

Identifying and Replicating Aspects of Prosody in Oral Reading for English Language Learners

Colin MACLEOD & Susana SANCHEZ

Abstract

Oral reading fluency is required for successful reading comprehension, as the ability to read accurately, automatically and with prosody is a signal that the reader is able to focus on the meaning of the text. Even though methodologies to identify and measure accuracy and automaticity are being successfully implemented, prosody presents a different challenge. For that reason, this paper proposes a way to identify, replicate, and measure the elements of prosody with Japanese subjects studying English as a second language.

1. INTRODUCTION

Skilled readers exhibit ‘fluency’, which is the ability to read accurately, automatically and with prosody. Accuracy in reading requires phonological word decoding. This is a bottom-up process where readers decode small units of language, such as letters, words, clauses, and phrases in order to understand what they mean, before then fitting them into the larger text (Gough, 1972). By contrast, automaticity, or rapid word recognition, refers to the properties of speed, effortlessness, autonomy and a lack of conscious awareness when reading; these properties develop in a continuum, but can also develop at different rates (Logan, 1997). Finally, prosody refers to the patterns of pause, stress and intonation in a language. Reading with prosody supports meaning, as appropriate phrasing, intonation and stress are indicators of reading fluency. Moreover, prosody is also thought to signal comprehension, as the ability to group phrases syntactically can be an indication of the reader’s understanding of the text they are reading (Kuhn and Stahl, 2003).

In assessing fluency, there are five critical areas in literacy development: phonemic awareness, phonics, articulation, vocabulary and comprehension. Vocabulary and comprehension are more complex and difficult to measure, as tests do not highlight difficulties in word-decoding or spelling that are related to the reader’s lack of vocabulary. Thus, efforts to objectively measure fluency have focused on accuracy and automaticity, and do

not assess prosody. For example, in the Curriculum-Based Measurements (CBMs) and Dynamic Indicator of Basic Early Literacy Skills (DIBELS), fluency is measured by counting the number of correct and incorrect words read aloud in one minute (Kuhn et al., 2010). This measure of fluency limits the assessment of fluency to reading speed, excluding important features of prosody.

2. PROSODY

Prosody is regarded as the music of the language and is concerned with the changes in pitch and rhythm, the use of stress and pauses, and the lengthening of words and phrases. For instance, prosodic patterns can convey emotions: fast speech rate, rising pitch and intonation variability can denote happiness (Hirschberg, 2002). Prosody can also carry discourse information as higher and variable pitch tones with longer pauses are interpreted as having greater importance within the discourse hierarchy (Wennerstrom, 2001). Intonation shows that the reader is assigning a syntactic role to words, and it is thought to be a key component in micro-processing text information (Kintsch, 1998). Prosody also provides a cognitive skeleton to hold auditory sequences in working memory until a more complete semantic analysis can be carried out (Swets et al., 2007; Koriat et al., 2002). An example of this is that readers have a better memory of a poem when it has enhanced prosodic features (Goldman et al., 2006). However, prosodic cues are not always well dictated by punctuation, as oral speech usually contains more pauses than indicated by the punctuation in a text. Thus, when reading aloud one has to abstract prosodic features to a greater extent than when reading in silence (Chafe, 1988).

2.1 National Assessment of Educational Progress (NAEP)

Fluency scales do exist that assess prosody in oral reading, such as the National Assessment of Educational Progress (NAEP) (see table 1). The NAEP assessment of oral reading considers that accuracy, automaticity, and prosody are related not only to one another, but also to overall reading comprehension. However, the assessment lacks precision as it involves qualitative judgment. Moreover, the measure of prosody in this fluency scale is far less detailed than the measure of accuracy and automaticity in the CBMs or DIBELS scales.

The scale was first developed in 1992 and is applied holistically, rather than analytically; the examiner categorizes each student's individual oral reading into one of the four level descriptions that best represents their overall performance. The scale evaluates prosody according to the following: phrasing, stress, intonation and pauses; syntax and structure to correctly transmit the ideas expressed in the text; and expressiveness, anticipation, and level of characterization.

Table 1. NAEP Oral Reading Fluency Scale (2005)

Fluent	Level 4	Reads primarily in larger, meaningful phrase groups. Although some regressions, repetitions, and deviations from text may be present, these do not appear to detract from the overall structure of the story. Preservation of the author’s syntax is consistent. Some or most of the story is read with expressive interpretation.
	Level 3	Reads primarily in three-or four-word phrase groups. Some small groupings may be present. However, the majority of phrasing seems appropriate and preserves the syntax of the author. Little or no expressive interpretation is present.
Non-fluent	Level 2	Reads primarily in two-word phrases with some three-or four- groupings. Some word-by-word reading may be present. Word groupings may seem awkward and unrelated to larger context of sentence or passage.
	Level 1	Reads primarily word-by-word. Occasional two-word or three-word phrases may occur – but these are infrequent and/or do not preserve meaningful syntax.

However, the NAEP Oral Fluency Scale can be perceived as an unsatisfactory measure of reading fluency for both examiners and students. For examiners, the scale lacks precision as it is split into four rudimentary levels. This over-simplistic scale does not adequately detail the strengths and weaknesses of the student’s levels of accuracy, automaticity or prosody, as it simply grades the student’s overall performance according to one of the four level bands. From the student’s perspective, the grading system can be uninformative and frustrating. For instance, a student graded at the lower end of Level 2 could make great strides in intonation over the following weeks, but fail to move up into the next level. In summary, the NAEP Oral Fluency Scale does not provide a detailed breakdown of which aspects of fluency the student lacks or has improved, making reflection on errors and efforts to improve difficult for both examiners and students. Moreover, the NAEP scale is unfit to objectively measure reading fluency, unlike the CBMs and DIBELS, which measure accuracy and automaticity more precisely. With this in mind, this paper introduces 1) a methodology for identifying whether students can recognize and accurately replicate the different aspects of prosody, and 2) how to break down these aspects more objectively.

3. METHODOLOGY TO IDENTIFY PROSODY

A group of 20 Japanese students (22-28 years old, 13 female), enrolled in an Academic English pre-session course to study a Master’s degree in the UK, were given a transcript of one story from the BBC radio program

'From Our Own Correspondent' (BBC, 2014). This radio program is a compilation of short, topical news reports from around the world read by locally-based BBC correspondents. The audio is available to download as a podcast or mp3 from the *From Our Own Correspondent* section of the BBC website, where a corresponding transcript of the text is also available (the text on the program's website actually differed slightly from what was broadcast, so the transcript was edited in order to represent the audio as closely as possible). The selected piece of audio was entitled "Beijing smog: When growth trumps life in China". The audio and transcript were then uploaded to a folder on the file-sharing website, Dropbox (2014), so that the material could be easily distributed to the students in two different stages, which will now be described.

3.1 Recording and Publishing

Students were asked to register with Audioboo (since re-named 'Audioboom'), a global, on-demand audio and podcasting distribution platform. The platform can be thought of as Twitter for audio. In Audioboo, each student created an account, which they used to 'follow' the class account (set up by the instructor). Audioboo can be used on tablets, smartphones, PCs and Macs and enables users to create and share audio files of up to ten minutes in length. For English language learners, Audioboo provides a space where they can practice language and pronunciation skills and then share their efforts with their instructor and peers and receive feedback.

Once registered, students were given a two-page, double-spaced printout of the transcript of the story being used. The first paragraph of the transcript, and the corresponding audio, were used to teach the students about prosodic features that this exercise would work on: stress, pause, and intonation. Students were taught to use different annotation marks for each of these features (see table 2) and were asked to consider how these features could be used in the opening paragraph. Once the students had made their predictions by annotating their transcripts, the students then listened to the audio of the native speaker and then annotated a fresh copy of the transcript in order to compare the native speaker's speech with their own initial predictions. To learn how to read with prosody, it was essential for the students to identify gaps between their predictions and how the text was actually read by the native speaker. The students were then encouraged to mimic the native speaker by reading the paragraph out loud.

Following the class introduction to the method, the students were allowed to take the transcript home and familiarize themselves with its contents. Once they felt comfortable with the text and had considered how they thought the text should be read, they were asked to record themselves reading it out loud using Audioboo.

Students were instructed that this should be done in a place and at a time where they could comfortably read out loud for three minutes without interruption or distraction. They were asked not to rush through the text, but to read as naturally as possible. This initial recording was published in their Audiobook profile as 'Beijing1' and was to be used as a baseline recording.

In the second stage, an mp3 file of the BBC radio program recording was shared with the students so that they could listen to the audio of the transcript. Students were asked to listen to the recording and make annotations on a printed transcript based on the prosodic features that they were able to identify. No right or wrong answers were given to the students; they had to identify features by themselves.

Table 2. Annotation marks

Stress	Pause	Intonation (increases and decreases in pitch)	
==	//		

Students were then asked to use their annotated transcript to improve their prosody in a second recording, and publish it as 'Beijing2' in their profile (see figure 1).

By Martin Patience, Beijing.

When I wake up in the morning // I pause briefly before opening my curtains // and what I see out of my window is likely to set the tone for the rest of the day. ↘

I'm not checking up on the weather // Instead // I want to know exactly how bad the pollution is going to be // On some mornings // it's truly appalling // It's as if the whole city has been turned into a smokers' lounge with a yellowish, nicotine colour staining the sky. ↘

And this month // pollution in Beijing went from bad to, well, dangerous.

Figure 1. Annotation method used by students.

4. BREAKING DOWN PROSODY

To address the subjectivity found in the NAEP fluency scale, a more objective methodology is proposed to evaluate the students' reading prosody. First, an annotation outline of the text's prosody, divided into pause,

stress and intonation, as interpreted by the instructor, was prepared (see figure 2).

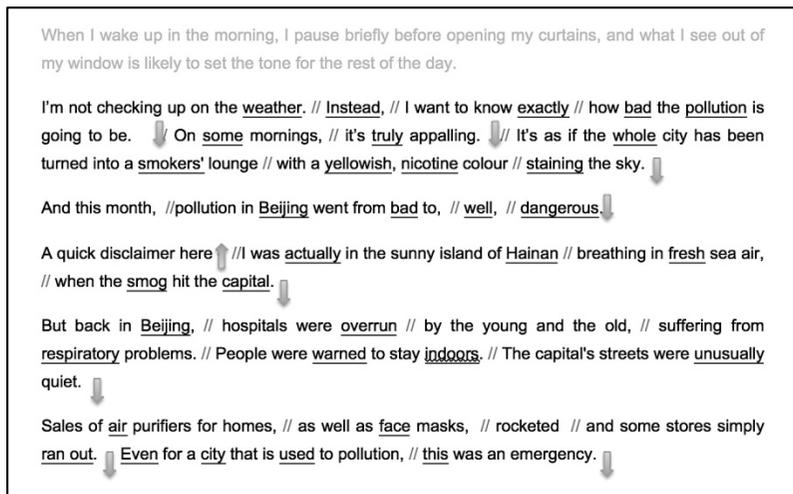


Figure 2. Annotation outline for first reading.

As can be seen, stress normally falls on content words, rather than grammatical words; pauses are usually used to identify clauses; and intonation is used for contrast, generally with pitch going down at the end of a clause, unless there is a reason to highlight contrast, in which case the pitch goes up. The annotations were done by the examiner, following the same annotation methodology as the students, resulting in an outline template with 21 instances of pause, 32 instances of stress, and finally 12 instances of intonation.

The examiner then evaluated pause, stress and intonation comparing the students' recordings with the template. One minute of reading was taken from the BBC podcast original recording for this evaluation. The minute began from the second paragraph of the transcript, as the first paragraph of transcript had been used to highlight and teach the students about the different features of prosody. The paragraphs on the transcript that were read within this minute in the original broadcast were used to evaluate the students. The evaluation consisted of counting the number of times the students paused correctly, stressed words appropriately, and used natural intonation.

5. RESULTS

Results show that students were able to replicate every feature of prosody after using the annotation methodology (see figure 3).

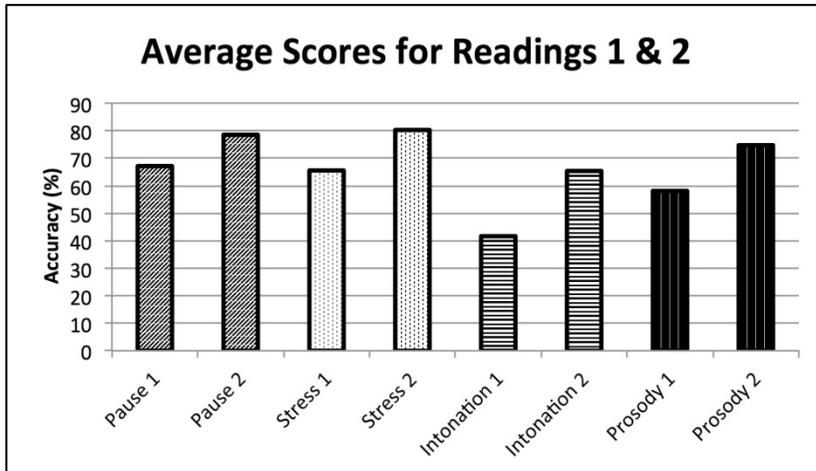


Figure 3. Graph comparing the group average scores in pause, stress, intonation and overall prosody.

A scatter plot of accuracy with a line of equality was drawn to compare the 'Beijing1' and 'Beijing2' conditions. This was done for overall prosody and also separately for each feature of prosody (see figure 4).

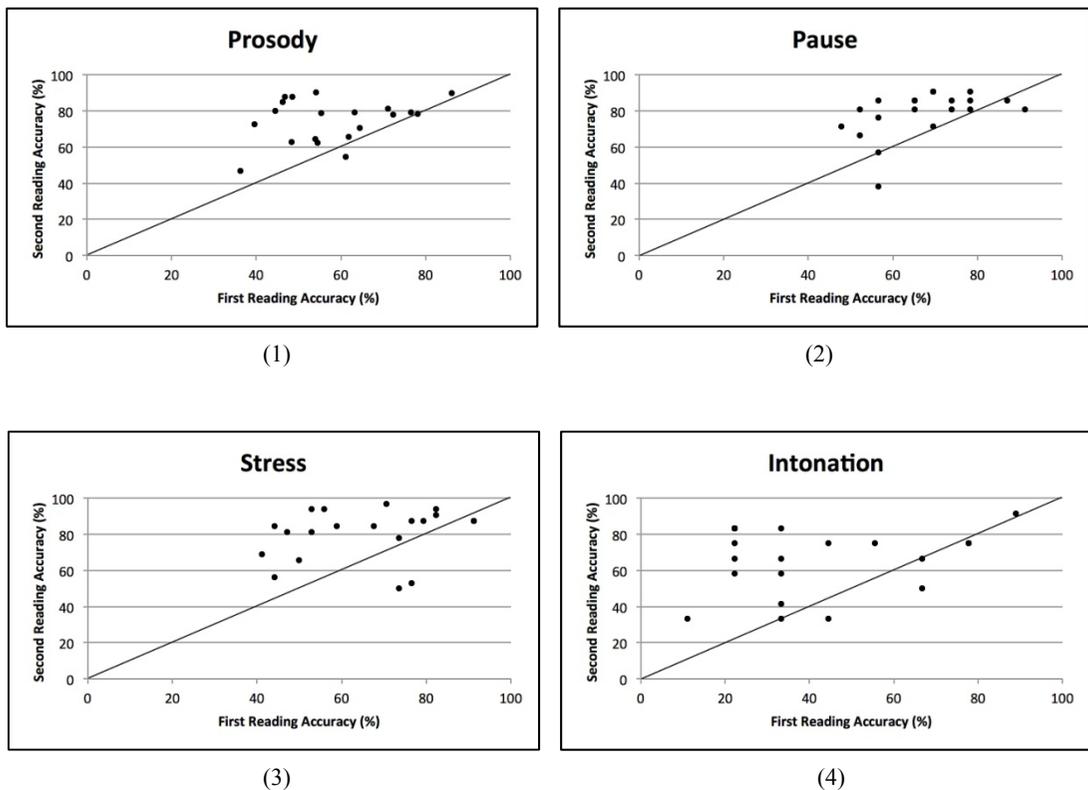


Figure 4. Scatter plot of accuracy percentage with a line of equality for the two variables: First Reading as baseline (Beijing1) and Second Reading with prosodic annotations (Beijing2). 1. Prosody: average of pause,

stress and intonation (top left); 2. Pause (top right); 3. Stress (bottom left); 4. Intonation: pitch up / pitch down (bottom right)

A Two-Sample T-Test was run to assess whether the student's improvement between the two conditions (baseline (Beijing1) and with annotations (Beijing2)) was significant. Prosody, determined as the combined average of pause, stress and intonation, was statistically significant ($t(38)=4.101$, $p<.001$) as were each of the constructs used to measure it: pause ($t(38)=2.906$, $p=.006$); stress ($t(38)=3.062$, $p=.004$); and intonation ($t(38)=3.629$, $p<.001$).

The raw scores for first (Beijing1) and second (Beijing2) readings were converted to average percentage of prosody, as an overall for the three components of pause, stress and intonation, but also broken down into each component for evaluation and comparison (Table 3).

Table 3. Average percentage of prosody, pause, stress, and intonation.

Subject	Prosody 1	Prosody 2	Pause 1	Pause 2	Stress 1	Stress 2	Intonation 1	Intonation 2
S1	61.88%	65.72%	86.96	85.71%	76.47%	53.13%	22.22%	58.33%
S2	48.36%	62.80%	56.52%	57.14%	44.12%	56.25%	44.44%	75.00%
S3	78.15%	78.37%	91.30%	80.95%	76.47%	87.50%	66.67%	66.67%
S4	61.06%	54.76%	65.22%	80.95%	73.53%	50.00%	44.44%	33.33%
S5	46.28%	85.02%	69.57%	90.48%	47.06%	81.25%	22.22%	83.33%
S6	48.58%	87.60%	56.52%	85.71%	55.88%	93.75%	33.33%	83.33%
S7	36.27%	46.73%	56.52%	38.10%	41.18%	68.75%	11.11%	33.33%
S8	39.50%	72.57%	52.17%	66.67%	44.12%	84.38%	22.22%	66.67%
S9	46.79%	87.60%	65.22%	85.71%	52.94%	93.75%	22.22%	83.33%
S10	64.51%	70.73%	73.91%	80.95%	52.94%	81.25%	66.67%	50.00%
S11	63.20%	79.27%	73.91%	85.71%	82.35%	93.75%	33.33%	58.33%
S12	54.46%	62.55%	56.52%	76.19%	73.53%	78.13%	33.33%	33.33%
S13	44.41%	80.11%	52.17%	80.95%	58.82%	84.38%	22.22%	75.00%
S14	53.86%	64.34%	78.26%	85.71%	50.00%	65.63%	33.33%	41.67%
S15	55.40%	78.92%	65.22%	85.71%	67.65%	84.38%	33.33%	66.67%
S16	54.13%	90.23%	69.57%	90.48%	70.59%	96.88%	22.22%	83.33%
S17	86.11%	89.88%	78.26%	90.48%	91.18%	87.50%	88.89%	91.67%
S18	71.08%	81.15%	78.26%	80.95%	79.41%	87.50%	55.56%	75.00%
S19	76.57%	79.02%	69.57%	71.43%	82.35%	90.63%	77.78%	75.00%
S20	72.26%	77.98%	47.8%3	71.43%	91.18%	87.50%	77.78%	75.00%

5.1 Discussion

The group's average improvement in prosody from the first reading to the second reading using the annotation methodology was 17% (see figure 3). In a breakdown of the different features of prosody, the scatter plots show most improvement is in identifying and replicating intonation (see figure 4). For first readings, out of all three of the aspects of speech being measured, students did best in pauses. This is not surprising as pauses are easier to identify than stress and intonation due to the presence of commas and full stops in writing, as well as the physical requirement of having to pause for breath at while speaking; whether a student chose to pause during the reading or was forced to as they had run out of breath is difficult to assess without asking the students after their readings.

The most striking result to emerge from the data is that all but one of the students (S4) showed an improvement in their overall prosody scores. A possible reason to explain why this student did worse overall on their second reading is distraction. While the students were encouraged to find a quiet and private area in which to go through the reading and recording procedure, the background in the recordings made by student 4 suggest that they were done in a busy coffee shop, which may have caused some distraction or perhaps hesitation due to the embarrassment of reading out loud.

Where students did worse in their second readings, it was rarely by a large amount and often followed a very high score in their first reading. Two students did, however, do significantly worse in stress in their second readings (see table 4). Reasons for this could be that they could not identify stress in the audio of the native speaker and were, therefore, unable replicate stress in their reading. Alternatively, we can see that this student did better in both pauses and intonation, which perhaps makes the drop in stress understandable; the students may have been unable to keep all three features of speech in mind while reading, finding pauses and intonation easier to identify and replicate (see figure 5). Future use of this procedure would perhaps focus only on one feature of speech at a time, rather than three, which would allow students to focus entirely on the feature being studied.

Subject	Prosody 1	Prosody 2	Pause 1	Pause 2	Stress 1	Stress 2	Intonation 1	Intonation 2
S1			86.96	85.71%	76.47%	53.13%		
S3			91.30%	80.95%				
S4	61.06%	54.76%			73.53%	50.00%	44.44%	33.33%
S7			56.52%	38.10%				
S10							66.67%	50.00%
S17					91.18%	87.50%		
S19							77.78%	75.00%
S20					91.18%	87.50%	77.78%	75.00%

Figure 5. Summary of students who showed a decrease in performance in overall prosody and / or the different features of prosody from first reading ‘Beijing1’ to second reading ‘Beijing2’

6. LIMITATIONS

While the data assessed whether students had successfully identified and replicated prosodic features, it did not account for false positives: situations where a student added stress, a pause, or a change in intonation where one should not exist. Over-generalisation of newly learned language features is to be expected, but the current analysis does not account for such instances, as the focus of this study was on whether or not students could identify the genuine instances of these language features. In hindsight, recording false positives would have given a better understanding of whether students were using a scattergun approach while annotating the transcript, or whether they were being more selective.

Furthermore, even though the methodology presented intends to address the subjectivity in identifying and evaluating prosody, the approach still involves a degree of subjectivity. There is inherent subjectivity in the students’ annotations when listening to the audio recording, as well as in the templates for examination prepared by the examiner. Moreover, the examiner must subjectively decide whether or not the students have achieved the desired stress, pause or intonation change.

7. CONCLUSION

The proposed procedure to identify and replicate aspects of prosody appears to be an effective approach for raising awareness of prosodic features that students may not have noticed before, and therefore not practiced. Students in Japan tend to do much of their language learning through textbooks rather than listening to or interacting with native English speakers, and pause, stress, and intonation cannot be easily gleaned from text alone. The methodology presented also allows for the specific analysis and measurement of key features of

reading prosody, helping examiners move beyond the somewhat vague terminology used by traditional systems of speech measurement. Removing this element of ambiguity in grading prosody may also be motivating for students; now, even small improvements in speech can be identified, which will encourage students to continue working to develop their reading skills.

This research is a first step towards a system whereby a computer programme can identify features of speech and then grade how accurately students can replicate them, providing the students with a score. This kind of “gamification” already exists in modern karaoke machines as singers can receive a score based on an analysis of their singing. In collaboration with the Korea Advanced Institute of Science and Technology (KAIST) and Keio Media Design (KMD), the authors hope to work towards creating a system for speech analysis which could potentially be a powerful tool for students of English in the future.

REFERENCES

- Chafe, Wallace. "Punctuation and the prosody of written language." *Written communication* 5, no. 4 (1988): 395-426.
- Goldman, Susan R., Peter M. Meyerson, and Nathalie Coté. "Poetry as a mnemonic prompt in children's stories." *Reading Psychology* 27, no. 4 (2006): 345-376.
- Gough, Philip B. "One second of reading." *Visible Language* 6, no. 4 (1972): 291-320.
- Hirschberg, Julia. "Communication and prosody: Functional aspects of prosody." *Speech Communication* 36, no. 1 (2002): 31-43.
- Kintsch, Walter. *Comprehension: A paradigm for cognition*. Cambridge University press, 1998.
- Koriat, Asher, Hamutal Kreiner, and Seth N. Greenberg. "The extraction of structure during reading: Evidence from reading prosody." *Memory & cognition* 30, no. 2 (2002): 270-280.
- Kuhn, Melanie R., Paula J. Schwanenflugel, and Elizabeth B. Meisinger. "Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency." *Reading Research Quarterly* 45, no. 2 (2010): 230-251.
- Kuhn, Melanie R., and Steven A. Stahl. "Fluency: A review of developmental and remedial practices." *Journal of educational psychology* 95, no. 1 (2003): 3.
- Logan, Gordon D. "Automaticity and reading: Perspectives from the instance theory of automatization." *Reading & Writing Quarterly: Overcoming Learning Difficulties* 13, no. 2 (1997): 123-146.
- Swets, Benjamin, Timothy Desmet, David Z. Hambrick, and Fernanda Ferreira. "The role of working memory in syntactic ambiguity resolution: a psychometric approach." *Journal of Experimental Psychology: General* 136, no. 1 (2007): 64.
- Wennerstrom, Ann. *The music of everyday speech: Prosody and discourse analysis*. Oxford University Press, 2001.