

Deliberative control in audiovisual sociolinguistic perception*

Kathryn Campbell-Kibler 

Department of Linguistics, Ohio State University, Columbus, OH, USA

Correspondence

Department of Linguistics, Ohio State University, 1712 Neil Avenue, Columbus, OH 43210-1219, USA.
Email: campbell-kibler.1@osu.edu

Abstract

Cognitive models of sociolinguistics must support a wide range of goal-oriented behavior (e.g. Eckert, 2000a) without suggesting unrealistic levels of deliberative control on the part of speakers. The current study investigates the limits of deliberative control in audiovisual face-voice perception. Perceivers evaluated co-present recorded speech and static face pictures, rating the stimuli on the scales ‘accented’ and ‘good-looking’ in one of three conditions: as a combined voice and face; evaluating the face while ignoring the voice; and evaluating the voice while ignoring the face. Perceivers’ ability to ignore social information from a face or voice upon instruction are taken as indicative of deliberative control in social evaluation. The results suggest that deliberative control and evaluative relevance both play a role in perception, but that available social information is difficult to ignore completely. They further suggest an asymmetry making voices more difficult to ignore than static faces and support a model of sociolinguistic perception and evaluation as a function of multiple competing processes under varying degrees of deliberative control.

KEYWORDS

implicit and explicit processes, sociolinguistic cognition, sociolinguistic perception, sociolinguistic variation

*Special thanks to Elizabeth McCullough for sharing stimuli and assistance with the preparation of the study and early versions of this manuscript. Also many thanks go to Katelyn Seitz, Rachel Castle, Samantha Arthurs, Michael de Matto, and Lauren Boyce for their excellent work collecting data on this project. Audiences at Sociolinguistics Symposium 21 and NWAV 44 contributed helpful comments and suggestions, as did the editors and several anonymous reviewers. All errors that remain are my own.

1 | INTRODUCTION

Sociolinguistic variation as a field examines the relationship(s) between complex social dynamics on the one hand and rapidly occurring linguistic cues on the other. The disconnect between apparently conscious social processing and apparently automatic linguistic processing has introduced an ongoing tension in attempts to model sociolinguistic cognition. Labov (1966) addressed this tension in two major ways. First, he posited a distinct cognitive module for sociolinguistic processing, external to the grammar (Labov, 1966, 1993; Labov et al., 2011), which depended at least partially on executive function-type resources, such that its ability to function was compromised by diminished attention. Second, he limited the evaluation and control powers of this module to a single dimension, namely socioeconomic prestige. The third wave of sociolinguistic variation studies (Eckert, 2012), along with work elsewhere in sociolinguistics, linguistic anthropology and language attitudes, challenged multiple aspects of this model, expanding the complexity required of sociolinguistic processing systems and modeling speaker/hearers as highly agentive.

Sociolinguists need a better understanding of this tension in order to develop adequate cognitive models of the full range of sociolinguistic behavior. Much of the resistance to third wave work, particularly in the early years, centered around concerns that the portrayal of the speaker/listener as an agent went improbably far. A better understanding of the deliberative/automatic interplay would help to resolve the apparent contradiction between speakers' complex and skillful use of linguistic variation, and the clear limitations they display in their ability to verbally report on the details of their own and others' linguistic behavior (see, e.g. Labov, 1966, p. 329).

The current study examines sociolinguistic perception, which allows more experimental control than is feasible in a production study. In addition, much of the work on sociolinguistic processing and the sociolinguistic monitor has been carried out through sociolinguistic evaluation tasks, making perception a logical place to start. Deliberative control was manipulated by offering information which could influence the evaluation, but which was verbally explained as irrelevant, probing whether participants were able to ignore this information when asked to do so. To the extent that participants could exert deliberative control over their perception and evaluation, information described as irrelevant to the task should not contribute to the resulting evaluation. To the extent that automatic processes contribute, this information should still influence participant responses.

Across three conditions, listeners provided ratings indicating how accented or good-looking they found stimuli presented as face-voice pairs. In the speaker condition, the face and voice were presented as a composite speaker and rated as a coherent unit. In the other two conditions, listeners were told that the visual and audio stimuli were of distinct people, and were asked to rate only the face (face condition) or only the voice (voice condition). The results indicate that listeners exhibit a great deal of deliberative control in choosing what information to favor, but that in most cases they are unable to entirely eliminate the irrelevant information. They further suggest that the social evaluation task (rating on the voice-focused accented scale or face-focused good-looking scale) strongly influences listener dependence on the two modalities in a way that interacts with explicit instruction.

2 | DELIBERATIVE AND AUTOMATIC PROCESSES IN SOCIOLINGUISTIC PERCEPTION

Scholars of language beliefs have theorized the multiple dimensions of awareness or knowledge in a few ways. Language attitudes research has grappled for decades with a distinction between 'overt' and 'covert' attitudes. Typically a method for distinguishing types of experimental tasks, 'overt' methods

are those which allow participants to know that attitudes towards language varieties are the subject of the research (Kristiansen, 2010a). A prototypical example would be a question such as 'Do you intend to teach your children French?' 'Covert' approaches, in contrast, seek to prompt behavior that is shaped by language attitudes through tasks that conceal or downplay this goal for participants, such as speaker evaluation tasks. Kristiansen (2009, 2010b) highlights the importance of the overt/covert distinction, finding that the rankings of different Danish varieties by Danish teens differ substantially when collected through these different avenues.

Some theorists have delved more specifically into different types of awareness. Silverstein (1981) laid out a set of principles to predict which aspects of sociolinguistic language behavior are more or less amenable to metalinguistic report. Preston (1996, 2016) explored the characteristics of the introspective knowledge that speakers are able to report and presented a model of the process by which language regard responses are generated, offering a distinction between automatic processes on the one hand and working memory on the other, along with a 'cognitorium', an independent module in which ideological relationships between concepts are stored.

The field of sociolinguistic variation has grappled with questions of agency, consciousness, awareness, and control since its inception (for an in-depth discussion, see Babel, 2016), including the struggle to develop clear and shared definitions of these terms. Often models of awareness models have melded multiple dimensions, such as metalinguistic commentary, situational use, and speaker evaluation, into a single dimension of 'more' versus 'less' awareness, for example, the widespread indicator/marker/stereotype division of Labov (1972). The most widespread theory of sociolinguistic cognition is the sociolinguistic monitor (Labov, 1993; Labov et al., 2011; Levon & Fox, 2014; Wagner & Hesson, 2014), a relatively high-effort cognitive module distinct from the grammar, which both monitors a speaker's own speech for social goal alignment and provides social assessment of the speech of others. In Campbell-Kibler (2016), I argued that sociolinguistic processing can be modeled without a distinct monitor, using independently motivated systems, namely linguistic grammar, self-regulation and person perception, with limited links between. This move would allow for a wider range of awareness or control levels for different sociolinguistic processes, but little is known about how awareness and control operate in sociolinguistic behavior.

A crucial first step is to name precisely what is meant by these terms. I will distinguish first between agency on the one hand and consciousness, awareness, and control on the other. Agency in sociolinguistics is most commonly used at the level of interaction, to refer to a speaker's ability to make choices and pursue goals. I take it thus to be independent of any cognitive models. A speaker whose sociolinguistic behavior is in alignment with their interactional goals may be taken to display agency, without any claims made as to the cognitive processes facilitating that behavior.

In contrast, consciousness, awareness, and control do distinguish different representations or processes. Scholars have theorized for centuries that some cognitive processes happen under the direction of a consciously aware system which demands attentional and memory resources, while others happen without its input or even knowledge (Frankish & Evans, 2009). Two current approaches suggest either that there are two types of systems, one more deliberative and another more automatic (Evans, 2008, 2019), or that a given task, like person perception, is carried out by a single system which proceeds iteratively, depending less on deliberative control towards the beginning of the process and allowing for more in later stages (Cunningham, Zelazo, Packer, & Van Bavel, 2007; Van Bavel, Jenny Xiao, & Cunningham, 2012). One challenge noted in this literature is that there are many characteristics which may intuitively distinguish processes, for example fast versus slow, associative versus propositional, or low effort versus high effort, but these dichotomies do not consistently align with one another when applied to specific processes (Evans, 2008). 'Consciousness' has been used to refer to many different aspects of these questions, as well as the explanatory post-hoc processes which

provide individuals with plausible explanations for their own behavior (Evans, 2008). I propose that for the purposes of sociolinguistic research, the term *consciousness* or *conscious* applies best to the broad set of questions, but not as a usable classifying term.

For practical discussion, I will focus here on the remaining two terms: *awareness* and *control*, which are operationalized with verbal report. *Introspective awareness* refers whether an individual is able to verbally report on a given piece of knowledge or a process. Thus, a speaker who verbally reports that they noticed and evaluated a specific form is displaying *introspective awareness* of some aspects of their sociolinguistics behavior. Similarly, *deliberative control* is the ability of an individual to engage in or refrain from engaging in a behavior when verbally asked to do so, or when they have verbally reported the intention to do so. Speakers may be able to *deliberatively control* processes to varying degrees, or be able to *deliberatively prevent* a process which does not require *deliberative control* to maintain, such as breathing. *Deliberative control* requires some amount of limited resources like attention and working memory and evokes a subjective sense of effort or difficulty under low-resource conditions like fatigue. I will use *automatic* as a contrast to *deliberative control*, meaning processes which can proceed without *deliberative effort* or, more strongly, those which cannot be prevented *deliberatively*, such as reading text in a familiar language (Stroop, 1935).

Linguistic processing for semantic comprehension is typically theorized as *automatic* and largely inaccessible to *introspective awareness* (Fodor, 1983). The socially goal-directed use of language, in contrast, has been characterized by many as dependent on *deliberative control*, even exclusively so. This is expressed at times explicitly, as in ‘We argue that sense of national or regional identity is necessarily *conscious*, and that unconscious accent accommodation falls below the level of *conscious* sense of identity. Brulard and Carr (2013, p. 151)’ At other times, it is expressed indirectly, as in ‘It has been noted that centralized diphthongs are not salient in the *consciousness* of Vineyard speakers. They can hardly therefore be the direct objects of social affect (Labov, 1972, p. 40)’. Both of these quotes reflect a widely held implicit model that positions social processing and goal pursuit as necessarily under *deliberative control* and available to *introspective awareness*. Furthermore, they suggest that social processing is unable to access or influence the *automatic processing* which underlies linguistic behavior.

This division of the *automatic linguistic* and the *deliberative social* has been challenged the third wave of variation research, which has shown that speakers and listeners can use linguistic variation to invoke highly complex and situationally dependent social meanings (see, for example, Eckert, 2000a). Third wave theorists have appealed to the semiotic concept of *indexicality* to theorize the links between linguistic cues and context which support these complex sociolinguistic tasks (Peirce, 1901; Silverstein, 1976). Through repeated exposure, speakers form associations between specific linguistic features and non-linguistic contextual elements, including speaker identity, affective stances, speaker-listener role alignments and others. Creative uses of *indexical links* are able to forge new associations which, over time, can solidify into established *indexical links* available for easy use (Silverstein, 2003).

Studies in the third wave have found that speakers are capable of producing highly nuanced linguistic styles which correlate with personal identities and ideologies (Eckert, 2000b), with situational features such as about-to-be-present interlocutors (Hay, Jannedy, & Mendoza-Denton, 1999), and with interactions between multiple social categories (Podesva & Van Hofwegen, 2015). Work in the perception of socially meaningful variation has likewise shown that listeners use overall linguistic style to contextualize the social meaning of variation (Pharao, Maegaard, Møller, & Kristiansen, 2014), that they incorporate extralinguistic information into their perception and evaluation process (Campbell-Kibler, 2010), that their own preferences or goals can influence their interpretation of a given linguistic cue (Campbell-Kibler, 2008), and that listeners are able to use features in person perception that they do not report having noticed (De, 2017). This work has established with certainty the agency of

sociolinguistic actors, and in so doing, opened up many questions about the awareness and control they exhibit over the processes involved. In particular, they suggest that at least some sociolinguistic behavior is likely to occur through processes inaccessible to awareness and control, given the attentional limits of human cognition and the inability of speakers to verbally articulate more than a handful of the resources they are able to interpret and deploy. Some scholars have presented models including such processes (Johnson, 2006; Sumner, Kim, King, & McGowan, 2014).

The existence of automatic sociolinguistic processes is further supported by a growing body of work in sociophonetics, which has shown that linguistic processes are susceptible to influence by social information, without speakers being aware of these effects. For example, Niedzielski (1999) presented Michigan listeners with recorded sentences from a Michigan talker whose speech exhibited Canadian Raising, the raising of the nuclei of /ay/ and /aw/ before voiceless obstruents. Despite the name, Canadian Raising is found in Michigan, but commonly believed by Michigan speakers to be limited to Canada. Listeners were asked to select from a synthesized vowel continuum the token which best matched that of the speaker. Some listeners were told the speaker was from Michigan while others were told she was from Canada. This manipulation influenced responses, such that listeners selected less-raised tokens when they knew they were listening to a fellow Michigander, only selecting the more closely matched raised tokens when told the speaker was Canadian. It appears that participants' ideological beliefs regarding Michigan speech are able to override their direct linguistic perceptions, during either perception or selection.

In Niedzielski's study and other similar work, we see evidence of sociolinguistic influences on behavior without explicit verbal instruction or report of intention. It is difficult to conclude with confidence, however, that this behavior is inaccessible to awareness or control. One tool for exploring the deliberative status of a behavior is to provide a participant with explicit verbal information countering the behavior in question. In Niedzielski's case, the information provided to the participants was presented as true information about the speaker. Thus its influence on the vowel selection task was in alignment with the participants' reportable beliefs and their understanding of the experimental task. If the same behavior emerges when their verbally reportable knowledge is in conflict with the behavior, we have stronger evidence that the effect is due to automatic processes.

This is exactly the data provided by Hay and Drager (2010). New Zealand participants were exposed to stuffed animals representing either New Zealand (kiwi birds) or Australia (kangaroos and koalas), then completed the same vowel-selection task used in Niedzielski (1999), using a New Zealand talker. The primary target vowel, /i/, is a highly stereotyped difference between the two countries. Participants were influenced by the presence of the stuffed animals, although in complex gendered ways which were further explored in Drager, Hay, and Walker (2010). For our purposes, the crucial element of this study is that the effect was obtained despite the irrelevance of the prompt to the task. The prompts were thematically related to the linguistic material, in that they dealt with national identity, a topic for which linguistic cues are a popular resource. The stuffed animals were, however, irrelevant to the task itself, in that no verbal connection was made from the stuffed animals or national identity to the task of assessing a computerized voice. Additionally, when probed during debriefing, participants self-reported the belief that the stuffed animals were unrelated to the task. Nonetheless, the New Zealand versus Australian orientation of the animals prompted shifts in participants' subsequent identification of tokens of the kit vowel, which is a stereotype of difference between the nations, although not of dress, which also differs across the two nations but is not as commonly remarked on.

This effect resonates with the earlier non-social finding of McGurk and Macdonald (1976), that perceivers integrate both audio and visual information in speech perception. This study showed that, when presented with audio of the sequence [baba] and the visual image of the speaker saying [gaga], a majority of the time listeners will report hearing a 'fused' response of [dada], a compromise, in terms

of place of articulation, between the two sources of information. As generations of linguistics instructors can report, knowing of the McGurk effect, and therefore knowing the visual stimulus' irrelevance to the comprehension of the audio, does not prevent the effect from occurring. To stop the effect, it is necessary to change the input, for example by looking away from the visual cue.

All of this work suggests that at least some of the elements of social perception are carried out by processes not fully under deliberative control, and that telling listeners a given stimulus is irrelevant may not, or may not always, be sufficient to eliminate its influence in the perceptual process. The current study likewise used information known to be irrelevant to probe control and automaticity, in this case in sociolinguistic perception.

3 | AUDIOVISUAL SOCIAL PERCEPTION

Previous work on faces and voices in sociolinguistic perception suggests that the two modalities can influence one another (Hansen, Rakić, & Steffens, 2018; Hansen, Steffens, Rakić, & Wiese, 2016; Hanulíková, 2018; Williams, 1973). Rubin (1992) found that the co-presence of an Asian-appearing face as compared to a white-appearing face prompted US undergraduates to report perceiving a foreign accent in a lecture recorded by a white speaker from Ohio. In addition to reporting an accent, listeners in that condition showed lower scores on a cloze test, suggesting that their comprehension or retention of the lecture material was impeded by the accent they perceived or by the mismatch between their expectations and the speech they were processing.

Other studies have supported this latter finding that social expectations triggered by a picture can interfere with linguistic comprehension. McGowan (2015) reported that when presented with Mandarin-accented speech in noise, US listeners were more accurate when they saw a picture of a Chinese person than a white person. Similarly, Canadians found Canadians of Chinese descent more difficult to understand than Canadians of European descent in a speech-in-noise task when their pictures were visible, suggesting that their facial phenotype combined with their Canadian English accent posed a processing difficulty for the listeners, themselves primarily either white or Asian Canadians (Babel & Russell, 2015). Yi, Phelps, Smiljanic, and Chandrasekaran (2013) found that video accompaniment assisted speech comprehension, as one might predict, but that this effect was greater for white native talkers of English than for Korean L1 Korean talkers speaking English. They also found that including video exaggerated the difference in perceived accentedness between the two. Gnevshva (2018) showed that Korean L1 Korean talkers were perceived by English-speaking listeners as accented based on audio, video and their combination, while white L1 German talkers were perceived as least accented based on video only, more based on audio and the most when in combination.

4 | SOCIAL EVALUATION OF FOREIGN ACCENT AND ATTRACTIVENESS

The current study examined social evaluation of faces and voices along two dimensions: how accented and how good-looking the speakers seemed. Throughout this discussion, accent and foreign accent will be used to refer to social percepts, rather than speech qualities. These two terms were chosen to create two more intuitive pairings (accented voice ratings and good-looking face ratings) and two less intuitive pairings (accented ratings based on faces and good-looking ratings based on voices). The less intuitive pairings were predicted to make it harder to ignore the explicitly irrelevant information, but not to lead participants to find the task nonsensical or that it would prompt arbitrary responses.

Language attitudes work has repeatedly documented the ease with which listeners are able to perform such tasks, not only providing responses without objection but showing consistent patterns, such as perceiving more powerful or standard accents as belonging to taller and better-looking speakers (Lambert, Hodgson, Gardner, & Fillenbaum, 1960). Participants in the current study reported little difficulty in performing the task when asked.

A previous work on the perception of foreign accent has focused largely on the acoustic details correlated with it, including VOT and vowel quality (McCullough, 2013; Munro, 1993; Riney & Takagi, 1999) and prosodic and global temporal information (Anderson-Hsieh, Johnson, & Koehler, 1992; Kang, 2010; Mareüil & Vieru-Dimulescu, 2006; Munro, Derwing, & Burgess, 2010). It has also established that listeners show fine-grained perceptions of foreign accent (Flege, 1984; Flege, Munro, & MacKay, 1995) and that listener percepts show correlations with a range of acquisition factors for the speakers, such as such as age of learning the second language and length of exposure (for an overview, see Piske, MacKay, & Flege, 2001).

A small body of work has suggested influence from the feelings and ideologies of the perceiver (Lindemann, 2002), showing, for example, that general qualities of ethnocentrism and bias predict negative responses to foreign-accented speakers (Neuliep & Speten-Hansen, 2013; Nguyen, Shaw, Tyler, Pinkus, & Best, 2015). These effects may be mediated by emotional and situational factors, however. Wang, Arndt, Singh, Biernat, and Liu (2013) found that accent bias towards customer service representatives emerged only when the valence of the call's outcome coincided with the valence of the accent in question: valued accents improved percepts of calls with favorable outcomes, while stigmatized accents worsened those of calls with unfavorable outcomes. In addition, pre-existing listener bias may be covert rather than overt, with implicit bias emerging even among participants reporting positive explicit attitudes (Pantos & Perkins, 2013).

Work on the perception of vocal attractiveness has primarily focused on men's and women's attraction to each other, or assessment of same sex targets as sexual competitors. Increased attractiveness in men's voices has been linked to lowered mean f_0 (Hodges-Simeon, Gaulin, & Puts, 2010; O'Connor et al., 2012), specific patterns of word choice (Hodges-Simeon et al., 2010), and increased perceived vocal tract length or formant dispersion (Babel, McGuire, & King, 2014; Collins, 2000; Hodges-Simeon et al., 2010). Similarly, women's voices have been rated as more attractive with higher f_0 (Collins & Missing, 2003). The field has been hampered somewhat by an essentialist view of attractiveness, seen in the search for links between vocal attractiveness and physical characteristics, such as shoulder to hip and waist to hip ratios (Hughes, Dispenza, & Gallup, 2004), body symmetry (Hughes, Harrison, & Gallup, 2002; Hughes & Pastizzo, 2008), or perceived facial attractiveness (Collins & Missing, 2003; Lander, 2008).

Overall, the literature provides good reason to believe that face and voice information will be mutually influential, particularly in the case of faces influencing foreign accent perception. The current study tests the following hypotheses:

Hypothesis 1 *Listeners will reduce their reliance on modalities they are informed are irrelevant and instructed to ignore.*

Hypothesis 2a *When listeners are instructed to ignore the face, it will still influence evaluations, but less so in the accented than the good-looking rating task.*

Hypothesis 2b *When listeners are instructed to ignore the voice, it will still influence evaluations, but less so in the good-looking than the accented rating task.*

Secondarily, an exploratory analysis will probe possible mediators of the influence of irrelevant information, specifically strength of the perceived match between the two modalities, self-report of

a subjective sense of influence, and self-reported familiarity with the accent backgrounds used in the study.

5 | METHODS

5.1 | Speech stimuli

The speech stimuli for the study were selected from those used in McCullough (2013), who generously allowed their use in this study. Recordings consisted of disyllabic English words produced by 15 male talkers. Three talkers had learned English from birth in the US and were screened, for the purposes of the previous study, to be perceived by US listeners as aregional and unaccented. The others all learned English at older ages after having learned Hindi, Korean, Mandarin, or Latin American Spanish from birth. The tokens from the latter group of talkers exhibited a range of features associated with a social percept of foreign accent including prosodic patterns, vowel variation, reduced VOT, retroflex stops, and the use of [β] for [b] (McCullough, 2013). As the pilot data below show, the stimuli were perceived as having different degrees of accentedness.

All speech stimuli were normalized for loudness prior to use. The stimuli were re-piloted to establish evaluations of accented ($N = 22$) and good-looking ($N = 23$) by listeners from the same population and using the same methods as the main study (see below), without the co-present face stimuli. The distribution and means of accented and good-looking ratings for each stimulus are shown in Figure 1. For each stimulus, the mean values of these standalone accented and good-looking ratings were used as predictors in the main study.

5.2 | Visual stimuli

Eighty-five still images of male faces were selected from the Collection of Facial Images,¹ an online database of facial images collected primarily for training computer vision systems in facial processing. Each person was shown from the shoulders or collar upwards, with varying color backgrounds and varying shirt styles. These pictures were rated on the scales accented, masculine and educated by 25 undergraduate participants at Ohio State University for course credit. Based on these ratings, 15

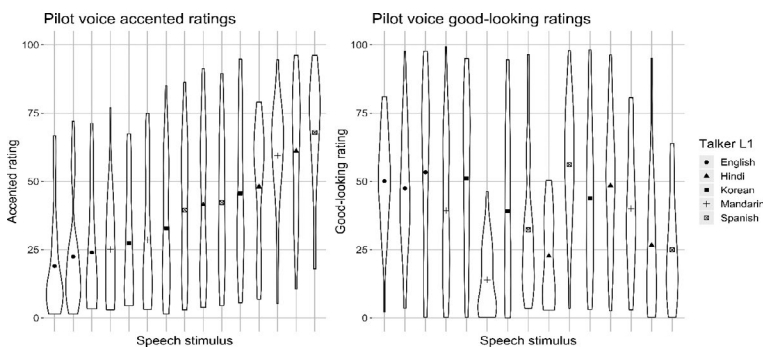


FIGURE 1 Pilot ratings for speech stimuli. Order of stimuli in both graphs is by mean accented rating

faces were selected which represented as wide a range of accented percepts as possible while minimizing variability for masculine and educated percepts.

Pilot data on these 15 faces were collected to establish accented ($N = 21$) and good-looking ($N = 21$) percepts by listeners from the same population as the main study, using the same methods as the main study (see below). The distribution and means of accented and good-looking ratings for each stimulus are shown in Figure 2. For each stimulus, the mean values of these standalone accented and good-looking ratings were used as predictors in the main study.

No demographic information is available for the people represented in the photos, but due to the composition of the original set, the selected faces disproportionately represented faces that are likely to be read by US evaluators as European, particularly western European, and east or southeast Asian.

5.3 | Procedure

All data for the pilots and the main study were collected at a research lab embedded in a large and popular science museum, with museum visitors recruited as participants [self-citation omitted]. Visitors were approached on the museum floor and invited to participate in a study about accents. This population provided a more diverse sample than that provided in university subject pools.

Visitors who agreed to participate were seated in a conference room with a large TV mounted on one wall, at a distance of 10–15 feet from the participants, depending on their position at the table. Each participant was given a paper response sheet with 15 linear scales and asked to indicate their ratings with a pen, as shown in Figure 3.

Faces were displayed on the TV screen, while each single-word recording was played through a speaker three times over the course of five seconds. The display was controlled by E-Prime software, and advanced by hand by the experimenter when all participants had completed their rating for each stimulus. In addition to the face stimulus, the TV displayed the word uttered by the voice stimulus

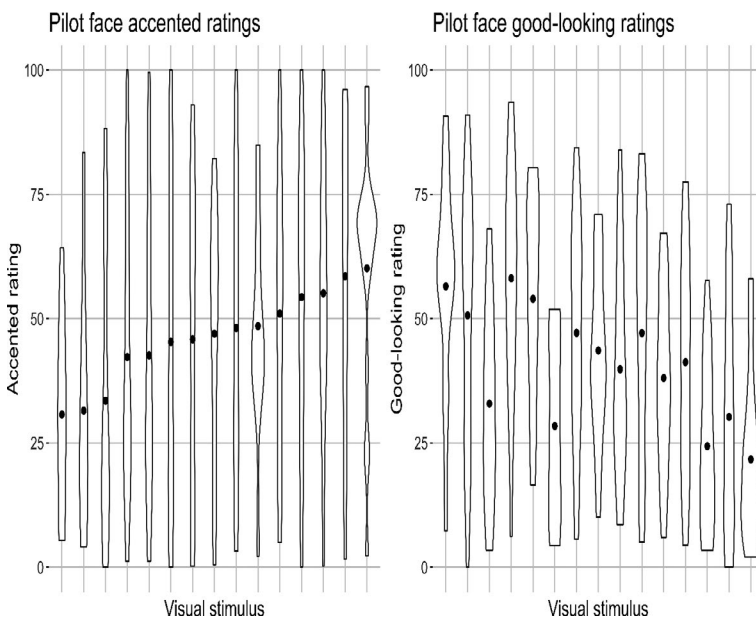


FIGURE 2 Pilot ratings for face stimuli. Order of stimuli in both graphs is by mean accented rating

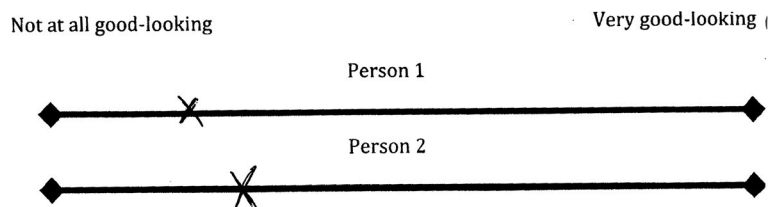


FIGURE 3 Excerpt of marked response sheet

to ensure that accentedness assessment was not impacted by misunderstanding. Visual and auditory stimuli were paired randomly, with every possible combination rated by between 25 and 49 listeners.

After all 15 stimulus pairs were presented, participants provided demographic information on the reverse side of the sheet. Information gathered was age, gender, regional background, racial and/or ethnic identification, having spoken English prior to 5 years of age and self-reported exposure to speakers with the five language backgrounds represented in the study, e.g. ‘native English speakers,’ ‘native Mandarin (Chinese) speakers’ (for a discussion of experience and accent perception, see McGowan, 2016).

After completing the demographic questionnaire, participants were verbally asked what cues they thought they had used to make their judgments and, in the face and voice conditions, whether they thought the irrelevant stimuli had influenced them. Responses were noted and later coded for inclusion in the analysis. The response sheets were scanned and the ratings marked by hand using PDF measuring tools in Adobe Acrobat and Revue Bluebeam, then exported to spreadsheet files and transformed to a 0–100 scale. Demographic responses were entered and coded by hand.

5.4 | Conditions

The study's three conditions were constructed to test the effects of instructing participants to ignore social information from a given source. The speaker condition functioned as a control, in which participants were not instructed to ignore any information. Instead, they were told that a voice would play and a face would be shown and instructed to look at the face while they listened to the auditory stimulus, then to indicate on the sheet how accented or good-looking the person they observed was. The response scales were labeled with ‘Person<NUMBER>’.

In the face and voice conditions, the verbal instructions explicitly included the information that the face and voice stimuli observed were not from the same person. Listeners were told that the goal of the study was to examine the influence of the irrelevant information on their ratings, and they were asked to attend to both modalities but to base their rating solely on the stimulus of the condition (face or voice, respectively). Response scales for these conditions were labeled ‘Face<NUMBER>’ or ‘Voice<NUMBER>’, respectively.

5.5 | Analysis

Eleven percent of the trials collected were excluded due to participants who had begun learning English after 5 years of age, or data collection issues observed and noted by the experimenter: technical difficulties conducting the trial, participants skipping individual questions, disruptions among

the participant group (e.g. a baby fussing) or individual participants who failed to follow the instructions to look at the screen prior to evaluation. This left 15,474 trials from 1,034 participants. These included 629 females, 399 males, and 4 nonbinary participants, as well as 2 for whom gender data were not collected. Ages ranged from 9 to 76 years, with a mean of 26.3 years and standard deviation of 14.3 years. 828 participants self-identified as white, 40 as Black, 43 as multiracial, 27 as Asian, 6 as Latinx, 5 as Middle Eastern and 2 as Native American. Eighty-three participants declined to provide their racial or ethnic self-identification or misunderstood the question. The large proportion of white participants makes it difficult to analyze differences across racial or ethnic groups in our data. Participants' raciolinguistic ideologies (Rosa & Flores, 2017) which underlie the mappings between individual stimuli and accented or good-looking perceptions must certainly vary in relation to how a perceiver is positioned with respect to those ideologies. It is also certainly possible that the interplay between face-based and voice-based racialized information may likewise differ. Our analysis must be performed to generalize primarily to white speakers and further work needed to understand the full picture.

The logit of each rating response was taken to better approximate a normal distribution. A linear mixed effects regression model was fit to each set of ratings (accented and good-looking), with the grouping factors of session, subject nested within session, and sound file. Random effects, both intercepts and slopes, were tested and eliminated only where they prevented model convergence. Fixed effects tested were listener age, sex, condition, and mean accented and good-looking ratings for each face and voice stimulus. To test Hypotheses 1, that the experiment instructions would influence the strength of contributions of the face and voice stimuli, interactions between condition and stimulus ratings were tested. The full model was constructed first and predictors eliminated if they did not significantly improve model fit.

If an interaction was found between condition and stimulus ratings, smaller models were fit to the data from each condition to examine the slope of the stimulus ratings in each case. This provided a test of Hypotheses 2, that the face would contribute less in the accented than good-looking task, and vice versa for the voice.

After hypothesis testing, exploratory models were fit that included interactions between the irrelevant (or in the case of the speaker condition, the less task-relevant) modality and goodness of fit between face and voice [self-citation omitted], as well as participants' self-reports of influence from the irrelevant stimuli and, for accented ratings, self-reported exposure to the language backgrounds of the talkers.

6 | RESULTS

The results supported Hypotheses 1 and Hypotheses 2, showing that both explicit instruction and the evaluative task strongly influenced the relative contributions of face and voice information. More specifically, evaluators were able to completely ignore visual face information when assessing the accentedness of a voice, but not when assessing how good-looking the person represented by the voice might be. Furthermore, voices seem to play a special role in these evaluations, in that listeners were unable to ignore voice information upon instruction, regardless of the evaluation task. No strong evidence was found to support the role of face-voice match, self-reported perception of influence or previous exposure to the language backgrounds heard in mediating the contribution of irrelevant information.

6.1 | Accentedness perception results

When evaluating accentedness, listeners overwhelmingly focused on voice information, which is the only significant contributor in the speaker and voice conditions and remains a contributor, although reduced, in the face condition. Table 1 shows the results of the overall regression model across conditions. Figure 4 shows the influence of the face and the voice, respectively, across the three conditions.

No significant difference is found between the speaker condition and the voice condition. Instead, listeners rely strongly on the voice and show little or no effect of face in both cases. The face condition; however, does differ significantly from the others, showing a reduced but not eliminated effect of voice and a stronger effect of face that is nonetheless less than two-thirds that of the voice in the other conditions. Note that our data offer only ambiguous evidence replicating previous findings (McGowan, 2015; Rubin, 1992; Rubin & Smith, 1990) that face information can shift evaluations of accentedness when both are presented together as a complete speaker. In the full model of the accented ratings, a significant but small effect of face is seen in the speaker condition, but this effect does not appear in a model fit to the speaker condition data alone. This reduction or loss of a previously-seen

TABLE 1 Coefficients table for accented regression model

Number of obs: 7,280, groups: Subject: Session, 493; Session, 180; SoundFile, 15					
	Estimate	SE	df	t value	Pr(> t)
(Intercept)	-0.576	0.127	21.95	-4.521	<0.001
Face-only rating	0.087	0.034	160.84	2.550	0.012
Voice-only rating	0.911	0.121	18.36	7.493	<0.001
Face condition	-0.044	0.114	42.91	-0.384	0.703
Voice condition	0.017	0.088	118.32	0.189	0.851
Face-only rating × Face condition	0.478	0.048	153.56	9.930	<0.001
Face-only rating × Voice condition	-0.070	0.048	143.25	-1.476	0.142
Voice-only rating × Face condition	-0.596	0.101	27.75	-5.913	<0.001
Voice-only rating × Voice condition	0.004	0.070	82.66	0.063	0.950

^aBolded values significant at $\alpha = 0.05$.

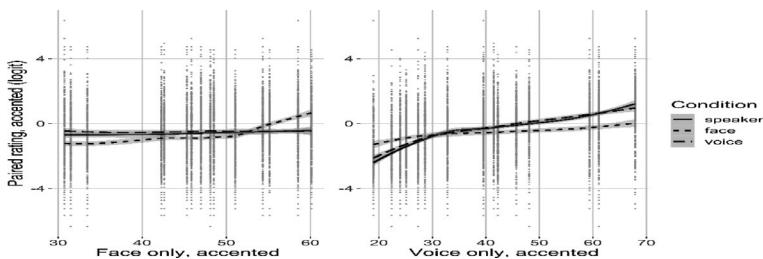


FIGURE 4 Influence of face and voice accented ratings on paired accented ratings (logit transformed) across conditions. Lines are loess smoothed

effect is possibly due to task differences, particularly that our participants had a shorter time of exposure than in those studies.

Additional models tested for interactions between the influence of the modality to be ignored and the exploratory factors face-voice match, self-report of influence and familiarity of language backgrounds. Goodness of face-voice match had no mediating effect. For subjective sense of influence, participants were split. In the voice condition, 35% reported perceiving some influence of the face on their ratings, while 17% reported none. The remaining 47% gave no clear answer one way or the other. In the face condition, a similar number perceived an influence of voice, 33%, while slightly more reported none, 21%. Reported influence did not correlate with ratings in either condition, nor did it mediate the influence of the task irrelevant information.

Listeners' levels of familiarity with the four non-English language backgrounds were averaged together to create a composite score, which showed no significant correlation with accented ratings nor any interaction with face or voice information.

Due to the wide range of ages among participants, additional analyses were conducted to ensure that age variability was neither introducing nor obscuring effects. First, the model in 1 was expanded by adding age in interaction with the two key interactions (Condition \times Voice Accent and Condition \times Face Accent). No age-based interaction reached significance. Only one marginal effect ($p = .051$) emerged, potentially suggesting a slight increase in the strengthening of the face-based effect in the face condition as participants grow older. Next, the model from 1 was refit on subset of the data from 20 to 40 year olds, 40 to 60 year olds, and over 60 year olds. These yielded similar effects as the main model, with the exception of the effect of face accent in the whole-speaker condition, which no longer reaches significance in any of the subsets. The estimated coefficients for this effect do not change substantially, however; all three are within one standard error of the original model's estimate. This suggests that the loss of significant effect is due to the reduced sample size and the overall small size of the effect, rather than age-related differences. Likewise, the oldest adults (60+) who represent the smallest subpopulation tested, also show no significance in the face condition's effect on the contribution of voice accentedness, but the estimate for this coefficient remains close to that of the original model.

6.2 | Good-looking perception results

Analysis of the good-looking ratings does not show the same voice-favoring asymmetry, indicating that the modality focus of the rating label is important. However, this analysis also does not show a face-favoring asymmetry, suggesting either that the voice/good-looking relationship is stronger than the face/accented relationship, or that voices carry an advantage in social perception over still photos.

Table 2 shows the results of the overall regression model for good-looking ratings across conditions. Figure 5 shows the influence of the face and the voice, respectively, across the three conditions.

Unlike in the accented evaluations, here we see robust effects of both face and voice across all three conditions. As in the accented evaluations; however, there is an asymmetry in condition effects, such that no significant difference is seen between the speaker and face conditions, but the voice condition does differ, showing reduced face and increased voice contributions.

In exploratory models, no effect of face-voice match was seen. In the face condition, 56% of participants self-reported influence of the voice information they had been instructed to ignore, and 36% reported experiencing none. In the voice condition, these reports were 64% and 24%, respectively. Such reports were not correlated with ratings in the face condition. In the voice condition, a significant effect did emerge, in which listeners reporting influence gave higher ratings than those who did

TABLE 2 Coefficients table for good-looking regression model

Number of obs: 8,178, groups: Subject: Session, 540; Session, 199; SoundFile, 15; PictureFile, 15					
	Estimate	SE	df	t value	Pr(> t)
(Intercept)	0.593	0.459	498.60	1.293	0.197
Age	0.014	0.002	507.80	7.844	<0.001
Sex: female	-1.789	0.447	470.20	-4.002	<0.001
Sex: male	-1.682	0.449	471.70	-3.745	<0.001
Face-only rating	0.383	0.074	17.20	5.171	<0.001
Face condition	-0.018	0.094	191.90	-0.193	0.847
Voice condition	0.280	0.113	113.50	2.484	0.014
Voice-only rating	0.112	0.032	29.00	3.487	0.002
Face-only rating × Face condition	-0.018	0.038	52.00	-0.462	0.646
Face-only rating × Voice condition	-0.259	0.063	20.20	-4.127	0.001
Voice-only rating × Face condition	-0.056	0.035	61.90	-1.595	0.116
Voice-only rating × Voice condition	0.332	0.051	22.80	6.441	<0.001

^aBolded values significant at $\alpha = 0.05$.

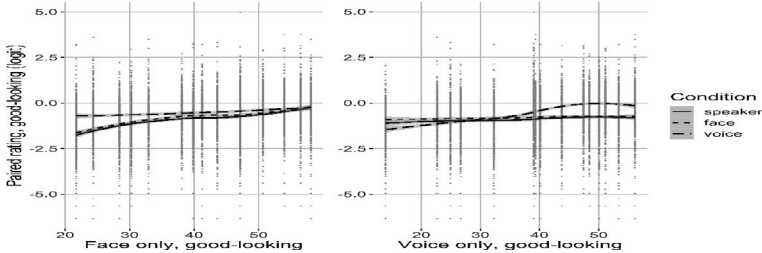


FIGURE 5 Influence of face and voice good-looking ratings on paired good-looking ratings (logit transformed) across conditions. Lines are loess smoothed

not ($\beta = 0.33$, $t = 2.225$, $p = .028$). It would be wise to treat this result with skepticism; however, given that a non-significant trend shows a similar effect for self-reported lack of influence in the same direction ($\beta = 0.31$, $t = 1.906$, $p = .059$), a group that one would expect to show an opposing effect, rather than a matching one. While it is possible that self-reports in either direction are prompted by self-observation of influence, the lack of similar effects in any of the other conditions makes it more likely that this is spurious rather than a real effect.

7 | DISCUSSION

Taken together, the results support the view that sociolinguistic perception relies on both processes under deliberative control and those outside of it, working simultaneously. More specifically, they

show that, as in the McGurk effect, introspectively available knowledge that a piece of information is irrelevant or misleading is not enough to fully block its use by the perception and/or evaluation systems. They also suggest that sociolinguistic perceptual processing is shaped by the evaluative task, regardless of additional deliberative evaluational goals.

There are three levels of effects. First, listeners are able to deliberately choose to rely more on one modality or the other when asked to do so. Second, the evaluative task, assessing how accented versus how good-looking the target is, privileges audio or visual information, respectively.

Finally, these results offer strong support for elements of sociolinguistic evaluation that are more difficult to reach with deliberately available knowledge. In three of the four single-modality conditions, listeners continue to show an effect of the other modality despite being explicitly aware of its irrelevance to the task at hand. The only condition in which they fully eliminate the irrelevant information (i.e. the voice condition for accented ratings) has the double assistance of an incongruent modality-task pairing and the social danger of displaying racial bias, a danger which debriefing discussions showed many participants to be aware of. Additionally, the use of still images instead of videos may have limited the possible contribution of the visual modality by significantly reducing the information available through this modality (Lander, 2008).

Social fear of displaying racial bias is worth exploring more explicitly as well. Some participants commented on the role that ethnolinguistic stereotyping plays in connecting perceptions of non-white-appearing faces and foreign accented voices. Some reported explicitly attempting to avoid response patterns that might indicate such stereotyping. This level of awareness may indeed have reduced effects that might have been bigger in a less apparent task. Conversely, it is possible that the attention itself may have heightened the effects by highlighting the relevant variation in faces and/or voices. Relatedly, since a large majority of the participants in this sample were predominantly white, their investment in and orientation to the relevant raciolinguistic ideological structures were likely to be quite different from perceivers who are themselves marginalized by those or similar other structures.

Finally, the lack of effects of face-voice match are intriguing here. [self-citation omitted], based on pilot data for the current study, argued that the match of the face and voice was a crucial element to consider in the construction of sociolinguistic studies using face-based information. Without entirely dismissing such a concern, note that in the current study, the perceived match of the face and voice turned out to play no role in the relative influence of modalities. This is reassuring for methodological concerns in future studies, and suggests that the evaluation of face-voice match may tap explicit processes not directly relevant for social perception. It's also possible that a different task might show effects of face-voice match. McGowan (2015) has suggested that a reduced face-voice match and genuine downstream effects of cognitive load in speech perception may contribute to the 'imagined accent' phenomenon found to be greater in his work than in this study. This is supported by Babel and Russell (2015), who found that Canadian listeners, regardless of ethnicity, implicit bias or ethnicity of network, showed reduced ability to understand speech in noise as a consequence of seeing a picture of the speaker, who was a Canadian English speaker of Chinese descent.

8 | CONCLUSIONS

These results support a broad literature in social psychology which suggests that many social cognitive processes operate to some extent outside of deliberative control. In the context of third wave

sociolinguistic variation research, they inform a long-standing debate about the appropriate degree of agency built into our models of speakers. Third wave models are necessarily agentive, in that speakers build and adjust their sociolinguistic worlds, including personal identities, speech activities, interactional roles, etc. However, there are two crucial counter-forces to speakers' agency in conducting their sociolinguistic business. The first is the competing agency of their interlocutors, who may have very different goals or beliefs. The second is the cognitive challenge in managing complex, rapidly occurring sociolinguistic information. These results suggest that one shortcut developed by our social perception systems is to allow ambient information to shape our sociolinguistic perceptions without always allowing our deliberative knowledge a full chance to review its relevance.

ORCID

Kathryn Campbell-Kibler  <https://orcid.org/0000-0002-2618-6988>

ENDNOTE

¹ Maintained by Dr. Libor Spacek. <http://cswww.essex.ac.uk/mv/allfaces/index.html>

REFERENCES

- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmental, prosody, and syllable structure. *Language Learning*, 42(4), 529–555.
- Babel, A. (Ed.). (2016). *Awareness and control in sociolinguistic research*. Cambridge, UK: Cambridge University Press.
- Babel, M., McGuire, G., & King, J. (2014). Towards a more nuanced view of vocal attractiveness. *PLoS ONE*, 9(2), e88616. <https://doi.org/10.1371/journal.pone.0088616>
- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *Journal of the Acoustical Society of America*, 137(5), 2823–2833. <https://doi.org/10.1121/1.4919317>
- Brulard, I., & Carr, P. (2013). Variability, unconscious accent adaptation and sense of identity: The case of RP influences on speakers of Standard Scottish English. *Language Sciences*, 39, 151–155.
- Campbell-Kibler, K. (2008). I'll be the judge of that: Diversity in social perceptions of (ING). *Language in Society*, 37(5), 637–659. <https://doi.org/10.1017/S0047404508080974>
- Campbell-Kibler, K. (2010). The effect of speaker information on attitudes toward (ING). *Journal of Language and Social Psychology*, 29(2), 214–223. <https://doi.org/10.1177/0261927X09359527>
- Campbell-Kibler, K. (2016). Toward a cognitively realistic model of meaningful sociolinguistic variation. In A. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 123–151). Cambridge, UK: Cambridge University Press.
- Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, 60(6), 773–780. <https://doi.org/10.1006/anbe.2000.1523>
- Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour*, 65, 997–1004. <https://doi.org/10.1006/anbe.2003.2123>
- Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, 25(5), 736–760. <https://doi.org/10.1521/soco.2007.25.5.736>
- De, Z. (2017). *Race, gender, and /u/: Social perceptions of a non-stereotype feature*. The Ohio State University Master's thesis.
- Drager, K., Hay, J., & Walker, A. (2010). Pronounced rivalries: Attitudes and speech production. *Te Reo*, 53, 27–53.
- Eckert, P. (2000a). *Linguistic variation as social practice: The linguistic construction of identity in Belten High*. New York, NY: Blackwell.
- Eckert, P. (2000b). *Linguistic variation as social practice: The linguistic construction of identity in Belten High*. New York, NY: Blackwell.

- Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of variation. *Annual Review of Anthropology*, 41, 87–100.
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>
- Evans, J. S. B. T. (2019). Reflections on reflection: the nature and function of type 2 processes in dual-process theories of reasoning. *Thinking & Reasoning*, 25(4), 383–415. <https://doi.org/10.1080/13546783.2019.1623071>
- Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America*, 76(3), 692–707. <https://doi.org/10.1121/1.391256>
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97(5), 3125–3134. <https://doi.org/10.1121/1.413041>
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Frankish, K., Evans, J. S. B. T. (2009). The duality of mind: An historical perspective. In K. Frankish & J. S. B. T. Evans (Eds.), *In two minds: Dual processes and beyond* (pp. 1–29). Oxford, UK: Oxford University Press.
- Gnevshva, K. (2018). The expectation mismatch effect in accentedness perception of Asian and Caucasian non-native speakers of English. *Linguistics*, 56(3), 581–598. <https://doi.org/10.1515/ling-2018-0006>
- Hansen, K., Rakić, T., & Steffens, M. C. (2018). Foreign-looking native-accented people: More competent when first seen rather than heard. *Social Psychological and Personality Science*, 9(8), 1001–1009. <https://doi.org/10.1177/1948550617732389>
- Hansen, K., Steffens, M. C., Rakić, T., & Wiese, H. (2016). When appearance does not match accent: Neural correlates of ethnicity-related expectancy violations. *Social Cognitive and Affective Neuroscience*, 12(3), 507–515. <https://doi.org/10.1093/scan/nsw148>
- Hanulíková, A. (2018). The effect of perceived ethnicity on text comprehension under clear and adverse listening conditions. *Linguistics Vanguard*, 4(1), 1–9.
- Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48(4), 865–892. <https://doi.org/10.1515/ling.2010.027>
- Hay, J., Jannedy, S., & Mendoza-Denton, N. (1999). Oprah and /ay/: Lexical frequency, referee design, and style. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1389–1392). Berkeley, CA: University of California.
- Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature*, 21(4), 406–427. <https://doi.org/10.1007/s12110-010-9101-5>
- Hughes, S. M., Dispenza, F., & Gallup, G. G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior*, 25(5), 295–304. <https://doi.org/10.1016/j.evolhumbehav.2004.06.001>
- Hughes, S. M., Harrison, M. A., & Gallup, G. G. (2002). The sound of symmetry: Voice as a marker of developmental instability. *Evolution and Human Behavior*, 23, 173–180. [https://doi.org/10.1016/S1090-5138\(01\)00099-X](https://doi.org/10.1016/S1090-5138(01)00099-X)
- Hughes, S. M., Pastizzo, M. J., & Gallup, G. G. (2008). The sound of symmetry revisited: Subjective and objective analyses of voice. *Journal of Nonverbal Behavior*, 32, 93–108. <https://doi.org/10.1007/s10919-007-0042-6>
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34, 485–499. <https://doi.org/10.1016/j.wocn.2005.08.004>
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38, 301–315. <https://doi.org/10.1016/j.system.2010.01.005>
- Kristiansen, T. (2009). The macro-level social meanings of late-modern Danish accents. *Acta Linguistica Hafniensia*, 41, 167–192. <https://doi.org/10.1080/03740460903364219>
- Kristiansen, T. (2010a). Attitudes, ideology and awareness. In R. Wodak, B. Johnstone, & P. E. Kerswill (Eds.), *The SAGE handbook of sociolinguistics* (Vol. 120, pp. 265–278). Thousand Oaks, CA: Sage.
- Kristiansen, T. (2010b). Conscious and subconscious attitudes towards English influence in the Nordic countries: Evidence for two levels of language ideology. *International Journal of the Sociology of Language*, 204, 59–95. <https://doi.org/10.1515/ijsl.2010.031>
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center For Applied Linguistics.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Labov, W. (1993). *The unobservability of structure and its linguistic consequences*. Paper presented at NWAV 22. Ottawa, Canada: University of Ottawa.

- Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., & Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, 15(4), 431–463.
- Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, 60(1), 44–51. <https://doi.org/10.1037/h0044430>
- Lander, K. (2008). Relating visual and vocal attractiveness for moving and static faces. *Animal Behaviour*, 75, 817–822. <https://doi.org/10.1016/j.anbehav.2007.07.001>
- Levon, E., & Fox, S. (2014). Social salience and the sociolinguistic monitor: A case study of ING and TH-fronting in Britain. *Journal of English Linguistics*, 42(3), 185–217. <https://doi.org/10.1177/0075424214531487>
- Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society*, 31, 419–441.
- Mareilil, P. B. D., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, 63(4), 247–267. <https://doi.org/10.1159/000097308>
- McCullough, E. A. (2013). *Acoustic correlates of perceived foreign accent in non-native English*. The Ohio State University PhD dissertation.
- McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech*, 58(4), 502–521. <https://doi.org/10.1177/0023830914565191>
- McGowan, K. B. (2016). Sounding Chinese and listening Chinese: Awareness and knowledge in the laboratory. In A. Babel (Ed.), *Awareness and control in sociolinguistic research* (p. 25). Cambridge, UK: Cambridge University Press.
- McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748. <https://doi.org/10.1038/264746a0>
- Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accent-ness ratings. *Language and Speech*, 36(1), 39–66. <https://doi.org/10.1177/002383099303600103>
- Munro, M. J., Derwing, T. M., & Burgess, C. S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, 52(7), 626–637. <https://doi.org/10.1016/j.specom.2010.02.013>
- Neuliep, J. W., & Speten-Hansen, K. M. (2013). The influence of ethnocentrism on social perceptions of nonnative accents. *Language & Communication*, 33(3), 167–176. <https://doi.org/10.1016/j.langcom.2013.05.001>
- Nguyen, N., Shaw, J. A., Tyler, M. D., Pinkus, R. T., & Best, C. T. (2015). Affective attitudes towards Asians influence perception of Asian-accented vowels. Proceedings of the International Conferences of Phonetic Sciences. 1–5.
- Niedzielski, N. A. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1), 62–85. <https://doi.org/10.1177/0261927X99018001005>
- O'Connor, J. J. M., Feinberg, D. R., Fraccaro, P. J., Borak, D. J., Tigue, C. C., Re, D. E., ... Tiddeman, B. (2012). Female preferences for male vocal and facial masculinity in videos. *Ethology*, 118, 321–330. <https://doi.org/10.1111/j.1439-0310.2011.02013.x>
- Pantos, A. J., & Perkins, A. W. (2013). Measuring implicit and explicit attitudes toward foreign accented speech. *Journal of Language and Social Psychology*, 32(1), 3–20. <https://doi.org/10.1177/0261927X12463005>
- Peirce, C. S. (1901). Sign. In J. M. Baldwin (Ed.), *Dictionary of philosophy and psychology* (vol 2, 527 pp). New York, NY: McMillan.
- Pharao, N., Maegaard, M., Møller, J. S., & Kristiansen, T. (2014). Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society*, 43(1), 1–31. <https://doi.org/10.1017/S0047404513000857>
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191–215. <https://doi.org/10.1006/jpho.2001.0134>
- Podesva, R. J., & Van Hofwegen, J. (2015). How conservatism and normative gender constrain variation in inland California: The case of /s/. University of Pennsylvania Working Papers in Linguistics 20(2). Article 15.
- Preston, D. R. (1996). Whaddayaknow?: The modes of folk linguistic awareness. *Language Awareness*, 5(1), 40–74.
- Preston, D. (2016). Whaddayaknow now? In A. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 123–151). Cambridge, UK: Cambridge University Press.
- Riney, T. J., & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning*, 49(2), 275–302. <https://doi.org/10.1111/0023-8333.00089>
- Rosa, J., & Flores, N. (2017). Unsettling race and language: Toward a raciolinguistic perspective. *Language in Society*, 46(5), 621–647. <https://doi.org/10.1017/S0047404517000562>

- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), 511–531. <https://doi.org/10.1007/BF00973770>
- Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, 14, 337–353.
- Silverstein, M. (1976). Shifters, linguistic categories, and cultural description. In K. Basso & H. Selby (Eds.), *Meaning in anthropology* (pp. 11–55). Albuquerque, NM: University of New Mexico Press.
- Silverstein, M. (1981). *The limits of awareness*. Sociolinguistic Working Paper Number 84, Austin, TX: Southwest Educational Development Laboratory.
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language and Communication*, 23, 193–229. [https://doi.org/10.1016/S0271-5309\(03\)00013-2](https://doi.org/10.1016/S0271-5309(03)00013-2)
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662. <https://doi.org/10.1037/h0054651>
- Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology*, 4, 1015. <https://doi.org/10.3389/fpsyg.2013.01015>.
- Van Bavel, J. J., Jenny Xiao, Y., & Cunningham, W. A. (2012). Evaluation is a dynamic process: Moving beyond dual system models. *Social and Personality Psychology Compass*, 6(6), 438–454. <https://doi.org/10.1111/j.1751-9004.2012.00438.x>
- Wagner, S. E., & Hesson, A. (2014). Individual sensitivity to the frequency of socially meaningful linguistic cues affects language attitudes. *Journal of Language and Social Psychology*, 33(6), 651–666. <https://doi.org/10.1177/0261927X14528713>
- Wang, Z. E., Arndt, A. D., Singh, S. N., Biernat, M., & Liu, F. (2013). “You lost me at hello”: How and when accent-based biases are expressed and suppressed. *International Journal of Research in Marketing*, 30(2), 185–196. <https://doi.org/10.1016/j.ijresmar.2012.09.004>
- Williams, F. (1973). Some research notes on dialect attitudes and stereotypes. In R. W. Shuy & R. W. Fasold (Eds.), *Language attitudes: Current trends and prospects ch. 9* (pp. 113–128). Washington, DC: Georgetown University Press.
- Yi, H.-G., Phelps, J. E. B., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America*, 134(5), EL387–EL393. <https://doi.org/10.1121/1.4822320>

How to cite this article: Campbell-Kibler K. Deliberative control in audiovisual sociolinguistic perception. *J Sociolinguistics*. 2021;25:253–271. <https://doi.org/10.1111/josl.12418>