Subphonemic detail, resyllabification, and Spanish speech segmentation

Helena Escalera and Miquel Simonet
University of Arizona
Second Language Acquisition and Teaching minor

In speech, words flow from one to another without pauses that delimit where they begin or end. In Spanish, words can both start and end with /s/, and they can also start and end with a vowel. This creates contexts that would seem to lead to ambiguity: *dices eso* and *dice seso*, for instance, could be homophonous, /diθeseso/. Furthermore, according to some phonological accounts, word-final consonants are resyllabified with the following word if the latter starts with a vowel (e.g., [1]). A strong interpretation of this claim predicts that the word-final /s/ in *dices eso* has identical phonetic characteristics as the word-initial /s/ in *dice seso*. Are these sequences indeed phonetically identical? How does word structure affect phonetic detail? If phonetic differences exist, are they exploited by listeners? Do fine acoustic differences, if found, impact speech comprehension, including word segmentation?

Speakers have been found to produce fine phonetic differences in their pronunciation of consonants preceding or following word junctures ([2]–[5]), and research suggests that listeners capitalize on these cues for disambiguation ([2], [6]). The present study revisits this theme and includes four experiments, two of them on speech segmentation.

A read-aloud production task with 10 L1 Spanish speakers from Spain focusing on ambiguous utterances examined whether different word affiliations led to phonetic differences in production. Speech recordings were acoustically analyzed for /s/ duration, /s/ spectral center of gravity, and vowel duration. Results revealed that /s/ duration was a reliable predictor of word affiliation, with word-initial /s/ being longer than word-final /s/. To investigate whether listeners exploit such duration differences, perception data were collected via a two-alternative forced-choice task (2AFC). In each trial in this task, participants (60 L1 Spanish listeners from Spain) were auditorily presented with one auditory stimulus ([diθes.eso]) and two written options (e.g., *dices eso* or *dice seso*). The findings confirmed that listeners' choices, while displaying low accuracy, were significantly above chance.

The remaining two experiments focus on speech segmentation. Sixty listeners participated in a word-monitoring task with the same materials as in the perception experiment above. In each trial, before playing the auditory stimuli, listeners were shown a word to monitor for (<seso>, <eso>, etc.) and they were instructed to press the space bar as soon as they recognized the word in the auditory stimulus, if at all present. Listeners were randomly played either a matching stimulus (<eso> [diθes.eso]), a mismatching stimulus (<eso> [diθe.seso]), or a distractor (<eso> [diθes.oso]). This was a go-nogo task for which we measured response times and accuracy. As expected, participants were found to react to both matching and mismatching trials (but not distractors). Most importantly, however, they were significantly faster when responding to matching than mismatching trials. This finding suggests that listeners exploit phonetic differences in the signal not only in perceptual identification but also speech segmentation. The last experiment was a fragment-monitoring task in which auditory stimuli had been cut to render them short and meaningless (e.g., rather than [diθes.eso], we had [θes.eso]), and participants were instructed to monitor for syllable-size fragments rather than word-sized fragments (e.g., <se>). Listeners accurately responded to matching and mismatching trials, but not distractors. However, there were no response-time differences between matching and mismatching trials in this experiment.

Evidently, listeners' phonological knowledge is lexically based, and their exploitation of subphonemic detail during speech comprehension is optimized for word-based speech segmentation. We conclude with a discussion of the theoretical relevance of our findings.

[1] S. Colina, *Spanish Phonology: A Syllabic Perspective*. Washington, DC: Georgetown University Press, 2009.
[2] K. Ito and W. Strange, "Perception of allophonic cues to English word boundaries by Japanese second language learners of English," *J. Acoust. Soc. Am.*, vol. 125, pp. 2348–2360, 2009.
[3] P. Strycharczuk and M. Kohlberger, "Resyllabification reconsidered: On the durational properties of word-final /s/ in Spanish," *Lab. Phonol.*, vol. 7, p. 19, 2016.
[4] M. Jiménez-Bravo and J. M. Lahoz-Bengoechea, "Durational cues to resyllabification in Spanish," *Loquens*, vol. 10, no. 1–2, p. e099, 2023.
[5] J. M. Lahoz-Bengoechea and M. Jiménez-Bravo, "Ambisyllabic characteristics of Spanish resyllabification: Beyond durational cues," *EFE*, vol. 33, pp. 153–168, 2024.
[6] J. M. Lahoz-Bengoechea and M. Jiménez-Bravo, "Spoken word boundary detection in ambiguous resyllabification contexts in Spanish," vol. under review, 2021.

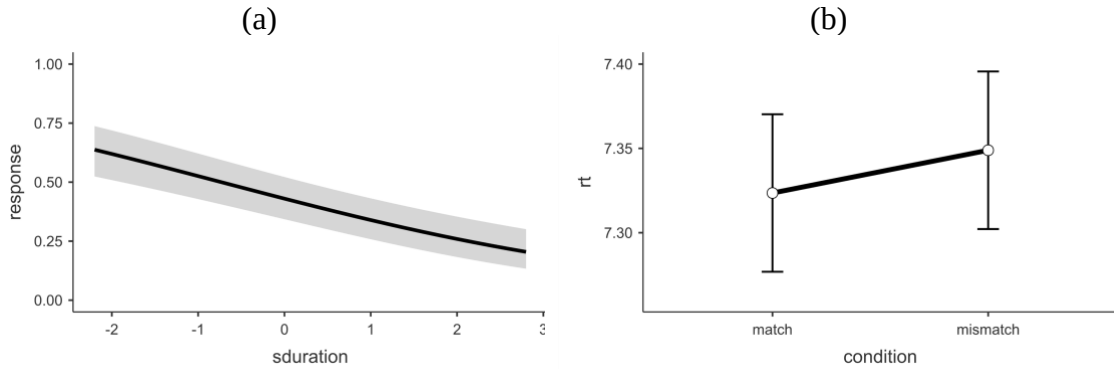

Figure 1. (a) Regression between duration of /s/ (*z*-scored) and probability of responding "word-initial vowel" (that is, word-final /s/), from experiment 2 (two-alternative forced-choice perceptual identification); and (b) estimated marginal means (means and 95% CI) of multilevel regression model with reaction times as response and condition (matching, mismatching) as fixed predictors, from experiment 3 (word-monitoring task).