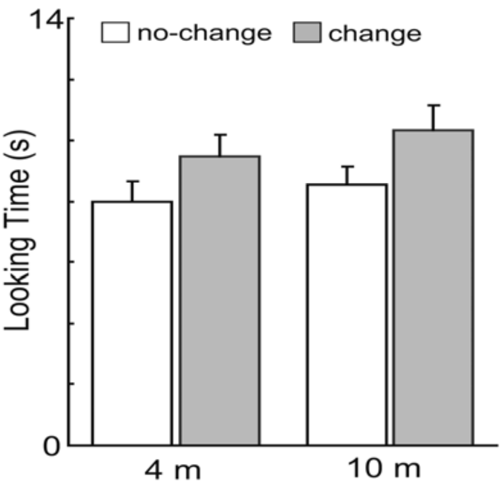

Figure 11. Both 4- and 10-month old Japanese infants showed significantly longer looking time for the change condition than no-change condition in visual habituation-dishabituation paradigm, showing that both group of infants were able to discriminate the stimuli behaviourally (From Sato, Sogabe & Mazuka, in press).

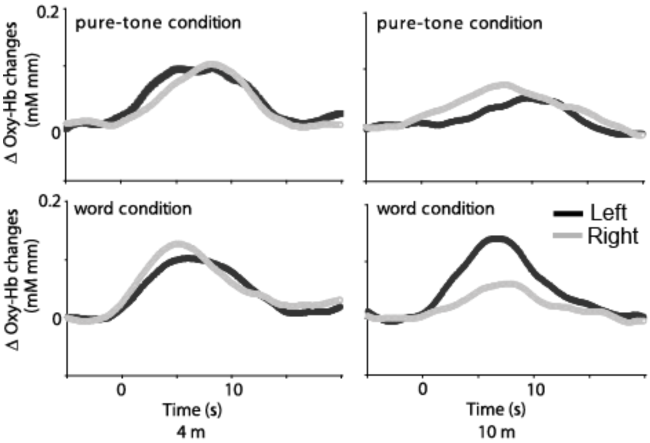

Figure 12. Changes to Oxy-hemoglobine level to HL vs LH pattern of pitch-accent in pure tone or word condition. Significantly higher left-activation was found for word condition only in 10 month olds (From Sato, Sogabe & Mazuka, in press).

In contrast to the behavioural results, infants' hemodynamic responses to the HL vs LH lexical pitch accent showed a different pattern between 4- and 10-months of age. When the pitch changes were presented in pure tone, both 4- and 10-month old infants showed bilateral activation (two top panels of Figure 12). In contrast, when the same pitch change was embedded in words, 10-month old infants showed higher activation in the left-hemisphere than the right (lower right-hand panel of Figure 12), while no such lateralization was found in 4-month old infants (lower left panel). The results show that the left and right hemispheres of 4-month old infants responded similarly to the HL vs LH pitch changes in pure tone and words. By 10 months of age, however, they begin to treat the HL/HL pattern in words differently from pure tone. Their activation is now left-lateralized, similar to that of adult Japanese native speakers. This pattern of results confirms the reorganization hypothesis.

Dialect differences

Lexical level prosody is so called because it contributes to the determination of lexical meaning. In this sense, lexical level prosody is similar to segments. The ERP study above confirmed that the lexical pitch-accent and segments are indeed processed in a similar manner at the initial stage of lexical processing in Japanese. At the same time, the acoustic cues for lexical level prosody are indistinguisheable from those for sentence level prosody, such as pitch, amplitude and duration. Thus, both the lexical and phrasal level prosody are represented simultaneously in an actual speech. The listener must be able to distinguish the two types of prosody in order for her/him to access the lexical and phrasal meaning accurately. Similarly, for infants who are trying to learn a language, it is a major task to distinguish the two types of prosody in the speech they hear.

One way to investigate the interaction between lexical and phrasal level prosody is to compare two languages with or without a lexical level prosody. Previous studies have compared Spanish and French, the former is a language with lexical stress while the latter without [30, 31]. When two separate languages are compared, however, the differences are not limited to the lexical level prosody as they typically differ in their lexicon and many other grammatical characteristics. If we can find a pair of dialects within a language one with and another without a lexical level prosody, the most straightforward comparison is possible.

Japanese is often classified as having lexical pitch-accent. But there are substantial variations among different dialects of Japanese regarding the lexical pitch-accent, and there are several "accentless" dialects are identified. 'Accentless' dialects are defined by the lack of pitch specification at the word level in Japanese dialectology.

In Utsugi, Koizumi and Mazuka [32], we tested the processing of lexical pitch-accent among the speakers of one of the accentless dialects spoken in Sendai area (northern Japan) and compared them to Tokyo dialect speakers. Native speakers of each of the dialects were first tested in an identification task and a discrimination tasks, both of which are typically used to test Categorical Perception.

Using a pair of nonsense word "manu," two sets of continuum were prepared. In the first set, the pitch of the second syllable was varied in ten steps from high-flat to falling, as shown in Figure 10. In the second set, the pitch of the first syllable varied in ten teps from high flat pitch to rising, as shown in Figure 11. The changes in the pitch in set 1 and 2 are in the mirror image of each other. In Tokyo Japanese, the pitch change in set 1 is phonemic. Rise-flat pattern of pitch is perceived as unaccented while rise-fall pattern of pitch is perceived as accented. In contrast, pitch changes in set 2 are not phonemic. Whether or not the first syllable has a rising or flat pitch, the word will be perceived as accented since there is a fall in the pitch in the second syllable. In accentless dialects, neither contrasts are phonemic.

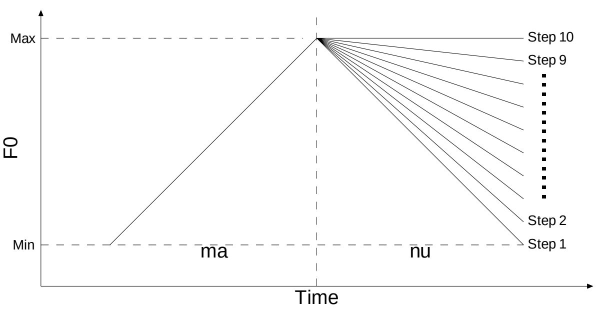

Figure 13. The stimuli in set 1. The difference between the two end points of the continuum is phonemic in Tokyo dialect of Japanese, while they are not phonemic in accentless dialects.

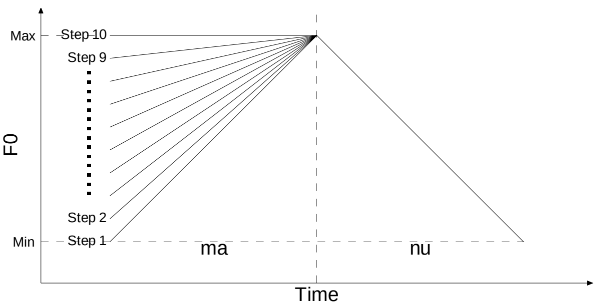

Figure 14. The stimuli in set 2. The difference between the two end points of the continuum is not phonemic in either Tokyo or accentless dialects of Japanese.

The results of discrimination and identification experiments revealed that neither group showed clear sign of categorical perception in either set of the stimuli, and we found no differences between the Tokyo dialect speakers and the accentless dialect speakers.

In the second experiment, we compared the two groups in a paradigm developed by Dupoux et al [30] that compared French and Spanish speakers. In this task, participants were presented a series 2, 3 or 4 stimuli and asked to report what they heard. For example, a participant is first trained to press "1" if they heard /manu/ and "2" if they heard /menu/. When they heard a sequence /manu-menu-manu-manu/, they need to respond 1-2-1-1. Participants in each dialect group were tested in 3 contrasts. In segmental contrast, the stimuli were a pair of nonsense words /ma'nu/ versus /me'nu/ that differed segmentally. In lexical pitch contrast, stimulus 1 and 10 in set 1 in Figure 10 above were used. In non-lexical pitch contrast, stimulus 1 and 10 in set 2 in Figure 11 were used. In all contrasts, stimuli were pronounced by 6 different speakers, 3 males and 3 females.

The results revealed significant difference between the two groups. The Tokyo dialect speakers were able to perform this task reliably both for segmental contrast and lexical pitch contrasts. They were at chance level for non-lexical pitch contrast. The speakers of accentless dialect were as good as Tokyo speakers in segmental contrast. Like Tokyo speakers, their performance in non lexical pitch contrast was at the chance level. But their performance in lexical pitch contrast was significantly less accurate than that of the Tokyo speakers in lexical pitch contrasts.

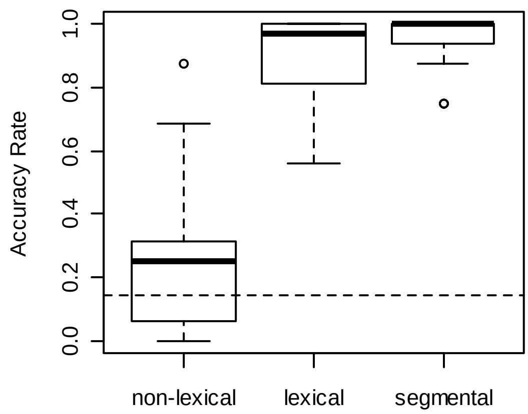

Figure 15. Accuracy rates for the Tokyo speakers (From Utsugi, Koizumi & Mazuka, 2010).

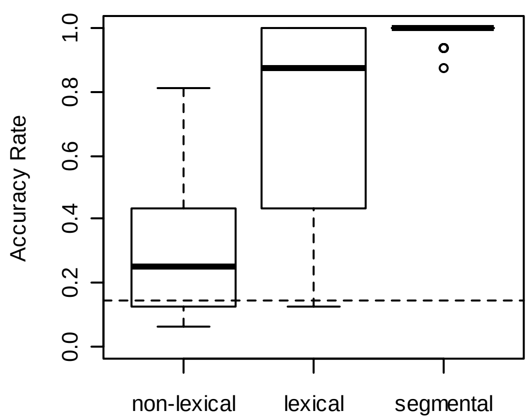

Figure 16. Accuracy rates for the accentless dialect speakers. (From Utsugi, Koizumi & Mazuka, 2010)

The results also revealed that the variance among accentless dialect speakers was significantly larger than that of Tokyo dialect speakers. This demonstrates the influence of the "standard" dialects among the speakers of non-standard dialects.

SEGMENTAL DISTRIBUTION

Lastly, Japanese is also a language whose segmental distribution may provide interesting research opportunities. For example, coronal obstruents (t/d) are known to occur at a higher frequency than dorsal counterparts (k/g) in many languages. The asymmetry is also observed in infants' word learning and production errors [33]. Some have proposed that the asymmetry arises from the markedness of dorsal segments in relation less marked coronal segments, while others have argued that the effect derives from the input frequency. In most languages, in which the dorsal segments occur less frequently, it is not possible to dissociate the two explanations.

The segmental distribution in Japanese exhibits an exceptional case. As shown in Figure 17, dorsal obstruents occur significantly more frequently than coronal or labial segments in Japanese[34]. This is observed not only in adult directed speech, but also in infant directed speech.

If as the markedness theory predicts, dorsal segments are more marked than coronal segments, the relative disadvantage of dorsal segments reported in English and Dutch infants should also be observed in Japanese infants. If, on the other hand, the input frequency is the source of the asymmetry observed in the previous studies, Japanese infants should not find disadvantage for the dorsal segments. Currently, which of the predictions would actualy bourne out is open to future investigation.

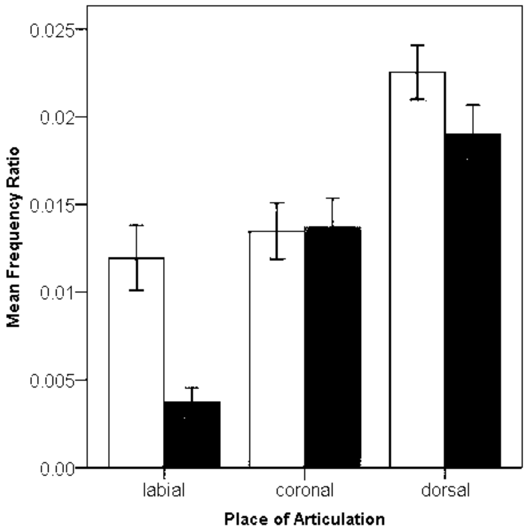

Figure 17. Mean type frequency ratio of labial (p/b), coronal (t/d), and dorsal (k/g) obstruents in Japanese. Black bars indicate adult directed speech and white bards indicate infant directed speech (From Tsuji, Nishikawa & Mazuka, 2010).

CONCLUSION

Cumulative knowledge of how infants acquire the phonological system of their language hasbeen gain almost exclusively from the research of languages such as English and other European languages. Since they represent only a part of phonological variations among the world's languages, lack of data from other languages is a serious gap in the field. Japanese is a language that differ in its phonological characteristics from English and other European languages such that the investigation into its acquisition could contribute significantly in this endeavour. In this paper, empirical data from four lines of research in Japanese are introduced to elucidate this point.

REFERENCES

1 Werker, J.F., & Yeung, H. H. "Infant speech perception bootstraps word learning". *Trends in Cognitive Sciences* **9**, 519-527 (2005).

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

3 Jusczyk, P.W., Cutler, A., & Redanz, N. J. "Infants' preference for the predominant stress patterns of English words". *Child Development* **64**, 675-687 (1993).

4 Jusczyk, P.W., Luce, P. A., & Charles-Luce, J. "Infants' sensitivity to phonotactic patterns in the native language". *Journal of Memory and Language* **33**, 630-645 (1994).

5 Mattys, S.L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. "Phonotactic and prosodic effects on word segmentation in infants". *Cognitive Psychology* **38**, 465-494 (1999).

6 Abercrombie, D. "Elements of General Phonetics". (1967).

7 Port, R.F., Dalby, J., & O'Dell, M. "Evidence for mora timing in Japanese". *Journal of the Acousitcal Society of America* **81**, 1574-1585 (1987).

8 Otake, T., Hatano, G., Cutler, A., & Mehler, J. "Mora or syllable? Speech segmentation in Japanese". *Journal of MEmory and Language* **32**, 258-278 (1993).

9 McQueen, J.M., Otake, T., & Cutler, A. "Rhythmic cues and possible-word constraints in Japanese speech segmentation". *Journal of Memory and Language* **45**, 103-132 (2001).

10 Nazzi, T., Bertoncini, J., & Mehler, J. "Language discrimination by newborns: Toward an understanding of the role of rhythm". *Journal of Experimental Psychology: Human Perception and Performance* **24**, 756-766 (1998).

11 Nazzi, T., & Ramus, F. "Perception and acquisition of linguistic rhythm by infants". *Speech Communication* **41**, 233-243 (2003).

12 Mazuka, R. "The rhythm-based prosodic bootstrapping hypothesis of early language acquisition: does it work for learning for all languages?". *言語研究, Journal of the Linguistic Society of Japan*, 1-13 (2007).

13 Sato, Y., Sogabe, Y. & Mazuka, R. "Discrimination of phonemic vowel length by Japanese infants". *Developmental Psychology* **46**, 106--119 (2010).

14 Houston, D.M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. "Cross-language word segmentation by 9-month-olds". *Psychonomic Bulletin and Review* **7**, 504-509 (2000).

15 Jusczyk, P.W., Houston, D. M., & Newsome, M. "The beginnings of word segmentation in English-learning infants". *Cognitive Psychology* **39**, 159-207 (1999).

16 Jun, S.-A. *Prosodic Typology: The Phonology of Intonation and Phrasing* (Oxford University Press, New York, 2005).

17 Igarashi, Y., & Mazuka, R. "Exaggerated prosody ini infant-directed speech? Intonational phonological analysis of Japanese infant-directed speech". *Proceedings of BUCLD 32*, 177-188 (2008).

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

19 Fernald, A. "Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages". *Child Development* **64**, 657-674 (1993).

20 Mazuka, R. Igarashi, Y. & Nishikawa, K. "Input for learning Japanese: RIKEN Japanese Mother-infant

Conversation Corpus". *IEICE Technical Report*, 11--15 (2006).

21 Gerken, L. "Early sensitivity to linguistic form". *Annual Review of Language Acquisition* **2**, 1-36 (2002).

22 Hyman, L.M. "How (not) to do phonological typology: the case of pitch-accent". *Language Sciences* **31**, 213-238 (2009).

23 Horie, R. et al. "Event-related potentials evoked by Japanese pitch-accent mismatch". *Neurosci Res* **61**, S199-S199 (2008).

24 Ito, K., Horie, R., Oishi, H., Hasegawa, M. & Mazuka, R. in Secondary "An incremental processing of lexical accent: An ERP study in Japanese" (Chicago, USA 2009/10).

25 Sato, Y., Sogabe, Y. & Mazuka, R. "Brain responses in the processing of lexical pitch-accent by Japanese speakers". *NeuroReport* **18**, 2001-2004 (2007).

26 Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., et al. " Activation of the left planum temporale in pitch processing is shaped by language experience". *Human Brain Mapping* **27**, 173-183 (2006).

27 Mattock, K., & Burnham, D. "Chinese and English infants' tone perception: Evidence for perceptual reorganization". *Infancy* **10**, 241-265 (2006).

28 Mattock, K., Molnar, M., Polka, L., & Burnham, D. "The developmental course of lexical tone perception in the first year of life". *Cognition* **106**, 1367?1381 (2008).

29 Sato, Y., Sogabe, Y. & Mazuka, R. "Development of hemispheric specialization for lexical pitch-accent in Japanese infants.". *Journal of Cognitive Neuroscience* (in press).

30 Dupoux, D., Peperkamp, S., & Sebastian-Galles, N. "A robust method to study stress ‘‘deafness’’". *Journal of the Acousitcal Society of America* **110**, 1606-1618 (2001).

31 Dupoux, E., Sebastian-Galles, N., Navarrete, E., & Peperkamp, S. "Persistent stress "deafness": The case of French learners of Spanish". *Cognition* **106**, 682-706 (2008).

32 Akira Utsugi, M.K., and Reiko Mazuka. "A robust method to detect dialectal differences in the perception of lexical pitch accent". *Proceedings of 20th International Congress on Acoustics* (2010).

33 Beckman, M.E., Yoneyama, K., & Edwards, J. . "Language-Specific and Language-Universal Aspects of Lingual Obstruent Productions in Japanese-Aquiring Children". *Journal of the Phonetic Society of Japan* **7**, 18-28 (2003).

34 Tsuji, S., Nishikawa, K. & Mazuka, R. in Secondary "Input frequency of place of articulation in Japanese infant-directed speech: An exceptional case" (ed. studies, I.s.o.i.) (Baltimore, USA 2010/3).