

The effects of dual-task interference in predicting turn-ends in speech and music

Nina K. Fisher^{1,2}, Lauren V. Hadley³, Ruth E. Corps^{1,4} & Martin J. Pickering¹

¹Department of Psychology, University of Edinburgh

²School of Applied Sciences, Edinburgh Napier University

³Hearing Sciences - Scottish Section, University of Nottingham

⁴Psychology of Language Department, Max Planck Institute for Psycholinguistics

Corresponding author:

Nina K. Fisher

e-mail: n.fisher@napier.ac.uk

School of Applied Sciences, Edinburgh Napier University, UK

Funding Sources

This research was supported by Leverhulme Trust Research Grant RPG-2017-239 (awarded to MJP and LVH), the Medical Research Council [grant number MR/S003576/1]; and the Chief Scientist Office of the Scottish Government.

The effects of dual-task interference in predicting turn-ends in speech and music

Abstract

Determining when a partner's spoken or musical turn will end requires well-honed predictive abilities. Evidence suggests that our motor systems are activated during perception of both speech and music, and it has been argued that motor simulation is used to predict turn-ends across domains. Here we used a dual-task interference paradigm to investigate whether motor simulation of our partner's action underlies our ability to make accurate turn-end predictions in speech and in music. Furthermore, we explored how specific this simulation is to the action being predicted. We conducted two experiments, one investigating speech turn-ends, and one investigating music turn-ends. In each, 34 proficient pianists predicted turn-endings while (1) passively listening, (2) producing an effector-specific motor activity (mouth/hand movement), or (3) producing a task- and effector-specific motor activity (mouthing words/fingering a piano melody). In the speech experiment, any movement during speech perception disrupted predictions of spoken turn-ends, whether the movement was task-specific or not. In the music experiment, only task-specific movement (i.e., fingering a piano melody) disrupted predictions of musical turn-ends. These findings support the use of motor simulation to make turn-end predictions in both speech and music but suggest that the specificity of this simulation may differ between domains.

Keywords: prediction, language, music, simulation, motor interference

1. Introduction

Both speakers and expert musicians are extremely adept at coordinating in time. For example, people are able to synchronize within about 40ms of each other both when speaking (Cummins, 2003) and when playing music (Keller & Appel, 2010; Rasch, 1979; Shaffer, 1984), and are able to take consecutive turns with very brief intervals in speech (Stivers et al., 2009) and music (Hadley et al., 2015). In the current study, we explore a potential mechanism for enabling such accurate turn-taking, namely the activation of the motor system, which might allow for a smooth turn change between agents when speaking or playing an instrument. In this section we first argue that people predict turn-ends in both speech and in music. We then discuss evidence for motor system involvement in this turn-end prediction. Finally, we consider the theory of prediction-by-simulation, which proposes that people motorically simulate the actions of the person that they are speaking or playing with in order to predict the actions that they are likely to produce next. This theory is discussed in order to motivate our study of different forms of motor interference in speech and music prediction.

During a conversation, modal intervals between speech turns are about 200ms (Stivers et al., 2009). However, according to Indefrey and Levelt (2004), it takes around 600ms to prepare to utter a single word. Together, these observations suggest that during conversation, listeners predict the end of their partner's speaking turn well in advance of their actual turn-end (Levinson, 2016), so they can prepare their response and produce it at the appropriate moment. In support of this proposal, De Ruiter et al. (2006) asked participants to indicate when they thought a speech turn would end. They found that participants were accurate to around the length of a syllable, and thus provided strong evidence that the ends of turns could be accurately predicted.

Turn-taking also occurs in music and may similarly rely on predictive mechanisms (Pecenka & Keller, 2011). According to Phillips-Silver and Keller (2012), musical turn-taking is “the ordering of communicative signals produced by separate individuals in such a way that there is little temporal overlap” (p. 2), and listeners are able to predict the end of some styles of musical turns with remarkable accuracy. Using the same paradigm as De Ruiter et al., (2006), Hadley et al., (2018) asked 45 listeners with a broad range of musical expertise (from none to over ten years of musical training) to press a button to indicate when they thought a musical improvisation turn was about to end. Similar to De Ruiter et al. (2006), they found that listeners predicted the musicians’ turn ends to the accuracy of approximately one beat. Given that turn-end predictions are made during both speech and music, we now consider how such accurate predictions are made.

1.1. Engagement of the Motor System in Prediction

Several researchers suggest that our motor systems are engaged both to understand and to predict the experiences of others. For example, the embodied simulation perspective suggests that the observer maps the actions of the other person onto their own motor system, this use of one’s own motor system contributes not only to physical action understanding, but also applies to understanding in social cognition (Gallese, 2005; 2007).

There are several different explanations for the engagement of the motor system in prediction. The common coding theory suggests a representational overlap between planning, execution, and perception of movement (Maes et al., 2014). In this view, inverse models allow intended or perceived sensory states to be mapped back onto corresponding motor commands. These sensory states can then be used to predict the sensory outcomes of planned actions (Maes et al., 2014, p1). The reverse, of identifying the motor command required to produce a particular sensory outcome, is implemented via the inverse model (Wolpert et al.,

2003). Similarly, Wolpert et al., (1995) proposed that when people produce a movement themselves, they predict the outcome of their actions by concurrently simulating them and running this simulation ahead.

An alternative view is the predictive coding account, which similarly relies on the idea of the forward model, but offers an alternative to the inverse model (Friston 2011; Friston et al., 2012). In predictive coding, perception relies on the identification of prediction errors resulting from differences between experienced and expected sensory input (Friston, 2005). According to this account, action is a direct consequence of prediction, and this complex set of predictions, which includes both proprioceptive and exteroceptive effects, removes the need for the inverse model (Clark, 2015; Pickering & Clark, 2014). Recently, this theory has been applied to both music and language (Blank & Davis, 2016; Koelsch et al., 2018; Okada et al., 2018).

Here we focus on the former approach, specifically focusing on the idea of simulating an action to generate predictions. When we observe others conducting an action, we use experience of what we have previously done under similar circumstances (i.e., our own motor experience) to predict what the observed individual will do (Wilson & Knoblich, 2005). For example, if we see someone starting to move their arm towards a target, we draw on our own motor experience to predict the timing and trajectory of the movement. Gallese (2005) suggested that social competence may depend on our capacity to comprehend the meaning of the actions we observe. If the observed actions are within or close to our motor repertoire, we can map them onto our own motor system, thus allowing understanding as well as predictions about the outcome of the observed actions.

This type of simulation theory has been applied to the prediction of speech by Pickering and Garrod (2013), and recently to the prediction of music by Hadley and

Pickering (2020). A variety of support for this proposal comes from evidence of motor engagement while listening to speech or to music, which we now consider in turn.

1.1.1. Motor Involvement in the Prediction of Language

The simulation theory of prediction proposes that listeners use their production systems to covertly imitate their partner's speech, which facilitates perception and allows them to predict what will come next (Pickering & Garrod, 2013; see also Dell & Chang, 2014; Pickering & Gambi, 2018). Several studies provide evidence for this proposal. Listening to speech, for example, activates areas of the brain associated with speech production (Londei et al., 2009; Skipper et al., 2005; Watkins et al., 2003; Watkins & Paus, 2004; Wilson et al., 2004). Moreover, the activated areas appear to be specific to what is being perceived; Fadiga et al., (2002) found that listening to words that involved using the tongue induced enhanced activation of the tongue region of the motor cortex compared to words that did not involve using the tongue. Furthermore, listening to real words induced a stronger motor facilitation effect compared to pseudowords, an effect further investigated and replicated by Roy et al. (2008). (See also Pulvermüller et al., 2006; Sundara et al., 2001).

In addition, manipulation of the production system during listening can affect people's perception of speech. Meister et al., (2007) found that disrupting the premotor cortex using transcranial magnetic stimulation (TMS) impaired speech perception. Furthermore, Möttönen and Watkins (2009) found that the impairment of speech perception was specific to the motor regions that were disrupted. They used an inhibitory TMS protocol to show that disruption of the lip area of the cortex impaired perception of speech that used those lip movements, but not of speech that did not use those lip movements, and also that disruption of the hand area had no effect. In addition, D'Ausilio et al. (2009) used a facilitatory TMS protocol in a double dissociation study to show that perception of speech sounds was

enhanced by stimulating the motor representation of the specific articulator responsible for production of that sound (i.e., the lip area enhanced lip-articulated phonemes, and the tongue area enhanced tongue-articulated phonemes; see also D'Ausilio et al., 2012; Möttönen et al., 2014; Smalle et al., 2014). These studies suggest that not only is the motor system activated while listening to speech, but that the motor system activation is finely tuned to the articulator in use during speech perception.

Further evidence suggests that this motoric activation is specifically related to prediction. Rommers et al., (2020) investigated the production effect (i.e., that reading a word aloud helps you to remember it better than if you read it silently) and found that it was smaller for predictable words than for unpredictable words. They suggest that this difference could be due to readers simulating predictable words during silent reading and therefore experiencing reduced benefit of actually speaking those words, in comparison to speaking the unpredictable (and un-simulated) words. This work supports the idea that predictions of speech are generated by the production system.

Furthermore, Drake and Corley (2015) demonstrated predictive motor engagement during listening by recording ultrasound images of tongue movement during a picture naming task. Participants heard sentence stems with a predictable final word (e.g., *When we want water, we just turn on the...*) and then immediately named a picture. In one condition, the picture matched the predictable word (here, *tap*), while in another condition just the initial consonant of the picture name differed from the predictable word (here, *cap*). In a control condition, participants were asked to name pictures without a sentence stem. When compared to movement in the control condition, there was a greater change in articulator positions (i.e., tongue movements) in the mismatch than the match condition 300-500ms before speech onset. They suggest that that predictions are made while we listen to another's speech, and these predictions affect the preparation of the motor execution of our own speech.

Finally, some research using event-related potentials suggests that motor interference affects listeners' ability to make linguistic predictions during reading. Martin et al., (2018) had participants read predictable sentences and found that the N400 (whose magnitude increases with unpredictability) was reduced when they concurrently produced the syllable *ta* (i.e., using their production systems) compared to when they concurrently tapped their tongue or when they listened to a recording of themselves producing *ta*. It appears that predictions were weaker when participants' production systems were engaged in the syllabic aspect of language production (and the effect was not simply due to comparable non-speech mouth movements). This evidence suggests that comprehenders use their production systems to predict upcoming language.

1.1.2. Motor Involvement in the Prediction of Music

The simulation theory of prediction has also been applied to music (Hadley & Pickering, 2020). As with the perception of speech, much evidence suggests that our motor systems are activated while we perceive music (e.g. Abrams et al., 2013; Buccino et al., 2004; Burunat et al., 2015; Chen et al., 2008; Popescu et al., 2004). It is argued that most people can experience and appreciate music even without having the skills to produce it (especially beyond Western conceptions of music-making and expertise). However, in music, in contrast to speech, people can have vastly different levels of motor experience, and it has been suggested that motor experience may contribute to how music is perceived/ experienced (Patel & Iversen, 2014; Overy & Molnar-Szakacs, 2009; see also Schiavio et al., 2014). This variability in musical ability in comparison to speech ability may provide us with new ways to address the links between motor experience and motor engagement during listening.

We now turn to research that investigates motor activation during listening to music. A number of studies suggest that the strength of motor activation while listening to music is

dependent on the listener's ability to play that music. A TMS study investigating the effect of rehearsal on motor engagement found that even after a very short time practicing, pianists show greater motor excitation when listening to a rehearsed compared to an unrehearsed piano piece (D'Ausilio et al., 2006; see also Lahav et al., 2007). Furthermore, the motor activation elicited during music listening appears to be related to the specific movements required to play the piece being heard, just like the articular-specific effects found for speech (D'Ausilio et al., 2006). Candidi et al., (2014) used TMS to investigate the motor excitability of pianists' hand muscles when observing silent videos of a pianist playing piano scales and occasionally making errors. They found that pianists showed motor activity in the hand area that was specific to the finger making the error, thus providing some evidence that simulation is used as a means of processing an observed musician's movements and thus musical output. Furthermore, Haueisen and Knösche (2001) used MEG to show that pianists activate their motor cortex predictively while listening to music, and that this activity is specific to the finger about to be used to play a note (see also Bangert et al., 2006; Baumann et al., 2005).

It has also been shown that use of the production system during music listening impacts music perception, just as use of the production system during speech listening impacts speech perception. For example, when pianists and non-pianists pressed piano keys to produce a sequence of two perceptually bistable tones (i.e., tones that can be perceived to form either a rising or falling interval), only the pianists tended to perceive the interval direction as complying with the direction of their finger movements at the piano (Repp & Knoblich, 2007; see also Taylor & Witt, 2014). Moreover, music perception affects music production. Drost et al., (2005) found that the ability to play music is impaired by the concurrent perception of incongruent music: Hearing an incongruent interval while making the finger movements necessary to play the congruent interval slowed pianists but not non-musicians (trained to play the chords for the experiment) in carrying out the task.

The above studies indicate that motor system activity can contribute to music perception, and that motor engagement is particularly likely when listeners are able to play the music themselves. This work does not, however, determine whether motor activation is involved in generating predictions. One study that begins to address this issue is Novembre and Keller (2011). Pianists imitated a hand playing a five-chord sequence, without hearing either the hand's performance or their own. Imitation of the final (i.e., fifth) chord in the sequence was faster when the penultimate chord provided a congruent harmonic context than an incongruent harmonic context. This finding is compatible with participants using the hand's movement sequence to predict the final chord. Moreover, the authors suggest this process explains how players predict each other during musical interaction – that is, monitoring of the other player influences how and when they themselves play. In addition, Novembre et al., (2014) used TMS to show that stimulating the right motor cortex of pianists during a duet (which is involved in simulating the left hand) disrupted their ability to coordinate with a left-handed part that they had previously rehearsed themselves, but not with a left-handed part that they had not previously rehearsed themselves. Finally, Hadley et al., (2015) used TMS in a similar paradigm but stimulated the dorsal premotor cortex to address its involvement in turn-taking timing. They found that stimulating this area during a duet also led to poorer turn-taking between players when they had previously rehearsed their partner's part than when they had not rehearsed their partner's part. Together, this work suggests that duetting pianists automatically represent heard music using their own motor system and use this representation to make predictions about their partner.

In sum, people's motor systems appear to be active while they perceive both speech and music. Furthermore, some studies suggest that people's motor systems are involved in making predictions while perceiving another person speak or play music, particularly when observing actions within their own motor repertoire. There is also evidence that motor

activation is specific to the act being perceived – that is, the articulator for the specific sound in speech, or the finger for the specific note in music.

1.2. Current Study

In our study, we hypothesise that if prediction relies on mechanisms involved in production, disrupting participants' ability to simulate heard stimuli (through concurrent engagement in another motor task) will interfere with their ability to make predictions of a turn end during listening. In other words, if prediction is based on production, then interfering with production will interfere with prediction. Furthermore, we have discussed the evidence that motor activity is specific to the task being observed (i.e., relating to the specific articulator or finger involved, and being greater for practiced stimuli). Thus, our second hypothesis is that task-specific, but not task-non-specific, engagement of the effector used to speak or play music will be detrimental to turn-end prediction.

To test these hypotheses we conduct two dual-task experiments, in which participants predict the end of a linguistic (i.e., a question) or musical (i.e., a piano melody) pre-recorded turn while conducting a secondary task. This secondary task involved either task-non-specific engagement of the relevant effector (i.e., a non-speech mouth movement or non-piano-playing hand movement) or task-specific engagement of the effector (i.e., a speech movement or a piano-playing movement), thus allowing us to test the specificity of motor activity. In both experiments, we compare these two conditions with a third baseline condition, in which participants predict the turn-ends while listening only. We predict that if motor engagement is at the level of the specific task, then the speech condition should impair predictions of speech but the mouth movement condition should not, and the piano movement condition should impair predictions of music but the hand movement condition should not. We additionally

report an exploratory analysis to address whether the impact of the secondary task depends on stimulus duration, based on prior work finding people generally respond earlier and less accurately for longer stimuli (De Ruiter et al., 2006; Hadley et al., 2018).

2. Results

Thirty-four participants aged 18-68 years (M age =30.0, $SD=14.6$) with a mean of 20.3 years ($SD = 16.7$) piano-playing experience and $M = 6.3$ hours ($SD = 7.2$) piano practice per week took part. Please note that four participants' music data were lost. More details on participants can be found in the methodology section below.

2.1. Speech Experiment Results

The mean response (across conditions) was 90 ms after stimulus offset, with a standard deviation of 260 ms (see Figure 1).

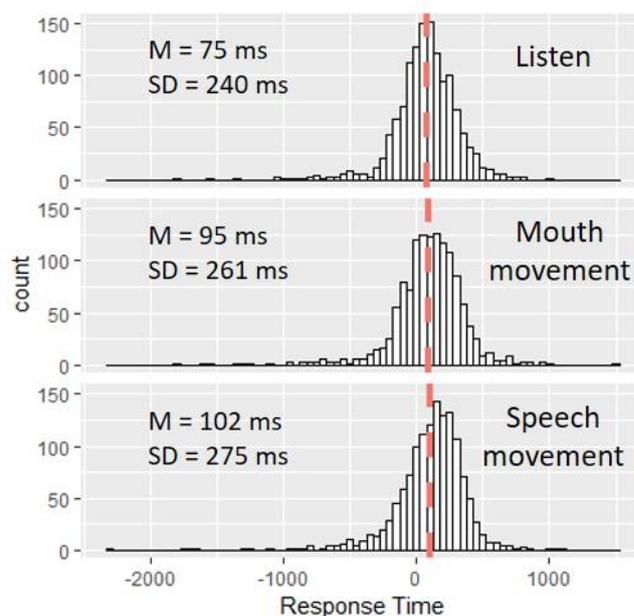


Figure 1. Distributions of relative response times in the three interference tasks for the speech experiment, with mean response times plotted as a red line. Means (M) and standard deviations (SD) are reported in the top left of each panel.

We found a main effect of Interference task ($F(2,3851.9)=6.85, p=.001$). Both the *Mouth Movement* condition and the *Speech Movement* conditions led to delayed responses in comparison to the baseline condition (*Mouth Movement vs Baseline*: difference=20 ms, $p=.003$, *Speech Movement vs Baseline*: difference=27 ms, $p=.005$), but did not differ from each other (*Mouth Movement vs Speech Movement*: difference=7 ms, $p=.97$). Hence using the mouth during speech listening disrupted prediction, but there was no difference according to whether the movement was task-non-specific or task-specific. There was a main effect of Stimulus Duration ($F(1,117.3)=73.28, p<.001$), such that participants responded earlier when questions were longer. No other effects were significant (Predictability: $F(1,114.8)=0.81, p=.37$; Predictability x Interference task: $F(2,3861.0)=0.54, p=.58$).

In an additional exploratory analysis, we used a median split to investigate shorter and longer stimulus responses separately (short stimuli ≤ 2740 ms, long stimuli >2740 ms). See Table 1. No effects were significant in the long stimuli ($ps>.34$). However, in the short stimuli there was a significant effect of Interference task ($F(2,1969.77)=8.42, p<.001$). For these short stimuli, participants responded earlier in the *Baseline* condition than either the *Mouth Movement* condition (difference=44 ms, $p=.002$) or the *Speech Movement* condition (difference=48 ms, $p<.001$), but responses did not differ between *Mouth Movement* and *Speech Movement* conditions (difference=4 ms, $p=.98$). Hence for the shorter stimuli only, any form of mouth movement delayed turn-end prediction compared to no mouth movement. No other effects were significant in the short stimuli (Predictability: $F(1,58.97)=0.22, p=.64$; Predictability x Interference task: $F(2,1974.276)=0.40, p=.67$).

	Short		Long	
	Mean	SD	Mean	SD
Baseline	114	193	31	279
Mouth Movement	157	223	36	280
Speech Movement	162	221	37	312

Table 1. Means and standard deviations of response times (in ms) for each interference condition in the speech experiment, for short and long stimuli separately (median split).

2.2. Music Experiment Results

The mean response (across conditions) was 1528 ms before stimulus offset (i.e., -1528ms), with a standard deviation of 2696 ms (see Figure 2).

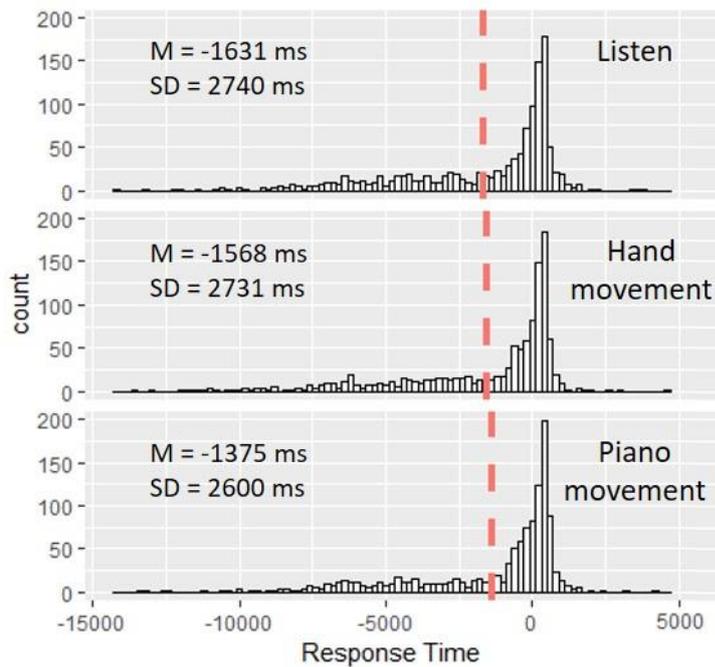


Figure 2. Distributions of response times in the three interference tasks for the music experiment, with mean response times plotted as a red line. Means (M) and standard deviations (SD) are reported in the top left of each panel.

We found a main effect of Interference task ($F(2,3353.6)=8.07, p<.001$). The *Piano Movement* condition led to delayed responses in comparison to the baseline condition (difference=256 ms, $p=.003$), and the *Piano Movement* condition also led to delayed responses in comparison to the *Hand Movement* condition (difference=193 ms, $p<.001$). The *Baseline* and *Hand Movement* conditions did not differ from each other (difference=63 ms, $p=.92$). As in the speech analysis, there was a main effect of Stimulus Duration ($F(1,157.4)=589.79, p<.001$), such that participants responded earlier to long than short stimuli.

In an additional exploratory analysis, we used a median split to investigate shorter and longer stimulus responses separately (short stimuli ≤ 7579 ms, long stimuli > 7579 ms). See Table 2. In the long stimuli there was a significant effect of Interference task ($F(2,1560.7)=9.50, p<.001$). For these long stimuli, participants responded later in the *Piano Movement* condition than either the *Baseline* condition (difference=361 ms, $p=.04$) or the *Hand Movement* condition (difference=595 ms, $p<.01$), but responses did not differ between *Baseline* and *Hand Movement* conditions (difference=234 ms, $p=.11$). In the short stimuli there was also a significant effect of Interference task ($F(2,1683.4)=3.64, p=.03$). However, for these short stimuli, while participants responded later in the *Piano Movement* condition than the *Baseline* condition (difference=66 ms, $p=.02$), there was no difference between the *Piano Movement* and the *Hand Movement* conditions (difference=39 ms, $p=.23$), or the *Hand Movement* and *Baseline* conditions (difference=27 ms, $p=.54$).

	Short		Long	
	Mean	SD	Mean	SD
Baseline	-465	1544	-2785	3147

Hand Movement	-438	1404	-3019	3282
Piano Movement	-399	1501	-2424	3081

Table 2. Means and standard deviations of responses times (in ms) for each interference condition in the music experiment, for short and long stimuli separately (median split)

3. Discussion

The aim of this study was to determine whether motor activity plays a causal role in turn-end predictions in speech and music. Based on the prediction-by-simulation model, which assumes that individuals use their own motor systems to predict the actions of others, we hypothesised that disrupting participants' ability to simulate heard stimuli (through engagement of the relevant effector in another motor task) would interfere with their ability to make turn-end predictions. Furthermore, on the basis that motor engagement is at the level of the specific task, we predicted that task-specific, but not task-non-specific, engagement of the effector would interfere with turn-end prediction. For speech, we found that either task-specific or task-non-specific mouth movement disrupted turn-end prediction of questions compared to listening only. For music, in contrast, we found that only the task-specific piano movement disrupted turn-end prediction compared to listening only – a task-non-specific hand movement did not.

In the speech experiment, we found that disrupting participants' ability to use their motor system in either a task-specific or non-task-specific way slowed their turn-end predictions. Our results support the idea that people's motor systems are used to make predictions about their partner's speech. These results are consistent with the prediction-by-simulation theory of speech, in which people covertly imitate the speaker in order to comprehend and make predictions about their own and their partners' upcoming speech (Dell

& Chang, 2014; Pickering & Gambi, 2018; Pickering & Garrod, 2013). Our speech findings, however, did not show this motor activation to be task-specific, suggesting that the general effector condition was as disruptive to the motor system as the task-specific effector condition.

In the music experiment, on the other hand, predictions were disrupted only by task-specific activity, suggesting that people simulate the specific motor movements that are needed to produce the movement being heard. It is curious that we see this difference only in the music and not in the speech experiment. It would be interesting to investigate whether non-pianists would show the same findings as the pianists tested here. If not, this finding would provide support for theories that suggest that motor activation during perception is finely tuned to the musical movement being perceived (Candidi et al., 2014), and support the proposal that simulation is stronger for movements that are in the listener's motor repertoire (Bangert et al., 2006; Baumann et al., 2005; D'Ausilio et al., 2006; Drost et al., 2005, 2007; Haueisen & Knösche, 2001; Proverbio & Orlandi, 2016). If non-pianists did show the same effects, it would instead provide support for Patel and Iversen's (2014) "action simulation for auditory perception" (ASAP) theory that suggests that simulation of rhythmic movements more broadly is used to help predict the timing of upcoming beats in humans (and perhaps non-human primates). Either way, together with the current findings, it would provide support for embodied theories of simulation as discussed earlier (see Schiavio et al., 2015, Maes et al., 2014; Gallese, 2005).

We see two more possibilities for the difference between the speech and music experiment findings (i.e., effector-specific interference in speech and task-specific interference in music). One possibility is that the general hand movement condition is further from the specific piano movement condition than the mouth movement is from the speech movement condition. Unfortunately, we have no way of assessing this, though future studies

could make use of different musical movements. A second possible explanation for more specific motor engagement for music than speech relates to the need to separate one's own actions from those of others. In music, it is common to be simultaneously playing your own part and listening to the output of others. On the other hand, it is much less common to simultaneously produce speech and listen to somebody else talk (while overlaps do occur, they are rarely for extended periods). Hence when listening to music, it is possible that this greater specificity of representation is a means of avoiding one's own movements being disrupted by the perception of others' parts.

Interestingly, people responded on average 90ms after stimulus offset for the speech stimuli, but 1528ms after stimulus offset for the music stimuli. This difference could be the result of the musical stimuli being much longer (on average) than the linguistic stimuli (7957 vs 2824ms). However, it is unlikely that such length differences explain the differing patterns of interference across domains, as duration was included as a fixed effect in each analysis. In our exploratory analyses that split short from long stimuli, we found the previously reported speech effects only in the shorter stimuli, whereas the music effects occurred regardless of stimulus duration. It is possible that the speech effects occurred only for the shorter stimuli because those stimuli were so brief (and the prediction task was presumably challenging) that any sort of motor engagement impaired the process.

While we have identified a potential key role of simulation in prediction of spoken and musical turn-ends, there are a variety of opportunities for further study. It is possible that the interference conditions in the speech experiment did not induce sufficiently different motor simulation from each other to produce differences in turn-end prediction. That is, opening and closing the mouth may have disrupted the simulation of the articulators responsible for producing speech in the same (or similar) way to the speech movement condition. A future study could include a motor task that is in no way related to the motor

system activated, for example hand movement during the speech stimuli, to ensure any effects were not simply due to dual-task interference. It is also possible that the random ordering of the conditions (switching from mouth movement to speech movement between trials for example) may have affected our results. The cognitive resources required to switch tasks at the start of each trial may have led to reduced resources to allocate to simulation, making the differences between conditions more striking (e.g., Jongman & Meyer, 2017). On the other hand, the random condition ordering could have led to enhanced motor activity on trials in which there was no motor task (carryover), leading to benefit in the baseline condition. A simple way of testing the impact of random ordering would be to run the experiments again using separate blocks per condition, though this may introduce practice effects, with participants becoming increasingly efficient at a particular task. Finally, it would be interesting to investigate how these findings extend to communication with a real partner, when accuracy of predicting the end of a spoken or musical turn impacts interaction success.

In summary, the results from both the speech experiment and from the music experiment support the claim that simulation is used during turn-end prediction. In the speech study, we found that task-specific and task-non-specific mouth movement disrupted turn-end prediction of questions compared to the listening only condition. In the music study, we found that only the task-specific piano movement disrupted turn-end prediction compared to task-non-specific hand movement and listening. These results suggest that the motor system is activated while we perceive speech and music, and this activation helps us to make accurate turn-end predictions, but that the specificity of the motor effects may differ across domains.

4. Experimental Procedure

4.1. Participants

Thirty-four pianists were paid £9 to take part in this study (19 female, 32 native English speakers). Participants were aged 18-68 years (M age =30.0, SD =14.6), with 26 reporting as right-handed, 6 as left-handed, and 2 as ambidextrous. Participants self-reported a mean of 20.3 years (SD = 16.7) piano-playing experience and a mean of 6.3 hours (SD = 7.2) of piano practice per week. Note that four participants' music data were lost (as a result of recording issues). This study was approved by the University of Edinburgh Psychology Research Ethics Committee.

4.2. Materials

Linguistic stimuli. We selected 120 questions of 3-15 words in length from Corps et al., (2018), such as *Why do people wear a coat in the winter?* They were recorded by a female speaker and had a mean duration of 2824ms (range 1183-5561ms). A full list of the questions can be found in the supplementary materials.

Musical stimuli. We selected 120 melodies of 4-14 bars length, varying in key and tempo, from Grade 4 and 5 sight-reading books (ABRSM, 2008a, 2008b). They were recorded by a pianist and had a mean duration of 7957ms (range 3320-17833ms). A full list of the melodies can be found in the supplementary materials.

4.3. Procedure

Items were presented on an 18-inch monitor, approximately 60cm from the participant. Participants listened to the items over a pair of 'Beyerdynamic DT 109' headphones and stimuli were presented using E-Prime 2.0. Reaction times were recorded using an SR Response Box, which was positioned in front of the monitor to the left (in both

the speech and the music experiments). All participants performed the hand and piano movement conditions on the right side of the table.

Before beginning the experiment, participants read the instructions and watched silent, five second instruction videos for the two movement conditions (detailed below). In the *Baseline* condition, participants were asked to simply listen.

Speech Experiment: In the *Mouth Movement* condition participants were instructed to open and close their mouths, and then they were shown a female demonstrating the action (see Figure 3). In the *Speech Movement* condition, participants were instructed to mouth the phrase “I am mouthing a phrase” and then saw the same female mouthing the phrase (see Figure 3). In both conditions, only the bottom half of the face was visible, and the videos looped for five seconds; this meant that the mouth opened and closed five times and the phrase was mouthed three times.

Music Experiment: In the *Hand Movement* condition, participants were instructed to trace a circle on the table with their index finger, and then they saw a video of a hand demonstrating the action (see Figure 3). In the *Piano Movement* condition, participants were instructed to finger a two-bar piano melody and then the two-bar melody was presented to them (see Figure 3) while they simultaneously watched a hand demonstrating that action (see Figure 3). Again, in both conditions the videos looped for five seconds so that the finger traced the circle a five times and the melody was fingered three times.



Mouth Movement Condition



Hand Movement Condition



Speech Movement Condition



Piano Movement Condition

Figure 3: Stills from the movement conditions. Top left: the mouth movement condition (opening and closing mouth); bottom left: the speech movement condition (mouthing a phrase); top right: the hand movement condition (tracing a circle); bottom right: the piano movement condition (fingering a piano melody)



Figure 4: Piano melody presented to participants for the piano movement condition.

Before each trial, a fixation cross was displayed on-screen until participants pressed a button on the SR-box to begin audio playback of the question or melody. The words 'LISTEN', 'MOVE', or 'PHRASE' were displayed on-screen at the beginning of each trial (and audio) during both experiments for the duration of playback (but not before) to indicate to participants which task they should complete while listening. Based on instructions used in De Ruiter et al. (2006), participants used their left hand to indicate when they thought the question or melody would end. They were asked to: 'Press the button when you believe the question/melody will end. Do not wait until the speaker/player has finished the

question/melody and stopped speaking/playing. Instead, you should press the button as soon as you expect the speaker/player to finish.’ Thus, participants were encouraged to predict the question or melody end, rather than simply waiting for the end of the audio before responding. Audio playback stopped and the trial terminated as soon as the participant pressed the button (as in De Ruiter et al., 2006). If the participant did not press the button 5s after the end of the stimulus, then the trial was recorded as missed and the next trial began.

The order of the speech and music experiments was counterbalanced, so that half of the participants completed the speech experiment before the music experiment, and vice versa for the other half. If participants began with the speech experiment, they first received full instructions and videos for the speech experiment just before beginning the speech experiment. Once the speech experiment was completed, they received full instructions and videos for the music experiment and then completed the music experiment. At the start of each experiment, participants completed one practice trial for each condition to familiarise themselves with the procedure. The instructions were also verbally repeated and physically demonstrated by the experimenter to ensure understanding at the beginning of each experiment.

Participants were presented with the 120 experimental trials that were randomly ordered within each condition (trials in each condition were counterbalanced across participants) for each participant (40 in *Baseline*, 40 in *Mouth/Hand Movement* condition, and 40 in *Speech/Piano Movement* condition). Each participant therefore saw each stimulus in one condition only, and the task alternated randomly trial by trial. Participants were given an opportunity to take a break half-way through each experiment. After the break, they were again presented with the relevant instruction videos to ensure they continued to carry out the movements in the movement conditions.

4.4. Analysis

We analysed the data from the speech and music experiments separately. Response times were calculated relative to stimulus offset, and so were negative when participants responded early (i.e., before the end of the question or melody) and positive when they responded late (i.e., after the end of the question or melody). Trials with responses more than 2.5SD from participants' means (in either direction) were then removed (2.92% of speech trials, 3.36% of music trials).

Analyses were conducted in R using the package lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017). We were interested in whether engaging the mechanisms of production would interfere with turn-end prediction. We tested this hypothesis in the speech experiment by fitting a model in which response times were predicted by Interference Task (baseline, mouth movement, or speech movement). But it is also worth noting that half of the questions were predictable in length, so that participants could estimate the number of words until turn-end, while the other half were unpredictable, so that they could not (see Corps et al., 2018). Thus, we also included Length Predictability (predictable vs. unpredictable) and its interaction with interference task as a fixed effect to determine whether listeners predict turn-endings by predicting the number of words the speaker will use. Finally, previous research using the button-pressing task has consistently found effects of stimulus duration, such that longer turns tend to elicit earlier button-presses than shorter turns (e.g., Corps et al., 2018; Corps, Pickering, & Gambi, 2018; De Ruiter et al., 2006; Hadley et al., 2018), and so we included (centred) Stimulus Duration in our model. To aid model convergence, this predictor was included as a fixed effect only. We included random effects for both by-participant and by-item intercepts (more complex models would not converge). In the music experiment, button-press times were predicted by Interference Task (baseline, hand movement, or piano movement) and Stimulus Duration as fixed effects. Again, we included

random effects for both by-participant and by-item intercepts (more complex models would not converge).

While we included Stimulus Duration as a fixed effect in our primary analyses, we conducted additional exploratory analyses to investigate participants' sensitivity to stimulus duration. Stimuli in both experiments varied in duration (range 1183-5561 ms in the speech experiment, and 3320-17833 ms in the music experiment), and we explored whether effects occurred in the short and long stimuli by using a duration median split and running separate analyses on short and long stimuli. These analyses used the same model structure as our main analyses, but stimulus duration was not included as a fixed effect.

Throughout, significant predictors are reported using ANOVA with degrees of freedom determined using the Satterthwaite method. Subsequent pairwise comparisons were performed using the package emmeans (Lenth et al., 2018), with Tukey adjustment.

References

- Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., & Menon, V. (2013). Inter-subject synchronization of brain responses during natural music listening. *European Journal of Neuroscience*, *37*(9), 1458–1469. <https://doi.org/10.1111/ejn.12173>
- ABRSM. (2008a). *Piano Specimen Sight-Reading Tests ABRSM Grade 4*. ABRSM.
- ABRSM. (2008b). *Piano Specimen Sight-Reading Tests ABRSM Grade 5*. ABRSM.
- Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H. J., & Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, *30*(3), 917–926. <https://doi.org/10.1016/j.neuroimage.2005.10.044>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Baumann, S., Koeneke, S., Meyer, M., Lutz, K., & Jäncke, L. (2005). A network for sensory-motor integration: what happens in the auditory cortex during piano playing without acoustic feedback? *Annals of the New York Academy of Sciences*, *1060*, 186–188. <https://doi.org/10.1196/annals.1360.038>
- Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS biology*, *14*(11), e1002577. <https://doi.org/10.1371/journal.pbio.1002577>
- Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H.-J., & Rizzolatti, G. (2004). Neural Circuits Underlying Imitation Learning of Hand Actions. *Neuron*, *42*(2), 323–334. [https://doi.org/10.1016/S0896-6273\(04\)00181-3](https://doi.org/10.1016/S0896-6273(04)00181-3)

- Burunat, I., Brattico, E., Puoliväli, T., Ristaniemi, T., Sams, M., Toiviainen, P., & Snyder, J. (2015). Action in perception: Prominent visuo-motor functional symmetry in musicians during music listening. *PLoS ONE*, *10*(9). <https://doi.org/10.1371/journal.pone.0138238>
- Candidi, M., Maria Sacheli, L., Mega, I., & Aglioti, S. M. (2014). Somatotopic Mapping of Piano Fingering Errors in Sensorimotor Experts: TMS Studies in Pianists and Visually Trained Musically Naïves. *Cerebral Cortex*, *24*(2), 435–443. <https://doi.org/10.1093/cercor/bhs325>
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to Musical Rhythms Recruits Motor Regions of the Brain. *Cerebral Cortex*, *18*, 2844–2854. <https://doi.org/10.1093/cercor/bhn042>
- Clark, A. (2015). Embodied Prediction. In T. Metzinger & J. M. Windt (Eds). *Open MIND*. Frankfurt am Main: MIND Group. doi: 10.15502/9783958570115
- Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition*, *175*, 77-95. doi:10.1016/j.cognition.2018.01.015.
- Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse processes*, *55*(2), 230-240. doi:10.1080/0163853X.2017.1330031.
- Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics*, *31*(2), 139–148. [https://doi.org/10.1016/S0095-4470\(02\)00082-7](https://doi.org/10.1016/S0095-4470(02)00082-7)
- D'Ausilio, A., Altenmüller, E., Olivetti Belardinelli, M., & Lotze, M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *European Journal of Neuroscience*, *24*(3), 955–958. <https://doi.org/10.1111/j.1460->

9568.2006.04960.x

D'Ausilio, Alessandro, Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, *48*(7), 882–887.

<https://doi.org/10.1016/j.cortex.2011.05.017>

D'Ausilio, Alessandro, Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The Motor Somatotopy of Speech Perception. *Current Biology*, *19*(5), 381–385.

<https://doi.org/10.1016/j.cub.2009.01.017>

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515–535.

https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_60156_3

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394. <https://doi.org/10.1098/rstb.2012.0394>

<https://doi.org/10.1098/rstb.2012.0394>

Drake, E., & Corley, M. (2015). Articulatory imaging implicates prediction during spoken language comprehension. *Memory & Cognition*, *43*(8), 1136–1147.

<https://doi.org/10.3758/s13421-015-0530-6>

Drost, U. C., Rieger, M., Brass, M., Gunter, T. C., & Prinz, W. (2005). Action-effect coupling in pianists. *Psychological Research*, *69*(4), 233–241.

<https://doi.org/10.1007/s00426-004-0175-8>

Drost, U. C., Rieger, M., & Prinz, W. (2007). Instrument specificity in experienced musicians. *Quarterly Journal of Experimental Psychology*, *60*(4), 527–533.

<https://doi.org/10.1080/17470210601154388>

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically

modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15(2), 399–402. <https://doi.org/10.1046/j.0953-816x.2001.01874.x>

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences* 29, 360 (1456), 815-836. 10.1098/rstb.2005.1622

Friston, (2011). What is optimal about motor control? *Neuron*, 72 (3), 488-498.
10.1016/j.neuron.2011.10.018

Friston, K., Adams, R. A., Perrinet, L. & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, 3 (151), 1-20.
10.3389/fpsyg.2012.00151

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, 4, 23-48. <https://doi.org/10.1007/s11097-005-4737-z>

Gallese V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362(1480), 659-669.
<https://doi.org/10.1098/rstb.2006.2002>

Hadley, L. V., Novembre, G., Keller, P. E., & Pickering, M. J. (2015). Causal role of motor simulation in turn-taking behavior. *Journal of Neuroscience*, 35(50), 16516–16520.
<https://doi.org/10.1523/JNEUROSCI.1850-15.2015>

Hadley, L. V., & Pickering, M. J. (2020). A neurocognitive framework for comparing linguistic and musical interactions. *Language, Cognition and Neuroscience*, 35(5), 559–572. <https://doi.org/10.1080/23273798.2018.1551556>

Hadley, L. V., Sturt, P., Moran, N., & Pickering, M. J. (2018). Determining the end of a

musical turn: Effects of tonal cues. *Acta Psychologica*, 182, 189–193.

<https://doi.org/10.1016/j.actpsy.2017.11.001>

Haruno, M., Wolpert, D. M., & Kawato, M. (2003). Hierarchical MOSAIC for movement generation. *International Congress Series*, 1250(C), 575–590.

[https://doi.org/10.1016/S0531-5131\(03\)00190-0](https://doi.org/10.1016/S0531-5131(03)00190-0)

Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, 13(6), 786–792.

<https://doi.org/10.1162/08989290152541449>

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144.

<https://doi.org/10.1016/j.cognition.2002.06.001>

Jongman, S. R., & Meyer, A. S. (2017). To plan or not to plan: Does planning for production remove facilitation from associative priming?. *Acta psychologica*, 181, 40-50.

<https://doi.org/10.1016/j.actpsy.2017.10.003>

Keller, P. E., & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception*,

28(1), 27–46. <https://doi.org/10.1525/mp.2010.28.1.27>

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, 23(1), 63-77.

<https://doi.org/10.1016/j.tics.2018.10.006>

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.

<https://doi.org/10.18637/jss.v082.i13>

- Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, 27(2), 308–314. <https://doi.org/10.1523/JNEUROSCI.4822-06.2007>
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Estimated marginal means, aka least-squares means. *R Package Version*, 1(1), 3.
- Levinson, S. C. (2016). Turn-taking in Human Communication - Origins and Implications for Language Processing. *Trends in Cognitive Sciences*, 20(1), 6–14. <https://doi.org/10.1016/j.tics.2015.10.010>
- Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Gratta, C. Del, Romani, G.-L., & Belardinelli, M. O. (2009). Sensory-motor brain network connectivity for speech comprehension. *Human Brain Mapping*, 31(4), NA-NA. <https://doi.org/10.1002/hbm.20888>
- Maes, P. J., Leman, M., Palmer, C., & Wanderley, M. (2014). Action-based effects on music perception. *Frontiers in psychology*, 4, 1008. doi: 10.3389/fpsyg.2013.01008.
- Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is Production: The missing link between language production and comprehension. *Scientific Reports*, 8(1), 1079. <https://doi.org/10.1038/s41598-018-19499-4>
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The Essential Role of Premotor Cortex in Speech Perception. *Current Biology*, 17(19), 1692–1696. <https://doi.org/10.1016/j.cub.2007.08.064>
- Möttönen, R., van de Ven, G. M., & Watkins, K. E. (2014). Attention fine-tunes auditory-motor processing of speech sounds. *Journal of Neuroscience*, 34(11), 4064–4069. <https://doi.org/10.1523/JNEUROSCI.2214-13.2014>

- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, 29(31), 9819–9825.
<https://doi.org/10.1523/JNEUROSCI.6018-08.2009>
- Novembre, G., & Keller, P. E. (2011). A grammar of action generates predictions in skilled musicians. *Consciousness and cognition*, 20(4), 1232-1243.
- Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2014). Motor simulation and the coordination of self and other in real-time joint action. *Social cognitive and affective neuroscience*, 9(8), 1062-1068.
- Okada, K., Matchin, W., & Hickok, G. (2018). Neural evidence for predictive coding in auditory cortex during speech production. *Psychonomic bulletin & review*, 25(1), 423
- Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music perception*, 26(5), 489-504.430.
<https://doi.org/10.1525/mp.2009.26.5.489>
- Patel, A.D., & Iversen, J.R. (2018) The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, 8:57. doi: 10.3389/fnsys.2014.00057
- Pecenka, N., & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research*, 211(3–4), 505–515.
<https://doi.org/10.1007/s00221-011-2616-0>
- Phillips-Silver, J., & Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience*, 6, 26.
<https://doi.org/10.3389/fnhum.2012.00026>
- Pickering, M. J., & Clark, A. (2014). Getting ahead: forward models and their place in

cognitive architecture. *Trends in cognitive sciences*, 18(9), 451-456.

<https://doi.org/10.1016/j.tics.2014.05.006>

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144(10), 1002–1044.

<https://doi.org/10.1037/bul0000158>

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347.

<https://doi.org/10.1017/S0140525X12001495>

Popescu, M., Otsuka, A., & Ioannides, A. A. (2004). Dynamics of brain activity in motor and frontal cortical areas during music listening: A magnetoencephalographic study.

NeuroImage, 21(4), 1622–1638. <https://doi.org/10.1016/j.neuroimage.2003.11.002>

Proverbio, A. M., & Orlandi, A. (2016). Instrument-specific effects of musical expertise on audiovisual processing (Clarinet vs. Violin). *Music Perception*, 33(4), 446–456.

<https://doi.org/10.1525/MP.2016.33.4.446>

Pulvermüller, F., Huss, M., Kherif, F., Del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 103(20), 7865–7870.

<https://doi.org/10.1073/pnas.0509989103>

Rasch, R. A. (1979). Synchronization in Performed Ensemble Music. *Acta Acustica United with Acustica*, 43(2), 121–131.

Repp, B. H., & Knoblich, G. (2007). Action can affect auditory perception: Short report.

Psychological Science, 18(1), 6–7. <https://doi.org/10.1111/j.1467-9280.2007.01839.x>

Rommers, J., Dell, G. S., & Benjamin, A. S. (2020). Word predictability blurs the lines

between production and comprehension: Evidence from the production effect in memory. *Cognition*, 198, 104206. <https://doi.org/10.1016/j.cognition.2020.104206>

Roy, A. C., Craighero, L., Fabbri-Destro, M., & Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. *Journal of Physiology - Paris*, 102(1–3), 101–105.
<https://doi.org/10.1016/j.jphysparis.2008.03.006>

Schiavio, A., Menin, D., & Matyja, J. (2014). Music in the flesh: Embodied simulation in musical understanding. *Psychomusicology: Music, Mind, and Brain*, 24(4), 340.
<https://doi.org/10.1037/pmu0000052>

Shaffer, L. H. (1984). Timing in Solo and Duet Piano Performances. *Quarterly Journal of Experimental Psychology*, 36A(4), 577–595.
<https://doi.org/10.1080/14640748408402180>

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: Motor cortical activation during speech perception. *NeuroImage*, 25(1), 76–89.
<https://doi.org/10.1016/j.neuroimage.2004.11.006>

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. <https://doi.org/10.1093/cercor/bhu218>

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K. E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10587–10592.
<https://doi.org/10.1073/pnas.0903616106>

- Sundara, M., Namasivayam, A. K., & Chen, R. (2001). Observation-execution matching system for speech: a magnetic stimulation study. *NeuroReport*, *12*(7), 1341–1344.
- Taylor, J. E. T., & Witt, J. K. (2014). Listening to music primes space: pianists, but not novices, simulate heard actions. *Psychological Research*, *79*(2), 175–182.
<https://doi.org/10.1007/s00426-014-0544-x>
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989–994.
[https://doi.org/10.1016/S0028-3932\(02\)00316-0](https://doi.org/10.1016/S0028-3932(02)00316-0)
- Watkins, K., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience*, *16*(6), 978–987.
<https://doi.org/10.1162/0898929041502616>
- Wilson, M., & Knoblich, G. (2005). The Case for Motor Involvement in Perceiving Conspicifics. *Psychological Bulletin*, *131*(3), 460–473. <https://doi.org/10.1037/0033-2909.131.3.460>
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Jacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701–702. <https://doi.org/10.1038/nn1263>
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882. <https://doi.org/10.1126/science.7569931>
- Wolpert, D. M., Doya, K. & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London*, *358* (1431), 593-602.
- Yeung, H. H., & Werker, J. F. (2013). Lip Movements Affect Infants' Audiovisual Speech

Perception. *Psychological Science*, 24(5), 603–612.

<https://doi.org/10.1177/0956797612458802>