



OPEN ACCESS

EDITED AND REVIEWED BY
Xiaolin Zhou,
Peking University, China

*CORRESPONDENCE

Christopher Carignan
✉ c.carignan@ucl.ac.uk
Joseph V. Casillas
✉ joseph.casillas@rutgers.edu
Eleanor Chodroff
✉ eleanor.chodroff@york.ac.uk
Georgia Zellou
✉ gzellou@ucdavis.edu

SPECIALTY SECTION

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

RECEIVED 30 November 2022

ACCEPTED 09 December 2022

PUBLISHED 23 December 2022

CITATION

Carignan C, Casillas JV, Chodroff E
and Zellou G (2022) Editorial: Fuzzy
boundaries: Ambiguity in speech
production and comprehension.
Front. Commun. 7:1112753.
doi: 10.3389/fcomm.2022.1112753

COPYRIGHT

© 2022 Carignan, Casillas, Chodroff
and Zellou. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Editorial: Fuzzy boundaries: Ambiguity in speech production and comprehension

Christopher Carignan^{1*}, Joseph V. Casillas^{2*},
Eleanor Chodroff^{3*} and Georgia Zellou^{4*}

¹Department of Speech, Hearing and Phonetic Sciences, University College London, London, United Kingdom, ²Department of Spanish and Portuguese, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States, ³Department of Language and Linguistic Science, University of York, York, United Kingdom, ⁴Department of Linguistics, University of California, Davis, Davis, CA, United States

KEYWORDS

speech, production, perception, sound change, data science, ambiguity

Editorial on the Research Topic

Fuzzy boundaries: Ambiguity in speech production and comprehension

Language is a system of discrete and abstract elements. Yet, we can rarely (if ever) identify predictable, linear, or clear one-to-one relationships between the speech signal and linguistic categories. Rather, the relationship between speech and language consists of fuzzy boundaries between categories and myriad sources of ambiguity. Early research may have attributed much of this ambiguity to equipment error, less than ideal recording conditions, population under-sampling, or other sources of spurious behavior in the data. Upon closer inspection, however, many researchers have identified a richness and systematicity in the fuzzy mapping from speech to language: ambiguity may play a crucial role in the development, evolution, and realization of language itself. Listeners may benefit from acoustic variability when learning phonological categories and generalizing from them across phonological contexts. Ambiguity about the source of acoustic effects can serve as a catalyst of sound change actuation. Speakers adapt their productions when the environment could make their speech ambiguous to listeners. Gradiency in linguistic representations could allow greater flexibility for listeners to adjust to cross-speaker and cross-situational variation.

The current research era presents opportunities for tackling this difficult topic in ways that have never before been possible or in some cases even imaginable. Recent trends and techniques involving co-registration of multiple data streams allow us to disentangle the articulatory source of observable acoustic effects of vocal tract dynamics, in spite of complicated many-to-one or even many-to-many articulatory-acoustic mappings. The interdisciplinary and trans-global collaborative research that is becoming increasingly popular in our virtual age encourages a wide range of interpretations and strategies for dealing with ambiguous data. Cutting edge machine learning techniques and statistical approaches can help dis-ambiguate fuzzy data patterns to uncover meaningful underlying structure. Virtual experiment platforms that have

flourished in recent times can be used to collect participant response data at a scale that was previously unthinkable, allowing novel insight into group-level patterns that characterize the cognitive processing of potentially ambiguous speech signals.

Rather than consider the ambiguous relationship between speech and language as mere noise, or even avoid it entirely in study design and the interpretation of study results, this Frontiers Research Topic seeks to highlight ambiguity itself as a central aspect of the research and object of observation. Our call for papers resulted in 11 original contributions that represent a range of perspectives within the topic of ambiguity in speech production and perception. The articles in this collection all present empirical research that centered around four major themes.

The first theme covers research in perceptual cue-weighting and cue-trading. Four contributions fall under this theme. Guo and Kwon examine the relation between stop aspiration and post-stop F0 in the production and perception of the laryngeal contrast in Mandarin Chinese. They find variations in F0 perturbations across tones which they explain as due to interactions between aerodynamic forces, vocal fold tension, and tonal targets. Yet, in perception, listeners associate high F0 with aspirated plosives. The contribution of this paper for fuzzy boundaries is a detailed exploration of mismatches between production and perception for contrasts that involve complex laryngeal gestures.

Phillips examines the time course for how listeners use anticipatory coarticulation on /s/ for an upcoming rhotic segment. Coarticulation has been considered by some as contributing to “noise” in the speech signal, variation that makes sound categories more “fuzzy”, yet this paper finds that listeners use coarticulatory variation immediately, as soon as those cues become available, and further that immediate integration strategies were strengthened when the coarticulatory cues of retraction were stronger and when they were more predictable.

Yu identifies top-down influences of the listener’s perception of the talker’s persona on the stop voicing contrast. The combination of the listener’s gender and the listener’s perception of the speaker’s socio-indexical properties, such as attractiveness, gayness, or confidence, significantly influences stop categorization, even for the same acoustic stimulus. Perceptual boundaries can therefore be a bit blurred before taking into consideration the listener’s in-the-moment perception of the speaker along various socio-indexical dimensions.

The final contribution under this theme comes from Lo in a study exploring the role of F0 as a cue to stop voicing in non-tonal and tonal languages. Lo analyzes the production and perception of stops in Mandarin-English bilinguals. F0 is considered a secondary cue to voicing in English, but serves as a critical acoustic correlate of tone in Mandarin. Participants completed two tasks: a reading production task and a two-alternative forced-choice identification task using stimuli drawn

from a bilabial stop continuum in which VOT and F0 were manipulated orthogonally. The results of the production task show that post-stop F0 is consistently higher for voiceless stops when compared with voiced stops. This F0 disparity is larger in the bilinguals’ English production than in Mandarin. Lo ascribes this difference to post-stop F0 receiving more weight in English. The perception data also reflect this weighting. Overall, stimuli with higher post-stop F0 are more likely to be identified as voiceless, but the probability of a voiceless response is even higher when the participants believe they are hearing English words. This study underscores a general flexibility, present not only in perceptual boundaries, but also in bilingual cue-weighting strategies, when producing and perceiving similar contrasts in typologically different languages.

The second theme of this collection targets the role of acoustic and/or perceptual ambiguity in sound changes in progress. Bi and Chen identify incomplete neutralization of two falling tones in Dalian Mandarin Chinese, tones 1 and 4. Though the phonetic form of these tones are typically transcribed with the same Chao tone numerals of 51, this study finds subtle but statistically significant differences in F0 contour and velocity profile across two generations of speakers. Lexical frequency and homophone neighborhood density also interact with the phonetic realization of each tone. These findings indicate incomplete neutralization, with additional fuzziness in the exact phonetic instantiation coming from influences of lexical frequency, homophone neighborhood density, as well as their interactions with speaker generation.

Zhang et al. evaluate the production-perception link in two marginal contrasts of Chicagoland English: [ɑ—ɔ] (“cot-caught”) and [ʌi—ai] (“writer-rider”). The former represents a phonological merger in this variety, and the latter a phonemic split. Individuals from this speech community provided production data by reading cot-caught and writer-rider pairs embedded in sentences and in isolation. The perception data was derived from ABX and two-alternative forced-choice tasks. Zhang et al. provide evidence suggesting that the production/perception link may follow a different trajectory depending on the type of sound change in question, i.e., a phonological merger vs. a phonemic split. This study highlights the manner in which data from fuzzy contrasts can contribute to our understanding of sound change and language acquisition processes.

Zahner-Ritter et al. investigate the form and function of three rising-falling contours—L + H*, (LH)*, and L* + H—found in German *wh*-questions across Northern and Southern varieties of German. The production results indicate reasonable separation among contours, but also some degree of fuzziness, especially for Southern German speakers with respect to the L + H* and (LH)* contrast. The perception results reveal very distributed and somewhat fuzzy meaning associations for each of the contour types: for both dialects, L + H* and L* + H accents are largely interpreted as information-seeking, whereas

(LH)* has a more distributed meaning, and is much more likely to be interpreted in both dialects as a negative attitude or aversion.

The **third** theme of this collection involves perceptual adaptation to speech that is variable in both time and space. Temporal boundaries of speech perception may be fuzzy: speech unfolds in time and variations in the duration and coordination of temporal events can affect how speech is perceived. Inappropriate gaps between syllables is a core diagnostic feature of childhood apraxia of speech (CAS), yet no baseline exists in the literature concerning how adults perceive inappropriate gaps in the speech of typically developing children. O'Farrell et al. address this issue by investigating the perceptual threshold for inter-syllabic temporal gaps from 84 adult listeners, using speech samples from typically developing children digitally altered to insert gaps. They find that 80% accuracy in detecting inappropriate gaps occurs for intervals between 100 and 125 ms, and 90% accuracy for intervals between 125 and 150 ms. This finding provides the first evidence of the perceptual limen of syllable segregation, which can provide a threshold for a therapy goal for treatment of CAS.

"Spatial" boundaries of speech perception may also be fuzzy: perceptual boundaries between categories are malleable and can shift as speech production traverses through myriad domains of sensory input. Previous studies have shown that repeated exposure to a particular acoustic stimulus can shift a listener's perceptual boundary toward that stimulus, a phenomenon known as selective adaptation. Ito and Ogane use orofacial skin stretching to investigate whether the category boundary between / ϵ / and /a/ is similarly affected by repeated *somatosensory* exposure. They find that exposure to a particular somatosensory stimulus (in this case, pulling the skin upward in a manner consistent with the production of / ϵ /) results in selective adaptation in the same way as acoustic exposure: participants perceive /a/ more than / ϵ / after repeated somatosensory training, suggesting that the perceptual boundary is shifted toward the repeated exposure stimulus, / ϵ /. These results may simulate the natural sensory pairing which occurs during speech production and, thus, support the idea that somatosensory inputs contribute to the formation of sound representations.

The **fourth** theme deals with the perception-production link specifically by looking at "own speech". Two contributions examine how listeners' perception of their own speech can shed light on questions of speech representation. This line of research stems from the fact that speakers are generally more accurate and efficient when processing familiar accents and voices. Cheung and Babel examine the own-voice benefit utilizing Cantonese-English bilinguals' productions of minimal pairs to generate personalized two-alternative forced-choice perception tasks. That is, the bilingual listeners identify instances of Cantonese words which were manipulations of their own voice, as well as productions of other speakers. Cheung and Babel find that the bilinguals are more successful identifying instances of their own

manipulated voice than when they are presented with tokens from other speakers, even when said speakers maintain the same degree of acoustically contrastive minimal pairs. Cheung and Babel conclude that phonological contrasts may be primarily shaped by the distributions of our own phonetic realizations. This study highlights the variability present in bilingual speech for producing contrasts. Importantly, it sheds light on how this variability relates to perception, particularly with regard to our understanding of how familiarity aids speech processing, even in presence of a more ambiguous signal.

Finally, Baxter et al. provide a partial replication study in which they evaluate the claim that one's own speech processing can be affected when interacting with L2 speakers. Specifically, this thread of research suggests that processing costs due to increased cognitive effort can affect one's memory of a conversation. In their study, L1 English speakers interact with other L1 English speakers as well as L2 English speakers of intermediate and advanced proficiency. The results suggest speakers display more accurate recall when interacting with L1 speakers in some conditions. The authors conclude that recall accuracy may be modulated by the degree of processing costs incurred and, in turn, result in fuzziest lexical/semantic representations of their own speech.

The contributions to this Research Topic provide wide-ranging and varied perspectives on ambiguity in speech production and perception. The contributions open questions and provide many ripe avenues for future research in this area.

Author contributions

CC, JC, EC, and GZ contributed equally to the conceptualization, writing, and article summaries of the editorial. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.