

Understanding Social Voice in Sociolinguistic Perception and Its Working Mechanism

Wei Wang, Lili Fan, Jiaxi Liu, Hui Yu, Qinghua Ni, and Levente Kovacs

Abstract—A cognitive model of “social voice” in sociolinguistic perception is proposed in this paper, which explores how language use reflects social identity and how listeners interpret these linguistic features to form social judgments. Moreover, social meaning involves the stereotypes or judgments people attach to certain ways of speaking. The social cognition processes how speakers perceive these linguistic cues to form those judgments. Linguistic features act as indexes of social categories. Context determines how these features are interpreted. This interplay between language, identity, and cognition reveals how societal norms, biases, and cognitive processes shape communication. Social voice in sociolinguistic perception reveals how language both reflects and reinforces social hierarchies. By understanding the cognitive mechanisms behind these processes, such as automatic stereotyping and contextual flexibility, we can challenge biases and promote equitable communication. Future research should prioritize intersectionality, technological ethics, and strategies to amplify marginalized voices, fostering a society in which linguistic diversity is celebrated rather than stigmatized.

Index Terms—Social vision, social voice, sociolinguistic perception, social meaning, social cognition, forensic phonetics, social context, sociophonetic variation

I. INTRODUCTION

IN the field of social psychology, the concept of social vision deals with how social factors (e.g., identity, culture, and context) influence visual perception and how people use visual cues to navigate social interactions [1]. Enlightened by social vision and social voice in sociolinguistic perception delves into how language use

conveys social meaning and how individuals cognitively process these meanings in relation to the social identity, group membership, and power dynamics. It bridges the gap between linguistic forms (e.g., accents, dialects, and speech styles) and the social interpretations that they evoke, offering insights into how language functions as a tool for constructing and perceiving social reality. The study of social voice in sociolinguistic perception highlights the intricate relationships between language, society, and cognition [2, 3]. By uncovering how social meaning is encoded in language and how it is perceived, researchers can better understand the role of language in shaping social identities, power structures, and cultural norms. This knowledge is essential for fostering greater linguistic equity and social justice.

II. UNDERSTANDING SOCIAL VOICE IN SOCIOLINGUISTIC PERCEPTION

The concept of social voice in sociolinguistic perception refers to how individuals use and interpret linguistic features to convey and infer social identity, group membership, and interpersonal dynamics. It encompasses the ways in which language variation (e.g., accents, dialects, and speech styles) carries social meaning and how listeners perceive and evaluate these variations within specific cultural and contextual frameworks. Social voice represents a critical lens for understanding how language functions as both a marker of identity and a tool for social interaction.

A. Forensic Phonetics

Forensic phonetics applies phonetic analysis to legal and investigative contexts, such as speaker identification, voice comparison, and authenticity verification of audio recordings. It examines how unique vocal features (e.g., pitch, formants, and speech rhythm) contribute to a speaker’s social voice. These features reveal demographic information (e.g., age, gender, and regional background) and even psychological states, such as stress and deception [4].

One typical example using forensic phonetics techniques is the Lindbergh case [5]. While it is common for part of the evidence in a case to be based on eyewitness testimony, which may or may not come from the victim, in most instances the perpetrator’s voice is not familiar to the witness. It is also typical that a significant amount of time, often several weeks or more, passes between the crime and the attempt to

Manuscript received: 15 June 2025; revised: 31 July 2025; accepted: 8 August 2025. (Corresponding author: Wei Wang.)

Citation: W. Wang, L. Fan, J. Liu, H. Yu, Q. Ni, and L. Kovacs, Understanding social voice in sociolinguistic perception and its working mechanism, *Int. J. Intell. Control Syst.*, 2025, 30(3), 218–226.

Wei Wang is with School of Languages and Communication Studies, Beijing Jiaotong University, Beijing 100044, China (e-mail: weiwang3@bjtu.edu.cn).

Lili Fan is with School of Artificial Intelligence, Beijing Institute of Technology, Beijing 100081, China (e-mail: lilifan@bit.edu.cn).

Jiaxi Liu is with School of Social Sciences, University of Manchester, Manchester M13 9PL, UK (e-mail: victorwv@sina.com).

Hui Yu is with School of Creative Technologies, University of Portsmouth, Portsmouth PO1 2DJ, UK (e-mail: hui.yu@port.ac.uk).

Qinghua Ni is with Faculty of Innovation Engineering, Macau University of Science and Technology, Macao 999078, China (e-mail: 3230001810@student.must.edu.mo).

Levente Kovacs is with John von Neumann Faculty of Informatics, Óbuda University, Budapest 1034, Hungary (e-mail: kovacs@uni-obuda.hu).

Digital Object Identifier 10.62678/IJICS202509.10235

determine whether a suspect's voice matches that of the perpetrator. In such cases, the accuracy of the witness's memory of the voice becomes a critical issue. A key question in this context is how the voice memory deteriorates over time. Early attempts to address this question were inspired by a testimony in the Lindbergh case [5].

Lindbergh testified under oath that he recognized Hauptmann's voice [6]. In legal settings, forensic phonetics can be regarded as one application field of social voice, to a broader sense. Studies in forensic phonetics reveal how listeners attribute identity and intent based on vocal cues. It also highlights the impact of linguistic stereotypes on legal outcomes (e.g., racial or ethnic profiling through voice) [7, 8]. There are four major methods in forensic voice perception, as shown in Fig. 1. Their detailed explanations are given below.

(1) Categorical-phonetic transcription and description

Categorical-phonetic transcription involves the systematic notation of speech sounds based on their phonetic properties, which include vocal setting and voice quality, pitch and intonation, articulation rate, rhythmical features, connected speech processes, consonantal features, vowel features, etc. It is widely used in speaker identification, accent and dialect analysis, voice comparison, and authentication of recordings [9].

(2) Acoustic phonetics

Fundamental frequency (f_0), also known as the pitch of a

voice, is a critical acoustic feature used in forensic phonetics. It refers to the rate of vibration of the vocal folds and is measured in Hertz (Hz). f_0 is a vital acoustic feature in forensic phonetics, providing critical insights for speaker identification, voice comparison, and emotional analysis [10].

(3) Holistic voice perception

"Holistic voice perception" in forensic phonetics refers to the approach of considering the overall impression and characteristics of a voice when analyzing and comparing speech samples, rather than focusing solely on individual phonetic features. This approach acknowledges that human listeners often perceive voices holistically, taking into account various aspects such as pitch, timbre, rhythm, and intonation, which together contribute to the unique quality of a speaker's voice [4].

(4) Automatic speaker identification

In forensic phonetics, "automatic speaker recognition (ASR)" system is an important technical mean used to compare and evaluate speech samples to determine whether the voice of an unknown speaker and that of a known suspect originate from the same person [4]. ASR systems typically involve four stage.

Feature extraction. Speech features are extracted from recordings. Commonly used features include mel-frequency cepstral coefficients (MFCCs) or log mel-filterbank energies.

Feature modeling. The extracted features are processed

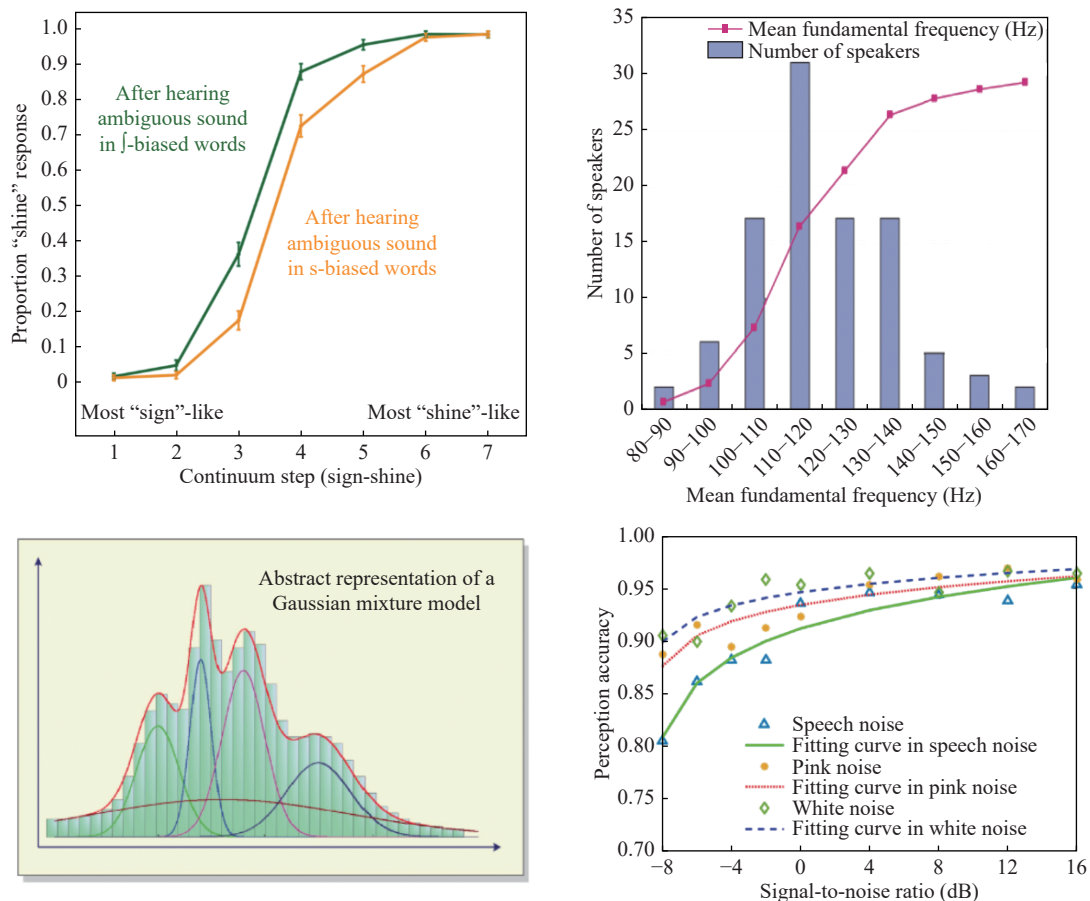


Fig. 1 Typical studies and their experimental results using categorical-phonetic transcription and description, acoustic phonetics, holistic voice perception, and automatic speaker identification in forensic voice perception [4, 9, 10].

into the speaker models. Popular models include Gaussian mixture model-universal background model (GMM-UBM), i -vectors, and x -vectors.

Scoring generation. Similarity scores between two recordings are calculated, often using probabilistic linear discriminant analysis (PLDA) or cosine similarity.

Likelihood ratio (LR) calculation. The scores are calibrated into interpretable likelihood ratios to evaluate the strength of speech evidence.

Forensic phonetics provides tools for analyzing how vocal features encode social information. It also highlights the role of bias in interpreting social voice, especially in high-stakes settings such as courtrooms. However, the use of voice as evidence raises ethical questions about the use of voice as evidence and the potential for discrimination.

B. Consciousness, Awareness, and Control

Consciousness and awareness are prerequisites for control [11]. If speakers are not aware of the sociolinguistic cues, they can not control their perception. But once speakers are aware, through education or training, they can exert control over their biases. Also, there is the aspect of automatic vs. controlled processing. Sociolinguistic perception starts as automatic (unconscious) based on ingrained stereotypes, but with awareness, one can engage in more controlled (conscious) processing to override those biases. But this requires cognitive effort, so in high-stakes situations or when motivated, people are more likely to exert that control [12]. Another angle is the speaker's control. Speakers consciously modify their speech (code-switching and style-shifting) to influence how they are perceived. This ties into audience design and accommodation theory. So, both speakers and listeners have roles in controlling sociolinguistic perception through awareness and deliberate control strategies [13].

Consciousness, awareness, and control in sociolinguistic perception emerge from a multi-level neural architecture that integrates sensory input, social cognition, reflective awareness, and self-regulation [11]. This framework highlights the capacity of brain to balance automatic biases with intentional adaptation, fostering more equitable and

adaptive social communication. By leveraging insights from this model, we can enhance both individual and societal responses to linguistic diversity. In summary, consciousness and awareness in sociolinguistic perception allow individuals to recognize and reflect on social language cues, while control enables them to adjust their responses. The multi-level neural network supports this through hierarchical processing, from basic sensory input to higher-order self-regulation, as shown in Fig. 2 [14].

The sensorimotor level handles the basic processing of sensory input (hearing speech) and motor responses (speaking). The cognitive level involves higher processing like understanding language, recognizing patterns, and making associations. The conscious level is where we are aware of these processes, and metacognitive is thinking about our own thinking, which monitors and adjusts our cognitive processes. When hearing a sound cue, at the sensorimotor level, the network processes the sound. At the cognitive level, it will recognize the sound as a particular signal and stored social meanings. Consciousness brings that recognition into awareness, and metacognition allows the network to reflect on and possibly control the biases. In sociolinguistic perception, all these levels work together: processing the speech sounds, recognizing social cues, becoming aware of those cues, and then using metacognitive control to manage responses.

Campbell-Kibler [13] once discussed the role of “deliberative control” in audiovisual sociolinguistic perception. The combination of seeing and hearing provides more cues that either enhance or hinder control. In the experiment, visual information reinforced stereotypes, making it harder to control perceptions, or conversely, seeing the speaker humanizing them and reducing bias. Consciousness and awareness enable individuals to recognize sociolinguistic cues, while control involves the deliberate management of perceptions and responses. Campbell-Kibler's research touches upon how these factors operate in an audiovisual context, affecting everything from first impressions to long-term social judgments [15]. Understanding this interplay can have implications for reducing linguistic prejudice and improving intergroup communication. Enlightened by

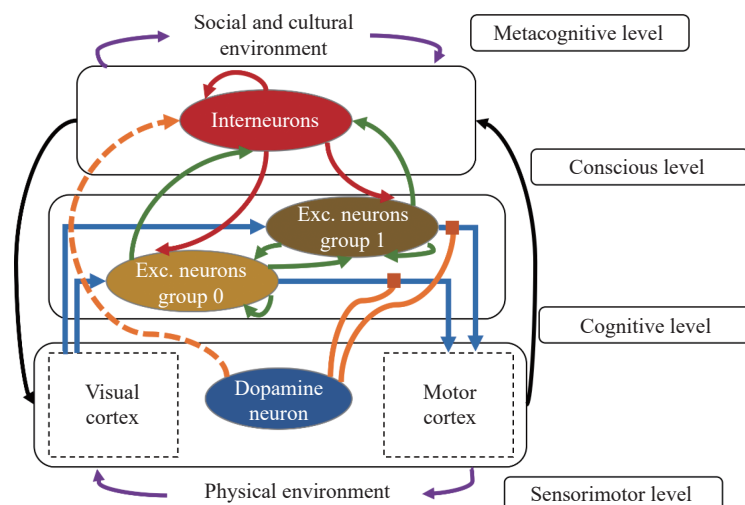


Fig. 2 Multi-level neural network processing cognitive tasks with increasing difficulties [14].

Campbell-Kibler's research [15], there are two major implications concerning the role of consciousness, awareness, and control framed in the general model of social voice.

(1) Consciousness and awareness

Sociolinguistic perception often begins with automatic processing of features such as accents, dialects, and speech styles, influenced by the ingrained social stereotypes. However, consciousness involves recognizing these cues as socially constructed markers. For example, becoming aware that a Southern USA accent is stigmatized in certain contexts. Awareness entails understanding the indexicality of linguistic features. For example, how vocal fry or uptalk may be linked to perceptions of youth, gender, and professionalism. Studies have shown that audiovisual input (e.g., seeing a speaker's appearance) can heighten awareness of these associations [13, 15].

(2) Deliberative control

Control refers to the ability to modulate sociolinguistic judgments through conscious effort. For instance, suppressing stereotypes about a non-native accent once aware of their bias. Campbell-Kibler's experiments suggest that while automatic perceptions are robust, deliberative control can alter evaluations when participants are motivated or instructed to reflect. Visual cues (e.g., race, gender, and attire) interact with auditory ones, complicating control [13]. Seeing a speaker might either reinforce stereotypes (e.g., linking a dialect to a racialized identity) or humanize them, reducing bias. Control depends on whether listeners can disentangle linguistic and visual signals.

The roles of consciousness, awareness, and control in sociolinguistic perception show that social voice is not fixed but dynamic and strategic. It explains how speakers navigate social expectations and power structures through linguistic choices. More importantly, the awareness and control constitute an important part in the model of sociolinguistic perception, which reveals the cognitive processes involved in monitoring and adjusting speech [11, 12].

C. Social Context

The concept of "social context" in sociolinguistic perception refers to the way, in which the surrounding social environment influences how language is perceived and interpreted. This context includes various factors such as the speaker's social identity, the listener's expectations, the setting in which the interaction occurs, and the broader cultural and social norms that shape communication [2, 16].

(1) Speaker identity and social cue

Social identity is the perceived social identity of a speaker (e.g., gender, age, race, and socioeconomic status) that can influence how their speech is interpreted. For example, a speaker with a higher social status may be perceived as more credible or authoritative, even if the linguistic content is the same. The accent a person uses can signal their regional or social background. Listeners often make judgments about a speaker's education, intelligence, and trustworthiness based on the cues [3, 17].

(2) Listener expectation and prejudice

Listeners bring their own expectations to a conversation

based on stereotypes or prior experiences. These expectations can affect how they perceive the speaker's intentions and the meaning of their speech. Pre-existing biases and prejudices lead to misinterpretations of speech purpose [18].

(3) Interactional setting

The level of formality in a setting (e.g., a casual conversation vs. a formal meeting) affects how language is perceived. In formal settings, certain speech patterns may be expected, and deviations could be interpreted negatively. The power dynamics between speakers also influence perception [19].

(4) Cultural and social norms

Different cultures have different norms for communication, such as politeness, turn-taking, and the use of silence. These norms shape how speech is perceived and understood. Social norms within a community dictate acceptable speech behaviors [20].

(5) Situational context

The situational context, such as the topic of conversation, provides additional cues that influence perception. For example, a speaker may use more technical language in a scientific discussion than in a casual chat. Nonverbal communication, such as gestures, facial expressions, and body language, also provides the context that affects how the speech is interpreted [21].

The social context in sociolinguistic perception highlights the importance of understanding, in which language is not only just about the words spoken but also about the social and cultural environment in which it is used. This context shapes how language is perceived, interpreted, and responded to, making it a crucial aspect of effective communication [22]. There are three examples regarding the role of social context in sociolinguistic perception.

Gender and speech perception. Studies have shown that listeners might perceive the same speech differently based on whether they believe that the speaker is male or female. For example, a female speaker may be perceived as more emotional or less assertive than a male speaker by saying the same thing [3].

Racial and ethnic stereotypes. Listeners may make assumptions about a speaker's intelligence or education based on their racial or ethnic background, even if these assumptions are not accurate [2, 17].

Socioeconomic status. A speaker's perceived socioeconomic status can influence how their speech is interpreted. For example, a speaker with a more "standard" accent may be perceived as more educated or competent than someone with a regional or working-class accent [19].

III. SOCIOPHONETIC VARIATION, SOCIAL MEANING, AND SOCIAL COGNITION

The model of social voice aims to explain the interplay between sociophonetic variations, social meaning, and social cognition. Understanding their interconnections lies at the heart of understanding how language functions as both social and cognitive phenomena. These concepts reveal how speech patterns are shaped by identity, how listeners interpret linguistic cues, and how cognitive processes mediate these interactions.

A. Sociophonetic Variations and Their Social Meaning

Sociophonetic variations can be simply understood as phonetic variations driven by social factors, ranging from social groups, regions, genders, ages, and other social variables. This term often appears in the study of loanword adaptation. Loanwords refer to words in the recipient language (L1) borrowed from donor language (L2). The borrowed words are usually naturalized by adapting their phonologically ungrammatical structure into a well-formed structure which conforms to the phonology of the L1 recipient language [23–25]. The adaptation strategies vary across languages. For example, there are two major types of loanwords in Mandarin, namely loan translations or calques, such as “hotdog” [‘hat,dɔŋ] literally translated as “re.gou” [ʃɤ.kou], and phonological loans, such as “bar” [‘bar] (American English) or [‘ba:] (British English) altered into “ba” [pʌ].

There used to be three major theoretical models to explain loanword adaptation [24, 26]. The first model deals with phonetic matching. This is represented by the model of perceptual assimilation (perceptual stance model) [27–29]. A process of phonetic decoding happens and largely accounts for the perceptual assimilation of L2 native sounds into L1 non-native sounds. The process of perceptual assimilation can be framed in a psycholinguistic model. During the process, segments or syllables in L2 are mapped onto their phonetically closest L2 counterparts. Such phonetically minimal transformations are determined by acoustic proximity between L1 native and L2 non-native phonetic categories [30–32]. The second model is based on phonological matching. This relates to category proximity principle and preservation (phonological stance model) proposed by LaCharité and Paradis [25]. The principle explicitly states that non-existing phonological category in L2 will be replaced by its closest phonological category in L1, but not the phonetically more similar sound in L1 inventory. Conversely, the phonological category shared in L1 and L2 will be preserved despite phonetic differences. However, L2

phonemes must change their structures into licit ones in L1. So, the phonemic proximity outweighs phonetic approximation [25, 32]. The third model focuses on the interaction between phonological grammar and phonetic perception. This is exemplified by perceptibility-map (P-map) theory proposed by Steriade [33]. The theory claims that neither pure phonological grammar nor a pure perceptual salience can fully account for loanword adaptation. Instead, perceptual similarity better explains the facts because it incorporates perceptual factors into grammatical constraints. So, this approach explicitly covers perceptual factors and phonetic details through highlighting the relative ranking of P-map based markedness and faithfulness constraints that determine the result of adaptation (perceptual similarity model) [24, 34–36]. However, the above models mainly examine the adaptation of loanwords from pure phonological, acoustic, and perceptual account. Social factors are either ignored or less studied [37]. Previous studies consider the social factors in following aspects: the level of bilingualism [28, 37–39], orthography [40, 41], degree of linguistic integration [42], age and social class, language dominance [39], and topics in speech and cultural images [23]. Wang [43] integrated the models into “sociolinguistic radar”, a relatively intensive consciousness and awareness required cognitive model in different sociophonetic processes [44]. The general working mechanism of sociolinguistic radar is shown in Fig. 3.

However, the phonetic variation is not just a linguistic phenomenon, but also has significant social implications. For example, certain phonetic features may be considered “standard” or “non-standard”, thereby influencing social evaluations of speakers. Additionally, the phonetic variation can be used to construct and express social identity. Sociophonetic variations can also reflect social stratification. For instance, the greater /s/ reduction is observed among older speakers, men, and lower socioeconomic classes in more informal styles. This suggests that certain phonetic features may be used to signal social identity and adherence to specific social norms. The social meanings of phonetic variants can be

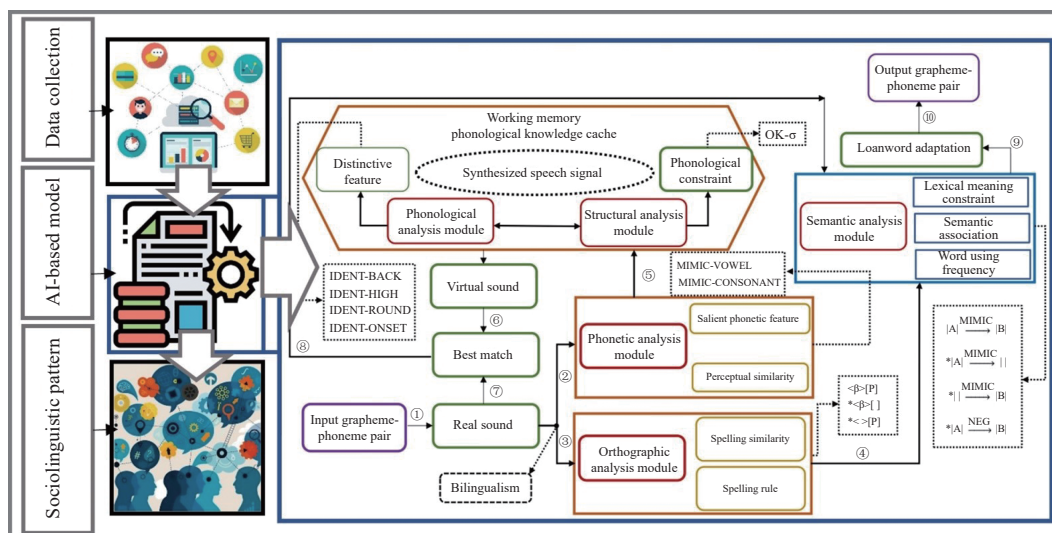


Fig. 3 General working mechanism of sociolinguistic radar [44].

seen as part of a fluid “indexical field”, where a single variant can activate multiple related social meanings depending on the context. For example, the variant [z] in intervocalic positions in Costa Rican Spanish was initially associated with physiological factors but later became linked to masculinity and confidence [2, 23]. Sociophonetic variations are crucial in understanding how language reflects and constructs social identity. These variations carry rich social meanings that influence how speakers are perceived by others. By studying sociophonetic variation, researchers can gain insights into the dynamic relationships between language and society, and how social meanings are constructed and interpreted through speech.

B. Social Meaning and Its Social Cognition

Every individual possesses a brain network that produces overarching social representations. This network encompasses regions such as the superior temporal sulcus (STS), the medial prefrontal cortex (mPFC), the medial parietal region (mPR), and the temporo-parietal junction, among others. These regions are intricately connected to other systems, including sensory, cognitive, and motor systems, which both supply and receive information from them [45]. Through these connections, the brain network generates appropriate predictions and behaviors that facilitate the coordination and interaction between two or more individuals [46, 47]. Social cognition processes influence how listeners interpret linguistic features, while sociolinguistic perception provides a rich source of social information that can shape and reinforce social attitudes and beliefs. Understanding this interaction is essential for studying how language functions in social contexts and how social meanings are constructed and interpreted through speech [48].

Linguistic forms can serve as indicators, or “indexes” of certain aspects of the social world, but these connections are not straightforward or unchanging. Instead, the social meanings tied to these forms shift depending on the context [49, 50]. For instance, in American English, pronouncing a /t/ sound with a clear release (as opposed to a tap) may signal clarity or precision [50]. Based on these associations, listeners may interpret the use of a released /t/ as the speaker attempting to sound clear, possibly implying that the speaker thinks the listener is struggling to understand. In other situations, these same associations with clarity and precision lead to broader social meanings, such as attributing intelligence or hyper-articulateness to the speaker [51].

However, linguistic forms do not always influence the conclusions listeners draw about speakers. Podesva and Hofwegen [52] found that variations in /t/ pronunciation significantly affected how less familiar American politicians were perceived but had no impact on perceptions of highly familiar figures like George W. Bush and Hillary Clinton. They suggest that listeners already have such strong and pre-existing associations with these politicians, Bush with inarticulateness and Clinton with clarity, that subtle variations in /t/ articulation do not alter these entrenched perceptions [15].

The case of George W. Bush and Hillary Clinton is not

exclusive to linguistic variation but rather reflects broader cognitive processes involved in forming impressions and attitudes. Research in experimental social psychology has shown that individuals rapidly and automatically deduce character traits [53] and social roles [54] based on others’ behavior. However, not all of these deductions hold the same cognitive significance. First impressions tend to be more easily recalled from memory compared with subsequent impressions [55], are harder to alter [56], and are often perceived as context-independent. Conversely, the information that conflicts with the first impressions is typically interpreted as context-dependent. For instance, if you initially encounter someone who is reserved and quiet at a social gathering, you may assume that they are generally shy. Yet, if you later see them being lively and outgoing while giving a lecture, you are likely to retain your original impression of them as shy and attribute their extroverted behavior to the specific teaching context.

Furthermore, the influence of contextual information on observers’ judgments depends on whether these judgments are based on strongly or weakly held attitudes. Strong attitudes, which are more enduring, stable, and resistant to change [57], seem to be formed through different cognitive processes compared with weak attitudes. The reason is that people tend to keep their strong attitudes in memory [58]. When a person relies on a strong pre-attitude retrieved from memory to make an evaluative judgment, the judgment will not be influenced by the context effect. In contrast, without strong attitudes, evaluative judgments tend to be formed immediately and are therefore more vulnerable to contextual influences [58].

The interplay between social meaning and social cognition lies at the core of sociolinguistic perception, shaping how individuals interpret and evaluate language variation. Social meaning provides the “content” (e.g., /t/ release = clarity), while social cognition determines how that content is processed (e.g., “Is the speaker trying to sound clear because I’m confused?”). Context mediates this interaction: a feature’s meaning shifts depending on the setting, participants, and cultural norms [16, 49]. Cognitive shortcuts (heuristics) amplify stereotypes. For instance, listeners may assume a speaker with a working-class dialect is less educated, even if the content of their speech contradicts this. Entrenched associations (e.g., linking Hillary Clinton to “clarity” or George W. Bush to “inarticulateness”) can override linguistic cues [52]. Speakers will strategically manipulate social meaning (e.g., code-switching to signal identity). Listeners can exercise deliberative control to override automatic biases, though this requires effort and awareness [15].

There are several factors influencing the interplay between social meaning and social cognition. The first factor is familiarity. Familiar speakers (e.g., celebrities, politicians, scholars, and workers) resist reinterpretation because their social meanings are already fixed. Modality also makes a difference. Audiovisual cues interact with linguistic ones, complicating the perception of people. Cultural ideologies are regarded as the third factor. Dominant norms privilege certain varieties while stigmatizing others. For future research directions in this field, it could be how overlapping personal

identities shape the perception of social meaning, how AI and speech recognition tools encode or challenge sociolinguistic biases, and how to map brain activity during sociolinguistic perception (e.g., how stereotypes activate neural networks) [13, 15].

Social meaning and social cognition form a feedback loop: linguistic features acquire social significance through cultural ideologies, and cognitive processes reinforce or challenge these associations. By studying their interplay, we uncover how language both reflects and perpetuates social hierarchies, and how we disrupt these patterns to foster equity. Understanding this dynamic is essential for addressing the linguistic discrimination.

IV. CRITICAL REFLECTION

The concept of social voice in sociolinguistic perception explores how language use reflects social identity and how listeners interpret these linguistic features to form social judgments. This interplay between language, identity, and cognition reveals how societal norms, biases, and cognitive processes shape communication.

A. Interaction between Social and Linguistic References

Linguistic variation is not merely a linguistic phenomenon. It also carries rich social meanings. Certain phonetic features (such as the “ing” variable in American English) can influence how listeners perceive speakers. Research shows that when speakers use specific phonetic variants, listeners may make judgments about the speakers’ education level, intelligence, and social class. These judgments are often based on social stereotypes and preconceived notions. Past researches found that the “ing” variable in American English could influence social perceptions of speakers. When speakers use “ing”, they may be perceived as having lower education levels or intelligence. This is primarily because listeners associate this phonetic feature with specific regions or social classes [2]. In multimodal interaction analysis, the language is treated as a multimodal construct, in which speech, intonation, and body language are all considered important symbolic resources. For example, the language and body language of police officers during law enforcement are analyzed together to reveal their meanings in social interactions [2, 3, 50].

The interaction between social and linguistic references is central to sociolinguistic perception. Linguistic variation not only reflects linguistic diversity, but also carries rich social meanings. Social cognitive processes further influence the perception and interpretation of language, and the interaction between language and other symbolic resources constitutes complex social interactions. Understanding these interactions is crucial for revealing how language is used.

B. Social and Linguistic Inferences under Uncertainty

Social references are the prior beliefs or expectations about social groups and their linguistic features. To be specific, social references can be stereotypes and biases which are pre-existing associations between social categories and linguistic features guiding interpretation [58]. Moreover, social references are also dependent on contextual cues

containing visual or situational information that primes specific expectations.

Linguistic references refer to the actual phonetic, syntactic, or lexical features presented in the speech signal. Speech signals are inherently noisy, requiring listeners to weigh cues probabilistically. Listeners could combine multiple sound cues (e.g., vowel quality, pitch, and tone) to reduce uncertainty [59].

Under uncertainty, listeners use these two types of references together to make the best guess about the speaker’s intended message or social identity. Uncertainty here can also relate to the variability within a speaker’s own speech. People do not always speak the same way. There is variation based on the context. Listeners have to account for this variability when interpreting speech, which adds another layer of uncertainty. Some listeners may rely more on social references, while others prioritize linguistic cues. Factors like the listener’s own social background, exposure to different dialects, or cognitive abilities will influence how they handle the uncertainty in sociolinguistic perception. Bayesian models are often used in cognitive science to model how people make inferences under uncertainty. In this sense, the sociolinguistic perception can be explained within the Bayesian framework [59, 60].

In summary, social voice in sociolinguistic perception reveals how language both reflects and reinforces social hierarchies. By understanding the cognitive mechanisms behind these processes, such as automatic stereotyping and contextual flexibility, we can challenge biases and promote equitable communication. The significance of social voice in sociolinguistic perception is profound because it moves beyond simply hearing language to interpreting it through a powerful social lens. It is the mechanism through which language becomes a tool for social navigation, identity formation, and social inequality.

(1) Social voice is our brain’s primary shortcut for understanding “who” a speaker is. We can not process every linguistic feature in isolation. The “social voice”, the package of accent, grammar, and style, allows us to quickly and subconsciously categorize a speaker into social groups (e.g., by region, social class, ethnicity, age, or education level). This categorization instantly activates the social stereotypes and attitudes that we hold about those groups. This is why we can form a detailed impression of a person’s background, character, and intentions within milliseconds of hearing them speak.

(2) Social voice shapes interpersonal and institutional judgments. A social voice can trigger systemic bias before a person being even seen. This is the auditory equivalent of racial profiling. Studies have shown that people make discriminatory decisions about housing, employment, and legal outcomes based solely on a voice heard over the phone.

(3) Social voice reinforces and challenges social power structures. The perception that one social voice (the “standard” dialect) is more “correct”, “logical”, or “beautiful” is a social construct, not a linguistic fact. This ideology grants social and economic power to speakers of the standard variety and marginalizes speakers of others. The “right” social voice acts as symbolic capital. It can open doors to education,

careers, and social mobility. Conversely, a stigmatized social voice can act as a barrier.

(4) Social voice drives language change and variations. Sociolinguistic perception creates a feedback loop. Speakers are not passive. They actively use their awareness of how social voices are perceived to manage their own identity. They may converge their speech toward a listener to build rapport or diverge to emphasize group belonging or distance.

Social voice is the bridge between the abstract system of language and the messy and power-laden reality of human social life. Its perception is a key mechanism through which society is ordered, maintained, and transformed. Future research will prioritize intersectionality, technological ethics, and strategies to amplify marginalized voices, fostering a society where linguistic diversity is celebrated rather than stigmatized.

REFERENCES

- [1] M. Weisbuch and N. Ambady, Thin-slice vision, in *The Science of Social Vision*. New York, NY, USA: Oxford University Press, 2010, 228–247.
- [2] W. Labov, S. Ash, M. Ravindranath, T. Weldon, M. Baranowski, and N. Nagy, Properties of the sociolinguistic monitor, *J. Socioling.*, 2011, 15(4), 431–463.
- [3] D. Hymes, The scope of sociolinguistics, *Int. J. Sociol. Lang.*, 2020, 2020(263), 67–76.
- [4] M. Jessen, Forensic phonetics, *Lang. Linguist. Compass*, 2008, 2(4), 671–711.
- [5] H. Hollien, *The Acoustics of Crime: The New Science of Forensic Phonetics*. New York, NY, USA: Plenum Press, 1990.
- [6] M. Jessen, Speaker classification in forensic phonetics and acoustics, in *Speaker Classification I: Fundamentals, Features, and Methods*. Berlin, Germany: Springer, 2007, 180–204.
- [7] Z. Song, M. Huang, Q. Miao, and F.-Y. Wang, Parallel learning for legal intelligence: A HANOI approach based on unified prompting, *IEEE Trans. Comput. Soc. Syst.*, 2024, 11(2), 2765–2775.
- [8] F.-Y. Wang, Parallel intelligence in metaverses: Welcome to Hanoi! *IEEE Intell. Syst.*, 2022, 37(1), 16–20.
- [9] S. Luthra, J. S. Magnuson, and E. B. Myers, Boosting lexical support does not enhance lexically guided perceptual learning, *J. Exp. Psychol. Learn. Mem. Cognit.*, 2021, 47(4), 685–704.
- [10] Y. Zhou, Y. Liu, and H. Niu, Perceptual characteristics of voice identification in noisy environments, *Appl. Sci.*, 2022, 12(23), 12129.
- [11] T. Kristiansen, Conscious and subconscious attitudes towards English influence in the Nordic countries: Evidence for two levels of language ideology, *Int. J. Sociol. Lang.*, 2010, 2010(204), 59–95.
- [12] K. B. McGowan, Sounding Chinese and listening Chinese: Awareness and knowledge in the laboratory, in *Awareness and Control in Sociolinguistic Research*. Cambridge, UK: Cambridge University Press, 2016, 25–61.
- [13] K. Campbell-Kibler, Deliberative control in audiovisual sociolinguistic perception, *J. Socioling.*, 2021, 25(2), 253–271.
- [14] K. Volzhenin, J. P. Changeux, and G. Dumas, Multilevel development of cognitive abilities in an artificial neural network, *Proc. Natl. Acad. Sci. USA*, 2022, 119(39), e2201304119.
- [15] K. Campbell-Kibler, The nature of sociolinguistic perception, *Lang. Var. Change*, 2009, 21(1), 135–156.
- [16] J. Wang, G. Jin, and W. Li, Changing perceptions of language in sociolinguistics, *Humanit. Soc. Sci. Commun.*, 2023, 10(1), 91.
- [17] L. Milroy, The social categories of race and class: Language ideology and sociolinguistics, in *Sociolinguistics and Social Theory*. London, UK: Routledge, 2014, 251–276.
- [18] J. J. Gumperz and J. Cook-Gumperz, Studying language, culture, and society: Sociolinguistics or linguistic anthropology, *J. Socioling.*, 2008, 12(4), 532–545.
- [19] L. Milroy and C. Llamas, Social networks, in *The Handbook of Language Variation and Change*. Malden, MA, USA: Wiley-Blackwell, 2013, 407–427.
- [20] M. Baynham, Cultural geography and the retheorisation of sociolinguistic space, in *Multilingualism, Discourse, and Ethnography*. New York, NY, USA: Routledge, 2012, 114–130.
- [21] Y. Wang and I. R. Olson, The original social network: White matter and social cognition, *Trends Cognit. Sci.*, 2018, 22(6), 504–516.
- [22] M. Bucholtz and K. Hall, Identity and interaction: A sociocultural linguistic approach, *Disc. Stud.*, 2005, 7(4&5), 585–614.
- [23] D. Hashimoto, Sociolinguistic effects on loanword phonology: Topic in speech and cultural image, *Lab. Phonol.*, 2019, 10(1), 11.
- [24] Y. Kang, Loanword phonology, in *The Blackwell Companion to Phonology*. Malden, MA, USA: Wiley-Blackwell, 2011, 1–25.
- [25] D. LaCharité and C. Paradis, Category preservation and proximity versus phonetic approximation in loanword adaptation, *Linguist. Inq.*, 2005, 36(2), 223–258.
- [26] E. Jessee and J. Calder, The cisgender listening subject in sociolinguistic perception: Transgender identity affects sibilant categorization in American English, *J. Socioling.*, 2025, 29(3), 168–181.
- [27] R. Daland, M. Oh, and L. Davidson, On the relation between speech perception and loanword adaptation, *Nat. Lang. Linguist. Theory*, 2019, 37(3), 825–868.
- [28] M. Aktürk-Drake, The role of perceptual salience in bilingual speakers' integration of illicit long segments in loanwords, *Lingua*, 2014, 143, 162–186.
- [29] S. Peperkamp, I. Vendelin, and K. Nakamura, On the perceptual origin of loanword adaptations: Experimental evidence from Japanese, *Phonology*, 2008, 25(1), 129–164.
- [30] S. Peperkamp and E. Dupoux, Reinterpreting loanword adaptations: The role of perception, in *Proc. of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 2003, 367–369.
- [31] D. Silverman, Multiple scansions in loanword phonology: Evidence from Cantonese, *Phonology*, 1992, 9(2), 289–238.
- [32] C. Paradis and A. Tremblay, Nondistinctive features in loanword adaptation: The unimportance of English aspiration in mandarin Chinese phoneme categorization, in *Loan Phonology*. 2nd ed. Amsterdam, the Netherlands: John Benjamins Publishing Company, 2009, 211–224.
- [33] D. Steriade, Directional asymmetries in place assimilation: A perceptual account, in *The Role of Speech Perception in Phonology*. San Diego, CA, USA: Academic Press, 2001, 219–250.
- [34] F.-F. Hsieh, M. Kenstowicz, and X. Mou, Mandarin adaptations of coda nasals in English loanwords, in *Loan Phonology*, 2nd ed. Amsterdam, the Netherlands: John Benjamins Publishing Company, 2009, 131–154.
- [35] M. Yip, The symbiosis between perception and grammar in loanword phonology, *Lingua*, 2006, 116(7), 950–975.
- [36] M. Kenstowicz, Salience and similarity in loanword adaptation: A case study from Fijian, *Lang. Sci.*, 2007, 29(2&3), 316–340.
- [37] S. Lev-Ari and S. Peperkamp, The influence of inhibitory skill on phonological representations in production and perception, *J. Phonet.*, 2014, 47, 36–46.
- [38] S. Lev-Ari, M. S. Giacomo, and S. Peperkamp, The effect of domain prestige and interlocutors' bilingualism on loanword adaptations, *J. Socioling.*, 2014, 18(5), 658–684.
- [39] M. Aktürk-Drake, Language dominance as a factor in loanword phonology, *Int. J. Biling.*, 2017, 21(5), 584–599.
- [40] R. Daland, M. Oh, and S. Kim, When in doubt, read the instructions: Orthographic effects in loanword adaptation, *Lingua*, 2015, 159, 70–92.
- [41] W. Wang, L. Fan, Y. Wang, Q. Ni, and F.-Y. Wang, Sociolinguistic radar of phonological variation and social meaning: Variables, quantitative methods, and prospects, *IEEE Trans. Comput. Soc. Syst.*, 2024, 11(6), 7734–7741.
- [42] S. Poplack, D. Sankoff, and C. Miller, The social correlates and linguistic processes of lexical borrowing and assimilation, *Linguistics*, 1988, 26(1), 47–104.

- [43] W. Wang, The orthography effect in loanword adaptation: Variable adaptation of English VNV sequences into mandarin Chinese, *J. Psycholing. Res.*, 2025, 54(4), 49.
- [44] L. Fan, C. Zeng, Y. Wang, J. Ma, and F.-Y. Wang, Social radars: Finding targets in cyberspace for cybersecurity, *IEEE/CAA J. Autom. Sin.*, 2024, 11(2), 279–282.
- [45] R. M. Whitaker, G. B. Colombo, L. Turner, Y. Dunham, D. K. Doyle, and E. M. Roy, The coevolution of social networks and cognitive dissonance, *IEEE Trans. Comput. Soc. Syst.*, 2022, 9(2), 376–393.
- [46] E. Brunet-Gouet and J. Decety, Social brain dysfunctions in schizophrenia: A review of neuroimaging studies, *Psychiatry Res.: Neuroimaging*, 2006, 148(2&3), 75–92.
- [47] P. Billeke and F. Aboitiz, Social cognition in schizophrenia: From social stimuli processing to social engagement, *Front. Psychiatry*, 2013, 4, 4.
- [48] Z. Lv, L. Qiao, and H. Lv, Cognitive computing for brain-computer interface-based computational social digital twins systems, *IEEE Trans. Comput. Soc. Syst.*, 2022, 9(6), 1635–1643.
- [49] M. Silverstein, Indexical order and the dialectics of sociolinguistic life, *Lang. and Commun.*, 2003, 23(3&4), 193–229.
- [50] P. Eckert, Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation, *Annu. Rev. Anthropol.*, 2012, 41, 87–100.
- [51] M. Bucholtz, The whiteness of nerds: Superstandard English and racial markedness, *J. Linguist. Anthropol.*, 2001, 11(1), 84–101.
- [52] R. J. Podesva and J. Van Hofwegen, How conservatism and normative gender constrain variation in inland California: The case of /s/, *Univ. Penn. Work. Papers Linguist.*, 2014, 20(2), 130–137.
- [53] L. Winter and J. S. Uleman, When are social judgments made? Evidence for the spontaneousness of trait inferences, *J. Pers. Soc. Psychol.*, 1984, 47(2), 237–252.
- [54] Y.-P. Chen, M. Shaffer, M. Westman, S. Chen, M. Lazarova, and S. Reiche, Family role performance: Scale development and validation, *Appl. Psychol.*, 2014, 63(1), 190–218.
- [55] B. Gawronski, Moral impressions and presumed moral choices: Perceptions of how moral exemplars resolve moral dilemmas, *J. Exp. Soc. Psychol.*, 2022, 99, 104265.
- [56] A. P. Gregg, B. Seibt, and M. R. Banaji, Easier done than undone: Asymmetry in the malleability of implicit preferences, *J. Pers. Soc. Psychol.*, 2006, 90(1), 1–20.
- [57] R. E. Petty and J. A. Krosnick, *Attitude Strength: Antecedents and Consequences*. New York, NY, USA: Psychology Press, 1995.
- [58] D. Nayakankuppam, J. R. Priester, J. H. Kwon, L. A. N. Donovan, and R. E. Petty, Construction and retrieval of evaluative judgments: The attitude strength moderation model, *J. Exp. Soc. Psychol.*, 2018, 76, 54–66.
- [59] D. F. Kleinschmidt, K. Weatherholtz, and T. Florian Jaeger, Sociolinguistic perception as inference under uncertainty, *Topics Cognit. Sci.*, 2018, 10(4), 818–834.
- [60] F.-Y. Wang, K. M. Carley, D. Zeng, and W. Mao, Social computing: From social informatics to social intelligence, *IEEE Intell. Syst.*, 2007, 22(2), 79–83.



Wei Wang received the MA degree in theoretical linguistics from University of Hong Kong, China, in 2018 and the PhD degree in linguistics and applied linguistics from Jilin University, China, in 2022. From 2021 to 2022, he was a visiting scholar at Department of Linguistics, University of Toronto, Canada. He is currently a lecturer at School of Languages and Communication Studies, Beijing Jiaotong University, Beijing, China. His research interests include sociolinguistics, sociophonetics, and the interface between phonetics and phonology.



Intelligence. Her research interests include computer vision tasks based on brain-inspired intelligence, various deep-learning based tasks, image retrieval, semantic segmentation, and object detection.



Jiaxi Liu received the BA degree in sociology from Tulane University, New Orleans, LA, USA, in 2021 and the MS degree in social network analysis from University of Manchester, Manchester, UK, in 2025. Her research interests include social network analysis, application of complex systems in sociology, structural composition, dynamic patterns, societal impacts within social operations, and corresponding countermeasures informed by complex systems approaches.



Hui Yu received the PhD degree in computing science from Brunel University London, Uxbridge, UK, in 2009. He was at University of Glasgow, Glasgow, UK and Queen's University Belfast, Belfast, UK, before joining University of Portsmouth, Portsmouth, UK, in 2012. He is currently a professor of visual computing at University of Portsmouth, UK, and leads the Visual Computing Group. He serves as an associate editor for *IEEE Transactions on Human-Machine Systems*, *IEEE Transactions on Intelligent Vehicles*, and *IEEE Transactions on Computational Social Systems*. His research interests include visual computing, cognition, machine learning, particularly 3D/4D facial expression reconstruction and perception, image analysis for human-machine/social interaction, and intelligent vehicles.



Qinghua Ni received the MS degree from Macau University of Science and Technology, China, in 2022. She is currently pursuing the PhD degree in intelligent science and systems study at Faculty of Innovation Engineering, Macau University of Sciences and Technology, Macao, China. Her research interests include parallel intelligence, blockchain and DAOs, and social computing.



Levente Kovacs received the MSc and PhD degrees from Budapest University of Technology and Economics, Hungary, in 2008 and 2011, respectively, and the Habilitation degree (Hons.) from Óbuda University, Hungary, in 2013. He is currently a professor at Óbuda University, Hungary. He is also a rector of Óbuda University, Hungary and the head of Physiological Controls Research Center. He has been the chair of the IEEE Hungary Section, since 2017. Since 2015, he has been the elected chair of the IEEE SMC Hungary Chapter. In 2018, he established the IEEE CSS Hungary Chapter. His research interests include modern control theory and physiological controls.