

Listeners feel the beat: Entrainment to English and French speech rhythms

Pascale Lidji · Caroline Palmer · Isabelle Peretz · Michele Morningstar

Published online: 13 September 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Can listeners entrain to speech rhythms? Monolingual speakers of English and French and balanced English–French bilinguals tapped along with the beat they perceived in sentences spoken in a stress-timed language, English, and a syllable-timed language, French. All groups of participants tapped more regularly to English than to French utterances. Tapping performance was also influenced by the participants' native language: English-speaking participants and bilinguals tapped more regularly and at higher metrical levels than did French-speaking participants, suggesting that long-term linguistic experience with a stress-timed language can differentiate speakers' entrainment to speech rhythm.

Keywords Speech rhythm · Sensorimotor synchronization · Entrainment · Stress-timed language · Syllable-timed language · Psycholinguistics · Speech perception · Phonology · Embodied cognition

Introduction

When people clap or tap to auditory events, they are often demonstrating entrainment (Large & Jones, 1999; Wilson & Wilson, 2005), a natural tendency to synchronize movements with the temporal regularities of external stimuli. These forms of entrainment entail a perceptual encoding of a stimulus periodicity, which increases the ability to predict the timing of upcoming events. Dynamical systems theories (Large & Jones, 1999; Large & Palmer, 2002) suggest that listeners' auditory–motor synchronization can be modeled with internal oscillators whose periods become aligned with the underlying beat of an auditory stimulus.

Listeners can entrain with rhythmic sequences at different hierarchical levels (Jones & Boltz, 1989). For example, one can find people on a dance floor who move with every musical beat (lower hierarchical level), while others move with every other musical beat (higher hierarchical level), and still others move every four beats (higher hierarchical level). The preferred hierarchical level at which people tap to simple auditory pulses is influenced by age (adults tap at higher hierarchical levels than do children) and by musical training (musicians synchronize their movements with higher hierarchical levels than do nonmusicians; Drake, Jones, & Baruch, 2000). Thus, auditory experience may influence the pattern of entrainment.

Most examples of motor entrainment refer to auditory events that convey a regular beat, such as music. Do people also entrain to stimuli that may not display a regular beat, such as speech (Patel, 2008)? Despite the heavily debated nature of temporal regularity in world languages, there is evidence that speakers exhibit

P. Lidji · C. Palmer (✉) · M. Morningstar
Department of Psychology, McGill University,
1205 Dr Penfield Ave,
QC H3A1B1, Montreal, Quebec, Canada
e-mail: caroline.palmer@mcgill.ca

P. Lidji · I. Peretz
International Laboratory for Brain, Music and Sound Research
(BRAMS), Université de Montréal,
Quebec, Canada

P. Lidji
UNESCOG, Department of Psychology,
Université Libre de Bruxelles,
Brussels, Belgium

entrainment to some extent; speakers can entrain to one another in conversational turn-taking (Wilson & Wilson, 2005), adjust their speech rate with that of another speaker (Cummins, 2009; Jungers, Palmer, & Speer, 2002), and synchronize their speech with regular nonlinguistic auditory cues (Cummins & Port, 1998). We examine here whether listeners can entrain their motor responses to the rhythms present in speech.

The rhythmic structure of language has been categorized in terms of stress-timed and syllable-timed distinctions (Abercrombie, 1967; Pike, 1945). Stress-timed languages, such as English, are thought to contain an alternation of stressed and unstressed syllables, with approximately equal time elapsed between stressed syllables. Syllable-timed languages, such as French, exhibit roughly equal stress on each syllable and, thus, display less systematic alternation of stressed and unstressed syllables. Sensitivity to stress-timed and syllable-timed distinctions appears early in life (Nazzi, Bertoni, & Mehler, 1998) and is thought to facilitate the segmentation of speech into lexical units (Cutler, Mehler, Norris, & Segui, 1986; Mersad, Goyet, & Nazzi, 2010). Although there is little empirical support for a strict dichotomy of these rhythmic classes (Cooper & Eady, 1986; Kelly & Bock, 1988; Roach, 1982), distinctions between stress-timed and syllable-timed languages are supported by other acoustic parameters. These include the durational contrast between successive vocalic intervals (normalized pairwise variability, or nPVI; Grabe & Low, 2002), the proportion of vocalic intervals (Ramus, Nespor, & Mehler, 1999), and the coefficient of variation of intervocalic intervals (White & Mattys, 2007). Stress-timed languages show a lower proportion of vocalic intervals, higher variability in the duration of these vocalic intervals, and higher vocalic nPVI than do syllable-timed languages. The higher vocalic nPVI of stress-timed languages is consistent with an alternation of long and short vowels (Dauer, 1983), which can lead to the perception of an alternation of strong and weak beats that form hierarchical rhythmic structures (Liberman & Prince, 1977). Stress-timed languages should, therefore, be perceived as more rhythmic and, thus, more regular than syllable-timed languages (Cutler, 1991). This difference in perceived regularity may translate into greater entrainment with the rhythms of stress-timed languages; we test this hypothesis here.

We measured listeners' entrainment to speech rhythms with a tapping task, an open-ended nonverbal task commonly used in studies of timekeeping mechanisms. Participants tapped along with the beat they perceived in English and French utterances. If stress-timed languages are perceived as more regular than syllable-timed languages, participants' taps should be more regular for English utterances than for French utterances. The coefficient of

variation of intertap intervals (ITIs; *SD*/mean ITI), a common index of tapping variability (Repp, 2005), was expected to be smaller for tapping to English than for tapping to French utterances. In addition, we examined the influence of linguistic experience on tapping behavior by comparing native monolingual speakers of French, native monolingual speakers of English, and French–English balanced bilinguals. Native language experience can influence the perception of speech rhythm in nonnative languages (Cutler, 2001), and listeners' cultural/linguistic familiarity can influence the hierarchical level at which they tap to song (Drake & Ben El Heni, 2003). Comparison of native and nonnative responses to speech rhythm allowed us to investigate whether experience with a stress-timed language increases perceived regularity and preferred hierarchical levels.

Method

Participants

Twenty-four monolingual native French speakers (mean age = 24 years, range = 19–45), 24 monolingual native English speakers (mean age = 23.6 years, range = 18–43), and 24 French–English balanced bilinguals (mean age = 24 years, range = 18–43) were recruited from the Montreal area. Monolinguals were defined as individuals who had been exposed to only one language at home, at school, and at work, from birth. Balanced bilinguals were defined as individuals who rated themselves as highly proficient in both French and English and who either had been raised in a French–English bilingual environment or were native speakers of one of these languages and had been schooled in the other language. Balanced bilinguals were screened with the Bilingual Dominance Scale (BDS; Dunn & Fox Tree, 2009), adapted to French, which determines whether participants are dominant in one language: A score of 0 would indicate a perfectly balanced bilingual (possible range of the BDS : –30 to +30). The bilingual participants had to score in a narrow range between +6 (French dominant) and –6 (English dominant) to ensure fluency with both languages; the mean score was –0.66 (range = +6 to –5).

Stimuli and equipment

Twelve sentences, each containing 13 monosyllabic words, were created in each language. French and English sentences were matched in terms of syntactic structure and word frequency. Sentences were constructed to present an alternation of content and function words,

permitting the interpretation of a *strong–weak* stress pattern in both languages and leading to nested metrical levels (see [Appendix](#) for a list of the sentences used). Each French sentence was spoken by three female native monolingual speakers of Quebec French, and each English sentence by three female native monolingual speakers of North American English at their normal speech rate (French, mean speech rate = 4.24 syllables/s, with sentence duration = 3.08 s; English, mean speech rate = 4.23 syllables/s, with sentence duration = 3.09 s; $p > .10$). Phoneme boundaries were marked using Praat (Boersma & Weenik, 2007), following rules described by Ramus et al. (1999).

The French and English utterances showed different acoustic properties typical of each language (Grabe & Low, 2002; Ramus et al., 1999), as shown in Table 1. French and English stimuli differed significantly in the proportion of vocalic intervals (%V), computed as the sum of vocalic intervals (onset to offset of vowel durations) divided by the total utterance duration. French and English utterances also differed in standard deviation of vocalic intervals (ΔV), but not in standard deviation of consonantal intervals (ΔC). Consistent with Grabe and Low, the normalized Pairwise Variability Index for Vocalic Intervals (V-nPVI) was significantly higher for English than for French stimuli, suggesting that English utterances exhibited a greater alternation of long and short vowels, as compared with French. We also computed a rate-normalized measure of stimulus rhythmic variability, the coefficient of variation (CV; $SD/mean$) of intervocalic intervals¹ (IVIs; vowel onset to onset, including pauses). As is shown in Table 1, the CV(IVI) tended to be smaller for English than for French utterances. The CV(IVI) computed across every other vowel (every two words) did not differ between the French stimuli ($M = .24$) and English stimuli ($M = .21$). In sum, the French and English utterances presented typical syllable-timed and stress-timed acoustic structures, respectively.

All participants were administered a questionnaire to assess their linguistic and musical experience; in addition, the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) was administered to the bilinguals to confirm their high level of fluency in English and French. Auditory stimuli were presented as audio files via Cubase SX software over AKG K271 studio headphones. Tapping responses

Table 1 Means (with standard errors) of the %V, ΔV , vocalic nPVI, and CV of intervocalic intervals for the French and English utterances

	French Utterances	English Utterances	$t(70)$
%V	51.51 (1.01)	44.58 (0.99)	-4.876***
ΔV	0.04 (0.0013)	0.05 (0.0025)	3.616***
V-nPVI	50.45 (2.82)	62.56 (2.05)	3.335***
CV(IVI)	0.372 (0.018)	0.330 (0.015)	-1.739 *

* $p < .1$

*** $p < .001$

were recorded on a silent Roland RD700 electronic keyboard, with a temporal resolution of 1 ms.

Design and procedure

Participants' spontaneous tapping rate was measured first, to identify any a priori rate preferences. They were asked to tap at a regular and comfortable pace on the (silent) keyboard with the index finger of their dominant hand, for 30 s. This was followed by the speech-tapping task, in which participants were instructed to tap along to the subjective beat they perceived in the speech segments they heard. All participants were presented with both French and English stimuli. Thus, participant group (French, English, bilingual) was a between-subjects variable, and stimulus language was a within-subjects variable. Stimuli were blocked by language; the order of language presentation was counterbalanced among participants.

On each experimental trial, a spoken sentence was presented 3 times. A 350-ms high-pitched tone signaled the beginning of a trial (and thus, of a new sentence), and a low-pitched tone (300 ms) separated each of the three repetitions of the sentence within the trial. Each sentence was followed by a 1-s silence, and each tone was followed by a 700-ms silence.

Participants were instructed to listen to the first presentation of the utterance and to tap along to the spoken stimulus on the second and third repetitions. There was a short break every 12 trials, whose duration was determined by the participant. Experimental trials were presented in a pseudorandom order within each language, with the restrictions that no sentence could appear twice within one subset of 12 trials and that no speaker could appear twice in a row. Thus, tapping data were collected from 72 trials in each language (12 sentences \times 3 speakers \times 2 repetitions).

Between language blocks, all participant groups completed the linguistic and musical background questionnaires. At the

¹ Consistent with previous research (Grabe & Low, 2002; Ramus et al., 1999), we assume that vowels are salient and measurable markers of speech rhythm. Analyses repeated for the CV of intersyllabic intervals showed the same pattern of results.

end of the speech-tapping task, the participants provided a second measure of their spontaneous tapping rate. Finally, they tapped along with a sounded metronome (IOI = 500 ms, for approximately 30 s) to assess their tapping regularity with a regular nonverbal stimulus. To control for potential effects of the language of experimental instructions on subsequent tapping performance, the experimenter (a native balanced bilingual) addressed the bilingual participants in the language in which they next heard stimuli in the tapping task, switching languages in the middle of the experiment. French and English monolinguals were addressed in their native language during the entire experiment.

Data analyses

Tapping regularity was assessed with two measures. First, the coefficient of variation (CV, SD/mean of ITIs) was computed from the tap onsets and indicates the temporal variability adjusted for tapping rate. Second, the hierarchical level at which participants aligned their taps was determined by raters who coded the alignment of taps relative to stimulus syllables (words), a subjective coding that may reflect P-centers (Morton, Marcus, & Frankish, 1976). P-centers are defined as the psychological moment of occurrence of syllables but have no consistent acoustic correlates (Marcus, 1981).

A French–English balanced bilingual coder (M.M.) first categorized each participant's tapping trials as belonging to one of the following three categories: (1) low hierarchical level, tapping with every word (11–13 taps per utterance, aligned with every word); (2) high hierarchical level, tapping with every second word (6–8 taps per utterance, aligned with every other word) or every fourth word (4 taps per utterance, aligned with every fourth word); or (3) other cases, in which none of these patterns was observed consistently across trials. The coder determined the category for 33% of each participant's trials in each language block. Interrater reliability was assessed by two other coders, one native speaker of English (C.P.) and one native speaker of French (P.L.), who rated the same data samples from a subset of 9 participants. The interrater agreement was 94%.

Results

The mean ITI was similar for English-speaking (405 ms), French-speaking (360 ms), and bilingual participants (411 ms), as well as for English stimuli (383 ms) and French stimuli (383 ms), with no interaction, all F s < 1. A two-way repeated measures ANOVA on the temporal variability of tapping to speech, measured by the CV(ITI), indicated main effects of both stimulus language and

participant group, as shown in Fig. 1. All participant groups tapped less variably to English speech (mean CV = 0.27) than to French speech (mean CV = 0.32), $F(1, 69) = 47.17$, $p < .001$. The main effect of participant group indicated that French monolinguals exhibited greater tapping variability than did English monolinguals and French–English bilinguals, $F(2, 69) = 3.202$, $p < .05$. No significant interaction was found between stimulus language and participant group, $F < 1$.

Next, we correlated the stimulus metrics shown in Table 1 with the tapping variability measures for each participant. The only acoustic variable that correlated with participants' tapping variability was the variability of the intervocalic intervals [CV(IVI); $r = .47$, $p < .001$]. An ANOVA on these individual correlations between stimulus and tapping CVs indicated a significant effect of stimulus language, $F(1, 69) = 33.05$, $p < .001$, and an interaction of stimulus language with participant group, $F(2, 69) = 3.75$, $p < .05$. As is shown in Fig. 2, participants' tapping variability was more strongly correlated with stimulus variability for French utterances than for English utterances, but this difference reached significance only for French monolinguals (Tukey post hoc HSD = 4.16, $p < .05$), who were less sensitive to (had lower correlations for) the temporal regularities of the English utterances.

Finally, we examined the hierarchical levels at which the participants tapped. Figure 3 shows the number of participants in each group ($n = 24$) who tapped hierarchically on alternating stressed words (every two or four words, high hierarchical level) and those who tapped with every word (low hierarchical level), which are the predominant tapping patterns (16.6% of participants used an alternate pattern). Participants were highly consistent in the hierarchical level at which they tapped: Fewer than 10% of participants switched hierarchical levels between trials or languages. A chi-squared

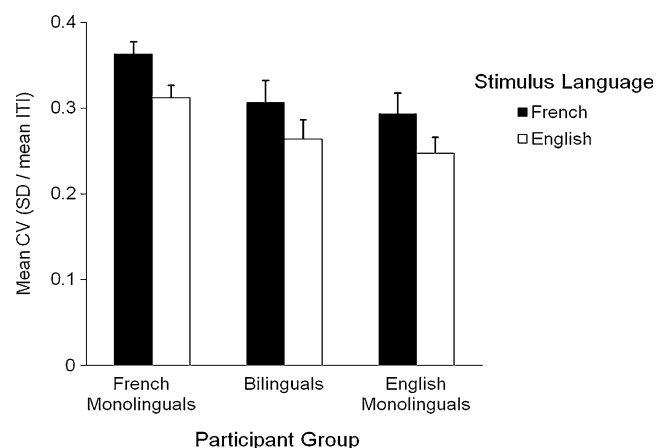


Fig. 1 Mean coefficient of variation (SD/mean) of intertap intervals (ITIs) by participant group and stimulus language. Error bars represent the standard error of the mean

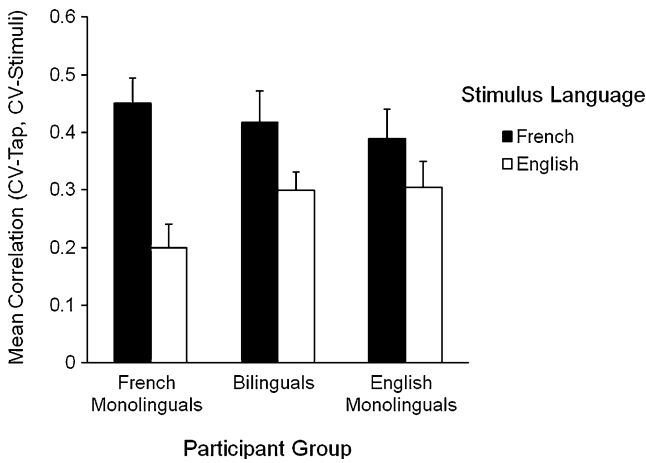


Fig. 2 Mean correlation of coefficients of variation for intertap intervals and stimulus intervocalic interval by participant group and stimulus language. Error bars represent the standard error of the mean

analysis on the number of individuals by stimulus language and participant group indicated a main effect of participant group, $\chi^2(4, 72) = 18.03, p < .01$; French participants were more likely to tap at low hierarchical levels (every word), whereas English participants were more likely to tap at higher hierarchical levels. Bilinguals were equally distributed in both categories. There was no effect of stimulus language and no interaction, $p > .10$; listeners within each participant group responded the same way to both languages.

Furthermore, the participant groups did not differ a priori in their metronomic tapping [mean ITI = 499 ms for each group, $F < 1$; French monolinguals, $CV(ITI) = .047$, English monolinguals, $CV(ITI) = .057$; bilinguals, $CV(ITI) = .047$; $F(2, 69) = 1.94, p > .10$]. They also did not differ in their spontaneous tapping rate (French monolinguals, mean ITI = 766 ms; English monolinguals, mean ITI = 695 ms; bilinguals, mean ITI = 760 ms, $F < 1$) or their spontaneous tapping variability [French monolinguals, $CV(ITI) = .077$; English monolinguals, $CV(ITI) = .079$; bilinguals, $CV(ITI) =$

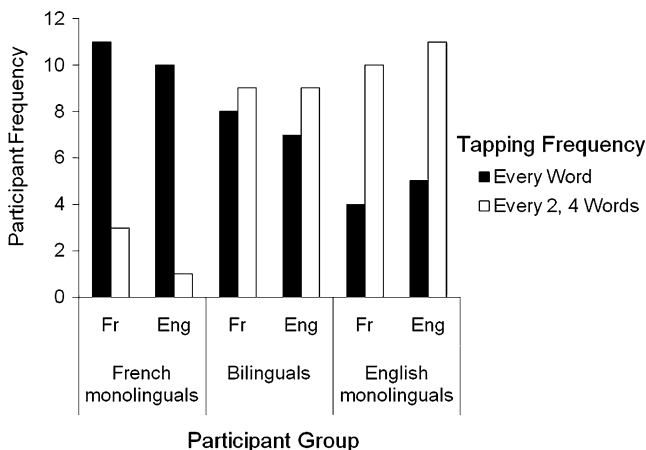


Fig. 3 Frequency of participants who tapped with every word or every two or four words by participant group and stimulus language

.060, $F < 1$], indicating that the group differences were specific to speech rhythms.

Discussion

Monolingual and bilingual speakers of English and French tapped more regularly to English than to French utterances, indicating that the stress-timed rhythms facilitated entrainment for all listeners. Even balanced bilinguals (highly proficient in both languages) showed increased tapping regularity for English utterances, as compared with French utterances; thus, the perceived rhythmic differences cannot be due to differential access to semantic information between the native and nonnative speakers. The utterances used in the present study were carefully controlled to allow the perception of alternating strong and weak stresses in both languages. This manipulation may have weakened natural rhythmic differences between French and English by artificially increasing the perceived regularity of the utterances. In spite of this, robust differences in listeners’ tapping regularity to French and English speech were found. Overall, these findings support the stress-timed/syllable-timed rhythmic distinction and suggest that speech rhythm differentially entrains listeners at acoustic levels prior to lexical comprehension.

Listeners’ tapping regularity increased for the more regular utterances within each language, suggesting that the acoustic stimulus regularity was sufficient to entrain tapping. Increased tapping variability was correlated with the variability of stimulus intervocalic intervals [CV(IVD)], an acoustic measure of temporal variability between vowel onsets (incorporating silent periods) that controls for speech rate. Tapping variability was not related to previous measures of vocalic durational variability that distinguish stress-timed from syllable-timed languages (%V and ΔV , Ramus et al., 1999; V-nPVI, Grabe & Low, 2002). Linguistic experience influenced the correlation between tapping variability and stimulus variability; French monolinguals showed less correspondence in tapping variability and stimulus variability for English, perhaps because they were less sensitive to the regularities present in English utterances.

Long-term experience with a stress-timed language seems to have heightened English speakers’ entrainment to stress regularities, as observed in their tendencies to tap regularly to both languages and to entrain to speech rhythms at higher hierarchical levels (tapping every two or four words). This finding challenges the idea that periodicity is specific to music and absent in speech (Patel, 2008). Speech entrainment might be driven by a perceived temporal regularity that depends in part on one’s familiarity with stress-timed languages. Native speakers of a stress-timed language may have greater expectations for rhythmic regularity in speech and may synchronize at higher levels with the “beat” of speech. These expectations

may transfer to the perception of foreign languages (Cutler, 2001). An effect of native language experience on the perception of speech stress converges with prior findings showing that the rhythmic properties of one's native language influence auditory grouping (Iversen, Patel, & Ohgushi, 2008), speech perception (Dupoux, Peperkamp, & Sebastian-Gallés, 2001), and speech segmentation (Cutler, 1991; Mehler, Dommergues, Frauenfelder, & Segui, 1981).

An ability to coordinate one's behavior with stimulus periodicities has many uses beyond speech perception and production (Cummins, 2009; Jungers et al., 2002; Wilson & Wilson, 2005). Cognitive processes of temporal prediction and movement timing are required for many behaviors, ranging from sharing the carrying of heavy objects to dancing with a partner. It is likely that processes underlying entrainment to speech rhythms would not be unique to language. The present findings suggest that experience with a stress-timed language influenced listeners' abilities to develop rhythmic expectations for other (nonnative) languages. Future research may examine whether an entrainment advantage conferred by experience with a stress-timed language extends beyond the carefully matched utterances of the present study to longer spans of spontaneous speech and to more complex rhythms, such as music.

Acknowledgments We gratefully acknowledge the assistance of Erik Koopmans, Jake Shenker, and Frances Spidle and suggestions about stimulus construction from Renée Béland and Laura Gonnerman. This research was supported by postdoctoral fellowships from the Belgian FRS-FNRS and the WBI-World Program to the first author and by Canada Research Chair and NSERC Grant 298173 to the second author. Correspondence should be sent to Caroline Palmer, Dept of Psychology, McGill University, 1205 Dr Penfield Ave, Montreal, QC H3A1B1, Canada, caroline.palmer@mcgill.ca.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

Table 2 List of French and English sentences

French	English
Chats et rats courent dans les champs, mais tous vont vers le Nord.	Rats and mice are in the grass, but some run through the house.
L'eau de source est bonne à boire quand l'air est chaud et lourd.	Days are bright with lots of light, but nights are cold and dark.
Dis aux filles de faire la fête à l'heure du cinq à sept.	Tell the girls to keep their poise when strange young men are near.
Gilles veut faire un tour au parc, mais là il	Jane and Tom were pals as kids, but then they

Table 2 (continued)

French	English
fait trop froid.	fell in love.
Gars et filles qui grimpent aux arbres peuvent voir les toits au loin.	Girls and boys who trip and fall can hurt their hands and knees.
Bière et vin sont bons au gout, mais trop en boire rend saoul.	Chips and cake are good as treat, but some will make you sick.
Biches et cerfs ont l'air très doux, mais chasse-les comme des proies.	Cats and dogs are pets of course, but treat them as your friends.
Bois du lait lui dit sa mère, car ça rend fort et grand.	Eat some bread, said Mom to Jeff, so you grow big and strong.
Tourne les yeux pour voir ma peine et sens comme je suis triste.	Turn your head to look at me, and tell me how you feel.
L'air du soir est bien trop frais pour mettre une jupe si courte.	Night in spring is much too cool to wear a dress that short
Tom est bon pour faire des blagues, mais Jean est bien plus drôle.	Snacks are good to bring to school, but too much food is bad.
L'or qui brille au doigt d'une femme veut dire qu'elle aime un homme.	Kids who play in mud and dirt will have to wash their pants.

References

- Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine.
- Boersma, P., & Weenink, D. (2007). *Praat: Doing phonetics by computer (Version 5.0)*. Retrieved December 10, 2007, from <http://www.praat.org>
- Cooper, W. E., & Eady, S. J. (1986). Metrical phonology in speech production. *Journal of Memory and Language*, 25, 369–384.
- Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37, 16–28.
- Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26, 145–171.
- Cutler, A. (1991). Linguistic rhythm and speech segmentation. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 157–166). London: Macmillan.
- Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, 5, 1–23.
- Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385–400.
- Dauer, R. (1983). Stress and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51–62.
- Drake, C., & Ben El Heni, J. (2003). Synchronizing with music: Intercultural differences. *Annals of the New York Academy of Science*, 999, 429–437.
- Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, 77, 251–288.
- Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient Bilinguals Dominance Scale. *Bilingualism: Language and Cognition*, 12, 273–289.

- Dupoux, E., Peperkamp, S., & Sebastian-Gallés, N. (2001). A robust method to study stress “deafness”. *Journal of the Acoustical Society of America*, *110*, 1606–1618.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Papers in laboratory psychology* (pp. 515–546). Cambridge: Cambridge University Press.
- Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, *124*, 2263–2271.
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and response to time. *Psychological Review*, *96*, 459–491.
- Jungers, M. K., Palmer, C., & Speer, S. R. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing*, *1*, 21–35.
- Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 389–403.
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*, 119–159.
- Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, *26*, 1–37.
- Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249–336.
- Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center). *Perception & Psychophysics*, *30*, 247–256.
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967.
- 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.
- Mersad, K., Goyet, L., & Nazzi, T. (2010). Cross-linguistic differences in early word-form segmentation: A rhythmic-based account. *Journal of Portuguese Linguistics*, *9*, 37–66.
- Morton, J., Marcus, S. M., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, *83*, 405–408.
- Nazzi, T., Bertoni, 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.
- Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.
- Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*, 969–992.
- Roach, P. (1982). On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73–79). London: Arnold.
- White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*, 501–522.
- Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*, 957–968.