

## **Time after time: The coordinating influence of tempo in music and speech**

MELISSA K. JUNGERS    CAROLINE PALMER    SHARI R. SPEER  
Ohio State University

**Abstract** - Music is one of the most complex and popular of human behaviors. People enjoy listening to music and respond to it with clapping, foot tapping, humming, and other natural behaviors. We consider the role of timing in how performers coordinate with other performers. Like speech, music is a joint activity; timing must be relatively consistent for musicians to coordinate their individual actions with those of others. We address the role of production rate or tempo as a coordinating device among producers and listeners. We discuss evidence that musicians' choices of tempi are influenced by contextual information, specifically, by tempi of previously heard sequences. We also describe a persistence of production rates and global phrase patterns in speakers' sentence productions. In addition to contextual influences, producers' preferred rates influence their choice of production rates in both music and speech; preferred rate affected speech more than primed rate, whereas primed rate affected music more than preferred rate. Tempo persistence is considered as a coordinating device for communication domains.

**Key words:** music, speech, rate, production.

Music is one of the most complex and popular of human behaviors. Without formal training, people enjoy listening to music and respond to it with foot tapping, humming, clapping, and other natural responses. Music performance is moreover a multifaceted cognitive and motor skill. Performance flows at a rate of about 2-10 musical tones per second, and musicians effortlessly synchronize their productions with those of others (Palmer, 1997). Most researchers propose that music is a form of communication among people, whether the message is structural, such as a particular melody or phrase, or emotional, such as a particular mood. We consider here the factors with which music performers and listeners communicate and how similar or different those factors are to other forms of human communication, in particular, speech.

The focus of this paper is on the role of timing in how performers coordinate with other performers. Like speech, music is a joint activity; it requires people to coordinate their individual actions in order to succeed. For communication to work, producers must attend closely to the timing of their own productions and

Address for correspondence: Melissa Jungers: Jungers.2@osu.edu - Caroline Palmer: Palmer.1@osu.edu  
Department of Psychology, Ohio State University, 142 Townshend Hall, 1885 Neil Avenue - Columbus,  
Ohio 43210 - phone: (614) 292-7718 - fax: (614) 292-5601



those of their recipients. Most musicians perform in groups, in which each musician tailors his production to the productions of other musicians. Speakers also perform in groups, and speakers often tailor the timing of their utterances in response to utterances from their audience (Clark, 1996). Both speakers and musicians synchronize their timing in turn-taking (such as in jazz styles); in addition, musicians need to fine-tune their timing to produce simultaneities within about 80 ms of the performances of others. Timing is central both to spontaneous forms of music, such as improvisation, and stylized or planned forms typical of Western concert music. Thus, time is particularly important for coordination among producers and recipients in music and speech.

We address the role of production rate or tempo as a coordinating device among producers and listeners. One of the fundamental principles of many forms of music is an underlying beat or pulse that provides a temporal framework by signaling the pace or rate of music. In “measured music”, the musical beat is an evenly paced timekeeper that aids synchronization among performers. It allows listeners to clap along, continuing even in the absence of performed events. Humans’ capacity to entrain to the underlying beat in an auditory sequence extends across a wide range of tempi (Fraisse, 1982; Merker, 2000). The functional utility of the beat is that it allows us to predict when future events will occur and thus synchronize listening with acoustic stimuli such as music performances (Jones, 1976, 1987; Large and Jones, 1999). For the beat to be useful, the tempo or overall rate of production must remain fairly consistent throughout a performance.

### **Tempo effects in music perception and memory**

Listeners display a relatively fine level of tempo discrimination for music. The relative JND (just noticeable difference) for tempo discrimination is on the order of 5-8% for nonmusician listeners (Drake and Botte, 1993; Ellis, 1991). Drake and Botte (1993) found that temporal thresholds depended on the number of intervals in the sequence; for single intervals (2 tones) the JND was around 6% and gradually decreased to 3% as the sequence length increased to 6 intervals. The optimal sensitivity was observed for interonset intervals in the range of 300-800 ms. In addition, preferences exist for some tempi over others. There is evidence for age-related tempo preferences (LeBlanc, Colman, McCrary, Sherrill, and Malin, 1988); younger listeners prefer faster tempi (Drake, Jones, and Baruch, 2000). Listener preferences have also been linked to internal tempo and preferred rate of activity, such as preferred tapping rate, walking rate, or speaking rate (Fraisse, 1982). People tend to judge melodies as occurring at a faster rate when they have fewer changes in informational content (Boltz, 1998). Melodies that contained more changes in pitch direction, pitch distances, and rhythmic accent structures were judged to be slower than those with fewer changes.

Listeners are also sensitive to subtle acoustic differences between music performances and they can retain these differences in memory for particular performances. Palmer, Jungers, and Jusczyk (2001) explored the role of memory for acoustic details in music performances. Musically trained listeners were familiar-

ized with one of two performances of the same short musical excerpt. The performances differed in articulation (the connectedness between notes, defined by offset and onset asynchronies), intensity (how loud the notes sound), and interonset interval (note duration) cues. At test, the listeners were presented with the original performances from familiarization as well as different performances of the same melodies (the same notated pitches and durations), but different intensities, articulations, and interonset intervals. Listeners were asked to identify which of the performances were present at familiarization. Listeners could recognize the performances of the melodies they had heard during familiarization, even though the categorical pitches and durations in the two versions were identical. Furthermore, non-musician listeners recognized the particular performances of melodies heard at familiarization as accurately as musically trained listeners, indicating that musical training is not necessary for memory for fine acoustic musical features (Palmer *et al.*, 2001).

Both the musically trained and untrained listeners in Palmer *et al.* (2001) had many years of exposure to music. To address whether musical acculturation is necessary for memory for musical features, Palmer *et al.* (2001) also tested 10-month-old infants' memory for performances with the same melodies, using a head-turn preference procedure (Kemler Nelson *et al.*, 1995). After being familiarized with one performance of each melody, infants oriented longer to the familiar performances during test than to other performances of the same melodies. Thus, even infants (with little music acculturation) can use acoustic cues that differentiate performances to form a memory for short melodies (Palmer *et al.*, 2001).

Although this study indicated that people are sensitive to subtle performance differences and can retain them in memory, it does not indicate which cues are most salient in perception and memory. In another study, musician listeners were tested for their ability to discriminate and remember music performances that differed in only one or two acoustic cues (Jungers and Palmer, 2000). In one experiment, musically trained listeners discriminated pairs of performances that differed in intensity, articulation, articulation with intensity, or interonset interval, while other variables remained constant. When articulation or articulation with intensity cues were present, listeners could accurately distinguish same from different pairs of performances of the same melody. In another experiment, musician listeners were familiarized with performances that varied in articulation, intensity, or articulation with intensity cues and later heard these performances as well as novel performances of the same melody. Listeners could more accurately identify performances they had heard before and were most accurate at identifying those performances that varied in articulation cues (Jungers and Palmer, 2000). Thus, listeners were particularly sensitive to the articulation cues in music performances; listeners discriminated musical sequences based on the timing between pitch events within the sequence.

Both musician and non-musician listeners can remember particular performance tempi over prolonged time periods. Musicians can reproduce performances of long musical pieces, such as an entire movement of a symphony, at the same tempo with very low variability (Clynes and Walker, 1986; Collier and Collier,

1994). Similarly, nonmusicians can reproduce popular songs from memory at tempi very close to the original tempo (Levitin and Cook, 1996). Furthermore, when people sang familiar songs as fast or as slow as possible, songs that lacked a tempo standard in original recordings were produced with a larger variability in tempo; this counters arguments that memory for the tempo of remembered songs was solely a function of articulatory constraints.

### **Rate effects in speech perception and memory**

Just as subtle acoustic differences among music performances can be recognized and remembered, similar acoustic differences in spoken sentences are recognized and represented during language understanding. In speech these acoustic differences and their perceptual consequences are referred to as “prosody”. Speech prosody has a wide variety of definitions that range from a structure that organizes sound to a phonological system that employs suprasegmental features such as pitch, timing, and loudness (for differing views, see Cutler, Dahan, and van Donselaar, 1997; Price, Ostendorf, Shattuck-Hufnagel, and Fong, 1991; Warren, 1999). Prosody refers to the “perceived stress, rhythm, and intonation in spoken sentences” (Kjelgaard and Speer, 1999). Prosody is important to the discussion at hand because prosody includes aspects of timing in speech.

Although tempo has not been the focus of much research in speech, many studies indicate that prosodic features in general influence listeners’ interpretation of sentence meaning. Word durations can disambiguate the meaning of ambiguous sentences (Lehiste, 1973; Lehiste, Olive, and Streeter, 1976). Listeners heard different versions of syntactically ambiguous sentences and were able to determine the intended meaning. Analysis of acoustic properties from the sentences suggested that timing and intonation were useful features for disambiguation (Lehiste, 1973). The placement and duration of pauses provides another perceptual cue to sentence meaning; speakers’ pause patterns tend to correlate with the syntactic structure of a sentence, with longer pauses near important structural boundaries (Lehiste, Olive, and Streeter, 1976). Speer, Crowder and Thomas (1993) presented listeners with sentences that contained different prosodic realizations of a single word in a syntactically ambiguous sentence, such as the sentences “They are FRYING chickens” and “They are frying CHICKENS”. Listeners’ paraphrasings of the sentences showed that the interpretation depended on the prosodic emphasis. In another experiment, listeners were familiarized with sentences and were later asked to recognize the sentences from familiarization, presented along with unfamiliar sentences. Recognition was higher for the sentences that retained the same prosody at study and test than for sentences that were syntactically identical, but had different prosodic cues at study and test. Furthermore, prosodic structure aided recognition even for nonsense sentences. Thus, prosody aided listeners’ memory for and differentiation of ambiguous sentences (Speer *et al.*, 1993).

The rate at which speech is produced influences its perception at the most basic of levels. Speech rate varies considerably during normal conversation (Miller, Grosjean, and Lomanto, 1984) and can substantially alter the acoustic information

that allows the discrimination of one speech sound from another. For example, syllable-initial voiced and voiceless stop consonants, such as the pair /b/ and /p/, are distinguished on the basis of voice-onset-time, the time at which vocal fold vibration starts relative to the release of the stop closure. This time is longer for voiceless stops like /p/ than for voiced stops like /b/. The rate at which syllables and sentences are spoken influences the production of these segments, so that as a talker speaks more slowly, and the duration of syllables and words increases, the voice-onset-time also increases. Listeners adjust to these differences in timing, so that the same acoustic signal may be perceived in a fast speech context as /b/, but in a slow speech context as /p/ (Volatis and Miller, 1992; Wayland, Miller, and Volatis, 1994).

Several studies suggest that temporal features of speech are incorporated in memory for language. Listeners can use extralinguistic information, including talker identity and talker's rate, to accurately identify previously presented words (Bradlow, Nygaard, and Pisoni, 1999). The rate of presentation affected listeners' abilities to recall items produced by different speakers. Listeners showed better recall for those items presented at the same rate in both familiarization and test than for items presented at different rates from familiarization to test (Nygaard, Sommers, and Pisoni, 1995). These findings suggest that speakers' rates influence memory for speech contents.

Temporal aspects of speech are important for communication not only in individual sentences, but also across conversations. When two people carry on a conversation, they must take turns speaking. They focus on the timing of their partner's utterances as well as their own (Clark, 2002). This turn-taking is often precisely timed so that one speaker begins at close to the same time that the other speaker has finished (Fox Tree, 2000). Pauses that interrupt the timing of turn-taking are not simply mistakes; they can carry information about the knowledge of the speaker (Fox Tree, 2000). To prevent misunderstandings, speakers use words such as "um" and "uh" to punctuate long silences and to hold their place in the conversation (Clark, 1996; Fox Tree, 2000; Levelt, 1989). In sum, although multiple acoustic dimensions are important in speech, timing is especially important for communication.

### **Persistence in music and language**

A few studies suggest that the tempo of music performances persists across sequences. Cathcart and Dawson (1928) instructed pianists to perform one melody at a particular tempo and another melody at a faster or slower tempo. When pianists attempted to perform the first melody again at the original tempo, their tempo drifted in the direction of the second melody. More recently, Warren (1985) reviewed studies of tasks that varied from color judgments to lifting weights. Each domain displayed a perceptual homeostasis, which Warren (1985) termed the "criterion shift rule": that the criterion for perceptual judgments shifts in the direction of stimuli to which a person has been exposed. Warren (1985) suggested that a criterion shift serves to calibrate perceptual systems so that behavior will be appropriate for environmental conditions.

As Warren (1985) noted, music is not the only domain in which persistence effects are found. Persistence also plays a role in speech; one element of speech that persists is syntactic form. When listeners were asked to repeat a sentence they had heard and then produce a description of a picture, they tended to use the same syntactic form as in the former sentence to describe the scene (Bock, 1986). For example, when subjects heard and repeated the sentence, “The referee was punched by one of the fans”, they were more likely to describe a picture with a church and a lightning bolt as “The church is being struck by lightning”, with both sentences in the passive form (Bock, 1986). Another aspect of speech that may persist is the rate. Kosslyn and Matt (1977) played a recording of two male speakers for listeners: one speaking at a fast rate and one at a slow rate. Then the subjects read a passage they were told was written by one of the speakers. The subjects imitated the rate of the speaker who supposedly wrote the passage, although they were not explicitly instructed to do so (Kosslyn and Matt, 1977). In that study, it is possible that subjects may have associated each written passage with a particular speaker and felt an expectation to reproduce the rate of that speaker.

### **Tempo persistence in music performance**

We next describe an experiment that examined whether pianists imitated the tempo of a short melody when they produced a subsequent melody. Pianists listened to melodies and then performed melodies. The pianists were instructed to pay attention to both the heard and performed melodies for a later memory test. If the pianists produced performances that were similar in rate to previously heard melodies, that would indicate that the tempo was remembered and incorporated into subsequent productions, even with no instructions to perform at the same tempo. If the pianists performed at a constant tempo throughout the experiment despite the melody tempi they had just heard, it would indicate the influence of a preferred performance rate.

Sixteen experienced adult pianists with an average of 8 years’ piano instruction (range: 6-13 years) performed 10 single-voiced melodies in the experiment. Two examples of the melodies are shown in Figure 1. Computer-generated versions of these melodies were created at two rates, fast (300 ms per eighth-note beat) and slow (600 ms per eighth-note beat), in legato style (with 0 ms between tone offsets and following onsets). These computer-generated performances, called prime melodies, were presented over headphones in blocks of slow or fast rates, with 5 melodies in each block. The melodies to be performed, called target melodies, were presented in musical notation. These melodies did not include bar lines or time signatures, so that there would be no indication of meter or any indirect indication of rate. The prime and target melodies were different, but each pair of prime and target melodies was matched for meter and length; both melodies in a prime/target pair were in a major or a minor key.

The pianists sight-read two melodies in order to establish their “preferred” performing rate. Pianists then alternated listening to and performing melodies. While listening to melodies, the pianists had a blank sheet before them; they were

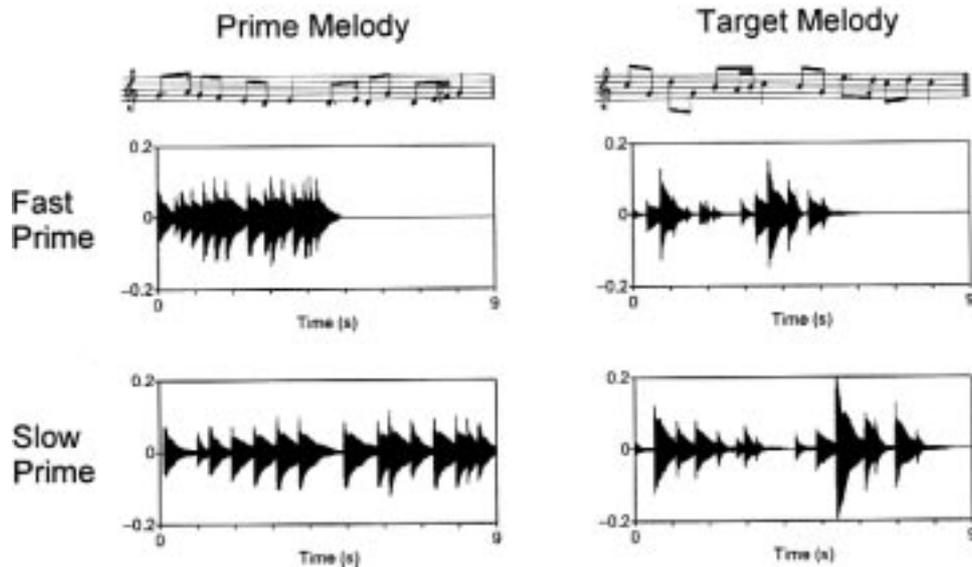


Fig. 1. Wave forms for sample prime and target melodies in the fast prime and slow prime conditions.

provided with musical notation for the performed melodies. In order to ensure that any differences between performances following fast primes or slow primes were not specific to the set of target pieces, the original prime melodies and target melodies were switched for half of the pianists. Thus, the prime melodies for the first group of pianists served as the target melodies for the second group. The pianists heard prime and target performances over headphones and performed on an electronic keyboard which recorded to computer. The duration of each performance was measured by the time of the initial event onset to the final event onset.

Pianists' total durations of performed melodies were significantly longer for those following a slow prime than a fast prime. A one-way analysis of variance (ANOVA) on the performance rates following fast or slow primes yielded a significant effect of prime tempo,  $F(1,14) = 68.5$ ,  $p < .01$ . This effect was found for all pianists. Figure 1 shows an example of a slow prime and a fast prime, with the pianists' performances of target melodies that followed the primes. The mean preferred melody duration was 6.1 seconds, falling between the mean of target durations that followed the slow prime (= 6.8 sec) and those that followed the fast prime (= 5.3 sec). Thus, the rate of music that pianists had just heard influenced the tempo at which they performed.

It is possible that the duration differences between the targets following the fast and slow primes were due to a particular event, such as a break between musical phrases. To assess this possibility, the melody durations were analyzed by quarter-note beat, a common measure of tempo (Gabrielsson, 1987; Palmer, 1989; Repp, 1994). Although the melodies contained different rhythms, all of the

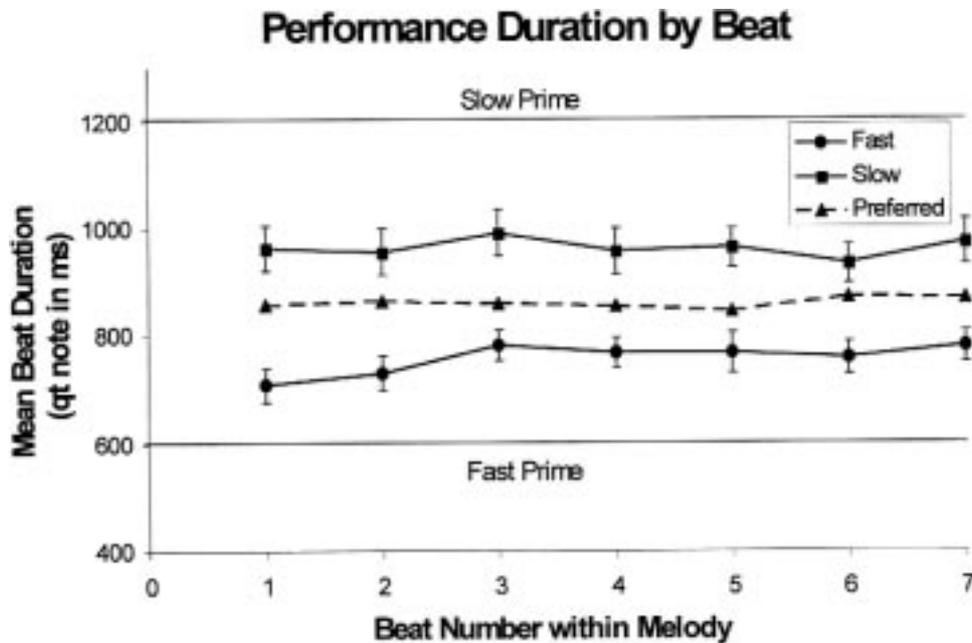


Fig. 2. Mean target beat duration in musical sequences by prime condition. Solid lines indicate prime durations; bold lines indicate mean target durations; dashed line indicates mean preferred beat durations.

melodies were composed of 8 quarter-note beats. Figure 2 illustrates the mean beat durations (measured by interbeat intervals) of the target performances following fast primes and slow primes, as well as the mean preferred beat durations. There was a significant change in target tempo across beats in the sequence,  $F(6, 90) = 4.6, p < .01$ , with the first two beats played fastest following the fast prime. However, there was also a main effect of prime,  $F(1, 15) = 67.6, p < .01$ , and no significant interaction of prime and beat. These findings indicate that pianists persisted in the tempo of the prime melodies across their entire performances. Although there was a clear difference between performances following fast and slow prime melodies, the pianists did not perfectly imitate the tempo of the melodies. Thus, the tempo at which pianists performed depended on other melodies they heard as well as, to a lesser extent, their preferred tempo.

### Rate persistence in speech production

The music experiment indicated that pianists persisted in the tempo of performances they heard, even in the absence of explicit rate instructions. The pianists had studied music for many years, and musicians often practice with a metronome in order to keep a consistent tempo. Does rate persistence occur for other forms of communication, such as speaking? Although people do not prac-

tice speaking with a metronome, conversational speech is learned early and practiced often. Speakers show persistence in the syntax of sentences they hear (Bock, 1986) and listeners encode the rate of a speaker in memory (Nygaard, Sommers, and Pisoni, 1995). Does the rate of sentences listeners hear influence the rate at which they will produce subsequent sentences?

We next describe an experiment that examined whether native English speakers imitated the rate of a previously heard sentence when they produced a sentence of analogous structure. The procedure for the speech experiment was designed to be as similar as possible to the music experiment in order to allow comparisons. As in the music experiment, the speakers were instructed to pay attention to both the heard and spoken sentences for a later memory test. If the speakers produced sentences similar in rate to the recently heard sentences, that would indicate that speakers incorporated the timing aspects from their memory of previous utterances into their own productions. If the speakers spoke at a constant rate throughout the experiment, despite the range of rates they had just heard, it would indicate the influence of a preferred speaking rate.

Sixty-four adult native English speakers produced 10 short sentences (6 to 7 words each) in the experiment. Two examples of the sentences are shown in Figure 3. The prime sentences were pronounced by a female speaker, who produced each sentence after hearing metronome clicks at a fast (375 ms per accent or 160 bpm) or slow (750 ms per accent or 80 bpm) tempo. No instructions were given to the speaker regarding intonation pattern. The timing of these prime sentences was less consistent than the timing of the prime melodies, but the advantage was the relatively natural sound of the sentences at the two rates. The prime and target sentences were matched for number of syllables, lexical stress pattern, and syntactic structure.

The speakers were seated in front of a computer screen and their productions were recorded using a head-mounted microphone. First, speakers read two sentences aloud from the computer screen as a measure of their preferred speaking rate. Next, speakers alternated listening to and reading sentences. The prime sentences were blocked by rate. As in the music experiment, the 10 original prime sentences and 10 target sentences were switched for half of the speakers. Thus, the prime sentences for the first group of speakers served as the target sentences for the second group. Speakers were instructed that they were to remember all of the sentences for a later memory test. The duration of their productions was measured by the time of the initial syllable onset to the final syllable offset.

The speakers' target sentence durations were significantly longer following a slow prime than a fast prime. A one-way analysis of variance (ANOVA) on the speaker's rates following fast or slow primes yielded a significant effect of prime tempo,  $F(1,63) = 11.7, p < .01$ . This effect was seen for 43 of 64 speakers. Figure 3 shows an example of a sentence that served as a slow and a fast prime and its corresponding waveforms. Two speakers' productions of the target sentence that followed those primes are also shown. There was a significant difference between target sentence durations in the fast and slow prime conditions, but subjects also were influenced by their own preferred production rate. Speakers' mean preferred

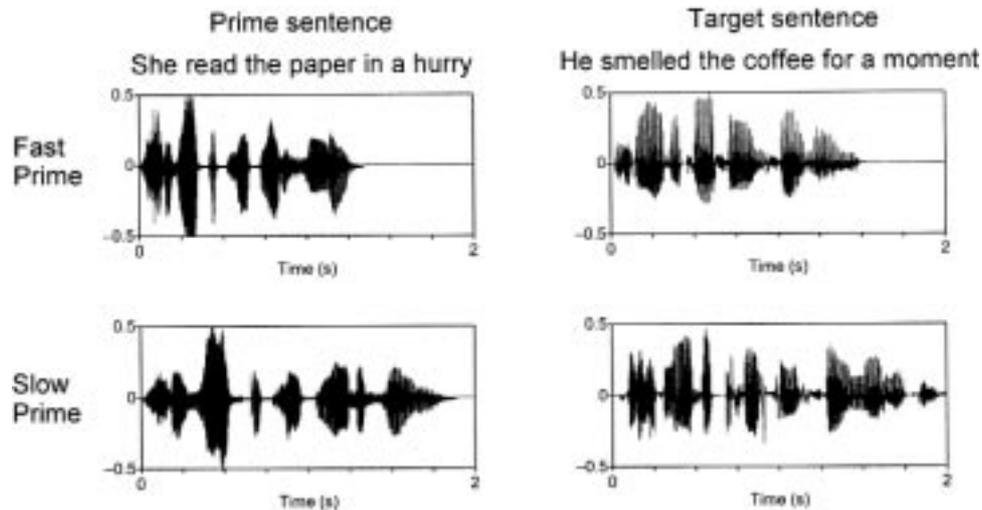


Fig. 3. Wave forms for sample prime and target sentences in the fast prime and slow prime conditions.

sentence durations averaged 1.8 seconds, falling between their mean target durations following the slow primes (= 1.72 sec) and the fast primes (= 1.81 sec). Thus, speakers were influenced by both their preferred rate and by the rate of the prime sentences they had just heard.

To investigate whether speakers persisted in more than overall tempo, an analysis of intonational and phrase break patterns was conducted, using the English ToBI (Tone and Break Indices) method to transcribe the prime and target utterances (Beckman and Elam, 1997). In the ToBI system, each utterance is assumed to be composed of at least one Intonational Phrase (IP), indicated by a phrase-final high or low tone (H% or L%) and given a break index of 4. Each intonational phrase is in turn composed of at least one intermediate phrase (ip), indicated by a high, downstepped, or low phrase accent (H-, !H-, or L-), and given a break index of 3. Each intermediate phrase contains at least one pitch accent, indicating sentence-level emphasis on the word. Pitch accents may be high, downstepped, low, or bitonal (e.g. H\*, !H\*, L\*, or L+H\*). Break indices of 2 and below indicate word-level boundaries (2 is precise speech, 1 is normal word boundary, and 0 is coarticulated boundary). Figure 4 shows the transcriptions for a prime and target example. On this trial, the speaker produced the target sentence with the same intonational pattern as the prime sentence.

If the participants persisted in the phrasing of the primes, they should tend to produce the target sentences with the same pattern of phrase breaks as they heard in the prime sentences. Utterance transcriptions were grouped into three possible patterns of global phrasing across the sentences: the biggest break following the verb (V), the biggest break following the noun (N), or equal phrase breaks after

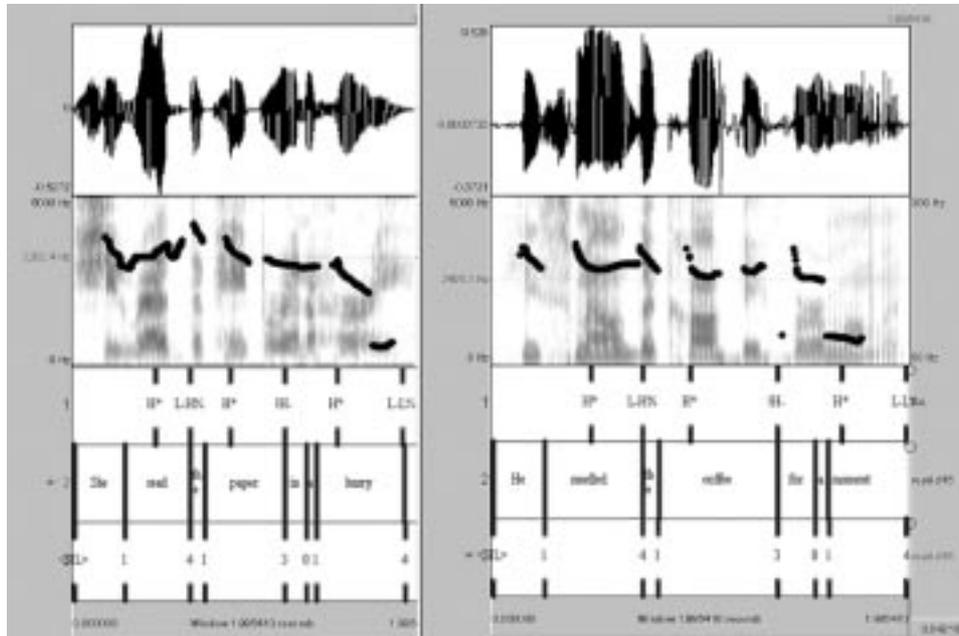


Fig. 4. ToBI analysis of prime and immediately following target speech utterances in sample trial (see text for explanation).

both the verb and noun (=VN). Transcriptions were grouped as follows (for further details of this type of grouping, see Schafer, Speer, Warren, and White, 2000): biggest break following Noun included VN] IP, V] ip N] IP, and VN] ip; biggest break following Verb included V] IP N, V] ip N, and V] IP N] ip. Equal breaks following Verb and Noun (=VN) included VN, V] ip N] ip, and V] IP N] IP. Table 1 shows the number of productions that fell in each of these 3 global phrasing patterns for 32 participants (16 represented each assignment of sentences to primes or targets), based on the global phrasing pattern heard in the immediately preceding prime. Nine productions that had speech errors were removed from this analysis. There was a significant interaction between the phrasing of the prime sentences and the phrasing of the target sentences (*chi-squared* (4) = 48.9,  $p < .01$ ). Thus, speakers most often persisted in producing the biggest break in the target utterance at the same global position as the biggest break in the prime sentence.

The priming effects on speakers' target sentences were smaller than priming effects on musicians' target productions, but effects of preferred rates were larger in speech than in music. To compare the relative roles of prime tempo and preferred tempo, a linear regression model was applied to both the speech and music experiments, predicting each producer's target durations from their preferred sequence durations and from the prime durations. The musicians' performances

Table 1. *Major Phrase Break Locations in Primes*

Phrase Break Locations In Targets Following Primes		
Verb (2)	Noun (15)	=VN (23)
Verb 4 (25%)	2 (2%)	12 (7%)
Noun 7 (44%)	85 (71%)	62 (35%)
=VN 5 (31%)	32 (27%)	102 (58%)

indicated a significant fit of the linear regression model ( $R = .78, p < .01$ ), with significant contributions of both the primed durations (standardized coefficient = .61,  $p < .01$ ) and the preferred durations (standardized coefficient = .49,  $p < .01$ ); the contributions of the primed durations were larger. The speakers' performances also indicated a significant fit of the linear regression model ( $R = .72, p < .01$ ), with significant contributions of both the primed durations (standardized coefficient = .22,  $p < .01$ ) and the preferred durations (standardized coefficient = .69,  $p < .01$ ); this time, the contributions of the preferred durations were larger. Overall, the music and the speech experiments demonstrated that preferred rate and prime rate both influence produced rate, but the importance of these two factors differed in the two domains.

## Conclusions

The studies described here indicate that rate or tempo persistence, in addition to other temporal aspects of sequence structure, contributes to music and speech production. Music-theoretic depictions often attribute tempo to other aspects of musical structure, including rhythm, melody, harmony, texture, and dynamics (Cooper and Meyer, 1960; Stein, 1989). Rate effects in speech have likewise been conceptualized as arising from particular structural relations such as contrastive emphasis and de-accenting, focus, and dominance relations among syntactic constituents. Yet the tempo chosen by musicians and speakers for each sequence persisted from the previous sequence, both when the preceding sequence was fast and slow in tempo. Thus, tempo persistence contributed above and beyond sequence structure in both music and speech. What is the source of this persistence? Producers may have implicitly learned the prime rate, given no explicit instructions. Reproducing the rate of previously heard productions may be easier than producing a new (different) rate. Implicit learning accounts have been proposed for syntactic persistence effects in speech (Bock, 1992; Bock and Griffin, 2000), based on experimental paradigms similar to those reported here. One important difference in the current studies is

that producers did not repeat the primes; thus, persistence of production rate was based solely on perceptual priming, not on production priming (see also Bock 2002).

Persistence of production rate may aid coordination among performers and listeners by providing temporal regularity and increasing the predictability of when future events will occur. Stressed or accented musical events often display a tendency toward equal spacing in time (or isochrony) that implies a regular beat or underlying period by which upcoming events can be measured or predicted. Musical patterns that are regular in their underlying beat are more readily perceived and remembered by listeners than irregular patterns (Essens and Povel, 1985; Povel, 1981). Although languages vary in their rhythmic organization, most are thought to have a basic level of prosodic organization that displays some tendency toward regularity (Cooper, Whalen, and Fowler, 1986; Munhall, Fowler, Hawkins, and Saltzman, 1992). Despite the fact that expressive utterances and music performances do not display temporal regularity, listeners tend to hear a regularity in stress patterns (Cooper and Eady, 1986; Lehiste, 1977; Martin, 1970). This regularity may increase the predictability of when future events will occur, a feature that has been incorporated in rhythmic theories of attending (Jones, 1976; Large and Jones, 1999).

Why might musicians persist more than speakers in their tempo? One reason may be the need to synchronize the performances of a large number of musicians in a group or ensemble. Musicians in bands and orchestras are taught to watch the conductor's baton and to subdivide rhythmically difficult passages in order for the group to synchronize. In contrast, even in group speaking situations such as a classroom, coordination among speakers is usually limited to one or two individuals, reducing demands of rate-matching. Musical compositions often have a prescribed tempo that varies widely, from *largo* or very slow to *prestissimo* or very fast. Conversational speech in contrast does not have a prescribed rate; speakers may be more constrained in choice of rate by feedback about the intelligibility of their utterances. Despite these differences in the constraints on speakers and performers, timing plays an integral role for communication in both domains.

#### *Acknowledgements*

Melissa K. Jungers and Caroline Palmer, Department of Psychology, 1885 Neil Ave., Ohio State University, Columbus OH 43210; and Shari R. Speer, Department of Linguistics, 1712 Neil Ave., Ohio State University, Columbus OH 43210. This research was supported by a Center for Cognitive Science Summer Fellowship to the first author, by NIMH Grant R01-45764 to the second author, and by NSF Grant SES-0088175 to the third author. We thank Grant Baldwin, Laurie Maynell, Beth Mechlin, and Annalisa Ventola for assistance. Correspondence can be sent to Melissa Jungers at jungers.2@osu.edu or to Caroline Palmer at palmer.1@osu.edu.

## References

- Beckman, M.E., and Elam, G.A. (1997). *Guidelines for ToBI labeling. (Version 3)*. Columbus, OH: Ohio State University.
- Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355-387.
- Bock, K. (2002). Persistent structural priming from language comprehension to language production. Paper presented at CUNY Sentence Processing Conference, New York.
- Bock, K., and Griffin, Z.M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*, 177-192.
- Boltz, M.G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics*, *60*, 1357-1373.
- Bradlow, A.R., Nygaard, L.C., and Pisoni, D.B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*, 206-219.
- Cathcart, E.P. and Dawson, S. (1928). Persistence: A characteristic of remembering. *British Journal of Psychology*, *18*, 262-275.
- Clark, H.H. (1996). *Using language*. NY: Cambridge University Press.
- Clark, H.H. (2002). Speaking in time. *Speech Communication*, *36*, 5-13.
- Clynes, M., and Walker, J. (1986). Music as time's measure. *Music Perception*, *4*, 85-119.
- Collier, G.L., and Collier, J.L. (1994). An exploration of the use of tempo in jazz. *Music Perception*, *11*, 219-242.
- Cooper, W.E. and Eady, S.J. (1986). Metrical phonology in speech production. *Journal of Memory and Language*, *25*, 369-384.
- Cooper, G. and Meyer, L.B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.
- Cooper, A.M., Whalen, D.H., and Fowler, C.A. (1986). P-centers are unaffected by phonetic categorization. *Perception and Psychophysics*, *39*, 187-196.
- Cutler, A., Dahan, D., and van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*, 141-201.
- Drake, C., and Botte, M.C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, *54*, 277-286.
- Drake, C., Jones, M.R., and Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, *77*, 251-288.
- Ellis, M.C. (1991). Thresholds for detecting tempo change. *Psychology of Music*, *19*, 164-169.
- Essens, P.J., and Povel, D.J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, *37*, 1-7.
- Fox Tree, J.E. (2000). Coordinating spontaneous talk. In L. Wheeldon (ed.), *Aspects of language production* (pp. 375-406). Philadelphia: Psychology Press.
- Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (ed.), *The psychology of music*. (pp. 149-180). New York: Academic Press.
- Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A Major (K331): A comparison of five performances. In A. Gabrielsson (ed.), *Action and perception in rhythm and music* (pp. 81-104). Stockholm: Royal Swedish Academy of Music.
- Grosjean, F.H., Grosjean, L., and Lane, H. (1979) The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, *11*, 58-81.
- Jones, M.R. (1987). Perspectives on musical time. In A. Gabrielsson (Ed), *Action and perception in rhythm and music* (pp. 153-176). Stockholm: Royal Swedish Academy of Music.
- Jones, M.R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*, 323-355.
- Jungers, M.K., and Palmer, C. (2000). Episodic memory for music performance. *Abstracts of the Psychonomic Society*, *5*, 105.
- Kemler Nelson, D.G., Juszczyk, P.W., Mandel, D.R., Myers, J., Turk, A., and Gerken, L.A. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, *18*, 111-116.
- Kjelgaard, M.M., and Speer, S.R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*, 153-194.
- Kosslyn, S.M., and Matt, A.M. (1977). If you speak slowly, do people read your prose slowly? Particular speech recoding during reading. *Bulletin of the Psychonomic Society*, *9*, 250-252.

- Large, E.W., and Jones, M.R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*, 119-159.
- Large, E.W., and Palmer, C. (2001). Perceiving temporal regularity in music. *Cognitive Science*, *26*, 1-37.
- LeBlanc, A., Colman, J., McCrary, J., Sherrill, C., and Malin, S. (1988). Tempo preferences of different age music listeners. *Journal of Research in Music Education*, *36*, 156-168.
- Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, *7*, 106-122.
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*, 253-263.
- Lehiste, I., Olive, J.P., and Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, *60*, 1199-1202.
- Levelt, W.J.M., (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.
- Levitin, D.J., and Cook, P.R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, *58*, 927-935.
- Martin, J.G. (1970). Rhythm-induced judgments of word stress in sentences. *Journal of Verbal Learning and Verbal Behavior*, *9*, 627-633.
- Merker, B. (2000). Synchronous chorusing and human origins. In N.L. Wallin, B. Merker, and S. Brown (eds.), *The origins of music* (pp. 315-327). Cambridge: MIT Press.
- Miller, J.L., Grosjean, F., and Lomato, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, *41*, 215-225.
- Munhall, K., Fowler, C.A., Hawkins, S., and Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics*, *20*, 225-239.
- Nygaard, L.C., Sommers, M.S., and Pisoni, D.B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics*, *57*, 989-1001.
- Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 331-346.
- Palmer, C. (1997). Music performance. *Annual Review of Psychology*, *48*, 115-138.
- Palmer, C., Jungers, M.K., and Jusczyk, P.W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language*, *45*, 526-545.
- Povel, D.J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception & Performance*, *7*, 3-18.
- Price, P., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*, 723-735.
- Repp, B.H. (1994). On determining the basic tempo of an expressive music performance. *Psychology of Music*, *22*, 157-167.
- Schafer, A.J., Speer, S.R., Warren, P., and White, S.D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*, 169-182.
- Speer, S.R., Crowder, R.G., and Thomas, L.M. (1993). Prosodic structure and sentence recognition. *Journal of Memory and Language*, *32*, 336-358.
- Stein, E. (1989). *Form and performance*. New York: Limelight.
- Streeter, L. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*, *64*, 1582-1592.
- Volaitis, L.E., and Miller, J.L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, *92*, 723-735.
- Wales, R. and Toner, J. (1979). Intonation and ambiguity. In W.E. Cooper and E.C.T. Walker (eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, N.J.: Erlbaum.
- Warren, P. (1999). Prosody and sentence processing. In S. Garrod and M. Pickering (eds.), *Language processing* (pp. 155-188). Hove: Psychology Press.
- Warren, R.M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review*, *92*, 574-584.
- Wayland, S.C., Miller, J.L., and Volaitis, L.E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America*, *95*, 2694-2701.

Received: July, 2002

Accepted: December, 2002