frontiers in Psychology

ORIGINAL RESEARCH
published: 09 April 2021
doi: 10.3389/fpsyg.2021.626662

Segmentation of Rhythmic Units in Word Speech by Japanese Infants and Toddlers

Yeonju Cheong[1]* *and Izumi Uehara*[1,2]*

[1] Department of Psychology, Ochanomizu University, Tokyo, Japan, [2] Institute for Education and Human Development, Ochanomizu University, Tokyo, Japan

When infants and toddlers are confronted with sequences of sounds, they are required to segment the sounds into meaningful units to achieve sufficient understanding. Rhythm has been regarded as a crucial cue for segmentation of speech sounds. Although previous intermodal methods indicated that infants and toddlers could detect differences in speech sounds based on stress-timed and syllable-timed units, these methods could not clearly indicate how infants and toddlers perform sound segmentation. Thus, the present study examined whether Japanese infants and toddlers could segment word speech sounds comprising basic morae (i.e., rhythm units similar to syllables), on the basis of concurrent basic mora units within syllable units, using the new intermodal matching procedure. The results indicated that, regardless of their ages and linguistic abilities, Japanese infants and toddlers aged 6–25 months tended to segment Japanese words comprising basic morae sounds on the basis of concurrent basic mora units within syllable units. This implies that infants' and toddlers' use of syllable units for segmentation of speech sounds at an early age could be evident among many infants and toddlers learning various languages. Although this finding should be interpreted carefully, the present study demonstrated the utility of the new intermodal matching procedure for examining segmentation of speech sounds and word sounds by infants and toddlers, on the basis of specific rhythm units.

Keywords: word segmentation, intermodal matching procedure, syllable, mora, Japanese infants

OPEN ACCESS

Edited by:
Anja Gampe, University of Duisburg-Essen, Germany

Reviewed by:
Antje Endesfelder Quick, Leipzig University, Germany
Louise Goyet, Université Paris 8, France

***Correspondence:**
Yeonju Cheong
cheongyeonju@gmail.com
Izumi Uehara
uehara.izumi@ocha.ac.jp

Specialty section:
This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology

Received: *06 November 2020*
Accepted: *08 March 2021*
Published: *09 April 2021*

Citation:
Cheong Y and Uehara I (2021) Segmentation of Rhythmic Units in Word Speech by Japanese Infants and Toddlers. Front. Psychol. 12:626662. doi: 10.3389/fpsyg.2021.626662

INTRODUCTION

When infants and toddlers are exposed to vocal speech, they are confronted with sequences of sounds and are required to segment the sounds into meaningful units to achieve sufficient understanding. Segmentation of speech might constitute a fundamental cognitive skill for an infant or a toddler to develop linguistic abilities. Previous research has focused on how infants and toddlers recognize the boundary of a sound within speech and when they begin to segment meaningful word sounds from the flow of speech. Segmenting speech sounds is presumably enabled by the perception of several linguistic cues, such as rhythmic cues (e.g., Echols et al., 1997; Johnson and Jusczyk, 2001), allophonic cues (e.g., Jusczyk et al., 1999a), phonotactic cues (e.g., Mattys and Jusczyk, 2001), and transitional probabilities between syllables (e.g., Saffran et al., 1996).

Infants begin to use these cues to segment speech sounds until ∼1 year of age. Notably, rhythmic cues have been identified as a crucial component of speech sounds (Echols et al., 1997; Johnson and Jusczyk, 2001; Goyet et al., 2013). There are reportedly two main types of rhythmic cues

(Dauer, 1983; Ramus et al., 1999; Arvaniti, 2009): stressed-timed and syllable-timed rhythms. Although mora-timed rhythm can be distinguished from these two rhythm categories (Abercrombie, 1967; Hoequist, 1983), this rhythm is also considered a type of syllable-timed rhythm (Otake et al., 1993; Arvaniti, 2009; Mazuka, 2009) according to indexes such as pairwise comparisons of successive vocalic and intervocalic intervals introduced by Grabe and Low (2002).

A mora is a segmental minimum unit of rhythm (Dan et al., 2013; Ogino et al., 2017) and is represented in the rhythmic structure of the Japanese language. A mora usually consists of a CV (consonant, vowel) or V syllable, which are called "basic morae." More than 70% of Japanese sound units are basic morae, and can be segmented in a manner identical to that of syllables. However, a small number of morae, termed "special morae," include special patterns of syllables "such as a nasal coda (CVN or VN), a geminate stop consonant (CVQ or VQ), a long vowel (CV: or V:), or a contracted sound (CjV)" (p. 113, right column, lines 6–8 from Ogino et al., 2017). Special morae are segmented in a manner that differs from that of syllables. For example, kitte (meaning stamp) includes three morae (ki, t, and te: "t" is a special mora unit, while "ki" and "te" are basic mora units) but two syllables: "t" is not counted as a syllable unit, while "kit" and "te" are considered constituent syllables.

Here, we briefly review past studies of speech sound segmentation in infants and toddlers for each of the three rhythm categories. Infants and toddlers whose native language belongs to the stress-timed category (e.g., English, German, and Dutch) use trochaic (strong-weak word stress pattern) or iambic rhythms (weak-strong word stress pattern) to segment speech sounds. Infants learning stress-timed languages begin to use or prefer trochaic units earlier than iambic units (Jusczyk et al., 1993). For example, English-learning 7.5-month-old infants segmented more often with a trochaic rhythm, rather than an iambic rhythm (Jusczyk et al., 1999b). Six-month-old German-learning infants also demonstrated this tendency (Höhle et al., 2009). Both Dutch- and English-learning 9-month-old infants could segment trochaic Dutch words in passages (Houston et al., 2000). The earliest age at which infants segment using an iambic rhythm as often as a trochaic rhythm is 10.5 months of age (Jusczyk et al., 1999b).

Infants and toddlers whose native language belongs to the syllable-timed category (e.g., French, Spanish, and Catalan) use syllable units to segment speech sounds. Nazzi et al. (2006) indicated that French-learning 12- and 16-month-old infants and toddlers (but not 8-month-old infants) could detect disyllabic words, which had been presented in isolation 15 times or embedded in several passages during the familiarization phase, although methods of detection or segmentation tended to differ between 12- and 16-month-old infants and toddlers. When the daily language environment involved Catalan only, Spanish only, or both languages, 6- and 8-month-old infants could detect monosyllabic words in passages (Bosch et al., 2013). Furthermore, Nishibayashi et al. (2015) indicated that, in a test phase, 6-month-old infants could identify monosyllabic words and syllables embedded in disyllabic words that had been presented during a preceding familiarization phase.

There is minimal evidence regarding how infants and toddlers perceive and discriminate mora units in a continuous speech stream. French-learning infants could discriminate English and Japanese sentences within 5 days of birth, despite low-pass filtering of the stimuli (Nazzi et al., 1998). Yoshida et al. (2010) examined the non-linguistic trochaic and iambic tones for Japanese- and English-learning infants. Although neither had any preferences for either tone type at 5–6 months of age, English-learning 7–8-month-old infants could differentiate the two tones on the basis of preferences for trochaic tones, whereas Japanese-learning 8-month-old infants did not show any preference for trochees or iambs. Nevertheless, there has been no direct examination involving whether or how Japanese-learning infants and toddlers segment speech sounds using rhythm cues in mora-timed language. As noted above, a mora-timed rhythm is also a presumptive type of syllable-timed rhythm (Otake et al., 1993; Arvaniti, 2009; Mazuka, 2009). Indeed, 2-month-old English-learning infants could differentiate passages in English from passages in Japanese, although they could not differentiate passages in French from passages in Japanese (Christophe and Morton, 1998).

While infants' and toddlers' responses toward specific rhythm types have been examined, several studies have proposed that infants and toddlers have broad speech perception in terms of syllables (e.g., Bertoncini et al., 1988; Jusczyk et al., 1995; Räsänen et al., 2018). Many languages are composed of a CVsound, which is a basic syllable unit. Nearly half of the sound forms in English are the CV, V, or VC form. Approximately 80% of the sound forms in French are CV, V, or VC, while more than 70% of the sound forms, so-called "basic morae" (as mentioned above) in Japanese are CV or V (Greenberg, 1999; Dankovičová and Dellwo, 2007). The syllable unit is considered a universal unit for examination of language structure (Mehler et al., 1981). Because French and Japanese, especially, have higher amounts of similarity in basic syllable structure and high degrees of phoneme regularity, French words could be relatively easily adapted to Japanese basic morae (Shinohara, 1996). Other previous studies have examined language learning by infants with respect to syllable units (e.g., Mehler, 1981; Bijeljac-Babic et al., 1993).

The prosodic bootstrapping hypothesis holds that prelinguistic infants acquire prosodic aspects of language including stress, rhythm, and intonation, which they use to identify speech sound boundaries and grammatical features (e.g., Gleitman and Wanner, 1982; Jusczyk, 1997; Soderstrom et al., 2003; Bernard and Gervain, 2012; Wellmann et al., 2012). Whereas some studies have indicated that infants younger than 12 months may perceive language-specific grammatical features such as word order (e.g., Japanese and Italian 8-year-olds in Gervain et al., 2008), some other studies have indicated that infants may develop skills regarding general phonological perception of speech sounds at an earlier age (e.g., Jusczyk, 1997; Soderstrom et al., 2003). For instance, a small number of Japanese studies have suggested that rhythm segmentation skills concerning Japanese "special morae" sounds may develop long after the infantile period, at around 4 years of age (Takahashi, 1997, 2001). Considering these findings, we assumed that more general basic segmentation skills that are widely shared,

regardless of language and linguistic skills, and can be framed in terms of syllables, would develop earlier than language-specific perceptual and segmenting skills (which might be more closely related to the acquisition of specific first words), and that they would remain after acquiring the first words. In the present study, we focused on the basic segmentation skills that are expected to develop earlier. However, general segmentation skills have not been directly examined or empirically confirmed using previous methodologies.

Past methods frequently used the headturn preference procedure (e.g., Houston et al., 2000; Nazzi et al., 2006; Höhle et al., 2009; Bosch et al., 2013; Nishibayashi et al., 2015). These methods could not directly demonstrate the method of speech sound segmentation by infants and toddlers. The typical headturn preference procedure is as follows. In the booth where the child sits on his or her parent's lap, one green lamp is on the center wall, one red lamp is on the left side wall, and another red lamp is on the right side wall. Each trial begins by lighting the green lamp, which is then switched off when the child looks at it. Subsequently, one of the red lamps begins to blink. When the child turns his/her head to look at the blinking red light, the speech sound presentation begins. The sound continues until the child looks (turns) away from the light for more than 2 s, or until the presentation duration ends. The total looking times toward the target sounds and the other sounds are compared. If a significant difference is observed in looking time toward the two stimuli, the child is presumed to detect a difference between the two sounds.

A few past studies of infants aged <6 months used a non-nutritive high-amplitude sucking procedure (e.g., Floccia et al., 1997; Christophe and Morton, 1998; Nazzi et al., 1998). The typical sucking procedure is as follows. Initially, the infant's usual sucking rate (i.e., baseline rate) is checked. To hear the sound continuously, the infant must suck with a high-amplitude rate. When the infant begins to suck with a high rate, the target sound is presented. The stimulus sound is continuously presented until the sucking rate becomes lower than a predetermined threshold (e.g., 75 or 80% of the max rate of sucking) during the familiarization session or the test trial for 2 consecutive minutes. Usually, the sucking rate decreases with repeated or continuous presentation of the same stimulus, whereas the sucking rate increases immediately after presentation of a different stimulus. If a significant difference is observed in sucking rate toward the two sound stimuli, the infant is presumed to detect a difference between the two sounds.

Either method can indicate whether children detect differences between two sounds and show that infants and toddlers separate target sounds from the flow of speech, although these methods cannot indicate how infants and toddlers segment speech sounds. To understand clearly the mechanism by which infants and toddlers segment speech sounds, a new method is necessary.

The purpose of the present study was to establish a new method that can show directly how infants and toddlers segment speech sounds. Considering that syllable units are the fundamental components of language, this study assessed the utility of the new method by examining whether infants and toddlers can segment speech sounds based on syllable units. This new method used the revised intermodal matching procedure developed by Kobayashi et al. (2005). The outline of the present study was as follows.

In each trial of the familiarization session, the child was shown each ball in the display, accompanied by a speech sound. The ball dropped from the top to the bottom of the display. There were four trials in the familiarization session: one ball with one-syllable word sounds, two balls with two-syllable word sounds, three balls with three-syllable word sounds, and four balls with four-syllable word sounds.

In each trial of the test session, an opaque square appeared first in the middle of the display and extended to the bottom of the display. One-word sounds (one, two, three, or four syllables) were presented and the square receded toward the bottom of the display, then disappeared. The child immediately saw either the match condition, where the number of balls (e.g., three) matched the number of syllables presented aurally immediately prior (e.g., three-syllable words, te-re-bi), or the non-match condition, where the number of balls (e.g., two) did not match the number of syllables presented aurally immediately prior.

If the infants or toddlers correctly perceived the number of syllables concerning the relevant word sounds, we expected that the amount of time they spent looking toward the ball stimuli would differ between the match and non-match conditions. A significant difference in looking time during two conditions (matched familiar and non-matched novel conditions) has been regarded as evidence that a child can distinguish between them, and thus, that they can correctly perceive the number of stimuli (Gerken et al., 2015; DePaolis et al., 2016). The condition during which the child spends more time looking at the stimuli is not clear: some studies have reported that familiar stimuli received more attention (i.e., longer looking time) from infants and toddlers during modality-matching in a visual-auditory setting (e.g., Starkey et al., 1983; Golinkoff et al., 1987), whereas other studies have reported that unexpected events received greater attention from infants (Mix et al., 1997; Kobayashi et al., 2005). However, in a recent meta-analysis, Bergmann and Cristia (2016) examined the direction or condition in which children looked longer toward during the intermodal matching procedure. The data, which were collected with the intention of examining segmentation of word sounds, indicated that young children often spent longer looking at the stimuli in the familiar matched condition. Indeed, previous studies concerning attention toward stressed word sounds or one- or two-syllable word sounds found that infants and toddlers almost always looked longer toward the familiar sounds (e.g., Jusczyk et al., 1999b; Houston et al., 2000; Nishibayashi et al., 2015). Thus, in the present study, we presumed that the children would look longer toward the stimuli in the match condition than that in the non-match condition when they correctly perceived the number of syllables (in this case, basic morae).

We used this method to examine whether basic early segmentation based on syllable units or basic morae could be observed in infants and toddlers aged 6–25 months, regardless of linguistic skill level. One reason why we selected 6 months of age as the approximate earliest age to examine is because this is the

point at which infants begin to detect differences in word sounds (Höhle et al., 2009; Bosch et al., 2013; Nishibayashi et al., 2015). The other reason is because this is the earliest age examined by studies of the bootstrapping hypothesis (e.g., Gleitman and Wanner, 1982; Jusczyk, 1997; Soderstrom et al., 2003; Bernard and Gervain, 2012; Wellmann et al., 2012). We selected 25 months as the approximate latest age at which children are expected to produce their first words for the following reason. Although the earliest children begin to speak is around 8 months, many children begin to speak after 1 year (Fenson et al., 1994). Word acquisition in typically developing Japanese children tends to be slightly delayed compared to English-learning children (Okumura et al., 2016), such that some typically developing Japanese children do not produce first words until around 24 months (Yoshioka and Tosa, 2014). Considering these data, it could be said that the age at which typically developing Japanese children acquire their first words ranges from 8 months to 2 years of age.

We assumed that, beginning at the age of ∼6 months, children would segment word sounds in the same manner as older children. We did not expect age, linguistic ability, or sex to influence the segmentation of word sounds considering previous findings concerning the perception of prosodic features of speech sounds in very young children (e.g., Nazzi et al., 1998; Christophe et al., 2003; Bernard and Gervain, 2012; Wellmann et al., 2012). Thus, all participants' looking patterns toward the stimuli would be similar if the new method were to be useful for examining segmentation of speech sounds in those children.

METHODS

Participants

Forty-five healthy monolingual Japanese infants or toddlers (21 boys and 24 girls; 6–25 months old; mean age = 13.6 months, standard deviation = 5.56 months) participated in this study, accompanied by their mothers. The data from five additional infants or toddlers were excluded from analysis because of fussiness during the task or < 30% looking time fixation during a test session (Tobii Pro Studio V3.2). All participants were recruited through flyers posted in various locations (e.g., nurseries and children's centers) and advertisements on the Internet homepage of our developmental laboratory. The participants and their parents lived in the Tokyo metropolitan area and had middle-class socioeconomic backgrounds. All participants, whose main caretakers were their mothers, were full-term born and monolingual. No child had been diagnosed with a developmental disability. All of the mothers were homemakers or on parental leave at the time of participation in this study.

Both verbal and written informed consent for the children to participate were provided by their mothers. In accordance with the Declaration of Helsinki, all procedures in the present study were approved by the Humanities and Social Sciences Research Ethics Committee of Ochanomizu University.

Material

To check a child's language ability, the child's mother was asked to complete the Japanese version of the MacArthur Communicative Development Inventory regarding the child [the original version is known as the CDI (Fenson et al., 1993), while the Japanese version is known as the JCDI]. The inventory "Words and Gestures," presumably suitable for assessing Japanese 8–18 month-old infants and toddlers (Ogura and Watamaki, 2004), was used for the present children aged 6–17 months (34 children including 15 boys and 19 girls; mean age = 10.8 months, standard deviation = 2.70 months). Although the inventory "Words and Gestures" was not developed for children as young as 6–7 months old, we used this inventory to confirm that the children in this age group had a lower linguistic level. Bergelson and Swingley (2012, 2015) used this inventory to check linguistic abilities in 6–9-month-old infants and 6–16-month-old infants in 2012 and 2015, respectively. The inventory "Words and Grammar," presumably suitable for assessing Japanese 16–36-month-old infants and toddlers (Watamaki and Ogura, 2004), was used for the present children aged 18–25 months (11 children including six boys and five girls; mean age = 22.2 months, standard deviation = 2.18 months).

Werker et al. (2002) and Song et al. (2018) counted the raw number of words produced by toddlers (according to their mothers) using the MacArthur CDI (Fenson et al., 1993) and regarded this number as the toddler's vocabulary size. Bates and Goodman (1997) discussed vocabulary size in relation to the number of words the child understood and produced, according to the mother's description on the MacArthur CDI (Fenson et al., 1993). We thus determined each infant's or toddler's vocabulary size according to these approaches. The vocabulary size of children aged 6–17 months was determined by counting words that the child understood and produced, according to the mother's description on the JCDI inventory "Words and Gestures" (Ogura and Watamaki, 2004), whereas the vocabulary size of the children aged 18–25 months was determined by counting words the child produced, according to the mother's description on the JCDI inventory "Words and Grammar" (Watamaki and Ogura, 2004).

The word stimuli, which were the sounds presented to the participants, were chosen by referring to the Japanese version of the MacArthur Communicative Development Inventory (JCDI) and "Tables of vocabulary obtained from Japanese children by association method" (The National Research Institute Research Report 69, Tokyo-Shoseki Ltd. 1981). Acoustic measurements of pitch, pitch range, and duration for each of the word stimuli are shown in **Table 1**. Considering the characteristics of rhythm counting in young children (Takahashi, 1997, 2001), differences in CV structure and pitch within the usual range were not expected to influence rhythm counting for morae. The other details, such as the number of syllables (morae) within the word presented, are mentioned in the Procedure section.

Apparatus and Stimuli

The infant or toddler was situated on their mother's lap in a chair. The stimuli were presented on a 17-inch laptop monitor (Dell Precision M6800), ∼50–70 cm from the infant or toddler and 70 cm from the floor. The experimenter asked the mother for the child's eyes to be located at the level of the center of the monitor. This was checked by the experimenter immediately before the start of both familiarization and test sessions. The mother was

TABLE 1 | Average acoustic measurements for test stimuli.

Stimuli	Mean pitch (Hz)	Mean pitch range (Hz)	Mean duration (s)
Ki	254.48	38.36	0.34
me	240.65	33.37	0.32
a-si	222.65	66.49	0.61
ha-na	233.00	57.33	0.55
te-re-bi	215.72	106.7	0.70
mi-ru-ku	224.14	111.5	0.69
mi-so-si-ru	231.80	96.32	0.97
ku-tu-si-ta	243.18	127.3	0.90

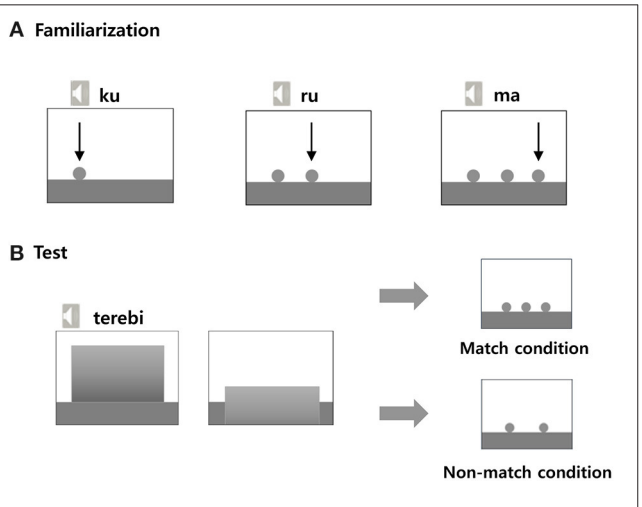

FIGURE 1 | Presentation sequences in the three-syllable condition during the familiarization trial **(A)** and test trial **(B)**.

also asked not to look at the display and not to interact with the child during the session. Presentation of the stimuli on the display were performed using Tobii Pro Studio V3.2. The child's fixation duration and gaze trajectory toward the stimuli or area of interest were recorded and measured using a Tobii X2-30 Compact Eye Tracker (display size 1920 × 1080 pixels) and Tobii Pro Studio V3.2. To confirm the child's visual attention and behavior for subsequent analyses, one video camera (Sony HDR-XR150) was also placed ∼30–45 cm behind the monitor to record the child's visual attention and behavior.

Speech and chime sounds were presented through the laptop speaker. All children were confirmed to hear the sounds during the familiarization session. All speech sounds were produced by a single female native speaker who spoke standard Japanese and did not know the purpose of this study. The sounds were created as wav. files. All of the sound stimuli were presented at an average level of 65 dB.

Procedure

The locations of experiments were determined by the children's mothers: either at the participants' homes or in our laboratory's playroom. Thus, the experiment was primarily conducted in our laboratory's playroom, but was conducted at the participants' homes in a few instances. In each situation, the child was relaxed and the surrounding environment was quiet. The child was accustomed to the playroom, the research setting, and the researcher through play interactions with the researcher before the experiment.

The revised version of the intermodal matching procedure developed by Kobayashi et al. (2005) was used in both familiarization and test sessions. In the first familiarization session of the experiment by Kobayashi et al. (2005), the infant saw two (or three) objects dropping from top to bottom in sequence, with a tone presented while each object dropped. In the second familiarization session of the experiment by Kobayashi et al. (2005), two (or three) auditory tones occurred in sequence during the period when two (or three) objects dropped from the top, although the movement of each object and the bottom remained hidden behind a screen. After the two (or three) objects had fully dropped, the screen was lowered and the two (or three) objects appeared. In this second familiarization session, the number of tones matched the number of objects that appeared after the screen was lowered. In the first portion of the test session in the experiment by Kobayashi et al. (2005), the infant heard two or three tones but the screen entirely covered the objects and their movement. In the second portion of the test session in the experiment by Kobayashi et al. (2005), the infant saw the screen lowered and two or three objects appeared. In that experiment, infants looked significantly longer at the unexpected event (e.g., two tones in the first portion of the test session and three objects in the second portion of the test session) than at the expected event (e.g., two tones in the first portion of the test session and two objects in the second portion of test session).

The present study used the revised versions of familiarization and test sessions from the experiment by Kobayashi et al. (2005) to examine whether children segmented speech sounds based on syllable units. The details of each session in the present study are explained below.

Familiarization Session

Five-point eye tracking calibration was initially conducted for each child, using Tobii Pro Studio V3.2. Following calibration, the familiarization session began. The familiarization session consisted of four trials. In each trial, the child was shown each sky blue ball dropping from the top to the bottom of the display (**Figure 1A**). Each ball was accompanied by a one-syllable speech sound while the ball dropped for 1 s. There were four types of words and trials. In the one-ball dropping trial, one ball dropped during presentation of a single one-syllable word sound. In the two-ball dropping trial, two balls dropped sequentially and each dropping ball was accompanied by a single one-syllable speech sound. Thus, two-syllable word sounds were presented with the two sequential dropping balls. In the three-ball dropping trial, three balls dropped sequentially and each dropping ball was accompanied by a single one-syllable speech sound at a rate of one-syllable sound per second. Thus, three-syllable word sounds were presented with the three sequential dropping balls, such that the first ball dropped with the sound [ku], the second ball dropped with the sound [ru], and the third ball dropped

TABLE 2 | Stimuli in familiarization session.

Stimuli	Words and breaks	Meaning	Number of balls dropped	Presentation duration (s)
1 syllable (1 mora)	e	Picture	1	5.5
2 syllables (2 morae)	u-mi	Sea	2	6.5
3 syllables (3 morae)	ku-ru-ma	Car	3	7.5
4 syllables (4 morae)	o-mu-re-tu	Omelet	4	8.5

with the sound [ma] (**Figure 1A**). The four-ball dropping trial was conducted in a manner similar to that of the two-ball and three-ball dropping trials.

Details of the one-syllable, two-syllable, three-syllable, and four-syllable word sounds presented in each trial are shown in **Table 2**. The sounds were presented to each child in the order shown.

Test Session

After the familiarization session, calibration was repeated and the test session began. The test session consisted of 16 trials: two one-syllable words × two conditions = four trials; two two-syllable words × two conditions = four trials; two three-syllable words × two conditions = four trials; and two four-syllable words × two conditions = four trials. There were three types of presentation orders during the 16 trials: A, B, and C (**Table 3**). Participants were randomized to each of the presentation orders (14 participants received order A, 15 participants received order B, and the remaining 16 participants received order C). Overall, each infant or toddler participated in 16 trials. The details of stimulus words and the numbers of presented balls in each ball presentation condition are shown in **Table 4**.

In each trial, a yellowish-green square appeared first in the middle of the display and extended to the bottom of the display (**Figure 1B**). One-word sounds (one, two, three, or four syllables) were presented at natural speed for 2 s and the square receded toward the bottom of the display, then disappeared. Immediately, the infant or toddler saw either the match condition where the number of balls (e.g., three) matched the number of syllables presented aurally just prior (e.g., three-syllable words, te-re-bi) or the non-match condition where the number of balls (e.g., two) did not match the number of syllables presented aurally just prior. This final presentation of balls (in both match and non-match conditions) lasted 6 s (**Figure 1B**).

The infant's or toddler's looking times toward the stimuli in the final presentation of each trial during the test session were measured by Tobii Pro Studio V3.2 and used for the analyses. In detail, to record the amount of time the infants and toddlers spent looking at the balls on the floor in the lower part of the display, the experimenter set a rectangle within which the stimuli appeared (2500 mm width × 1450 mm height located right above the floor), using the built-in function in Tobii Pro Studio V3.2. The looking time included all traces recorded by the eye tracker that indicated that the children were looking inside the rectangle.

TABLE 3 | Presentation orders.

Order A		Order B		Order C	
Words and breaks	**Condition**	**Words and breaks**	**Condition**	**Words and breaks**	**Condition**
mi-ru-ku	Match	ku-tu-si-ta	Match	ki	Non-match
ki	Non-match	ha-na	Non-match	mi-ru-ku	Match
a-si	Match	ki	Match	a-si	Match
te-re-bi	Non-match	mi-so-si-ru	Match	te-re-bi	Non-match
ku-tu-si-ta	Non-match	mi-ru-ku	Non-match	me	Match
ha-na	Match	te-re-bi	Match	ha-na	Non-match
a-si	Non-match	me	Non-match	te-re-bi	Match
me	Match	mi-so-si-ru	Non-match	mi-ru-ku	Non-match
ku-tu-si-ta	Match	mi-ru-ku	Match	mi-so-si-ru	Match
ha-na	Non-match	ki	Non-match	ku-tu-si-ta	Non-match
ki	Match	a-si	Match	mi-so-si-ru	Non-match
mi-so-si-ru	Match	te-re-bi	Non-match	me	Non-match
mi-ru-ku	Non-match	ku-tu-si-ta	Non-match	a-si	Non-match
te-re-bi	Match	ha-na	Match	ha-na	Match
me	Non-match	a-si	Non-match	ki	Match
mi-so-si-ru	Non-match	me	Match	ku-tu-si-ta	Match

TABLE 4 | Stimuli and ball-presentation conditions in test session.

Stimuli	Words and breaks	Meaning	Number of balls in final presentation	
			Match condition	**Non-match condition**
1 syllable (1 mora)	ki	Tree	1	3
	me	Eye	1	4
2 syllables (2 morae)	a-si	Leg	2	3
	ha-na	Flower	2	4
3 syllables (3 morae)	te-re-bi	Television	3	2
	mi-ru-ku	Milk	3	1
4 syllables (4 morae)	mi-so-si-ru	Miso soup	4	2
	ku-tu-si-ta	Socks	4	1

If the children looked away from the rectangle, the gaze traces were not counted as part of the looking time.

As we explained in the Introduction section, considering meta-analyses of studies concerning young children's segmentation of word sounds (Bergmann and Cristia, 2016) and previous data concerning the detection of stressed word sounds and one- or two-syllable word sounds (e.g., Jusczyk et al., 1999a; Houston et al., 2000; Nishibayashi et al., 2015), we assumed that the children would spend more time looking toward the stimuli in the match condition than in the non-match condition if they correctly perceived the number of syllables (basic morae).

Statistical Analysis

Most analyses were conducted using IBM SPSS Statistics, version 25. Cohen's d was calculated using Excel 2013. For the

analyses, we used the ratio data, i.e., looking time toward the matched (or non-matched) stimuli/total looking time toward the stimuli (including looking time toward the matched and non-matched stimuli), to exclude the influence of differences in total looking time among the children and to ensure the homogeneity of variance between the variables to the greatest possible degree.

RESULTS

Although we did not expect to find any effects of sex, age, or linguistic ability, based on previous studies, we confirmed that these variables did not have significant effects.

We confirmed that there were no significant effects of sex in the ratio of looking time in the match condition [24 girls: mean = 0.54, standard deviation = 0.09; 21 boys: mean = 0.56, standard deviation = 0.14; $t_{(43)} = 0.78$, $p = 0.44$, d = 0.23, 95% confidence interval (−0.04, 0.10)]. Thus, we combined the data from the girls and boys for further analysis.

To check whether there was an effect of age, we calculated the correlation coefficient between age and the ratio of looking time toward the stimuli in the match condition. We found no significant correlation ($r = -0.09$, $p = 0.56$). Therefore, there appeared to be no effect of age on infants' and toddlers' segmenting of words based on syllable units.

Next, to check whether there was an effect of linguistic ability, we calculated the correlation coefficient between the linguistic score and the ratio of looking time toward the stimuli in the match condition. For the 34 children aged 6–17 months, the numbers of words understood and produced from the JCDI inventory "Words and Gestures" were used to determine a linguistic score. When we calculated the correlation coefficient between the linguistic score and the ratio of looking time toward the stimuli in the match condition, no significant correlation was observed ($r = 0.13$, $p = 0.48$). For the 11 children aged 18–25 months, the numbers of words understood and produced from the JCDI inventory "Words and Grammar" were used to determine a linguistic score. When we calculated the correlation coefficient between the linguistic score and the ratio of looking time toward the stimuli in the match condition, no significant correlation was observed ($r = -0.003$, $p = 0.99$). Thus, there appeared to be no effect of linguistic ability on infants' and toddlers' segmenting of words based on syllable units.

We compared the ratio of looking time toward the ball stimuli between the match and non-match conditions during the final presentation in the test session. The children looked significantly longer at the stimuli in the match condition (mean = 0.55, standard deviation = 0.12) than those in the non-match condition (mean = 0.45, standard deviation = 0.12) [$t_{(44)} = 2.75$, $p = 0.009$, d = 0.82, 95% confidence interval (0.026, 0.166)] (**Figure 2**). This indicated that the children appropriately perceived the number of syllables (concurrent basic morae) included in Japanese words comprising basic morae.

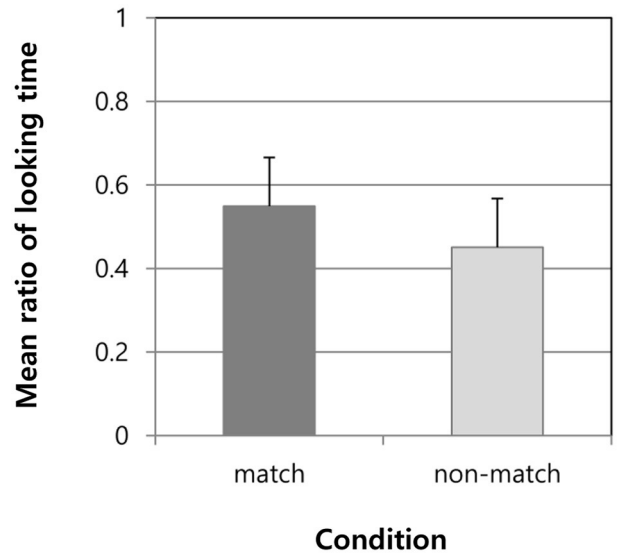

FIGURE 2 | Mean ratio of looking times during test trials in matched and non-matched conditions.

Thus, these analyses implied that children aged 6 months to 2 years perceived word segmentation based on syllables or basic morae, regardless of sex, age, and linguistic ability.

DISCUSSION

The overall results of this study were as follows. Using a new intermodal matching procedure, we found that Japanese-learning infants and toddlers aged 6–25 months tended to segment Japanese words on the basis of concurrent basic mora units within syllable units because they looked significantly longer at visual stimuli in the matched condition than in the non-matched condition. No significant differences in age or linguistic ability were found in looking time toward the stimuli, thus indicating that the children segmented the words in a similar manner, regardless of age or linguistic ability. All the words comprised only basic morae, which are segmented in a manner similar to that of syllables. Thus, each word included the same numbers of basic morae and syllables. The results have three main implications.

First, the results suggest that this new intermodal matching procedure is more useful for determining whether and how young children segment words, compared with previous methods (e.g., the headturn preference procedure) because the new intermodal matching procedure can indicate the number of units that an infant or toddler perceives within a single word. Although the previous methods could indicate whether young children discriminate familiar monosyllabic or disyllabic words from novel monosyllabic or disyllabic words, they could not indicate the number of units perceived within the test words. This new method will facilitate expansion of knowledge concerning the segmentation of words by young children.

Second, the results indicate that the infants aged ≥6 months have begun to segment words using rhythmic cues in a manner similar to that of toddlers, suggesting that Japanese infants' and toddlers' use of syllable or mora units for word segmentation depends less on language abilities or experiences. The present results demonstrate this point more clearly, including the number of perceived units, compared with prior literature, thereby suggesting that infants and toddlers can detect sound differences based on rhythmic units. Additionally, the present results are consistent with a previous finding that 6-month-old infants could differentiate monosyllabic words and syllables embedded in disyllabic words based on syllable units (Nishibayashi et al., 2015).

Third, the results suggest that children tend to segment Japanese words, which include only basic morae, on the basis of concurrent basic mora units within syllable units. The syllable has been regarded as a salient unit of rhythmic structure for perception of speech sounds (Räsänen et al., 2018) and thus has been regarded as a universal unit during speech perception (Mehler, 1981; Dumay et al., 2002). The syllable structure is present in Japanese and many other languages (Greenberg, 1999). The results of the present study imply that the syllable serves as a basic unit for perception of speech sounds by Japanese-learning infants and toddlers. Accordingly, the use of syllable units for segmentation of speech sounds at an early age could be evident among many infants and toddlers learning various languages.

However, these results should be interpreted carefully. First, to gain a more comprehensive understanding of basic segmenting skills, further studies with larger samples and wider age ranges are necessary. Second, the present study used Japanese words that included only basic morae. As we mentioned in the Introduction section, although more than 70% of Japanese sound units are basic morae, which can be segmented in a manner identical to that of syllables, the other sound units, i.e., "special morae," are segmented in a manner that differs from that of syllables. As we mentioned in the Introduction section, kitte (meaning stamp) includes three morae (ki, t, and te: "t" is a special mora unit, while "ki" and "te" are basic mora units) but two syllables: "t" is not counted as a syllable unit. Although Japanese has been included in a subcategory of syllable-timed rhythm languages (Otake et al., 1993; Arvaniti, 2009; Mazuka, 2009), different rhythms should be considered for special morae, compared with syllables. Perception and/or segmentation of special morae should be examined in Japanese-learning young children in future studies, especially as these young children begin to segment special morae based on unique mora units. Notably, Japanese-learning 4-year-old children may perceive special mora unit sound-like syllables because they often clap their hands based on syllable units (e.g., two claps for the word "kitte"), not on special mora units (e.g., three claps for the word "kitte"), whereas they vocalize words that include special morae sounds (Inagaki et al., 2000; Takahashi, 2001). Accordingly, Japanese-learning young children may segment special morae based on syllable units, rather than unique mora units, until 4 years of age. Infants and toddlers may therefore segment speech sounds based on syllable units, rather than rhythms specific to their mother tongue.

Moreover, it has not been fully investigated whether infants and toddlers whose native language belongs to the stress-timed category, such as English and German, segment speech sounds based on syllable-timed units. Previous methods only assessed whether infants and toddlers could detect differences in stress patterns between words. The new method described in this study will allow assessment of whether infants and toddlers learning a stress-timed language also segment speech sounds based on syllable-timed units. If those infants and toddlers are found to segment speech sounds based on syllable-timed units, that finding would imply that the segmentation of speech sounds based on syllable-timed units is a more innate tendency in young children.

Some consideration may be needed regarding the direction of preferential looking in the intermodal matching procedure. In the contexts of visual learning and visual testing, infants often prefer to look at novel stimuli (e.g., Fantz, 1961, 1964). However, the direction of preferential looking was not clearly established, especially in the contexts of visual learning and auditory (or tactile) testing. Familiar stimuli received greater attention (i.e., longer looking time) from infants in the usual modality-matching procedure (Starkey et al., 1983; Golinkoff et al., 1987), whereas unexpected novel events received greater attention from infants (Mix et al., 1997; Kobayashi et al., 2005). Although the direction of preferential looking toward stimuli is presumably influenced by experimental settings and conditions (Gerken et al., 2015; DePaolis et al., 2016), the biased direction toward stimuli observed using the new intermodal method also has important implications concerning infants' and toddlers' segmentation of speech sounds in future studies.

Finally, we would like to speculate regarding the relationship between the present segmenting skills and other prosodic skills and word acquisition. The segmentation of speech sounds is deeply linked to the perception of boundaries in sounds, which has led researchers to develop theories about sound units such as morphemes, strings, and phrases. As Soderstrom et al. (2003) suggested, the development of prosodic perception of morphemes, strings, and phrases may progress from more general perception of sounds to that of more language-specific sounds such that children acquire word sounds first and then develop knowledge about language. In the present study, we proposed a new method, which we found to be appropriate for examining the segmentation of speech sounds by children. This method may be useful in examining the relationships between skills regarding the segmenting and structuring of sound-units. During the course of development, perceptual skills regarding speech sounds likely move from more general to more language-specific as a child acquires words and develops knowledge about language. However, this process is ambiguous, and so further studies are needed to clarify these mechanisms.

To the best of our knowledge, this study is the first to demonstrate empirically that the new intermodal matching procedure is useful for examining segmentation of speech sounds by young children, and that Japanese-learning children aged 6–25 months segmented Japanese words (comprising basic mora sounds) on the basis of concurrent basic mora units

within syllable units, regardless of their ages and linguistic abilities. The findings will lead to further developments concerning infants' and toddlers' speech sound perception and language development.

DATA AVAILABILITY STATEMENT

Datasets generated for this study are available on request to the corresponding authors.

ETHICS STATEMENT

The procedures of the present study were approved by the Humanities and Social Sciences Research Ethics Committee of Ochanomizu University (Ethics approval number: 2017-63). Both verbal and written informed consent for the children to participate were provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

YC and IU: substantial contributions to the concept and design of the study and to the acquisition, analysis, and interpretation of data. YC and IU: drafting of the manuscript and revising it critically for important intellectual content. Both authors contributed to manuscript revision and have read and approved the submitted version.

FUNDING

This research was supported by JSPS KAKENHI (Grant Number: JP 18H05524).

ACKNOWLEDGMENTS

The authors thank the infants, toddlers, and their families who participated in this study.

REFERENCES

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica* 66, 46–63. doi: 10.1159/000208930

Bates, E., and Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: evidence from acquisition, aphasia and real-time processing. *Lang. Cogn. Processes* 12, 507–584. doi: 10.1080/016909697386628

Bergelson, E., and Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3253–3258. doi: 10.1073/pnas.1113380109

Bergelson, E., and Swingley, D. (2015). Early word comprehension in infants: replication and extension. *Lang. Learn. Dev.* 11, 369–380. doi: 10.1080/15475441.2014.979387

Bergmann, C., and Cristia, A. (2016). Development of infants' segmentation of words from native speech: a meta-analytic approach. *Dev. Sci.* 19, 901–917. doi: 10.1111/desc.12341

Bernard, C., and Gervain, J. (2012). Prosodic cues to word order: what level of representation? *Front. Psychol.* 3:451. doi: 10.3389/fpsyg.2012.00451

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., and Mehler, J. (1988). An investigation of young infants' representation of speech sounds. *J. Exp. Psychol.* 117, 21–33. doi: 10.1037/0096-3445.117.1.21

Bijeljac-Babic, R., Bertoncini, J., and Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Dev. Psychol.* 29, 711–721. doi: 10.1037/0012-1649.29.4.711

Bosch, L., Figueras, M., Teixid,ó, M., and Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: evidence from infants acquiring syllable timed languages. *Front. Psychol.* 4:106. doi: 10.3389/fpsyg.2013.00106

Christophe, A., and Morton, J. (1998). Is Dutch native English? Linguistic analysis by 2 month-old infants. *Dev. Sci.* 1, 215–219. doi: 10.1111/1467-7687.00033

Christophe, A., Nespor, M., Guasti, M. T., and van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: the case of the head-direction parameter. *Dev. Sci.* 6, 213–222. doi: 10.1111/1467-7687.00273

Dan, H., Dan, I., Sano, T., Kyutoku, Y., Oguro, K., Yokota, H., et al. (2013). Language-specific cortical activation patterns for verbal fluency tasks in Japanese as assessed by multichannel functional near-infrared spectroscopy. *Brain Lang.* 126, 208–216. doi: 10.1016/j.bandl.2013.05.007

Dankovičová, J., and Dellwo, V. (2007). Czech speech rhythm and the rhythm class hypothesis. *International Congress of Phonetic Sciences* (Saarbrüecken) 1241–1244.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalysed. *J. Phonetics* 11, 51–62. doi: 10.1016/S0095-4470(19)30776-4

DePaolis, R. A., Keren-Portnoy, T., and Vihman, M. (2016). Making sense of infant familiarity and novelty responses to words at lexical onset. *Front. Psychol.* 7:715. doi: 10.3389/fpsyg.2016.00715

Dumay, N., Frauenfelder, U. H., and Content, A. (2002). The role of the syllable in lexical segmentation in French: word-spotting data. *Brain Lang.* 81, 144–161. doi: 10.1006/brln.2001.2513

Echols, C. H., Crowhurst, M. J., and Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *J. Memory Lang.* 36, 202–225. doi: 10.1006/jmla.1996.2483

Fantz, R. L. (1961). The origin of form perception. *Sci. Am.* 204, 66–72. doi: 10.1038/scientificamerican0561-66

Fantz, R. L. (1964). Visual experience in infants: decreased attention to familiar patterns relative to novel ones. *Science* 146, 668–670. doi: 10.1126/science.146.3644.668

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., et al. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, 5–173. doi: 10.2307/1166093

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D. J., Bates, E., Hartung, J. P., et al. (1993). *The MacArthur Communicative Development Inventories: Users Guide and Technical Manual*. San Diego, CA: Singular Publishing Group.

Floccia, C., Christophe, A., and Bertoncini, J. (1997). High-amplitude sucking and newborns: the quest for underlying mechanisms. *J. Exp. Child Psychol.* 64, 175–198. doi: 10.1006/jecp.1996.2349

Gerken, L., Dawson, C., Chatila, R., and Tenenbaum, J. (2015). Surprise! Infants consider possible bases of generalization for a single input example. *De. Sci.* 18, 80–89. doi: 10.1111/desc.12183

Gervain, J., Nespor, M., Mazuka, R., Horie, R., and Mehler, J. (2008). Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study. *Cogn. Psychol.* 57, 56–74. doi: 10.1016/j.cogpsych.2007.12.001

Gleitman, L. R., and Wanner, E. (1982). "Language acquisition: the state of the state of the art," in *Language Acquisition: The State of the State of the Art*, eds E. Wanner and L. Gleitman (Cambridge: Cambridge University Press), 3–48.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., and Gordon, L. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *J. Child Lang.* 14, 23–45. doi: 10.1017/S030500090001271X

Goyet, L., Nishibayashi, L.-L., and Nazzi, T. (2013). Early syllabic segmentation of fluent speech by infants acquiring French. *PLoS ONE* 8:e79646. doi: 10.1371/journal.pone.0079646

Grabe, E., and Low, L. (2002). "Durational variability in speech and the rhythm class hypothesis," in *Laboratory Phonology*, Vol. 7, eds C.Gussenhoven and N. Warner (Berlin: Mouton de Gruyter), 515–546.

Greenberg, S. (1999). Speaking in shorthand - a syllable-centric perspective for understanding pronunciation variation. *Speech Commun.* 29, 159–176. doi: 10.1016/S0167-6393(99)00050-3

Hoequist, C. (1983). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica* 40, 203–237. doi: 10.1159/000261692

Höhle, B., Bijeljac-Babicb, R., Herold, B., Weissenborn, J., and Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behav. Dev.* 32, 262–274. doi: 10.1016/j.infbeh.2009.03.004

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., and Cutler, A. (2000). Cross-language word segmentation by 9-month-old infants. *Psychonomic Bull. Rev.* 7, 504–509. doi: 10.3758/BF03214363

Inagaki, K., Hatano, G., and Otake, T. (2000). The effect of Kana literacy acquisition on the speech segmentation unit used by Japanese young children. *J. Exp. Child Psychol.* 75, 70–91. doi: 10.1006/jecp.1999.2523

Johnson, E. K., and Jusczyk, P. W. (2001). Word segmentation by 8-month-old infants: when speech cues count for more than statistics. *J. Memory Lang.* 44, 548–567. doi: 10.1006/jmla.2000.2755

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., Cutler, A., and Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Dev.* 64, 675–687. doi: 10.2307/1131210

Jusczyk, P. W., Hohne, E. A., and Bauman, A. (1999a). Infant's sensitivity to allophonic cues for word segmentation. *Perception Psychophys.* 61, 1465–1476. doi: 10.3758/BF03213111

Jusczyk, P. W., Houston, D. M., and Newsome, M. (1999b). The beginnings of word segmentation in English-learning infants. *Cogn. Psychol.* 39, 159–207. doi: 10.1006/cogp.1999.0716

Jusczyk, P. W., Kennedy, L. J., and Jusczyk, A. M. (1995). Young infants' retention of information about syllables. *Infant Behav. Dev.* 18, 27–41. doi: 10.1016/0163-6383(95)90005-5

Kobayashi, T., Hiraki, K., and Hasegawa, T. (2005). Auditory-visual intermodal matching of small numerosities in 6-month-old infants. *Dev. Sci.* 8, 409–419. doi: 10.1111/j.1467-7687.2005.00429.x

Mattys, S. L., and Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78, 91–121. doi: 10.1016/S0010-0277(00)00109-8

Mazuka, R. (2009). Acquisition of linguistic rhythm and prosodic bootstrapping hypothesis. *J. Phonetic Soc. Japan* 13, 19–32. doi: 10.24467/onseikenkyu.13.3_19

Mehler, J. (1981). The role of syllables in speech processing: infant and adult data. *Philos. Trans. R. Soc. London Ser. B* 295, 333–352. doi: 10.1098/rstb.1981.0144

Mehler, J., Dommergues, J. Y., Frauenfelder, U., and Segui, J. (1981). The syllable's role in speech segmentation. *J. Verbal Learn. Verbal Behav.* 20, 298–305. doi: 10.1016/S0022-5371(81)90450-3

Mix, K. S., Levine, S. C., and Huttenlocher, J. (1997). Numerical abstraction in infants: another look. *Dev. Psychol.* 33, 423–428. doi: 10.1037/0012-1649.33.3.423

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Perception Perform.* 24, 756–766. doi: 10.1037/0096-1523.24.3.756

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., and Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: emerging evidence for crosslinguistic differences. *J. Memory Lang.* 54, 283–299. doi: 10.1016/j.jml.2005.10.004

Nishibayashi, L.-.L., Goyet, L., and Nazzi, T. (2015). Early speech segmentation in French-learning infants: monosyllabic words versus embedded syllables. *Lang. Speech* 58, 334–350. doi: 10.1177/0023830914551375

Ogino, T., Hanafusa, K., Morooka, T., Takeuchi, A., Oka, M., and Ohtsuka, Y. (2017). Predicting the reading skill of Japanese children. *Brain Dev.* 39, 112–121. doi: 10.1016/j.braindev.2016.08.006

Ogura, T., and Watamaki, T. (2004). *Technical Manual of the Japanese MacArthur Communicative Development Inventory: Words and Gesture*. Kyoto: Kyoto International Social Welfare Exchange Center.

Okumura, Y., Kobayashi, T., and Oshima-Takane, Y. (2016). Child language development: what is the differences between Japanese and English? *NTT Technical J.* 28, 21–25. Available online at: https://www.ntt.co.jp/journal/1609/files/jn20160921.pdf

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *J. Memory Lang.* 32, 258–278. doi: 10.1006/jmla.1993.1014

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265–292. doi: 10.1016/S0010-0277(99)00058-X

Räsänen, O., Doyle, G., and Frank, M. C. (2018). Pre-linguistic segmentation of speech into syllable-like units. *Cognition* 171, 130–150. doi: 10.1016/j.cognition.2017.11.003

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926

Shinohara, S. (1996). The roles of the syllable and the mora in Japanese: adaptations of French words. *Cahiers de Linguistique Asie Orientale* 25, 87–112. doi: 10.3406/clao.1996.1493

Soderstrom, M., Seidl, A., Nelson, D. G. K., and Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: evidence from prelinguistic infants. *J. Memory Lang.* 49, 249–267. doi: 10.1016/S0749-596X(03)00024-X

Song, J. Y., Demuth, K., and Morgan, J. (2018). Input and processing factors affecting infants' vocabulary size at 19 and 25 months. *Front. Psychol.* 9:2398. doi: 10.3389/fpsyg.2018.02398

Starkey, P., Spelke, E. S., and Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science* 222, 179–181. doi: 10.1126/science.6623069

Takahashi, N. (1997). A developmental study of wordplay in preschool children: Japanese game of "shiritori." *Jpn. J. Dev. Psychol.* 8, 42–52.

Takahashi, N. (2001). "Knowledge of letters and phonological awareness," in *Introduction to Language Development*, ed E. Hatano (Tokyo: Taishukan Publishing), 196–218.

Watamaki, T., and Ogura, T. (2004). *Technical Manual of the Japanese MacArthur Communicative Development Inventory: Words and Grammar*. Kyoto: Kyoto International Social Welfare Exchange Center.

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., and Höhle, B. (2012). How each prosodic boundary cue matters: evidence from German infants. *Front. Psychol.* 3:580. doi: 10.3389/fpsyg.2012.00580

Werker, J. F., Fennell, C. T., Corcoran, K. M., and Stager, C. L. (2002). Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy* 3, 1–30. doi: 10.1207/S15327078IN0301_1

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., and Werker, J. F. (2010). The development of perceptual grouping biases in infancy: a Japanese–English cross-linguistic study. *Cognition* 115, 356–361. doi: 10.1016/j.cognition.2010.01.005

Yoshioka, Y., and Tosa, K. (2014). First word of typically developing children and language impaired children. *Niigata J. Health Welfare* 13, 15–19. Available online at: https://core.ac.uk/download/pdf/70372738.pdf

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.