Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Five-month-old infants' discrimination of unfamiliar languages does not accord with "rhythm class"

Laurence White [1], *Claire Delle Luche* [2] *& Caroline Floccia* [1]

[1] School of Psychology, Plymouth University, UK
[2] Centre for Research in Language Development throughout the Lifespan, Department of Language and Linguistics, University of Essex, UK
laurence.white@plymouth.ac.uk

Abstract

Young infants are sensitive to prosody, which they use to distinguish between speakers and between languages. Indeed, patterns of language discrimination have been interpreted as supporting a "rhythm class" typology, where the rhythmic nature of earliest language exposure determines the primary perceptual units of speech segmentation. In previous studies, five-month-old infants discriminated languages from different putative rhythm classes, but only within a class where one or both languages are familiar. Thus, English infants did not distinguish "syllable-timed" Italian and Spanish, nor "stress-timed" Dutch and German, but Spanish infants distinguished "syllable-timed" Catalan and Spanish. In three head-turn preference experiments, we tested whether English five-month-olds could discriminate – pairwise – between French, Spanish and Finnish. Although all have been categorized as "syllable-timed", they differ in their realization and distribution of strong syllables. In each experiment, we familiarized infants to one of two languages and then exposed them to new speakers from the same or the other language. With French and Spanish, infants looked longer to the new language than the familiar one, indicating discrimination. They failed, however, to discriminate French/Finnish or Spanish/Finnish. These results strongly suggest that – rather than sensitivity to discrete classes – discrimination reflects exploitation of gradient prosodic differences.

Index Terms: rhythm class, language development, language discrimination, speech perception.

1. Introduction

Young infants pay particular attention to the prosody of spoken utterances. Prosodic variation – notably in the hyper-articulated form found in infant-directed speech – may serve to promote infants' engagement and interaction with the conversational world around them [1]. Furthermore, prosodic marking of prominences and boundaries is believed to bootstrap word segmentation and hence the acquisition of vocabulary and syntax [2].

Infant perceptual orientation to variation in intonation and timing naturally entails sensitivity to prosodic differences between speakers, between accents and between languages. Indeed, infant listeners' patterns of discrimination are taken as support for a division of the world's languages into a small number of "rhythm classes" [3]. This rhythmic typology was initially an elaboration of metaphorical descriptions of the "machine-gun rhythm" of Spanish contrasted with the "Morse-code" rhythm of English [4]. The first formal description of rhythm class was grounded in claims that the temporal flow of speech is organized according to isochronous principles [5]. The fundamental constituent for temporal organization was held to differ between "syllable-timed" languages like French, Spanish and Italian – claimed to have isochronous syllables – and "stress-timed" languages like English, Dutch and German, with isochronous stress-delimited feet. Significantly, it was further claimed that all languages manifest one of these two rhythmic types [6]. Later work, noting the importance of moraic structure in languages such as Japanese, postulated a third "mora-timed" rhythm class [7]. The existence of isochronous timing units – the fundamental motivation for the rhythm class hypothesis – was subsequently disproved [8], but invocation of a three-way categorical distinction remains remarkably persistent throughout the speech science literature [e.g., 9] and into the domain of second language pedagogy [e.g., 10]. Furthermore, this typology has fundamental implications for theories of speech perception: the rhythm class of an infant's dominant language is held to determine the nature of the primary units – morae, syllables, stress-delimited feet – into which the speech stream is segmented [11,12,13,14].

Given the manifest lack of phonetic support for categorical rhythm classes, it has been argued [13,15] that the strongest evidence for the ternary typological distinction comes from perceptual studies, in particular, patterns of language discrimination by infants and adults. Most notably, young infants have been shown to discriminate between languages from different putative rhythm classes, but not between languages within classes, except in the case that one or both languages is familiar to the infant. Thus, for example, presented with low-pass filtered speech – used to retain prosodic characteristics whilst removing most segmental information – French newborn infants distinguished English from Japanese, but not English from Dutch [3]. Furthermore, in the same series of experiments, French newborns discriminated sets of English and Dutch *vs* Spanish and Italian utterances, but did not show discrimination when English and Italian utterances were contrasted with Spanish and Dutch [3].

With experience, infants are able to discriminate two languages within a rhythm class if one is familiar to them: for example, Spanish- and Catalan-learning four-month-olds distinguished these two languages [16]. However, where languages are unfamiliar, patterns of discrimination have been claimed to persist in accordance with rhythm class. For example, English five-month-olds infants did not distinguish between "syllable-timed" Italian and Spanish nor between

 doi: 10.21437/SpeechProsody.2016-116

"stress-timed" Dutch and German [17], but did discriminate both English and Italian from "mora-timed" Japanese. Utterances in these experiments were presented intact, rather than low-pass filtered, but despite the availability of segmental as well as prosodic cues, five-month-olds failed to discriminate within putative rhythm classes. Indeed, such evidence, based on paradigms exploiting infant novelty *vs* familiarity preferences, represents the strongest remaining support for the ternary typological rhythm class distinction.

The testing of adult sensitivity to prosodic differences is confounded by mature phonemic awareness, which makes distinguishing languages a more straightforward task. Thus, adult discrimination experiments have exposed listeners to stimuli modified to eliminate segmental information and focus solely on timing (e.g., *sasasa* speech, created by replacing all vowels with /a/ and all consonants with /s/, but retaining the durational values of the original intervals). Initial results suggested that adult discrimination was also governed by categorical rhythm class distinctions [18], but this interpretation was challenged by studies showing within-class and within-language discrimination on the basis of timing cues alone [19,20]. Such results suggest that – rather than discrete categories – adult language discrimination relies on gradient variation along prosodic dimensions, with differences in speech rate and utterance-final lengthening shown to be particularly salient [19, see also 21 for infant discrimination between accents of a familiar language]. This raises the possibility that infant language discrimination is also based on degree of similarity along prosodic and/or segmental dimensions. Notably, the key infant evidence for rhythm class is negative, based on failures to discriminate within a putative class, which could be due to similarities of the particular languages tested rather than to any categorical distinction.

The endurance of the rhythm class concept may lie in part in the subjective phenomenology of linguistic experience, harking back to the earlier observations of salient prosodic distinctions: for example, "The difference in rhythm of 'machine-gun' languages and 'Morse-code' languages is, however, an undeniable fact" [13, p. 1150]; also, "perceptually salient differences include the three-way distinction between […] staccato, rapid rhythm [,…] regular "machine gun" rhythm […] and the "Morse code" alternations of strong and weak syllables" [9, p. 1]. Certainly Castilian Spanish and standard southern British English, for example, sound distinct in their flow, in particular, in the alternation of strong and weak syllables. There are also, however, salient prosodic differences within a "rhythm class". Notably, both French and Spanish have been held, by various authors, to be exemplars of syllable-timing, but they differ strikingly in their prosodic properties. Stress is lexically contrastive in Spanish, being predominantly word-penultimate, but with many minimal pairs differing only in the placement of stress (e.g., ***sa****co* "I remove" vs sa**có** "she removed"). On the other hand, French stress does not serve a lexically contrastive function: word-final syllables have consistent stress, but this is only realized phonetically in phrase-final position [22].

The prosodic heterogeneity of languages described as syllable-timed extends beyond this pair of paradigmatic cases. Finnish, for example, has fixed word-initial stress, and a realization of phrasal accent that depends on the moraic structure of the first two syllables [23]. Unsurprisingly then, the rhythm class status of Finnish – originally classified as syllable-timed – has been disputed [24]. This ambiguous typological position echoes proposals that certain languages have mixed rhythm classes [25]. Given the diversity of stress systems and their phonetic realization, as well as the existence of languages such as Korean that may lack stress altogether [26], an obvious conclusion would be to abandon categorical classes. Certainly, studies of infant perception appear to represent the last bastion of evidence for rhythm class: discrimination of unfamiliar languages in the first six months of life has so far only been seen between, and not within, putative classes.

We used the head-turn preference paradigm to test whether English five-month-old infants could discriminate between French, Finnish and Spanish in three pairwise experiments. We familiarized infants to one language in each pair (French/Spanish; French/Finnish; Finnish/Spanish) and then exposed them to samples of new speakers, either from the same or the other language. In line with previous studies, where infants showed more interest in utterances from the unfamiliar language, we take such behaviour as evidence that they perceive it as different from the familiarized language.

2. Experiment 1: French vs Spanish

2.1. Method

2.1.1. Participants

Twenty-four British English 5-month-olds (14 females and 10 males) from monolingual homes completed the study. They were aged 5 months; 12 days on average (range 4;22 to 6;12). An additional 17 infants were not included, 1 being an outlier (defined by an average looking time above or below 2SD of the global mean), 9 because of inattentiveness, 3 for experimenter or computer error, and 4 who cried.

2.1.2. Stimuli

The French and Spanish stimuli consisted of 8 passages each. Each passage contained 5 sentences repeated once, with 400ms between each sentence. The Spanish ones were taken from a set used in an earlier infant discrimination study [3]; the French sentences were translations of the English sentences from the same study. Sentences were selected so that all passages had comparable mean syllable numbers (French: 18.4 syllables per sentence, Spanish: 18.2), and comparable durations within languages (French: 36.2s per passage; Spanish: 33.0s).

Four native female speakers recorded the sentences for each language. The recordings were distributed so that one infant would hear 2 speakers and 4 passages of one language in the familiarization, followed by 2 new speakers and 4 new passages of that language in the test phase, along with 2 speakers and 4 passages from the unfamiliar language. Recordings were conducted in a soundproof booth with an amplified mono microphone at 41KHz. Utterances were extracted and normalized for amplitude using Praat [27].

2.1.3. Procedure

The experiment took place in a sound-attenuated head-turn preference booth. The infant was seated on lap of his/her caregiver, who wore headphones with masking music so as not to hear the stimuli being played. The experimenter viewed the infant's head remotely via a camera facing the participant. Initially, a green light at eye level flashed to attract the infant's

attention to a midline point. Red lights were mounted just above speakers on each side of the infant, and after sufficient fixation – defined as a voluntary and sustained orientation rather than just a passing glance – the experimenter triggered one red side light, selected randomly by the software. Once the infant's head was oriented toward this light, the experimenter triggered a sound-file that played until the infant looked away or reached the end of the passage. Trials were repeated if looks did not reach a minimum of 1.5 seconds.

Infants were randomly assigned to French or Spanish as the familiarization language (12 for each). Familiarization was complete when all four passages from one language were listened to for at least 20 seconds each. The test phase consists of 8 trials, randomly alternating between 4 passages for each language. The experimenter was unaware of the familiarized language and test trial order.

2.1.4. Statistical analysis

We constructed mixed-effects regression models with random factors of subject and trial, looking time in test phase trials as the dependent variable, and Familiarization Language (French/Spanish) and Status (Same/New) of test trial languages as predictors. Regression models with and without each predictor, and their interaction, were compared with χ^2 tests.

2.2. Results and discussion

Infants familiarized to Spanish looked longer in test trials to French, whilst those familiarized to French looked longer in test trials to Spanish (Figure 1). There was accordingly a main effect of Status, $\chi^2(1) = 9.14$, $p = .0025$, with longer looking times to utterances from the new language than from the same, familiarized language. There was no effect of Familiarization Language, $\chi^2(1) = 0.11$, $p = 0.74$, and no interaction between Status and Familiarization Language, $\chi^2(1) = 0.68$, $p = 0.41$.

Thus, contrary to the rhythm class hypothesis and to previous patterns of language discrimination in five-month-olds, the infants in our study clearly discriminated two unfamiliar languages within the same ("syllable-timed") rhythm class. We develop this in Experiments 2 and 3 to examine whether these two canonically "syllable-timed" languages can each be distinguished from another language, Finnish, whose status within the same class is more debatable.

3. Experiment 2: Finnish vs French

3.1. Method

3.1.1. Participants

Twenty-four British English 5-month-olds (13 females and 11 males) from monolingual homes completed the study. They were aged 5;15 on average (range 4;23 to 6;16). An additional 13 infants were not included, 3 because of inattentiveness, 3 for experimenter or computer error, and 7 who cried.

3.1.2. Stimuli

The Finnish sentences were also translations from the English set in an earlier infant discrimination study [3]. The Finnish utterances had a mean length of 18.7 syllables and the Finnish passages had a mean duration of 33.0 seconds. The French passages were those used in Experiment 1.

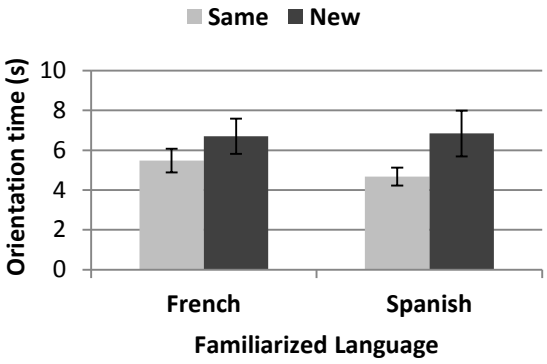

Figure 1. *French vs Spanish: Experiment 1 orientation times for familiarized (same) and new languages.*

3.1.3. Procedure

The familiarization and test trial procedure was the same as in Experiment 1, but in this case, 12 infants each were familiarized to Finnish and French. Because a noise occurred 7 seconds into one Finnish passage and not detected until after the experiments had been run, it was decided to remove all trials when infants listened to that passage for more than 7s (rejecting 2/198 test trials).

3.2. Results and discussion

There was no effect of Status (same/new), $\chi^2(1) = 0.04$, $p = .84$, no effect of Familiarization Language, $\chi^2(1) = 0.57$, $p = .45$, and no interaction between Status and Familiarization Language, $\chi^2(1) = 0.00$, $p = .98$. Thus, as shown in Figure 2, five-month-old infants did not discriminate between French and Finnish, despite, as outlined above, the significant prosodic differences between them. Taken in isolation, this result is congruent with the "rhythm class" account of discrimination, but that explanation is not available for the French *vs* Spanish discrimination found in Experiment 1.

4. Experiment 3: Finnish vs Spanish

4.1. Method

4.1.1. Participants

Twenty-four British English 5-month-olds (10 females and 14 males) from monolingual homes completed the study. They were aged 5;7 on average (range 4;23 to 5;30). An additional 11 infants were not included, 1 being an outlier (see Experiment 1), 4 because of inattentiveness, 3 for experimenter or computer error, and 3 who cried.

4.1.2. Stimuli

The passages were those used in Experiment 1 (Spanish) and Experiment 2 (Finnish).

4.1.3. Procedure

The familiarization and test trial procedure was the same as in Experiment 1, but in this case, 12 infants each were familiarized to Finnish and Spanish. As in Experiment 2, 2/198 test trials were rejected because of a noise 7 seconds into one Finnish passage.

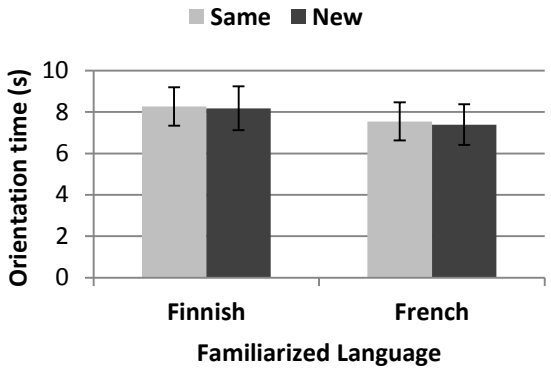

Figure 2. *Finnish vs French: Experiment 2 orientation times for familiarized (same) and new languages.*

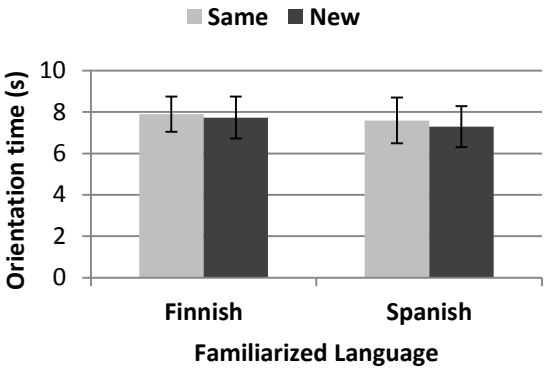

Figure 3. *Finnish vs Spanish: Experiment 3 orientation times for familiarized (same) and new languages.*

4.2. Results and discussion

There was no effect of Status (same/new), $\chi^2(1) = 0.81$, p =.78, no effect of Habituation Language, $\chi^2(1) = 0.07$, p = .79, and no interaction between Status and Habituation Language, χ^2 (1) = 0.01, p = .94. Thus, as shown in Figure 3, five-month-old infants did not discriminate between Finnish and Spanish.

5. General discussion

In three pairwise experiments testing English five-month-olds' discrimination of intact Finnish, French and Spanish speech, we found that infants were able to distinguish French and Spanish, but not Finnish and French, nor Finnish and Spanish. Overall, this selective pattern of discrimination strongly argues that infant perception is not determined by intrinsic sensitivity to discrete rhythm classes, as all three languages are usually characterized as "syllable-timed". Moreover, of the three languages, Finnish is the one whose status within that class has been the most disputed. It is therefore perhaps surprisingly that Finnish utterances were not distinguished from either French or Spanish, results which point to the difficulty of the task for infants, even with the five months' post-natal language experience – in this case, English. This difficulty challenges the argument that infants at this age are too phonologically sophisticated to manifest a pre-existing categorical sensitivity. Indeed, key studies have taken the pattern of discrimination at this age – in particular, failures to discriminate within a particular "rhythm class" – as a strong argument in favour of the rhythm class hypothesis [17].

Previous discrimination studies with adults have suggested that the ability to do the task relies on the perception of functionally important – but gradient – differences in prosodic features [19]. Adult studies using monotone *sasasa* speech have pointed to the primary role of speech rate as a discrimination cue, with other cues including variations in utterance-final lengthening and in vocalic and consonantal interval durations [19]. Infant studies of accent discrimination within a language pointed to the importance of final lengthening cues [21]. The division of utterances into vowels and consonants, as measured by the metric %V, has also been argued to be perceptually salient for young infants, particularly neonates lacking finer phonological categories [13].

Given our current pattern of results, salient cues used by infants should be most different between French and Spanish, with Finnish having intermediate values. We looked at a range of timing features, including speech rate, final lengthening and variation in vowel and consonant intervals. In all cases, values for Finnish were higher or lower than both those for French and Spanish, suggesting that no simple set of timing features can account for the observed behaviour. There is some evidence pointing to the role of *f0* variation in adult discrimination tasks, though secondary to that of speech rate [20]. It should also be noted that intact natural speech was used here, as in previous discrimination experiments with five-month olds [17], and so segmental cues – such as the sharply contrasting vowel inventories of French and Spanish – are also available. Further work in progress, including adult experiments with these and other languages, is intended to determine the salient cues that facilitate the discrimination of French and Spanish, whilst also accounting for the surprising finding that Finnish is not successfully distinguished from the other two.

Prosody is doubtless central to the infant experience of spoken interaction and how such experience supports the development of both phonological awareness and lexical and syntactic acquisition. Given that patterns of infant language discrimination provide the only remaining support for "rhythm class", however, our results reinforce recent phonetic, perceptual and theoretical studies that argue that the notion of rhythm class is redundant in accounts of speech perception and language learning.

6. Acknowledgements

This work was supported by ESRC grant code ES/K010123/1: *Lexical development in bilingual toddlers*.

7. References

[1] Schachner, A., & Hannon, E.E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology*, *47*, 19-25.

[2] Johnson, E.K., & Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory and Language*, *44*, 548-567.

[3] Nazzi, T., Bertoncini, J., & Mehler, J., (1998). Language discrimination by newborns: toward an understanding of the role

of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 756-766.
[4] Lloyd James, A. (1940). *Speech Signals in Telephony*. London: Pitman & Sons.
[5] Pike, K. (1945). *The Intonation of American English.* Ann Arbor: University of Michigan Press.
[6] Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
[7] Ladefoged, P. (1975). *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.
[8] Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51-62.
[9] Peelle, J.E., & Davis, M.H. (2011). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3:320*.
[10] Intuition Languages. (2013). Stress-timed and syllable-timed languages. Retrieved from http://www.intuitionlang. com/blog/stress-timed-and-syllable-timed-languages.
[11] Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*, *92*, 81-104.
[12] 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.
[13] Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs syllable-timed languages. In M. van Oostendorp, C.J. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell Companion to Phonology. Volume 2. Suprasegmental and Prosodic Phonology* (pp. 1147-1157). Hoboken, New Jersey: John Wiley & Sons.
[14] Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 258-278.
[15] Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Probus*, *13*, 247-275.
[16] Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, *65*, 33-69.
[17] Nazzi, T., Jusczyk, P. W., & Johnson, E. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*, 1-19.
[18] Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. In M.J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* (pp. 337–342). Barcelona: Causal Productions.
[19] White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*, 665-679.
[20] Arvaniti, A., & Rodriquez, T. (2013). The role of rhythm class, speaking rate, and F0 in language discrimination. *Laboratory Phonology*, *4*, 7-38.
[21] White, L., Floccia, C., Goslin, J., & Butler, J. (2014). Utterance-final lengthening is predictive of infants' discrimination of English accents. *Language Learning*, *64*, 27-44.
[22] Jun, S. A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, *14*, 147-172.
[23] Suomi, K. (2009). Durational elasticity for accentual purposes in Northern Finnish. *Journal of Phonetics*, *37*, 397-416.
[24] Iivonen, A. (1998). Intonation in Finnish. In Hirst, D., & Di Cristo, A. (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 311-327). Cambridge: Cambridge University Press.
[25] Nespor, Marina. (1990). On the rhythm parameter in phonology. In I. Roca (Ed.), *The Logical Problem of Language Acquisition* (pp. 157-175). Dordrecht: Foris.
[26] Jun, S.A. (1995). A phonetic study of stress in Korean. *Journal of the Acoustical Society of America*, *98*, 2893-2893.
[27] Boersma, Paul & Weenink, David (2015). Praat: doing phonetics by computer [Computer program]. Version 6.0.06, retrieved 29 November 2015 from http://www.praat.org/