The social life of voices: studying the neural bases for the expression and perception of the self and others during spoken communication

Carolyn McGettigan *

Department of Psychology, Royal Holloway, University of London, Egham, UK

Keywords: voice, social neuroscience, speech perception, speech production, functional neuroimaging, identity

Introduction

In 2013, London Underground reinstated the actor Oswald Laurence’s famous “Mind the gap” announcement at Embankment station, having learned that the widow of the actor had been regularly visiting this station since her husband’s death in order to hear his voice again (Hope, 2013). Even in the absence of a personal connection to the couple, it is easy to find this an emotionally affecting story. Anecdotally, “It’s so nice to hear your voice” is commonly encountered in telephone conversations with loved ones, yet there is relatively little known about the cognitive and neural underpinnings of this expression. Similarly, a sense of ownership of one’s voice has important implications—companies like VocalID (www.vocalid.co) have recognized the impact of providing individualized voices to patients who rely upon synthesizers for communication—but, to date, the neuroscience of speech production has been predominantly concerned with the accurate formulation of linguistic messages.

Although there are relatively unchanging aspects of every voice, due to the anatomical constraints of the talker’s vocal tract as well as body size and shape (Kreiman and Sidtis, 2011), it is also important to note that the voice is not a static object. There is no such thing as a passive voice; the voice (like all sounds) demands an action to occur for it to exist at all (Scott and McGettigan, 2015). Much of our vocal expression is the result of voluntary motor acts, which can be modified consciously in response to changes in acoustic, informational and social demands (McGettigan and Scott, 2014; Scott and McGettigan, 2015). Sidtis and Kreiman (2012) write that the voice is “revelatory of ‘self’, mental states, and consciousness,” reflecting “both the speaker and the context in which the voice is produced” (p. 150). It is thus a dynamic self that is modified according to the talker’s goals, affecting both the talker and the addressee in their roles as perceivers and producers of verbal and non-verbal vocal signals.

Disruption to paralinguistic aspects of voice perception and production has implications for psychosocial wellbeing. Most reports of Foreign Accent Syndrome—where patients produce altered speech that perceptually resembles a non-native accent (e.g., due to brain injury, or orofacial surgery)—concentrate on the phonetic, perceptual and neurological correlates of the disorder, yet there is evidence that there can also be significant impacts on the patient’s sense of self-identity (Miller et al., 2011; DiLollo et al., 2014). In voice perception, difficulties in the recognition of emotional and attitudinal prosody have implications for effective psychosocial function in healthy aging, schizophrenia, and autism (Mitchell and Ross, 2013). It is thus crucial that neurobiological accounts of speech and voice processing consider not just what is said, but how it is said, in order to characterize the human aspects of vocal communication behaviors.
Listening to Spoken Selves—The Importance of Personally Familiar Voices

An influential early functional MRI study compared the neural responses to the voices of human men, women and children with those to non-human sounds (Belin et al., 2000). This revealed enhanced activation to voices in bilateral regions of the superior temporal cortex, which became known as the “Temporal Voice Areas” (TVAs). Further work exploring the perceptual processing of individual vocal identities has typically implicated right-dominant activation in the anterior superior temporal sulcus (Belin and Zatorre, 2003; von Kriegstein et al., 2003, 2005; Kriegstein and Giraud, 2004; Schall et al., 2015). More recently, temporal activations were associated with the perception of acoustic differences between voices, while purely identity-related responses were found in right inferior frontal cortex (Latinus et al., 2011). Similar profiles of right-dominant temporal activation have been also observed in the perception of acoustic cues in affective vocal signals, with additional engagement of prefrontal cortex and the limbic system (including the amygdala, and dorsolateral and medial prefrontal cortices, depending on specific task demands; see Brueck et al., 2011). However, the neuroscience of voices and emotion has not yet considered the emotional consequences of hearing other vocal identities, in particular those of highly familiar and valued others.

Presumably due to methodological constraints, the majority of work on paralinguistic voice processing has involved the perception of unfamiliar or newly learned vocal identities. This overlooks the social and emotional salience associated with hearing the voices of trusted friends and loved ones. Sidtis and Kreiman (2012) write: “Personally relevant voices, by definition, are represented in memory with emotional reference to the self” (p. 154). Mechanistically, Kreiman and Sidtis (2011) suggest that unfamiliar voice perception is based on distinguishing local acoustic features, whereas identification of familiar voices involves comparing a heard stimulus to representations in long-term memory, and this dissociation is supported by neuropsychological evidence from cases of phonagnosia (Van Lancker et al., 1988; Garrido et al., 2009). There are also implications from neuroimaging studies that known voices engage higher order responses that could reflect their social salience. For example, personally familiar voices have engaged responses in anterior regions of the temporal lobe, the precuneus and frontal poles (e.g., Nakamura et al., 2001; Shah et al., 2001), which in other literatures have been associated with autobiographical memory and the “social brain” (concerned with the processing of mental states and intentions in others; Blakemore, 2008). To date, however, the use of sets of “commonly familiar” voices including a mix of friends, colleagues, relatives or celebrities has enabled the identification of overall responses to familiarity in vocal signals, but has limited the investigation of the higher-order meaning of those individual voices as social signals for the listener (Sugiura, 2014). Thus, to literature on familiar voice processing has so far offered no clues as to the neural basis for the significance of voices as described in the London Underground story above.

Therefore, there are remaining questions about how familiar voices of different types—family members, friends, colleagues, romantic partners—engage the brain during the perception of vocal signals. A number of existing studies have shown evidence for heightened release of oxytocin—a hormone associated with parental and interpersonal bonding—when participants hear the voice of a loved one (Seltzer et al., 2010). Seltzer and colleagues additionally showed that vocal communication between mothers and daughters is more effective at reducing blood cortisol levels (a marker of stress) than text communication (Seltzer et al., 2012). Abrams and colleagues (Abrams et al., 2013) showed evidence for reduced structural and functional connectivity between posterior temporal regions associated with speech perception and the brain’s reward systems (involving ventral tegmental area (VTA), nucleus accumbens, left insula, orbitofrontal cortex, ventromedial prefrontal cortex) in children with autism, suggesting that this may be related to this group’s relative lack of engagement with spoken signals in day-to-day interactions. There is, however, no detailed account of the behavioral correlates of this functional interaction of speech and reward systems in typical populations. To understand the voice as a social signal, it is essential to investigate the interplay between linguistic and non-linguistic (affective, reward, social) networks during vocal communication. To date, the considerable methodological demands of obtaining controlled, participant-specific vocal recordings of personally valued others has precluded such an investigation (Sidtis and Kreiman, 2012).

Producing the Self Voice—Speaking in a Social Context

Recent neurobiological models of speech production have adopted a forward models approach, in which the brain aims to reduce the error between the predicted and actual sensory consequences of a spoken utterance (Guenther, 2006; Hickok et al., 2011; Guenther and Vladusich, 2012; Hickok, 2012). Here, the goal is the accurate delivery of spoken language at the phonemic and syllabic level. Although what we say (i.e., the choice of words) is important for informational and social exchanges, so too is the way we say it. In vocal communication, we adopt a variety of “selves,” which we use flexibly to achieve the social and informational goals of the conversation, even if the linguistic message itself remains constant across contexts (McGettigan et al., 2013; Hughes et al., 2014)—consider the tone of voice used with a close family member vs. that used with colleagues at work. We carried out the first functional imaging study of this voluntary modulation of the spoken self, using an impersonations task (McGettigan et al., 2013). We found evidence engagement of the left anterior insula and the frontal operculum in the modulation of vocal identity during spoken sentence production, with stronger interaction of these regions with right-dominant superior temporal voice perception areas supporting the emulation of specific target identities. A similar study asking participants to voluntarily introduce a phonological...
Future Directions for the Neuroscience of Vocal Communication

Pardo (2012) writes: “Talkers speak to be understood, and understanding means more than intelligibility” (p. 764). I suggest that this should act as a starting point for the onward development of the neuroscience of human vocal behavior, and propose the following considerations for future work in the area:

- Studies of vocal identity perception should make more regular and selective use of familiar voices, in order to interrogate the interaction of speech/voice perception systems with other response networks relevant to social interactions. It is important to consider that there are different types of familiar person, for whom the perceptual response may systematically differ. Alternatively, Sugiura (2014) suggests that more tightly controlled investigation of the social significance of familiar others could be achieved experimentally by training participants to associate particular social attributes with virtual agents.

- Studies of speech production mechanisms should consider the intended recipient of the spoken message and their relationship to the talker. Here, neuroimaging techniques may offer a means of investigating the interaction of speech perception and production systems with affective, reward and motivational responses, in both the presence and absence of measurable behavioral changes in the phonetic realization of speech.

- The advent of improved methodological approaches to brain imaging during speech production (e.g., Xu et al., 2014) presents mounting pressure to examine vocal behavior in its most typical context: conversation. Garrod and Pickering (2004, 2009) advocate this approach in their Interactive Alignment model of dialog, where it is proposed that conversation proceeds smoothly via the alignment of the interlocutors’ mental models. A recent investigation of brain-to-brain correlations during storytelling has identified significant and extensive coupling between the producer and the comprehender in regions previously associated with higher-order mentalizing and theory of mind tasks (Silbert et al., 2014). Application of such approaches using dyads varying in the type and quality of the interlocutors’ relationship could form a promising avenue to investigate how speech networks interact with social and affective networks during exchanges with familiar others.

References

Abrams, D. A., Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., et al. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. Proc. Natl. Acad. Sci. U.S.A. 110, 12060–12065. doi: 10.1073/pnas.1302982110

Adank, P., Stewart, A. J., Connell, L., and Wood, J. (2013). Accent imitation positively affects language attitudes. Front. Psychol. 4:280. doi: 10.3389/fpsyg.2013.00280

Belin, P., and Zatorre, R. J. (2003). Adaptation to speaker’s voice in right anterior temporal lobe. Neuroreport 14, 2105–2109. doi: 10.1097/00001756-200311140-00019

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature 403, 309–312. doi: 10.1038/35002078

Blakemore, S.-J. (2008). The social brain in adolescence. Nat. Rev. Neurosci. 9, 267–277. doi: 10.1038/nrn2353

Bruck, C., Kreifelts, B., and Wildgruber, D. (2011). Emotional voices in context: a neurobiological model of multimodal affective information...
