Acoustic Prominence and Reference Accessibility in Language Production

Duane Watson¹, Jennifer E. Arnold², & Michael K. Tanenhaus³

Department of Psychology^{1,2} & Department of Brain and Cognitive Science³

University of Illinois, Urbana-Champaign¹; University of North Carolina, Chapel Hill²; & University of Rochester³ dgwatson@uiuc.edu

Abstract

Two experiments explored discourse and communicative factors that contribute to the perceived prominence of a word in an utterance, and how that prominence is realized acoustically. In Experiment 1 two hypotheses were tested: (1) acoustic prominence is a product of the given-new status of a word and (2) acoustic prominence depends on the degree to which a referent is accessible, where greater acoustic prominence is used for less accessible entities. In a referential communication task, speakers used acoustic prominence to indicate referent accessibility change, independent of givennew status. In Experiment 2 a variant of Tic Tac Toe was used to investigate whether effects of accessibility are driven by a need to signal the importance of a word or to indicate the word's predictability. The results indicate that both importance and predictability contribute to the prominence of a word, but in different ways.

1. Introduction

It is widely assumed that prosodically prominent words play a role in signaling information about the discourse status of entities in a discourse ([1],[2],[3],[4]).

Although information status is clearly related to whether or not a word receives a pitch accent, the nature of this relationship is less clear. A number of researchers have proposed that prominence signals the *givenness* of referents. Words that present some new or salient event are accented while words that refer to a previously mentioned referent are not accented ([5], [1], [3]).

In this paper, we explore an alternative hypothesis: accenting is a product of the shifting accessibility of information in the discourse. Specifically, we argue that accenting occurs on information that has become more accessible or activated in the discourse. Information that is already highly accessible is less likely to receive an accent.

The notion of cognitive accessibility has typically been used by researchers to account for choice of referring expression in a discourse ([6], [7]). It refers to the degree to which information in a discourse is activated. The accessibility of an entity at a given point in the discourse can depend on whether it is given or new, since given information tends to be more highly activated than new information. However, accessibility can depend on a host of other factors such as topicality, syntactic position, and recency of mention ([8]).

Recent evidence from a comprehension experiment by Dahan, Tanenhaus, and Chambers suggests a link between accenting and accessibility change [9]. Dahan et al. proposed that words are accented when they refer to items that are not in discourse focus, and that listeners would be sensitive to the resulting distributional pattern. Dahan et al. provided evidence for this proposal in a visual world eye movement study ([10], [11]). They found that participants identified the referent of "BED" in the second instruction more quickly in sequences like (1) than (2), where "BED" is accented in both cases.

- Theme Condition

 Put the bed above the triangle.
 Now put the BED above the square.
- 2) Goal conditiona. Put the sock below the bed.b. Now put the BED above the square.

They argue that the bed is highly activated in the theme condition (as in 1a) because this position is associated with topic-hood in imperatives. Thus, an accent on *bed* in the following sentence (1b) is infelicitous since bed is already highly activated, slowing down fixations. When *bed* occurs in the less activated goal position (as in 2a), an accent in the subsequent instruction (2b) is appropriate because *bed* is referring to an entity that is shifting to a more highly activated slot

In this study, we investigate whether speakers are actually sensitive to accessibility when producing a word. We hypothesize that reference to highly accessible entities will be produced less prominently than those that are not highly accessible. This hypothesis rests on the assumption that entities vary in accessibility along a continuum. Those that are not a part of the current discourse or situation (i.e., "new" entities) are extremely inaccessible. Furthermore, given information can vary in its relative levels of discourse accessibility.

1. Experiment 1

1.1. Task

The goals of Experiment 1 was to investigate whether information that shifts in accessibility is produced with higher acoustic prominence than information that does not. In order to achieve these goals, it was necessary to elicit natural productions in a context-rich linguistic environment, while at the same time maintaining enough experimental control to manipulate the properties of interest. We devised a variant of a referential communication task in which two participants worked together to match object locations on their computer screens [12]. Two participants were seated at two separate displays. Each display contained six objects that were initially in the same position on the director's and the follower's screen. Objects then changed locations on the director's screen and the director's task was to describe the change to his partner, so that the partner could replicate the event.



Figure 1: Example display from Experiment 1.

Each trial consisted of three sequential moving events involving the same set of six items. The first two events created a discourse context for the final, target event The theme NP of the third event, the word of interest, could have a) moved twice, b) moved once, c) been mentioned as a location, or d) be completely new. The following represent what trials might have sounded like¹:

a) 2 Theme Condition	b) 1 Theme Condition	
Put the bed above the flag.	Put the piano above the flag.	
Put the bed above the house.	Put the bed above the house.	
Put the bed above the pineapple. Put the bed above the pineapple.		
c) 1 Goal Condition	d) New Condition	
Put the piano above the flag.	Put the piano the flag	
Put the house above the bed	Put the house above the bell	

Put the **bed** above the pineapple. Put the **bed** above the pineapple.

Under a given-new theory, the target word *bed* should be accented in condition (d) and unaccented in all others because it the only condition in which the target word *bed* is new. Under an accessibility shift theory, both the 1 goal condition and the new condition should be more acoustically prominent than the 1 theme and 2 theme conditions because in both cases, the target word is shifting from low accessibility in the discourse to high accessibility.

1.1. Items and Analysis

- 16 Trials (displays) were used.
- The items were divided into four lists, counterbalanced using a Latin square design.
- Each list was presented to participants in blocks, so that each participant produced every condition for every item in the experiment at least once. This within-subject design was necessary to control for between speaker variance in utterance productions.
- Item order was randomized across blocks and across subjects.
- Each speaker utterance was recorded and the target word in the third instruction was labeled using Praat, a speech analysis program [13]. For each of the target words, duration, intensity, and pitch were extracted over the length of the word.
- In order to investigate whether differences in pitch, duration and intensity were detectable by the human ear, a research assistant who was naïve to the purpose of the experiment as well as the conditions from which each utterance originated, labeled the

accent on the target word using the ToBI intonational coding system.

In addition, the coder rated the subjective prominence of each of the target words in relation to the other words in the sentence on a scale of 1 to 7.

1.3. Results

.

The results of the acoustic analysis and the prominence rating suggest that the accessibility account provides a better account of the data then the given-new hypothesis.



Figure 2: Results from Experiment 1.

The conditions in which accessibility shifted (i.e. the 1 Goal and the New Condition) had significantly higher pitch $F_2(1,10)=48.6$, p < 0.001; $F_1(1,11)=4.61$, p=.06; higher intensity $F_1=5.37$, p<0.05, $F_2=11.20$, p<0.01; and higher duration $F_1(1,7)=14.9$, p < .01, $F_2(1,15)=17.64$, p < .001.

Interestingly, the absence of an accent occurred relatively infrequently (9.3%) and did not differ between any of the four conditions, F's <1. These results are consistent with those reported by Bard and Aylett (1999) which suggest that de-accenting is relatively rare. However, when a naïve coder rated the acoustic prominence of the target word (how "strong" the word sounded), the differences between the focus shift and non-focus shift conditions were statistically significant, $F_1(1,10)=7.44$, p<.05; $F_2(1,15)=43.74$, p<.001.

1.4. Discussion

This data suggests that accessibility shift plays a role in acoustic prominence. Effects that have been traditionally attributed to the given-new status of words may have been due to differences in the change in activation of information in a discourse.

¹ Because this was a free form production task, actual utterances varied across subjects.

3. Experiment 2

Why might speakers place acoustic emphasis on word that shift to positions of high accessibility in a discourse? There are two possibilities that have been discussed in the literature.

One possibility is that information that is important to the conversation or its goals are produced with more acoustic prominence ([2],[14]). Shifts in topic are important, so words that shift from low to high levels of accessibility are accented. Importance is also correlated with degree of givenness. In most contexts, important information is going to be new and unimportant information is going to be given.

A second possibility is that predictability guides the acoustic realization of a word. There is support for this conjecture in the literature. Gregory and colleagues have found that when a word is statistically predictable from the preceding linguistic context, it tends to be produced with a shorter duration, one of the acoustic factors correlated with prominence ([15],[16]). In addition, words that have been topics in a previous discourse tend to remain topics in subsequent utterances ([17],[18]). And, Arnold has shown that discourse cues like subject status may be cues to the likelihood of an entity being mentioned again, and when these cues make a re-mention predictable, speakers tend to use pronouns more than more explicit lexical items [19]. Thus, all things being equal, a subsequent reference to a theme is likely to be more predictable than a subsequent reference to a goal or to a discourse-new entity. If predictability modulates acoustic prominence, we would expect less prominence for more predictable items, and thus greater prominence for reference to words that involve a topic shift.

Because predictability and conversational importance make very similar predictions within the bounds of a discourse, it is difficult to know whether they contribute independently to the acoustic realization of a word, or whether they are simply two ways of characterizing a single phenomenon (accessibility shift). However, both importance and predictability can be defined in terms of nonlinguistic, task-based factors. Experiment 2 uses a variation of the game of Tic Tac Toe to separate importance and predictability.

3.1. Task

Experiment 2 uses a variation of the game of Tic Tac Toe to separate importance and predictability. Tic Tac Toe is traditionally played on a 3 x 3 grid. Players take turns placing a mark in one of the cells of the grid. The goal of the game is for players to position their marks so that they make a continuous line of three cells vertically, horizontally, or diagonally. An opponent can prevent a win by blocking the completion of the opponent's line. In order to induce participants to produce utterances that were usable for analysis, participant had their own playing boards and faced away from each other, so that verbal communication was required. Each player had a group of red objects and a group of blue objects, and was randomly assigned a color. Each of the cells in the grid was labeled with a number from 1 to 9, so that the players could indicate cell position by number. The number of the cell was the critical target word used in the analysis.

This variant of Tic Tac Toe is a useful domain for separating effects of importance and predictability on acoustic prominence. First, because each numbered cell is only mentioned once per game, all moves in a game are discourse

new. Thus, any difference in acoustic prominence between moves cannot be attributed to discourse status. Second, conversational importance is a relatively difficult notion to formalize. However, within the context of Tic Tac Toe, defining importance is straightforward. The utterance of a game move that wins or blocks the win of a game can be defined as more important than the utterance of a move that does not win or block the win of a game. Finally, Tic Tac Toe allows us to separate contributions of predictability and importance to the acoustic realization of the word because they make differing predictions. Moves that are important are highly predictable. An importance-based account would predict that a move that is important should have relatively high acoustic prominence. On the other hand, a predictabilitybased account would predict that such a move should have relatively low acoustic prominence because it is highly predictable.



Figure 3. Example games states from Experiment 2

Thus, an importance based account would predict higher prominence on the word "one" in the game state in Figure 3B than 3A because it is important (i.e. it blocks the win of a game.). A predictability-based account makes the opposite prediction because the word "one" in the context of 3A is less predictable, and should therefore be more prominent than in the context of 3B.

3.2 Items and Analysis

- Nine pairs played ten games.
- The target word was the cell location of a move.
- Moves that won a game, or blocked the win of a game were coded as being important. Those that did not were labeled unimportant.
- Pitch, intensity, duration, were extracted from the target word.
- A naïve coder rated the prominence of the target word. For labeling, moves were randomized across subjects and games, so that game context did not create any biases.

3.3 Results

	Ratings	Pitch (Hz)	Duration (ms)
Important (predictable)	3.84 (±.095)	177 (±14)	450 (±18)
Unimportant (non-predictable)	3.68 (±.095)	166 (±13)	481 (±18)

 Table 1. Prominence ratings, pitch, and duration from

 Experiment 2. Standard errors are in parentheses.

The data show that important moves (wins and blocks) were produced more prominently than the other moves in the rating and pitch data. In an analysis by participants and items, there was a significant effect of importance on prominence ratings, with important moves rated as more prominent than unimportant moves, $F_1(1,15)=5.94$, p<.05, $F_2(1,8)=8.07$, p<.05. Similar results were found in the case of pitch. Important moves had significantly higher pitch than non-important moves $F_1(1,15)=5.89$, p<.05 in subject analyses. This effect was not significant in the items analysis ($F_2 < 1$); however this is likely due to variance introduced by different contributions of each subject to each item condition (cell number).

Interestingly, numerical effects of duration were present in the opposite direction, with important (predictable) moves having shorter duration than non-important (unpredictable) moves. This trend was marginally significant in subject analyses $F_1(1,15)=3.85$, p < .07, but not significant in an items analysis ($F_2=1.13$). This trend was likely due to difficulty in planning. Data from disfluency and intonational boundary placement support this. Utterances were more likely to be disfluent in the unpredictable condition, $F_1(1,15) = 6.498$; p < .05, F₂ (1,8) = 17.132, p < .005. Similarly, in the unpredictable condition participants were more likely to pause before the object phrase, pause before the number phrase, and to have a longer duration on the object phrase. This result extends earlier findings that predictable lexical items have shorter durations ([15], [16]), establishing that predictability on a task level can impact message planning, and thus result in more intonational breaks, longer durations and higher levels of disfluency.

These data suggest that both importance and predictability contribute to the acoustic realization of words, and that these effects occur independently. Words used in the context of important moves had higher pitch and were rated as more prominent than words that were not important. However these important moves also tended to have shorter durations, fewer disfluencies, and fewer intonational phrase boundaries than non-important moves, presumably because the latter require more planning.

4. Conclusions

The results from Experiments 1 and 2 suggest that acoustic prominence is not simply a matter of labeling given and new information. The results from Experiment 1 demonstrate that words that shift in accessibility are more likely to be produced with acoustic prominence than those that do not. The results from Experiment 2 suggest that effects of accessibility shift are likely due to speakers marking these as important discourse events. Length effects associated with accessibility shift are likely due to an increased need for planning associated with referring to information that is not readily accessible.

5. References

- [1] Chafe, W. (1974). Language and consciousness. Language, 50, 111-133.
- [2] Bolinger, D. (1986). Intonation and Its Parts: Melody in Spoken English. Stanford, CA: Standford University Press.
- [3] Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (ed.), Radical Pragmatics, New York, NY: Academic Press. 223-255.

- [4] Pierrehumbert, J. & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P.R. Cohen, J. Morgan, and M.E. Pollack (eds.), *Intentions in Communication*. Cambridge, MA: MIT Press, 271-311.
- [5] Brown, G. (1983). Prosodic structure and the given/new distinction. In Cutler, A. & R. Ladd (eds.). *Prosody: Models and Measurements*. 67-77. Berlin: Springer Verla
- [6] Ariel, M. (1990). Accessing Noun-Phrase Antecedents. Routledge, London.
- [7] Gundel, J.K. (1997) "Centering Theory and the Givenness Hierarchy: Towards a Synthesis." In A. Joshi, E. F. Prince and M. Walker, eds. Centering Theory in Discourse. Oxford University Press.
- [8] Grosz, B. & Sidner, C. (1986). Attentions, intentions, and the structure of discourse. *Computational Linguistics*, 12, pp. 175-204
- [9] Dahan, D., Tanenhaus, M.K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47, 292-314.
- [10] Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. Cognitive Psychology, 6, 84-107.
- [11] Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.
- [12] Kraus, R. & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4, 343-346.
- [13] Boersma, P., & Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.3.01) [Computer program]. Retrieved from <u>http://www.praat.org/</u>.
- [14] Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, 48, 633-644.
- [15] Gregory, M. L., Raymond, W., Bell, A., & Jurafsky, D. (1999). Effects of informativeness on word duration in conversation. Vancouver, B.C.: Society for Text and Discourse.
- [16] Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2001). "Effects of disfluencies, predictability, and utterance position on word form variation in English conversation," *Journal of the Acoustical Society of America* 113:2, pp. 1001-1024, February, 2003.
- [17] Givón, T. 1983. Topic continuity in discourse: A quantitative cross-language study. Amsterdam: John Benjamins.
- [18] Grosz, B., Weinstein, S., & Joshi, A. (1995). Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics* 2(21), 203-225.
- [19] Arnold, J. (1998). Reference Form and Discourse Patterns. Unpublished doctoral dissertation, Stanford University.