4pSCb10. French listeners' perceptions of prominence and phrasing are differentially affected by instruction set

Caroline Smith*

*Corresponding author's address: Linguistics, University of New Mexico, MSC 03 2130, Linguistics, Albuquerque, New Mexico 87131-0001, caroline@unm.edu

Listeners' perception of prosodic structure may differ depending on whether they are instructed to attend to the meaning of a spoken passage, or to the acoustics. Real-time perceptions of prominence and phrasal boundaries were obtained from Rapid Prosody Transcription (Cole et al. 2010). Twenty naive French listeners were divided into two groups that were given either meaning- or acoustically-based instructions for listening to passages of spontaneous speech. While listening, they read an orthographic unpunctuated transcript of the speech. Half of each group first labeled words they perceived as prominent in five passages, then phrasal boundaries in five different passages; the other half performed the tasks in the opposite order. Consistent with previous results, listener agreement (measured by kappa) was higher for labeling boundaries than for prominence. The mean kappa was 0.80 for both meaning- and acoustically-based responses, but the difference between prominence and boundaries was greater in meaning-based responses. Listeners labeled more words as prominent than they labeled boundaries, with a greater difference in meaning-based responses than acoustically-based, although the frequency of labeling did not differ overall between acoustic- and meaning-based instructions. The divergence between prominence and boundary labeling challenges the assumption that prominence in French derives from pre-boundary position.

Published by the Acoustical Society of America through the American Institute of Physics
Rapid Prosody Transcription is a technique that has been used to investigate prosody perception in several languages (e.g., English: Cole et al. 2010a,b; French: Portes 2000, Smith 2011; Dutch: Streefkerk et al. 1997; Kabyle and Hebrew: Mettouchi et al. 2007). Although the details of implementation have varied among different investigators, the basic idea is to elicit perceptual reactions to prosody from naïve listeners. These listeners are presented with unpunctuated orthographic transcriptions of speech that is then played to them. As they listen, they mark on the transcription either words they perceive as prominent or locations where they perceive phrase boundaries. These concepts are explained in non-technical terms; in most such studies the emphasis is on accessing the listeners’ intuitions and avoiding theoretical concepts or meta-linguistic analysis.

These studies have yielded impressive rates of agreement among the listeners in each study despite the non-specific instructions that were given. However, prosody is known to reflect the interplay of numerous factors that in linguistic theory are assumed to arise from different parts of the grammar. The present study deals with French, and is a pilot attempt to distinguish how two aspects of prosody contribute to listeners’ perceptions: the acoustic expression of prosody, and the meaning that listeners extract from an utterance. These were investigated by directing listeners’ attention to one or the other as they listened to recordings of spontaneous speech. We then compared their responses to investigate how attending to different aspects of prosody might give rise to differing interpretations of spoken material.

Listener responses to the two sets of instructions were compared in several ways:

- Rates of agreement in marking prominence and boundaries;
- Frequency with which listeners marked prominence and boundaries;
- Overall correlations between listener responses under the two sets of instructions;
- Specific words where listeners were consistent or discrepant in marking prominence and boundaries.

The rates of agreement were high and similar under the two sets of instructions. The other comparisons showed differences between the groups of listeners who received different sets of instructions. The paper concludes with a discussion of these differences, and what they suggest about the perception of prosody.

**METHODODOLOGY**

**Speech materials**

Ten extracts was taken from a discussion of television advertising that was broadcast on a current affairs program on the France Info radio station. These extracts consist of single-speaker passages of spontaneous conversational speech; the speakers are journalists and public figures. Their conversation was intended for broadcast and thus illustrates a somewhat formal register. The selected samples include two extracts from each of the five speakers who participated in the discussion. The duration of these extracts is from 26 to 53 seconds; the number of individual words per extract varied from 97 to 206.

Orthographic transcriptions of the extracts were prepared by the experimenter (a fluent non-native speaker of French), then edited by a native speaker with phonetic training. These transcriptions were prepared for use in the listening test by removing punctuation and line breaks except as necessary to fit on the page, in order to avoid providing any hints as to the structure. Disfluencies such as repeated or partial words were included in the transcripts but filled pauses (“euh…”) were not indicated. An additional extract with similar style of speech but from a different radio program and a different speaker was prepared to serve as practice.

**Listeners**

Twenty listeners were recruited from the student population of a French university. None had advanced training in linguistics, and all were native speakers of French. They were tested individually in a sound-attenuated room. Each listener was presented with a packet containing instructions and the printed transcripts of the practice and text extracts. They marked their responses on these print-outs.

Listeners were randomly assigned to one of four groups, with five listeners per group. Two groups received instructions directing their attention to the meaning of the extracts they heard. The other two groups were instructed to pay attention to the sound, with the suggestion that they attend especially to variation in pitch and timing. The listeners performed two tasks, marking either boundaries or prominent words. When marking boundaries, listeners were asked to mark a vertical line between words at locations where the speech could be interrupted with minimal effect on comprehension (meaning-based instructions), or at locations where they heard a break or discontinuity in the flow of speech (acoustically-based instructions). When marking prominence, listeners were instructed to...
underline words that were particularly important for the meaning (meaning-based instructions), or that stood out from the surrounding speech by virtue of some property of their sound (acoustically-based instructions). Instructions were to mark a word even though prominence in French is associated with syllables, as the need to respond rapidly would make it too hard to mark individual syllables.

For each set of instructions, one group marked boundaries first in half the extracts (one from each of the five speakers) and then prominence in the remaining five extracts, while the second group did their marking in the reverse order, first prominence in the extracts where the first group had marked boundaries, and then boundaries in the other half.

<table>
<thead>
<tr>
<th>Instructions</th>
<th>Listener Group</th>
<th>First task</th>
<th>Second task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meaning</td>
<td>A</td>
<td>Boundaries, extracts 1-5</td>
<td>Prominence, extracts 6-10</td>
</tr>
<tr>
<td>Acoustics</td>
<td>B</td>
<td>Prominence, extracts 1-5</td>
<td>Boundaries, extracts 6-10</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Boundaries, extracts 1-5</td>
<td>Prominence, extracts 6-10</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Prominence, extracts 1-5</td>
<td>Boundaries, extracts 6-10</td>
</tr>
</tbody>
</table>

Each listener heard the extracts in a different random order. Before starting, they all practiced their first task (boundaries or prominence) on the practice extract. After completing the task for five extracts, they then practiced the other task on the same practice extract before responding to the remaining five extracts.

Compared to previous studies using this methodology (e.g., Cole et al. 2010a,b, Smith 2011), the number of listeners is small for this pilot study. Recent experimentation (Jennifer Cole, pc) has suggested that smaller numbers of listeners can give reliable results when they participate under conditions that promote attentiveness. The listeners in the present study listened alone using headphones in a sound-attenuated room, which was judged to be more favorable than the group testing which was used in the study reported in Smith (2011).

Statistical analyses

Most of the results reported here are counts and correlations. These include the calculation of a boundary score (b-score) for each word, equal to the proportion of listeners who marked that word as being followed by a boundary, and a prominence score (p-score), equal to the proportion of listeners who marked that word as being prominent. As proportions, the b-scores and p-scores vary between 0 (no listeners marked the word) and 1 (all listeners marked it). These scores and the counts and correlations were calculated using Microsoft Excel.

Agreement among listeners was assessed using a modified form of Cohen’s Kappa, which takes into account the amount of agreement that can be expected by chance. Kappa values can vary between 0 and 1. The particular form of Kappa used here is based on Brennan & Predinger 1981. Calculations were made using the Online Kappa Calculator (Randolph 2012). Kappa values were determined for each extract, pooling across the five listeners in each listener group.

RESULTS

The goal of this study was to examine the extent to which listeners responded differently if they were attending to the meaning of what they were listening to, or the acoustics. The initial analyses examined the rates of agreement among the listeners within each group, to ensure that they are high enough to justify drawing conclusions, and to see if they differed depending on the instructions the listeners had received.

Agreement

Four kappa values were calculated for each extract, one for each of the four listener groups. Kappas for individual extracts varied from a low of 0.57 (one extract with meaning-based prominence marking) to a high of 0.96 (one extract with meaning-based boundary marking). The overall average was 0.80, which is within the range considered to indicate good agreement. The means for extracts marked under meaning-based instructions and under acoustically-based instructions were also both 0.80, indicating that globally, the two sets of instructions resulted in comparable agreement among listeners. Boundary-marking, however, had a higher kappa (0.86) than prominence-marking (0.74), which is consistent with previous findings for both French and other languages.
Frequency of marking

Listeners marked on average 11.5 words per extract when they had received instructions to attend to the acoustics, and 12.0 words per extract when their instructions were to attend to the meaning of the speech. There was a larger difference between tasks, with listeners marking more words for prominence (mean = 13.9 per extract) than they did for boundaries (mean = 9.6).

![Figure 1](image)

**FIGURE 1.** Mean number of words per extract marked by listeners for prominence or boundary following the word, under different instruction sets.

Most interesting was the interaction between instructions and labeling task, which can be seen in Figure 1. More words were marked under the meaning-based instructions for prominence, but more for boundaries under the acoustically-based instructions. The difference between instruction sets was not significant for prominence-marking (difference = 3.0; t=1.9, df=9, ns), but approached significance for boundary-marking (difference = 2.1; t=2.2, df=9, p=.05).

Further examination showed that listeners marked consecutive words as prominent far more often than they marked consecutive words as being followed by boundaries, particularly under the meaning-based instructions. This is not surprising, given that the consecutive words marked as prominent often grouped together syntactically, e.g., a noun phrase. If multi-word sequences marked as prominent are treated as a single unit, there is even less difference between instruction sets in terms of the number of words marked (difference = 0.6 words marked per extract).

**Correlation between marking under different instructions**

For each extract, the correlation was calculated between the p-scores for each word that were obtained under the two sets of instructions. Similarly, correlations were calculated for the two sets of b-scores for each extract. These two sets of correlations differed significantly in a paired t-test (t=3.27, df=9, p<.01) as the two sets of b-scores were more highly correlated than the two sets of p-scores. The mean correlation for prominence-marking was 0.67; for boundary-marking it was 0.80. This suggests greater concordance between the listeners who received different instructions in the case of boundary-marking than for prominence-marking. Possibly the acoustic cues to boundaries are more closely aligned with syntactic/semantic cues to boundaries than are the cues for prominence.

Under the acoustically-based instructions, there were 19 words where a majority of listeners marked the same word both as prominent and as followed by a boundary. There were 20 words where a majority marked both prominence and boundary under the meaning-based instructions. However, there were only eight words where a majority marked the same word for both prominence and boundary under both sets of instructions. In other words, the two sets of instructions resulted in comparable rates of agreement in marking, but they gave different results as to what words merited marking by listeners.
Words where there was a discrepancy between marking under different instructions

For prominence-marking, there were 29 tokens where there was a difference of at least three (out of five participants in the group) between the number who marked a word under the acoustically-based instructions and the number who marked it under the meaning-based instructions. For boundary-marking, there were 16 tokens with a difference of at least three. The most extreme case is that shown in Figure 2. All five participants who marked boundaries in this extract with acoustically-based instructions marked a boundary following the word été, [ete], ‘been’. Very likely the long filled pause following the word influenced their responses. In contrast, none of the participants who marked boundaries in this extract under meaning-based instructions indicated a boundary at this location, which seems reasonable as the sentence is incomplete and cannot be interpreted at this point.

FIGURE 2. Token where one of the largest discrepancies in marking boundaries was observed between listeners responding under acoustically-based instructions and listeners with meaning-based instructions.

In prominence-marking the greatest discrepancy occurred in the token illustrated in Figure 3. All participants marking prominence under acoustically-based instructions indicated that the word seuls, [sœl], ‘only (ones)’, was prominent, but only one participant with meaning-based instructions did so. The waveform shows that the amplitude of the vowel in this word is greater than the preceding vowels. The fundamental frequency also increases abruptly between the preceding vowel and the vowel of seuls. These factors may have influenced the participants who had acoustically-based instructions. From the perspective of the meaning-based instructions, the word seuls seems potentially salient, but it does not occur in a construction such as a cleft which would be a particularly strong cue to prominence.

FIGURE 3. Token where the largest discrepancy in marking prominence was observed between listeners responding under acoustically-based instructions and listeners with meaning-based instructions.

These two examples have in common that there were strong acoustic cues to a boundary or prominence in the absence of obvious syntactic cues. As would be expected if listeners are able to distinguish the different types of cues, differing instructions resulted in divergent reactions to these stretches of speech.
DISCUSSION

In their survey article from 1996, Shattuck-Hufnagel and Turk define prosody “as both (1) acoustic patterns of F0, duration, amplitude, spectral tilt, and segmental reduction, and their articulatory correlates, that can be best accounted for by reference to higher-level structures, and (2) the higher level structures that best account for these patterns.” (1996:196) This quote highlights the dual nature of prosody: it is both conceptual and physical. The pilot study reported here investigated whether listeners are able to concentrate their attention on one or the other of these aspects. The different responses observed in listener groups who received different instructions suggests that they were able to make this distinction.

Agreement among listeners and frequency of marking were very similar for the two sets of instructions, whereas they differed for the two marking tasks (prominence and boundaries), as had been found in previous studies that used a single fairly vague set of instructions, which different listeners might have interpreted as referring to either the physical or conceptual aspects of prosody. These results suggest that listeners did not find one set of instructions much easier or harder to implement, whereas there does seem to be a consistent finding across this and other studies that boundary-marking is more reliable than prominence-marking.

In addition, there seemed to be some interactions between the instruction sets and the marking tasks. While the overall frequency of marking words did not differ between the two instruction sets, they had differing effects for the two marking tasks. Listeners marked more words under the acoustically-based instructions when marking boundaries, but more words under the meaning-based instructions when marking prominence. One interpretation of this result is that different cues are more or less salient for the two marking tasks, more or less easy to access. The difference between instructions sets in prominence marking essentially disappeared if we treat consecutive words marked as prominent as a single unit. Particularly for meaning-based instructions, this seemed likely to be how listeners were responding, as they often marked all words in a structural unit as prominent (for example, all three words in *fabricant de contenu*, ‘content creator’, were marked as prominent).

The traditional description of French prosody is that prominence derives from position in the phrase: the final syllable in an accentual phrase, and optionally an earlier syllable, receive prominence (Di Cristo 2000). Thus the default expectation is that words perceived as prominent (having a prominent syllable) are likely to be also perceived as preceding a boundary. This expectation may not hold up under the methodology used here (Smith 2012), since only the strongest prominences – essentially, those receiving focus or emphasis – are likely to be marked by a majority of listeners, but it is worth considering. A very similar number of words were marked as both prominent and preceding a boundary under the two sets of instructions (acoustically-based: 19 words, meaning-based: 20 words), suggesting that either set is equally effective, or not, at bringing out this correspondence in French. What is more intriguing is that for the most part, words that a majority of listeners attending to the acoustics heard as both prominent and pre-boundary were not the same words that listeners attending to the meaning heard as both prominent and pre-boundary. (Recall that different listeners marked prominence and boundaries under each set of instructions, so this is not simply a consequence of one set of listeners paying attention to a word that another set ignored.) Thus, the different aspects of prosody may highlight different parts of an utterance. Or, different cues to prosodic structure may highlight different dimensions. Conceivably, the acoustic cues were most salient in words or breaks being emphasized for rhetorical affect, whereas listeners attending to the meaning would have paid the most attention to the basic structure – which locations might correspond to a sentence break, which words might be the subject of a sentence. (Although this was a broadcast debate, many utterances did not correspond to complete grammatical sentences, as is typical for spontaneous speech.)

This was a small pilot study, so much remains to be explored in terms of listeners’ access to different aspects of prosody. The differences uncovered here seem to suggest that the conceptual and physical aspects of prosody may not always coincide in conveying information about the same parts of an utterance. Listeners may need to attend to all aspects in order to comprehend the full richness of what the speaker is saying.

ACKNOWLEDGMENTS

Many thanks to GIPSA-LAB for permission to use their facilities, and especially to Nathalie Vallée for her assistance in recruiting participants.
REFERENCES


