LANGUAGE PRIMING BY MUSIC AND SPEECH: EVIDENCE OF A SHARED PROCESSING MECHANISM

MELISSA K. JUNGERS, JULIE M. HUPP, & SARA D. DICKERSON
The Ohio State University at Newark

SOME SOUNDS FROM MUSIC AND LANGUAGE FOLLOW domain-specific syntactic rules and may be processed by distinct brain regions. However, domain-general mechanisms may explain some connections between music and language. The current study tested how rate differences in music and language perception influence the subsequent production of speech. In two experiments, participants were primed with fast or slow sentences (Experiment 1) or melodies (Experiment 2) and then produced novel picture descriptions. Participants’ rate of production was influenced by the rate of the prime, both within-domain (language-language) and cross-domain (music-language). A shared mechanism for domain-general rate processing is proposed.

Received: December 20, 2014, accepted October 3, 2015.

Key words: cross-domain priming, music, psycholinguistics, temporal processing

ALTHOUGH MUSIC AND LANGUAGE ARE OFTEN compared, research has not yet established whether or not these domains represent distinct types of knowledge with separable processing and production systems. Fodor (1983) argues for the modularity of mind, with specific cognitive functions, such as language processing, encapsulated and not influenced by other areas. This debate continues, as researchers theorize about the cognitive structures of the mind (Gottschling, 2009). For example, neuropsychological evidence supports this notion of domain specificity for music and language. Music deficits (amusia) and language deficits (aphasia) have been demonstrated with damage to the right and left temporal lobes, respectively (Peretz & Coltheart, 2003). Imaging evidence from healthy adults also shows distinct processing regions for music and language (Rogalsky, Rong, Saberi, & Hickok, 2011; Zatorre, Belin, & Penhume, 2002).

Despite these differences, there is neurological and behavioral evidence for connections between the domains. Violating either musical or linguistic expectations by presenting ungrammatical sentences or out of key chords results in similar P600 activation (Patel, Gibson, Ratner, Besson, & Holcomb, 1998). The beginnings of temporal sequences are processed similarly in music, language, and vision to identify the sequences (Hupp, Sloutsky, & Culicover, 2009). Melodies and lyrics are integrated in memory recognition, as demonstrated by verbal responses (Serafine, Crowder, & Repp, 1984; Serafine, Davidson, Crowder, & Repp, 1986, but also see Besson, Faita, Peretz, Bonnel, & Requin, 1998) and reaction time to primed words or melodies (Peretz, Radeau, & Arguin, 2004). The connections between music and language are apparent in Western tonal music: linguistic and musical accents align (Palmer & Kelly, 1992). Also, the language background of the composer correlates with the rhythmic pattern of the composer’s music (Patel & Daniele, 2003). Evidence from singers’ vocal imitations support a theory of speech and music integration (Mantell & Pfordresher, 2013). There is some evidence of connections between these domains, but further research is needed to understand possible shared processing mechanisms.

One domain-general mechanism that might explain some of the interactions between music and language is temporal processing. Both domains rely on regularly placed events in time, whether these are notes in music or syllables/stress in language. One way to test temporal processing is by perceptually priming the rate and then testing the rate of production of a novel sequence. There is evidence of within-domain rate persistence in speech. Both adults and children transfer the tempo of speech from sentences they hear to new sentences they produce as they describe pictures (Hupp & Jungers, 2009; Jungers & Hupp, 2009). Similar persistence is found in music. Pianists play new melodies faster after a fast music prime than a slow prime (Jungers, Palmer, & Speer, 2002). Because these speech and music studies show rate persistence within-domain, it is not clear whether the persistence is due to structural priming of the specific stimulus (music or speech) or a more general timing mechanism that underlies both domains. Evidence suggests there may be a shared temporal processing mechanism across a variety of domains (e.g.,
Hupp & Jungers, 2013; Hupp et al., 2009; Shintel & Nusbaum, 2007); however, this has yet to be fully explored across the language and music domains.

The current study examines rate persistence across the language and music domains. It is unique in testing whether perceptual priming in one domain (music) will lead to production differences in another domain (language). If temporal aspects of music and speech are processed within-domain, there should not be an influence across the two types of stimuli. If there is a domain-general processing mechanism, the two types of stimuli should influence each other such that a faster prime (regardless of domain) should lead to faster speech.

**Experiment 1: Language Prime**

To establish within-domain priming of speech rate, undergraduates listened to priming sentences and then gave picture descriptions. They were told to focus on their memory for these priming sentences.

**METHOD**

**Participants.** Twenty-five undergraduates (9 males, 16 females; mean age = 18.60 years, \( SD = 4.56 \)) participated to fulfill a course requirement. On average, participants had little formal music training (mean years of private instruction = 1.56, \( SD = 3.01 \)).

**Materials.** The experiment included 20 transitive sentence primes that described the action in a cartoon picture. The experiment also included 20 additional target pictures that participants described aloud and five practice trials. See Figure 1 for example stimuli. Fast and slow prime sentences were recorded by a female speaker who listened to a metronome to create the tempi. The slow sentences sounded natural, with the fast sentence primes averaging 0.40 s/syllable (2.5 syllables/s). The fast prime sentences were approximately 1.6 s long, and the slow prime sentences were approximately 2.6 s long. Half of the sentences described the picture in active form while the other half described the picture in passive form.

PowerPoint on a desktop PC was used to present the participants with the stimuli and directions necessary to complete the task. To record the participants’ speech, a TASCAM DR-03 digital recorder was connected to a head-mounted microphone. To familiarize the participants with describing the picture stimuli used to record language production, each participant produced five practice trial picture descriptions. Only the first two included a written description, leaving the participants to create their own descriptions of the last three pictures. A 16-item memory test included 8 foil sentences, which contained similar words to the original sentences, and 8 sentences from the study. The memory test sentences were either identical to the sentences from the study or varied by rate (fast/slow) or syntax (active/passive). The memory test also included an open-ended question about the purpose of the study.

**Design and procedure.** Rate was an independent, within-subject variable which was blocked across trials; participants heard fast sentences followed by slow sentences or vice versa. Rate was blocked so participants were not aware of the rate manipulation. An individual sentence was heard at a fast rate for half of the participants, but at a slow rate for the rest. The syntactic form was randomized across trials, so that there were equal numbers of active and passive sentence primes across the study. Rate of participants’ produced sentences was the dependent variable.

Participants began with five practice trials. On each experimental trial, participants listened to a prime sentence while looking at the picture being described, and then saw a second picture which they described aloud. Participants were told to pay close attention as there would be a memory test at the end of the experiment. There were 20 self-paced experimental trials, presented in two blocks of 10 trials. Between blocks, participants completed a paper and pencil questionnaire about their music and language background and demographic information. After the picture description task, participants took a 16-item memory test and indicated whether they had heard each sentence during the experiment.

**RESULTS & DISCUSSION**

Two trained research assistants determined utterance length of each trial based on both the visual depiction of the onset and offset of the sound wave and listening to the sound at a comfortable volume over headphones. The research assistants were blind to the task and the potential differences across trials. The speaking rate was calculated by dividing the number of syllables produced by utterance length to get a syllable per second rate. Each research assistant coded approximately 70% of the participants, which included 7 participants (28%) coded by both people; inter-rater reliability on trial lengths was \( r = .92 \).

Preliminary analyses on speech rate revealed no significant effects or interactions involving gender or condition (fast or slow first), \( F's < 1.40, p's > .25, \eta_p^2's < .08, \)
so these variables were eliminated from subsequent analyses. A repeated measures ANOVA determined that participants’ rate of speech was faster after fast primes (M = 4.31 syllables/second, SD = .51) than after slow primes (M = 4.09 syllables/second, SD = .59), F(1, 24) = 4.40, p < .05, η² = .16. See Figure 2 for means across each prime condition. The average rate of speech on practice trials (M = 4.13, SD = .77) was between the speech rates following slow and fast primes. Almost all produced sentences were in the active voice (95.4%). An analysis of the memory test “yes” responses, which indicated that a sentence was from the experiment, showed that foil items had fewer “yes” responses than items previously heard in the experiment, F(1, 24) = 135.84, p < .001, η² = .85. This memory result shows that participants were paying attention to the stimuli. In the open-ended question about the purpose of the study, no participants reported an awareness of the rate manipulation.

FIGURE 1. Example stimuli for Experiment 1 (Language Prime) and Experiment 2 (Music Prime). For each trial, the prime sentence/melody is in either a fast or slow rate, and then the participants produce a description of the Target picture.
Thus, speakers’ production rate was influenced by the rate of the sentences they heard, showing within-domain priming. This replicates previous work by Jungers and Hupp (2009), who demonstrated rate priming in new picture descriptions. These results support a model of language processing in which comprehension and production systems align (Pickering & Garrod, 2004). If language and music share common processing mechanisms, priming rate in one domain should lead to rate variation in the other domain. Evidence of priming from music to speech would support theories that extend beyond language and may include dynamical systems models in which listeners entrain to the acoustic environment (Large & Jones, 1999).

Experiment 2: Music Prime

The goal of this experiment was to test the effect of music rate priming on speech production. If language and music are processed separately, the rate of one domain should not influence the other. If language and music are processed using a common mechanism, it is more likely that perception in one domain will influence production in the other domain.

METHOD

Participants. Thirty-one undergraduate students (8 males, 23 females; mean age = 19.03 years, SD = 3.58) participated to fulfill a course requirement and did not participate in Experiment 1. On average, participants had little formal music training (mean years of private instruction = 0.65, SD = 1.36).

Materials. The prime stimuli were 20 short melodies presented at fast or slow rates. The prime songs were initial phrases from familiar melodies digitally produced using a piano timbre with Finale software (e.g., It’s Raining, It’s Pouring or Three Blind Mice). The stimuli were pretested with independent listeners to be similarly familiar. The slow music primes were produced at a tempo of 60 bpm (8 s long) and the fast music primes at 120 bpm (4 s long). Half of the melodies included staccato notes (separated) and half included legato notes (connected).

The use of PowerPoint and the voice recording device were the same as in Experiment 1. The practice trials and pictures were also the same as in Experiment 1. The questionnaire with music, language, and demographic information was the same as in Experiment 1. A 16-item memory test included 8 foil melodies and 8 melodies from the study and included a question about the purpose of the study.

Design and procedure. Rate was an independent, within-subject variable that was blocked across trials; participants heard fast melodies followed by slow melodies or vice versa. Rate was blocked so participants were not aware of the rate manipulation. An individual melody was heard at a fast rate for half of the participants, but at a slow rate for the rest. The note connectedness (staccato/legato) was randomized across trials. Rate of participants’ produced sentences was the dependent variable.

Participants began with five practice trials to familiarize them with producing picture descriptions. For each experimental trial, participants listened to a prime melody while looking at the name of the song, and then saw a picture that they described aloud. Participants were told to pay close attention as there would be a memory test at the end of the experiment. There were 20 experimental trials, presented in two blocks of 10 trials. Between blocks, participants completed a paper and pencil questionnaire about their music and language background and demographic information. After the picture description task, participants took a 16-item memory test over the melodies where they indicated whether or not they had heard the melody in the experiment. The final question was about the purpose of the study.

RESULTS & DISCUSSION

The same two research assistants from Experiment 1 determined utterance length of each trial. They were blind to the task and potential differences across trials. Each research assistant coded approximately 60% of the participants, which included six participants (19%) coded by both people; inter-rater reliability on trial
lengths was \( r = .96 \). Preliminary analyses on speech rate revealed no significant effects or interactions involving gender or condition (fast or slow first), \( F_{1,30} = 2.97, p's > .09, \eta^2_{p} = .12 \), so these variables were eliminated from subsequent analyses. A percentage of "yes" responses was calculated for each memory trial type. Overall, a repeated measures ANOVA determined that participants' rate of speech was faster after fast music primes (\( M = 4.35 \) syllables/s, \( SD = .85 \)) than after slow music primes (\( M = 4.13 \) syllables/s, \( SD = .81 \)), \( F_{1,30} = 9.22, p < .001, \eta^2_{p} = .24 \). The average rate of speech on practice trials (\( M = 4.13, SD = .81 \)) was between the speech rates following slow and fast music primes. Thus, speakers' production rate was influenced by the rate of the melodies they hear, showing cross-domain priming.

**Discussion**

Speakers' sentence production rates were influenced by the rate of the priming auditory stimuli, both for sentences and melodies. Speakers produced faster sentences after faster primes than after slower primes within-domain (language-language) and cross-domain (music-language). The unique contribution of this study is the cross-domain influence from perception to production. Priming from perception to production within the domain of language is not surprising; the persistence of rate and syntax replicates previous literature and can be explained by shared, domain-specific mechanisms (Bock, 1986; Jungers & Hupp, 2009). However, rate persistence between music perception and language production indicates there must be shared processing mechanisms across domains.

The idea of a shared processing mechanism for music and speech runs counter to previous work suggesting modular representation for speech (Fodor, 1983) and music (Peretz & Coltheart, 2003). Neuropsychological and case study evidence supports specific brain regions dedicated to music and speech (Cuddy et al., 2012; Zatorre et al., 2002), although this modularity may emerge developmentally (McMullen & Saffran, 2004). Peretz and Coltheart (2003) propose a modular model of music processing that includes processing components for different aspects of music. Their model is supported by evidence of modularity for pitch components, but the modularity of temporal components is debated. Patel (2003) proposes a model with shared syntactic processing for music and speech, but separable representations for domain-specific knowledge. The current study, with the transfer of timing from music perception to language production, argues for shared processing mechanisms for temporal aspects of music and speech as has been found with speech and a variety of other domains suggesting a domain-general temporal processing mechanism (Hupp & Junger, 2013; Hupp et al., 2009; Shintel & Nusbaum, 2007).

What is the nature of this shared processing mechanism for music and speech? Perhaps this result can be connected to the more general notion of behavioral coordination. People coordinate their actions with external stimuli, themselves, and each other (Turvey, 1990). Pickering and Garrod (2004) argue that alignment between conversation partners is due to a shared conceptual representation. A similar shared representation of sequence structure is proposed to guide the connection between perception and action in music (Pfordresher, 2006). Mirror neurons, which fire during an action or the perception of that action, offer a possible explanation for within-domain priming results (Rizzolatti, Fogassi, & Gallese, 2001). However, mirror neurons and shared cognitive representations do not explain the cross-modal results, since music and speech are representationally dissimilar at multiple levels (i.e., syntax, semantics).

Coordination between music and speech could be conceptualized more globally using a dynamical systems approach, which posits a coordinative structure that is self-organized and temporary (Shockley, Baker, Richardson, & Fowler, 2007). Coordination of gaze and sway between conversation partners can be accounted for by this joint action system (Shockley et al., 2007). Such a low-level coordinative structure does not presuppose a shared cognitive representation, but it could give rise to a higher-level representation.

Another perspective that might explain the priming from music to language is an oscillator model. Listeners' internal oscillators synchronize to an external rhythm in a dynamical systems theory (Jones & Boltz, 1989; Large & Jones, 1999; Large & Palmer, 2002). This model could account for the difference in effect size between the music and speech results because it posits greater entrainment with more regular timing. In the current study, music stimuli were created using Finale software so each event was precisely timed compared to the speaker-produced sentences. Also, the sentences were 1.6-2.6 s long, while the melodies were 4.0-8.0 s long. An oscillator model would predict greater entrainment for the longer, more precise music stimuli than the
speech stimuli, which is supported by the current findings. The extension of oscillator models from perception to production is an area open for further exploration.

Shared neural resources for music and language may also explain the cross-modal priming result. Several studies propose specific brain regions involved in auditory-oral motor integration. For example, Hickok, Buchsbaum, Humphries, and Muftuler (2003) used fMRI to test the perception and production of music and speech. The bilateral prefrontal cortex, bilateral inferior frontal gyrus, and an area in the left posterior part of the Sylvian fissure (area Spt) were active during all tasks and were identified as part of the audio-motor network (Hickok et al., 2003). This replicates previous work with speech perception and production implicating the area Spt as part of the audio-motor network (Buchsbaum, Hickok, & Humphries, 2001). Several areas are implicated in the perception and production of auditory timing for speech and music, including the cerebellum (Chen, Pehhume, & Zatorre, 2008; Janata & Grafton, 2003; Kotz & Schwartz, 2010), basal ganglia (Ferrandez, et al., 2003; Grahn & Brett, 2007; Kotz & Schwartz, 2010), premotor cortex, and supplementary motor area (Chen et al. 2008; Geiser, Zaehele, Jancke, & Meyer, 2008; Grahn & Brett, 2007; Macar, Anton, Bonnet, & Vidal, 2004). Also, the same neural regions are involved in timing perception and movement coordination (Schubotz, Friederici, & Cramon, 2000). These studies as a whole point to a shared neurological network linking auditory perception and production. Neurological and dynamical systems explanations for cross-domain priming are not mutually exclusive and instead may represent two levels of explanation for the observed phenomena.

Cross-domain transfer of primed rate demonstrates evidence for a common temporal processing mechanism for music and language as has been demonstrated across a variety of domains. Future work should include other rhythmic, but nonauditory sequences to further test domain-general temporal processing. Although there are differences between music and language, studying the areas where the domains converge, such as temporal processing, offers unique opportunities to better understand human cognition.

**Author Note**

Portions of this research were presented at the Midwestern Psychological Association conference in Chicago (May, 2012). We wish to thank the participants and the research assistants who made this research possible: Alexander Chien and Jennifer Watkins.

**Correspondence concerning this article should be addressed to** Melissa K. Jungers, Department of Psychology, The Ohio State University at Newark. 1179 University Drive, Newark, OH 43055. E-mail: jungers.2@osu.edu

---

**References**


