The right superior temporal sulcus (STS) is the primary locus of the human voice perception system (Belin et al., 2000). However, methodological issues have prevented prior neuroimaging studies from being able to demonstrate the functional connection between speech and voice perception systems:

Stimulus-Based Approaches contrast activation due to speech vs. non-speech (Belin et al., 2000; Fecteau et al., 2004).

This approach cannot reveal a functional integration because speech and voice are confounded into a single condition.

Task-Based Approaches contrast activation due to attending vocal identity vs. verbal content (Stevens, 2004; von Kriegstein et al., 2003; von Kriegstein & Giraud, 2004).

This approach cannot reveal a functional integration because identical speech information is present in both conditions.

The right superior temporal sulcus (STS) is the primary locus of the human voice perception system (Belin et al., 2000). However, methodological issues have prevented prior neuroimaging studies from being able to demonstrate the functional connection between speech and voice perception systems:

Stimulus-Based Approaches contrast activation due to speech vs. non-speech (Belin et al., 2000; Fecteau et al., 2004).

This approach cannot reveal a functional integration because speech and voice are confounded into a single condition.

Task-Based Approaches contrast activation due to attending vocal identity vs. verbal content (Stevens, 2004; von Kriegstein et al., 2003; von Kriegstein & Giraud, 2004).

This approach cannot reveal a functional integration because identical speech information is present in both conditions.

This study demonstrates increased functional integration between brain regions responsible for speech and voice perception during the identification of talkers speaking a native versus foreign language. Behavioral studies have demonstrated that individuals are more accurate at identifying voices when they understand the language being spoken. Previous neuroimaging studies of voice perception have predominantly contrasted either speech vs. non-speech or voice vs. verbal content. However, these contrasts preclude detecting ways in which speech- and voice-perception systems work in conjunction in a talker identification task. By contrasting voice identification in a native vs. foreign language, we demonstrate that neural systems responding to encoding auditory information onto meaningful structure (i.e. the phonology of one’s native language) contribute more to identification of voices in one’s native language.

Understanding speech requires mapping between high-variability auditory information (intra- and inter-speaker variation) and meaningful (invariant) phonological units.

Specificity to voice has a well-established effect on speech perception abilities.

Specific linguistic experience affects speech perception abilities. Listeners demonstrate a “Language-Familiarity Effect” on talker identification tasks. Voices speaking a familiar language are more accurately identified than those speaking an unfamiliar language (Goggin et al., 1991)

Perrachione & Wong (2007a): The Language-Familiarity Effect is a True Linguistic Effect:

- Exposure to foreign-language voices not enough; specific linguistic knowledge (proficiency) is necessary to over come the Language-Familiarity Effect.
-Listeners of different language backgrounds find different voices more confusable – may attend to different cues.

Suggests a bidirectional integration between speech & voice.

Neuroevolution evidence from talker identification in a dichotic listening paradigm provides early evidence confirming this link. Right-aud (left cerebral hemisphere) accuracy is a better predictor of overall accuracy when identifying voices in a native versus foreign language (Perrachione & Wong, 2007b)