

# Comprehension and resynthesis of duration and pitch in ambiguous words

Angela M. Isaacs and Duane G. Watson  
University of Illinois at Urbana-Champaign

## ABSTRACT

Past research shows that listeners interpret deaccented words as referring to Given (previously mentioned) words and accented words as referring to New (not previously mentioned) words (Dahan et al., 2002). The present study tests the comprehension effects of independently altering pitch and duration through resynthesis. An interaction is found between pitch and duration such that for more extreme pitch slopes, high pitch is interpreted as referring to New objects and low pitch is interpreted as referring to Given objects. When slopes are less extreme (in the long duration conditions), no preference is found.

## PREVIOUS RESEARCH

Dahan, Tanenhaus, and Chambers (2002) found that listeners rapidly use prosodic information to extract information from ongoing speech.

- [1] Move the candle below the circle.
- [2] NOW move the camel/CAMEL above the triangle.

In a visual world paradigm study, when the critical word was deaccented (camel), fixations are higher to the Given cohort (candle). When the critical word was accented (CAMEL), fixations were higher to the New cohort (camel).

It is unclear how listeners are performing this task. Most theories have concentrated on listeners' use of pitch information (e.g. Pierrehumbert and Hirschberg, 1990) but little is known about the role of duration or the interaction of duration and pitch.

## CURRENT RESEARCH

The current study uses a variant of the task used by Dahan et al. (2002). Participants' fixations are recorded as they listen to critical utterances which differ by pitch contour (high/low) and duration (long/short).

Critical utterances were created using PSOLA resynthesis (Moulines & Charpentier, 1990). PSOLA resynthesis changes the duration and pitch of naturally-produced audio based on acoustic parameters calculated from the source audio set. This allows for both a high degree of control and a high degree of naturalness.

## METHODS DESIGN AND PARTICIPANTS

- \* 8 conditions total: Context (Given/New) x Duration (Long/Short) x Pitch (High/Low)
- \* 20 items
- \* 48 participants

## PROCEDURE

Participants view a computer display of four possible targets and four stationary geometric figures. On critical trials, the display contained two cohorts (e.g. camel/candle).

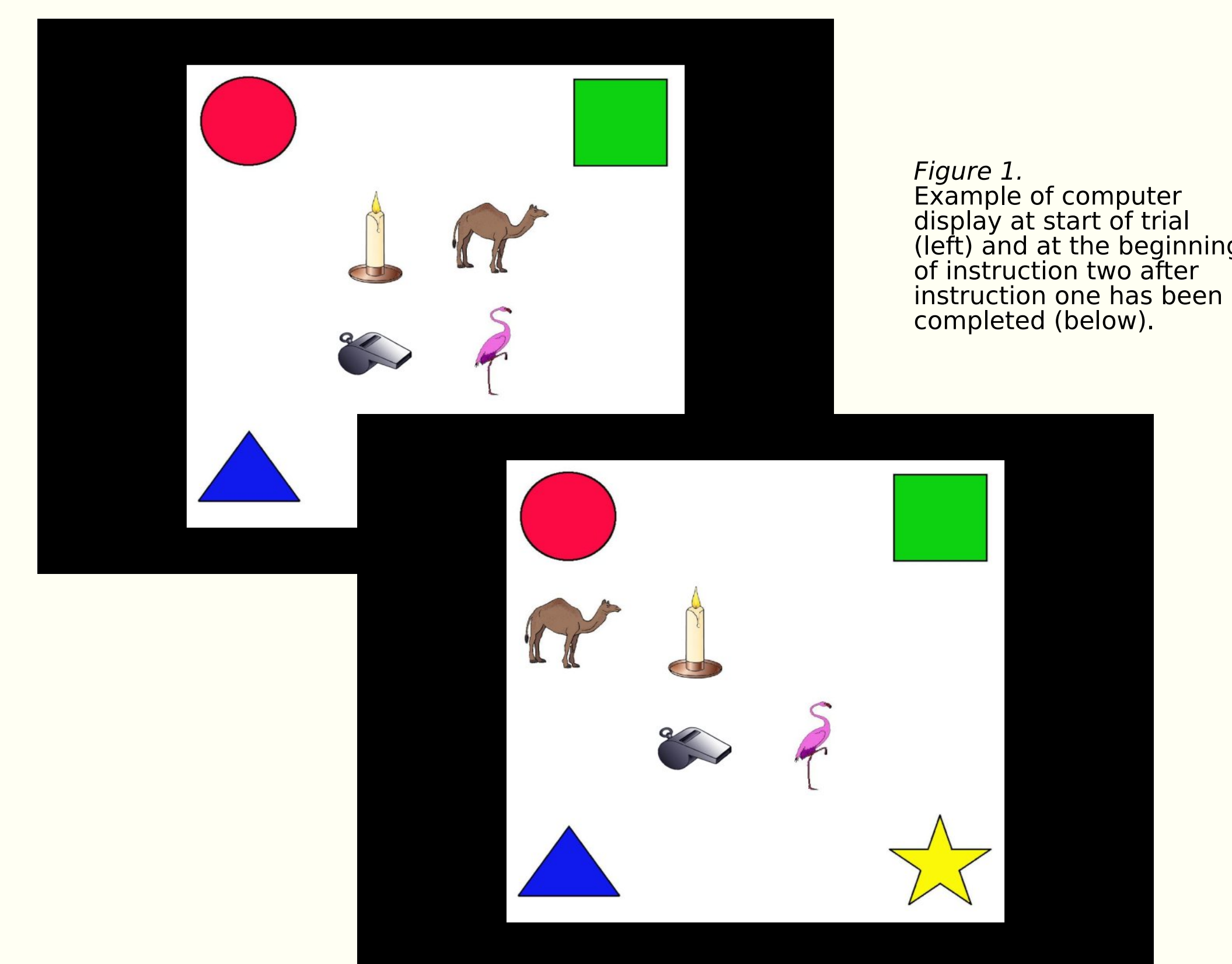


Figure 1. Example of computer display at start of trial (left) and at the beginning of instruction two after instruction one has been completed (below).

Example Instructions:

- [1] Given: Move the camel below the circle.  
New: Move the candle below the circle.
- [2] Low-Long: Now move the camel above the triangle.  
Low-Short: Now move the CAMEL above the triangle.  
High-Long: Now move the camel above the triangle.  
High-Short: Now move the CAMEL above the triangle.

Participants are given two instructions to move objects within the display. On critical trials, the first instruction directs the participant to move one of the cohorts. This sets the context (Given/New) of the critical target in the second instruction.

The second instruction directs the participant to move one of the cohorts. The duration and pitch of the critical cohort were resynthesized from naturally-produced utterances.

## METHODS CONT. STIMULI

### Source Audio

Experimental stimuli were created by resynthesizing (changing the acoustic parameters of) naturally-produced audio. The acoustic parameters used in the resynthesis were based on the parameters present in the source audio.

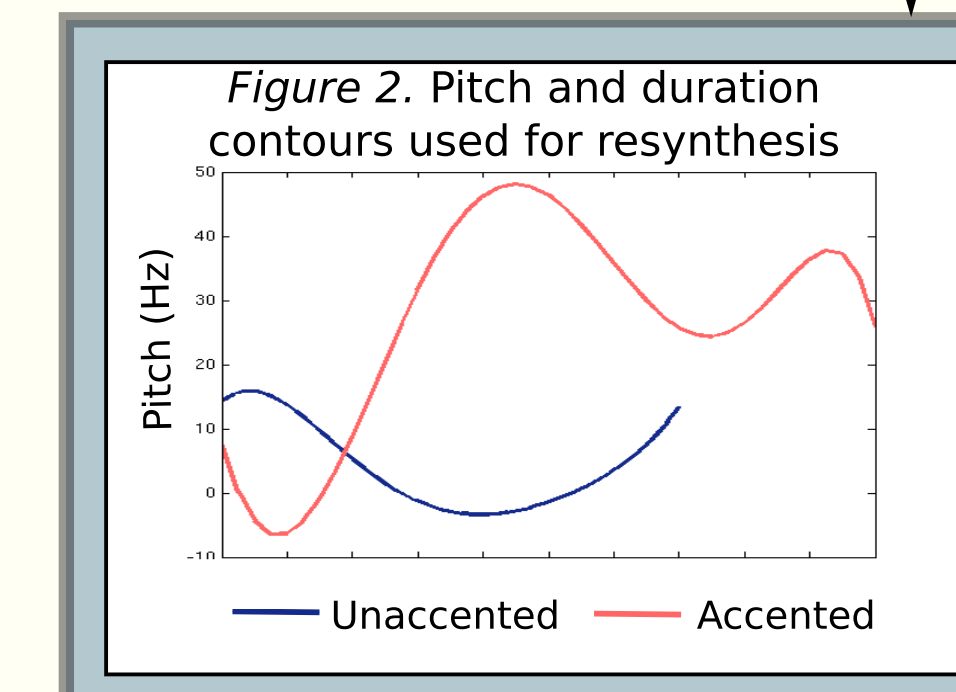
Example Source Audio:

- [1] Given: Move the camel below the circle.  
New: Move the candle below the circle.

- [2] Unaccented: Now move the camel above the triangle.  
Accented: Now move the CAMEL above the triangle.

### Calculate Resynthesis Parameters

- \*Measured pitch contours for all critical words.
- \*Calculated median pitch contour for accented and unaccented conditions.
- \*Measure the duration of each critical word



### Resynthesis

- \*Use Accented version as the basis for resynthesis
- \*Pitch contour:  
High pitch = Accented  
Low pitch = Unaccented
- \*Duration:  
Long duration = No change  
Short duration = 0.7\*Accented

- [2] Low-Long: Now move the camel above the triangle.  
Low-Short: Now move the CAMEL above the triangle.  
High-Long: Now move the camel above the triangle.  
High-Short: Now move the CAMEL above the triangle.

The resynthesized critical words were cross-spliced into the Unaccented carrier sentence.

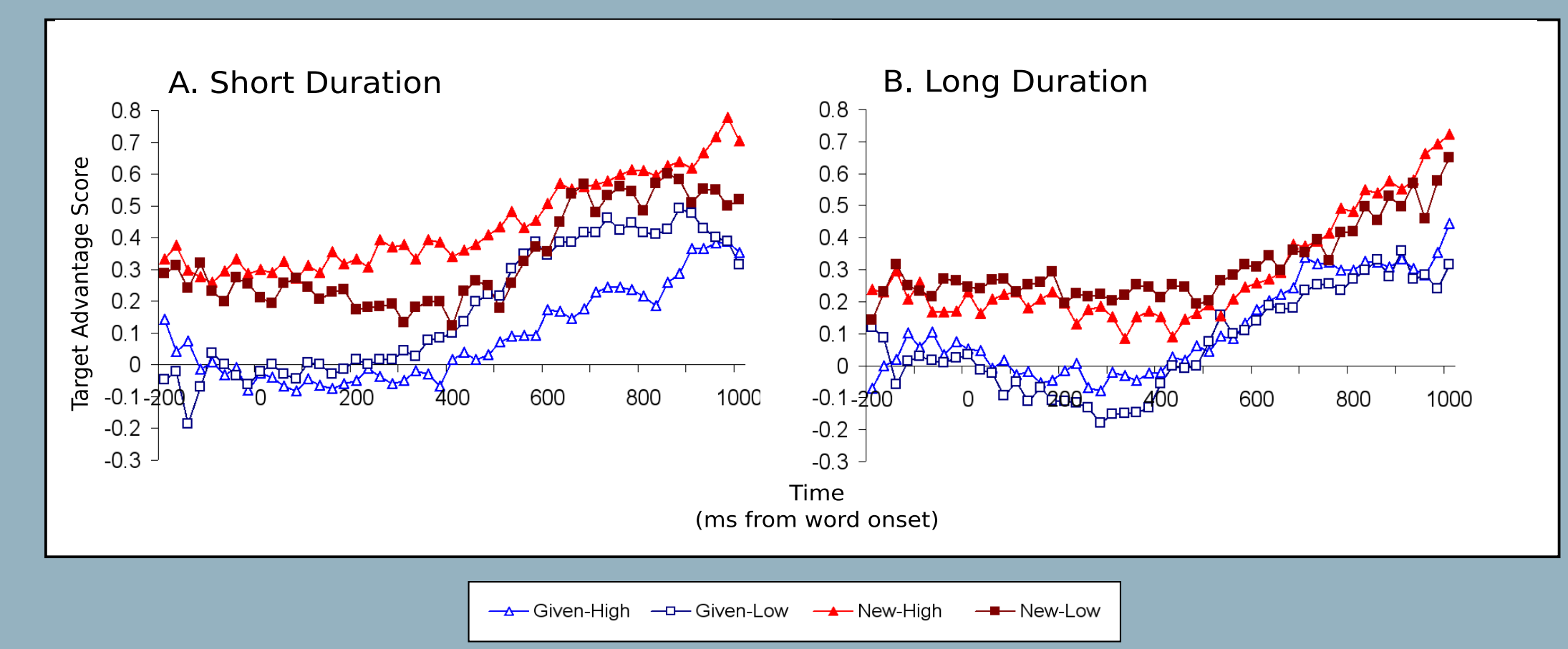
## RESULTS

Baseline difference in Given/New conditions present at word onset (consistent with Dahan et al (2002)).

Significant Context x Pitch x Duration x Time region interaction for Target Advantage Scores (TAS; proportion of looks to target minus proportion of looks to competitor)

- \*F1 (1,46)=4.873, p<.05
- \*F2 (1,38)=7.310, p<.05

Figure 4. Context x Duration x Pitch Interaction Graphed by Duration



In Short duration (Fig. 4A), higher TAS for New object in High pitch condition and higher TAS for Given object in Low pitch condition.

In Long duration (Fig. 4B), no difference.

## DISCUSSION

Listeners' showed fixation preferences similar to those found by Dahan et al. (2002) but only when duration was short. When duration is short, the rate of change in pitch (i.e. slope) is most extreme so listeners may accumulate data about the direction and magnitude of pitch change more quickly. When duration is long, the slope is less extreme and listeners do not show any fixation preferences. This data suggest that listeners are utilizing the slope of pitch change to infer the presence of an accent and make judgements about the probable target of an utterance based on this inference.