Morae and Syllables: Rhythmical Basis of Speech Representations in Neonates*

JOSIANE BERTONCINI, CAROLINE FLOCCIA, THIERRY NAZZI, and JACQUES MEHLER

Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS-EHESS

KEY WORDS

cue transparency

infant speech perception

rhythm

speech units

ABSTRACT

Are neonates sensitive to the different rhythmical units that are used in different spoken languages? And do they use these units to represent and discriminate multisyllabic words? In the present study, we used the High-Amplitude Sucking procedure to test whether 3-day-old French infants discriminate lists of Japanese words. The lists of words differed either in the number of syllabic units or in the number of sub-syllabic units such as morae. In Experiment 1, infants heard bisyllabic versus trisyllabic words (e.g.: *iga* vs. *hekiga*); in Experiment 2, they were presented with bimoraic versus trimoraic bisyllabic words (e.g.: *iga* vs. *iNga*). The results corroborate those obtained by Bijeljac-Babic, Bertoncini, and Mehler (1993), providing further evidence that neonates discriminate bisyllabic from trisyllabic words. In contrast, neonates do not appear to discriminate bisyllabic words that vary in number of sub-syllabic units. It is proposed that syllables are particularly salient units during the initial stage of speech processing, irrespective of which language and rhythmical structure is heard.

INTRODUCTION

During the process of word recognition, adult listeners appear to use perceptual strategies that are related to the phonological properties of their language. Native speakers of French have been shown to rely mainly on the syllable to segment and access spoken words (Mehler, Dommergues, Frauenfelder, & Segui, 1981). In contrast, native speakers of English do not segment the speech stream into syllables (Cutler, Mehler, Norris, &

*Acknowledgments: This research was supported by grants from the Human Frontier Scientific Program, the European Communities Human Capital Program and the CNET (Convention 837BD28 00790 9245 LAA/TSS/CMC). We would like to thank Kazuaki Miyagishima for his help in constructing the experimental stimuli. We would also like to thank Brit van Ooyen for her comments on a previous version of the manuscript.

Address for correspondence: Josiane Bertoncini, Laboratoire de Sciences Cognitives et Psycholinguistique, 54, Boulevard Raspail, 75270 Paris Cedex 06, FRANCE. Fax number: (33 1) 45.44.98.35. E-mail: josiane@lscp.msh-paris.fr

Segui, 1986) but rather seem to use stress rhythm, that is, alternations of stressed and unstressed syllables, to locate word boundaries (Cutler & Norris, 1988; Cutler & Butterfield, 1990, 1992). Considering that French is described as syllable-timed while English is described as stress-timed, such results indicate that speakers develop language-specific strategies that rely on rhythmic properties to retrieve lexical items. In this view, the rhythmic structure of the language forms the basis for segmenting the continuous speech stream into meaningful units. This hypothesis received further support from recent studies using Japanese, a language whose rhythmic structure is based on another unit: the mora (Otake, Hatano, Cutler, & Mehler, 1993). Their results suggest that the mora acts as the perceptual segmentation unit for Japanese listeners.

Moreover, listeners use the perceptual strategies developed for their own language even when listening to foreign-language inputs (Cutler et al., 1986; Otake et al., 1993; Cutler & Otake, 1994). These strategies that reflect the use of the rhythmic properties of the native language would be part of speakers' language-specific competence, which may have developed from the earliest stages of language acquisition (Mehler, Dupoux, & Segui, 1991). If adults base their speech segmentation strategies on different rhythmic units, we need to know how infants discover the unit most appropriate for the language they are in the process of acquiring. Is it the case that infants are initially sensitive to any rhythmic unit and can use this to represent spoken utterances? Or do they first use some kind of universal unit of representation (which still needs to be determined), from which they will specify the characteristics of the unit of their native language?

Recent data have shown that newborns do not process multisyllabic utterances as undecomposable entities, but that they are able to detect common structural properties shared by multisyllabic strings. In particular, Bijeljac-Babic et al. (1993) showed that neonates are sensitive to the number of syllabic units in multisyllabic utterances and that they use this capacity to discriminate varied sets of 2-CV (Consonant-Vowel) from 3-CV French pseudowords. These results suggest that newborns represent complex multisyllabic utterances in terms of syllables (or some correlated unit, such as the vowels that form the nucleus of syllables). Would such a result hold only for French stimuli or would it generalize to every natural language, and especially to those that are not syllable-timed? If so, this would make the syllable a good candidate for being the primary unit of infants' initial processing system. Alternatively, if newborns were sensitive to any kind of rhythmic unit, they should be able to represent speech in terms of whichever unit was most appropriate to the language heard.

The present study was designed to distinguish between these two alternatives. Two experiments were conducted with French neonates who were presented with multisyllabic Japanese words. The Japanese language (Tokyo standard) was chosen because its phonological structure allowed us to compare infants' sensitivity to syllabic and sub-syllabic units such as certain types of morae. In Japanese, syllables can be composed of either one mora (V, CV), or two morae. There are different types of bimoraic syllables: long vowels (VV), CVVs, and closed syllables (V or CV followed by a nasal consonant (N), or a geminate consonant (represented as (Q)). Therefore, using Japanese words composed of monomoraic or bimoraic syllables permits us to compare infant perception of syllabic and sub-syllabic units within a single phonological system.

To date, there is no evidence that infants represent speech in terms of units smaller than syllables. Studies that have specifically addressed this issue failed to find evidence that young infants' representations are structured in terms of phonetic segments (Jusczyk & Derrah, 1987; Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Bijeljac-Babic et al., 1993). However, while morae can be smaller than syllables, they might play a role different from that of phonemes in speech perception. Morae are rhythmic units, in that it has been shown that the duration of a word is a function of the number of mora components (Hoequist, 1983; Port, Dalby, & O'Dell, 1987), and that each mora receives a pitch value (High or Low). Thus, although sub-syllabic, the mora could be salient enough to be processed by infants as a perceptual unit.

As in Bijeljac-Babic et al. (1993), infants were presented with two large sets of phonetically varied Japanese words. All words within a set shared the same number of components (two or three), which could be either syllabic or moraic. In a discrimination experiment involving lists of items (rather than two single items), discrimination can be performed only if infants extract two different overall perceptual patterns from the two lists. Such patterns will have to contain the property (or properties) shared by all items in the list, while discarding irrelevant variations. In other words, in order to discriminate the two lists, infants will have to build a representation that includes the property on which the lists are contrasted. If infants can represent words equally well in terms of syllables and morae, then discrimination should be obtained in both cases, that is, when words differ in number of syllables as well as when they differ in number of morae. In contrast, if words are represented primarily in terms of syllable-sized units, then discrimination should be obtained only when the property distinguishing the lists is the number of syllabic components of the words. Such a result would extend the Bijeljac-Babic et al. (1993) findings to a different phonological system, and would provide new evidence that infants are prepared to process certain kinds of speech units preferentially.

The French and Japanese languages differ not only in their rhythmic structure but also in another prosodic characteristic. In French, prosody is not contrastive at the lexical level. However, in Japanese words are characterized by their intonational contour, each mora being pronounced as either high or low pitched. The sequences of high and low morae follow certain rules that depend on the number of morae and the position of the mora that bears the pitch accent. In the present study, using real Japanese words, we decided to include two possible types of contour, that is, descending (H-L or H-L-L) and ascending (L-H or L-H-H), which correspond to different accentuation patterns. In Japanese (Tokyo standard), the accent corresponds to a pitch fall, thus located on the last high mora before a low one. It follows that in words with a descending contour, the accent is on the first mora, while there is no accent location in words with an ascending contour as they present no pitch fall (Shibatani, 1990). In Bijeljac-Babic et al.'s study mentioned above, all the stimuli consisted of French pseudowords naturally produced with the same global contour (ascending with a lengthened final syllable). In order to test discrimination of Japanese words in a similar context of minimal prosodic variation, we chose to present each subject with only one contour, that is, either ascending or descending.

EXPERIMENT 1

Method

Stimuli. Four lists of Japanese words were recorded by a female native speaker. Two lists were made of bisyllabic words (VCV or CVCV), one constituted of 29 words with an ascending intonational contour (L-H) such as *iga*, *tema*, and one of 29 words with a descending intonational contour (H-L) such as *rika*, *tomi*. The other two lists were each made up of 29 trisyllabic words, one with an ascending intonational contour (L-H-H) such as *hekiga*, *temari* and one with a descending intonational contour (H-L-L) such as *erika*, *tomita*. In these words, the syllabic and the moraic structures strictly matched: Each syllable corresponded to one mora.

The mean duration of the stimuli was 457 ms (SD = 62.1, range 351 – 562 ms) for the H-L words; 477 ms (SD = 52.5, range 370 – 580 ms) for the L-H words; 630 ms (SD = 59.4, range 536 – 789 ms) for the H-L-L words; and 725 ms (SD = 54.8, range 606 – 908 ms) for the L-H-H words. The lists of stimuli with their durations are displayed in Appendix 1. The distributions of durations for the four lists of words are given in Figure 1.

Subjects. Eighty-three infants were tested within the first five days after delivery. They were healthy, full-term newborns recruited at the Baudelocque Maternity Hospital in Paris, France. They had suffered no complication during pregnancy and delivery, and they were classified as "normal" after neurological evaluation on their first or second day of life (Amiel-Tison, 1977). All subjects weighed more than 2,700 g at birth and had 1-min Apgar scores of more than 8, and 5-min Apgar scores of 10.

Forty-three infants were excluded for the following reasons: falling asleep (11), rejecting the pacifier (8), irregular or insufficient sucking (11), failing to achieve the familiarization criterion within 12 min (5), crying (2), becoming agitated (2), technical problems (4). Forty infants (16 girls and 24 boys) completed the experimental session. They had a mean age of 2.5 days (SD = 0.75) and a mean gestational age of 39.5 weeks (SD = 1.12). Their mean birth weight was 3336 g (SD = 345.8).

Procedure and Design. Infants were tested individually in a sound-attenuated booth. They were awakened about two hours after feeding and roused until a quiet, active state was obtained. They were then placed in a reclining position in a special chair that reduces head movements. The pacifier was held by an adjustable mechanical arm to avoid any intervention by the experimenter during the experimental session.

The experimental session started with a two-minute silent baseline, during which the infants' spontaneous sucking rate was registered. Then the familiarization period started, during which the presentation of stimuli was made contingent upon high-amplitude (HA) sucking. The criterion for selecting HA sucks was fixed at the same value for all subjects, and corresponded to an average of about 80% of sucks being considered as HA sucks. The familiarization period lasted at least five minutes, and was completed when a 25% decrement in sucking rate was registered over two consecutive minutes compared to the rate of the immediately preceding minute. When this criterion was reached, stimuli were changed for infants in the experimental group, while infants in the control group continued to be presented with the same stimuli. The postshift period lasted four minutes.

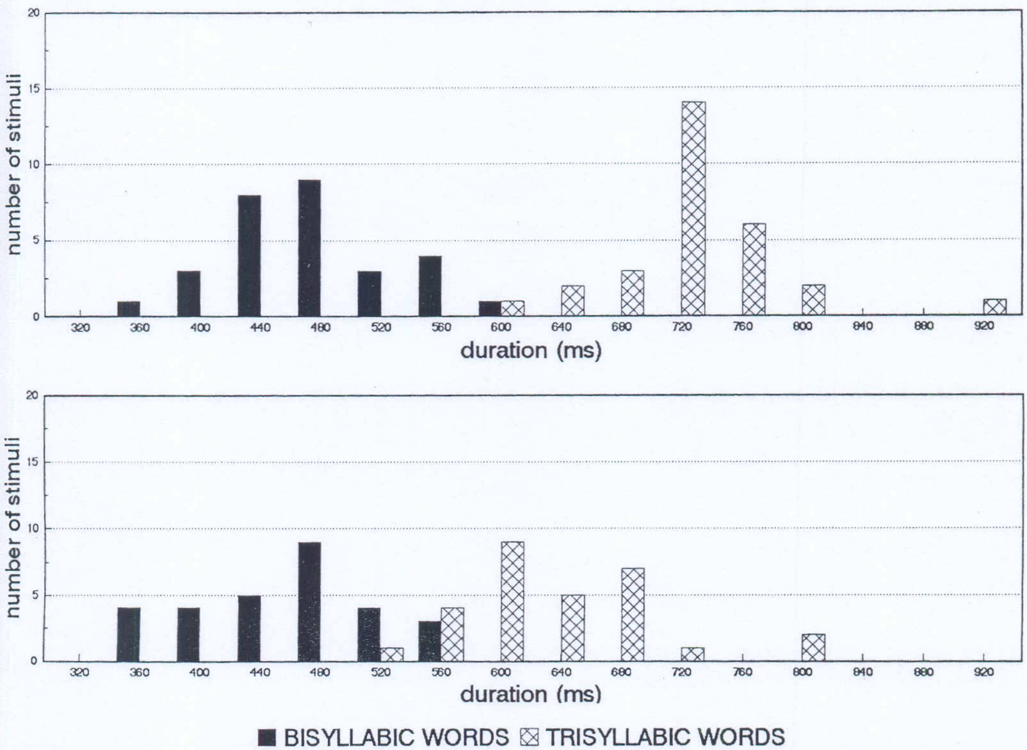

Figure 1

Distribution of durations of bisyllabic and trisyllabic words used in Experiment 1. (a) stimuli with ascending intonational contour. (b) stimuli with descending intonational contour. (Values correspond to the center of each 40-ms-wide category range.)

All the stimuli were stored in a digital form on the computer. For each period, the corresponding 29 stimuli were first presented in a random order, then randomly reordered and played again, and so on until the end of the period. The maximum rate of stimulus presentation was of one stimulus every 1350 ms.

Infants were randomly assigned to one of the eight subgroups resulting from the crossing of the three main factors: Preshift Stimulus condition (bisyllabic words or trisyllabic words), Intonational Contour (ascending or descending) and Experimental Condition (control or experimental). There were five subjects in each subgroup.

Apparatus. A BABISOL pacifier was connected to a GOULD pressure transducer which in turn was connected to an IBM-PC 386 compatible computer via an analog to digital board (Data Translation 2814). The computer detected each sucking response, and registered its amplitude and the time between its occurrence and that of the preceding sucking. When a HA suck is detected, a stimulus is delivered by an OROS AU22 board, via a stereo amplifier (ROTEL RA820B*3), and two loudspeakers (MARTIN Control Monitor DB92), at a normal level of intensity.

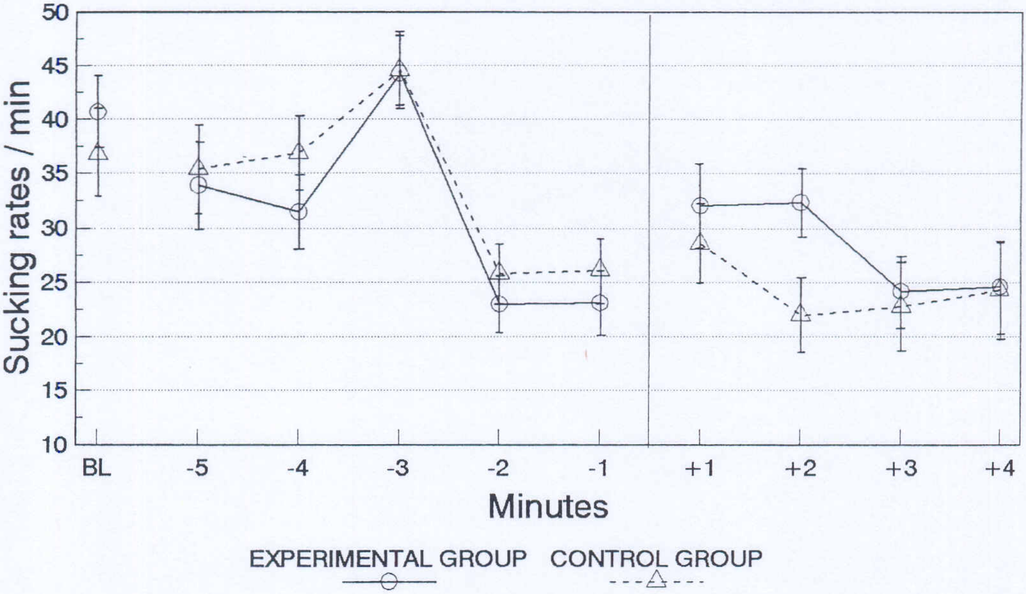

Figure 2

Sucking-rate averages during baseline (BL), last five minutes of familiarization (– 5 to – 1), and four minutes of test (+ 1 to + 4) for experimental and control groups in Experiment 1. The bars above and below each point indicate the standard error of the mean.

Results

For the 40 infants who completed the test, the mean duration of the familiarization phase was eight minutes. There was no difference between the experimental group and the control group (8.5 and 7.5 min respectively; $F(1, 38) = 1.42$, $p = .24$). The average sucking rates of experimental and control groups are displayed in Figure 2.

An overall analysis of variance was performed on HA sucking rates during baseline and the last five preshift minutes with the main factors of Condition (experimental vs. control), Preshift Stimulus (bisyllabic vs. trisyllabic word) and Contour (ascending vs. descending). During these two periods, there was no significant effect of Condition and Contour, $F(1, 32) < 1$ in all cases). A significant effect of the Preshift Stimulus factor was observed on the baseline, $F(1, 32) = 4.35$, $p = .045$, between subjects who were to hear bisyllabic words (mean sucking rate: 43.8 per minute) and subjects who were to hear trisyllabic words (mean sucking rate: 33.8 per minute). However, this difference disappeared during presentation of the stimuli, and was not significant on the last five preshift minutes, $F(1, 32) = 1.59$, $p = .22$. Moreover, the Preshift Stimulus factor did not interact with any other factor on the preshift measures, $F(1, 32) < 1$ in all cases.

To evaluate the effect of the change of stimuli, the last two preshift minutes were compared with the first two postshift minutes. These comparisons revealed a significant increase in sucking rates in the experimental group, $F(1, 19) = 9.39$, $p < .01$, and no difference in the control group, $F(1, 19) < 1$. Moreover, there was a significant interaction between Shift (two last preshift min vs. two first postshift min) and Condition,

TABLE 1

Means of differences between the sucking rates in the last two preshift minutes and the first two postshift minutes for the different groups in Experiment 1. Numbers in parenthesis indicate the standard error of the mean

	Contour	
Group	*Ascending*	*Descending*
Experimental	9.20 (3.7)	9.15 (4.5)
Control	– 2.55 (2.7)	1.15 (4.6)

$F(1, 38) = 5.87$, $p = .02$, indicating that the experimental group's recovery was significantly larger (+ 9.17 sucks) than that of the control group (– 0.7 sucks).

No interaction between Condition, Shift and the two other main factors (Preshift Stimulus and Contour) reached significance, $F(1, 32) < 1$ in all cases. This indicates that infants in all experimental subgroups reacted similarly to the stimulus change. The means of the differences between the sucking rates in the last two preshift minutes and the first two postshift minutes are given in Table 1 for the experimental and control groups that heard either ascending or descending contours.

Discussion

The results of Experiment 1 indicate that infants can discriminate lists of bisyllabic versus trisyllabic Japanese words. It is assumed here that discrimination reflects the fact that infants have built a common representation of the phonetically varied words in a list, and that this representation includes the property on which the two lists were contrasted, that is, the number of syllabic components of the words. These findings converge with those obtained by Bijeljac-Babic et al. (1993) who showed (1) that infants discriminate 2-CV from 3-CV French pseudowords and (2) that this discrimination is not merely based on differences in the durational characteristics of the stimuli. In the present study, as we used only naturally spoken words, the bisyllabic and trisyllabic words have different mean durations and barely overlap (see Figure 1). However, we observe no difference between the experimental groups according to the intonational contour they were presented with, whereas bisyllabic and trisyllabic words with a descending contour have less different durational characteristics (12% overlap and 173 ms difference between duration means) than those with an ascending contour (no overlap and 248 ms difference between duration means).

Thus, the present results provide further evidence that infants can extract a syllable-based representation of multisyllabic utterances. Moreover, this ability does not apply exclusively to French, a language for which the syllable would be the most appropriate perceptual unit, but also to a language with a completely different phonological system.

Our initial intention was to compare infants' ability to process syllabic and sub-syllabic moraic units perceptually: Would infants also be able to discriminate Japanese words that are composed of two morae from words composed of three morae, when these words have identical length in number of syllables? Experiment 2 was designed to answer this question. All stimuli were bisyllabic Japanese words. Half of them were bimoraic words, each syllable being monomoraic (as in *kago*, *tomi*, *seki*, *buke*); the other half was constituted of trimoraic words, in which one syllable was bimoraic and one monomoraic (as in *kaNgo*, *tomiN*, *seQki*, *buuke*). As in the previous experiment, neonates were tested on their reaction to a change from a set of bimoraic words to a set of trimoraic words (or vice versa).

If infants not only represent multisyllabic words in terms of number of syllabic units, but if in addition they have access to sub-syllabic information, as long as this corresponds to a rhythmic speech unit, they should discriminate bimoraic from trimoraic words. Alternatively, if infants represent multisyllabic words only on the basis of syllabic structure, and are not sensitive to sub-syllabic rhythmic information, they should represent both bimoraic and trimoraic words as being composed of two syllables. In this case, no discrimination should be obtained.

EXPERIMENT 2

Method

Stimuli. Four lists of 30 bisyllabic words were constructed. Two lists were made up of bimoraic words, one with a H-L contour (ex: *mika*, *buke*), the other with a L-H contour (ex: *kago*, *seki*), and two lists were made up of trimoraic words, one with H-L-L words and the other with L-H-H words. In the two lists of trimoraic words, four types of sub-syllabic morae were included: nasal in medial position (10 items, ex: *kaNgo*), nasal in final position (five items, ex: *mikaN*), geminate consonant (10 items, ex: *seQki*) and long first vowel (five items, ex: *buuke*).

The mean duration was 405 ms (SD = 49.3, range 319 – 550 ms) for the H-L words; 379 ms (SD = 39.1, range 304 – 497 ms) for the L-H words; 537 ms (SD = 61.5, range 393 – 682 ms) for the H-L-L words; and 525 ms (SD = 41.7, from 448 – 633 ms) for the L-H-H words. The lists of stimuli with their durations are displayed in Appendix 2. The distributions of durations for the four lists of words are given in Figure 3.

Subjects. Eighty-three infants were tested within the first five days after delivery. They were healthy, full-term newborns recruited at the Baudelocque Maternity Hospital in Paris, France. They were selected according to the same criteria as those used in the previous experiment.

Forty-three infants were excluded for the following reasons: falling asleep before the switch (16), rejecting the pacifier (2), irregular or insufficient sucking (7), failing to achieve the familiarization criterion within 12 min (11), crying (3), becoming agitated (4). Forty infants (19 girls and 21 boys) completed the experimental session. They had a mean age of 3.04 days (SD = 0.92) and a mean gestational age of 39.4 weeks (SD = 0.96). Their mean birth weight was 3490 g (SD = 395).

Procedure, Apparatus and Design. The procedure and the apparatus were identical to those described in Experiment 1, except that the maximum rate of stimulus presentation

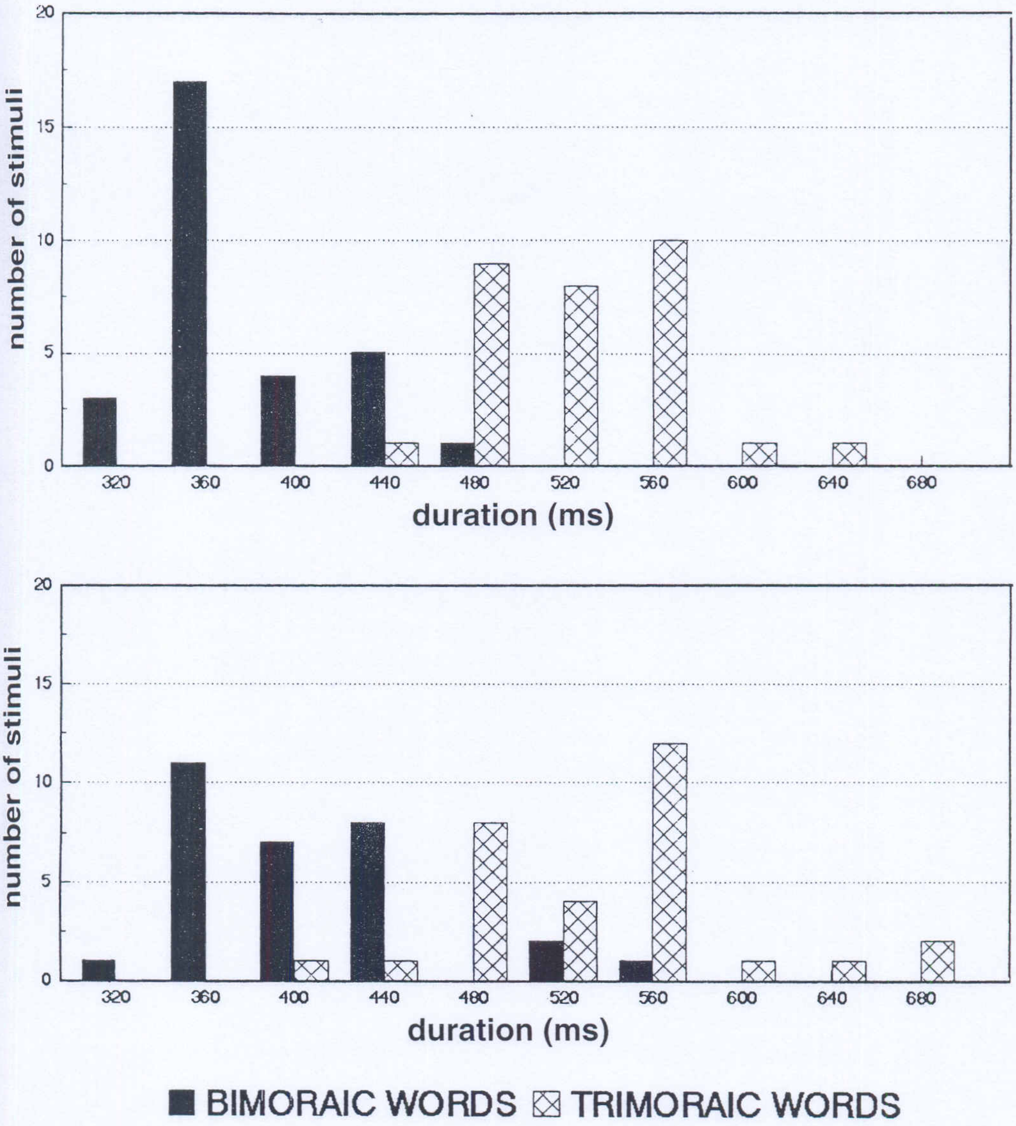

Figure 3

Distribution of durations of bimoraic and trimoraic words used in Experiment 2. (a) stimuli with ascending intonational contour. (b) stimuli with descending intonational contour. (Values correspond to the center of each 40-ms-wide category range.)

was of one stimulus every 1200 ms (instead of 1350 ms) as stimuli were in average shorter in Experiment 2 than in Experiment 1.

Infants were randomly assigned to one of the eight subgroups resulting from the crossing of the three main factors: Preshift Stimulus condition (bimoraic vs. trimoraic words), Intonational Contour (ascending or descending) and Experimental Condition (control or experimental). There were five subjects in each subgroup.

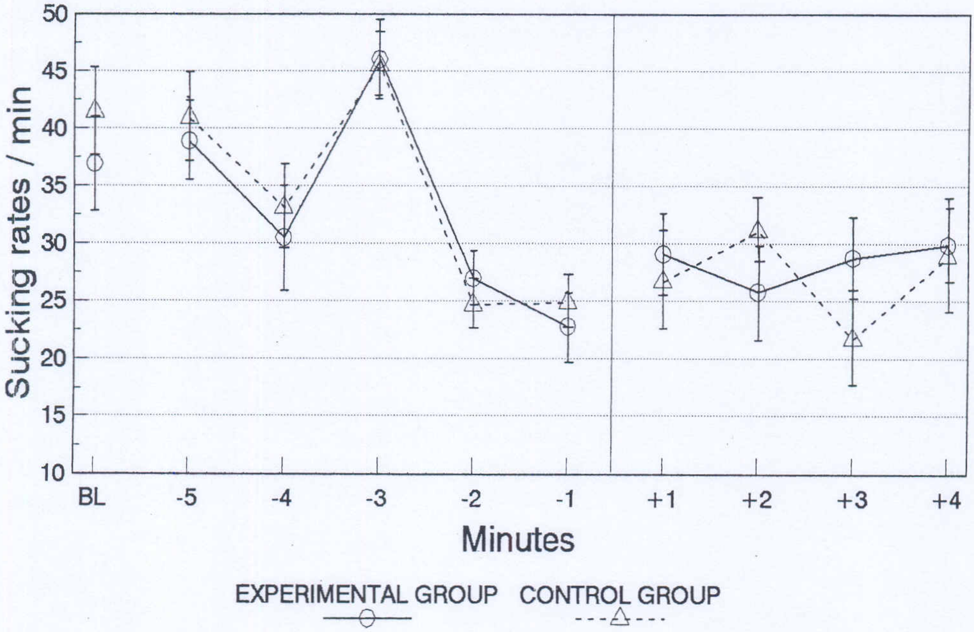

Figure 4

Sucking-rate averages during baseline (BL), last five min of familiarization (– 5 to – 1), and four min of test (+ 1 to + 4) for experimental and control groups in Experiment 2. The bars above and below each point indicate the standard error of the mean.

Results

For the 40 infants who completed the test, the mean duration of the familiarization phase was 8.8 minutes. There was no significant difference between the experimental group and the control group (8.7 vs. 8.95 minutes; $F(1, 38) < 1$). The average sucking rates of experimental and control groups are displayed in Figure 4.

As for Experiment 1, an overall analysis of variance was performed on HA sucking rates during baseline and the last five preshift minutes with Condition (experimental vs. control), Preshift Stimulus (bimoraic vs. trimoraic words) and Contour (ascending vs. descending) as main factors. The analysis revealed no difference between groups on both periods, $F(1, 32) < 1$ in all cases.

The comparisons between the last two preshift minutes and the first two postshift minutes revealed no significant difference in sucking rates for both the experimental group, $F(1, 19) = 1.45, p = .24$, and the control group, $F(1, 19) = 1.72, p = .21$. Accordingly, there was no interaction between Shift (two minutes before vs. two minutes after the shift) and Condition, $F(1, 38) < 1$. In addition, neither the interaction among Shift, Condition and Preshift Stimulus, $F(1, 32) = 2.16$, $p = .15$, nor the one among Shift, Condition and Contour reached significance, $F(1, 32) < 1$. Table 2 displays the means of the differences between the sucking rates in the last two preshift minutes and the first two postshift minutes for the experimental and control groups that heard either ascending or descending contours.

TABLE 2

Means of differences between the sucking rates in the last two preshift minutes and the first two postshift minutes for the different groups in Experiment 2. Numbers in parenthesis indicate the standard error of the mean

	Contour	
Group	*Ascending*	*Descending*
Experimental	1.60 (2.3)	3.55 (3.4)
Control	6.70 (4.3)	1.50 (4.2)

Discussion

These results suggest that newborns do not discriminate between varied sets of bimoraic and trimoraic Japanese words that share the same number of syllables. This indicates that not all rhythmic units are equally salient to newborns: While words can be discriminated on the basis of the number of syllables (or possibly on some correlated unit such as the vocalic nucleus), they cannot be discriminated on the number of morae when these do not correspond to syllables.

One difference between Experiment 1 and Experiment 2 could potentially account for the absence of discrimination between bi- and trimoraic words. In Experiment 1, the mean durations of the bi- and trisyllabic words differed by 248 ms and 173 ms for the ascending and descending contours respectively, while the corresponding differences between the bi- and trimoraic words were only 146ms and 132 ms. Moreover, the duration distributions only slightly overlapped in Experiment 1 (0% and 12% for the ascending and descending contours respectively), while the corresponding overlaps were greater in Experiment 2 (17% and 68%). However, Bijeljac-Babic et al. (1993) have shown that discrimination of bisyllabic versus trisyllabic utterances could be obtained even when the mean duration difference was less than 120 ms and the overlap was about 75%. Moreover, the absence of effect of contour on the experimental groups in both experiments further suggests that mean duration difference and degree of overlap do not explain the discrimination in Experiment 1 and the absence of discrimination in Experiment 2.

Rather, the present results can be taken as an indication of the limits of syllable representation by infants, at least for bisyllabic utterances. Trimoraic words have more complex syllabic structures than bimoraic words. However, this difference in internal complexity does not seem to be perceived by newborns, suggesting that, at least in a context of phonetically varied bisyllabic words, infants can only extract global syllabic information, presumably based on the presence of vocalic nuclei (see also, Bertoncini et al. (1988) and Jusczyk, Jusczyk, Kennedy, Schomberg, & Koenig (1995) for arguments in favor of global syllable representation).

This suggestion ties in with the results of one of Bijeljac-Babic et al.'s (1993) experiments, which showed that infants were unable to discriminate two sets of bisyllabic French pseudowords that differed on their number of phonemes (four phonemes: CVCV, VCCV, VCVC versus six phonemes: CCVCCV, CVCCVC, CVCCCV). In other words, infants do not seem to be sensitive to a difference between bisyllabic words composed of simple CV or VC syllables and words composed of more complex syllables as long as all the words share the same number of syllabic components (or the same number of vocalic nuclei). In the present study, most of the trimoraic words (25 out of 30) used in Experiment 2 involved an additional consonantic mora, either a nasal or a geminate consonant. Although the role of phonemes and morae might not be equivalent with respect to the rhythmic structure of the language and presumably with respect to the way in which they are processed, the absence of discrimination in both cases yields information on the nature of syllabic representation.

One possibility is that sub-syllabic segments are represented as an integral part of the syllable, and not as extra elements. Thus the variation in syllabic composition may have been less salient than the property common to all the words, that is, identical number of syllables. Perhaps the procedure was not sensitive enough to show that simple and complex syllables were differently represented in a context of bisyllabic words which included a variety of bimoraic syllables (initial CVV and CVQ, initial or final CVN). This question remains open and could be addressed in future experiments contrasting Japanese monosyllabic words such as simple CVs with CVNs or CVVs.

Another possibility is that infants establish similarly simplified representations of syllables regardless of whether they are simple CVs, or more complex syllables, including bimoraic ones. In this view, multisyllabic utterances would be perceived in terms of syllable-like units, based simply on the presence of a vocalic nucleus, and regardless of finer information about the structure of the syllable.

In any case, sub-syllabic morae do not appear to be as salient as syllabic morae. When embedded in bisyllabic words, bimoraic syllables, despite the fact that they are longer and more complex than monomoraic syllables, are not represented by infants as consisting of two units. In other words, infants do not seem to represent trimoraic words as containing one more unit than bimoraic words.

GENERAL DISCUSSION

The principal result of the present study indicates that infants seem to represent and to discriminate phonetically varied multisyllabic words on the basis of the number of syllabic components rather than on the basis of the number of moraic units. Thus, this study corroborates the evidence that French infants represent speech in terms of syllables, or of a property correlated with number of syllables, for example, number of vowels or number of energy peaks. The present results confirm those previously obtained by Bijeljac et al. (1993) by showing that high variability in phonetic composition does not impede representation and discrimination of words that differ in the number of syllables, at least when syllables are constituted of Vs or CVs. However, the rhythmic unit used by Japanese speakers, the mora, does not appear to be similarly salient to newborns. When sub-syllabic morae are used in order to dissociate the number of syllables and the number of morae in the words, infants behave as if they were only reacting to the number of syllabic components.

In the present study, our aim was to investigate whether infants possess universal sensitivity to rhythmic structures of words from any natural language. Recent data indicate that speech segmentation by adult listeners is largely determined by the rhythmic properties of their native language (Cutler et al., 1986; Otake et al., 1993; Sebastiàn-Gallès, Dupoux, Segui & Mehler, 1992). One hypothesis would be that these language-specific strategies have been selected from the set of universal abilities present at birth. However, our results do not support such a hypothesis. Instead, our findings are compatible with the hypothesis that neonates, when presented with sets of varied stimuli, form global representations of syllables (Bertoncini et al., 1988), in which the vowels are the most prominent element, whatever language is heard. In this view, infants first use some kind of universal underspecified unit of representation, which seems to be syllable-like, and later refine their primary representations according to linguistic experience. It is thus possible that more experience with the Japanese language is required for infants to be able to represent consonantic morae as equivalent to syllabic morae. Such a possibility would be supported by data showing that Japanese neonates first behave like French neonates, but rapidly develop representations of complex utterances in terms of morae, whether syllabic or nonsyllabic.

We acknowledge that this hypothesis is based on the assumption, of which we are in favor, that all infants initially process speech in the same universal way. However, we cannot exclude the possibility that it is not the case. It could be that Japanese neonates would react to a change in number of morae, whether syllabic or sub-syllabic, as a possible consequence of prenatal exposure to this distinctive rhythmic information. Such a finding would have extremely important implications on both what would be considered as initial universal abilities and what would be the relevant experience that tunes these abilities to language-specific processing. In any case, the answer is empirical and will require similar experiments with neonates from different linguistic backgrounds.

If adult listeners rely on different units of segmentation, it is important to understand how infants converge onto the one appropriate for their native language. Two accounts of this process of convergence are worth considering in their relation to the process of lexical acquisition. According to the first account, infants begin by building the lexicon, and only afterwards do they discover, on the basis of their primary representations of words, the regularities of their native language, that is, that French is a syllable-timed language or Japanese a mora-timed one. According to the second, or pre-lexical, account, neonates start to represent the speech signal on the basis of salient acoustic information, principally located in vowels (Cutler & Mehler, 1993). Rhythmic information, related to consistent variations in vowel duration and intensity, would be sufficiently salient to characterize the type of language to which utterances belong (Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996). In this view, the rhythmic structure of a language would be acquired at a very early age, much before infants start to build a lexicon. Thus, hearing French utterances will bring infants to process speech in terms of syllables, while hearing Japanese utterances will induce infants to process speech in terms of morae.

More data are obviously needed to improve our comprehension of the time course of lexical acquisition and its relationship to the development of phonological processing (Jusczyk, 1993). The experiments reported here provide further evidence that neonates extract only the most salient properties of the sound pattern of complex utterances. This rather simple ability, driven by robust acoustic information, might allow infants to develop a particular sensitivity to the sound pattern of their native language very early on, before any lexical learning as such has begun. If the acquisition of language-specific skills is viewed as a process by which universal abilities that allow infants to acquire any language progressively converge onto abilities specialized in the processing of one particular language (Mehler & Christophe, 1995), then this study may contribute to a better characterization of the initial state for phonological processing.

Received June 26, 1995; revised manuscript received December 18, 1995; accepted December 22, 1995.

REFERENCES

AMIEL-TISON, C. (1977). Standardizing the physical examination during the first year. In L. Gluck (Ed.), *Intrauterine asphyxia and the developing fetal brain* (pp. 473 – 485). Chicago: Year Book Medical.

BERTONCINI, J., BIJELJAC-BABIC, R., JUSCZYK, P. W., KENNEDY, L., & MEHLER, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, **117**, 21 – 33.

BIJELJAC-BABIC, R., BERTONCINI, J., & MEHLER, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology*, **29**, 711 – 721.

CUTLER, A., & BUTTERFIELD, S. (1990). Durational cues to word boundaries in clear speech. *Speech Communication*, **9**, 485 – 495.

CUTLER, A., & BUTTERFIELD, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, **31**, 218 – 236.

CUTLER, A., & MEHLER, J. (1993). The periodicity bias. *Journal of Phonetics*, **21**, 103 – 108.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25**, 385 – 400.

CUTLER, A., & NORRIS, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 113 – 121.

CUTLER, A., & OTAKE, T. (1994). Mora or phoneme? Further evidence for language specific listening. *Journal of Memory and Language*, **33**, 824 – 844.

HOEQUIST, C. (1983). Syllable duration in stress-, syllable-, and mora-timed languages. *Phonetica*, **40**, 203 – 237.

JUSCZYK, P. W. (1993). From general to language-specific capacities: The WRAPSA Model of how speech perception develops. *Journal of Phonetics*, **21**, 3 – 28.

JUSCZYK, P. W., & DERRAH, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology*, **23**, 648 – 654.

JUSCZYK, P. W., JUSCZYK, A. M., KENNEDY, L. J., SCHOMBERG, T., & KOENIG, N. (1995). Young infants' retention of information about bisyllabic utterances. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 822 – 836.

MEHLER, J., & CHRISTOPHE, A. (1995). Maturation and learning of language in the first year of life. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 943 – 954). Cambridge, MA: MIT Press.

MEHLER, J., DOMMERGUES, J. Y., FRAUENFELDER, U., & SEGUI, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, **20**, 298 – 305.

MEHLER, J., DUPOUX, E., NAZZI, T., & DEHAENE-LAMBERTZ, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax : Bootstrapping from speech to grammar in early acquisition* (pp.101 – 116). Mahwah, NJ: Lawrence Erlbaum Associates.

MEHLER, J., DUPOUX, E., & SEGUI, J. (1991). Constraining models of lexical access: The onset of word recognition. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 236 – 262). Cambridge, MA: Bradford Books.

OTAKE, T., HATANO, G., CUTLER, A., & MEHLER, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, **32**, 258 – 278.

PORT, R. F., DALBY, J., & O'DELL, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, **81**, 1574 – 1585.

SEBASTIAN-GALLES, N., DUPOUX, E., SEGUI, J., & MEHLER, J. (1992). Contrasting syllabic effects in Catalan and Spanish: The role of stress. *Journal of Memory and Language*, **31**, 18 – 32.

SHIBATANI, M. (1990). *The languages of Japan*. Cambridge, U.K.: Cambridge University Press.

APPENDIX 1

Stimuli used in Experiment 1 Classified by Intonational Contour.

1. Words with Ascending Contours

Bisyllabic words	*Duration (ms)*	*Trisyllabic words*	*Duration (ms)*
iga	403	hekiga	707
toʃi	466	togutʃi	731
eki	438	ekaki	736
kago	468	kagome	727
ʃime	580	ʃimetsu	908
taka	458	mitaka	807
tema	370	temari	640
maki	523	mateki	755
miwa	473	miwaku	765
ike	467	ikeda	606
iʒi	415	ikoʒi	683
kabi	422	okabu	731
keʃi	435	kokeʃi	673
saka	536	sakago	751
muʃi	547	muʃizu	793
muda	496	amida	646
sjaka	547	osjaka	707

aʃi	436	ataʃi	712
itʃi	491	itatʃi	715
oto	473	osato	673
kate	459	karite	704
seki	571	koseki	733
toki	489	togaki	726
hake	443	hakike	704
hiʃi	450	higaʃi	744
gaka	554	agaki	701
nagi	472	kodaʃi	747
deki	538	osaki	753
aka	416	kokage	733

2. Words with Descending Contours

Bisyllabic words	*Duration (ms)*	*Trisyllabic words*	*Duration (ms)*
iʃi	413	inotʃi	595
ketʃi	421	ketsugi	548
ʃiro	498	hirosa	613
nika	492	konika	623
nega	461	megami	689
moʃi	470	monaka	669
rika	456	erika	677
kaʒi	383	kaʒino	543
koʒi	391	ikuʒi	612
tomi	351	tomita	609
iki	420	hibatʃi	657
kakjo	487	kakjoku	707
koʃi	442	kokuʃo	664
satʃi	553	asahi	652
noto	544	namida	629
hika	458	ehime	536
maki	534	meguro	584
baku	562	bakuhu	789
buʃi	463	nobuʃi	607
kado	361	kanoko	587
kuri	368	ikura	577
kotʃi	490	ikuʒi	614

ʃiru	515	kuʃiro	605
uka	380	otaru	572
suʒi	537	nasubi	678
doki	454	noʒuku	762
kobu	358	joduri	671
mika	522	wasabi	627
hato	465	ibuse	575

APPENDIX 2

Stimuli used in Experiment 2 Classified by Intonational Contour and Additional Mora.

1. Words with Ascending Contours

Bimoraic words	*Duration (ms)*	*Trimoraic words*	*Duration (ms)*
a. MEDIAL NASAL			
iga	363	iNga	495
eki	379	eNki	546
kago	363	kaNgo	492
ʃime	497	ʃiMme	633
taka	369	taNka	503
tema	358	teMma	503
nagi	422	naNgi	566
maki	432	maNki	510
b. FINAL NASAL			
ike	371	ikeN	490
keʃi	380	keʃiN	559
saka	439	sakaN	610
toʃi	383	toʃiN	497
muda	375	mudaN	490
c. GEMINATE CONSONANT			
aka	362	aQka	489
aʃi	379	aQʃi	493
itʃi	377	iQtʃi	506
oto	367	oQto	500

kate	380	kaQte	554
seki	444	seQki	569
toki	365	toQki	501
hiʃi	435	hiQʃi	573
gaka	359	gaQka	501
daʃi	370	daQʃi	506
d. LONG FIRST VOWEL			
oki	371	ooki	576
kori	319	koori	494
tori	304	toori	448
juki	378	juuki	564
juri	314	juuri	491

2. Words with Descending Contours

Bimoraic words	*Duration (ms)*	*Trimoraic words*	*Duration (ms)*
a. MEDIAL NASAL			
iʃi	384	iNʃi	552
uka	354	uNka	492
ketʃi	438	keNtʃi	569
ʃiro	500	ʃiNro	673
suʒi	501	suNʒi	682
nika	428	niNka	563
nega	428	neNga	511
moʃi	380	moNʃi	571
rika	418	riNka	494
doki	375	doNki	576
b. FINAL NASAL			
kobu	361	kobuN	488
kaʒi	370	kaʒiN	507
koʒi	359	koʒiN	499
tomi	358	tomiN	492
mika	418	mikaN	433

c. GEMINATE CONSONANT

iki	383	iQki	499
kakjo	382	kaQkjo	501
koʃi	378	koQʃi	554
satʃi	550	saQtʃi	575
noto	373	noQto	555
hato	420	haQto	488
hika	434	hiQka	547
maki	417	maQki	548
baku	365	baQku	552
buʃi	364	buQʃi	393

d. LONG FIRST VOWEL

kado	378	kaado	505
kotʃi	439	kootʃi	574
ʃiru	448	ʃiiru	622
nasu	434	naasu	616
buke	319	buuke	481