# Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production

**Duane G. Watson**,
*Department of Psychology, Beckman Institute for Advanced Science and Technology, University of Illinois Urbana-Champaign*

**Jennifer E. Arnold**, and
*Department of Psychology, University of North Carolina Chapel Hill*

**Michael K. Tanenhaus**
*Department of Brain & Cognitive Sciences, University of Rochester*

## Abstract

Importance and predictability each have been argued to contribute to acoustic prominence. To investigate whether these factors are independent or two aspects of the same phenomenon, naïve participants played a verbal variant of Tic Tac Toe. Both importance and predictability contributed independently to the acoustic prominence of a word, but in different ways. Predictable game moves were shorter in duration and had less pitch excursion than less predictable game moves, whereas intensity was higher for important game moves. These data also suggest that acoustic prominence is affected by both speaker-centered processes (speaker effort) and listener-centered processes (intent to signal important information to the listener).

It is widely assumed that prosodically prominent words play a role in signaling the information status of entities in a discourse (Chafe, 1974; Bolinger, 1986; Prince, 1981; Jackendoff, 1972; Pierrehumbert & Hirschberg, 1990). For example, the word *violin* in (1b) receives emphasis, and this is related to it being the information requested by the speaker.

1)

    **a.**   What does Alessandra play?

    **b.**   Alessandra plays the VIOLIN.

Although information status is clearly related to whether or not a word receives a pitch accent, the nature of this relationship is less clear. Two different factors have been claimed to provide an account of this relationship: 1) the importance of the information to the goals of the interlocutors and 2) the predictability of the information in a given context.

Bolinger (1972, 1986) argued that the most informative words in a sentence receive an accent and some version of this view has been used to understand differences in accent type (e.g. Pierrehumbert & Hirschberg, 1990) as well as the ways in which prominence signals the

---

information structure of a sentence (e.g. Gussenhoven, 1983; Selkirk, 1996; Schwarszchild, 1999). The word "violin" in (1b) receives an accent because, as the answer to the question in (1a), it is the most important part of the sentence.

Researchers have also argued that acoustic phenomena associated with pitch accenting, especially duration, are a function of how predictable information is in a discourse. Words that are statistically predictable from the preceding linguistic context tend to be produced with shorter duration (Bell et al., 2003; Gregory et al., 1999). In addition, Gregory (2002) found that words were more likely to be produced without a pitch accent when they were predictable given their context. Related findings have shown that words that are in predictable contexts tend to have lower intelligibility (e.g. Bard & Aylett, 1999; Lieberman, 1963; Fowler & Housum, 1987), which is associated with a reduction in duration.

Importance and predictability might be two aspects of the same phenomenon. Information that is predictable tends to be less important and information that is not predictable tends to be more important (Shannon, 1951). Our primary goal is to explore whether these factors are really two aspects of the same thing or whether one or both of these factors contribute independently to the acoustic realization of a word.

A second goal is to understand whether acoustic prominence is speaker or listener centered. By marking information that is important, the speaker may help the listener coordinate the information structure of the utterance. In contrast, effects of predictability, could be the result of either speaker-centered or listener-centered processes. On the one hand, speakers might aim to produce intelligible language when there is less information from the context to help the listener, as in the case of unpredictable words (e.g., Lieberman, 1963). On the other hand, predictable words might be less prominent because they require less effort by the speaker. Speakers face the challenge of preparing and uttering their conversational contributions in real time, often while also completing other, nonlinguistic tasks. When these demands require speakers to plan complex utterances or prepare upcoming words, they tend to produce disfluent words like "um" or "uh" (Clark & Fox Tree, 2002), repeat themselves (Clark & Wasow, 1998), and produce intonational phrase boundaries (Watson & Gibson, 2004). These demands also result in longer durations for words (Bell et al., 2003; Gregory et al., 1999), as does the production of lower-frequency words (Gahl, 2006). This suggests that when speech is effortful, word durations are longer, contributing to the impression of acoustic prominence.

We used a variation of the game of Tic Tac Toe to separate the predictions of importance and predictability as well as to understand the cognitive processes underlying prominence. Tic Tac Toe is traditionally played on a $3 \times 3$ 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 our variant of the game, players placed objects on a board. In order to induce participants to produce utterances that were usable for analysis, participants 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 was the critical target word used in the analysis.

There are two reasons why this variant of Tic Tac Toe is useful for separating effects of importance and predictability on acoustic prominence. First, defining importance within the context of Tic Tac Toe is straightforward. An utterance that introduces a game move that wins or blocks the win of a game can be defined as more important than an utterance introducing a move that does not win or block the win of a game. This is particularly advantageous because

defining importance is difficult in most conversations. Importance can vary depending on the task, conversation, and intentions of the interlocutor--a point that Bolinger (1972) aptly summarized with the title of his classic article: "Accent is predictable (if you are a mind reader)". Within the context of Tic Tac Toe, however, importance is easily operationalized.

Second, Tic Tac Toe allows us to separate contributions of predictability from contributions of importance in acoustic prominence because moves that are important are highly predictable. An importance-based account predicts that a move that is important should have relatively high acoustic prominence. In contrast, a predictability-based account predicts that such a move should have relatively low acoustic prominence because it is highly predictable.

Consider the following example, illustrated in Figures 1 and 2. Figure 1 displays a game state in which the utterance of the blue player's move is of relatively little importance. Because the red player has only one object on the board, the blue player can place the balloon at almost any location and not lose (or win) the game. At the same time, the move the blue player is going to make is not very predictable.

In contrast, in a game state such as the one in Figure 2, a move to cell one would be very important since it is crucial in blocking a win by the red player. At the same time, the game state makes this move highly predictable. Under an importance-based account, one would expect the utterance introducing this move to be acoustically prominent because it is critical to not losing the game. By contrast, the predictability account predicts the utterance introducing the move to be less acoustically prominent. Over a series of games, we measured how players pronounced the numbers of the square they were placing their marker in (*one* through *nine*), and compared these pronunciations in cases where the moves were important/highly predictable and non-important/less predictable.

To understand whether prominence is associated with speaker-centered or listener-centered processes, linguistic events associated with planning such as disfluencies and intonational boundaries were measured. If the acoustic marking of either predictability or importance is speaker-centered, they should co-occur with indicators of speaker difficulty.

In sum, within the context of this game, predictability and importance make differing predictions. If predictability is the primary factor in acoustic prominence, then moves that are highly important (but highly predictable) should be less acoustically prominent than moves that are less important (but less predictable). If importance is the critical factor, than we expect the opposite pattern. Of course, because predictability has been found to correlate with duration (e.g. Gregory, 2002) and importance has been argued to correlate with pitch, duration, and intensity (e.g. Bolinger, 1972, 1986), it is possible that one might see contributions of both of these factors either in concert or in opposition.

## Method

### Participants

10 pairs (20 participants) of students from the University of Illinois Urbana-Champaign participated in the experiment for course credit.

### Materials and Procedure

Participants were seated at separate tables so that they could not see each other. Each participant had their own game board and both were given two groups of blue and red paper cut outs of objects so that they could mark their moves and the moves of their partner. The players were assigned to the red and blue objects randomly. In order to vary the starting game state, participants were given a card that instructed them where to place their first piece. These initial

trials were not included in the analysis. Both participants were free to place their objects at any location after this initial turn. All participants were familiar with the game Tic Tac Toe, and played 20 games.

16 bit recordings were made of the utterances at 44.1 KHz and were analyzed using the Praat analysis software (Boersma & Weenik, 2005). A move was labeled as important if it was necessary for winning the game or preventing a loss on the next turn. All "important" moves were also more predictable, and the less important moves were less predictable. There were a total of 642 important moves and 715 non-important moves. In order to control for idiosyncratic acoustic differences between words and speakers, utterances were only included in the analysis if the speaker had produced the move in both important and non-important contexts. Using these criteria, 9.3% of the cell numbers went unmatched. These were not included in the analysis.

We hypothesized that if importance or predictability effects result from speaker-interrnal processing constraints, they should correlate with markers of production difficulty. We therefore examined each utterance for the following three factors: 1) was there an intonational boundary or pause immediately before the prepositional phrase denoting the move location (e.g., "[boundary] on number five"), or before the number word ("on [boundary] five"); 2) was there any disfluency in the utterance (um, uh, repeats, repairs, saying "thiy" for "the" or "ay" for "a", or hesitations ("hmmm", "mmm", "ooooh"); and 3) the duration of the object phrase.

## Results

The means for maximum F0, minimum F0, intensity and duration are presented in Table 1. In this analysis, there was a reliable effect of intensity in the direction predicted by importance. Cell numbers produced in the context of important moves had higher overall intensity than numbers produced in non-important moves, $F_1(1,19)=9.92$, $p < .01$; $F_2(1,8)=20.53$, $p < .01$.

The opposite pattern was found for duration, in keeping with the predictability hypothesis. Moves that were not predictable contained cell numbers that were produced with longer duration than predictable moves, $F_1(1,19)=55.51$, $p <.001$; $F_2(1,8)=109.78$, $p < .001$.

Measures of F0 change and amplitude also patterned with the duration data. The pitch excursion (maximum F0) was higher when the move was not predictable $F_1(1,19)=4.80$, $p <. 05$; $F(1,8) =15.08$, $p < .01$. There was also a greater difference in pitch range (the difference between the maximum and minimum F0) for non-predictable moves than predictable moves, $F_1(1,19) =7.18$, $p < .05$; $F_2(1,8)=24.54$, $p < .01$. One might argue that the source of these F0 differences is attributable to duration. If the rising F0 for two words of unequal length have the same slope and starting point, the pitch excursion will be higher for the longer word because the time over which the F0 has to rise is greater. However, the minimum F0 in the non-important/non-predictable conditions was lower than in the important/predictable conditions, $F_1(1,19)=8.22$, $p < .05$; $F_2(1,8)=10.85$, $p < .05$, suggesting that the larger pitch difference for non-important moves was driven by an expanded pitch range rather than just differences in duration.

One consequence of Tic Tac Toe is that the predictability of game moves changes as the game progresses. As more and more spaces are filled on a game board, the predictability of any potential move increases. Thus, it is possible that the differences between important and non-important moves reflect the changing predictability of moves across the game rather than differences between the predictability and importance of a move given a game state. To see if this was in fact the case, data from the sixth move of the game and higher were examined in a sub-analysis. All of these moves occur relatively late in the game when the majority of the cells have been filled, and consequently, should be less affected by predictability based on the time point in the game. The F0, duration, and intensity of the cell numbers in these moves are in

presented in Table 2. After the sixth move of the game, cell numbers produced in non-predictable contexts were significantly longer than moves produced in predictable contexts F1 (1,19)=38.45, p < .001; F2(1,8)=22.61, p< .01. There were also significant effects of intensity, with predictable/important moves having a greater intensity than non-predictable/non-important moves, F1(1,19)=17.97, p < .001; F2(1,8)=11.59, p < .01. There were no differences in F0.

These data suggest that both predictability and importance contribute to the acoustic realization of a word. While important moves were produced with higher intensity, non-predicable, non-important moves were produced with higher F0 and duration. An analysis of the measures of production difficulty suggests that the latter may be due to effortful planning. Speakers had more production difficulty in the non-predictable/non-important moves, as shown in Table 3. There was a higher likelihood of pausing immediately before the target number word ($F_1(1,19)$ = 35.94, p < .001; $F_2(1,8)$ = 52.07, p < .001). Speakers were also more likely to produce a disfluency in non-predictable than predictable moves ($F_1(1,19)$ = 35.60, p < .001; $F_2$ (1,8) = 24.71, p < .001). Even though the target word always came at the end of the utterance, we expected that some planning would occur earlier in the utterance (cf. Clark & Wasow, 1998). We therefore examined the object phrase for evidence of planning and production difficulty, assuming that more words would be produced and that the produced words would have longer duration when more planning time was necessary. As predicted, the object phrase was longer in duration when the move was non-predictable than when it was predictable, ($F_1(1,15)$ = 19.71, p < .01, $F_2(1,5)$ = 28.91, p < .001).

## Discussion

These results suggest that both importance and predictability play a role in the acoustic realization of a word. Duration is longer and pitch movement is greater for non-predictable words while intensity is greater for important words.

One potential concern is that the effects of intensity may have been the result of paralinguistic factors such as emotional excitement related to winning the game. How one might distinguish between the linguistic and paralinguistic factors that drive prominence is a question of considerable debate (see Ladd, 1996 for a discussion). However, it is important to keep in mind that important trials did not consist of only emotionally exciting wins, but also the relatively more routine cases where a win was blocked. Moreover, the speaker's personal reaction to the importance of their utterance is not inconsistent with the proposal that importance drives acoustic prominence.

These results also suggest that acoustic prominence is the result of both speaker and listener-centered processes. Disfluencies and intonational boundaries, which are linked to production difficulty, occurred more frequently in the non-predictable moves. This suggests that both duration and F0, which were also higher in these conditions, are linked to speaker-centered processes. It is unknown whether duration and pitch effects are the byproducts of planning processes, such that they are the result of more effortful production, or whether they actually facilitate the production of lexical items that are difficult to access. In the case of duration, a longer word might facilitate production by providing more processing time. These data also suggest that acoustic prominence may be used overtly to assist listeners. The intensity of the target word, which marked important information, did not correlate with disfluencies and intonational boundaries, suggesting only a weak link with production processes. Therefore, marking the target word with higher intensity may have been done to assist the listener.

The more general question of whether speaker preferences are listener centered or speaker centered is a topic of considerable debate in the psycholinguistics literature (see Arnold, in

press). Recent work suggests that speakers' production choices are often made without listener needs in mind, both with respect to prosody (Schafer et al., 2001; Kraljic & Brennan, 2005), and syntactic choices (Arnold, Wasow, Asudeh, & Alrenga, 2004; Ferreira & Dell, 2000, but see Haywood, Pickering, & Branigan, 2005). However, research also suggests that listeners pay attention to prominences (Arnold, 2007; Dahan, Tanenhaus & Chambers, 2002), perhaps because a speaker's production choices create systematic patterns in the input that listeners learn implicitly. Moreover, many researchers have argued that speaker's and listener's mental states are closely aligned in conversation (Clark, 1996; Pickering & Garrod, 2004; Brown-Schmidt, Campana & Tanenhaus, 2005). Studying acoustic prominence in this context may shed light on understanding the interplay between speaker and listener requirements.

An important challenge for future research is to understand how listeners are able to compute prominence, given that prominence can be conveyed by multiple acoustic dimensions, some of which may fractionate under different circumstances. Ladd (1996), among others, has warned against interpreting any one acoustic factor such as F0, for example, as a transducer of discourse status. The data presented here are consistent with this claim. They suggest that a complex gestalt of acoustic features work together to convey prominence, with different factors affecting the acoustic signal in different ways (e.g. importance affects intensity and predictability affects duration and F0).

### Acknowledgements

# References

Arnold JE. Reference Production: Production-internal and Addressee-oriented Processes. Language and Cognitive Processes. in press

Arnold, JE. THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. University of North Carolina; Chapel Hill: 2007. Unpublished manuscript

Arnold JE, Wasow T, Asudeh A, Alrenga P. Avoiding attachment ambiguities: The role of constituent ordering. Journal of Memory and Language 2004;51:55–70.

Bard, EG.; Aylett, MP. The dissociation of deaccenting, givenness and syntactic role in spontaneous speech. Proceedings of ICPhS-99; San Francisco. 1999.

Bell A, Jurafsky D, Fosler-Lussier E, Girand C, Gregory M, Gildea D. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. Journal of the Acoustical Society of America 2003;113(2):1001–1024. [PubMed: 12597194]

Boersma, P.; Weenink, D. Praat: Doing phonetics by computer (Version 4.5.01) [Computer program]. Amsterdam: Institute of Phonetic Sciences; 2005. Retrieved from http://www.praat.org/

Bolinger D. Accent is predictable (if you're a mind-reader). Language 1972;48:633–644.

Bolinger, D. Intonation and Its Parts: Melody in Spoken English. Stanford, CA: Stanford University Press; 1986.

Brown-Schmidt, S.; Campana, E.; Tanenhaus, MK. Real-time reference resolution by naïve participants during a task-based unscripted conversation. In: Trueswell, JC.; Tanenhaus, MK., editors. World-situated language processing: Bridging the language as product and language as action traditions. Cambridge: MIT Press; 2005. p. 153-171.

Chafe W. Language and consciousness. Language 1974;50:111–133.

Clark, HH. Using language. Cambridge University Press; 1996.

Clark HH, Fox Tree JE. Using *uh* and *um* in spontaneous speech. Cognition 2002;84:73–111. [PubMed: 12062148]

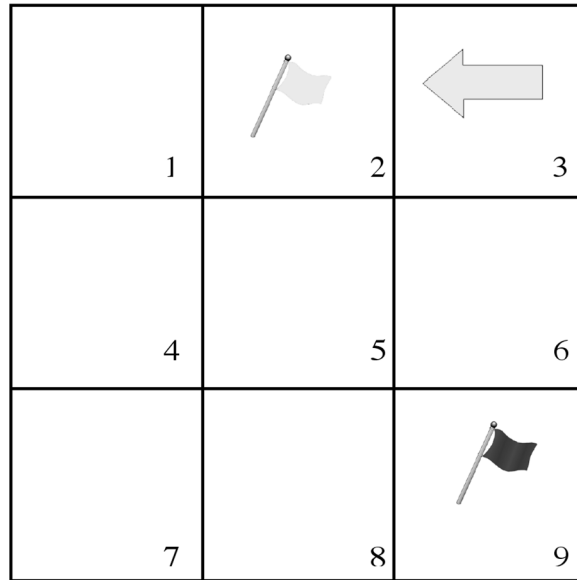Clark HH, Wasow T. Repeating words in spontaneous speech. Cognitive Psychology 1998;37:201–242. [PubMed: 9892548]

Dahan D, Tanenhaus MK, Chambers CG. Accent and reference resolution in spoken-language comprehension. Journal of Memory and Language 2002;47:292–314.

Ferreira VS, Dell GS. Effect of ambiguity and lexical availability on syntactic and lexical production. Cognitive Psychology 2000;40:296–340. [PubMed: 10888342]

Fowler CA, Housum J. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. Journal of Memory and Language 1987;26:489–504.

Gahl, S. Is frequency a property of phonological forms? Evidence from spontaneous speech. Paper presented at the19th annual CUNY conference on human sentence processing; New York City, NY. 2006.

Gregory, ML. Linguistic Informativeness and Speech Production: An Investigation of Contextual and Discourse Pragmatic Effects on Phonological Variation. University of Colorado Boulder; 2002. Unpublished doctoral dissertation

Gregory, ML.; Raymond, W.; Bell, A.; Jurafsky, D. Effects of informativeness on word duration in conversation. Vancouver, B.C: Society for Text and Discourse; 1999.

Gussenhoven, C. A semantic analysis of the nuclear tones of English. Bloomington, Ind: Indiana University Linguistics Club; 1983.

Haywood S, Pickering MJ, Branigan HP. Do Speakers Avoid Ambiguities During Dialogue? Psychological Science 2005;16(5):362–366. [PubMed: 15869694]

Jackendoff, R. Semantic interpretation in Generative Grammar. Cambridge, MA: MIT Press; 1972.

Kraljic T, Brennan SE. Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? Cognitive Psychology 2005;50:194–231. [PubMed: 15680144]

Ladd, DR. Intonational Phonology. Cambridge: Cambridge University Press; 1996.

Lieberman P. Some effects of the semantic and grammatical context on the production and perception of speech. Language and Speech 1963;6:172–175.

Pickering MJ, Garrod S. Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences 2004;27

Pierrehumbert, J.; Hirschberg, J. The meaning of intonational contours in the interpretation of discourse. In: Cohen, PR.; Morgan, J.; Pollack, ME., editors. Intentions in Communication. Cambridge, MA: MIT Press; 1990. p. 271-311.

Prince, EF. Toward a taxonomy of given-new information. In: Cole, P., editor. Radical Pragmatics. New York, NY: Academic Press; 1981. p. 223-255.

Schafer, AJ.; Speer, SR.; Warren, P.; White, SD. Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. Fourteenth Annual CUNY Conference on Human Sentence Processing; Philadelphia, PA. 2001.

Schwarzschild R. GIVENness, Avoid F and other constraints on the placement of focus. Natural Language Semantics 1999;7:141–177.

Selkirk, EO. Sentence prosody: Intonation, stress and phrasing. In: Goldsmith, JA., editor. The handbook of phonological theory. Cambridge, Mass., USA: Blackwell; 1996. p. 550-569.

Shannon CE. Prediction and entropy of printed english. Bell System Technical Journal 1951;30:50–64.

Watson D, Gibson E. The relationship between intonational phrasing and syntactic structure in language production. Language and Cognitive Processes 2004;19:713–755.
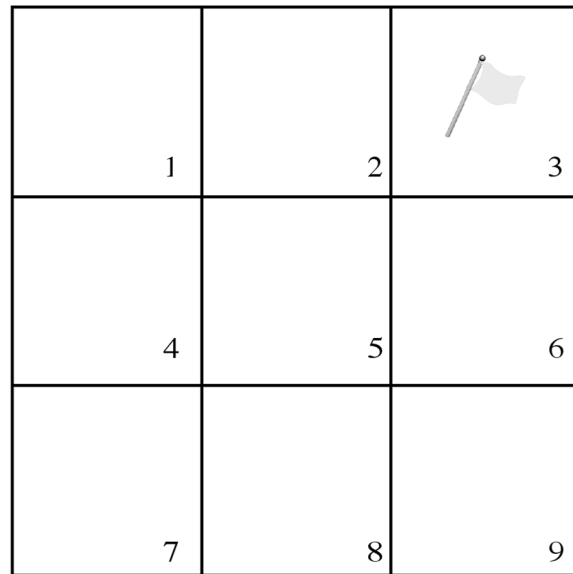
"Put the blue balloon in one."

**Figure 1.**
A Tic Tac Toe game state in which the blue player's move can be described as of relatively low importance for the goals of the game since the placement of the blue player's balloon cannot win or prevent the win of a game.

"Put the blue ballon in one"

**Figure 2.**
A Tic Tac Toe game state in which the move blue player's move is important. In order to prevent a loss of the game, the blue player must place the balloon in cell 1.

**Table 1**

The F0, intensity and duration of the cell numbers for each game move. Standard errors are in parentheses.

| | Important/Predictable | Non-important/Non-predictable |
|---|---|---|
| Maximum F0 (Hz) | 224.27 (10.38) | 240.67 (12.10) |
| Minimum F0 (Hz) | 105.41 (6.18) | 99.86 (5.20) |
| Intensity (db) | 67.16 (.83) | 66.50 (.79) |
| Duration (msec) | 1617 (64) | 2585 (149) |

**Table 2**

The F0, intensity and duration of the cell numbers after the sixth move of the game. Standard errors are in parentheses.

|  | Important/Predictable | Non-important/Non-predictable |
| --- | --- | --- |
| Maximum F0 (Hz) | 227.66 (12.75) | 222.20 (12.98) |
| Minimum F0 (Hz) | 103.51 (5.79) | 101.79 (5.92) |
| Intensity (db) | 67.09 (.79) | 66.00 (.87) |
| Duration (msec) | 1557 (76) | 2206 (115) |

**Table 3**

Percentage of disfluencies and pauses in the game moves. Standard errors are in parentheses.

| | Important/Predictable | Non-important/Non-predictable |
|---|---|---|
| % disfluent utterances | 9.79 (2.15) | 25.36 (3.68) |
| % pauses before target word or target PP | 15.10 (2.86) | 37.28 (3.53) |
| Duration of object phrase (msec) | 552 (51) | 889 (91) |