Laboratory Phonology
Journal of the Association for Laboratory Phonology

Rathcke, T., et al. 2021 Tapping into linguistic rhythm. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1): 11, pp.1–32. DOI: https://doi.org/10.5334/labphon.248

JOURNAL ARTICLE

Tapping into linguistic rhythm

Tamara Rathcke[1,2,3], Chia-Yuan Lin[3], Simone Falk[4,5] and Simone Dalla Bella[5,6,7,8]

[1] Department of Linguistics, University of Konstanz, DE
[2] MARCS Institute for Brain, Behavior and Development, Western Sydney University, AU
[3] English Language and Linguistics, University of Kent, UK
[4] Department of Linguistics and Translation, University of Montreal, CA
[5] International Laboratory for Brain, Music and Sound Research (BRAMS), Canada
[6] Department of Psychology, University of Montreal, Canada
[7] Centre for Research on Brain, Language and Music, Montreal, Canada
[8] Department of Cognitive Psychology, University of Economics and Human Sciences in Warsaw, Warsaw, Poland

Rhythmic properties of speech and language have been a matter of long-standing debates, with both traditional production and perception studies delivering controversial findings. The present study examines the possibility of investigating linguistic rhythm using movement-based paradigms. Informed by the theory and methods of sensorimotor synchronization, we developed two finger-tapping tasks (synchronization and reproduction), and tested them with English participants. The synchronization task required participants to tap along with the beat of a looped sentence while the reproduction task asked them to tap out the perceived beat patterns after listening to a sentence loop. The results showed that both tasks engaged participants in period tracking of a beat-like structure in the linguistic stimuli, though synchronization did so to a greater extent. Patterns obtained in the reproduction task tended to converge toward participants' spontaneous tapping rates and showed a degree of regularization. Data collected in the synchronization task displayed a consistent anchoring of taps with the vowel onsets. Overall, synchronization performance with language resembled many well-established findings of sensorimotor synchronization with metronome and music. We conclude that our setting of the sensorimotor synchronization paradigm—finger tapping along with looped spoken phrases—is a valid experimentation tool for studying rhythm perception in language.

Keywords: Speech rhythm; sensorimotor synchronization; motor reproduction; vowel onset; p-centre

1. Introduction

Is language rhythmic? For decades, this seemingly simple but profoundly important question that connects language with other aspects of human cognition has been controversially debated (Cummins, 2012; Roach, 1982). Early accounts of linguistic rhythm suggest that it relies on some acoustic isochrony in spoken language—either at the level of syllables or inter-stress intervals (Abercrombie, 1967). However, this idea did not stand up to acoustic measurement as no temporal analyses of speech have ever provided any evidence for an isochrony-based rhythm (Dauer, 1983; Fowler & Tassinary, 1981; Pointon, 1980; Roach, 1982; Uldall, 1971; van Santen & Shih, 2000). The failure to find isochrony in speech has led to the development of alternative approaches. One of the most prominent proposals has suggested that linguistic rhythm may be ascribed to the durational variability present in consonantal and vocalic intervals, with languages being more or less variable and thus sounding rhythmically different (Dellwo & Wagner, 2003; Deterding, 2001; Grabe & Low, 2002; ling Low, Grabe, & Nolan, 2000; Ramus, Nespor, & Mehler, 1999; White

& Mattys, 2007). Initially appealing and fuelling much research into the cross-linguistic study of rhythm, this approach has been recently critiqued as empirically inadequate and misrepresenting the issue at heart (Arvaniti, 2009, 2012; Arvaniti & Rodriquez, 2013; Barry, Andreeva, & Koreman, 2009; Kohler, 2009; Rathcke & Smith, 2015a, 2015b; Wiget et al., 2010). The latest attempts at capturing rhythmic properties of spoken language involve analyses of the prosodic hierarchy and its temporal signatures, i.e., durational implementation of the hierarchical structure of linguistic utterances (Rathcke & Smith, 2015a; White, Payne, & Mattys, 2009), though this proposal has also seen some counterevidence (Mairano, Santiago, & Romano, 2015). Other recent approaches are less theoretically than acoustically driven, and rely on signal analyses such as properties of the amplitude envelope to capture some rhythmic properties in language (Goswami et al., 2002; Port, Cummins, & Gasser, 1995; Šturm & Volín, 2016; Tilsen & Arvaniti, 2013). Following on from these diverse and controversial accounts of rhythm, ideas have been put forward that language has a scale of rhythmicity (Kohler, 2009), is only occasionally rhythmic (White, Mattys, & Wiget, 2012), or even anti-rhythmic (Nolan & Jeon, 2014).

However, it has also been noted that previous production and perception studies of linguistic rhythm do not capture one of the core features of rhythm—its ability to entrain movement (Cummins, 2009, 2012). The idea that rhythm perception and movement are closely interconnected looks back at a long history (Bolton, 1894), and is supported by a growing body of evidence showing that beat and rhythm perception involve motor regions of the brain (e.g., the basal ganglia and premotor cortex) and their connections to auditory regions (e.g., Brett & Grahn, 2007; Grahn & Rowe, 2009; Patel & Iversen, 2014; Zatorre, Chen, & Penhune, 2007). Behavioural research has exploited the potential of external, rhythmically structured events to entrain movement, with the goal to gain a better understanding of rhythmic mechanisms and their underpinnings. The sensorimotor synchronization (SMS) paradigm has been developed and successfully utilized to study rhythm perception and the properties of the human timing system by observing how a motor action is temporally coordinated with an external auditory event (Aschersleben, 2002; Repp, 2005; Repp & Su, 2013). Such coordination of a perceived rhythm and a motor action gives insights into mechanisms underlying our capacity to achieve complex coordination in time when we dance, jointly sing, or chant, such as perceptual beat tracking and the generation of temporal expectancies (Repp, 2005). In the present study, we test the potential of SMS to provide new insights into the rhythmic organization of language.

The simplest way to measure SMS is to record finger tapping in time with an auditory stimulus, such as repeated sounds of a metronome or more complex musical sequences (Aschersleben, 2002; Repp, 2005; Repp & Su, 2013; for batteries of tests involving SMS see Dalla Bella et al., 2017; Iversen & Patel, 2008). Typically, the task consists in synchronizing finger taps produced with the dominant hand to the beat perceived in the auditory signal. Measures of the temporal asynchrony between the stimulus and the tap, the duration, and the variability of inter-tap intervals quantify the synchronization performance and motor stability during the task.

According to this research, the most stable patterns of synchronization arise when participants tap at 1:1 or other multiple integer ratios in-phase with the beat whereas more complex ratios and anti-phase tapping are generally more difficult (Bouvet, Varlet, Dalla Bella, Keller, & Bardy, 2019; Repp, 2005). Synchronized tapping on temporal scales longer than 800 ms is often more difficult and breaks down completely when inter-onset intervals exceed 1.8–2 seconds (Engström, Kelso, & Holroyd, 1996; Mates, Müller, Radil, & Pöppel, 1994). The fastest tapping rates occur at the inter-onset intervals of 150–200 ms (Repp, 2005; Truman & Hammond, 1990), and are faster in musicians than non-musicians (Repp, 2003). In addition to motor constraints, upper and lower rate limits reveal general cognitive constraints on temporal processing (Repp, 2005) and have been related to the working memory capacity (Pöppel, 1997).[1] Importantly, both upper and lower SMS limits of manual synchronization with a pacing signal have meaningful temporal counterparts in speech acoustics. Lower limits correspond to the average duration of a vowel or a syllable, and upper limits can correspond to the duration of larger units such as a prosodic or a syntactic phrase. Existing evidence further demonstrates that SMS can take place even in syncopated signals (Large & Palmer, 2002) and in auditory stimuli with complex metrical structures (Madison, 2014). That is, SMS responds to those features that bear the closest resemblance to language. Unlike traditional speech perception experiments that rely on listeners' metalinguistic conceptualization of rhythm, and delivered inconclusive results in the past (Miller, 1984; Rathcke & Smith, 2015b), SMS appeals as more intuitive to non-specialists and taps into the motor routines that are known to sharpen sensory rhythmic representations in music (Morillon, Schroeder, & Wyart, 2014; Ravignani et al., 2019).

Although it has been noted that the ability to synchronize movement to an external timekeeper is predominantly human and might have even played an important role in the evolution of language and music (Merker, 2000; Ravignani et al., 2019), SMS paradigms have so far inspired relatively little interest within the linguistic field. Some early work on speech rhythm (Allen, 1972) utilized a version of sensorimotor synchronization with speech by asking listeners to tap to a designated syllable in a spoken sentence that was repeated and played back to them 50 times. The results suggested that the paradigm was able to unveil the 'beat location' of a syllable which was close to a vowel onset and varied with prominence and syllable structure. Similarly, Falk, Rathcke, Dalla Bella, and Bella (2014) used sentence looping and demonstrated that SMS was highly sensitive to a low-level, within-language timing variation, thus suggesting that the method is well suited for the study of subtle rhythmic differences in spoken signals, despite their high complexity and temporal variability. Lidji, Palmer, Peretz, and Morningstar (2011) compared finger tapping performance of monolingual French and English participants as well as French-English bilinguals when tapping to French and English spoken sentences with a regular metrical structure of strong and weak syllables. The study revealed that both the listener's native language and the language-specific acoustics affected the obtained tapping patterns in terms of tapping frequency and inter-tap-interval variability.

Falk and Dalla Bella (2016) used finger tapping with metrically regular sentences to examine potential benefits on speech and language processing that might arise from a concurrent motor activity while listening. Tapping congruently (i.e., in-phase) with accented syllables was found to enhance speech processing compared to incongruent (i.e., anti-phase) tapping, or listening without the motor activity. Such linguistic processing advantages may be supported by increased attentional resources being available through the coupling of perception and action that is typical of SMS tasks (cf. Hommel, 2015; Large & Jones, 1999). Overall, the recent SMS studies with language suggest that movement-based paradigms tap into language-specific rhythmic properties of speech.

Most of the studies above reported measures of motor rate and variability (including duration of inter-tap-intervals, ITI, and the coefficient of their variation, CV). Dalla Bella and colleagues (Dalla Bella, Białuńska, & Sowiński, 2013) note that SMS with language displays a relatively high amount of variation in contrast to SMS with music (reflected in a CV of 30% versus 4%, respectively), though it is unclear if this variability arises from the fact that language entrains movement less than temporally more regular stimuli (e.g.,

[1] Synchronization abilities can also vary across individuals (Dalla Bella et al., 2017) and across music-cultural environments (Polak et al., 2018).

music) as the authors suggest, or is rather reflective of the unique temporal properties of language such as lack of isochrony in its acoustic signal (cf. Dauer, 1983; Pointon, 1980; Roach, 1982; Uldall, 1971; van Santen & Shih, 2000).

Only a few studies to date have addressed the question of potential SMS anchors in the acoustic signal of speech (Allen, 1972; Falk, Volpi-Moncorger, & Dalla Bella, 2017; Rathcke, Lin, Falk, & Dalla Bella, 2019). An answer to this question crucially hinges on empirical evidence that would demonstrate whether or not listeners attempt to systematically synchronize their movement with some specific points in the time course of an acoustic signal. Falk et al. (2017) defined SMS-anchors to coincide with the so-called 'perceptual centres' (or p-centres, Marcus, 1981). A p-centre describes the subjective moment of occurrence of an event (typically a syllable in speech). More often than not, the p-centre and the acoustic onset of the corresponding event do not co-occur in time (Cooper, Whalen, & Fowler, 1986; Marcus, 1981; Morton, Marcus, & Frankish, 1976). The original interest in p-centres arose from a search for some temporal constancy in language, and led to the hypothesis that temporal isochrony in language might be perceptual, and not acoustic, in nature (Lehiste, 1977; Morton et al., 1976). There have been attempts to localize the p-centre at the midpoint of the amplitude rise-time at the onset of nuclear accented vowels (following Cummins & Port, 1998; see also Morton et al., 1976). However, neither kinematic nor any acoustic properties of speech signals seem to consistently capture the essence of the p-centre location (De Jong, 1994; Patel, Löfqvist, & Naito, 1999), and after 40 years of p-centre research, a comprehensive and reliable account of the phenomenon still remains a desideratum (Villing, Repp, Ward, & Timoney, 2011).

In our own study (Rathcke et al., 2019), we systematically examined several potential anchors by measuring SMS with very simple verbal stimuli, namely regularly spaced sequences containing alternations of syllables /bi/ and /bu/. The results of the study suggest that vowel onsets are likely to serve as attractors of individual taps. Moreover, SMS accuracy with similarly structured verbal and tonal stimuli did not significantly differ, if SMS in verbal stimuli was measured at vowel onsets. The latter result echoes previous findings obtained using a different tapping task (Dalla Bella et al., 2013). When linguistic stimuli closely resemble the metrical structure of music, the discrepancy between music and speech in their ability to attract movement disappears. However, natural speech is rarely metrical and never isochronous. Thus, rhythmic motor entrainment with language is yet to be demonstrated.

In contrast, a movement task that does not involve synchronization avoids the challenge of locating a tapping anchor in the acoustic signal. This non-synchronized motor reproduction (henceforth, NMR) paradigm has occasionally been used in previous speech research (Donovan & Darwin, 1979; Scott, Isard, & de Boysson-Bardies, 1985; Wagner, Cwiek, & Samlowski, 2019). In this task, listeners are asked to tap or drum a perceived rhythmic pattern after listening to an auditory prompt. NMR is somewhat similar to the synchronization-continuation paradigm which is common in the timing literature (e.g., Wing, 2002), with the difference that a synchronized tapping phase is missing. In NMR, listeners' tapping performance can be quantified by the interval duration between their taps (ITI) and by the variability in the interval duration (CV of the ITIs). When measuring period tracking of a beat in a linguistic input by means of such a non-synchronized task, ITI and CV could reflect some meaningful timing properties of the corresponding speech signal, e.g., syllable, word, or phrase duration, and the number of taps could reflect the number of rhythmically relevant events. Since this beat tracking ability of NMR has not been explicitly demonstrated in previous work, it is as yet unclear if, and how well, this motor paradigm can assist with understanding rhythm perception in language.

2. Aims and hypotheses

The aims of the present study were two-fold: (1) to test whether or not motor paradigms can help to tap into rhythm perception in natural language, and (2) to identify which paradigm would optimally support this.

While it seems natural and easy to synchronize to music, prolonged motor synchronization to speech is at first sight a less obvious and widespread activity (Dalla Bella et al., 2013), although previous studies have provided some evidence that it is not impossible (e.g., Allen, 1972; Falk et al., 2014; Lidji et al., 2011; Rathcke et al., 2019). It has its natural precursors in motor engagement with nursery rhymes (Cardany, 2013), clapping to political oratory (Tanaka & Rathcke, 2016) or co-speech gesturing (Wagner, Malisz, & Kopp, 2014). There are different implementations of a laboratory SMS paradigm with speech that can be found in the literature. For example, Lidji et al. (2011) asked their participants to listen to three repetitions of a spoken sentence in total, and to synchronize their finger taps with the beat during the second and the third presentation of the sentence. In the present study, we decided to use a larger number of repetitions, and participants were instructed to tap along with the perceived beat throughout a sentence loop.

Capitalizing on the general perceptual phenomenon of repetition (Margulis & Simchy-Gross, 2016; Rowland, Kasdan, & Poeppel, 2019), looped speech has the potential to reveal underlying rhythmic structures of sentences (Falk et al., 2014; Rathcke, Falk, & Dalla Bella, 2018). After having listened to repetitions of a sentence, listeners are no longer engaged in cognitively demanding semantic and syntactic processing. Instead, they can attend to the prosodic structure of the sentence and extract its rhythmic properties more easily. Looped speech is also known to sometimes induce the so-called 'speech-to-song illusion' (S2S). S2S describes a perceptual phenomenon in which an originally spoken phrase switches to being perceived by many people as a song if it is embedded in a loop (Deutsch, 2003; Deutsch, Henthorn, & Lapidis, 2011). However, not all phrases are equally likely to transform into song (e.g., Falk et al., 2014), and we controlled for this phenomenon in our materials. Early work by Allen (1972) also utilized looped sentences, though participants of this experiment were only asked to synchronize with one designated syllable on each repetition and not tap along with the beat of the looped sentence as in this study.

The present approach is also different from the NMR paradigm implemented in previous research in which listeners were either presented with one repetition of each stimulus and could hear the stimulus again if needed (Wagner et al., 2019), or presented with 10 repetitions of a stimulus but were asked to tap after each repetition (Donovan & Darwin, 1979; Scott et al., 1985). The looped version of both SMS and NMR represents a principled way of expanding and applying current movement-based paradigms to language.

Given the differences in the nature of the SMS and NMR tasks, different scenarios would indicate (the degree of) the success of the paradigm in representing perceived rhythmic structures. In the case of SMS, motor entrainment necessarily involves the presence of a consistent anchor of synchronization in the speech signal. Lack of temporal consistency between a tap occurrence and an acoustic landmark will suggest lack of entrainment. In the case of NMR, beat tracking and rhythmic reproduction would be considered successful if tapping rates diverge from participants' natural and preferred tapping rate and converge towards the IOI of meaningful durational intervals in the linguistic input. If patterns resulting from NMR deviate from this prediction, beat tracking cannot be assumed to have been successful during the task. We further expect to find individual variation in both tasks, which should be at least partially explainable by individual musicality, timekeeping, and synchronization abilities (Dalla Bella et al., 2017).

In summary, this study set out to assess two movement-based paradigms that had been previously used with language, synchronization and reproduction, with the aim of providing empirical evidence on a methodology that would be best suited for studying rhythmic properties of spoken language.

3. Method

3.1 Experimental stimuli

Six English sentences were chosen from an existing database that we had previously created to investigate S2S. Every sentence in this database has been tagged for S2S-likelihood based on perception data obtained from 40 healthy native listeners (Rathcke et al., 2018, forthcoming; see **Table 1**). The sentences were read by a female native Standard Southern British English speaker (22 years old at the time of the recording).

The materials of the present study comprised six sentences, two with four syllables, two with seven, and two with ten syllables. The ten-syllable sentences were syntactically more complex than the shorter ones. To control for the possibility that repetition might induce musical interpretation of speech and thereby bias synchronization or reproduction in unexpected ways, the chosen pairs varied in their probability to induce S2S, with one high-transforming and one low-transforming sentence in each pair (see **Table 1**).

The stimuli were repeated 20 times for the SMS task and 10 times for the NMR task (see 3.7). A 400 ms pause separated the repetitions.

3.2 Sentence annotations

A trained phonetician (first author) annotated onsets of vowels and syllables in the test sentences. Vowels were defined as syllabic nuclei identified by the presence of voicing, formant structure, and relatively high intensity. Accordingly, pre-aspiration was excluded. Segmentation of vowel onsets in post-sonorant contexts combined acoustic and auditory criteria that were guided by impressions of the intended vowel quality. There were no segmental reduction or deletion phenomena, given that the recordings comprised of clear, read speech samples. There were also no cases of glottalization in these recordings. Segmentation examples are given in **Figures 2** and **3**. All materials are available from https://osf.io/3dh4m/. An independent annotator segmented vowel onsets in all test sentences following the criteria given above, and reached a cross-annotator agreement of .999945 ($p < .001$) in Pearson's correlation coefficient.

Additionally, each syllable (and its vowel) was specified with respect to its metrical status (strong or weak) and phrasal prominence (accented or unaccented). From these annotations, we derived acoustic timings of the two potential synchronization anchors (syllable or vowel) and linguistic prominence of the underlying units (0 for metrically weak, 1 for metrically strong but unaccented, 2 for accented syllables/vowels).

Table 1: Summary of the materials used in the experiment. Accented syllables are underlined.

Sentence length	Sentence	S2S likelihood	Duration (seconds)
short: 4 syllables	S1: *I wove a yarn.*	68% (high)	1.1
	S2: *I took the prize.*	30% (low)	1.2
medium: 7 syllables	M1: *Ann won the yellow award.*	73% (high)	1.5
	M2: *Grandpa did not eat the cake.*	35% (low)	1.8
long: 10 syllables	L1: *The incident occurred last Friday night.*	34% (low)	2.1
	L2: *As the boy sneezed, the door closed suddenly.*	58% (high)	3.0

3.3 Acoustic pre-processing

To define potential acoustic SMS anchors, linguistically informed data preparation above was complemented by analyses of the amplitude envelope (cf. Goswami et al., 2002; Port, Cummins, & Gasser, 1995; Tilsen & Arvaniti, 2013) and signal energy derivatives (Šturm & Volín, 2016). Amplitude envelopes were created by employing the envelope function in Matlab (2018b). Accordingly, envelopes were derived from the absolute signal amplitude, and smoothed using a spline interpolation with a window of at least 500 samples (amounting to approximately 11 ms). Smoothed energy contours were derived following the procedure developed by Šturm and Volín (2016) which was based on the calculation of energy averages across 40-ms segment windows with a 44-sample shift and the 6th-order moving-average filter.

Figure 1 compares the amplitude envelope and the smoothed energy contour for the test sentence M2, and demonstrates core differences between the two contours. The amplitude envelope (shown in blue) closely follows the waveform, apart from the sections where there is a discrepancy between positive and negative amplitude values (since the envelope is based on an average of the absolute values). In contrast, the energy function shows multiple deviations from the original waveform, especially in regions of low sonority. Moreover, energy contours of open vowels at the beginning of the sentence are more closely matched to the waveform than the contours of close vowels towards the end of the sentence.

Additionally, local energy dynamics of voiced parts of the acoustic signal were described following the formula in Šturm and Volín (2016). This function operates on the smoothed energy contour, calculates differences between two neighbouring samples (disregarding samples with zero-crossing rates higher than 4000 that are typical of voiceless fricatives), and smooths the difference values via a moving-average filter of order 10. **Figure 2** displays the two energy contours in comparison (red versus green lines). Subsequently, local maxima of the smoothed energy function (maxE) and local maxima of the energy difference function (maxD) were identified and localized within each syllable of the test sentences. According to Šturm and Volín (2016), maxD represents a close approximation of the p-centre in Czech.

Both values (maxE and maxD) were subsequently examined with respect to their ability to serve as SMS anchors, along with the syllable amplitude maxima derived from amplitude

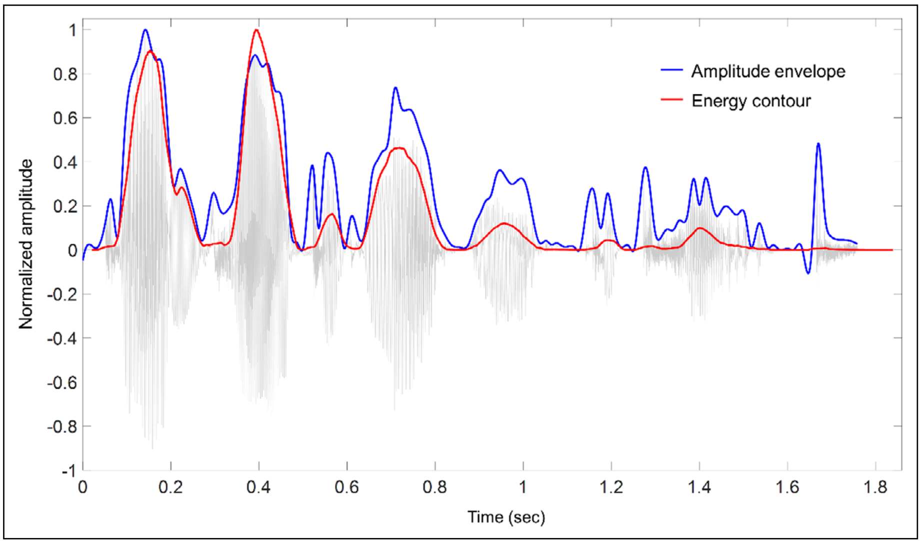

Figure 1: Waveform (grey), amplitude envelope (blue), and energy contour (red) of the test sentence M2 ("Grandpa did not eat the cake").

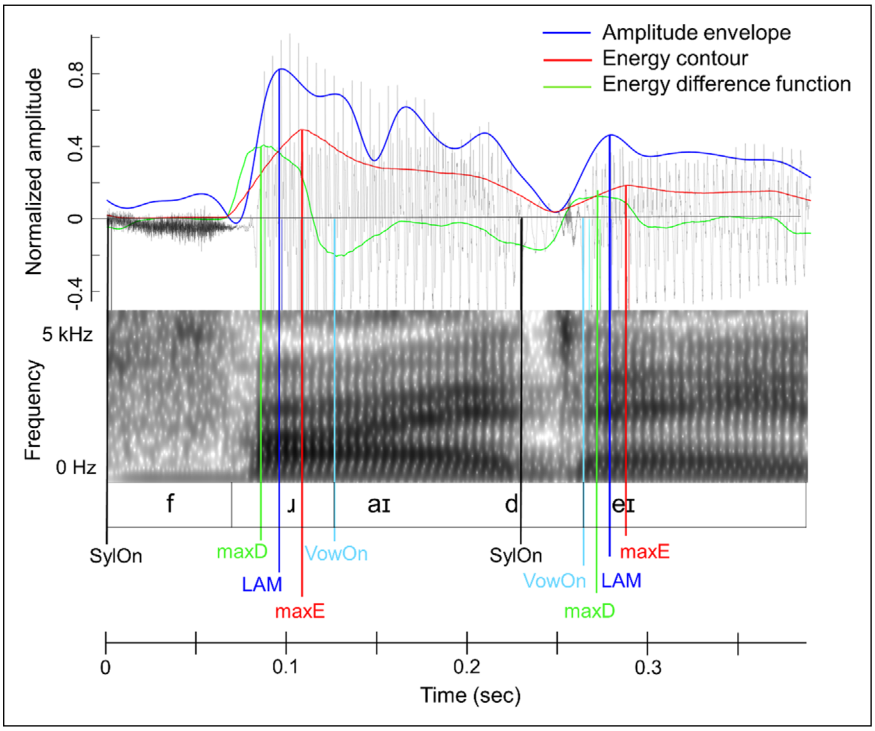

Figure 2: Waveform, spectrogram, and annotation of the example word "Friday" comparing temporal locations of the five landmarks under investigation: (1) syllable onset (SylOn, identified manually), (2) vowel onset (VowOn, identified manually), (3) local amplitude maximum (LAM, derived from the amplitude envelope, indicated in blue), (4) local energy maximum (maxE, derived from the energy contour, indicated in red), (5) fastest local energy increase (maxD, derived from the energy difference function, indicated in green).

envelopes and the syllable/vowel onsets identified manually. **Figure 2** compares temporal locations of the five potential SMS anchors for the bisyllabic word "Friday" (taken from L1). The figure illustrates our general finding that distances between the five temporal landmarks varied depending on the properties of the syllable, and could be rather small (as in the second syllable of the example word) or large (as in the first syllable).

Each temporal landmark was then further described on the basis of the acoustic properties in their local amplitude envelopes. More specifically, two measures of local changes in the signal amplitude were identified: (1) rise-time, i.e., the temporal distance between a local amplitude minimum and a maximum (Goswami et al., 2002; Goswami & Leong, 2013) and (2) rise-slope, i.e., the steepness of a change in the envelope measured as the amplitude differences between a minimum and a maximum, divided by the duration of their rise-time. These measures are illustrated in **Figure 3**.

To derive these measures, neighbouring maxima and minima in the amplitude envelope were located using the *findpeaks* function in Matlab 2018b. First, a local maximum was found in the closest proximity to the event onset (in vowels, it could precede or follow the identified vowel onset but in syllables, the temporal location was restricted by syllable boundaries). Second, the algorithm searched for a preceding minimum in the amplitude envelope. The local minima were mostly located around the 'valleys' between local maxima in the amplitude envelope (see **Figure 3**). A local threshold was adjusted to each individual case, based on a combination of two parameters (duration of the sampling time window and average amplitude decrease over a series of consecutive sampling intervals). Automatically detected turning points of the amplitude envelope were manually checked by a trained phonetician (first author).

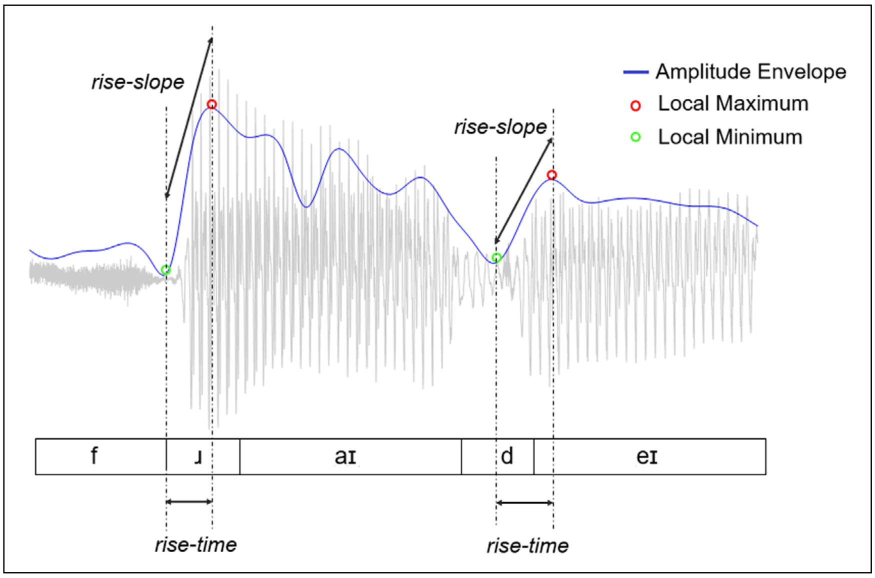

Figure 3: Waveform and amplitude envelope of the example word "Friday" (taken from L1). Local amplitude maxima (in red) and minima (in green), and their derived measures of rise-time and rise-slope are indicated.

3.4 Individual data

All participants had to fill in an online questionnaire prior to their scheduled experimental session. The form asked about musical training, ongoing and past musical activities, and dancing experience. A general musicality index was derived from these data (similar to the approach by Šturm & Volín, 2016). The index was an aggregate score based on years of musical training (from 0 to 12 in the present sample), current regular music practice (0 for non-active and 1 for active participants), number of musical instruments (which included singing and dancing, from 0 to 4 in the sample), and finally the age at which participants started their musical training (below the age of 10 was coded as 2, from 10 up to 20 years as 1, above 20 years as 0). The resulting musicality indices varied from minimally 0 (no musical experience) to 18 (a high level of musical experience and skills). There were no professional musicians or dancers among the participants of this study, though 69% had received some musical training, taken dancing classes, or danced regularly. Given the aims and hypotheses of the present study, the questionnaire only included questions about active music practice and did not collect information about passive experience of listening to specific music styles.

We assessed individual SMS abilities with the Battery for the Assessment of Auditory Sensorimotor Timing Abilities (BAASTA, Dalla Bella et al., 2017). Six tasks were selected from the battery, including two unpaced tapping tasks, two paced tapping-to-tones tasks, and two paced tapping-to-music tasks. Unpaced tapping tasks measured the speed of participants' spontaneous and fast tapping rates without a stimulus, and their ability to sustain the regular motor activity. In these tasks, participants were instructed to either tap at their most comfortable speed for 60 seconds, or to tap at their fastest possible speed for 30 seconds, paying attention to maintaining a constant speed for the whole duration of a trial. In the paced tapping tasks, participants' synchronization abilities were measured with a simple regular sound (here, a piano tone with a frequency of 1319 Hz or E6) and computer-generated excerpts of classical music. When tapping to piano tones, participants were presented with 60 repetitions of a tone presented at a faster (450 ms) and a slower (600 ms) IOI, and asked to tap in synchrony with the tones throughout the repetitions. When tapping to music, participants were instructed to synchronize with what they perceived as the beat in musical excerpts from Bach's "Badinerie" and Rossini's "William Tell Overture." Both music extracts consisted of 64 beats with a quarter note of 600 ms IOI (see Dalla Bella et al., 2017 for more detail). The task order was counterbalanced across experimental sessions, following a Latin square design. The order within each task category was fixed though: In the unpaced tapping task, participants first tapped spontaneously then fast; in the paced tapping task, they first synchronized with the metronome of 450 ms IOI and then 600 ms IOI; in the music synchronization task, they first tapped to Bach and then to the Rossini piece.

After the experiment, all participants filled in a brief questionnaire that collected confidence ratings for their self-evaluated experimental performance. Participants indicated how easily they were able to extract beat patterns from looped test sentences, and how well they were able to replicate them in the NMR task. In addition, they were asked to self-report how confident they felt about tapping precisely in time with the stimuli they experienced in the SMS task. A 9-point Likert scale (with 9 being the highest level of confidence) was used to collect the ratings.

3.5 Participants

Thirty-one native speakers of Southern British English (21 female; mean age 23.1 years, range 18 – 36 years) participated in the study. They gave informed consent and received a small fee in compensation for their time and efforts. The data of two participants were removed from the sample because they self-declared as dyslexics. All remaining participants had no existing history of language impairments or motor disorders that could affect their rhythmic processing or SMS abilities (e.g., dyslexia: Leong & Goswami, 2014; apraxia: Park, 2017; dystonia: Liu et al., 2008), and no hearing impairments at the time of testing. Moreover, their individual performance with the metronome tasks of BAASTA did not indicate any issues with their general synchronization abilities (cf. Dalla Bella et al., 2017).

3.6 Tasks, procedure, and apparatus

The study obtained ethical approval from the Ethics Committee of the University of Kent, and was conducted in a quiet behavioral testing room of the Kent Linguistics Laboratory. Each experimental session consisted of one SMS and one NMR task with the experimental stimuli. During the SMS task, participants were presented with 20 repetitions of each target sentence and asked to start synchronizing with what they perceived as the beat structure of the sentence as soon as they felt able to, while the repeated auditory sequence was still ongoing. In the NMR task, participants were asked to listen to 10 repetitions of a test sentence first, and then to replicate the beat pattern they had heard. No instruction was given as to how many taps or cycles they should reproduce. During each task, test sentences were presented in increasing order of complexity, i.e., the short sentences were tested first, the long ones last. Test sentences were played binaurally through Sennheiser HD 380 headphones. The order of the SMS and NMR tasks was counterbalanced across participants. At the start of an experimental session, participants familiarized themselves with the equipment and had an opportunity to clarify their questions about the procedure. The session ended with the BAASTA tests and the post-test questionnaire, and took 35–45 minutes in total to complete.

Tapping responses were collected using a Roland HandSonic drum pad (HPD-20) and a Dell Latitude 7390 laptop in the CakeWalk MIDI software (BandLab). BAASTA was implemented as an app running on an Acer tablet (Iconia One 10 B3-A40FHD 32GB) with an Android 7.0 system. Participants were free to adjust the sound volume to a comfortable level.

3.7 Preparation of SMS data

Collected taps can be analyzed with regards to different aspects of their distribution in time. **Figure 4** shows a hypothetical example of two taps produced in time with the bisyllabic word Friday. The first tap follows the syllable onset (resulting in positive asynchronies measured with this landmark) but precedes all other landmarks (resulting in negative asynchronies indicative of anticipation of different magnitudes: maxD can be considered less anticipated than the vowel onset in this example). The second tap shows positive asynchronies for all landmarks, though the magnitude of the time lag is landmark-specific—here, it is the smallest for maxE and the largest for the syllable onset. Moreover, the distance between the taps (or the inter-tap interval, ITI) and the variability of these intervals can provide insightful information about the sychronization performance with each sentence as a whole.

We extracted the tapping data using Matlab MIDI toolbox (Eerola & Toiviainen, 2004), and corrected the timing of taps by subtracting the delay of the MIDI device (here, 5 ms). For each sentence and participant, we calculated the temporal distribution of individual taps within the temporal window of the sentence duration and then aggregated the available taps across all repetitions of the same sentence. Using GGPLOT2 (Wickham, 2016), a Gaussian kernel estimation with a bandwidth adjustment of ⅛ was applied to the aggregated data. This procedure allowed us to obtain a smoothed distribution for each participant and sentence while retaining salient peaks of the aggregated taps. **Figure 5** shows an example of such density functions for the test sentence S1. Individual densities were obtained from the SMS data and aggregated across all participants. The resulting distribution in **Figure 5** is clearly multimodal, with one tapping peak per syllable of this test sentence.

Figure 6 displays density functions created for the group tapping performance with the same sentence during the NMR task. The NMR task seems to increase variability at both individual and group level, and lacks the clearly defined quadrimodality observed in the SMS task with this sentence.

The temporal location of the density peak maxima (see **Figure 5**) was used to quantify the individual SMS performance. To derive this measure, the *findpeaks* function from

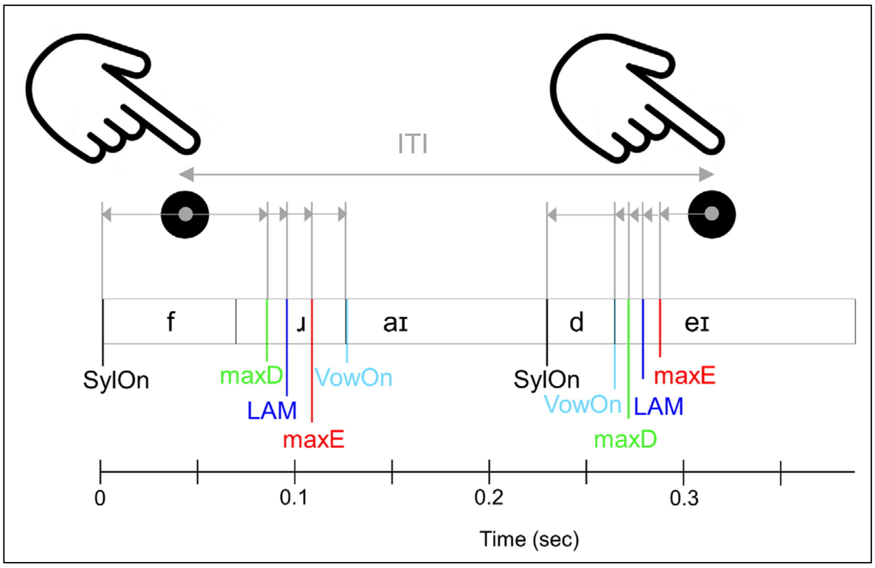

Figure 4: A hypothetical example of two taps (identified by black circles) produced in time with the bisyllabic word Friday. Distances between the taps (or inter-tap intervals, ITI) and distances between taps and the nearby landmarks are indicated by grey arrow lines.

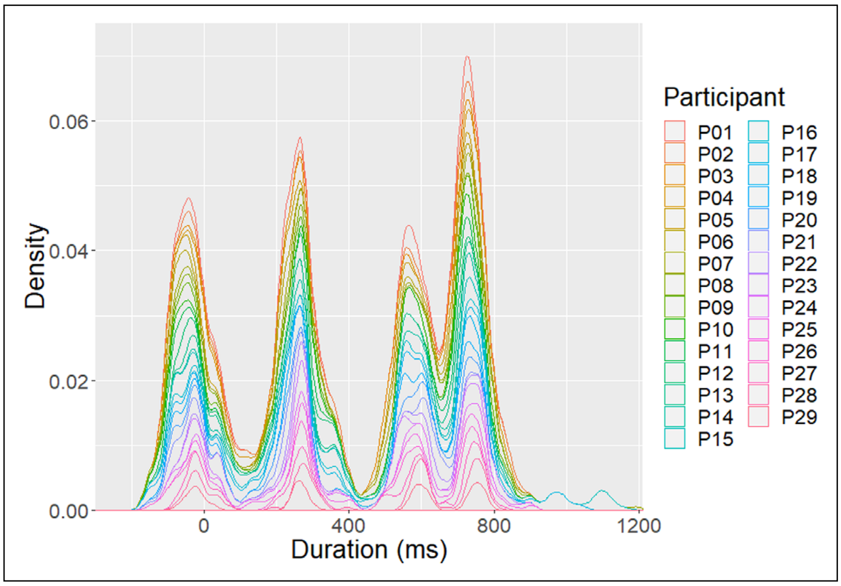

Figure 5: An aggregated density function of the group SMS-performance with the four-syllable test sentence S1 ("I wove a yarn").

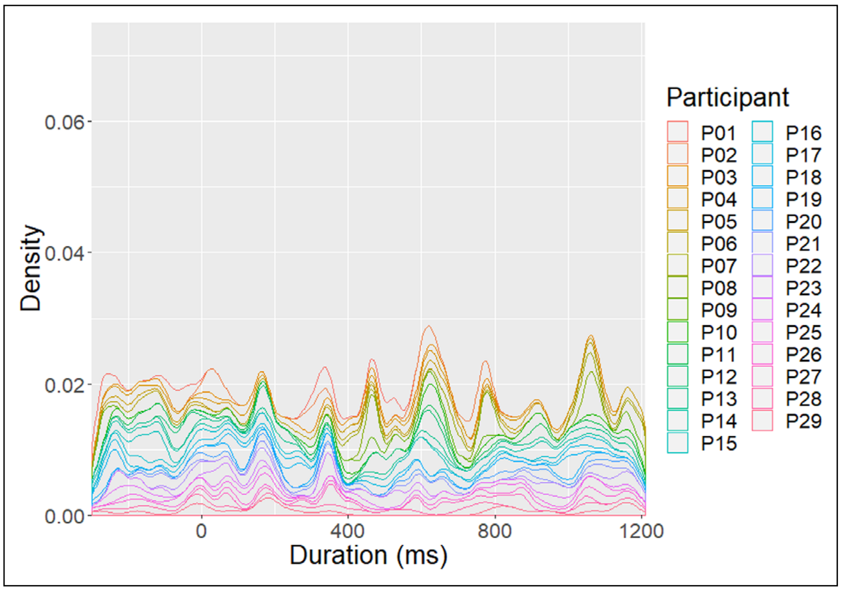

Figure 6: An aggregated density function of the group NMR-performance with the test sentence S1 ("I wove a yarn").

the R-package PRACMA (Borchers, 2018) was applied. It identified all peaks using a 40%-threshold of the maximum peak value of each sentence, separated by at least 100 ms distance. The timepoints of the density peaks and locations of the temporal landmarks under investigation (see 3.3) were then compared. Asynchronies between the taps and the temporal landmarks were calculated for those density peaks which occurred within a ±120 ms window of the corresponding landmark location (cf. Repp, 2004).

3.8 Period tracking in SMS and NMR

To compare the properties of period tracking in SMS and NMR, we calculated ITIs (in ms) that participants produced in each target sentence during the very first movement cycle as well as mean ITIs upon completion of a trial. Variability of the interval duration between taps was expressed by the coefficient of variation CV, calculated as SD(ITI)/mean(ITI).

4. Results

All analyses below were conducted in Rstudio running R version 3.5.1., using packages LME4 (Bates, Mächler, Bolker, & Walker, 2015), LMERTEST (Kuznetsova, Brockhoff, & Christensen, 2017), CHANGEPOINT (Killick & Eckley, 2014), GGPLOT2 (Wickham, 2016), and SJPLOT (Lüdecke, 2019).

4.1 Motor activity in SMS and NMR

First, one-sample Kolmogorov-Smirnov tests confirmed that tapping data were not uniformly distributed in both tasks. That is, participants did not tap randomly in either task, or any of the test sentences (see **Table 2**; all *p* values were < .01). Moreover, confidence ratings did not differ significantly between the two tasks. Participants felt equally confident (median: 6, interquartile range: 5–7) about their ability to extract the beat patterns in NMR and to synchronize with the beat in SMS.

4.2 Comparisons between landmarks as potential SMS anchors

To find the most appropriate SMS-anchor, we fitted linear mixed-effects models to absolute asynchronies between the tapping peaks and the temporal landmarks. The model included a five-level predictor landmark and two random effects: participant (P1–P29) and sentence (1–6). We started with a maximal random effect structure recommended by Barr, Levy, Scheepers, and Tily (2013), and iteratively removed random effects if the model failed to converge or produced a singular fit. A change of the default optimizer (to 'optimx,' John et al., 2020) helped to resolve the model convergence issues and keep the random effect structure maximal. The likelihood ratio tests were run to determine the best-fit models.

Figure 7 displays estimates and standard errors of absolute asynchronies for the five landmarks under investigation. Smaller asynchronies indicate a higher accuracy of a tap

Table 2: The D-statistic of one-sample Kolmogorov-Smirnov tests for taps collected with the six target sentences in the SMS and NMR task.

Sentence/Task	S1	S2	M1	M2	L1	L2
SMS	.280	.291	.182	.119	.073	.127
NMR	.203	.148	.135	.270	.346	.288

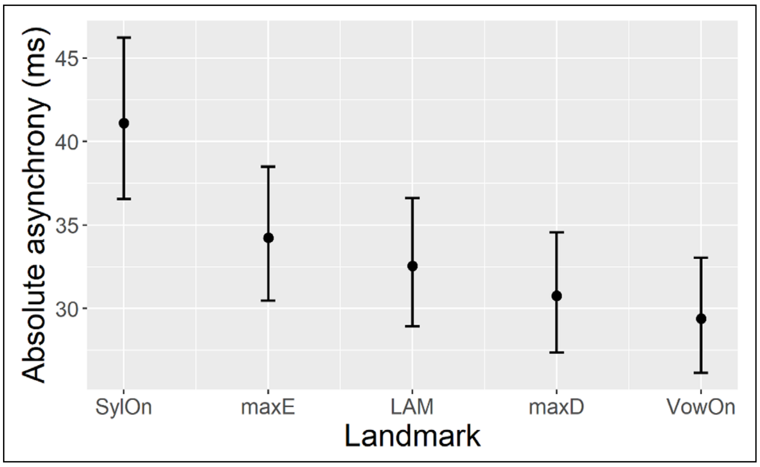

Figure 7: Estimated means and standard errors of absolute asynchrony, comparing participants' SMS performance with the five temporal landmarks—syllable onset (SylOn), local energy maximum (maxE), local amplitude maximum (LAM), maximal difference in the local energy contour (maxD), and vowel onset (VowOn). The model contained log-transformed absolute asynchronies which were back-transformed to the original ms-scale in the plot.

in the proximity of the corresponding landmark (the ±120 ms window applied across all landmarks). Here (and below), raw duration measurements in ms were logarithmically transformed to reduce or remove the skewness of the distribution that is typically observed in durational data (Baayen, 2008, pp. 31ff). Visual inspection of the estimates in **Figure 7** led to the conclusion that vowel onsets demonstrated the smallest temporal discrepancy between SMS-peak locations and the nearby landmarks. Vowel onsets were thus taken as the reference level for the pairwise comparisons with the Bonferroni-corrected α-level set to 0.0125 (0.05/4). Accordingly, syllable onsets ($t = 6.80, p < .001$) and maxE ($t = 3.15$, $p < .01$) differed significantly from the asynchronies measured with vowel onsets while LAM did not reach significance at the Bonferroni-corrected α-level of 0.0125 ($t = 2.18$, $p = 0.03$). Despite the numerical difference observed in **Figure 7**, maxD did not produce significantly longer absolute asynchronies in comparison to vowel onsets ($t = 0.91$, n.s.).

Based on the analyses above, we conclude that vowel onsets constitute the best anchor of SMS in these data. **Figure 8** displays an example of the group performance with the vowel onsets of the test sentence S1. Despite some individual variation, cumulative tapping peaks shown in the graph are temporally well aligned with the vowel onsets (indicated by vertical dashed blue lines). While the sentence-initial vowel demonstrates large negative asynchronies typical of SMS performance with a metronome (e.g., Aschersleben, 2002), all following vowels seem much less anticipated in this example, i.e., display smaller or no negative mean asynchronies.

4.3 Probability of a tap in the SMS task

A logistic mixed-effects regression was performed to test for the likelihood of a tapping peak being present (1) or absent (0) in the proximity of a vowel onset (±120 ms around the temporal landmark; see 3.7). Metrical status (i.e., the vowel being nucleus of a metrically weak, strong, or pitch-accented syllable), rise-time and rise-slope of the amplitude envelope, S2S likelihood of the sentence (high/low), and participant-specific characteristics were entered as predictors. We also tested if the order of tasks (SMS first versus NMR first) had an impact on tapping with vowels. Participant (P1–P29) and

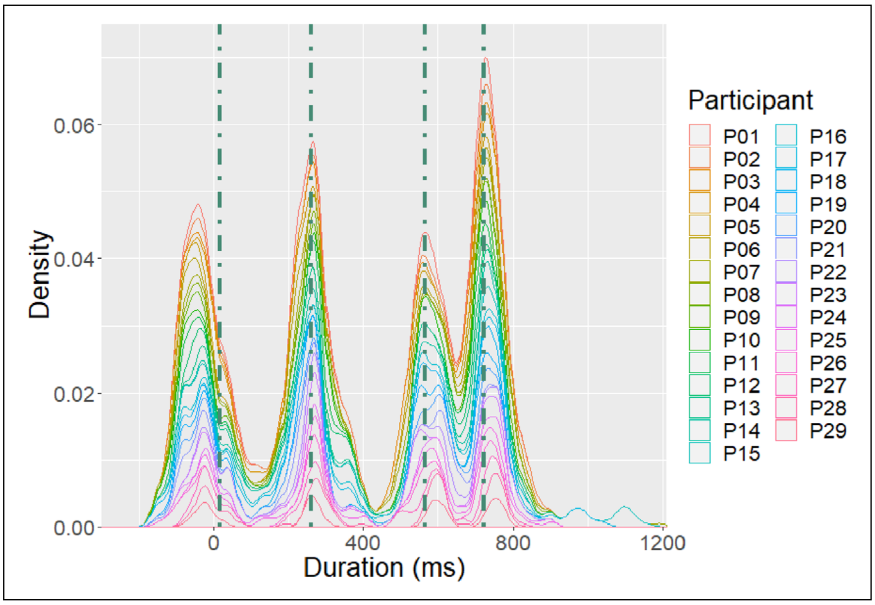

Figure 8: Accumulated tapping frequencies of 29 participants of this study, found in the experimental sentence S1 (“I wove a yarn”). Dashed vertical lines indicate vowel onsets in the acoustic signal of the sentence.

sentence (1–6) were fitted as random effects. Again, we started with a maximal random effect structure and retained those random effects that allowed the models to converge. To combat the model convergence issues of the mixed-effects logistic regressions, we changed the default optimizer (to 'bobyqa') and increased the number of iterations from default 10,000 to 100,000. Summary of the best-fit model established by the likelihood ratio tests can be found in the supplementary materials.

The best-fit model produced two main effects, including the metrical status of the vowel and S2S likelihood of the sentence (see **Table 3**). Accordingly, metrically weak vowels were less likely to attract a tap, in comparison to either metrically strong ($z = 3.28, p < .01$) or accented vowels ($z = 4.50, p < .001$). Although accented vowels were slightly more often tapped to than metrically strong but phrasally unaccented vowels, the difference between them was not significant. Sentences identified as high-transforming in previous S2S-experiments (Rathcke et al., 2018, forthcoming) were also more likely to induce a higher number of taps, forming a tapping peak in the density map around a vowel onset ($z = 2.37, p < .05$). These effect estimates are summarized in **Table 4**.

4.4 SMS accuracy

SMS accuracy was measured as absolute asynchronies between SMS peaks and vowel onsets (on a logarithmic scale), and entered linear mixed-effects modelling as the dependent variable. Again, we tested the predictive power of metrical status (i.e., the vowel being nucleus of a metrically weak, strong, or pitch-accented syllable), rise-time and rise-slope of the amplitude envelope, S2S likelihood of the sentence (high/low), and participant-specific characteristics. We further added the order of tasks (SMS first/NMR first) to check if SMS improved in those participants who first performed NMR. Individual participant (P1–P29) and sentence (1–6) were fitted as random effects. Starting with the maximal random effect structure and changing the default optimizer (to 'optimx,' John et al., 2020), random effects were iteratively removed if they produced convergence or singular-fit issues. The best-fit model established by the likelihood ratio tests is given in the supplementary materials.

Table 5 displays the best-fit model which included two factors, (1) the rise-time of the amplitude envelope around the vowel onset and (2) the musicality score of participants.

Table 3: Summary of the logistic mixed-effects model best fitting the SMS probability data.

Factor	***AIC***	***df***	χ^2	***p***
Metrical status	771.42	2	24.68	<.001
S2S likelihood	754.23	1	5.48	<.05

Table 4: Estimates of fixed effects for the SMS probability data (reference level of metrical status is weak, and reference level of S2S is high-transforming).

	Estimate	***SE***	***z***	***p***
Intercept	2.26	0.30	7.52	<.001
Metrical status: strong	1.00	0.30	3.28	<.01
Metrical status: accented	0.98	0.22	4.50	<.001
S2S: low	−0.47	0.20	−2.37	<.05

Table 5: Summary of the linear mixed-effects model best fitting the SMS accuracy data.

Factor	*Sum Sq.*	*Mean Sq.*	*df*	*F*	*P*
Log(rise-time)	7.63	7.63	1	6.23	<.05
Participant musicality	6.34	6.34	1	5.17	<.05

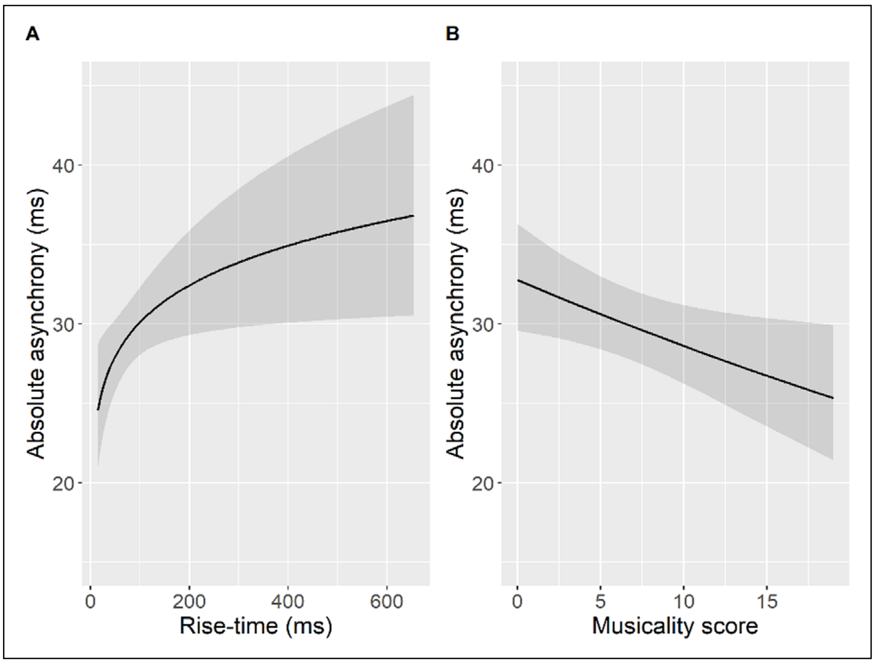

Figure 9: Estimated effects for the two factors that predict SMS accuracy: **(A)** amplitude rise-time around the vowel onset and **(B)** musical training of participants. The model contained log-transformed absolute asynchronies which were back-transformed to the original ms-scale in the predictions.

Effect estimates from the best-fit model are plotted in **Figure 9**. Vowels with shorter amplitude rise-times displayed smaller asynchronies ($t = 2.50, p < .05$). Higher levels of musical training also improved SMS accuracy ($t = –2.27, p < .05$).

4.5 Anticipation during SMS

To see if participants displayed anticipation in SMS with language, we analyzed signed asynchronies between tapping peaks and vowel onsets. Here, negative values indicated that a tap preceded (i.e., anticipated) a vowel onset. Linear mixed-effects models tested four stimulus-related predictors, including metrical status (weak, strong, or accented), rise-time and rise-slope of the amplitude envelope, S2S likelihood of the sentence (high/low), and participant-specific characteristics. As observed before, targets that occurred at the beginning of a sentence seemed more anticipated than any of the subsequent targets, i.e., they show larger negative mean asynchronies. To see how systematically this effect occurred in our data, we included the serial order of targets within a sentence as a covariate. We also fitted the order of tasks (SMS first/NMR first) as a fixed effect to see if the anticipation of the upcoming vowels is reduced after participants had experienced the sentence during the NMR-task. The model further contained two random effects: participant (P1–P29) and sentence (1–6). The maximal random effect structure initially included random slopes and was iteratively simplified if convergence or singular-fit issues persisted despite the change in the optimizer (John et al., 2020). The final model established by the likelihood ratio tests is shown in the supplementary materials.

The best-fit model retained three covariates related to the acoustic and positional properties of sentence targets (see **Table 6**). Both rise-time and rise-slope of the amplitude envelope around the vowel onset showed a strong influence.[2] More specifically, vowels with longer rise-times ($t = -4.21, p < .001$) and steeper rise-slopes ($t = -3.55, p < .001$) were more anticipated than vowels with shorter rise-times and shallower rise-slopes (see **Figure 10A–B**). As far as the serial order of a vowel in a sentence was concerned, our preliminary observations were confirmed. Each subsequent vowel showed smaller negative asynchronies and was thus less anticipated than its predecessor ($t = 2.29, p < .05$). That is, SMS accuracy increased incrementally and was particularly high for sentence-final vowels (see **Figure 10C**).

4.6 Number of repetitions in SMS

Given that our SMS task involved a total of 20 stimulus presentations, we examined participants' tapping behaviour across the repetitions. In particular, we were interested in answering the two main questions. Firstly, when did participants start tapping, and how might this have been influenced by their self-reported confidence in the personal synchronization performance? Secondly, assuming that SMS improved with practice, how many tapping cycles were needed for participants to achieve their best, stable SMS performance in this task?

On average, most participants started to tap during the third repetition of a sentence (median: 3, interquartile range: 2–3). The first tap was recorded slightly later at the very beginning of the SMS task (interquartile range: 2–4) and generally shifted to an earlier repetition cycle at the end of the SMS session (interquartile range: 2–3). Only on a few

Table 6: Summary of the linear mixed-effects model best fitting the SMS anticipation data.

Factor	Sum Sq.	Mean Sq.	df	F	P
Log(rise-time)	44253	44253	1	17.76	<.001
Rise-slope	31469	31469	1	12.63	<.001
Serial order	13242	13242	1	5.31	<.05

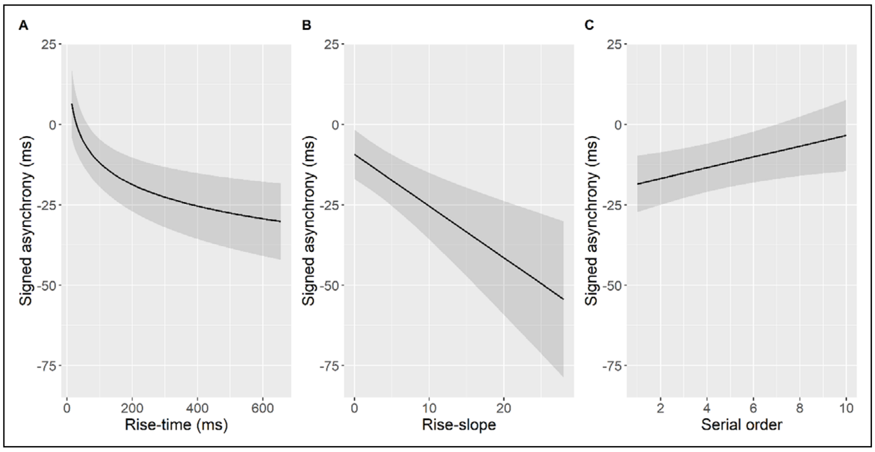

Figure 10: Model estimate plots for the three factors that best explain SMS anticipation with language: **(A)** amplitude rise-time, **(B)** amplitude rise-slope, and **(C)** serial order of a vowel in a sentence. The temporal onset of a vowel is indicated as 0 ms. Log-transformed rise-times are back-transformed to the original ms-scale.

[2] A correlation test (see supplementary materials) indicated that rise-time and rise-slope were not correlated in these data (r = –0.12, n.s.), i.e., each affected the participants' SMS-performance independently.

trials, participants started to synchronize as late as during the 10th or the 11th repetition cycle. The location of the first synchronization attempt within the loop was unaffected by the self-reported confidence ratings participants provided upon completion of the task, or by their musicality score. The order of tasks (SMS first or NMR first) did not have any impact, either.

The time-point at which tapping performance stabilized also differed across participants. **Figure 11** displays examples of the time-series data collected for the participants P02 (A) and P12 (B) during an SMS trial with the target sentence S2 ("I took the prize"). All taps collected for each participant are plotted along the x-axis where 1 demarcates the first tap recorded. If participants had tapped to every single vowel from the very first repetition of the sentence in the loop, the total number of taps would be (4 vowels × 20 repetitions =) 80, which was not the case for either participant in the example. Instead, there was a lot of individual variability. The overall number of taps available per trial differs, depending on when the participant started tapping and how many vowels they sychronized with (P02 produced more taps than P12).

The y-axis in **Figure 11** displays signed asynchronies (in ms) where 0 represents the vowel onset. The two chosen examples suggest that P02 started off tapping 10–20 ms ahead of the vocalic targets in this sentence and became consistently more accurate in synchronization with the vowel onset after s/he had produced 12 taps, while P12 started off lagging behind the vowel targets by 20–40 ms and reached a stably improved performance after s/he had tapped 17 times. Alternatively, these asynchronies could be interpreted as stable from the start of synchronization but timed with a different landmark at the beginning versus toward the end of a trial. Yet, this interpretation seems very unlikely. As shown in **Figure 2**, timing of every landmark varied quite substantially with respect to the vowel landmark. For example, maxE could occur before or after a vowel onset in two successive syllables. Such variability means that trajectories plotted in 10-A or 10-B would show little systematicity prior to the identified point of stability when synchronization with the vowel onset begins (which was clearly not the case).

These time-series data were analyzed using R-library CHANGEPOINT (Killick & Eckley, 2014). For each participant and sentence, we identified the individual point of change in the synchronization accuracy by examining global fluctuations in the mean and variance. We further calculated the number of sentence repetitions that participants required as input as well as the number of tapping cycles that participants performed until they achieved stable synchronization (as measured by the mean and variance in their signed asynchronies). In **Figure 11**, horizontal red lines show estimated means of asynchronies. The local discontinuity between the two fitted lines indicates the location of the change

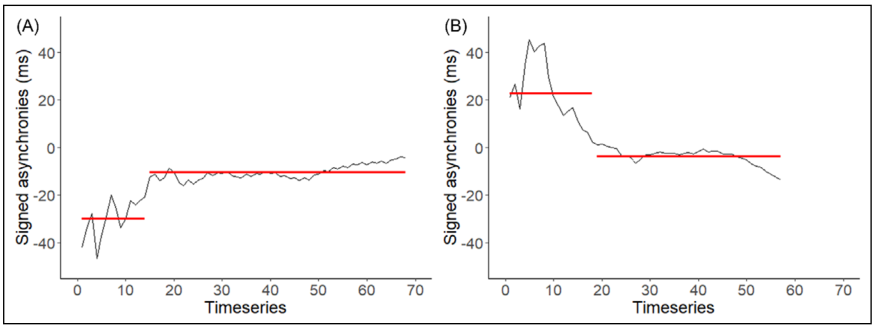

Figure 11: Time-series plots of signed asynchronies measured from vowel onset for each tap produced by participants P02 **(A)** and P12 **(B)** during their synchronization trial with the sentence S2. Black lines show raw data, red lines represent estimated means fitted by change-point analyses.

point. On average, participants performed 5 tapping cycles of each sentence (interquartile range: 3–7) until they reached the point of stability in their synchronization. Each of these tapping cycles could consist of 10–20 taps, depending on the participant's performance. None of the hypothesized predictors (confidence ratings, musicality, order of tasks) had an influence on the individually achieved point of synchronization stability.

4.7 Period tracking in SMS versus NMR

To understand how well participants were able to track the beat period in speech stimuli in the two tasks under investigation, we compared SMS and NMR in two aspects: (1) how successful the two tasks were in making participants deviate from their spontaneous tapping rates (measured by BAASTA, Dalla Bella et al., 2017); and (2) if both tasks induced convergence between participants' tapping rates and IOI of the intervocalic intervals of the test sentences. A mixed-effects regression was fitted to the dependent variable mean ITI per sentence and participant. We tested for two interactions, namely (1) between the task and the participant's unpaced spontaneous tapping rate and (2) between the task and vocalic IOI. We also fitted the order of tasks (SMS first/NMR first) as a predictor to control for a potential task order effect. Participant and sentence were defined as random effects. Again, we started with a maximal random effect structure and iteratively simplified it if the model failed to converge or produced a singular fit. A change of the default optimizer (to 'optimx,' John et al., 2020) counteracted some of the convergence issues. The likelihood ratio test helped to determine the best-fit model which is given in the supplementary materials.

Both interactions were significant in the best-fit model (see **Tables 7** and **8**). Accordingly, larger intervocalic intervals significantly increased ITI, with a positive linear relationship in both tasks ($t = 4.76, p < .01$). However, an increase in vocalic IOI showed a notably smaller effect on the increase of ITI in NMR than in SMS ($t = -2.63, p < .01$). While individual tapping rates did not have any effect on ITI obtained in the SMS task, ITI in the NMR task tended to show longer durations if participants' spontaneous tapping tempo also had longer ITI ($t = 2.08, p < .05$). These findings demonstrated that period tracking was present in both tasks, though it had a subtler effect in NMR whose ITI drifted toward the participant's preferred individual tapping tempo in the absence of a simultaneous auditory signal. Crucially, an ITI regularization could also be observed in

Table 7: Summary of the linear mixed-effects model best fitting the ITI data.

Factor	*Sum Sq.*	*Mean Sq.*	*df*	*F*	*p*
*task * log(vocalic IOI)*	.23	.23	1	6.93	<.01
*task * log(spontaneous tapping rate)*	.15	.15	1	4.31	<.05

Table 8: Estimates of fixed effects for the ITI data (reference level of task is SMS).

	Estimate	*SE*	*t*	*p*
Intercept	1.65	1.12	1.48	n.s.
Task	0.38	0.60	0.64	n.s.
log(vocalic IOI)	0.61	0.13	4.76	<.01
log(spontaneous tapping rate)	0.13	0.14	0.96	n.s.
*task * log(vocalic IOI)*	−0.22	0.09	−2.63	<.01
*task * log(spontaneous tapping rate)*	0.13	0.06	2.08	<.05

the NMR task. According to an additional model fit to the CV of ITI (see **Tables 9** and **10**), this dependent variable differed significantly across the two tasks, showing that NMR led to less variability across ITI than SMS did ($t = -2.55$, $p < .05$). That is, taps were paced more regularly in NMR.

To test whether or not the above effects were merely consequences of a self-sustained, repeated movement in the SMS task, in contrast to the NMR task which generally led to fewer taps, we compared SMS and NMR data collected on the very first tapping trial. However, these comparisons produced comparable results.

5. Discussion

The present study was conducted to examine the suitability of two movement-based paradigms—synchronization (SMS) and reproduction (NMR)—for the study of rhythm perception in natural language, and to provide empirical evidence on the settings of such a paradigm. Below, we discuss the results with reference to our original research aims and hypotheses and comment on how they compare to previous research with other types of auditory stimuli.

5.1 Suitability of motor paradigms for linguistic rhythm research

The present study demonstrates that motor paradigms are suitable tools to investigate rhythm perception in language. Our results suggest that particularly SMS is informative and better suited than NMR to support rhythm research in language. Our version of SMS with natural language produced consistent patterns of synchronization with vowel onsets, thus replicating our previous results with simpler verbal stimuli (Rathcke et al., 2019). In our version of NMR, a certain level of period tracking could also be observed. However, the NMR results showed weaker relations to linguistic stimuli and an overall trend to converge towards participants' spontaneous tapping rates. Alongside this shift toward individually preferred tapping rates, the overall variability of inter-tap intervals was reduced, suggesting that participants tapped more regularly. This result parallels previous research with a similar motor reproduction paradigm (see Donovan & Darwin, 1979; Scott et al., 1985), as well as with other paradigms that involve beat tracking during speech production (Jungers, Palmer, & Speer, 2002; Port et al., 1995).

When movement is not synchronized in time with an auditory signal, temporal regularization of ITI prevails in linguistic stimuli, though not in other types of stimuli (see Donovan & Darwin, 1979; Scott et al., 1985). Such regularization is likely to arise due to a high level of rhythmic complexity in language that lacks temporal isochrony (Dauer, 1983; Pointon, 1980; Roach, 1982; Uldall, 1971; van Santen & Shih, 2000) while employing a highly intricate hierarchy of nested constituents (cf. Nespor & Vogel, 1986; Selkirk, 1984) and prominence alternations (Liberman & Prince, 1977; Prince, 1983). In the context of such complexity, the superiority of SMS is in line with previous research

Table 9: Summary of the linear mixed-effects model best fitting the CV of ITI data.

Factor	***Sum Sq.***	***Mean Sq.***	***df***	***F***	***p***
task	0.13	0.13	1	6.52	<.05

Table 10: Estimates of fixed effects for the CV of ITI data (reference level of task is SMS).

	Estimate	***SE***	***t***	***p***
Intercept	0.47	0.04	12.21	<.001
task: NMR	–0.05	0.02	–2.55	<.05

that demonstrated that movement along a complex rhythm facilitates the discovery of its beat (Su & Pöppel, 2012). Such advantage of a synchronized movement possibly arises due to an enhanced internal representation of an auditory rhythm that accompanies movement (Chemin, Mouraux, & Nozaradan, 2014). In contrast, movement without a concurrent auditory signal relies more heavily on an internal representation of temporal patterns, thus increasing working memory load and the associated processing cost (Repp & Su, 2013). Recent evidence suggests that tapping without a concurrent auditory signal might be a more demanding task than SMS (Koch, Oliveri, & Caltagirone, 2009; Lewis, Wing, Pope, Praamstra, & Miall, 2004). More specifically, reproduction or continuation tasks with a metronome, whose settings are quite similar to the NMR task with language in our experiment, seem to be placing higher demands on both working memory (Jantzen, Oullier, Marshall, Steinberg, & Kelso, 2007; Koch et al., 2009) and motor timing abilities (Serrien, 2008). In contrast, SMS is likely to enhance basic perceptual abilities (Valdesolo, Ouyang, & DeSteno, 2010). These findings provide some explanations as to why participants might have tended to converge toward their spontaneous tapping rates in the NMR but not in the SMS task, as well as why SMS might be superior to NMR in the context of beat perception and rhythmic processing in language.

5.2 SMS anchors and acoustic influences on SMS

As hypothesized, SMS with language can produce systematic responses to the temporal structure of natural spoken sentences. The present study tested five potential anchors of synchronization, including onsets of linguistic units (syllables, vowels) and acoustic landmarks (local maxima of energy or amplitude, local changes in the smoothed energy contour). The shortest asynchronies were observed between SMS peaks and nearby vowel onsets, followed by the moment of the fastest change in the smoothed energy contour (maxD) and the local amplitude maximum (LAM). The numerical difference in synchronization accuracy with these three landmarks was not significant at the Bonferroni-corrected α-level, though vowel onsets produced smallest asynchronies. In contrast, anchoring taps to syllable onsets and local maxima in the energy contours led to a significantly deteriorated accuracy in the participants' performance. These results indicate that vowel onsets seem to reliably attract taps not only in simple verbal prompts (Rathcke et al., 2019), but also in complex temporal patterns of natural spoken language. Recent evidence from naturally evolved drummed languages like Amazonian Bora further corroborates this finding. In drummed Bora, rhythmic units have also been shown to consistently match intervocalic intervals, irrespective of syllable complexity (Seifart, Meyer, Grawunder, & Dentel, 2018).

Vowels play an important role in shaping the trajectory of the sonority contour in speech signal, frequently constituting local sonority peaks (Morgan & Fosler-Lussier, 1998; Wang & Narayanan, 2007). The sonority contour reflects variable degrees of energy emanating from the vocal tract during speech production and is particularly high for open vowels. The cyclical production of vowel gestures in connected speech has been previously highlighted as one of the potential reasons why spoken language might be rhythmic in nature (Fowler, 1983; Fowler & Tassinary, 1981). Local fluctuations in signal sonority related to vowel acoustics have also been argued to guide speech segmentation and to assist first language acquisition (Räsänen, Doyle, & Frank, 2018). It thus does not seem surprising that beat perception (at least in English) locks on to vocalic and not to syllabic onsets, though more research is needed to determine if beat perception in English involves tracking of vowels per se or rather tracking of nuclear constituents within larger units such as syllables.

Importantly, SMS-performance with the linguistic stimuli of the present study demonstrated anticipation of vowel targets (as indicated by negative mean asynchrony, cf. Aschersleben, 2002). Especially the first synchronization target within a sentence, i.e., the target that occurred after an acoustic silence, showed larger negative asynchronies and was thus more anticipated than all subsequent vocalic targets. This finding is in keeping with the existing evidence on the properties of SMS with other types of auditory stimuli. Anticipation seems to be a characteristic of non-musicians' syncronizing with a metronome signal where regular auditory prompts are interspersed with silences (Aschersleben, 2002; Repp, 2005). The negative mean asynchrony is reduced, or even disappears completely in more complex rhythmic contexts where synchronization targets are not separated by silences, e.g., in music (Thaut, Rathbun, & Miller, 1997; Wohlschläger & Koch, 2000).

SMS in the present study was more precise with those vowels that had shorter rise-times. Unfortunately, rise-time (and also rise-slope) of an amplitude envelope is a complex acoustic measure that is influenced by many aspects of speech production. These properties can change depending on the manner and the place of articulation of the onset consonant(s), levels of syllabic prominence (weak, strong, accented, or emphatic), degrees of coarticulation, and syllable reduction. This underlying complexity impedes a meaningful interpretation of the rise-time contribution to beat perception (cf. Peelle & Davis, 2012), though interestingly, once again we find parallels to SMS with a metronome where shorter rise-times of tones have been shown to improve synchronization accuracy (Vos, van Kruysbergen, & Mates, 1995).

5.3 SMS sensitivity to the metrical structure

One of the crucial findings of the present study is that SMS is sensitive to the metrical structure of spoken sentences. In the present study, native English participants were more likely to tap to metrically strong than metrically weak syllables. These results are somewhat comparable with the findings by Allen (1972: 89) who concluded that English participants tend to "tap before the nuclear vowels of rhythmically accented syllables." Given that the prosodic system of English incorporates word-level stress and sentence-level alternations of strong and weak syllables (Liberman & Prince, 1977; Prince, 1983), it is highly likely that the prosodic system of a listener's native language plays a major role in inducing their feeling of a beat in speech and that rhythm perception in language might be a constructive perceptual process.

5.4 On the role of repetition in SMS

In our view, repetition is a crucial aspect of the success of the SMS-paradigm with language. Despite being a laboratory task, looping resonates with the idea that linguistic rhythm arises on large temporal scales through repeated experience with one's native language. Unlike other approaches that rely on special cases of language use like poetry, mantra, or chant (cf. Cummins, 2012) or on short, metrical, or regularized speech (Lidji et al., 2011), looping can be applied to any natural spoken utterances, leading to an increased ecological validity of the proposed paradigm (cf. Allen, 1975). The present SMS method creates a unique situation for unlocking the rhythmic structure of natural, unmanipulated language while bypassing other mechanisms of sentence processing (cf. Rathcke et al., 2018).

In the present experiment, most participants appeared to have created an internal representation of the sentence beat structure after two repetitions and could start synchronizing during the third presentation cycle of the sentence. As our results indicate, a total of three repetitions used in previous research (Lidji et al., 2011) is not quite sufficient to fully capture stable SMS patterns. For example, the Kernel density fitting procedure relies on the presence of at least two events and can lead to missing data in shorter sentences or in participants who might require longer to entrain. Given the results of our time-series analyses, we recommend using at least 10 repetitions of a sentence to produce stable, consistent, and representative patterns across individual participants (e.g., Gérard & Rosenfeld, 1995; Pressing & Jolley-Rogers, 1997; Repp, 2005; Repp & Penel, 2002).

Finally, the results of the present study exclude the possibility that the speech-to-song illusion interferes with SMS patterns in a significant way. Both high- and low-transforming sentences tested in the present study produced similar results in terms of synchronization accuracy and targets. The only difference between high- and low-transforming sentences consisted in the overall number of recorded taps. Accordingly, sentences that led to more S2S transformations (Rathcke et al., 2018, forthcoming) also induced more taps. The reason for this effect is as yet not quite clear, though it might be related to a higher level of the overall signal sonority in the high-transforming set (Rathcke et al., 2018, forthcoming). It is, however, clear that the speech-to-song illusion is not a core prerequisite for a successful application of the SMS paradigm to language, which is in line with our previous work showing that S2S transformations rely more heavily on pitch- than on time-related features of speech (Falk et al., 2014; Falk & Rathcke, 2010).

5.5 Individual variability in SMS with language

As expected, we found some individual variation in the SMS task. Some aspects of this variation, e.g., SMS accuracy, could be partially explained by individually varying levels of musical training. Participants produced lower asynchronies if they had higher levels of musical training and experience (which included playing an instrument, singing, and dancing). Musical sophistication is also known to decrease error and variability in synchronization with a metronome in non-professional musicians (e.g., Gérard & Rosenfeld, 1995; Pressing & Jolley-Rogers, 1997; Repp & Penel, 2002). However, measures of general synchronization performance with music and metronome employed in the present study (BAASTA, Dalla Bella et al., 2017) did not help to explain individual variability in SMS with speech. Reasons for this might be multiple (e.g., different mechanisms of rhythm perception in isochronous versus non-isochronous signals or in non-speech versus speech) and require further investigation.

6. Conclusions and outlook

The two movement-based paradigms that were elaborated and tested in the present study view language rhythm as a consequence of general internal timekeeping mechanisms that allow us to synchronize, anticipate, and adapt our behaviour in response to an external stimulus (Repp, 2005). We showed that SMS performance can be successfully used with linguistic stimuli and that SMS patterns resemble well-documented findings of SMS with metronome and music (Aschersleben, 2002; Repp, 2005; Repp & Su, 2013) in listeners of various degrees of musical training (Gérard & Rosenfeld, 1995; Pressing & Jolley-Rogers, 1997; Repp & Penel, 2002). Like music, beat perception in language can be linked to temporal expectancy and prediction of upcoming events, and we showed that such expectancies can be elicited during SMS with spoken sentences presented in a loop. Our study further demonstrated that vowels constitute the most likely rhythmic anchors in language, though more work is required with diverse languages to establish if the present finding generalizes beyond English. An alternative movement task, NMR, showed some potential to engage listeners' capacity to extract rhythmic patterns from speech, though it also tended to evoke motor regularization arising from preferred individual finger-tapping rates.

In sum, the present study demonstrates that natural language can entrain movement. Our setting of the SMS paradigm is a valid experimental paradigm to study beat tracking and rhythm extraction in linguistic stimuli of different degrees of complexity, and can be used in future work to answer many open questions on rhythm perception and cognition across prosodically diverse languages.

Data Accessibility Statement

The data and materials for the experiment reported here can be made available upon request. Speech materials and annotations can be downloaded from OSF (see https://osf.io/3dh4m/). The experiment was not preregistered.

Additional File

The additional file for this article can be found as follows:

- **Supplementary Material.** Summary of statistical analyses and best-fit models. DOI: https://doi.org/10.5334/labphon.248.s1

Acknowledgements

This research was supported by a research grant from the Leverhulme Trust (RPG-2017-306) to the first author. We would like to express our deep gratitude to Georg Lohfink for his help with the equipmental set-up, testing, and data collection at Kent Linguistics Lab, and to Roger Dean for his help with the time-series analyses reported in 4.6. Our thanks also go to Mona Franke (University of Montreal) who prepared the annotations of vowel onsets for the cross-validation of the annotator agreement reported in 3.2. We are further indebted to Naeem Komeilipoor, Mélody Blais, and Simon Rigoulot from the Centre for Research on Brain, Language and Music Montreal for their support with the preparation of the BAASTA data. Discussions at IPS Munich, MARCS Sydney, and CLAS Macquarie helped to develop the ideas presented in the paper while the input from two anonymous reviewers and the associate editor helped to improve the manuscript. Our special thanks go to all the tireless tappers who participated in the experiment.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

Tamara Rathcke designed the experiment, annotated the sentences to identify the syllable and vowel onsets, devised the acoustic analyses of the sentences, conducted the statistical analyses, was the primary author of all sections of the manuscript, and dealt with the manuscript revisions. Chia-Yuan Lin set up the experiment, collected the tapping data, conducted the pre-processing of the data, and performed the acoustic analyses of the sentences. Simone Falk coordinated annotations of the second annotator and offered comments on the manuscript. Simone Dalla-Bella guided the development of the pre-processing procedure of the tapping data and commented on the manuscript. The study evolved from extensive discussions between TR, SF, and SDB.

References

Abercrombie, D. (1967). *Elements of general phonetics* (E. U. Press, ed.).

Allen, G. D. (1972). The Location of Rhythmic Stress Beats in English: An Experimental Study I. *Language and Speech, 15*(1), 72–100. DOI: https://doi.org/10.1177/002383097201500110

Allen, G. D. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*. DOI: https://doi.org/10.1016/S0095-4470(19)31351-8

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica, 66*(1–2), 46–63. DOI: https://doi.org/10.1159/000208930

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*(3), 351–373. DOI: https://doi.org/10.1016/j.wocn.2012.02.003

Arvaniti, A., & **Rodriquez, T.** (2013). The role of rhythm class, speaking rate, and F0 in language discrimination. *Laboratory Phonology, 4*(1), 7–38. DOI: https://doi.org/10.1515/lp-2013-0002

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition*, *48*(1), 66–79. DOI: https://doi.org/10.1006/brcg.2001.1304

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511801686

Barr, D. J., Levy, R., Scheepers, C., & **Tily, H. J.** (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Barry, W., Andreeva, B., & **Koreman, J.** (2009). Do rhythm measures reflect perceived rhythm? *Phonetica*, *66*(1–2), 78–94. DOI: https://doi.org/10.1159/000208932

Bates, D., Mächler, M., Bolker, B., & **Walker, S.** (2015). Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology*, *6*, 145–238. DOI: https://doi.org/10.2307/1410948

Borchers, H. W. (2018). *pracma: Practical Numerical Math Functions*. https://cran.r-project.org/package = pracma

Bouvet, C. J., Varlet, M., Dalla Bella, S., Keller, P. E., & **Bardy, B. G.** (2019). Accent-induced stabilization of spontaneous auditory–motor synchronization. *Psychological Research*. DOI: https://doi.org/10.1007/s00426-019-01208-z

Brett, M., & **Grahn, J. A.** (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906. papers2://publication/uuid/9405ACE2-E77B-48A6-9CAC-98818CC2CB87. DOI: https://doi.org/10.1162/jocn.2007.19.5.893

Cardany, A. B. (2013). Nursery Rhymes in Music and Language Literacy. *General Music Today*, *26*(2), 30–36. DOI: https://doi.org/10.1177/1048371312462869

Chemin, B., Mouraux, A., & **Nozaradan, S.** (2014). Body Movement Selectively Shapes the Neural Representation of Musical Rhythms. *Psychological Science*, *25*(12), 2147–2159. DOI: https://doi.org/10.1177/0956797614551161

Cooper, A. M., Whalen, D. H., & **Fowler, C. A.** (1986). P-centers are unaffected by phonetic categorization. *Perception & Psychophysics*, *39*(3), 187–196. DOI: https://doi.org/10.3758/BF03212490

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, *37*(1), 16–28. DOI: https://doi.org/10.1016/j.wocn.2008.08.003

Cummins, F. (2012). Looking for rhythm in speech. *Empirical Musicology Review*, *7*(1), 28–35. DOI: https://doi.org/10.18061/1811/52976

Cummins, F., & **Port, R.** (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171. DOI: https://doi.org/10.1006/jpho.1998.0070

Dalla Bella, S., Białuńska, A., & **Sowiński, J.** (2013). Why Movement Is Captured by Music, but Less by Speech: Role of Temporal Regularity. *PLoS ONE*, *8*(8), 1–16. DOI: https://doi.org/10.1371/journal.pone.0071945

Dalla Bella, S., Farrugia, N., Benoit, C. E., Begel, V., Verga, L., Harding, E., & **Kotz, S. A.** (2017). BAASTA: Battery for the Assessment of Auditory Sensorimotor and Timing Abilities. *Behavior Research Methods*, *49*(3), 1128–1145. DOI: https://doi.org/10.3758/s13428-016-0773-6

Dauer, R. M. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51–62. https://psycnet.apa.org/record/1983-29886-001. DOI: https://doi.org/10.1016/S0095-4470(19)30776-4

De Jong, K. J. (1994). The correlation of P-center adjustments with articulatory and acoustic events. *Perception & Psychophysics*, *56*(4), 447–460. DOI: https://doi.org/10.3758/BF03206736

Dellwo, V., & **Wagner, P.** (2003). Relations between language rhythm and speech rate. *Proceedings of International Congress of Phonetic Science*, 471–474. https://pub.uni-bielefeld.de/record/1785384

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, *29*(2), 217–230. DOI: https://doi.org/10.1006/jpho.2001.0138

Deutsch, D. (2003). *Phantom words, and other curiosities.* Philomel Records.

Deutsch, D., Henthorn, T., & **Lapidis, R.** (2011). Illusory transformation from speech to song. *The Journal of the Acoustical Society of America*, *129*(4), 2245–2252. DOI: https://doi.org/10.1121/1.3562174

Donovan, A., & **Darwin, C.** (1979). The perceived rhythm of speech. *Proceedings of the 9th International Congress of Phonetic Sciences*, *2*, 268–274.

Eerola, T., & **Toiviainen, P.** (2004). MIDI Toolbox: MATLAB Tools for Music Research. In *University of Jyväskylä: Kopijyvä, Jyväskylä, Finland.* http://www.jyu.fi/musica/miditoolbox/

Engström, D. A., Kelso, J. A. S., & **Holroyd, T.** (1996). Reaction-anticipation transitions in human perception-action patterns. *Human Movement Science*, *15*(6), 809–832. DOI: https://doi.org/10.1016/S0167-9457(96)00031-0

Falk, S., & **Dalla Bella, S.** (2016). It is better when expected: Aligning speech and motor rhythms enhances verbal processing. *Language, Cognition and Neuroscience*, *31*(5), 699–708. DOI: https://doi.org/10.1080/23273798.2016.1144892

Falk, S., & **Rathcke, T.** (2010). On the Speech-To-Song Illusion: Evidence from German. *Speech Prosody*, *169*. https://www.isca-speech.org/archive/sp2010/papers/sp10_169.pdf

Falk, S., Rathcke, T., Dalla Bella, S., & **Bella, S. D.** (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1491–1506. DOI: https://doi.org/10.1037/a0036858

Falk, S., Volpi-Moncorger, C., & **Dalla Bella, S.** (2017). Auditory-motor rhythms and speech processing in French and German listeners. *Frontiers in Psychology*, *8*(MAR), 1–14. DOI: https://doi.org/10.3389/fpsyg.2017.00395

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General*, *112*(3), 386–412. DOI: https://doi.org/10.1037/0096-3445.112.3.386

Fowler, C. A., & **Tassinary, L.** (1981). Natural measurement criteria for speech: The anisochrony illusion. In J. Long & A. Baddeley (Eds.), *Attention and performance, IX* (pp. 521–535). Erlbaum.

Gérard, C., & **Rosenfeld, M.** (1995). Pratique musicale et régulations temporelles. *L'année Psychologique*, *95*(4), 571–591. DOI: https://doi.org/10.3406/psy.1995.28856

Goswami, U., & **Leong, V.** (2013). Speech rhythm and temporal structure: Converging perspectives. *Laboratory Phonology*, *4*(1), 67–92. DOI: https://doi.org/10.1515/lp-2013-0004

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & **Scott, S. K.** (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(16), 10911–10916. DOI: https://doi.org/10.1073/pnas.122368599

Grabe, E., & **Low, E. L.** (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology* (pp. 515–546). Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110197105

Grahn, J. A., & **Rowe, J. B.** (2009). Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat perception. *Journal of Neuroscience*, *29*(23), 7540–7548. DOI: https://doi.org/10.1523/JNEUROSCI.2018-08.2009

Hommel, B. (2015). The theory of event coding (TEC) as embodied-cognition framework. *Frontiers in Psychology*, *6*, 1318. DOI: https://doi.org/10.3389/fpsyg.2015.01318

Iversen, J. R., & **Patel, A. D.** (2008). The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. In K. Miyazaki, M. Adachi, Y. Hiraga, Y. Nakajima & M. Tsuzaki (Eds.), *Proceedings of the 10th International Conference on Music Perception & Cognition* (pp. 465–468). Causal Productions.

Jantzen, K. J., Oullier, O., Marshall, M., Steinberg, F. L., & **Kelso, J. A. S.** (2007). A parametric fMRI investigation of context effects in sensorimotor timing and coordination. *Neuropsychologia*, *45*(4), 673–684. DOI: https://doi.org/10.1016/j.neuropsychologia.2006.07.020

John, N. C., Varadhan, R., & **Gabor, G.** (2020). *Package "optimx."* (pp. 1–87).

Jungers, M. K., Palmer, C., & **Speer, S. R.** (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing*, *2*(614), 21–35.

Killick, R., & **Eckley, I. A.** (2014). Changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, *58*(3), 1–19. DOI: https://doi.org/10.18637/jss.v058.i03

Koch, G., Oliveri, M., & **Caltagirone, C.** (2009). Neural networks engaged in milliseconds and seconds time processing: Evidence from transcranial magnetic stimulation and patients with cortical or subcortical dysfunction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1907–1918. DOI: https://doi.org/10.1098/rstb.2009.0018

Kohler, K. J. (2009). Rhythm in speech and language: A new research paradigm. *Phonetica*, *66*(1–2), 29–45. DOI: https://doi.org/10.1159/000208929

Kuznetsova, A., Brockhoff, P. B., & **Christensen, R. H. B.** (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Large, E. W., & **Jones, M. R.** (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159. DOI: https://doi.org/10.1037/0033-295X.106.1.119

Large, E. W., & **Palmer, C.** (2002). Perceiving temporal regularity in music. *Cognitive Science*, *26*(1), 1–37. DOI: https://doi.org/10.1016/S0364-0213(01)00057-X

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*(3), 253–263. DOI: https://doi.org/10.1016/S0095-4470(19)31139-8

Leong, V., & **Goswami, U.** (2014). Assessment of rhythmic entrainment at multiple timescales indyslexia: Evidence for disruption to syllable timing. *Hearing Research*, *308*, 141–161. DOI: https://doi.org/10.1016/j.heares.2013.07.015

Lewis, P. A., Wing, A. M., Pope, P. A., Praamstra, P., & **Miall, R. C.** (2004). Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. *Neuropsychologia*, *42*(10), 1301–1312. DOI: https://doi.org/10.1016/j.neuropsychologia.2004.03.001

Liberman, M., & **Prince, A.** (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry*, *8*(2), 249–336. DOI: https://doi.org/10.2307/4177987

Lidji, P., Palmer, C., Peretz, I., & **Morningstar, M.** (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin and Review*, *18*(6), 1035–1041. DOI: https://doi.org/10.3758/s13423-011-0163-0

ling Low, E., Grabe, E., & **Nolan, F.** (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, *43*(4), 377–401. DOI: https://doi.org/10.1177/00238309000430040301

Liu, X., Wang, S., Yianni, J., Nandi, D., Bain, P. G., Gregory, R., Stein, J. F., & **Aziz, T. Z.** (2008). The sensory and motor representation of synchronized oscillations in the globus pallidus in patients with primary dystonia. *Brain*, *131*(6), 1562–1573. DOI: https://doi.org/10.1093/brain/awn083

Lüdecke, D. (2019). *sjPlot: Data Visualization for Statistics in Social Science*. DOI: https://doi.org/10.5281/zenodo.1308157

Madison, G. (2014). Sensori-motor synchronisation variability decreases as the number of metrical levels in the stimulus signal increases. *Acta Psychologica*, *147*, 10–16. DOI: https://doi.org/10.1016/j.actpsy.2013.10.002

Mairano, P., Santiago, F., & **Romano, A.** (2015). Cross-linguistic differences between accented vs. unaccented vowel durations. *International Congress on Phonetic Sciences ICPhS*. https://halshs.archives-ouvertes.fr/halshs-01440315/

Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics*, *30*(3), 247–256. DOI: https://doi.org/10.3758/BF03214280

Margulis, E. H., & **Simchy-Gross, R.** (2016). Repetition enhances the musicality of randomly generated tone sequences. *Music Perception*, *33*(4), 509–514. DOI: https://doi.org/10.1525/mp.2016.33.4.509

Mates, J., Müller, U., Radil, T., & **Pöppel, E.** (1994). Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience*, *6*(4), 332–340. DOI: https://doi.org/10.1162/jocn.1994.6.4.332

Merker, B. (2000). Synchronous chorusing and human origins. In N. L. Wallin, B. Merker & S. Brown (Eds.), *The origins of music* (pp. 315–327). MIT. http://www.biolinguagem.com/ling_cog_cult/merker_2009_synchronouschorusing_humanorigins.pdf

Miller, M. (1984). On the perception of rhythm. *Journal of Phonetics*, *12*, 75–83. https://scholar.google.com/scholar?hl = zh-TW&as_sdt = 0%2C5&as_ylo = 1970&as_yhi = 2000&q = On + the + perception + of + rhythm + miller + 1984&btnG = . DOI: https://doi.org/10.1016/S0095-4470(19)30852-6

Morgan, N., & **Fosler-Lussier, E.** (1998). Combining multiple estimators of speaking rate. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings*, *2*, 729–732. DOI: https://doi.org/10.1109/ICASSP.1998.675368

Morillon, B., Schroeder, C. E., & **Wyart, V.** (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications*, *5*, 1–9. DOI: https://doi.org/10.1038/ncomms6255

Morton, J., Marcus, S., & **Frankish, C.** (1976). Perceptual centers (P-centers). *Psychological Review*, *83*(5), 405–408. DOI: https://doi.org/10.1037/0033-295X.83.5.405

Nespor, M., & **Vogel, S. E.** (1986). *Prosodic Phonology*. Foris.

Nolan, F., & **Jeon, H. S.** (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130396. DOI: https://doi.org/10.1098/rstb.2013.0396

Park, J. E. (2017). Apraxia: Review and update. In *Journal of Clinical Neurology (Korea)*, *13*(4), 317–324. DOI: https://doi.org/10.3988/jcn.2017.13.4.317

Patel, A. D., & **Iversen, J. R.** (2014). The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*(May), 1–14. DOI: https://doi.org/10.3389/fnsys.2014.00057

Patel, A. D., Löfqvist, A., & **Naito, W.** (1999). The acoustics and kinematics of regularly timed speech: A database and method for the study of the p-center problem.

Proceedings of the 14th International Congress of Phonetic Sciences, 1, 405–408. https://pdfs.semanticscholar.org/28c8/b08e2f68db0261a2fc17e6df59d27750967b.pdf

Peelle, J. E., & **Davis, M. H.** (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3*(SEP), 1–17. DOI: https://doi.org/10.3389/fpsyg.2012.00320

Pointon, G. E. (1980). Is Spanish really syllable-timed? *Journal of Phonetics, 8*(1), 293–304. https://eric.ed.gov/?id=EJ236834. DOI: https://doi.org/10.1016/S0095-4470(19)31479-2

Polak, R., **Jacoby, N.**, **Fischinger, T.**, **Goldberg, D.**, **Holzapfel, A.**, & **London, J.** (2018). Rhythmic Prototypes Across Cultures. *Music Perception: An Interdisciplinary Journal, 36*(1), 1–23. DOI: https://doi.org/10.1525/mp.2018.36.1.1

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences, 1*(2), 56–61. DOI: https://doi.org/10.1016/S1364-6613(97)01008-5

Port, R., **Cummins, F.**, & **Gasser, M.** (1995). A Dynamic Approach to Rhythm in Language: Toward a Temporal Phonology. In B. Luka & B. Need (Eds.), *Proceedings of the Chicago Linguistic Society* (pp. 375–397). University of Chicago, Department of Linguistics. http://arxiv.org/abs/cmp-lg/9508007

Pressing, J., & **Jolley-Rogers, G.** (1997). Spectral properties of human cognition and skill. *Biological Cybernetics, 76*(5), 339–347. DOI: https://doi.org/10.1007/s004220050347

Prince, A. S. (1983). Relating to the Grid. *Linguistic Inquiry, 14*(1), 19–100. DOI: https://doi.org/10.2307/4178311

Ramus, F., **Nespor, M.**, & **Mehler, J.** (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265–292. DOI: https://doi.org/10.1016/S0010-0277(99)00058-X

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

Rathcke, T., **Lin, C.-Y.**, **Falk, S.**, & **Dalla Bella, S.** (2019). When language hits the beat: Synchronising movement to simple tonal and verbal stimuli. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia* (pp. 1505–1509). Australasian Speech Science and Technology Association Inc. https://www.researchgate.net/profile/Tamara_Rathcke/publication/332060854_When_language_hits_the_beat_Synchronising_movement_to_simple_tonal_and_verbal_stimuli/links/5c9d11b292851cf0ae9da23e/When-language-hits-the-beat-Synchronising-movement-to-simple-ton

Rathcke, T. V., **Falk, S.**, & **Dalla Bella, S.** (2018). Linguistic structure and listener characteristics modulate the "speech-to-song illusion." *15th International Conference on Music Perception and Cognition.*

Rathcke, T. V., & **Smith, R. H.** (2015a). Speech timing and linguistic rhythm: On the acoustic bases of rhythm typologies. *The Journal of the Acoustical Society of America, 137*(5), 2834–2845. DOI: https://doi.org/10.1121/1.4919322

Rathcke, T. V., & **Smith, R. H.** (2015b). Rhythm class perception by expert phoneticians. In T. S. C. for Icp. 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1–5). London: International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0403.pdf

Ravignani, A., **Dalla Bella, S.**, **Falk, S.**, **Kello, C. T.**, **Noriega, F.**, & **Kotz, S. A.** (2019). Rhythm in speech and animal vocalizations: a cross-species perspective. In *Annals of the New York Academy of Sciences*. DOI: https://doi.org/10.1111/nyas.14166

Repp, B. H. (2003). Rate Limits in Sensorimotor Synchronization With Auditory and Visual Sequences: The Synchronization Threshold and the Benefits and Costs of

Interval Subdivision. *Journal of Motor Behavior, 35*(4), 355–370. DOI: https://doi.org/10.1080/00222890309603156

Repp, B. H. (2004). On the nature of phase attraction in sensorimotor synchronization with interleaved auditory sequences. *Human Movement Science, 23*(3–4 SPE. ISS.), 389–413. DOI: https://doi.org/10.1016/j.humov.2004.08.014

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review, 12*(6), 969–992. DOI: https://doi.org/10.3758/BF03206433

Repp, B. H., & **Penel, A.** (2002). Auditory Dominance in Temporal Processing: New Evidence from Synchronization with Simultaneous Visual and Auditory Sequences. *Journal of Experimental Psychology: Human Perception and Performance, 28*(5), 1085–1099. DOI: https://doi.org/10.1037/0096-1523.28.5.1085

Repp, B. H., & **Su, Y.-H. H.** (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin and Review, 20*(3), 403–452. DOI: https://doi.org/10.3758/s13423-012-0371-2

Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (Ed.), *Linguistics Controversies* (pp. 73–79). Edward Arnold. http://w3.salemstate.edu/%7B~%7Djaske/courses/readings/On_the_distinction_between_stress-timed_and_syllable-timed_languages_By_Peter_Roach.pdf

Rowland, J., Kasdan, A., & **Poeppel, D.** (2019). There is music in repetition: Looped segments of speech and nonspeech induce the perception of music in a time-dependent manner. *Psychonomic Bulletin and Review, 26*(2), 583–590. DOI: https://doi.org/10.3758/s13423-018-1527-5

Scott, D. R., Isard, S. D., & **de Boysson-Bardies, B.** (1985). Perceptual isochrony in {E}nglish and {F}rench. *Journal of Phonetics, 13*, 155–162. DOI: https://doi.org/10.1016/S0095-4470(19)30743-0

Seifart, F., Meyer, J., Grawunder, S., & **Dentel, L.** (2018). Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *Royal Society Open Science, 5*(4). DOI: https://doi.org/10.1098/rsos.170354

Selkirk, E. (1984). On the major class features and syllable theory. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure*. MIT Press.

Serrien, D. J. (2008). The neural dynamics of timed motor tasks: Evidence from a synchronization-continuation paradigm. *European Journal of Neuroscience, 27*(6), 1553–1560. DOI: https://doi.org/10.1111/j.1460-9568.2008.06110.x

Šturm, P., & **Volín, J.** (2016). P-centres in natural disyllabic Czech words in a large-scale speech-metronome synchronization experiment. *Journal of Phonetics, 55*, 38–52. DOI: https://doi.org/10.1016/j.wocn.2015.11.003

Su, Y. H., & **Pöppel, E.** (2012). Body movement enhances the extraction of temporal structures in auditory sequences. *Psychological Research, 76*(3), 373–382. DOI: https://doi.org/10.1007/s00426-011-0346-3

Tanaka, H., & **Rathcke, T.** (2016). Then, what is charisma? The role of audio-visual prosody in L1 and L2 political speeches. *Proceedings of Phonetik & Phonologie Im Deutschsprachigen Raum, Munich: LMU* (pp. 294–306).

Thaut, M. H., Rathbun, J. A., & **Miller, R. A.** (1997). Music versus metronome timekeeper in a rhythmic motor task. *International Journal of Arts Medicine, 5*, 4–12. https://psycnet.apa.org/record/1998-00056-001

Tilsen, S., & **Arvaniti, A.** (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America, 134*(1), 628–639. DOI: https://doi.org/10.1121/1.4807565

Truman, G., & **Hammond, G. R.** (1990). Temporal regularity of tapping by the left and right hands in timed and untimed finger tapping. *Journal of Motor Behavior, 22*(4), 521–535. DOI: https://doi.org/10.1080/00222895.1990.10735526

Uldall, E. T. (1971). Isochronous stresses in RP. In L. Hammerich, R. Jacobson & E. Zwirner (Eds.), *Form and substance* (pp. 205–210). Akademisk Forlag.

Valdesolo, P., **Ouyang, J.**, & **DeSteno, D.** (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *Journal of Experimental Social Psychology, 46*(4), 693–695. DOI: https://doi.org/10.1016/j.jesp.2010.03.004

van Santen, J. P. H., & **Shih, C.** (2000). Suprasegmental and segmental timing models in Mandarin Chinese and American English. *The Journal of the Acoustical Society of America, 107*(2), 1012–1026. DOI: https://doi.org/10.1121/1.428281

Villing, R. C., **Repp, B. H.**, **Ward, T. E.**, & **Timoney, J. M.** (2011). Measuring perceptual centers using the phase correction response. *Attention, Perception, and Psychophysics, 73*(5), 1614–1629. DOI: https://doi.org/10.3758/s13414-011-0110-1

Vos, P. G., **van Kruysbergen, N. W.**, & **Mates, J.** (1995). The Perceptual Centre of a Stimulus as the Cue for Synchronization to a Metronome: Evidence from Asynchronies. *The Quarterly Journal of Experimental Psychology Section A, 48*(4), 1024–1040. DOI: https://doi.org/10.1080/14640749508401427

Wagner, P., **Cwiek, A.**, & **Samlowski, B.** (2019). Exploiting the speech-gesture link to capture fine-grained prominence impressions and listening strategies. *Journal of Phonetics, 76*, 100911. DOI: https://doi.org/10.1016/j.wocn.2019.07.001

Wagner, P., **Malisz, Z.**, & **Kopp, S.** (2014). Gesture and speech in interaction: An overview. *Speech Communication, 57*, 209–232. DOI: https://doi.org/10.1016/j.specom.2013.09.008

Wang, D., & **Narayanan, S. S.** (2007). Robust speech rate estimation for spontaneous speech. *IEEE Transactions on Audio, Speech and Language Processing, 15*(8), 2190–2201. DOI: https://doi.org/10.1109/TASL.2007.905178

White, L., & **Mattys, S. L.** (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*(4), 501–522. DOI: https://doi.org/10.1016/j.wocn.2007.02.003

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*(4), 665–679. DOI: https://doi.org/10.1016/j.jml.2011.12.010

White, L., **Payne, E.**, & **Mattys, S. L.** (2009). Rhythmic and prosodic contrast in Venetan and Sicilian Italian. In M. Vigario, S. Frota & M. J. Freitas (Eds.), *Phonetics and Phonology: Interactions and Interrelations* (pp. 137–158). John Benjamins. https://www.researchgate.net/profile/Laurence_White3/publication/256944638_Rhythmic_and_prosodic_contrast_in_Venetan_and_Sicilian_Italian/links/57b19c0208ae15c76cbb163d/Rhythmic-and-prosodic-contrast-in-Venetan-and-Sicilian-Italian.pdf. DOI: https://doi.org/10.1075/cilt.306.07whi

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-319-24277-4_9

Wiget, L., **White, L.**, **Schuppler, B.**, **Grenon, I.**, **Rauch, O.**, & **Mattys, S. L.** (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America, 127*(3), 1559–1569. DOI: https://doi.org/10.1121/1.3293004

Wing, A. M. (2002). Voluntary timing and brain function: An information processing approach. *Brain and Cognition, 48*(1), 7–30. DOI: https://doi.org/10.1006/brcg.2001.1301

Wohlschläger, A., & **Koch, R.** (2000). Synchronization error: An error in time perception. In P. Desain & L. Windsor (Eds.), *Rhythm Perception and Production* (pp. 115–127). Swets & Zeitlinger.

Zatorre, R. J., **Chen, J. L.**, & **Penhune, V. B.** (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*(7), 547–558. DOI: https://doi.org/10.1038/nrn2152

How to cite this article: Rathcke, T., Lin, C.-Y., Falk, S., & Dalla Bella, S. D. 2021 Tapping into linguistic rhythm. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1): 11, pp. 1–32. DOI: https://doi.org/10.5334/labphon.248

Submitted: 22 November 2020 **Accepted:** 12 April 2021 **Published:** 28 May 2021

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS