“Sorry, what was that?”: The roles of pitch, duration and amplitude in the perception of reduced speech

Ryan Podlubny, Benjamin V. Tucker, and Terrance M. Nearey

Introduction

- Context (phonetic, syntactic and semantic) plays an important role in the processing of reduced speech forms (Ernestus et al., 2002; Ven et al., 2011; Eyck, 2011)
- The present study considers the level of reduction and the importance of specific phonetic information available in reduced productions, relative to perception

Research goal

- By manipulating selected acoustic elements we aim to clarify the conditions under which listeners find reduced speech intelligible
- The present study examines the contributions of pitch, amplitude and duration relative to perception

Method

Stimuli

- Native speaker of Western Canadian English
- Twenty-five minute telephone conversation
- Interlocutor not recorded
- Speech segmented into 71 usable frames of varying sizes (one vs. many words) and degrees of reduction (minor to massive)
- Level of reduction: proportion of actual phonemes to total “expected” number of phonemes (deletion ratio)

Manipulation/Procedure

Participants were exposed to five variants of each signal

Task: Fill in the blank. “So I like take ______ box.”

1. Visual cloze test (Taylor, 1953): No auditory stimuli
2. Amplitude: Signal correlated noise matches intensity contour (Fig1)
3. Duration: PSOLA used to stretch target to 1.5 times length (Fig2)
4. Pitch: LPC residual modulated by short-term averaged amplitude contour of original (Fig3) (Note: Windowing alignment errors may be present)
5. Original: The removed and reintegrated, unmodified signal (Fig4)

Visual context accompanies subsequent auditory manipulations

Blocks 1 and 5 occur first and last, respectively; blocks 2-4 are randomized

All signals are randomized within blocks

Participants

- 22 native speakers of western Canadian English (6M/16F)
- One participant excluded for not completing the task

Results

- Our modeling shows speech rate (syllables/sec), trial, and deletion ratio as not significant
- Significant improvement when phonetic context is supplied
- Location of target is important (medial vs. initial vs. final)
  - Medial scores are significantly higher than both initial and final
  - Participants improve as experiment progresses, regardless of Block Order
  - Block Order does contribute to accuracy of responses eg. Dur → SCN → F0 elicits higher overall scores than F0 → SCN → DUR

Accuracy:

1. Original 1397
2. Summed correct responses (1562 trials for each manipulation) 1386
3. Duration 671
4. Fundamental Frequency 602
5. Signal Correlated Noise 129

Sample manipulations: “So I like take them out of this box.”

Discussion

- Block manipulations may be likened to puzzle pieces
- Phonetic components are absorbed and reassembled to shape a more detailed realization of the missing target
- Not all “puzzle pieces” offer an equal amount of information

Further experiment/analysis

- Continued examination of these manipulations as proportions (as opposed to signals that have been completely replaced)
- Exploration of socio-cultural factors that may contribute to better scoring on particular Blocks

References:

Acknowledgements:
D. Kyle Danielson: technical support, Kaillin Mackie: RA, Erika Walker: RA, Adrienne Findlay: RA. This project funded by Faculty of Arts Roger S. Smith Student Research Award, The Faculty of Arts Student Life and Enrichment Fund, University of Alberta Department of Linguistics, and SSHRC

Contact: podlubny, bvtucker@ualberta.ca

Nijmegen Spontaneous Speech Workshop, Nijmegen, NL December 2011