Bilinguals' Segmentation of Unfamiliar Speech Rhythm

Poorani Vijayakumar, Laurence Bruggeman, Mark Antoniou

The MARCS Institute for Brain, Behaviour and Development
poorani.v@westernsydney.edu.au, l.bruggeman@westernsydney.edu.au,
m.antoniou@westernsydney.edu.au

Abstract

Listeners rely on their native language rhythm to segment speech. Bilinguals are thought to have certain advantages in linguistic processing, but little is known about their speech segmentation strategies. Here, we investigated bilinguals' segmentation of speech with an unfamiliar rhythm. Monolingual speakers of a stress-timed language (English) and early bilingual speakers of two stress-timed languages (English-Arabic) completed a fragment detection task in a language with an unfamiliar rhythm (Japanese; mora-timed). Preliminary findings showed that the bilinguals did not outperform the monolinguals. Furthermore, neither group applied a moraic segmentation strategy when segmenting Japanese.

Index Terms: language rhythm, speech segmentation, bilingualism

1. Introduction

The regularity of timing units in speech (i.e., the rhythm of a language) tends to influence the way listeners of a language segment speech [1]. In French, for example, words have predictable syllable boundaries, and therefore a regular syllable structure, allowing listeners to use this regularity for speech segmentation [2, 3]. Conversely, English has many possible syllable structures, which makes the syllable an unsuitable unit for segmentation. Rather, English syllables can be classified as strong or weak, and listeners of English may use this regularity to segment speech [4, 5]. In Japanese, the rhythmic structure is based on the mora, which is a unit of syllabic weight. Light syllables consist of one mora, while heavy syllables are bimoraic [1]. The importance of the mora as a rhythmic unit is illustrated in Japanese poetry. A haiku, for instance, has three lines, with five, seven and five morae, respectively, regardless of the number of syllables [1, 6]. Studies on the segmentation of Japanese have found that native listeners of Japanese use a mora-based segmentation strategy for their language [6, 7]. During speech segmentation, listeners leverage whatever rhythm best characterises the phonological regularities of their own native language [1], which makes segmenting the native language very efficient. However, listeners tend to apply their native segmentation strategy to other languages as well, regardless of whether it is an appropriate strategy for that language [8, 9, 10, 11]. This does not have to be a problem, as long as the other language has the same language rhythm as their native language [1, 12]. And sometimes it may even give second-language (L2) listeners an advantage over native listeners, at least in the confines of an experimental task. For instance, Dutch and German L2 listeners of English may use suprasegmental cues of lexical stress in English more accurately than native listeners of that language [9, 10, 11] (but see [13]).

Unfortunately, the speech segmentation literature has thus far mainly focused on monolingual listeners. Consequently, it is as yet unknown how bilinguals, with the knowledge of two languages, might use language rhythm when segmenting speech in familiar and unfamiliar languages. Many bilinguals are flexible language users and switch between their languages seamlessly, while never seeming to compromise on effective communication [14]. Bilinguals tend to develop their languages as required by the environment in which they use them, so that their language knowledge and proficiency are not static and evolve according to necessity and use [15]. During speech comprehension, both of a bilingual's languages are activated and influence one another to varying degrees, so that bilinguals parse speech signals using the linguistic and contextual knowledge of both of their languages [14, 15]. It has been suggested that, perhaps due to this flexibility, bilinguals may possess advantages over monolinguals in several aspects of linguistic processing. Examples of such advantages include metalinguistic awareness [16], phonetic learning [17] and novel word learning [18]. Nevertheless, bilingual advantages do not extend to every aspect of linguistic processing. For example, monolinguals often outperform bilinguals in tasks of speech production, such as naming tasks (e.g., where participants are asked to name drawings), and fluency tasks (e.g., name as many animals starting with the letter F within a limited time) [19].

These mixed findings raise the question whether the 'bilingual advantage' may extend to speech segmentation as well. One study examined a group of balanced early French-English bilinguals, using a French fragment detection task and an English word spotting task [20]. Based on their answer to the question "Suppose you developed a serious disease, and your life could only be saved by a brain operation which would unfortunately have the side effect of removing one of your languages Which language would you choose to keep?" [20, p. 390], the bilinguals were divided into two language dominance groups. Participants who were thus categorised as French-dominant appropriately used a syllable-based segmentation strategy for French, which is syllable timed[1]. However, they also, inappropriately, used this same strategy for stress-timed English. Conversely, the English-dominant listeners used a stress-based segmentation strategy for both languages.

---

[1] The importance of the rhythmic properties of a language for the strategies its listeners use during segmentation initially led to the assumption that all languages could be categorised into one of three rhythmic classes. French was thus seen as a syllable-timed language, English as stress-timed, while Japanese was classed as a mora-timed language. Although it is now generally agreed upon that this classification oversimplifies a much more complex reality and is therefore considered problematic [e.g., [28]], we opted to use these categorical terms in the present paper in the interest of brevity.

Although the bilinguals were fluent in both of their languages, they appeared unable to switch between their language rhythms and thus resorted to using only one over the other.

In this study, we aim to investigate if being bilingual benefits speech segmentation in an unknown language with an unfamiliar rhythm. If bilinguals have a ‘bilingual advantage’, we would expect that they might segment speech in an unfamiliar language with an unfamiliar rhythm more efficiently than monolinguals, perhaps even using language-appropriate segmentation strategies. Monolingual speakers of a stress-timed language (English) and early bilingual speakers of two stress-timed languages (English-Arabic) completed a fragment detection task in a language with an unfamiliar rhythm (Japanese; mora-timed). If bilinguals indeed have an advantage over monolinguals, we hypothesised that the bilinguals would detect the target fragments quicker and more accurately than the monolinguals. If the bilinguals use a language-appropriate mora-based segmentation strategy, we also predicted that they would detect the target fragments that were aligned with mora boundaries faster and more accurately than those that were unaligned (see section 2.2 for more details regarding fragment alignment), suggesting a language-appropriate mora-based segmentation strategy. The monolingual listeners were not predicted to respond differently to aligned and unaligned target fragments.

2. Method

2.1 Participants

Participants were 36 Australian English monolinguals (34 female, M_{age} = 26.0, *SD* = 9.3) and 13 English-Arabic bilinguals (10 female, M_{age} = 21.9, *SD* = 7.7), recruited from the participant pool of undergraduate psychology students at Western Sydney University. To assess participants’ language experience and proficiency, we used the Language Experience and Proficiency Questionnaire (LEAP-Q) [21]. The bilingual participants acquired English (M_{age} = 3.0; *SD* = 2.9) and Arabic (M_{age} = 3.8; *SD* = 4.0) in early childhood. All but one of the bilinguals reported English as their dominant language. The bilinguals’ mean self-reported proficiency was 9.7 out of 10 (*SD* = 0.6) in English and 6.8 out of 10 (*SD* = 1.5) in Arabic. These self-report ratings were confirmed by participants’ scores on the Lexical Test for Advanced Learners of English (LexTALE) [22] and the LexArabic [23], with mean scores of 80.8% (*SD* = 6.8) and 65.4% (SD = 12.5) respectively. No participants reported any problems with their speech or hearing, nor any uncorrected vision problems. All participants gave informed consent before completing the experiment. Upon completion of the experiment, participants were reimbursed with course credits.

2.2 Stimuli

Auditory stimuli were taken directly from Experiment 3 of [6] and included eight pairs of meaningful Japanese target words: *tanishi-tanshi; monaka-monka; kanoko-kanko; sanaka-sanka; nanoka-nanka; kinori-kinri; haneda-handa; shinigao-shingao.* Words in each pair had identical initial and final morae and differed only in their medial mora. For example, *tanishi* and *tanshi* both start with the mora *ta* and end with the mora *shi*. The medial mora is *ni* in *tanishi* and *n* in *tanshi*. Pitch accent patterns were matched across the words within in each pair, so that both words were either accented (fall from a high pitch to a low pitch) or unaccented (no fall in pitch). Each target word was combined with two to five additional words to form a word sequence, for a total of 32 word sequences. Each word sequence only consisted of one target word and the rest were filler words. Sequence length varied from three to six words, with the target word always appearing in the penultimate position (e.g., fourth in a sequence with five words). The target fragments that participants were asked to detect always formed part of both target words in a pair, with the mora boundaries of the target fragment aligning with those of only one of the words in the pair. For instance, the fragment *tan* (/ta/ /n/; forward slash indicates a mora boundary) aligns with the morae of the target word *tanshi* (/ta/ /n/ /shi/) and is unaligned with the morae of the target word tanishi (/ta/ /ni/ /shi/). This manipulation was included in the experiment because previous findings have shown that listeners who use mora-based segmentation detect aligned fragments faster and more accurately than fragments that are not aligned [12, 6].

Apart from the 32 word sequences that contained a target word, there were another 32 trials that were catch trials. These did not contain any target fragments and participants were expected to abstain from pressing a button on these trials. The trials were presented in two blocks of 32 word sequences (16 with embedded target words and 16 catch trials with no embedded target words) each. Target words were divided so that one word in each pair appeared in the first block, the other in the second block.

2.3 Procedure

Participants completed all tasks online, using PsychoPy [24] and the Pavlovia platform. The demographics and language background questionnaire was administered via Qualtrics. Participants were instructed to wear headphones throughout the task. Of the monolinguals, 21 reported using in-ear headphones, and 15 wore over-ear headphones while eight of the bilingual participants reported using in-ear headphones, and five wore over-ear headphones. The experiment began with a practice phase to familiarise participants with the task. The participants then completed the fragment detection task. At the start of each trial, participants first saw a fixation point and after 1s were presented with the audio of a target fragment they had to detect (e.g., *tan*). This was then followed by a 3s silence, and then the start of the word sequence (e.g., “*nazo, ekubo, kengaku,* ***tan****ishi, mamushi*”). Each word in a sequence started 1s after the offset of the previous word. Participants were instructed to press the spacebar as soon as they heard the target fragment within the presented word sequence. Button presses did not terminate trials, and word sequences were played in full for each trial. The following trial then started 3s after the end of the previous trial.

3. Results

Response times (RTs) were measured from the onset of the target fragment. If a button press was recorded after the entire word sequence had finished, it was counted as a miss. Miss rates were calculated by obtaining the percentage of missed responses for each participant, overall and for both the different target alignment conditions. We removed as outliers all trials that had response times faster than 200 ms.

To account for the trade-off between response speed and accuracy (fast button responses can lead to lower accuracy, while high accuracy may come with slower responses) [25], we chose to combine response time and accuracy into a single dependent variable: the inverse efficiency score (IES; [26]). The IES takes into account a participant's RTs as well as their accuracy, and can thus be seen as the RT corrected for the number of errors committed. IES was calculated using the formula IES = (RT/(1-PE)), where RT is the mean response time of correct responses, and PE is the miss rate. For each participant, we calculated one overall IES (based on all experimental trials in the experiment), as well as one IES each for both alignment conditions (i.e., one IES for the trials in which the target fragments aligned with the mora boundaries of the target word, and one IES for the trials with unaligned target fragments). Raw miss rates, raw mean response times, and mean IES are presented in Table 1.

Table 1: *Raw miss rates and raw mean response times for the monolinguals and bilinguals, by condition and overall*

		Condition		**Overall**
		Aligned	Unaligned	
Monolinguals	Miss rates (%)	20.8	12.8	18.8
	RT (ms)	786	1028	783
	IES (ms)	1038	924	997
Bilinguals	Miss rates %	29.5	30.8	29.8
	RT (ms)	981	777	953
	IES (ms)	1123	1595	1567

Figure 1 shows the overall mean IES for the bilinguals and the monolinguals. A comparison between both participant groups using a non-parametric Mann-Whitney U-test in R [27] showed that there were no significant differences between the IES of the monolinguals and bilinguals ($W = 181$, $p = .235$).

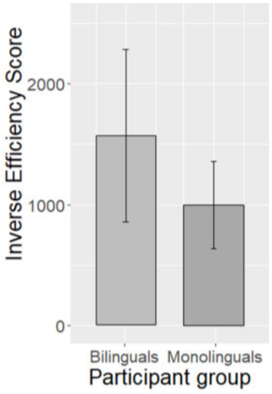

Figure 1: *Bilinguals' (left) and monolinguals' (right) overall IES). Error bars depict standard deviation of the means.*

As mentioned, faster and/or more accurate detection of aligned than unaligned fragments indicates moraic segmentation. Thus, to assess whether participants had used a mora-based segmentation strategy, we further compared the IES for target fragments that aligned with a mora boundary to the IES for targets that did not. Figure 2 shows the IES for monolinguals (left panel) and bilinguals (right panel). A paired-samples Wilcoxon test was used to analyse the IES. Monolinguals showed significantly faster IES ($V = 533$, $p = .001$) for unaligned fragments than aligned fragments. The bilinguals did not show any difference in IES between the types of fragment alignment ($V = 62$, $p = .273$).

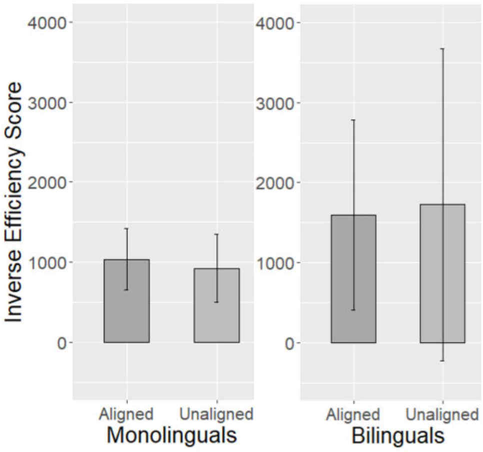

Figure 2: *Monolinguals'(left) and bilinguals' (right) IES by fragment alignment. Error bars depict standard deviation of the mean.*

4. Discussion

The present study investigated how monolingual and bilingual participants segment speech in a language that is unfamiliar to them. Firstly, we had predicted that English-Arabic bilinguals would segment language with an unfamiliar rhythm better than English monolinguals. This hypothesis was not borne out, as bilingual participants were not quicker nor more accurate when detecting the target fragments than the monolinguals. Even though the bilingual listeners possess knowledge of two stress-timed languages (English and Arabic) and the monolinguals know just one stress-timed language (English), this does not seem to provide them with a 'bilingual advantage' when detecting fragments in Japanese, with its unfamiliar mora-timed rhythm.

Our second hypothesis – that bilinguals would use mora-based segmentation and detect targets that align with mora boundaries better than those that do not – was also not supported by our results. This suggests that bilinguals may not have an advantage over monolinguals when it comes to speech segmentation. Our results are in line with previous findings from French-English bilinguals [20] completing similar target fragment detection tasks as the one used here. These bilinguals did not employ an appropriate segmentation strategy in both of their languages. Instead, they relied on only one rhythmic segmentation strategy (i.e., that of their dominant language) for both languages [20]. In contrast to that study, the bilinguals tested here completed the task in a language they did not speak at all. They did not have any knowledge of the appropriate moraic segmentation strategy, and it is therefore perhaps unsurprising that they could not exploit the moraic structure of Japanese during the fragment detection task.

Our results from the English monolinguals in this study showed that they detected unaligned fragments better than aligned fragments. This confirms that the monolingual English listeners did not use a mora-based segmentation strategy, in line with previous studies [e.g., 6]. We had, however, predicted that there would not be any difference in their detection accuracy of these two fragment types, since previous findings with English monolinguals had found no significant differences in the detection speed or accuracy for these types either [6]. On the other hand, other findings have shown that English monolinguals detect fragments faster in words beginning with consonant-vowel-consonant-vowel (CVCV) than in words beginning with consonant-vowel-consonant-consonant (CVCC), regardless of the target fragment itself [3]. This corresponds exactly to the unaligned condition of our study, in which all target words start with CVCV.

This study is a work in progress, and at present may be underpowered due to the small sample of bilingual participants, so it is not possible to draw firm conclusions regarding a potential advantage for bilinguals in speech segmentation. Nevertheless, our preliminary results offer a valuable addition to an understudied research area. Further research is, as always, still needed. The bilinguals in the present study spoke two languages with the same rhythm (stress timing). It would be interesting to see the segmentation strategies used by bilinguals whose languages are from two different rhythmic categories. This could tell us whether knowledge of two different language rhythms may afford bilinguals an advantage when segmenting an unfamiliar language with a rhythm unfamiliar to the bilinguals (e.g., English-Korean bilinguals [stress-timed and syllable-timed, respectively] segmenting mora-timed Japanese). Would this kind of bilingual, unlike the one tested here, employ a language-appropriate segmentation strategy?

In conclusion, this study aimed to investigate if English-Arabic bilinguals detect target word fragments quicker and more accurately in Japanese than English monolinguals. The bilinguals did not segment speech in Japanese, a language unfamiliar to them, better than the monolinguals. Furthermore, neither group applied a mora-based segmentation strategy to Japanese, suggesting that, like the monolinguals, the bilinguals relied on a single segmentation strategy, based on their native language, and do not have an advantage over monolinguals when segmenting speech in an unfamiliar language.

5. References

[1] Cutler, A., "Native listening: Language experience and the recognition of spoken words", The MIT Press, 2012.

[2] Cutler, A., Mehler, J., Norris, D. G. and Seguí, J., “A language-specific comprehension strategy,” Nature., 304:159-160, 1983.

[3] Cutler, A., Mehler, J., Norris, D. G. and Seguí, J., “The syllable’s differing role in the segmentation of French and English.,” Journal of Memory and Language., 25(4):385-400, 1986.

[4] Cutler, A. and Norris, D. G., “The role of strong syllables in segmentation for lexical access,” Journal of Experimental Psychology: Human Perception and Performance.,14(1):113-121, 1988.

[5] Cutler, A. and Butterfield, S., “Rhythmic cues to speech segmentation: Evidence from juncture misperception,” Journal of Memory and Language., 31(2):218-236, 1992.

[6] Otake, T., Hatano, H., Cutler, A. and Mehler, J., “Mora or syllable? Speech segmentation in Japanese,” Journal of Memory and Language., 32:258-278, 1993.

[7] Otake, T., Yoneyama, K., Cutler, A. and van der Lugt, A., “The representation of Japanese moraic nasals,” Journal of the Acoustical Society of America., 100(6): 3831–3842, 1996.

[8] Cutler, A. and Otake, T., “Mora or phoneme? Further evidence for language-specific listening,” Journal of Memory and Language., 33(6):824 – 844, 1994.

[9] Cooper, N., Cutler, A. and Wales, R., “Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners,” Language and Speech., 45:207–228, 2002.

[10] Cutler, A., “Greater sensitivity to prosodic goodness in non-native than in native listeners,” Journal of the Acoustical Society of America., 125(6):3522–3525, 2009.

[11] Yu, J., Mailhammer, R. and Cutler, A., “Vocabulary structure affects word recognition: Evidence from German listeners,” in Proc. Speech Prosody 2020., 2020.

[12] Murty, L., Otake, T. and Cutler, A., “Perceptual tests of rhythmic similarity: I. Mora rhythm,” Language and Speech., 50(1):79-99, 2007.

[13] Bruggeman, L. and Cutler, A., “Listening like a native: Unprofitable procedures need to be discarded,” Bilingualism: Language and Cognition., 26:1093–1102, 2023.

[14] Antoniou, M. “Speech Perception,” in F. Grosjean and K. Byers-Heinlein (Eds.), The Listening Bilingual, Wiley-Blackwell, 43-64, 2018.

[15] Grosjean, F. and Byers-Heinlein, K., “Bilingual Adults and Children: A Short Introduction,” in F. Grosjean and K. Byers-Heinlein (Eds.), The Listening Bilingual, Wiley-Blackwell, 4-24, 2018.

[16] Bialystok, E. “Metalinguistic dimensions of bilingual language proficiency,” in E. Bialystom (Ed.), Language Processing in Bilingual Children, Cambridge, Cambridge University Press, 113-140, 1991.

[17] Antoniou, M., Liang, E., Ettlinger, M. and Wong, P. C. M., “The bilingual advantage in phonetic learning,” Bilingualism: Language and Cognition.,18(4):683–695, 2015.

[18] Kaushanskaya, M. and Marian, V., “The bilingual advantage in novel word learning,” Psychonomic Bulletin & Review.,16:705–710, 2009.

[19] Bialystok, E., Craik, F. and Luk, G. “Cognitive control and lexical access in younger and older bilinguals,” Journal of Experimental Psychology: Learning, Memory, and Cognition., 34(4):859–873, 2008.

[20] Cutler, A., Mehler, J., Norris, D. and Segui, J., “The monolingual nature of speech segmentation by bilinguals,” Cognitive Psychology., 24(3):381–410, 1992.

[21] Kaushanskaya, H., Blumenfeld, K. and Marian, V., “The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later,” Bilingualism: Language and Cognition., 23(5):945-950, 2019.

[22] Lemhöfer, K. and Broersma, M., “Introducing LexTALE: A quick and valid lexical test for advanced learners of English,” Behav Res., 44(2):325-343, 2012.

[23] Alzahrani, A., “LexArabic: A receptive vocabulary size test to estimate Arabic proficiency,” Behavior Research Methods., 56:5529–5556, 2024.

Table of Contents for this manuscript

[24] Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E. and Lindeløv, J. “PsychoPy2: Experiments in behavior made easy,” Behavior Research Methods., 51:195-203, 2019.

[25] Vandierendonck, A. “A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure,” Behav Res, 49:653–673, 2017.

[26] Townsend, J. T. and Ashby, F. G., “Methods of modeling capacity in simple processing systems.,” in J. Castellan and F. Restle (Eds.), Cognitive theory, New York, Lawrence Erlbaum Associates,199-239, 1978.

[27] R Core Team, R: A language and environment, Vienna, Austria: R Foundation for Statistical Computing, 2022.

[28] White, L. and Malisz, Z., “Speech rhythm and timing,” in C. Gussenhoven and A. Chen (Eds.), Oxford Handbook of Language Prosody, 167-182, Oxford University Press, 2020.