Are listeners sensitive to the phonological form of prosody or its phonetic encoding?

Jennifer Cole
University of Illinois at Urbana-Champaign, USA

Stefanie Shattuck-Hufnagel
Massachusetts Institute of Technology, USA

Hansjörg Mixdorff
BHT University of Applied Sciences, Berlin, Germany

Memory for spoken words includes memory of the speaker’s voice and other contextually specified phonetic properties, but there is also evidence that memory encoding is in terms of abstract phonological segments or features.

Broad Research Question: What are the roles of phonological and phonetic form in the perception, memory encoding, and production of spoken utterances?

Our focus: the phonology and phonetics of prosody (prominence and phrasing) in American English

Methods
We elicited imitations of pre-recorded, spontaneously produced utterances. Subjects heard each target utterance once and imitated it three times in succession.

Stimuli
32 utterances of spontaneous speech (7-15 words each) were extracted from 4 young adult, female speakers (Eastern dialect) from the American English Maptask corpus.

Pairs of Maptask speakers engaged in a cooperative task, with one speaker acting as a direction-giver conveying the detailed location of a path on one map to their partner who has a similar map that differs in some landmarks.

Sample utterance: “so you’re gonna go between the mill wheel and the mountain”

Imitators
Imitators were 10 young adult females (18-30 years old), students at the University of Illinois and native speakers of American English (Midland dialect), who were paid $10. Data from four (up to six) subjects are presented here.

Task
Target utterances were presented through headphones, with no text presentation. Imitators were instructed to listen to each utterance and then to “repeat the words and the way the utterance was said.”

Listener sensitivity. Listeners are sensitive to the acoustic cues to prosodic, and adapt their perceptual response to the cue pattern of an individual speaker.

Listening – production link. The phonetic detail perceived by the listener can influence their subsequent production. Evidence from speech shadowing studies shows that speakers imitate subphonetic, phonetic detail, at least when it is linguistically relevant.

Imitation of the canonical phonological form. Although listeners may perceive phonetic detail of a specific stimulus signal, in shadowing tasks speakers show a bias towards canonical forms. Canonical forms are shadowed at faster response times than reduced forms, and reduced forms are often restored to a canonical, full form in the imitated production.

Phonological prosodic labeling
Target utterances were independently transcribed by two of the authors (JC, SSH) using ToBI, with differences resolved through discussion to obtain a consensus ToBI labeling.

Phonetic encoding is in terms of abstract phonological segments or features.

Hypothesis 1: listeners are sensitive to the phonological detail of the prosody of spontaneous utterances, but what they store to guide future production is the phonological form.

Hypothesis 2: when asked to reproduce an utterance they have heard, speakers will replicate the broad phonological form of prosody, but using their own phonetic exponents.

Results
Descriptive statistics
Total number of Accent and Boundary labels over all utterances (approx. 180 words), for each speaker and for stimulus (target) utterances.

Statistical analysis of target-imitation similarity
Target-Imitation. Cohen’s Kappa statistic measures the agreement between the target utterance and each imitation. The analysis here is limited to the third of the three successive imitations subjects produced for each target utterance.

Cohen’s Kappa: Agre− Agre+
Agre− is observed agreement over all labels, Agre+ expected agreement (chance). Kappa scores of .41–.60 can be taken as moderate agreement, .61–.80 as substantial, and .81–1 as almost perfect agreement.

Agreement in phonological prosodic structure
Imitations show substantial agreement with target utterances for the location of Accents and Boundaries.

Agreement in acoustic-phonetic correlates of prosody
There are significant though moderate correlations between target and imitated utterances for F0 parameters and duration.

Pairwise agreement between subjects
Could agreement between target and imitation be driven by a “default” prosody, determined by the lexical and syntactic context? If so, agreement between pairs of subjects should be comparable to agreement between individual subjects and the target utterance.

Discussion
Results are consistent with Hypothesis 1:
Although listeners are sensitive to sub-phonemic detail in an utterance, they store the phonological structure.

Results are also consistent with Hypothesis 2:
Speakers use that phonological structure to guide imitations produced with their own phonetic exponents.

What is the role of subphonemic detail in this task?
• provides cues to prosodic structure, speaker specificity, etc.
• unlikely to lead to ‘re-tuning’ of phonemic category boundaries ( calorie 2009) because stimuli include only a few tokens from each speaker.

Significance of these results
• Although listeners are sensitive to and influenced by sub-phonemic detail, and shadowers can reproduce it, our imitators more reliably reproduce the linguistically-significant contrastive phonological structures of prosodic form.

Do shadowing and imitation tasks tap into different aspects of the perception-storage-production process?

Questions for Future Work
• Do these findings hold for other phonetic exponents? e.g., silence duration, F0 alignment
• How consistent and how detailed are individual speakers in their use of phonetic exponents? e.g., placement of irregular pitch periods within a syllable
• Are individual speakers more consistent for some exponents than for others?
• Are there significant differences from Imitation 1 to Imitation 3, as the memory of the original stimulus fades?

References

Broad research question: What are the roles of phonological and phonetic form in the perception, memory encoding, and production of spoken utterances? Our focus: the phonology and phonetics of prosody (prominence and phrasing) in American English.