Psychonomic Bulletin and Review
2005, 12 (6), 957-968

THEORETICAL AND REVIEW ARTICLES

An oscillator model of the timing of turn-taking

MARGARET WILSON
University of California, Santa Cruz, California

and

THOMAS P. WILSON
University of California, Santa Barbara, California

When humans talk without conventionalized arrangements, they engage in *conversation*—that is, a continuous and largely nonsimultaneous exchange in which speakers take turns. Turn-taking is ubiquitous in conversation and is the normal case against which alternatives, such as interruptions, are treated as violations that warrant repair. Furthermore, turn-taking involves highly coordinated timing, including a cyclic rise and fall in the probability of initiating speech during brief silences, and involves the notable rarity, especially in two-party conversations, of two speakers' breaking a silence at once. These phenomena, reported by conversation analysts, have been neglected by cognitive psychologists, and to date there has been no adequate cognitive explanation. Here, we propose that, during conversation, endogenous oscillators in the brains of the speaker and the listeners become mutually entrained, on the basis of the speaker's rate of syllable production. This entrained cyclic pattern governs the potential for initiating speech at any given instant for the speaker and also for the listeners (as potential next speakers). Furthermore, the readiness functions of the listeners are counterphased with that of the speaker, minimizing the likelihood of simultaneous starts by a listener and the previous speaker. This mutual entrainment continues for a brief period when the speech stream ceases, accounting for the cyclic property of silences. This model not only captures the timing phenomena observed in the literature on conversation analysis, but also converges with findings from the literatures on phoneme timing, syllable organization, and interpersonal coordination.

When humans talk to one another, they overwhelmingly engage in *conversation*—that is, speech exchange between two or more parties where there is no external imposition of procedures on the flow of talk. This is in contrast with conventionally established forms of speech exchange, such as debates, interviews, ceremonies, and courtroom proceedings. In the latter cases, the order of speaking, length of turns, or content of what is said is managed by preestablished arrangements. Conversation, on the other hand, is managed *locally* by the participants, turn by turn, in terms of who speaks when, for how long, and about what.

Given its anarchistic nature, conversation proceeds remarkably smoothly. Typically, conversational partners take *turns*, with one person being treated by coparticipants as having the right and also the obligation to speak (while coparticipants have the right and obligation to attend to the speaker), and then smoothly relinquishing that status as another person begins talking. Thus, in most cases of conversation, speech is exchanged nearly continuously by speakers taking turns, with minimal gaps in the talk and minimal overlaps.

Simultaneous talk and silences do occur in conversation, but such events are themselves organized relative to turn-taking. Sometimes, this involves competition for the turn, as when two potential speakers start simultaneously, the previous speaker adds to an already completed utterance, or a listener directly interrupts in an attempt to take over the turn (French & Local, 1983; Schegloff, 2000). Such skirmishes are typically brief, usually settled by one party's dropping out. Other cases of simultaneous talk are treated as nonproblematic by participants. These include collaborative turn completions (Lerner, 1991, 1996) and recognition point interruptions (Jefferson, 1973, 1984), in which a listener begins speaking to indicate comprehension or affiliation, and back-channeling, in which brief interjections (e.g., "yeah" or "wow") occur that are not treated by either party as an attempt to take the turn. There are even situations in which *choral* coproduction can properly occur, such as mutual greetings when guests arrive at a party (Lerner, 2002). Finally, a silence can be treated by listeners as a positive action on the part of the current

This work was partially supported by the Max Planck Institute for Human Cognition and Neuroscience, Munich. The authors thank Emanuel Schegloff, Don Zimmerman, and two anonymous reviewers for their insights. Address correspondence to M. Wilson, Department of Psychology, University of California, Santa Cruz, CA 95064 (e-mail: mlwilson@ucsc.edu).

turn-holder (see, e.g., Heritage, 1984, pp. 273–274). For example, a direct question transfers the turn immediately to the recipient (see below, Option 1 of Sacks, Schegloff, & Jefferson's [1974] model), and for this reason, a pause before the question is answered may be interpreted by others as an act on the part of the recipient, reflecting mental work load (D'Urso & Zammuner, 1990) or honesty or comfort level (Fox Tree, 2002). Similarly, when the recipient of a request or invitation pauses before answering, this may be heard as portending some kind of awkwardness, such as a denial of the request or refusal of the invitation, whereas conversely, a *lack* of a pause can be interpreted as rudeness (Heritage, 1984, p. 268). The point to note in all these cases is that participants treat simultaneous talk and silences in ways that presuppose the normalcy of smooth turn-taking.

The features of turn-taking described above are ubiquitous. They hold across cultures and social classes, despite differences in the specifics of the verbal and nonverbal regulators employed (Caspers, 1998; Hafez, 1991; Kjaerbeck, 1998; La France, 1974; Lerner & Takagi, 1999; Murata, 1994; Robbins, Devoe, & Wiener, 1978; Sidnell, 2001; Streeck, 1996; Tanaka, 2000a, 2000b). Some authors have proposed, to the contrary, that there is a conversation style known as the *collaborative floor*, in which the turn-taking norms of the *one-at-a-time floor* do not hold (e.g., Coates, 1994, 1997); however, careful inspection of the examples offered suggests that although these conversations may be more collaborative in terms of *content*, the turn is generally held by one speaker at a time and apparent violations are, in fact, examples of the practices described above. To our knowledge, no culture or group has been found in which the fundamental features of turn-taking are absent. This is true even when the physical substrate of conversation is radically different from that of ordinary speech, as in the cases of sign language used by the deaf (e.g., Coates & Sutton-Spence, 2001, especially Conversation 1) and tactile sign language used by the deaf-blind (Mesch, 2000, 2001). Finally, turn-taking emerges early in human development, manifested in the structure of babies' vocal and bodily interactions with caregivers (Beebe, Alson, Jaffe, Feldstein, & Crown, 1988; Bloom, Russell, & Wassenberg, 1987; Crown, Feldstein, Jasnow, Beebe, & Jaffe, 2002; Elias, Hayes, & Broerse, 1986; Holmlund, 1995; Jasnow & Feldstein, 1986; Keller, Schölmerich, & Eibl-Eibesfeldt, 1988; Masataka, 1993, 1995; Rutter & Durkin, 1987; Stevenson, Ver Hoeve, Roach, & Leavitt, 1986).

Taken together, these findings point to the grounding of turn-taking in fundamental human cognitive processes. As Sidnell (2001) has argued, turn-taking in conversation may constitute a species-specific biological adaptation.

This raises a puzzle that has been oddly neglected in the cognitive sciences, which have otherwise taken great interest in issues of language, discourse, and social interaction. The puzzle is this: What are the cognitive mechanisms responsible for the smooth operation of turn-taking? As we shall see below, turn-taking involves a fine-tuned coordination of timing between speakers. In this article, we propose a model that can account for these timing phenomena.

The Phenomenon of Turn-Taking

The phenomenon of turn-taking in conversation was first explored systematically by Sacks et al. (1974; see Wilson, Wiemann, & Zimmerman, 1984, for a brief account of previous work). Since then, turn-taking has become a major focus of the field of *conversation analysis*, an interdisciplinary offshoot of sociology that also includes scholars from communication studies, discourse-oriented linguistics, and other fields (Heritage, 1984, chap. 8; Levinson, 1983, chap. 6).

The key insight is that because the order of speakers, length of turns, and content of what is said in conversation are not specified in advance, these matters must be managed by the participants during the interaction itself. Sacks et al. (1974) argued that turns consist of syntactically demarcated *turn-constructional units*, which can be anything from a word (e.g., "yeah") to a sentence, depending on the context. A turn can continue through many of these units, but the end of a unit that completes an *action* (a behavior treated by participants as pragmatically meaningful and complete) constitutes a *transition relevance place*, where a change of speakers could properly occur. (By *properly*, here and throughout the article, we mean that participants in the conversation treat such an event as acceptable and nonproblematic.)

The individual participant in a conversation is faced with a number of complex tasks: determining when a transition relevance place is coming up, determining whether the speaker intends to continue into another turn-constructional unit, preparing what they themselves might say, judging whether other listeners are intending to take the turn, speaking up quickly enough to take the turn, but also shutting down quickly if someone else gets there first. Given these demands, one of the extraordinary features of turn-taking is the precision of its timing.

The timing of turn-taking. Turn transitions are commonly so tightly synchronized that the next speaker begins speaking with virtually no gap following the end of the prior speaker's utterance. This is not to say that silences of longer durations are abnormal occurrences, but as was discussed earlier, such silences are treated by participants as noticeable events, with pragmatic implications (see Jefferson, 1989). Furthermore, as we will see below, there are reasons to believe that even quite short gaps may result from the next speaker's temporarily declining to speak, rather than being "off the mark" in producing a no-gap transition. Bearing these disclaimers in mind, it is nevertheless the case that turn transitions with virtually no gap are a common occurrence in ordinary conversation.

Conversation analysts, in fact, distinguish multiple varieties of these very minimal gaps that occur at turn transitions. The most lengthy of these are termed *micropauses*, transcribed as (.), which are gaps long enough to be clearly hearable by experienced transcribers but too brief to be measured by manual methods. A shorter and

more common form of transition is not given any special marking by transcribers but is assumed as the default. These transitions, which we will refer to as *simple transitions*, are ones in which there appears to the casual listener to be essentially no gap but, in fact, on careful listening there is what transcribers refer to as a *beat* between the two turns (Jefferson, 1984). Finally, there are the somewhat less common *latched* transitions (transcribed with = between the two items), in which the next turn is jammed up against the prior turn with no discernable space but, also, no discernable overlap with the prior turn.

Unfortunately, it is not clear exactly what durations transcribers are assigning to the categories of micropause, simple transition, and latch, and it is likely that these vary across transcribers. Furthermore, because studies of discourse processes typically either rely solely on transcriber judgment or, at best, use approximate measures of timing, such as stopwatch measurement (e.g., Beebe et al., 1988), stopwatch measurement of audiotapes played at quarter speed (e.g., Trimboli & Walker, 1984), or digital sampling every 250 msec of the soundstream (e.g., Crown et al., 2002), precise data on the duration of between-speaker transitions are scant; nevertheless, it is clear that these gaps can be strikingly brief. Walker and Trimboli (1984) reported that observers who are not practiced transcribers show a threshold of approximately 200 msec for detecting between-speaker gaps. It is likely, then, that the simple transitions and latches reported by experienced transcribers are somewhere below this 200-msec mark.

A more accurate picture of the frequencies of between-speaker gaps of various durations can be obtained from a reexamination of data originally reported in Wilson and Zimmerman (1986). In that study, a 9-min segment from each of 7 two-party conversations was analyzed, always beginning with the 2nd min of the conversation. For each segment, the durations of all between-speaker silences were measured with a sampling interval of 10 msec. (Very few gaps longer than 1 sec occurred, and those that did were excluded from Wilson and Zimmerman's analysis. No analysis was made of cases of overlap.) Table 1 shows frequency counts for silences in 100-msec increments. It is clear from these data that very short gaps are routine. Of the total number of between-speaker silences, 30% are less than 200 msec and over 70% are less than 500 msec. Furthermore, Table 2 shows that the shortest gaps (in the 10- to 100-msec range in Table 1) span the range down to 20 msec.

For purposes of contrast, note that simple reaction time for initiating a vocal response to a variably timed cue under maximally favorable conditions (the participants are highly trained, the participants are highly alert, cue initiation is under the participant's control, cue is a stimulus onset rather than an offset, variation in cue timing is small, and the response consists of a neutral vowel) is approximately 200 msec (Izdebski & Shipp, 1978, Table 1). Under more complex and less predictable stimulus conditions (but still with simple reaction time—that is, a prechosen response), vocal reaction times can be in the range of approximately 500–800 msec (Kuriki, Mori, & Hirata, 1999, Table 1). Thus, it seems clear that turn-taking does not proceed by listeners' waiting for the speaker to fall silent before initiating their own speech. Instead, listeners must be projecting the upcoming transition relevance place and making the physiologically required preparations for speech well before the actual cessation of the current talk (see Walker, 1982).

Table 1
Frequencies of Durations of Between-Speaker Silences, in Increments of 100 msec up to 1,000 msec, From the Conversations Analyzed in Wilson and Zimmerman (1986)

Duration	Frequency
10–100	73
110–200	198
210–300	169
310–400	117
410–500	89
510–600	68
610–700	58
710–800	46
810–900	55
910–1,000	32

Table 2
Frequencies of Durations of Between-Speaker Silences, in Increments of 10 msec up to 100 msec, From the Conversations Analyzed in Wilson and Zimmerman (1986)

Duration	Frequency
10	0
20	5
30	9
40	8
50	9
60	6
70	11
80	11
90	5
100	9

Unfortunately, even less is known about cases of brief overlap between the current speaker and the next speaker. Transcribers do not have a vocabulary of symbols for these, instead indicating overlap with actual overlap in the horizontal position of typed lines in transcripts. The data collected by Wilson and Zimmerman (1986) included only cases in which the acoustic signal fell below a preset threshold, indicating silence from both participants. Thus, their data did not include any cases of minimal overlaps (and also, incidentally, may have missed some cases of near-zero silences as well). Clearly, a full picture of the timing of turn-taking would require some knowledge of the durations of overlaps. Nevertheless, one fact is known that is of relevance here, which is that overlaps are particularly likely when the current speaker unexpectedly draws out his or her final syllable (Sacks et al., 1974, p. 707). This accords with the claim made above that listeners mentally project the probable moment at which the current utterance will end.

Between-speaker silences. There is a further, rather sophisticated feature of the timing of turn-taking. We described earlier how turn transitions occur at syntactically and pragmatically demarcated transition relevance places. At such points, a new speaker may take the turn, but it is also possible that the current speaker will continue. Alternatively, neither of these options may be invoked for some interval, resulting in a brief silence. Such a silence can then be broken by either the previous speaker or the listener. Despite this variety of possibilities, there is a striking orderliness to turn-taking; in particular, when there is a gap in a two-party conversation, it is rare for both participants to then begin speaking at the same time. To account for these facts, Sacks et al. (1974) proposed a model in which allocation of the turn can occur by any of the following processes.

1. *Current selects next*. The speaker explicitly passes the turn to someone else—for example, by directing a question or a request to a particular individual. The turn goes to the selected next speaker immediately on the current speaker's finishing his or her turn.

2. *Listener selects self*. A listener may choose to begin speaking. The person who talks first properly gains the turn.

3. *Current selects self*. The current speaker may resume speaking.

According to the model, these options arise in a specific order. The *current selects next* option is available to the current speaker during the current turn; if that does not happen, a current listener can self-select; if no listener self-selects, the current speaker can continue; and finally, if none of these occurs, the process recycles to Option 2. Thus, the options for listener or speaker to self-select occur sequentially and occur cyclically if no one self-selects on the first pass. The model is formulated as a process in real time, so that repeated cycling between Options 2 and 3 results in a gap in the talk (Sacks et al., 1974; Wilson & Zimmerman, 1986).

This assumption of serially ordered options was based on phenomena observed from transcripts and tapes, such as the rarity of simultaneous starts following brief silences in two-party conversations, as was noted above. (Simultaneous starts are more likely with three or more participants, since multiple listeners may then compete at Stage 2.) This phenomenon suggests that, at any given instant, either the previous speaker or the listener has the "right" to speak, but not both. In terms of Sacks et al.'s (1974) model, this indicates that Option 3 cannot be collapsed into Option 2, since the current speaker does not appear to compete with the listener for the right to speak. In other words, the current speaker's starting to speak again after other potential speakers have declined is different from the logically possible alternative of the current speaker's being involved in the competition among listeners allowed for by Option 2.

Subsequently, a more fine-grained temporal analysis of silences revealed exactly such an orderly progression in the exercising of options. As was described earlier, Wilson and Zimmerman (1986), motivated by the considerations above, analyzed the durations of between-speaker silences in two-party conversations. They found that these gaps are not of arbitrary length but, instead, tend to be multiples of some unit length of time, designated as *S* (the value of *S* varying from conversation to conversation). This was shown by the use of an autoregressive integrated moving average analysis, which showed a periodicity in the lengths of the between-speaker silences.

This counterintuitive result is, in fact, exactly what is predicted by Sacks et al.'s (1974) model, in which the options of *listener selects self* and *speaker selects self* occur serially and cyclically. In two-party conversations in which the silence is resolved by the listener's taking the turn, the option to speak will have been passed from the listener back to the original speaker (when the listener fails to exercise Option 2 and back again to the listener (when the speaker fails to exercise Option 3 some integral number of times. Since this passing back and forth takes place in real time, this translates into integral multiples of some measurable duration *S*. In Wilson and Zimmerman's (1986) data, *S* ranged from 80 to 180 msec across the seven conversations, with an average of 120 msec.

This leads to a strong theoretical claim at the cognitive level: Each participant must be keeping track (not necessarily consciously) of whose option it is to speak and exercising his or her own option only at the appropriate moment. The counterintuitive nature of this claim increases when we consider multiple cycles of silence. (The cyclical passing of the option to speak does break down eventually, resulting in a *lapse*, which can be terminated by any participant at any moment. Exactly how long it takes for the cycling of options to break down is unclear, but it is probably a matter of a few seconds.)

Cues that regulate turn-taking. A number of authors have proposed that listeners project an upcoming end of a turn by using semantic, syntactic, prosodic, eye gaze, or body movement cues produced by the speaker (Bavelas, Chovil, Coates, & Roe, 1995; Beattie, 1979; Beattie, Cutler, & Pearson, 1982; Caspers, 1998, 2003; Clark & Fox Tree, 2002; Craig & Gallagher, 1982; Koiso, Horiuchi, Tutiya, Ichikawa, & Den, 1998; Miura, 1993; Robbins et al., 1978; Schaffer, 1983; Stephens & Beattie, 1986a, 1986b; Wells & MacFarlane, 1998; for reviews, see Ford & Thompson, 1996, and Fox Tree, 2000). Conversely, a variety of devices have been proposed by which listeners can indicate their desire to take the turn, such as movements, audible inbreath, or interjected words (Bavelas et al., 1995; Dittmann & Llewellyn, 1968; Duncan & Niederehe, 1974; Harrigan, 1985; Thomas & Bull, 1981). Furthermore, speakers can respond to these attempts with adjustments in their own production, such as a *rush-through* or the interjection of semantic cues, such as "then" or "anyway," to indicate an intention to continue (e.g., Fox Tree, 2000, p. 387; Schegloff, 1982, p. 76).

In fact, it is likely that the identifying of potential turn transitions is an opportunistic process, with listeners and speakers exploiting any and all available cues, which may vary across cultures, languages, and situations. For example, Tanaka (2000b) has argued that the grammar of Japa-

nese results in a later point in time at which a turn ending can be projected than in English but that this is compensated for by a greater degree of certainty in Japanese, due to devices that specifically mark transition relevance places. As another example, visual cues may be employed in face-to-face conversations but will clearly be of no use for telephone calls or conversations in the dark, in which normal turn-taking nevertheless occurs (Sellen, 1995).

However, note that enumerating the relevant cues can explain how listeners know *that* a turn is ending but is of less help in explaining how listeners know precisely *when* a turn is ending. Given the precision of turn transitions, it seems likely that listeners form increasingly precise predictions of the end of a turn as the utterance proceeds through its final syllables (see Walker & Trimboli, 1984). Still absent is a concrete cognitive theory of how the listener projects the time course of the unfolding speech stream into the immediate future in order to begin, before the unfolding speech has been completed, the initiation of his or her own speech production process.

The phenomena discussed in this section—the normality of gaps too small to hear and the passing of the option to speak at regular, extremely short intervals—imply an entrainment of timing between participants. Below, we will develop an account of how this is accomplished.

Oscillators and Turn-Taking

In order for listeners and speakers to show this kind of precision in mutual timing, it is necessary, at a minimum, that there be some form of cyclic patterning in the cognitive processes of the speaker that influences the timing of the cognitive processes of the listener.

This suggests the involvement of endogenous oscillators known to exist in the human brain. Endogenous oscillators are populations of neurons that collectively show periodicity in their activity and serve timing-related functions in the brain. These oscillators have been implicated in a range of cognitive processes, including perception, motor control, attention, memory, and consciousness (for reviews, see Buzsáki & Draguhn, 2004; Penttonen & Buzsáki, 2003; Ward, 2003), as well as perhaps more generally providing a temporal framework for information processing, resulting in discrete processing epochs (Burle & Bonnet, 1999; Körner, Gewaltig, Körner, Richter, & Rodemann, 1999). The periodicity of these oscillators varies, ranging across three bands of slow frequencies (covering 0.25–1.5 Hz), four intermediate frequencies known as *delta* (1.5–4 Hz), *theta* (4–10 Hz), *beta* (10–30 Hz), and *gamma* (30–80 Hz), and two bands of fast frequencies (covering 80–600 Hz). The periodicity involved in between-speaker silences, from 180 msec down to 80 msec (5.5–12.5 Hz), is compatible with the operation of endogenous oscillators falling roughly in the theta range. (We should note, though, the paucity of data on which this estimate is based. The range of periodicity across individuals, cultures, and conversational contexts remains unexplored.)

One useful property of oscillators is that they act as timing devices and are, therefore, ideal for solving the problem discussed earlier—that of making the leap from identifying cues that signal an upcoming transition to knowing precisely when the transition will occur. In general, encoding periodicity of a signal increases the ability to predict the future course of events, particularly with respect to their timing (see Jungers, Palmer, & Speer, 2002, p. 33; Large & Jones, 1999, p. 123). As Buzsáki and Draguhn (2004) put it, "feedforward and feedback networks predict well what happens next. Oscillators are very good at predicting when" (p. 1929).

A further property of oscillators that is critical for our purposes is that, when they are allowed to influence one another, they tend to become phase locked. This is true across a wide range of physical systems, including not only populations of neurons, but also fireflies, pendulums, electrons, and asteroids (Strogatz, 2003). We propose, then, that during conversation, periodicity within the information-processing system of the listener tends to synchronize with that of the speaker. Still needed, though, is a channel by which the oscillatory signal is transmitted from person to person. The cyclic pattern must be present not just in the brain of the speaker, but in the behavior as well, and must be detectable by the listener.

Entrainment of cyclic behaviors during conversation is already known to exist. Breathing patterns, for instance, become entrained between conversational participants, with listeners' breathing cycles coming to resemble the sawtooth function of speakers (short inhale and long exhale), rather than the more uniform cycle of quiet breathing (McFarland, 2001). Furthermore, the breathing cycles of the speaker and the listener tend to become phase locked to one another in the few seconds surrounding a turn transition (McFarland, 2001). Interestingly, this phase locking can be either in-phase (peak co-occurring with peak) or counterphase (peak co-occurring with trough), although one pattern or the other tends to dominate within a particular conversational dyad. The speechlike pattern of breathing does not occur for passive listening, in which the listener does not have the opportunity to participate in conversation with the speaker (Shea, Walter, Pelley, Murphy, & Guz, 1987). This suggests that the imminent opportunity to take the turn from the speaker is responsible for the change in breathing pattern.

However, the timing involved in breathing is extremely coarse, as compared with that involved in turn-taking. In McFarland's (2001) data, the duration of breathing cycles during conversation was in the range of 3–7 sec. In contrast, as was noted above, the cyclicity of timing during between-speaker silences is in the range of 80–180 msec. We must look, then, for another source of cyclic patterning.

The obvious contender is the mandibular oscillations, or jaw cycles, that make up the opening and closing pattern of syllables. A syllable consists of a consonant or consonant cluster and an accompanying vowel or dipthong (vowel cluster). Consonants are moments of relative closure of the vocal tract, and vowels are moments of relative opening. Thus, as the vocal tract produces a stream of ongoing speech, the progression of syllables constitutes a cycling of the jaw between closure and opening. It has

been speculated that this cyclic pattern of speech evolved from older cyclic uses of the mouth, including ingestive behaviors, such as chewing and licking, and communicative behaviors, such as lip smacks and teeth chatters (MacNeilage, 1998; MacNeilage, Davis, Kinney, & Matyear, 2000). MacNeilage proposes that this cyclic pattern provides an underlying syllabic *frame* for speech, which is then filled in with specific phonemes, and that it furthermore forms the basis for infant babbling.

A frequently repeated estimate of syllable rate in American speech is five syllables per second, or 200 msec per syllable, but this estimate is an informal one. Furthermore, much of the relevant research concerns the reading of sentences under laboratory conditions, and in fact, speech rates in natural conversation may be considerably faster than this. H. Nusbaum (personal communication) has estimated that syllable rates in natural speech may, in fact, be in the range of 100–150 msec per syllable. This corresponds nicely to the estimate of 80–180 msec per cycle in Wilson and Zimmerman's (1986) study.

Explaining Timing of Turn-Taking With an Oscillator Model

On the basis of the considerations above, we propose a model of the timing of turn-taking with the following assumptions.

1. The timing of turn-taking in conversation is based on an oscillatory function of readiness to initiate speech, occurring in both the speaker and the listener.

2. The frequency of this oscillation is determined by the speaker's syllable rate. More specifically, a cycle of the production of a syllable corresponds inversely to a cycle of the speaker's readiness for initiating a new syllable. Readiness is at a minimum in the middle of syllable production, at the point of greatest syllable sonority (roughly, greatest *openness* or *vowel-like-ness*). As sonority lessens and the coda of the syllable is produced, the ability to initiate a new syllable rises. This appears at first to be not so much an assumption of the model as a consequence of the physical realities of syllable production. However, as will be seen in Assumption 5, the model necessitates that this cyclic preparation be instantiated in the speaker independently of whether syllable production is currently occurring.

3. The listener also engages in an oscillator-based cycle of readiness to initiate a syllable. This cycle is entrained to that of the listener, via the medium of the speaker's rate of syllable production. The listener is maximally prepared to begin speaking (to take the turn, or to interrupt) at the peaks of this cycle.

4. The listener's cycle is counterphased to that of the speaker. The listener's potential to initiate is at a minimum when the speaker begins a syllable and is at a maximum when the speaker is mid-syllable.

5. If the listener does not begin speaking in the first cycle following the previous speaker's completion, the oscillators in speaker and listener continue to be mutually entrained for some short period of time. Eventually, due to the lack of signal transmission that could maintain the entrainment, the two oscillator populations drift apart, and the mutuality of the timing metric is lost.

Let us now examine how this model plays out in the negotiation of turn transitions and how it accounts for the known phenomena. When the speaker finishes the final syllable of a turn-constructional unit, he or she may already be in the process of initiating a new syllable, if he or she plans to continue into a new unit. However, if the speaker does not seamlessly initiate further syllabic production, he or she cannot initiate such production at any arbitrary instant after completion. Instead, for some number of cycles, he or she will be mentally *mid-syllable* at certain points in time and, hence, unprepared to initiate. He or she must wait until he or she reaches the next peak of the readiness cycle, the next virtual syllable beginning, before initiating again.

Conversely, the listener who detects an approaching transition relevance place will be maximally prepared to initiate as the final syllable of the speaker emerges or in the first half-cycle after the speaker has finished. This generates the counterintuitive prediction that latches, as described in the conversation analysis literature, are not, in fact, cases of zero silence and zero overlap; instead, they are cases of silence or overlap so brief that they sound like latches to the unaided human observer. To take a concrete example, if the speaker and the listener are mutually entrained to a rate of 150 msec per syllable, the listener, according to this model, has the option to initiate speech at approximately 75 msec *before* or 75 msec *after* the speaker's final syllable is complete. Either or both of these cases may be recorded by transcribers as latches. (Recall that untrained listeners do not reliably hear gaps of less than 200 msec.) Note that this prediction is supported by the data in Table 2, as far as they go. Although these data aggregate across conversations, each with its own periodicity, none of the conversations shows a tendency toward transitions of zero duration, and in fact, no transitions at all were found at the shortest duration measured. (Missing, of course, from Table 2 are the critically interesting data of turn transitions where there is brief overlap.)

To return to our account of turn allocation, if the listener does not exercise the option to speak in the half-cycle immediately preceding or following the speaker's completion, a brief silence will ensue. During this silence, both the speaker and the listener will be going through cycles of rising and falling potential for initiating speech. Because they are counterphased to one another, the probability of simultaneous starts will be relatively low. However, because there is no ongoing signal to keep the participants calibrated to one another, the entrainment will eventually break down, resulting in a lapse.

Support for the Role of Syllables

We have argued that syllables form the basis of the coordinated timing of turn-taking. However, we should ask whether there are other forms of signaling that might make equally plausible candidates. One possibility is some form of nonverbal signaling, such as subtle body movements. However, two lines of evidence argue against this.

First, Shockley, Santana, and Fowler (2003) found that coordination between partners was carried by a verbal and not by a visual signal. Although in this study body sway was examined, rather than turn-taking, the results may be relevant to the present argument. Shockley et al. found mutual influence of body sway between dyads who were talking to each other, even when not facing each other, but not between dyads who were each talking to other parties, even when facing each other. That is, seeing the other person's body sway does not influence the observer's body sway, but conversing with the other person (even without seeing) does. In short, the medium through which the coordination was transmitted was verbal, not visual.

Unfortunately, due to the complexities of measuring body sway, the technique used by Shockley et al. (2003) does not reveal what exactly has been transmitted between the two participants nor at what time-scale the mutual influence is occurring. One possibility, for example, is that entrainment of breathing is driving coordination of body sway. Nevertheless, although these data do not speak directly to the coordination of turn-taking, they do suggest that visually observing the other person's body movement does not provide the basis for mutual entrainment.

Second, there is the simple and well-established fact that the vocal signal alone is sufficient for participants to coordinate smooth turn-taking. As was noted earlier, turn-taking proceeds normally in a variety of situations in which there is no visual contact, including conversations in the dark, over the telephone, between blind people, or when visual attention is allocated elsewhere (e.g., while one is driving or chopping vegetables).

Thus, the medium by which speakers and listeners influence each other to coordinate turn-taking appears to be present in the vocal signal itself. (We should note that these observations are based on spoken language. In sign language, transmission will, of course, be visual, but on this account would presumably likewise be carried in the linguistic signal, rather than in other aspects of body movement.) This conclusion makes sense from an evolutionary standpoint. Language is fundamentally grounded in social interaction, and it is likely that the mechanisms of language production and turn-taking coevolved, perhaps building on the same preexisting cognitive structures.

Assuming, then, that the timing information is carried in the speech stream itself, are there possible vehicles for this other than the syllable? In particular, one could argue that the primary rhythmic unit of the language might be a more salient feature. Languages are generally described as being stress timed (e.g., English), syllable timed (e.g., French), or mora timed (e.g., Japanese). Thus, one might argue that the basis of mutual entrainment could be the stress foot, syllable, or mora, depending on the language being spoken. However, whereas both the syllable and the mora (which is a syllable or part of a syllable) are close to the appropriate range of timing, a stress foot can contain several syllables and, thus, can be considerably longer, which places it beyond the range of our estimate of cycle frequency for turn-taking. This particularly raises problems since that estimate was derived from English, a stress-timed language. Thus, the syllable, as a universal feature of language, seems the most likely basis for entrainment.

Additional Support for the Model

The model offered here, which was developed to explain the turn-taking data reported in the conversation analysis literature and, particularly, in Wilson and Zimmerman (1986), in fact fits well with a variety of other findings from a range of literatures. Of particular note is the fact that many of these findings would not be predicted by alternative accounts of turn-taking. Such alternatives include noncyclic models, wherein each participant merely projects forward in time to anticipate the onset or offset of the other's speech on the basis of various predictive cues, and models that appeal in general to cyclic processes but do not share the specific commitments to mutual entrainment, counterphasing, and the role of syllable rate.

Converging speech rates. One prediction of the model is that the speech rates of participants in a conversation should be similar. The entrainment of the syllable rates of speakers and listeners not only should facilitate the smooth handing over of the turn (similar to runners in a relay race matching their footfalls as they pass the baton), but also should have the additional consequence that the next speaker will begin speaking at a rate similar to that of the previous speaker. Evidence on this has been reported by Street (1984), who found convergence of the speech rates of participants in two-party conversations. (Additional evidence comes from a study by Jungers et al., 2002, in which speakers' rates were influenced by the rate of an immediately preceding heard sentence, although this study did not involve turn-taking in a conversational context.) There are, of course, many reasons why conversational partners might match their speech rates, such as signaling mutual affiliation. However, it is worth noting that whereas nonoscillatory accounts remain silent on this point, the phenomenon follows naturally from the present model. Furthermore, the present model offers a mechanism by which such matching occurs "for free," without conscious effort.

Oscillators and variation of timing. Although the syllable is the strongest contender for the behavioral signal that allows speaker and listener to become time locked, one difficulty arises: Syllables, although cyclic, vary widely in duration, due to many factors. This variation is not just occasional; it is endemic in natural speech. Long words show syllable compression, relative to short words; syllables are lengthened or slight pauses inserted before syntactic breaks; conveying emphasis or emotion with affective prosody can result in both shortening and lengthening of syllables; and cognitive demands on the listener can slow the overall rate of speech production. (These problems are equally true for stress feet and morae; Cutler, 1991; Warner & Arai, 2001.) Is it possible for the shared timing between speaker and listener to be extracted from so inconsistent a source of cyclic signal?

Surprisingly, this problem can be solved with an oscillator model. Brain-based oscillators show properties of *relaxation oscillators*, rather than *harmonic oscillators*

(Buzsáki & Draguhn, 2004). The former are susceptible to outside influence only during one phase of their cycle and, therefore, are highly stable and predictable in their timing properties. This stability means that such oscillators can tolerate some degree of perturbation in the incoming signal. In contrast, a timing device based on a clock or temporal *grid*, such as that proposed by Povel (1981; Povel & Essens, 1985), is quite rigid with respect to deviations of timing (Large & Jones, 1999, p. 122). Vousden, Brown, and Harley (2000, p. 161) argued that stimulus-driven oscillators can, in fact, accommodate variation of up to 50% in the rate of the stimulus signal while still maintaining the underlying rhythm of the oscillation. A similar point has been made for perceiving temporal regularity in music and other temporally structured events, despite ubiquitous deviations from actual strict temporal regularity, and this has been successfully modeled with an oscillator model (Large & Jones, 1999; Large & Palmer, 2002). Thus, a stable and regular oscillator-driven beat may underlie the many surface variations in the rate of syllable production in the speaker; and on the listener's side, the cyclic but variable nature of perceived syllables may entrain an oscillator in the listener that likewise establishes a stable and regular beat. Moreover, such a mechanism does not preclude the possibility of oscillations that are stable in the face of short-run input perturbations yet are responsive to longer run shifts in pacing over the course of a conversation.

Phoneme identification based on fine-grained timing. A further strength of the present theory is that it helps to account for other phenomena in speech production and perception that demand a highly sensitive and reliable timing device, despite all the variation that exists in the signal. For example, some languages use duration of a phoneme as a distinctive feature, so that a long and a short version of the "same" vowel are in fact two distinct phonemes that can result in two entirely different words when one is substituted for the other. The same holds true for geminate (lengthened) consonants and nongeminate consonants. With all the sources of variation in speech rate discussed above, how does a listener know whether a speaker intends a short phoneme or a long phoneme? Along similar lines, in many languages, listeners must distinguish consonants with an early voice onset time (VOT), such as /d/ and /g/, from consonants with a late VOT, such as /t/ and /k/. Again, the question arises of how listeners make this distinction, which depends on fine-grained timing, in a speech environment with high variability in timing. The fact that listeners have little trouble making these distinctions in natural speech suggests the existence of an underlying timing metric shared by speaker and listener.

In fact, it is known that VOT varies with speech rate, so that slower speech rates correspond to longer VOTs, and furthermore, listeners are attuned to this, in that they are influenced by speech rate when identifying voiced versus voiceless consonants (see Allen, Miller, & DeSteno, 2003, for a review). Wayland, Miller, and Volaitis (1994) have identified two sources of this influence. The first, of less interest here, is the duration of the individual syllable in which the phoneme occurs. Wayland et al. described this effect as intrinsic, in the sense that the relevant acoustic property is not actually VOT alone but, rather, the relationship between VOT and syllable duration, which constitutes a higher order acoustic property. However, Wayland et al. also found an independent effect of sentence-level speaking rate.

There has been little, though, in the way of proposals of how listeners actually encode speech rate (i.e., with what metric), in order to apply it during phoneme identification. According to the account being offered here, listeners encode speech rate via oscillations entrained to the speaker's overall rate of syllable production. This may well provide the mechanism by which speech rate influences both the production and the perception of duration-based phonemic features to coincide with the intended phoneme category.

Furthermore, the use of oscillators tuned to syllable production can help to explain how speech rate, which occurs at a larger time scale than do time-based features of phonemes, can nevertheless be playing a role in fine-grained temporal processing of these phonemes. The answer lies in the fact that, within the mammalian brain, lower frequency oscillators can influence higher frequency oscillators. As Buzsáki and Draguhn (2004) put it, "perturbations occurring at slow frequencies can cause a cascade of energy dissipation at higher frequencies and . . . widespread slow oscillations modulate faster local events" (p. 1926). Thus, syllable rate can provide a larger scale metric that helps to regulate and stabilize the timing of events on a smaller scale, such as the temporal properties of individual phonemes. Note that this differs from an account considered and rejected by Wayland et al. (1994), in which sentence-level timing sets a clock that determines VOT.

One final point to note is that, as was suggested above, the timing metric must be shared by the speaker and the listener. In order for language to be a robust communicative signal, the speaker and the listener must share an understanding of what the stimulus is intended to be. A speaker who intends a short vowel must, on the whole, be heard as producing a short vowel. Thus, the data on phoneme discrimination necessitate not only a representation of speech rate that can influence timing at other levels, but also a mechanism by which that representation is shared between a speaker and a listener.

This argument on logical grounds is supported by evidence that disturbances in the timing of speech production and of speech perception go hand in hand—specifically, in patients with cerebellar damage (see Ackermann, Mathiak, & Ivry, 2004, for a review). Disruptions of production include increased variability of syllable duration at faster production rates, whereas controls actually show increased uniformity at faster rates (Kent, Kent, Rosenbek, Vorperian, & Weismer, 1997), and show greater variability than do controls in VOT (Ivry & Gopal, 1992). Disruptions of perception include difficulty in the discriminating of duration-based phonemic distinctions (Ackermann, Gräber, Hertrich, & Daum, 1997). On the basis of this and

other evidence, Ackermann et al. (2004) proposed that the cerebellum governs the tempo of articulation of a planned sequence of syllables. However, their proposal does not provide an account of why timing functions for speech perception and speech production are linked, other than to propose that the cerebellum provides a *common platform* for these two functions. The present proposal provides an answer to this question, by postulating a functional link between the tempo of perceived speech and the tempo of produced speech.

The literature on phoneme identification, then, lends support to the present claims of a metric for speech rate shared between a speaker and a listener. Furthermore, the influence across time scales suggests that entrainment of oscillators is a plausible mechanism.

Oscillator models of syllable representation. A further strength of an oscillator-based account of turn-taking is that it dovetails with a class of models of speech production and perception derived from an entirely different set of empirical considerations (Harris, 2002; Hartley, 2002; Vousden et al., 2000; see also Brown, Preece, & Hulme, 2000, Burgess & Hitch, 1992, 1999, and Hartley & Houghton, 1996, although the latter works focus on phonological representation in short-term memory, rather than during online perception and production). These authors have proposed oscillator-based models of the serial ordering of phonemes, in which the syllable is a basic unit of oscillation. According to these models, ordered sequencing of phonemes is governed in part by a set of oscillators that represent the *repeating portion* of the signal—that is, the cyclic nature of syllable production, with one cycle corresponding to one syllable. Such oscillation successfully explains patterns of errors in phoneme production, such as the tendency for misplaced phonemes to show up in the wrong syllable but in the correct syllabic position (onset, peak, or coda). Thus, we have convergence from two separate literatures—the speech error literature and the turn-taking literature—toward an oscillation-based account of the representation of syllables.

A role for entrained breathing. Above, we described data on the entrainment of breathing between conversational participants—particularly, near turn transitions (McFarland, 2001). We abandoned this as a primary explanation for the timing of turn-taking, due to the time scale involved. However, this leaves the puzzling question of why breathing should become entrained. What function does this fulfill?

One possible role is suggested by the fact, noted above, that oscillators at slower time scales can influence oscillators at faster time scales (Buzsáki & Draguhn, 2004). This suggests that the coordination of breathing around turn transitions, despite being at the "wrong" time scale, may nevertheless play a role in the smooth management of turn-taking, by helping to regulate the timing of faster paced events. That is, entrained breathing may help to stabilize and amplify the entrainment of syllable rate, resulting in a more robust coordination between the parties at the finer-grained time scale. In support of this possibility, failure to coordinate breathing around a turn transition is associated with simultaneous speech—that is, a departure from smooth turn-taking (McFarland, 2001). This suggests that when the coordination of breathing is disrupted, there is an increased likelihood that coordination at the time scale required for turn-taking will also cease to function smoothly.

Predictions of the Model

A number of further predictions emerge from the model proposed here that set it apart from alternative accounts and indicate ways in which the model can be verified.

Matched rates of syllable production. The prediction that conversational participants should match their rates of syllable production, particularly near turn transitions, can be examined in greater detail than in the study by Street (1984). Street's measure of speech rate was not directly related to syllable rate and is, therefore, a somewhat indirect measure for the present purposes. In addition, a further prediction emerges: Not only should syllable rates of partners correlate with one another, but in addition, both should also correlate with *S*, the duration of one cycle passing the right to speak during a silence, as identified by Wilson and Zimmerman (1986). That is, in conversations with short *S*, participants should have rapid syllable rates, and in conversations with long *S*, participants should have slow syllable rates.

Moreover, the model predicts that, because well-matched speech rates should be the normal state of affairs, departures from this will be noticeable events that can be commented on and can have pragmatic implications and furthermore, that sustained badly matched speech rates should be associated with problems in smooth turn-taking.

Timing of near-zero gaps and overlaps. As was noted above, the model predicts that the very shortest turn transitions (i.e., latches) should not show a tendency toward true zero but, rather, should cluster on either side of zero as extremely short gaps and overlaps. This is because of the counterphasing of the readiness cycles of speaker and listener. In addition, each of these two clusters should deviate from zero by approximately half of the value of *S*.

Verifying the existence of lapses. In the field of conversation analysis, there is widespread informal agreement as to the existence of lapses—that is, cases in which, after the passage of sufficient time, the cycling of options in Sacks et al.'s (1974) model breaks down and any party may begin speaking at any moment. If this is true, simultaneous starts by two parties should become much more common after the threshold for a lapse has been reached. However, surprisingly few data exist on this point. Jefferson (1989), in examining silences of approximately 1 sec, noted that simultaneous starts are rare (p. 192), which suggests that the mutual entrainment and cycling of options lasts at least this long. To our knowledge, though, there is no published analysis of the frequency of simultaneous starts as a function of duration of the preceding silence. Both Sacks et al.'s model and the present model predict a discontinuity in the function, with a notable increase in simultaneous starts after some point. Furthermore, the point that marks the increase in simultaneous

starts should also be the point at which a coherent cyclic pattern of silence durations breaks down.

Within-speaker silences. The data in Wilson and Zimmerman (1986) included only between-speaker silences—that is, cases in which Party A stops speaking and Party B starts speaking. However, if the present account of their data is correct, it should hold true for within-speaker silences as well. According to the present account, the cyclic nature of the duration of silences should not be a peculiarity of transfers of the turn from A to B but should occur any time that the turn is up for grabs. This should include, then, cases in which A finishes a turn and then happens to be the next party to take the turn again. Furthermore, the cyclic pattern should obtain even in cases in which semantic or other factors indicate that the turn is *not* up for grabs, that Speaker A still holds the turn and is pausing for cognitive or pragmatic reasons. This is because, even though not in a state of alternating the right to speak, as specified in Options 2 and 3 of Sacks et al.'s (1974) model, the speaker's readiness function will continue to rise and fall. Note that this is a prediction that does not follow from Sacks et al.'s formulation but is necessitated by the present, more specific account.

Simultaneous starts in larger conversations. Finally, recall that according to Sacks et al.'s (1974) model, in a conversation with three or more participants, if the current speaker does not select the next speaker, the potential next speakers can compete for the turn, and this results in a higher probability of simultaneous starts than in dyadic conversations. This is a natural consequence of the present model as well, for in such a larger conversation, the readiness functions of all the listeners are counterphased to that of the speaker and, consequently, are in phase with one another. This, then, allows for the possibility of simultaneous starts. Moreover, the present account goes further: The distribution of lengths of silences ending with such simultaneous starts should show the same cyclic distribution as that shown in Wilson and Zimmerman's (1986) data for between-speaker silences in dyadic conversations.

Conclusions

The present account addresses a long-neglected void in our understanding of turn-taking, grounding the timing phenomena observed by conversation analysts in a theory of the cognitive processes of the individual participants. The phenomena to be accounted for include intriguing details, such as latches, relative absence of simultaneous starts in two-party conversations, and cyclic patterning of the probability of breaking a silence. We have offered here an account of these phenomena based on oscillatory representation of syllable rate and counterphased mutual entrainment between speaker and listener.

The model described here is intended as a more detailed specification of that proposed by Sacks et al. (1974). Specifically, it offers a mechanism to explain the second and third options in Sacks et al.'s model and, in particular, their sequential and nonoverlapping nature. In addition, we suggest that it is compatible with proposals offered in the literature on the temporal organization of phonemes and syllables and that one of the strengths of the present account is its potential to be integrated into a more general theory of speech production and speech perception. We note the further possibility that other activities that are complex, rhythmic, and interpersonally coordinated—notably, music performance—may be governed by similar principles (see Jungers et al., 2002; Large & Jones, 1999; Large & Palmer, 2002).

Finally, it is important to note that although this model offers a mechanistic account of how timing is coordinated between conversational partners, this does not imply a simplistically determinist account of when the next speakers will choose to begin speaking. Instead, speakers and listeners bring to bear a variety of higher order cognitive and motivational considerations in deciding when to speak. Thus, how soon a self-selecting next speaker moves to take the turn may depend on interactional concerns, such as the pragmatic implications of a silence, the degree of competitiveness in the tone of the conversation, the speaker's confidence in what he or she has to say, and the relative social status of the participants. What *is* mechanistically governed, we claim, is the precise timing of the speech onset, at the moment that the speaker does begin to speak.

REFERENCES

Ackermann, H., Gräber, S., Hertrich, I., & Daum, I. (1997). Categorical speech perception in cerebellar disorders. *Brain & Language*, **60**, 323-331.

Ackermann, H., Mathiak, K., & Ivry, R. B. (2004). Temporal organization of "internal speech" as a basis for cerebellar modulation of cognitive functions. *Behavioral & Cognitive Neuroscience Reviews*, **3**, 14-22.

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, **113**, 544-552.

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality & Social Psychology Bulletin*, **21**, 394-405.

Beattie, G. W. (1979). Contextual constraints on the floor-apportionment function of speaker-gaze in dyadic conversations. *British Journal of Social & Clinical Psychology*, **18**, 391-392.

Beattie, G. W., Cutler, A., & Pearson, M. (1982). Why is Mrs. Thatcher interrupted so often? *Nature*, **300**, 744-747.

Beebe, B., Alson, D., Jaffe, J., Feldstein, S., & Crown, C. (1988). Vocal congruence in mother–infant play. *Journal of Psycholinguistic Research*, **17**, 245-259.

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language*, **14**, 211-227.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, **107**, 127-183.

Burgess, N., & Hitch, G. (1992). Towards a network model of the articulatory loop. *Journal of Memory & Language*, **31**, 429-460.

Burgess, N., & Hitch, G. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, **106**, 551-581.

Burle, B., & Bonnet, M. (1999). What's an internal clock for? From temporal information processing to temporal processing of information. *Behavioral Processes*, **45**, 59-72.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, **304**, 1926-1929.

Caspers, J. (1998). Who's next? The melodic marking of question vs. continuation in Dutch. *Language & Speech*, **41**, 375-398.

Caspers, J. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics*, **31**, 251-276.
Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, **84**, 73-111.
Coates, J. (1994). No gap, lots of overlap: Turn-taking patterns in the talk of women friends. In D. Graddol, J. Maybin, & B. Stierer (Eds.), *Researching language and literacy in social context* (pp. 177-192). Philadelphia: Multilingual Matters.
Coates, J. (1997). The construction of a collaborative floor in women's friendly talk. In T. Givón (Ed.), *Conversation: Cognitive, communicative and social perspectives* (pp. 55-89). Philadelphia: Benjamins.
Coates, J., & Sutton-Spence, R. (2001). Turn-taking patterns in deaf conversation. *Journal of Sociolinguistics*, **5**, 507-529.
Craig, H. K., & Gallagher, T. M. (1982). Gaze and proximity as turn regulators within three-party and two-party child conversations. *Journal of Speech & Hearing Research*, **25**, 65-75.
Crown, C. L., Feldstein, S., Jasnow, M. D., Beebe, B., & Jaffe, J. (2002). The cross-modal coordination of interpersonal timing: Six-week-olds infants' gaze with adults' vocal behavior. *Journal of Psycholinguistic Research*, **31**, 1-23.
Cutler, A. (1991). Linguistic rhythm and speech segmentation. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 157-166). London: Macmillan.
Dittmann, A. T., & Llewellyn, L. G. (1968). Relationship between vocalizations and head nods as listener responses. *Journal of Personality & Social Psychology*, **9**, 79-84.
Duncan, S., & Niederehe, G. (1974). On signaling that it's your turn to speak. *Journal of Experimental Social Psychology*, **10**, 234-247.
D'Urso, V., & Zammuner, V. (1990). The perception of pause in question–answer pairs. *Bulletin of the Psychonomic Society*, **28**, 41-43.
Elias, G., Hayes, A., & Broerse, J. (1986). Maternal control of co-vocalization and inter-speaker silences in mother–infant vocal engagements. *Journal of Child Psychology & Psychiatry*, **27**, 409-415.
Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 134-184). New York: Cambridge University Press.
Fox Tree, J. E. (2000). Coordinating spontaneous talk. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 375-406). Philadelphia: Taylor & Francis.
Fox Tree, J. E. (2002). Interpreting pauses and *um*s at turn exchanges. *Discourse Processes*, **34**, 37-55.
French, P., & Local, J. (1983). Turn-competitive incomings. *Journal of Pragmatics*, **7**, 17-38.
Hafez, O. M. (1991). Turn-taking in Egyptian Arabic: Spontaneous speech vs. drama dialogue. *Journal of Pragmatics*, **15**, 59-81.
Harrigan, J. A. (1985). Listeners' body movements and speaking turns. *Communication Research*, **12**, 233-250.
Harris, H. D. (2002). Holographic reduced representations for oscillator recall: A model of phonological production. In W. D. Gray & C. D. Schunn (Eds.), *The proceedings of the 24th Annual Meeting of the Cognitive Science Society* (pp. 423-428). Hillsdale, NJ: Erlbaum.
Hartley, T. (2002). Syllabic phase: A bottom-up representation of the structure of speech. In J. A. Bullinaria & W. Lowe (Eds.), *Connectionist models of cognition and perception: Proceedings of the Seventh Neural Computation and Psychology Workshop* (pp. 277-288). Singapore: World Scientific.
Hartley, T., & Houghton, G. (1996). A linguistically constrained model of auditory verbal short-term memory. *Journal of Memory & Language*, **35**, 1-31.
Heritage, J. (1984). *Garfinkel and ethnomethodology.* Cambridge: Polity.
Holmlund, C. (1995). Development of turntakings as a sensorimotor process in the first 3 months: A sequential analysis. In K. E. Nelson & Z. Reger (Eds.), *Children's language* (Vol. 8, pp. 41-64). Hillsdale, NJ: Erlbaum.
Ivry, R. B., & Gopal, H. S. (1992). Speech production and perception in patients with cerebellar lesions. In D. E. Meyer & S. Kornblum (Eds.), *Synergies in experimental psychology, artificial intelligence and cognitive neuroscience* (pp. 771-802). Hillsdale, NJ: Erlbaum.
Izdebski, K., & Shipp, T. (1978). Minimal reaction times for phonatory initiation. *Journal of Speech & Hearing Research*, **21**, 638-651.
Jasnow, M., & Feldstein, S. (1986). Adult-like temporal characteristics of mother–infant vocal interactions. *Child Development*, **57**, 754-761.
Jefferson, G. (1973). A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica*, **9**, 47-96.
Jefferson, G. (1984). Notes on some orderliness of overlap onset. In V. D'Urso & P. Leonardi (Eds.), *Discourse analysis and natural rhetorics* (pp. 11-38). Padova: CLEUP.
Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a "standard maximum" silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation: An interdisciplinary perspective* (pp. 166-196). Clevedon: Multilingual Matters.
Jungers, M. K., Palmer, C., & Speer, S. R. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing*, **1-2**, 21-35.
Keller, H., Schölmerich, A., & Eibl-Eibesfeldt, I. (1988). Communication patterns in adult–infant interactions in Western and non-Western cultures. *Journal of Cross-Cultural Psychology*, **19**, 427-445.
Kent, R. D., Kent, J. F., Rosenbek, J. C., Vorperian, H. K., & Weismer, G. (1997). A speaking task analysis of the dysarthria in cerebellar disease. *Folia Phoniatrica et Logopædica*, **49**, 63-82.
Kjaerbeck, S. (1998). The organization of discourse units in Mexican and Danish business negotiations. *Journal of Pragmatics*, **30**, 347-362.
Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., & Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language & Speech*, **41**, 295-321.
Körner, E., Gewaltig, M.-O., Körner, U., Richter, A., & Rodemann, T. (1999). A model of computation in neocortical architecture. *Neural Networks*, **12**, 989-1005.
Kuriki, S., Mori, T., & Hirata, Y. (1999). Motor planning center for speech articulation in the normal human brain. *NeuroReport*, **10**, 765-769.
La France, M. (1974). Nonverbal cues to conversational turn taking between black speakers. *Personality & Social Psychology Bulletin*, **1**, 240-242.
Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, **106**, 119-159.
Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, **26**, 1-37.
Lerner, G. H. (1991). On the syntax of sentences-in-progress. *Language in Society*, **20**, 441-458.
Lerner, G. H. (1996). On the "semi-permeable" character of grammatical units in conversation: Conditional entry into the turn space of another speaker. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 238-276). Cambridge: Cambridge University Press.
Lerner, G. H. (2002). Turn-sharing: The choral co-production of talk-in-interaction. In C. E. Ford, B. A. Fox, & S. A. Thompson (Eds.), *The language of turn and sequence* (pp. 225-256). New York: Oxford University Press.
Lerner, G. H., & Takagi, T. (1999). On the place of linguistic resources in the organization of talk-in-interaction: A co-investigation of English and Japanese grammatical practices. *Journal of Pragmatics*, **31**, 49-75.
Levinson, S. C. (1983). *Pragmatics*. New York: Cambridge University Press.
MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral & Brain Sciences*, **21**, 499-546.
MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (2000). The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development*, **71**, 153-163.
Masataka, N. (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- to four-month-old Japanese infants. *Journal of Child Language*, **20**, 303-312.
Masataka, N. (1995). The relation between index-finger extension and

the acoustic quality of cooing in three-month-old infants. *Journal of Child Language*, **22**, 247-257.

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, & Hearing Research*, **44**, 128-143.

Mesch, J. (2000). Tactile Swedish Sign Language: Turn taking in signed conversations of people who are deaf and blind. In M. Metzger (Ed.), *Bilingualism and identity in deaf communities* (pp. 187-203). Washington, DC: Gallaudet University Press.

Mesch, J. (2001). *Tactile sign language: Turn taking and questions in signed conversations of deaf-blind people*. Hamburg, Germany: Signum.

Miura, I. (1993). Switching pauses in adult–adult and child–child turn takings: An initial study. *Journal of Psycholinguistic Research*, **22**, 383-395.

Murata, K. (1994). Intrusive or co-operative? A cross-cultural study of interruption. *Journal of Pragmatics*, **21**, 385-400.

Penttonen, M., & Buzsáki, G. (2003). Natural logarithmic relationship between brain oscillators. *Thalamus & Related Systems*, **2**, 145-152.

Povel, D.-J. (1981). Internal representations of simple temporal patterns. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 3-18.

Povel, D.-J., & Essens, P. J. (1985). Perception of temporal patterns. *Music Perception*, **2**, 411-440.

Robbins, O., Devoe, S., & Wiener, M. (1978). Social patterns of turn-taking: Nonverbal regulators. *Journal of Communication*, **28**, 38-46.

Rutter, D. R., & Durkin, K. (1987). Turn-taking in mother–infant interaction: An examination of vocalizations and gaze. *Developmental Psychology*, **23**, 54-61.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, **50**, 696-735.

Schaffer, D. (1983). The role of intonation as a cue to turn taking in conversation. *Journal of Phonetics*, **11**, 243-257.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 71-93). Washington, DC: Georgetown University Press.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, **29**, 1-63.

Sellen, A. J. (1995). Remote conversations: The effect of mediating talk with technology. *Human–Computer Interaction*, **10**, 401-444.

Shea, S. A., Walter, J., Pelley, C., Murphy, K., & Guz, A. (1987). The effect of visual and auditory stimuli upon resting ventilation in man. *Respiration Physiology*, **68**, 345-357.

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 326-332.

Sidnell, J. (2001). Conversational turn-taking in a Caribbean English Creole. *Journal of Pragmatics*, **33**, 1263-1290.

Stephens, J., & Beattie, G. (1986a). On judging the ends of speaker turns in conversation. *Journal of Language & Social Psychology*, **5**, 119-134.

Stephens, J., & Beattie, G. (1986b). Turn-taking on the telephone: Textual features which distinguish turn-final and turn-medial utterances. *Journal of Language & Social Psychology*, **5**, 211-222.

Stevenson, M. B., Ver Hoeve, J. N., Roach, M. A., & Leavitt, L. A. (1986). The beginnings of conversation: Early patterns of mother–infant vocal responsiveness. *Infant Behavior & Development*, **9**, 423-440.

Streeck, J. (1996). A little Ilokano grammar as it appears in interaction. *Journal of Pragmatics*, **26**, 189-213.

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, **11**, 139-169.

Strogatz, S. (2003). *Sync: The emerging science of spontaneous order*. New York: Hyperion.

Tanaka, H. (2000a). The particle *ne* as a turn-management device in Japanese conversation. *Journal of Pragmatics*, **32**, 1135-1176.

Tanaka, H. (2000b). Turn projection in Japanese talk-in-interaction. *Research on Language & Social Interaction*, **33**, 1-38.

Thomas, A. P., & Bull, P. (1981). The role of pre-speech posture change in dyadic interaction. *British Journal of Social Psychology*, **20**, 105-111.

Trimboli, C., & Walker, M. B. (1984). Switching pauses in cooperative and competitive conversations. *Journal of Experimental Social Psychology*, **20**, 297-311.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, **41**, 101-175.

Walker, M. B. (1982). Smooth transitions in conversational turn-taking: Implications for theory. *Journal of Psychology*, **110**, 31-37.

Walker, M. B., & Trimboli, C. (1984). The role of nonverbal signals in co-ordinating speaking turns. *Journal of Language & Social Psychology*, **3**, 257-272.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, **7**, 553-559.

Warner, N., & Arai, T. (2001). The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America*, **109**, 1144-1156.

Wayland, S. C., Miller, J. L., & Volaitis, L. E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America*, **5**, 2694-2701.

Wells, B., & MacFarlane, S. (1998). Prosody as an interactional resource: Turn-projection and overlap. *Language & Speech*, **41**, 265-294.

Wilson, T. P., Wiemann, J. M., & Zimmerman, D. H. (1984). Models of turn-taking in conversational interaction. *Journal of Language & Social Psychology*, **3**, 159-180.

Wilson, T. P., & Zimmerman, D. H. (1986). The structure of silence between turns in two-party conversation. *Discourse Processes*, **9**, 375-390.

(Manuscript received August 6, 2004;
revision accepted for publication May 17, 2005.)