Talking heads: how our perceptions of face and voice melt together.

Experts seem to agree that the ability to recognize faces is rooted deeply in the genes. Our genes not only contain information about the identity of other familiar humans, but also include facial cues such as attractiveness, sex, emotional expression and last but not least, the human voice. The ability to extract these features could have been an important condition in the survival of humans in the course of evolution. Magnetic resonance studies and studies of neurological patients have further presented evidence that the *fusiform face area* (FFA) could be a domain-specific 'hot spot' in our brains, specialized in recognition of faces. In addition, twin studies have shown that competence in face recognition is also an individual heritable trait, comparable to that of general intelligence.

How about voices? In naturalistic conditions the muscles of our head that control mouth, eye movements, eyebrows etc are moving constantly and, more importantly, producing articulated sounds. The human head is thus typically a dynamic 'talking' head, and because movements of the face and articulated speech always occur in temporal alignment, they provide a powerful interactive cue that they belong to the same person.

Belin a.o. have further suggested that in line with face recognition, voice recognition also seems to depend on a specific area or network in our brain. This is the *temporal voice area* (TVA) in the right superior temporal sulcus. Similar to *prosopagnosia* after damage of the FFA, *phonoagnosia* is found after damage of the TVA. Another exciting feature is that these areas or networks seem to be tightly coupled. This implies that when we recognise a familiar face we activate in synchrony the unique vocal signature of the person behind that face. Recognizing a voice would then also have an evolutionary benefit, in helping us to identify a person in suboptimal visual viewing conditions or when visual acuity declines in old age.

Strong face-voice coupling could also explain why our memory of faces is often accompanied by a clear representation of the associated voice . Even when we think about a familiar person we have not seen for many years, we often not

only ‘see’ the face of that person in our minds eye, but also ‘hear’ his or her voice and its typical intonations. That these phenomena are often preserved in old age has to do with the *reminiscence bump*: the increased recollection of older adults to remember events that took place, or friends made during adolescence or early adulthood. This could lead to the strange experience at a school reunion when our old school pall seems to have transformed in another person not matching at all with the vivid and well preserved identity of the younger person in our memory. It’s like we first have to ‘peel of’ the disguise before we can communicate with our friend like we did in the old days.

Coupling between the senses, in particular vision and audition, has been examined in many studies and is often attributed to supra-modal areas in the brain. For instance, the famous *ventriloquism effect* has been ascribed to visual-spatial attention, creating the impression that the voice of the ventriloquist comes from the puppets mouth. Using ERPs, Durk Talsma and I described some 10 years ago how visual-spatial attention modulated ERPs to trains of randomly presented visual as well as auditory stimuli.

Recent evidence however suggest that when faces are coupled with a voice a different mechanism is involved, by which multisensory information is integrated and shared across modalities more directly. For instance, von Kriegstein using brain imaging techniques found that when a subject heard a familiar voice the FFA was automatically activated, even when there was no visual input. Von Kriegstein believes that, at least in naturalistic coupling of faces with voices, we use a network with direct structural connections between face and voice areas to identify a person, independent of the specific sensory input. Schweinberger et al. using ERPs, came to a similar conclusion. He found that audiovisual pairings of a face with a voice produced larger N170 (a specific ERP response to faces) than faces alone. So, taken together these finding indicate that the ability to recognize a person depends on a face/voice network in our temporal cortex. This network is not only capable of storing many (hundreds? thousands?) of unique face/voice pairings over a long period of time, but also of detecting rapidly a person’s identity.

References

Maguinness, C. & Newell, F. N. (2014). Recognising others: Adaptive changes to person recognition throughout the lifespan. In Schwartz, B. L., Howe, M. L., Toglia, M. P., & Otgaar, H. (Eds). *What is adaptive about adaptive memory?* Oxford University Press

von Kriegstein, K., & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology*, *4*(10), e326. v Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129-135.

Von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A.-L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, *17*(3), 367-376. Schweinberger, S. R., Walther, C., Zäske, R., & Kovács, G. (2011). Neural correlates of adaptation to voice identity. *British Journal of Psychology*, *102*(4), 748-764.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129-135.