Acoustic cues in perception of reduced speech

In everyday conversations, speech sounds are not produced in a canonical form, undergoing reduction and deletion (e.g. [fɹɛ̃ː] for “Friday night”, audio athttps://sites.arizona.edu/nwarner reduced-speech-examples/). To decode reduced forms, native listeners use various information, such as semantic/syntactic contextual information (Ernestus et al., 2002) and acoustic cues (Van de Ven et al., 2012). Warner et al. (2022) argue that acoustic cues predominate over other information in a study of listeners’ perception of words/phrases such as “he’s” vs. “he was” (which can both be realized [ɨz] in spontaneous speech). The current study investigates whether listeners use duration and the second formant (F2) as perceptual cues to this verb tense distinction by analyzing the stimuli and listeners’ responses from a previous perception experiment (Warner et al., 2022).

Duration of these past forms (e.g., “he was”) is expected to be longer than the corresponding present tense (e.g., “he is”) due to an additional consonant /w/ and the potential for “he is” to contract to “he’s”. If /w/ is present, it would lower the F2 (Ladefoged & Johnson, 1975). Ambiguous [ɨz] with low F2 or long duration might be perceived as past more often than a token with higher F2 or shorter duration. For this project, we measured formant and duration information from the 184 stimuli used in Warner et al. (2022). (Forms such as “we’re/we are” and “we were” were also used in this study, and the same predictions hold for duration and formants as for “is/was”.) Formant frequencies were converted to Bark scale distances F2-F1 and F3-F2 for speaker normalization (Syrdal and Gopal, 1986; Traunmüller, 1990). Low F2 is expected to result in smaller F2-F1 and larger F3-F2.

In the current study, two analyses were conducted. First, we investigated whether the stimuli show acoustic differences in the predicted direction despite the potential for the tense distinction to be neutralized in reduced speech. That is, do the current stimuli show smaller F2-F1, larger F3-F2, and longer duration in past tense forms, or are is/was and are/were statistically the same in this reduced speech (reduced to the point of being homophonous)? Second, we analyze whether the formant and duration measures correlate with listeners’ identification of tense. This second analysis addresses whether listeners use these acoustic cues to distinguish tense even in highly reduced speech where those cues may not be reliably present.

In the first analysis, across all three acoustic measurements (duration, F2-F1, and F3-F2), past stimuli did not differ significantly from present ones, indicating that the recording stimuli are not acoustically different, indicating that past and present forms are statistically homophonous in this conversational speech due to reduction. In other words, “he’s/he is” and “he was” are statistically homophonous in this conversational speech because of reduction, as are “we’re/we are” and “we were.” However, the non-significant trends were in the predicted directions, suggesting that some of the more clearly produced tokens do contain these potential acoustic cues. In the second analysis, among singular (“is/was”) verbs, listeners were significantly more likely to perceive past when F2-F1 was smaller. Among plural verbs (“are/were”), listeners recognized stimuli with longer duration as past tense significantly more often. Thus, listeners used the formant cue to perceive “is/was” and the durational cue to perceive “are/were” even though the stimuli were too reduced to contain these cues consistently. Even when speech is highly reduced, making these function words homophonous, listeners still use the acoustic cues to comprehend speech. The results complement Warner et al.’s (2022) finding that listeners prioritize acoustic cues over semantic/syntactic information in perceiving reduced speech even when reduction creates misleading acoustic cues.

References

Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81(1–3), 162–173. https://doi.org/10.1006/brln.2001.2514

Ladefoged, P., & Johnson, K. (2011). *A Course in Phonetics* (6th ed). Wadsworth. https://doi.org/10.1080/07268600600885494

Syrdal, A. K., and Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of Acoustical Society of America*, 79(4), 1086-1100.

Traunmüller, H. (1990). 'Analytical expressions for the tonotopic sensory scale. *Journal of Acoustical Society of America*, 88, 97-100.

Van De Ven, M., Ernestus, M., & Schreuder, R. (2012). Predicting acoustically reduced words in spontaneous speech: The role of semantic/syntactic and acoustic cues in context. *Laboratory Phonology*, 3(2).

Warner, N., Brenner, D., Tucker, B.V., and Ernestus, M. (2022). Native Listeners' Use of Information in Parsing Ambiguous Casual Speech. *Brain Sciences*, 12(7):930. https://doi.org/10.3390/brainsci12070930