

Comprehension and resynthesis of duration and pitch in ambiguous words
Angela M. Isaacs and Duane G. Watson
University of Illinois at Urbana-Champaign
aisaacs2@uiuc.edu

Previous research suggests that F0, intensity, and duration play a role in signaling the presence of a pitch accent (e.g. Pierrehubert & Hirschberg, 1990). However, less is known about the precise contribution of each. The current work examines whether duration, F0 excursion, and F0 slope can independently drive the perception of an accent.

In a visual world experiment, Dahan, Tanenhaus, and Chambers (2002) found that listeners rapidly interpret pitch accents in on-line processing. In their study, speakers followed sets of two instructions as in [1] to move objects on a computer screen while their fixations were monitored. On critical trials two of the pictured objects were phonetic cohorts (e.g. camel/candle). The first instruction directed the participant to move one of the cohorts. In the second instruction the participant was directed to move either the previously mentioned cohort (given) or the unmentioned cohort (new). They found that listeners were more likely to fixate the given cohort when the cohort was unaccented and were more likely to fixate the new cohort when it was accented.

Because only the overall presence vs absence of an accent was manipulated, it is impossible to know which aspect of the acoustic signal was driving fixations in the Dahan et al. study. To investigate this, a modification of their design was implemented using more precisely controlled stimuli. Three factors were manipulated: the duration (long vs short), the pitch excursion (high vs. low), and the discourse status (given vs new) of the cohort in the second instruction.

The PSOLA re-synthesis algorithm (Moulines & Charpentier, 1990) was used to independently manipulate the F0 and duration of the critical word. Because re-synthesis alters the acoustic properties of naturally produced speech, it has the advantage of being more natural than synthesized speech, but offers more control than spontaneous speech. The source audio was recorded by a male speaker. To determine what durations and pitch contours to use as targets of the re-synthesis, the speaker produced all the critical utterances in accented and de-accented contexts. The overall ratio of the duration of the accented to unaccented words in these productions was used to generate the stimuli in the short and long conditions. The median pitch contours in the accented and de-accented contexts were used to generate the high and the low conditions.

An analysis by time region of the Target-Advantage scores revealed a significant context (given/new) x duration (long/short) x pitch (high/low) x time region (-200to200/200to600 ms after word onset) interaction $F_1(1,46)=4.873$, $p<.05$, $F_2(1,38)=7.31$, $p<.05$. In the short duration condition, fixations were higher to New cohorts when the pitch was high and looks were higher to the Given cohorts when the pitch was low. However, effects of pitch were not present in the long duration condition. This suggests that the perception of accenting is driven in large part by F0 slope. Conditions with short duration had more extreme slopes as compared to those with long duration. Pitch and duration by themselves did not appear to drive the perception of a pitch accent.

Example [1] a. Context instruction: *Move the camel/candle below the circle.*
b. Test instruction: *Now move the camel above the triangle.*

References

Dahan, D., Tanenhaus, M. K., & Chambers, C. C. (2002). Accent and reference resolution in spoken-

language comprehension. *Journal of Memory and Language*, 47, 292-314.

Moulines, E. & Charpentier, F. (1990). Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9, 453-467.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271). Cambridge: MIT Press.