

Removing the disguise: the matched guise technique and listener awareness

Kyler Laycock¹ and Kevin B McGowan²

¹The Ohio State University

²The University of Kentucky

Author Note

Kyler Laycock  <https://orcid.org/0000-0002-3731-0841>

Kevin B McGowan  <https://orcid.org/0000-0002-8214-1901>

Correspondence concerning this article should be addressed to Kyler Laycock, Email:
laycock.21@buckeyemail.osu.edu

Abstract

Sociophonetic perception is often studied using versions of the matched guise technique. Linguists using this technique appear united in the methodological assumptions that participants believe the manipulation and that this belief influences perception below the level of introspective awareness. We report an audiovisual matched guise experiment with a novel ‘unhidden’ instruction condition. The basic task is a replication of the Strand effect (?; ?). Participants in the ‘unhidden’ condition were instructed that the man or woman in the photo did not represent the voice they were listening to. Participants in both guises exhibited the Strand effect to nearly numerically identical extents. This result suggests that participants need not believe a link exists between a voice and a purported social category for visually-cued social information to influence segmental perception. We explore the implications of this result for the MGT and for theories of social awareness and speech perception more broadly.

Removing the disguise: the matched guise technique and listener awareness

Introduction

A great deal of attention has been paid in the phonetics and sociophonetics literatures to the perception of the voiceless fricatives [] and [s] in English. To a first approximation, these fricatives differ in the distance between the point of lingual articulation and the teeth, which give them their characteristic sibilance Shadle (?). English [s] has a short resonating chamber behind the teeth; it is typically produced by holding the tongue blade near enough to the alveolar ridge to cause turbulent airflow. English [] has a comparatively larger resonating chamber; it is typically produced with a more posterior, palato-alveolar tongue position and lip rounding both of which serve to reinforce this posteriority. But listeners do not perceive via {++this type of++} {»KM«} first approximation{++; we are sensitive to fine phonetic details far beyond these gross, categorical, differences++} {»KM«}. Indeed, these two fricatives have been exciting to researchers precisely because of the sensitivity listeners bring to their perception and how that perception interacts with both linguistic and social knowledge.

Coarticulatory and Social Information Influence []-[s] perception

{»KL: ok, this is the specific section the editors didn't like I think vis-a-vis "rewriting the introduction to state more strongly why this study is important to sociolinguistics, and not mainly interesting to cognitive linguists"«} Listeners are sensitive to articulatory mismatches between the fricatives []-[s] and neighboring sounds. Whalen (?) conducted a series of experiments to investigate listeners' responses to articulatory mismatches in synthetic speech. Overall, the result of these investigations was that subcategorical phonetic mismatches slow phonetic judgments. In onset position, in isolation, or in coda position, misleading coarticulatory information inhibited reaction times. Listeners, Whalen cautions in the conclusion, are sensitive to articulatory patterns that are below the level of conscious awareness and not available to direct experimenter scrutiny. While listeners will readily fill-in missing or ambiguous information, the presence of actively *conflicting* articulatory information is inhibitory.

A commonly used methodology involves the creation of synthetic fricative continua.

These continua have endpoints in prototypical examples of [] and [s] with some number of equal-sized acoustic steps generated, synthesized, or even mixed between these. Somewhere in the middle of such a continuum will be fricative-like noise that is ambiguous as to category membership: not clearly a [] and not clearly an [s]. ? paired a continuum from [] (2.9 kHz) to [s] (4.4 kHz) with synthetic [æ] vowels to form CV pairs. May found that listeners perceived a higher proportion of the fricative continuum as [] when paired with vowel stimuli from a smaller vocal tract. The logic here is that smaller resonating chambers between the lingual articulation and teeth will have a higher mean frequency than larger resonating chambers. Listeners' use of apparent vocal tract size in perception reflect their knowledge of this variation (?).

Mann and Repp (?) replicated this finding, extending it to natural productions of vowels spoken by a male or female-identified talker. Similar to May's results with simulated vocal tract size, Mann & Repp found a higher proportion of the fricative continuum was heard as [] when paired with the speech of the female talker. This early work, like others of the period (?), theorized size as being a relatively deterministic feature of talker sexual dimorphism. One consequence of this view is that gender-related variation in the speech signal is considered mechanistic, universal, and following from purely physical laws. Vocal tract size is presumably not available for individual performance and so listener knowledge of this variation can be correspondingly simple. Vocal tract size may influence perception, but it does so implicitly, automatically, and below the level of introspective awareness.

{++ Mann and Repp (?) also replicated and extended previous work (?; ?) demonstrating that listeners report hearing more of the synthetic fricative continuum as [s] when followed by a rounded vowel quality such as English [u] than when followed by an unrounded quality such as [i] or [a]. Listeners experience the fricative continuum differently in the presence of anticipatory coarticulation. The presence of nasal coarticulation on a vowel similarly allows listeners to make a lexical decision between words like *bend* and *bed* as soon as that information is present in the acoustic signal (?; ?). Mann & Repp's participants in this study experienced auditory evidence of posteriority in the ambiguous portion of the fricative continuum as the presence of coarticulation