



10-1-2015

# Investigating an Acoustic Measure of Perceived Isochrony in Conversation: Preliminary Notes on the Role of Rhythm in Turn Transitions

Shannon Mooney

Grace C. Sullivan

---

# Investigating an Acoustic Measure of Perceived Isochrony in Conversation: Preliminary Notes on the Role of Rhythm in Turn Transitions

## **Abstract**

In a preliminary investigation of isochrony, the rhythmic integration of talk, we evaluated rhythmic phenomena previously theorized to coordinate turn-transitions for correlates in the acoustic signal. Rhythmic sequencing is one of many elaborate contextualization cues regarded as facilitating a successful turn-transition. Previous studies of rhythm in conversation have attended only to its perceptual and interactional facets. In addressing this gap, our study finds quantitative justification for such claims of rhythmic turn-taking. We selected for acoustic analysis the twelve non-task-based, dyadic conversations of the Santa Barbara Corpus of Spoken American English (SBCSAE). Following Marcus's (1981) assertion that the onset of the vowel is the closest acoustically-measurable location to the perceptual center of the syllable where the rhythmic downbeat occurs, duration was measured between vowel onsets to create prosodic syllables. Not all prosodic syllables can contain a rhythmic beat, and those that can are characterized as "prominent" in nature (Couper-Kuhlen 1993). Out of 42,807 prosodic syllables measured, our methods yielded 15,972 prominent prosodic syllables. The units of duration between prominent syllables, hereafter intervals, were judged to form an isochronous sequence when the durations between at least three consecutive intervals varied by less than the conservative measure of the perceptual threshold for tolerance of isochrony, up to a 30% variance (Couper-Kuhlen 1993). This measure revealed 564 rhythmic sequences across the twelve SBCSAE conversations, which ranged in duration between one and ten seconds and consisted of up to eleven intervals. Of these, 208 or 37% appeared within turn-transitions, and results from our preliminary analysis indicated that rhythmic sequencing was significantly more likely to appear within a turn-transition than outside of one. Our analysis shows that isochrony is not simply perceptual in nature, but that it has a quantifiable correlate in the acoustic signal. Our findings of significant isochrony in the turn-transitions of the SBCSAE, a corpus often used in discourse analysis, confirms what many interactional sociolinguists have long argued: that rhythmic cues aid the coordination of talk between speakers in turn-transitions. We can confirm that these rhythmic cues are a component of turn-transitions not only perceptually, but acoustically as well.

# Investigating an Acoustic Measure of Perceived Isochrony in Conversation: Preliminary Notes on the Role of Rhythm in Turn Transitions

Shannon Mooney and Grace C. Sullivan\*

## 1 Introduction

The possibility of a periodic signal in human speech that directly or indirectly leads to the rhythmic coordination of conversation by speakers has been pursued over the years by scholars across various subfields of linguistics with, as yet, no clear consensus. While the interest in rhythm as a conversational feature in the subfields of ethnomusicology and interactional sociolinguistics (e.g., Erickson 1982, Scollon 1982) has dwindled over the past twenty years, interest in neurological periodicity and its role in speech processing and production has only increased as more and more complex research tools and procedures have become widely available in the same amount of time. This paper revisits the conversational phenomenon of rhythmic integration and engages with claims made decades ago about its function in interaction. Our analysis relies on a quantitative methodology in order to investigate an assumption still maintained in some interactional sociolinguistic research: specifically, that rhythmic integration, termed *isochrony*, indeed underlies everyday talk (Erickson 1982, Erickson and Shultz 1982, Scollon 1982).

Our study assumes, following Couper-Kuhlen 1993, that isochrony can manifest either as the simultaneous synchronization of rhythmic beats between two or more speakers or as the maintenance of tempo across two or more speakers' turns. The paper will first introduce the interdisciplinary literature motivating the current study across the subfields of experimental phonetics, metrical phonology, discourse analysis, interactional sociolinguistics, and variationist sociolinguistics, then describe our data and methods for exploring the potential acoustic basis for the phenomenon of isochrony, and finally conclude with some preliminary results and indicate future areas for quantitative inquiry into the discursive significance that isochrony in its various forms may play.

## 2 Isochrony

### 2.1 Periodicity in Speech Processing

Speech rhythm is, in its most basic form, conceived of as “hierarchically temporally coordinated prosodic units” (Cummins and Port 1998). Despite this, the idea of rhythmic organization of language is traditionally only thought of in phonological theory in terms of classifying languages into different rhythmic classes (Abercrombie 1967). Until recently, there have been very few attempts to analyze the hierarchical, prosodic, and sometimes pragmatic structure and functioning of temporal coordination of prosodic syllables, but in the past few years researchers have begun to question the notion that rhythmic class is a clear-cut, basically binary system. The most commonly used rhythmic classes are *stress-timed* and *syllable-timed* languages. A prototypical stress-timed language is English; it has a wide variety of syllable structure types, a high frequency of complex syllable structure types, vowel reduction in unstressed positions, and final lengthening (Wightman et al. 1992). Spanish, on the other hand, is typically classified as a syllable-timed language, meaning that it has predominantly CV syllable structure with low syllable complexity and less vowel reduction and final lengthening than English. Despite this characterization, there are dialects within English that exhibit what is sometimes referred to as syllable timing (Coggshall 2008), so that it is not at all clear that this classification into rhythmic classes is straightforward or holds across an entire language. What recent work has argued is that the classification into rhythmic classes may be decomposed into different strategies for rhythmicity in conversation, what we call *isochrony* in the current paper. Recent work in cognitive psychology suggests a large role for periodicity in the perception and processing of sound (Ghitza 2011, Large 2010, Goswami and Leong 2013). There

---

\*We are grateful to Heidi Hamilton, Jen Nycz, Cynthia Gordon, Fred Erickson, Youngah Do, and Mark Sicoli in the Department of Linguistics at Georgetown University for their insights. We also received many helpful comments from our colleagues attending *NWAV 43* in Chicago, where this poster was presented.

is much ongoing debate about whether this periodicity, especially as it applies to linguistic phenomena, is “real”, e.g., present in the acoustic signal, or only perceived, e.g., aperiodic utterances are processed in a way that somehow involves their reinterpretation as periodic. This debate recently occupied an entire issue of the journal *Laboratory Phonology* (2013). If rhythm is indeed “real” and present in the acoustic signal, what is it about speech production that leads utterances to be characterized by periodicity?

An interesting proposal by Cummins and Port (1998) is that the rhythmic systematicity of a language is the result of “entrainment between metrical levels” (which has also been shown to underlie other human and animal behaviors outside of language). What is meant by this is that the hierarchical structure is indeed present in the functioning and production of rhythm, rather than simply existing as a tool to analyze the present of rhythm and to classify languages according to their prosodic properties. A theory such as this relies on Liberman’s (1975) Metrical Organizational Hypothesis, which assumes that languages are prosodically structured through different metrical tiers, both abstractly, as in within syllabic and suprasegmental structures, and concretely, as in how language is produced by its speakers. Work which followed up on this theory, such as Liberman and Prince 1977 and Hayes and Puppel 1985, refined it to lesser and greater extents. In particular, Hayes and Puppel 1985 found no real phonetic basis for the presence of an underlying metrical structure, and thus argue that although it may belong to the abstract structuring and organization of a language into units, it is not replicated in production of the language by its speakers. Interestingly, many recent experiments and quantitative analyses have indeed shown evidence for a real phonetic basis to the functioning of rhythmic organization of language. Even very early work within metrical phonology has shown the existence in English of roughly temporally equal intervals between stressed prosodic syllables (Jones 1918, Pike 1945, Abercrombie 1967). This, combined with the largely uncontested presence of a metrical hierarchy which organizes syllabic and suprasegmental structure abstractly in languages (though the structure varies cross-linguistically, as discussed in the previous paragraphs), points to the potential role of rhythm in the organization of language as it is spoken and perceived.

Abercrombie 1967 typifies English as a stress-timed language. Despite this work and other early studies (such as Pike 1945) that divide languages into classes based on how they behave “rhythmically”, it was originally unclear what predictions would follow from this classification for each of the languages studied. Much of the debate on the existence of language-specific differences in rhythmic patterning that has occupied the past few years falls out from this early work. In the current study, we find acoustic evidence for the existence of periodic intervals in a corpus of American English conversations. Specifically, we find that rhythmically coordinated speech tends to occur during the transition between one speaker and another, and immediately following speech errors. This points to the potential role of speech timing as aiding perception and benefiting the communicative function of speech, but left unresolved is the question of whether there is anything inherently rhythmic about the arrangement of the segmental and suprasegmental (i.e., prosodic) phonology of English and other languages.

## 2.2 The Interactional Value of Isochrony

An important aspect of the analysis of isochrony lies in the notion of contextualization. Contextualization cues are linguistic features, such as prosody, lexical choice, code and paralinguistic signs that are considered to signal to the speakers in the interaction how the conversation is going (Gumperz 1977). Gumperz himself considered conversational synchrony to be one such paralinguistic feature that may act as a contextualization cue. The subsequent literature on isochrony has long maintained this to be a contextualization cue signaling that the conversation is “working” and the participants are comfortable in the interaction (Couper-Kuhlen 1993). In the interactional sociolinguistic field, scholars find that there is a relationship between rhythm and cohesion. Scollon (1981:342) considers rhythm in conversation to be “used as the means of negotiating the interaction between speakers.” He finds that conversational partners will utilize linguistic features (e.g., stressed syllables, coughs, laughs, sneezes) to establish what he calls a “rhythmic ensemble”. Similarly, Erickson (1982:65) adopts this notion and states, “rhythm seems to be the fundamental social glue by which cohesive discourse is maintained in conversation.” He continues to state that through rhythm, interactional partners can work together in the conversation.

The work of Erickson, Shultz, and Scollon have examined isochrony as it relates to the musicality of conversation and how speakers work in a sort of *ensemble* to unconsciously signal to each other that the conversation is going well. This analysis illustrates an acoustic basis for the phenomenon of isochrony and rhythmic integration in conversation and is motivated by this seminal research. The idea that isochrony acts as a contextualization cue to speakers is supported in the literature from the acoustic phonetic perspective as well.

### 3 Data and Methods

#### 3.1 Data

In order to carry out an analysis of the possibility of rhythmic coordination between speakers in conversation, we selected for the current study a subset of conversations from the *Santa Barbara Corpus of Spoken American English* (DuBois et al. 2000–2005), which we will refer to in this paper by its acronym, *SBCSAE*. The sound files and transcripts which make up this corpus were extremely valuable to our purposes, as they offered us the unique ability to carry out our quantitative acoustic methodology on a dataset that is widely known and studied among discourse analysts and interactional sociolinguists. In this sense, the corpus itself enabled us to directly address claims of a discursive function of isochrony made by the aforementioned research in sociolinguistics.

The *SBCSAE* is comprised entirely of naturally occurring recorded conversations between two or more participants. Some conversations in the *SBCSAE* are task-based or occur in a specific institutional context, while others spontaneously occur between friends and family. It was important to us to limit ourselves to recordings of conversations, in lieu of the traditional sociolinguistic interview, so that we could avoid any interviewer effects of performativity or more general monologic effects that may be present to a greater or lesser extent in other types of recordings that have previously been used for variationist inquiry into the rhythmic patterning of speech. For consistency, we selected only dyadic conversations that were naturally occurring and not task-based. This resulted in a subset of twelve recordings, each approximately thirty minutes in duration, which comprised an exhaustive list of *SBCSAE* conversations that are not task-based and take place between only two participants.

<i>SBCSAE</i> File number	<i>SBCSAE</i> Name	Participants	Length
SBC005	<i>A Book About Death</i>	couple	20:28
SBC006	<i>Cuz</i>	female cousins	27:16
SBC007	<i>A Tree's Life</i>	sisters	23:05
SBC017	<i>Wonderful Abstract Notions</i>	male friends	20:18
SBC028	<i>Hey Cutie Pie</i>	couple	25:17
SBC034	<i>What Time Is It Now?</i>	couple	24:40
SBC043	<i>Try A Couple Spoonfuls</i>	mother and daughter	25:02
SBC044	<i>He Knows</i>	male friends	29:07
SBC045	<i>The Classic Hooker</i>	friends	30:15
SBC047	<i>On The Lot</i>	male cousins	20:13
SBC058	<i>Swingin' Kid</i>	mother and son	25:47
SBC060	<i>Shaggy Dog Story</i>	male friends	24:50

Table 1: The twelve non task-based and dyadic conversations of the *SBCSAE* that were selected for analysis in the current study.

#### 3.2 Methods

We aligned each of the twelve conversation transcripts to their respective sound files using FAVE, and then extracted every vowel, preserving the timestamp of each. Taking the discussion provided by Couper-Kuhlen (1993) into account in determining our criteria for prosodically prominent syllables that can potentially hold rhythmic beats, we decided that for this preliminary investigation we would assume, as she does, that the most direct correlate of prominence in a prosodic syllable

was a long duration. To be consistent with previous methodologies concerning prosodic syllables (e.g., Marcus 1981), we defined the prosodic syllable as occurring between vowel onsets, which yielded 64,314 prosodic syllables in all.

We implemented the somewhat arbitrary designation of prosodic prominence to syllables with a duration value greater than the overall mean syllable duration of the conversation as a whole, so as to broadly account for speech rate. In this way, we narrowed our dataset down to 23,292 prosodically prominent syllables, with the intervals between the onsets of prominent syllables being potential candidates for rhythmic synchronization with the closest neighboring intervals. Again relying on observations cited in Couper-Kuhlen 1993, we assumed perceptible isochrony within the conservative threshold of 30% variance between exact interval durations and defined a rhythmic sequence as the presence of at least three consecutive intervals with variance in duration of less than 30%. This strategy resulted in 2,369 rhythmic intervals with a mean duration of 82 milliseconds making up 776 rhythmic sequences.

We then analyzed these sequences for participation in conversational contextualization and coordination, focusing specifically on co-occurrence with a transition of the speaker floor from one conversant to the other, which we refer to as a turn transition after Sacks et al. 1974.

#### 4 Results and Analysis

Upon analysis, we found that while isochrony is relatively rare overall in the conversations under the current methodological assumptions, when an isochronous sequence does occur, it is significantly more likely to be during a turn transition between speakers. This positive correlation between a turn transition and isochrony is significant under the non-parametric Kendall's tau test both overall in the dataset and in ten out of the twelve conversations individually.

<i>SBCSAE</i> Conversation	Proportion of rhythmic intervals	Rhythmic intervals ( <i>N</i> )	Proportion of rhythmic intervals in turns	Turns ( <i>N</i> )	Kendall's tau
005	0.11	1,331	0.41	249	.17***
006	0.13	2,384	0.14	195	.05**
007	0.08	1,462	0.20	219	.06*
017	0.13	1,951	0.39	319	.23***
028	0.12	1,973	0.44	446	.17***
034	0.10	1,106	0.36	170	.17***
043	0.09	2,710	[0.18]	370	[.03]
044	0.12	2,840	0.18	205	.12***
045	0.11	1,951	0.30	230	.17***
047	0.14	2,104	0.40	381	.17***
058	0.09	1,624	0.31	335	.10***
060	0.12	1,844	[0.10]	116	[.02]
Total	0.11	23,280	0.28	3,236	.12***

Table 2: Analysis of Isochrony in *SBCSAE* Conversations with (1) Conversation identifying code, (2) Proportion of intervals across the conversation that are isochronous, (3) Number of rhythmic intervals across the conversation, (4) Proportion of rhythmic intervals that appear in speaker turns, (5) Number of turns across the conversation, and (6) Statistical significance of the proportion of rhythmic intervals that appear in turns.

Overall, we found that turn transitions are significantly more likely to contain rhythmically coordinated intervals than non-turn utterances. We consider these findings to suggest a functional role of isochrony in the transfer of floor between speakers. Upon further investigation, we found that these rhythmically integrated sequences may also be working as a contextualization cue, signaling to the speakers that there is cohesion in the conversation.

Consider two instances of isochronous sequences, in particular. Figure 1 shows rhythmic in-

tegration between speakers during backchannelling, or listener-responses that signal to the floorholder that they comprehend what is being said and are allowing the speaker to continue their turn (Yngve 1970). This image illustrates the rhythmic beats on the point tier in Praat and shows that the rhythm was established by speaker Jim on the second syllable in the word *matheMATics* and this rhythm is met by Michael's backchannelled *yeh*. Jim's rhythm is maintained during the remainder of his utterance on syllables *so* and *know*.

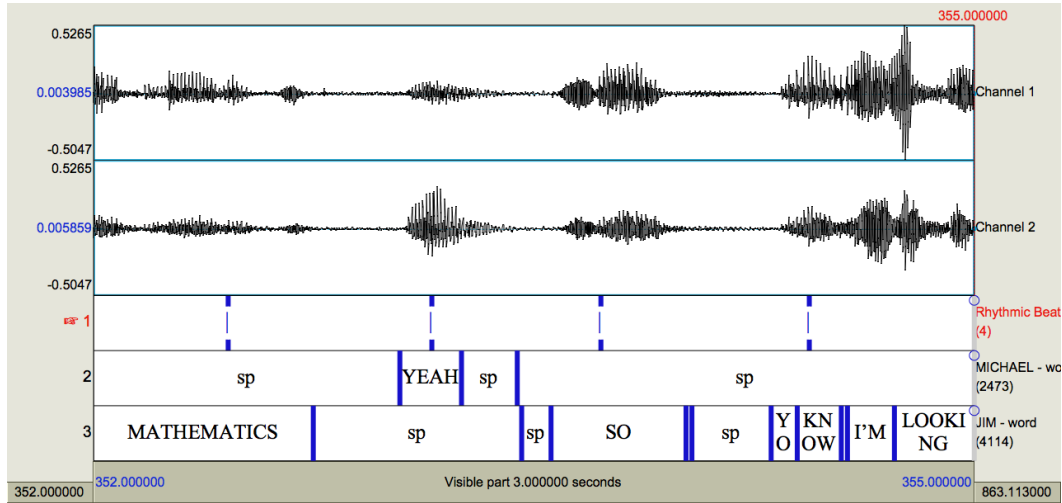


Figure 1: An Excerpt from *Wonderful Abstract Notions* (SBC017).

The second instance of note here (Figure 2) occurs during a change of floor that does not include an instance of backchannelling but rather during a point of collaborative sentence building. Similarly, the visualization of this image shows rhythmic beats on the point tier. This time, speaker Lenore asks Alina a question *Who does he work with now?* and initiates the rhythmic beat on the word *now* which is then followed by a prolonged pause. Alina answers with *Kevin*, which meets Lenore's established rhythm, and then maintains this rhythm in her subsequent turn.

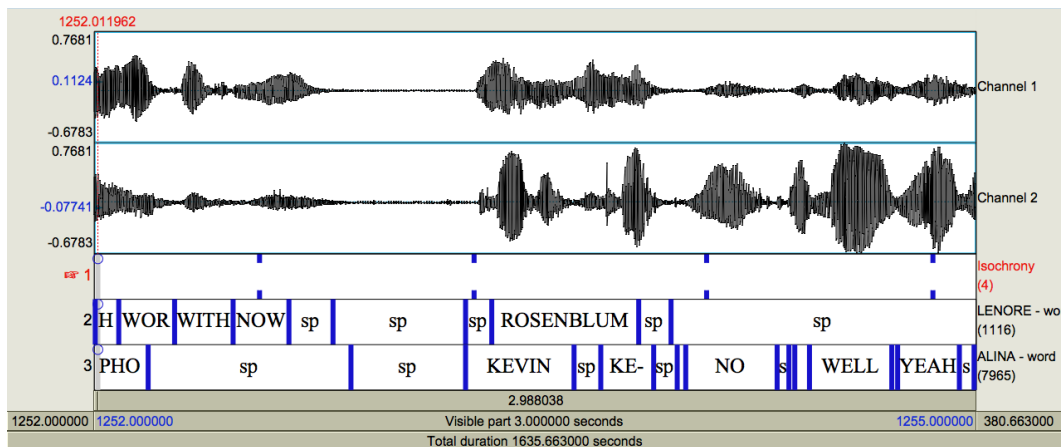


Figure 2: An Excerpt from *Cuz* (SBC006).

Our findings do not conform to the predictions made by sociolinguists in the seminal work on rhythmic integration in interaction as we found that isochrony is actually rare in conversation. However, this is not disheartening because our findings suggest there may be a more specialized interactional function of isochrony as we have defined it. The two examples above are illustrative of the interactional function of isochrony as they show that rhythmic integration can occur alongside a transfer of the speaker floor and may be acting as a contextualization cue to signal to the

speakers that there is cohesion in the conversation and that they are working together discursively.

## 5 Discussion and Suggestions for Future Research

According to Tannen 2005, conversational style plays an important role in the treatment of turn transitions. In Tannen 1981, the features that are found more frequently with high involvement speakers include pacing that is defined as (a) faster rate of speech, (b) avoidance of inter-turn pauses, (c) faster turn-taking, and (d) cooperative overlap and participatory listenership (p. 137). In our methodology, there was no way to determine an individual's typical conversational style, but our findings support the notion that rhythmic integration during turn transitions is a strategy found in oral discourse that signals involvement and participatory listenership, perhaps regardless of (or in addition to) conversational style. We argue that integration of isochrony over turn transitions is a clear indication of participatory listenership, which demonstrates that the listener is attentive and understanding of the speaker's perspective (Tannen 1986). Therefore, our findings of rhythmic integration over turn transitions further support the previous literature in various sub-fields of sociolinguistics that isochrony is a contextualization cue by suggesting that listeners utilize this rhythmic strategy to signal to the speaker that they are involved and attentive in the conversation.

This methodology was adapted for a subsequent analysis by Sullivan (2015) who examined the interactional function of rhythmic integration within narratives. Sullivan's (2015) findings conformed to our findings above that rhythmic integration may function as a contextualization cue signaling to the speaker that they are working as a team, or a conversational duet (Falk 1987). Our investigation into finding an acoustic basis for isochrony is preliminary, and as such we will continue examining this phenomenon. The main focus of further research will be to refine and formalize the methodology used in the current study, as well as to draw from more diverse theoretical motivating perspectives for the underlying phenomena of natural speech rhythms. Specifically, the methodology will be improved through further exploration into how to operationalize prosodic prominence of syllables. We continue to test different arguments for what makes a syllable prominent prosodically and will motivate our argumentation with subsequent research. Furthermore, it will be beneficial to perform a perception study, as this phenomenon is assumed to be dependent upon interaction, and as it is also assumed in the literature (Couper-Kuhlen 1993) that participants can perceive isochrony within a threshold of 30% variance between interval durations. Therefore, we will ultimately wish to examine whether isochrony over turn transitions relies on the interaction aspect of conversations and participatory listenership or if conversational isochrony results from other cognitive factors.

## 6 Conclusion

This analysis sought to re-examine the interactional phenomenon of isochrony, or rhythmic integration, with a focus on a quantifiable, replicable methodology. Motivated by a seminal work on rhythmic integration in acoustic phonetics and interactional sociolinguistics, we analyzed twelve conversations from the SBCSAE through the scope of sociophonetic and discourse analytic frameworks. We found that although rhythmic integration may not be as ubiquitous in conversation as previously thought, when it does occur, it may be during a discursively significant instance. This analysis opened up a bevy of potential research opportunities; the most pressing is the finalization of the methodology.

## References

- Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Coggshall, Elizabeth L. 2008. The prosodic rhythm of two varieties of native American English. In *U. Penn Working Papers in Linguistics 14.2: Selected Papers from NWAV 36*, ed. K. Gorman, 1–9.
- Couper-Kuhlen, Elizabeth. 1993. *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*. John Benjamins Publishing.



- Cummins, Fred, and Robert Port. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26:145–171.
- DuBois, John W., Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson, and Nii Martey. 2000–2005. *Santa Barbara Corpus of Spoken American English*. University of California, Santa Barbara Center for the Study of Discourse.
- Erickson, Frederick. 1982. Money tree, lasagna bush, salt and pepper: social construction of topical cohesion in a conversation among Italian-Americans. In *Analyzing Discourse: Text and Talk. Georgetown University Round Table on Languages and Linguistics 1981*, ed. D. Tannen, 43–70. Washington, DC: Georgetown University Press.
- Erickson, Frederick, and Jeffrey J. Shultz. 1982. *The Counselor as Gatekeeper: Social Interaction in Interviews*. Waltham, MA: Academic Press.
- Falk, Jane Lois. 1979. *The Duet as Conversational Process*. Doctoral dissertation, Princeton University.
- Ghitza, Oded. 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology* 2:130.
- Goswami, Usha, and Victoria Leong. 2013. Speech rhythm and temporal structure: converging perspectives. *Laboratory Phonology* 4:67–92.
- Gumperz, John J. 1977. Sociocultural knowledge in conversational inference. *Linguistics and Anthropology* 76:785–98.
- Hayes, Bruce, and Stanislaw Puppel. 1985. On the rhythm rule in Polish. In *Advances in Nonlinear Phonology*, ed. H. van der Hulst and N. Smith, 59–81. Dordrecht: Foris.
- Jones, Daniel. 1918. The use of experimental phonetics to the linguist. *Modern Language Teaching* 14:125–132.
- Large, Edward W. 2010. Neurodynamics of music. In *Music Perception*, eds. M. Riess Jones, R.R. Fay, and A.N. Popper, 201–231. New York: Springer.
- Liberman, Mark Yoffe. 1975. *The Intonational System of English*. Doctoral dissertation, Massachusetts Institute of Technology.
- Liberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry*, 8:249–336.
- Marcus, Stephen Michael. 1981. Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics* 30:247–256.
- Pike, Kenneth L. 1945. *The Intonation of American English*. Ann Arbor, MI: University of Michigan Press.
- Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50:696–735.
- Scollon, Ron. 1982. The rhythmic integration of ordinary talk. In *Analyzing Discourse Text and Talk. Georgetown University Round Table on Languages and Linguistics 1981*, ed. D. Tannen, 335–349. Washington, DC: Georgetown University Press.
- Sullivan, Grace C. 2015. Alignment as a "with": an investigation on rhythmic integration in narratives. Paper presented at the 21<sup>st</sup> Annual Meeting of Language, Interaction, and Social Organization, Santa Barbara, California.
- Tannen, Deborah. 1981. *Analyzing Discourse: Text and Talk*. Georgetown University Press, Washington, DC.
- Tannen, Deborah. 1986. Introducing constructed dialogue in Greek and American conversational and literary narrative. *Direct and Indirect Speech* 3:11–32.
- Tannen, Deborah. 2005. *Conversational Style: Analyzing Talk among Friends*. New York: Oxford University Press.
- Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America* 91:1707–1717.
- Yngve, Victor H. 1970. On getting a word in edgewise. Paper presented at the 6<sup>th</sup> Annual Meeting of the Chicago Linguistics Society, University of Chicago.

Department of Linguistics  
 Georgetown University  
 Washington, DC 20057  
 sm2842@georgetown.edu  
 gcs39@georgetown.edu