




It is better when expected: aligning speech and motor rhythms enhances verbal processing

Simone Falk & Simone Dalla Bella


To cite this article: Simone Falk & Simone Dalla Bella (2016): It is better when expected: aligning speech and motor rhythms enhances verbal processing, *Language, Cognition and Neuroscience*, DOI: [10.1080/23273798.2016.1144892](https://doi.org/10.1080/23273798.2016.1144892)



To link to this article: <http://dx.doi.org/10.1080/23273798.2016.1144892>

 [View supplementary material](#) 

 Published online: 04 Apr 2016.

 [Submit your article to this journal](#) 

 Article views: 16

 [View related articles](#) 

 [View Crossmark data](#) 

It is better *when* expected: aligning speech and motor rhythms enhances verbal processing

Simone Falk^{a,b} and Simone Dalla Bella^{c,d,e}

^aInstitute of German Philology, Ludwig-Maximilians-University, München, Germany; ^bLaboratoire Parole et Langage, CNRS, UMR 7309, Aix-Marseille Université, Aix-en-Provence, France; ^cEuroMov, University of Montpellier, Montpellier, France; ^dInstitut Universitaire de France, Paris, France; ^eDepartment of Cognitive Psychology, WSFiZ, Warsaw, Poland

ABSTRACT

Rhythm is a powerful way to shape the processing of complex sounds such as speech or music by generating temporal expectancies in the listener. Here, we investigated if multisensory expectancies generated by aligning speech and motor rhythms may enhance verbal processing. Participants listened to rhythmically regular German sentences and detected word changes occurring on stressed or unstressed syllables. Participants were cued to produce finger taps simultaneously with the auditory speech rhythm. Finger taps were aligned or misaligned with stressed syllables. Detection of word changes was facilitated when manual movements were temporally aligned with the auditory speech rhythm. Moreover, motor alignment enhanced sensitivity to detect changes on stressed syllables compared to a perceptual control condition. Thus, rhythmic speech structure reinforced by concurrent movement in multisensory contexts has beneficial effects on verbal processing. This finding lends support to models of expectancy-driven speech processing.

ARTICLE HISTORY

Received 4 November 2015
Accepted 12 January 2016

KEYWORDS

Speech perception;
multisensory rhythm; verbal
processing; temporal
expectancies; temporal
prediction

When we expect something to happen, we react quicker and more efficiently to the expected event. Expectancies can boost human behaviour. They enable us to anticipate *what* is going to happen and also *when* it will occur. There is growing evidence that the rhythmic properties of music or speech can generate *temporal* expectancies in the listener (Jones, 2009; Kotz & Schwartze, 2010; Zheng & Pierrehumbert, 2010). The recurrence of prominent events such as beats in music or accented syllables in speech leads listeners to develop expectancies about the time of occurrence of upcoming events (Dalla Bella, Białńska, & Sowiński, 2013; Large & Jones, 1999). Typically, information at expected times is better attended to, and processed more efficiently and rapidly than at unexpected times (e.g. Bolger, Coull, & Schön, 2014; Jones, Moynihan, MacKenzie, & Puente, 2002; Quené & Port, 2005). Beneficial effects of temporal expectancies on verbal processing are reported, for example, when monitoring phonetic form or resolving syntactic ambiguities (Cason & Schön, 2012; Roncaglia-Denissen, Schmidt-Kassow, & Kotz, 2013).

Interestingly, the benefits of temporal rhythmic expectancies are not confined to a single modality. For example, reactions to visual targets at expected times

are facilitated by an auditory rhythmic sequence, and vice versa (Bolger, Trost, & Schön, 2013; Su, 2014). Multisensory rhythms such as an auditory and motor rhythm that are temporally aligned can also enhance expectancies. Subtle timing deviations from a regular tone sequence are better detected when participants tap to the sequence before making their judgment (Manning & Schutz, 2013). Moreover, synchronised movement to a musical beat shapes and enhances encoding and later retrieval of musical structures (e.g. Brown & Palmer, 2012; Chemin, Mouraux, & Nozaradan, 2014). These benefits of synchronous movement on auditory processing are possibly underpinned by coupling of neural oscillations originating in distant motor and auditory areas in the brain (Nozaradan, Zerouali, Peretz, & Mouraux, 2015). Thus, benefits of auditory-motor synchronisation in music may emerge because sensory and auditory predictions coincide (Maes, Leman, Palmer, & Wanderley, 2014).

Compared to music, little is known about rhythmic auditory-motor facilitation in verbal processing. However, a close link is observed between the temporal coordination of speech production and upper limb motor functions. Manual pointing gestures or head

movements are temporally well coordinated with the articulation of prominent syllables in speech (e.g. Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004; Rochet-Capellan, Laboissière, Galván, & Schwartz, 2008). Moreover, rhythmic movements like finger tapping co-vary in amplitude with the amplitude of simultaneous articulations (Kelso, Tuller, & Harris, 1983; Parrell, Goldstein, Lee, & Byrd, 2014). Given this close link in the motor system, it is likely that temporally aligned movement also affects speech perception (Gentilucci & Dalla Volta, 2008).

Coordinated movement during speech production is a widespread phenomenon in oral tradition, across groups and cultures. It is common to align rhythmic movements (e.g. hand clapping, finger movements or rope jumping in children's games) with metrical speech rhythms in poems or nursery rhymes (Ong, 2002). In languages such as English or German, metrical speech consists of recurring patterns of strong and weak positions filled with stressed and unstressed syllables, respectively. These features of metrical speech are shown to affect perceptual processes. When listening to metrical speech, expectancies are generated that are typically directed more towards strong positions than weak positions (Cutler, 1976; Pitt & Samuel, 1990). For example, Zheng and Pierrehumbert (2010) showed that participants, listening to English metrical sentences, attended more to variations in vowel duration when these occurred in metrical strong than in weak positions. That metrical speech patterns in languages such as English or German direct attention particularly towards prominent stressed syllables in the speech stream has also been underpinned by recent electroencephalography (EEG) studies (Kotz & Schwartze, 2010; Schmidt-Kassow & Kotz, 2009). In summary, there are indications that expectancies induced by metrical speech are strongest at expected points in time (e.g. stressed syllables), and weaken at less-expected points (e.g. unstressed syllables). Here, we investigate whether verbal processing is enhanced at times when expectancies driven by a verbal and a motor rhythm temporally coincide. Participants are cued to align or misalign a motor rhythm (i.e. finger tapping) to the stressed syllables of a metrical spoken German sentence alternating strong (i.e. stressed syllables) and weak (i.e. unstressed syllables) positions (e.g. **look for rain before the night**, stress in bold). Their task is to detect a word change occurring in either of these positions (see Sturt, Sanford, Stewart, & Dawydiak, 2004; Tillmann & Dowling, 2007). In a perceptual control condition, the same task is performed but without tapping. We expect enhanced performance (i.e. higher word change detection) in the motor task compared to the perceptual control condition, and when rhythmic movement

is temporally aligned with strong ("expected") positions in speech.

Methods

Participants

One hundred twenty-eight German native speakers (32 males, $M = 23.8$ years, $SD = 4.1$; $M = 5.9$ years of musical training, range = 0–21 years), all students from the Ludwig-Maximilians-University in Munich volunteered to participate in the study.

Materials

Twenty-four German speech stimuli with alternating strong and weak syllables were constructed. Each stimulus consisted of two short sentences (eight syllables each), one of which contained the target word (i.e. a verb) that served to test change detection (see Figure 1A). Two versions of each stimulus were created (24×2). In one version, the verb occurred in a metrically strong position, preceded by a bisyllabic subject. In the other version, the verb was in a metrically weak position, preceded by a monosyllabic subject noun and followed by an adverbial (Figure 1A). The stimuli were recorded by a female speaker reading at a regular pace (100 beats/min), cued by a metronome prior to each recording. The recordings were adjusted, when necessary, using PRAAT software (Boersma, 2001), to obtain intervals of 600 ms on average ($SE = 0.84$ ms) between the perceptual centres of metrically strong syllables (Cummins & Port, 1998). Acoustically, target verbs in metrically strong positions were on average 47.5% longer, 1.8 dB louder and had the same pitch height than immediately adjacent weak syllables while target verbs in metrically weak positions were on average 20.1% shorter, 2.6 dB softer and had lower pitch than adjacent strong syllables.

A trial was created in which a stimulus was repeated with a 2 s pause between presentations. In the second presentation ("detection phase"), the verb was replaced by another verb that had the same morpho-syntactic structure and was very similar in meaning (e.g. *jault* "yowls" – *heult* "howls"; Figure 1A; Sturt et al., 2004).

Semantic closeness of the verbs (on a scale from 1 = very distant to 10 = very close in meaning) was confirmed in a pilot experiment ($n = 14$, see Table 1). Both verbs were monosyllabic and comparable in number of phonemes and frequency (Table 1). Filler trials were created to ensure that participants did not pay selectively attention to verbs. In 24 fillers, a noun was changed, and in 12 fillers, there was no change.

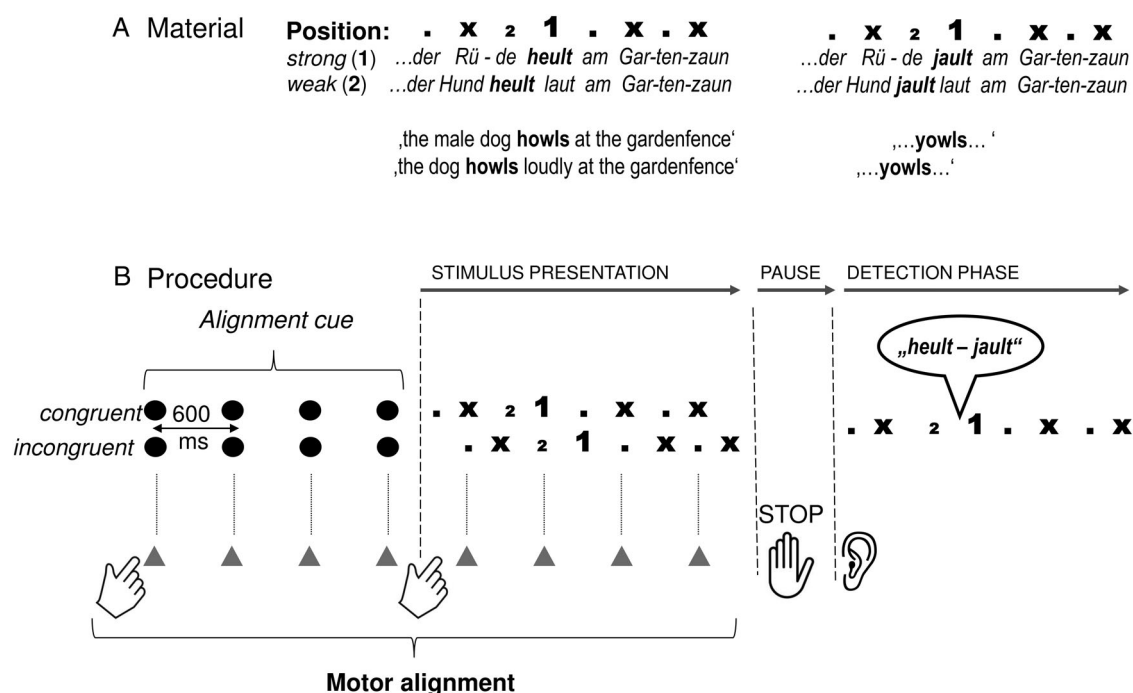


Figure 1. (A) Material and (B) procedure used in the Experiment. The metrical speech pattern is displayed with strong positions marked by “x”, and weak positions by “.”. The verb change occurred in a strong (1) or weak (2) position. In panel (B) the alignment cue (big black dots) and motor alignment (i.e. finger taps as grey triangles) with the positions in the speech stimuli are displayed.

Procedure

In the multisensory task (“tapping”), participants ($n = 64$) performed a rhythmic finger tapping task while listening to the verbal stimuli with the goal of detecting a change when the stimulus was repeated. The alignment of the motor rhythm with the verbal rhythm was controlled by using a synchronisation-continuation task (Wing, 2002, see Figure 1B). Prior to each stimulus, participants were cued by 12 isochronous metronome tones (“alignment cue”; duration of tone = 30 ms, Inter-Onset-Interval = 600 ms) which stopped when the speech started. Tones were either congruently or incongruently timed with the following speech stimulus. When the alignment cue was *congruently* timed, the first metrically strong syllable started 600 ms after the last tone (i.e. it occurred at the expected time based on the metronome). When it was *incongruently* timed, the speech stimulus was delayed by 300 ms, as compared to the congruent cue (see Figure 1B). Participants tapped in synchrony with the metronome using the index finger of their dominant

hand on the left panel of a Roland SPD-6 MIDI percussion pad. When the cue stopped and the speech started, they were instructed to keep tapping at the rate previously indicated by the cue. Their taps continuing at the pace of the congruently timed metronome coincided with metrically strong positions, while they fell between strong positions when the metronome was incongruently timed (Figure 1B). During the pause, participants stopped tapping and prepared for detecting a verbal change. In case they perceived a change, they tapped as fast as possible on the right panel of the percussion pad, thereby stopping the stimulus. Participants then recalled the original and the changed word. Verbal answers were recorded via microphone and written down by the Experimenter. To ensure that the whole sentence was processed, participants summarised its content every three trials on average. A block of 30 trials (12 stimuli, 12 fillers, and 6 no-change stimuli) per alignment was preceded by three practice trials. Stimuli were organised in eight randomisation lists and were presented equally often under both alignment conditions in counterbalanced order across participants.

A perceptual task was run. Another group of participants ($n = 32$) were presented to the same stimuli as described above. They were asked to perform the same task as above, but without finger tapping (“cueing”). Finally, a perceptual baseline (“no cues”) for word change detection with the same verbal material was

Table 1. Mean phoneme number, frequency and semantic distance of the verb pairs.

Verb pair	Original verb	Changed verb
Number of phonemes	4.1 (0.58)	4.3 (0.38)
Frequency (Dlexdb)	11 per million (28)	16 per million (49)
Semantic distance	7.62 (1.25) “semantically close meaning”	

Standard deviations are displayed in parentheses.

obtained in another group of participants ($n = 32$) who performed the task in the absence of alignment cues.

In the aforementioned tasks and in the baseline, the positions in which the verbal change occurred (i.e. metrically strong – weak) were varied between-subjects to avoid carry-over effects from previous stimulus presentations. Half of the participants were randomly assigned to detect changes in strong positions, the other half in weak positions. The Experiment lasted approximately 45 min. It was run on an IBM-compatible computer using MaxMSP 5.1.9. Auditory stimuli were presented via Beyer-Dynamic DT-770 Pro 250 headphones.

Results

Participants' verbal responses were analysed by calculating sensitivity (d') and response bias (C , Macmillan & Creelman, 2005). A Hit occurred when both the changed and the original verb were provided in their semantically and phonetically accurate form. A False alarm occurred when a change was reported in a no-change trial. Detection time (DT), for Hits only, was calculated from the perceptual centre of the verb to the motor response.¹ Data were log-transformed as a measure to reduce skewness of DT data. Statistic analyses were run with IBM SPSS 22.0 Software.

In the tapping task, performance was analysed for all trials. When cued by the metronome, participants tapped at the intended tempo (inter-tap-interval, ITI, $M = 600.1$ ms, $SE = .73$ ms) with a normal variability (coefficient of variation of the inter-tap-interval, CV ITI = .05, $SE = .002$). Their taps anticipated the metronome tones, as indicated by a negative mean asynchrony between the taps and tone onsets (-46.94 ms, $SE = 2.77$ ms), a common finding in tapping studies (e.g. Repp, 2005). The participants kept tapping at the tempo indicated by the metronome while they were presented with the speech stimulus, in both the congruent (mean ITI = 597.4 ms, $SE = 1.4$ ms) and the incongruent conditions (mean ITI = 5964.0 ms, $SE = 2.9$ ms). They displayed comparable variability with both alignments (CV ITI = .06 and .07, respectively). In summary, the participants performed the tapping task as instructed.

We examined whether expectancies derived from motor alignment affected change detection in metrically strong and weak positions compared to perceptual cueing. Results for verbal responses are displayed in Figure 2 as well as the perceptual baseline performance which served as a reference. The mean baseline performance was obtained by averaging the results across strong and weak positions (positions did not significantly differ, $p > .40$). The assumption of normality (tested with Shapiro–Wilk tests) was met for all but one condition

(C in the incongruent motor condition) due to an extreme observation. Therefore, the ANOVAs on C reported below were run with and without this outlier. Discarding the outlier did not affect the results.

Results for verbal responses were entered in three separate $2 \times 2 \times 2$ mixed-design analyses of variance (ANOVAs), taking sensitivity (d'), response bias (C), and DT as dependent variables, with Subject as the random variable. Alignment (congruent vs. incongruent) was the within-subject factor, Position (strong vs. weak) and Task (tapping vs. cueing) the between-subjects factors. As group sizes differed for tapping and cueing,² homogeneity of variance was confirmed by Box's M test and Levene's test. Results showed a main effect of Task for DT ($F(1, 86) = 5.46$, $MSE = .03$, $p = .022$). No main effect of alignment was found for either dependent variable ($ps > .48$), but a significant Alignment \times Task interaction was present with sensitivity and DT (d' : $F(1, 92) = 4.12$, $MSE = .47$, $p = .045$; DT: $F(1, 86) = 5.73$, $MSE = .01$, $p = .019$). The interactions were decomposed by computing simple effects of Alignment while keeping Task constant. In the tapping task, Congruent alignment of a motor rhythm with the verbal rhythm resulted in higher detection rates and faster detection than when they were incongruently aligned (d' : $F(1, 94) = 4.82$, $MSE = .46$, $p = .031$; DT: $F(1, 88) = 6.62$, $MSE = .01$, $p = .012$). In contrast, no differences between the presentation of congruent and incongruent alignment cues were found in the perceptual cueing task ($ps > .29$). Response bias differed for the two alignments in both tasks (C : $F(1, 92) = 6.49$, $MSE = .11$, $p = .013$), showing a lower criterion in the congruent alignment than in the incongruent alignment. The three-way Alignment \times Task \times Position interaction did not reach significance with either of the dependent variables ($p > .47$). Finally, changes in metrically strong positions were detected overall more accurately (d' : $F(1, 92) = 12.66$, $MSE = .95$, $p = .001$), more reliably (C : $F(1, 92) = 14.52$, $MSE = .23$, $p < .001$) and faster (DT: $F(1, 86) = 35.53$, $MSE = .03$, $p < .001$) than in metrically weak positions in both tapping and cueing tasks.

To determine if motor alignment enhances detection performance, we compared the performance for each alignment condition in each position to the perceptual baseline. This was done by using Bonferroni-corrected two-tailed t -tests (adjusted $p = .00625$).³ Homogeneity of variance was tested by Levene's tests, and degrees of freedom were adjusted whenever homoscedasticity was not met. Congruent motor alignment resulted in significantly higher sensitivity (d' : $t(62) = 2.99$, $SEM = .21$, $p = .004$), smaller response bias (C : $t(62) = 4.11$, $SEM = .08$, $p < .001$) and faster DT ($t(62) = 6.69$, $SEM = .03$, $p < .001$) than in the baseline when changes were in

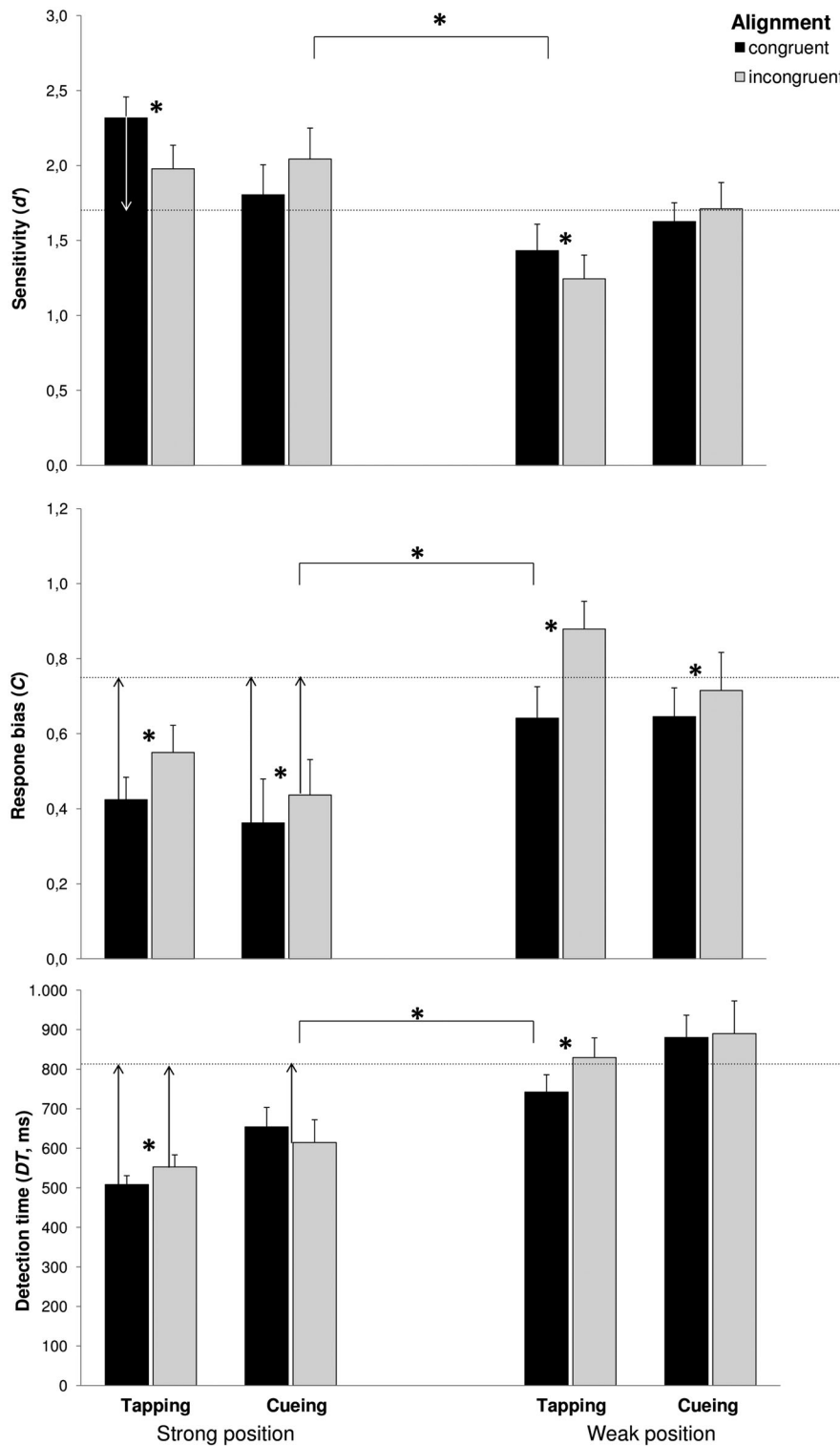


Figure 2. Sensitivity (upper panel), response bias (middle panel) and raw DT data (lower panel) in the change detection task with and without motor alignment when the target verb was in metrically strong or weak position. Baseline performance (i.e. without any alignment cues) is indicated by the dotted line. Arrows mark performances significantly different from baseline. Stars indicate significant effects. Error bars represent SE of the mean.

strong positions. Moreover, unexpectedly, incongruent motor alignment yielded faster DT than in the baseline ($t(62) = 5.20$, $SEM = .03$, $p < .001$). No differences were found for detection performance in weak positions. The

perceptual cueing data (averaged for alignment for d' and DT) were similarly compared to the baseline showing faster DT ($t(46) = 2.88$, $SEM = .04$, $p = .006$) and smaller response bias (congruent: $t(46) = 3.52$,

$SEM = .11, p = .001$; incongruent: $t(46) = 3.17, SEM = .10, p = .003$), but only in response to changes in strong positions. Finally, to confirm the unique benefit of motor alignment, we performed a comparison between congruent motor alignment and the pooled perceptual data (i.e. cueing + baseline) separately for each position. This comparison showed that in strong positions, congruent motor alignment enhanced both sensitivity ($t(62) = 2.52, SEM = .21, p = .014$, one-tailed test) and speeded up DT ($t(55.5) = -4.51, SEM = .03, p < .001$, one-tailed test) compared to perception. No differences were found for weak positions. Moreover, no difference in response bias was found between tapping and the perceptual tasks.

Discussion

Multisensory temporal alignment reinforces rhythmic expectancies that improve verbal processing. When finger tapping was congruently aligned with strong metrical positions, verb changes were easier and faster to detect than when movement was incongruently aligned. Alignment effects were found exclusively with concurrent motor performance for all variables, except for response bias. Congruent motor alignment particularly enhanced sensitivity in detecting changes that occurred in strong metrical positions, as compared to perceptual tasks (with and without cueing). This was not the case, though, for weak metrical positions. Although perceptual cueing was also superior to a perceptual baseline without cueing in lowering response bias and speeding up DT in strong positions, sensitivity remained unchanged, in contrast to the results in the multisensory task.

These results show for the first time that temporal alignment of an auditory verbal and a concurrent motor rhythm is beneficial for verbal processing. This idea is compatible with models of the linking of perception and action in time through shared predictions (Maes et al., 2014; Nozaradan et al., 2015). Accordingly, synchronised perceptual and motor cues in speech may tap the same expectancy-driven mechanism (Chemin et al., 2014). The aligned motor rhythm likely enhances the strength of rhythmical expectancies and thereby increases attending to expected times in the speech signal (Jones, 2009; Large & Jones, 1999; Schröger, Kotz, & SanMiguel, 2015). A similar mechanism has been advocated to account for benefits in speech processing (e.g. quicker phoneme detection, improved computation of semantic and syntactic information) due to a rhythmic cue that precedes metrical speech (Cason & Schön, 2012; Kotz & Gunter, 2015; Quené & Port, 2005). The present findings add to our understanding on how

temporal expectancies in speech emerge in the listener. In the presence of an aligned motor rhythm, the expectancies were clearly driven by the speech rhythm. Indeed, detection of a word change was enhanced (with higher d' and lower DT) when the movement coincided with prominent syllables in strong metrical positions compared to incongruent tapping.

These findings highlight the crucial role of prominent syllables in building temporal expectancies. Prominent syllables in speech carry heightened acoustic information (i.e. they are more high-pitched, louder and longer; Beckman, 1986), and they are generated through increased oromotor kinematics, such as larger and longer jaw–lip displacement and higher peak velocity than less prominent syllables (Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985; Vatikiotis-Bateson & Kelso, 1993). Interestingly, the acoustic signature of prominent syllables is not only supported by enhanced oromotor kinematics but is also visible in the kinematics of concurrently produced finger taps (Parrell et al., 2014; Smith, McFarland, & Weber, 1986). This transfer between articulatory and manual motor domains may also be relevant for the perceptual results in our study. In perception, the heightened motor features of prominent syllables may be simulated through forward predictions about the interlocutor's productions (Pickering & Garrod, 2013). Hence, when aligning manual movements with prominent syllables, multisensory integration may be more successful than with non-prominent syllables because of underlying correspondences in motor dynamics.

The differences between multisensory and perceptual tasks were found only for prominent syllables. However, the results differed for sensitivity, response bias and DT. It is likely that these variables represent different underlying processes. Sensitivity, appears to be most affected by the amount of attentional resources allocated to the speech signal. It was higher in strong compared to weak positions and it was uniquely altered by motor alignment, as the perceptual tasks with and without cueing did not differ. The absence of an alignment effect in perceptual cueing also suggests that, independently of the timing of the cue, participants were entrained by the speech rhythm over the course of stimulus presentation. It may even be a possibility that longer delays between the cue and the speech stimulus (as it was the case for the incongruent condition) rather help to adapt to the speech rhythm, as suggested by the slight, but non-significant, tendency for more accurate change detection in the incongruent condition during cueing (see Figure 2). These suggestions deserve further investigation.

Response bias was especially influenced by the timing of the alignment cue. Incongruent timing of the cue

made participants adopt a more conservative criterion than congruent timing, in the perceptual and in the multisensory tasks. Response bias was lower in the cueing and tapping tasks than in the perceptual baseline (i.e. in strong positions). With respect to metrical positions, participants clearly adopted a more conservative criterion for weak positions than for strong positions. Note that response bias, as indicated by C , can be considered as being independent of d' (Stanislaw & Todorov, 1999). It is worth noting that recent neuroimaging and evidence from studies using transcranial magnetic stimulation indicates that the magnitude of response bias correlates with activations in primary motor (e.g. hand motor cortex) and sensorimotor brain regions during syllable discrimination tasks (Smalle, Rogers, & Möttönen, 2015; Venezia, Saberi, Chubb, & Hickok, 2012). The present results suggest that decision biases in a word change detection task may also be modulated by temporal aspects of rhythmic cueing. This interpretation is compatible with current models of perceptual and sensorimotor timing which hypothesise that rhythmic cueing engages the motor network through a cerebello-thalamo-cortical circuitry (Dalla Bella, Benoit, Farrugia, Schwartz, & Kotz, 2015; Kotz & Gunter, 2015; Kotz & Schwartz, 2010). In summary, our findings point towards effects of rhythmic expectancies on decision-making, a result which needs further investigation.

Finally, DT was faster in tapping and cueing than in the perceptual baseline. However, it is with tapping that participants showed the fastest DTs. Interestingly, incongruent tapping yielded also faster DTs compared to baseline while we expected slower DTs. Thus, it is possible that speeded response times may be due to a more general effect of predictive rhythms in enhancing the temporal preparation prior to action (e.g. de la Rosa, Sanabria, Capizzi, & Correa, 2012; Sanabria, Capizzi, & Correa, 2011). Even in the case of incongruent tapping, which occurred systematically off-beat, and not randomly in between accented syllables, the rhythmic speech structure may have been partly enhanced, as compared to the baseline perceptual task. An alternative view would be that motor processes are generally facilitated through rhythmic tapping or cueing. However, in this case, we would have expected to find a general facilitation of DTs across positions and not a selective enhancement in strong positions which appear to be the anchors for rhythmic expectancies.

How do the results relate to previous findings on expectancy-driven speech processing in the auditory modality? Previous research has consistently shown that listening to alternating metrical strong-weak patterns in sentences, such as used in the present study,

creates metrical expectancies in the listener about the upcoming structure (e.g. Bohn, Knaus, Wiese, & Domahs, 2013; Magne et al., 2007; Schmidt-Kassow & Kotz, 2009). This result has also been confirmed with less common metrical patterns (e.g. strong-weak-weak-weak, Kotz & Schmidt-Kassow, 2015) and for temporally precise and less precise recurrences of these patterns. In speech processing, metrical expectancies can affect the early interpretation of prosodic features (Brown, Salverda, Dilley, & Tanenhaus, 2015) and, subsequently, word segmentation strategies (Dilley & McAuley, 2008). EEG and neuroimaging studies also showed that metrical speech compared to more irregular speech lowers processing demands when participants listen to complex linguistic structures such as sentences containing syntactic ambiguities or semantic violations (Roncaglia-Denissen et al., 2013; Rothermich, Schmidt-Kassow, & Kotz, 2012). In line with results from Rothermich et al. (2012) and Rothermich and Kotz (2013), our results (i.e. improved detection of subtle word changes) suggest that the beneficial effects of expectancies in metrical speech may extend to semantic processes. However, our task also involved working memory. There is an ongoing debate about whether rhythmic contexts such as metrical speech or music serve as contextual cues that foster multiple encoding and enhance memory processes (Baddeley, 2004; Purnell-Webb & Speelman, 2008; Tillmann & Dowling, 2007; Tulving, 1972). Although our findings shed light on potential benefits for memory through multisensory encoding involving rhythmic movement, further research is needed in order to clarify which memory processes are engaged in this task and under which conditions.

Finally, the findings support the idea that multisensory information sharing similar temporal dynamics fosters binding and cross-modal integration, a process possibly underpinned by enhanced coherence of neural activity in distinct brain areas (e.g. Damasio, 1989; Nozaradan, Peretz, & Mouraux, 2012; Sapkota, Pardhan, & van der Linde, 2013). It may seem paradoxical that participants' performance increased while performing a dual task (i.e. a verbal task and a motor task) as compared to a single-task baseline. Thus, aligning auditory and motor rhythms created optimal conditions for efficient integration of verbal and motor information. The observed effects may relate to other cases of temporally aligned multisensory (e.g. audiovisual or somatosensory) integration observed during speech perception (e.g. Gick & Derrick, 2009; Ito, Gracco, & Ostry, 2014). For example, speech processing is enhanced when participants visually see simple linear hand movements (i.e. beat-gestures, McNeill, 1992) that are simultaneously presented with spoken words (Biau & Soto-Faraco, 2013; Wang & Chu,

2013). Such multimodal effects may be driven by an enhancement of timing cues via consistent auditory speech and visual movement information (Munhall et al., 2004), consistent with theories of action-based effects on sound perception (Maes et al., 2014).

The present study is limited to the effects of multisensory rhythms on metrical speech perception. However, we do not exclude a priori that similar auditory-motor processes may play a role during conversational speech, when interlocutors move their heads, fingers or feet along with expected prominent syllables while listening to their conversational partner. Another issue is the language-specificity of the present results in light of the fact that German as well as English are stress-based language in which rhythmic expectancies play an important role in speech perception. Although a few studies on non-stress languages such as French (e.g. Cason & Schön, 2012; Magne et al., 2007) point in a similar direction, more evidence is needed to extend our conclusions on rhythmic expectancies and multisensory binding to such languages. Finally, we see a link between the present results and efforts in speech therapy to use auditory-motor mappings to stimulate language functions. For instance, tapping together with speech productions is used in variants of melodic intonation therapy (Albert, Sparks, & Helm, 1973) to treat non-fluent aphasia resulting from a brain insult (e.g. Stahl, Kotz, Henseler, Turner, & Geyer, 2011), and to aid autistic children recover elementary verbal production skills (Wan et al., 2011). In light of our results, it may be interesting to consider training of verbal perceptual skills by using multisensory rhythmic stimulation.

In summary, moving along to a speech rhythm is particularly efficient in reinforcing rhythmic expectancies in speech perception. Our findings underscore the close link between rhythm in the verbal and motor domains (e.g. Cummins, 2009; Schmidt-Kassow et al., 2014). They point to verbal prominences as predictive anchors allowing for multisensory coupling and possibly, joint temporal expectancies in both the listener and the speaker (e.g. Lidji, Palmer, Peretz, & Morningstar, 2011; Shockley, Santana, & Fowler, 2003).

Notes

1. A MIDI delay of 81 ms was subtracted from all motor data. One per cent of the DT data were discarded due to particularly slow responses (>3 s), 9% were not available because of failures in recording participants' motor response (five participants were therefore excluded from analysis).
2. In order to assure that the unequal sample sizes did not impact on differences observed between groups, we ran the same ANOVAs with equal sample sizes by randomly

choosing a sample of 32 participants in the tapping condition (half of the participants received stimuli in weak, half in strong positions; half of the participants began with congruent alignment, half with incongruent alignment). The ANOVAs yielded the same effects and interactions as with the larger sample except for the main effect of Task in DT. Main effects of Position were found for d' , C and DT (d' : $F(1, 60) = 8.33, p = .005$; C : $F(1, 60) = 5.80, p = .019$; DT: $F(1, 60) = 29.63, p < .001$). A main effect of Alignment was found for C ($F(1, 60) = 5.10, p = .028$) and Task \times Alignment interactions for d' and DT (d' : $F(1, 60) = 7.17, p = .01$, DT: $F(1, 60) = 6.52, p = .013$).

3. As the baseline had no alignment condition (no cue was present), t -tests were the most appropriate means of statistical comparison.

Acknowledgements

We thank the IPS (LMU Munich), Franziska Flieger, Franziska Holzheimer, Elena Maslow, Judith Oebels and Alina Schmajew for help with stimulus preparation and testing.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007–2013) under REA grant agreement n° 327586 to S.F., the LMU Munich's Institutional Strategy LMUexcellent within the framework of the German Excellence Initiative to S.F., the Institut Universitaire de France (IUF) to S.D.B. Both authors were financially supported by a bilateral travel grant of the DAAD and Polish Ministry of Science and Higher Education.

References

- Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology*, 29, 130–131. doi:10.1001/archneur.1973.00490260074018
- Baddeley, A. D. (2004). *Your memory: A user's guide*. London: Carlton.
- Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.
- Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, 124, 143–152. doi:10.1016/j.bandl.2012.10.008
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341–345.
- Bohn, K., Knaus, J., Wiese, R., & Domahs, U. (2013). The influence of rhythmic (ir)regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia*, 51(4), 760–771. doi:10.1016/j.neuropsychologia.2013.01.006
- Bolger, D., Coull, J. T., & Schön, D. (2014). Metrical rhythm implicitly orients attention in time as indexed by improved target detection and left inferior parietal activation. *Journal*

- of *Cognitive Neuroscience*, 26, 593–605. doi:10.1162/jocn_a_00511
- Bolger, D., Trost, W., & Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychologica*, 142, 238–244. doi:10.1016/j.actpsy.2012.11.012
- Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance*, 41(2), 306–323. doi:10.1037/a0038689
- Brown, R. M., & Palmer, C. (2012). Auditory-motor learning influences auditory memory for music. *Memory and Cognition*, 40, 567–578. doi:10.3758/s13421-011-0177-x
- Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, 50, 2652–2658. doi:10.1016/j.neuropsychologia.2012.07.018
- Chemin, B., Mouraux, A., & Nozaradan, S. (2014). Body movement selectively shapes the neural representation of musical rhythms. *Psychological Science*, 25, 2147–2159. doi:10.1177/0956797614551161
- Cummins, F. (2009). Rhythm as an affordance for the entrainment of movement. *Phonetica*, 66, 15–28. doi:10.1159/000208928
- Cummins, F., & Port, R. F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26, 145–171. doi:10.1006/jpho.1998.0070
- Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics*, 20, 55–60. doi:10.3758/BF03198706
- Dalla Bella, S., Benoit, C. E., Farrugia, N., Schwartz, M., & Kotz, S. A. (2015). Effects of musically cued gait training in Parkinson's disease: Beyond a motor benefit. *Annals of the New York Academy of Sciences*, 1337, 77–85. doi:10.1111/nyas.12651
- 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, e71945. doi:10.1371/journal.pone.0071945
- Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25–62. doi:10.1016/0010-0277(89)90005-X
- Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59, 294–311. doi:10.1016/j.jml.2008.06.006
- Gentilucci, M., & Dalla Volta, R. (2008). Spoken language and arm gestures are controlled by the same motor control system. *Quarterly Journal of Experimental Psychology*, 61, 944–957. doi:10.1080/17470210701625683
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, 462, 502–504. doi:10.1038/nature08572
- Ito, T., Gracco, V. L., & Ostry, D. J. (2014). Temporal factors affecting somatosensory-auditory interactions in speech processing. *Frontiers in Psychology*, 5, 1198. doi:10.1093/cercor/bhl174
- Jones, M. (2009). Musical time. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The oxford handbook of music psychology* (pp. 81–92). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199298457.013.0008
- Jones, M., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, 13(4), 313–319. doi:10.1111/1467-9280.00458
- Kelso, J. A. S., Tuller, B., & Harris, K. S. (1983). A “Dynamic Pattern” perspective on the control and coordination of movement. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 137–173). New York, NY: Springer.
- Kelso, J. A., Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America*, 77, 266–280. doi:10.1121/1.392268
- Kotz, S. A., & Gunter, T. C. (2015). Can rhythmic auditory cuing remediate language-related deficits in Parkinson's disease? *Annals of the New York Academy of Science*, 1337, 62–68. doi:10.1111/nyas.12657
- Kotz, S. A., & Schmidt-Kassow, M. (2015). Basal ganglia contribution to rule expectancy and temporal predictability in speech. *Cortex*, 68, 48–60. doi:10.1016/j.cortex.2015.02.021
- Kotz, S. A., & Schwartz, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences*, 14, 392–399. doi:10.1016/j.tics.2010.06.005
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106, 119–159. doi:10.1037/0033-295X.106.1.119
- Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin & Review*, 18, 1035–1041. doi:10.3758/s13423-011-0163-0
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory. A user's guide*. Mahwah, NJ: Laurence Erlbaum.
- Maes, P.-J., Leman, M., Palmer, C., & Wanderley, M. M. (2014). Action-based effects on music perception. *Frontiers in Psychology*, 4, 1008. doi:10.3389/fpsyg.2013.01008
- Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martin, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cerebral Cortex*, 17(11), 2659–2668. doi:10.1093/cercor/bhl174
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- Manning, F. C., & Schutz, M. (2013). “Moving to the beat” improves timing perception. *Psychonomic Bulletin & Review*, 20, 1133–1139. doi:10.3758/s13423-013-0439-7
- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, 15, 133–137. doi:10.1111/j.0963-7214.2004.01502010.x
- Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Steady-state evoked potentials as an index of multisensory temporal binding. *NeuroImage*, 60, 21–28. doi:10.1016/j.neuroimage.2011.11.065
- Nozaradan, S., Zerouali, Y., Peretz, I., & Mouraux, A. (2015). Capturing with EEG the neural entrainment and coupling underlying sensorimotor synchronization to the beat. *Cerebral Cortex*, 25, 736–747. doi:10.1093/cercor/bht261
- Ong, W. J. (2002). *Orality and literacy: The technologizing of the word*. New York, NY: Routledge.
- Parrell, B., Goldstein, L., Lee, S., & Byrd, D. (2014). Spatiotemporal coupling between speech and manual motor actions. *Journal of Phonetics*, 42, 1–11. doi:10.1016/j.wocn.2013.11.002

- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral Brain Sciences*, 36(4), 329–347. doi:10.1017/S0140525X12001495
- Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance*, 16(3), 564–573. doi:10.1037/0096-1523.16.3.564
- Purnell-Webb, P., & Speelman, C. P. (2008). Effects of music on memory for text. *Perceptual Motor Skills*, 106(3), 927–957. doi:10.2466/pms.106.3.927-957
- Quené, H., & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, 62, 1–13. doi:10.1159/000087222
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12, 969–992. doi:10.3758/BF03206433
- Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J. L. (2008). The speech focus position effect on jaw-finger coordination in a pointing task. *Journal of Speech, Language and Hearing Research*, 51, 1507–1521. doi:10.1044/1092-4388(2008)07-0173
- Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS ONE*, 8, e56000. doi:10.1371/journal.pone.0056000
- de la Rosa, M. D., Sanabria, D., Capizzi, M., & Correa, A. (2012). Temporal preparation driven by rhythms is resistant to working memory interference. *Frontiers in Psychology*, 2, 308. doi:10.3389/fpsyg.2012.00308
- Rothermich, K., & Kotz, S. A. (2013). Predictions in speech comprehension: fMRI evidence on the meter–semantic interface. *NeuroImage*, 70, 89–100. doi:10.1016/j.neuroimage.2012.12.013
- Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, 50, 232–244. doi:10.1016/j.neuropsychologia.2011.10.025
- Sanabria, D., Capizzi, M., & Correa, A. (2011). Rhythms that speed you up. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 236–244. doi:10.1037/a0019956
- Sapkota, R. P., Pardhan, S., & van der Linde, I. (2013). Manual tapping enhances visual short-term memory performance where visual and motor coordinates correspond. *British Journal of Psychology*, 104, 249–264. doi:10.1111/j.2044-8295.2012.02115.x
- Schmidt-Kassow, M., & Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience*, 21(9), 1693–1708. doi:10.1162/jocn.2008.21153
- Schmidt-Kassow, M., Zink, N., Mock, J., Thiel, C., Vogt, L., Abel, C., & Kaiser, J. (2014). Treadmill walking during vocabulary encoding improves verbal long-term memory. *Behavioral and Brain Functions*, 10, 24. doi:10.1186/1744-9081-10-24
- Schröger, E., Kotz, S. A., & SanMiguel, I. (2015). Bridging prediction and attention in current research on perception and action. *Brain Research*, 1626, 1–13. doi:10.1016/j.brainres.2015.08.037
- Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 326–332. doi:10.1037/0096-1523.29.2.326
- Smalle, E. H., Rogers, J., & Möttönen, R. (2015). Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation. *Cerebral Cortex*, 25(10), 3690–3698. doi:10.1093/cercor/bhu218
- Smith, A., McFarland, D., & Weber, C. (1986). Interactions between speech and finger movements: An exploration on the dynamic pattern perspective. *Journal of Speech, Language, and Hearing Research*, 29, 471–480. doi:10.1044/jshr.2904.471
- Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: Why singing may not hold the key to recovery from aphasia. *Brain*, 134(10), 3083–3093. doi:10.1093/brain/awr240
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavioral Research Methods*, 31(1), 137–149. doi:10.3758/BF03207704
- Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, 11, 882–888. doi:10.3758/BF03196716
- Su, Y.-H. (2014). Visual enhancement of auditory beat perception across auditory interference levels. *Brain and Cognition*, 90, 19–31. doi:10.1016/j.bandc.2014.05.003
- Tillmann, B., & Dowling, W. J. (2007). Memory decreases for prose, but not for poetry. *Memory & Cognition*, 35, 628–639. doi:10.3758/BF03193301
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization and memory* (pp. 381–403). New York, NY: Academic Press.
- Vatikiotis-Bateson, E., & Kelso, J. A. S. (1993). Rhythm type and articulatory dynamics in English, French, and Japanese. *Journal of Phonetics*, 2, 231–265.
- Venezia, J. H., Saberi, K., Chubb, C., & Hickok, G. (2012). Response bias modulates the speech motor system during syllable discrimination. *Frontiers in Psychology*, 28(3), 157. doi:10.3389/fpsyg.2012.00157
- Wan, C. Y., Bazen, L., Baars, R., Libenson, A., Zipse, L., Zuk, J., ... , & Schlaug, G. (2011). Auditory-motor mapping training as an intervention to facilitate speech output in non-verbal children with autism: A proof of concept study. *PLoS ONE*, 6(9), e25505. doi:10.1371/journal.pone.0025505
- Wang, L., & Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: An ERP study. *Neuropsychologia*, 51, 2847–2855. doi:10.1016/j.neuropsychologia.2013.09.027
- Wing, A. M. (2002). Voluntary timing and brain function: An information processing approach. *Brain and Cognition*, 48, 7–30. doi:10.1006/brcg.2001.1301
- Zheng, X., & Pierrehumbert, J. B. (2010). The effects of prosodic prominence and serial position on duration perception. *Journal of the Acoustical Society of America*, 128(2), 851–859. doi:10.1121/1.3455796