Title: Understanding audio-visual communication at the level of the single neuron

Nick E. Barraclough, Dengke -K. Xiao, David I. Perrett

Correspondence to nick.barraclough@st-andrews.ac.uk Research Fellow, St Andrews, UK.

Abstract

Communication takes place through visual and auditory modalities. We show here that the sight and sound of vocalisation gestures are integrated by single brain cells in one region of the monkey brain. This brain region is important in humans for decoding speech through visual and auditory modalities. Thus the cellular mechanisms we study in the monkey may form a basis for comprehending communication signals in primates and ultimately the evolution of language in humans.

Document

Primates (including humans) communicate using different techniques. We can threaten an individual with a mean look, call for our mother using our voice, or develop friendships using touch. Often we make use of a combination of our different senses to aid our understanding of communicative signals. For example, watching the lips of someone talking (lip-reading) can markedly improve our perception of speech [1], particularly when we are in a noisy environment. Moreover incongruent information from two senses can upset the combined percept. The combination of heard speech and the sight of a talking face with lips moving incongruently to the speech sound results in a percept that is different from the sight or sound alone [2]. In this article we will discuss recent experiments indicating that both humans and monkeys make use of combined audio-visual cues to understand communicative gestures and present some of our recent data showing responses from single neurones in the rhesus macaque that may process communicative signals.

Monkeys make different calls under different circumstances. Two common calls used by rhesus macaques are the pant-threat and coo, aggressive and affiliative vocalisations respectively [3].

Different calls are often associated with different facial gestures, the facial gesture can help convey the message and is in part a product of the muscle movements used in producing the call. Ghazanfar and Logothetis tested where rhesus macaques looked when they were presented with movies of other macaques making either pant-threats or coos [4]. They edited the movies so that visual pictures of a face performing a specific facial gesture were paired with either the corresponding vocalisation sound or were paired with a sound of a different incongruent vocalisation. Monkeys chose to look at the congruent pairings of facial gestures and vocalisations more often than incongruent facial gesture and vocalisation pairings. This parallels the ability of young human infants to look at film of a face moving congruently with the speech sound [5]. These findings indicate that rhesus macaques and can match the sight and sound of conspecific communicative signals. Ghazanfar and Logothetis make the argument that rhesus macaques detect vocal expression information invariantly across the visual and auditory modalities.

In 1874 Carl Wernike described a syndrome in which patients were unable to understand what was said to them [6]. Subsequent post-mortem analysis showed that his patients had damage to cortex at the top of their temporal lobes. More recently human brain scanning experiments have showed that this region is critical to the processing of audio-visual speech [7, 8]. In an experiment to determine regions of the brain involved in understanding audio-visual speech [8], Calvert looked for a phenomenon in neural responses called ‘supra-additivity’. For those who study how different senses are integrated, supra-additivity is the ‘Holy Grail’ of true multisensory integration. In short multisensory integration occurs when the neural response to the combination of two sensory inputs is greater than the sum of the responses to the two sensory inputs presented alone [9]. Effectively this response signature shows transformation of the different modality stimuli into an integrated product. Calvert looked for the property of supra-additivity in brain activity measured using functional brain resonance imaging (fMRI) when human subjects observed the sight and sound of a

face talking. Human subjects observed audio-visual stimuli of people talking when the sight of the moving lips were matching or not matching the heard speech. The only area of the brain that showed true integration of sight and sound (supra-additivity) was a region or the cortex in the superior temporal sulcus (STS). This region of the STS was selectively active when the sight of the talking face matched the heard speech. This finding shows that the STS is critical for forming neural representations of audio-visual communicative signals, and would be a promising candidate for examining the underlying cellular mechanisms in monkeys.

Many neurons recorded in the STS of rhesus macaques will respond when the monkey is presented with visual, auditory and somatosensory (touch) stimuli [10, 11]. Neurons are sensitive to moving faces and in fact show a fair degree of selectivity for different facial expressions [12-15]. The region also receives substantial input from regions of auditory cortex [16] that contains neurons that are selective for different rhesus macaque vocalisations [17]. Recording the responses of single neurons in the STS cortex when the monkey is presented with audio-visual stimuli can tell us about the neural processing of communicative signals in the monkey's brain.

Using standard single cell recording techniques, we examined responses of STS neurons in rhesus macaques when presented with visual, auditory and combined visual and auditory stimuli. We found the signature of true multisensory integration (supra-additive responses) in a substantial proportion of cells [18, 19]. Figure 1 illustrates the responses of a single STS neuron that integrates sight and sound and could be used for processing communicative signals. The neuron responds to the sight of a monkey face performing a coo gesture, but does not respond to the sound of the coo itself. When these stimuli are combined then the response of the neuron is greater than the sum of the individual unimodal responses, supra-additive. There is little effect of combining the sound of a pant threat with the sight of the face making a coo gesture compared to the response to the coo

gesture alone. In effect this neuron codes the correct matching of the facial gesture of a coo with the specific sound of a coo vocalisation.

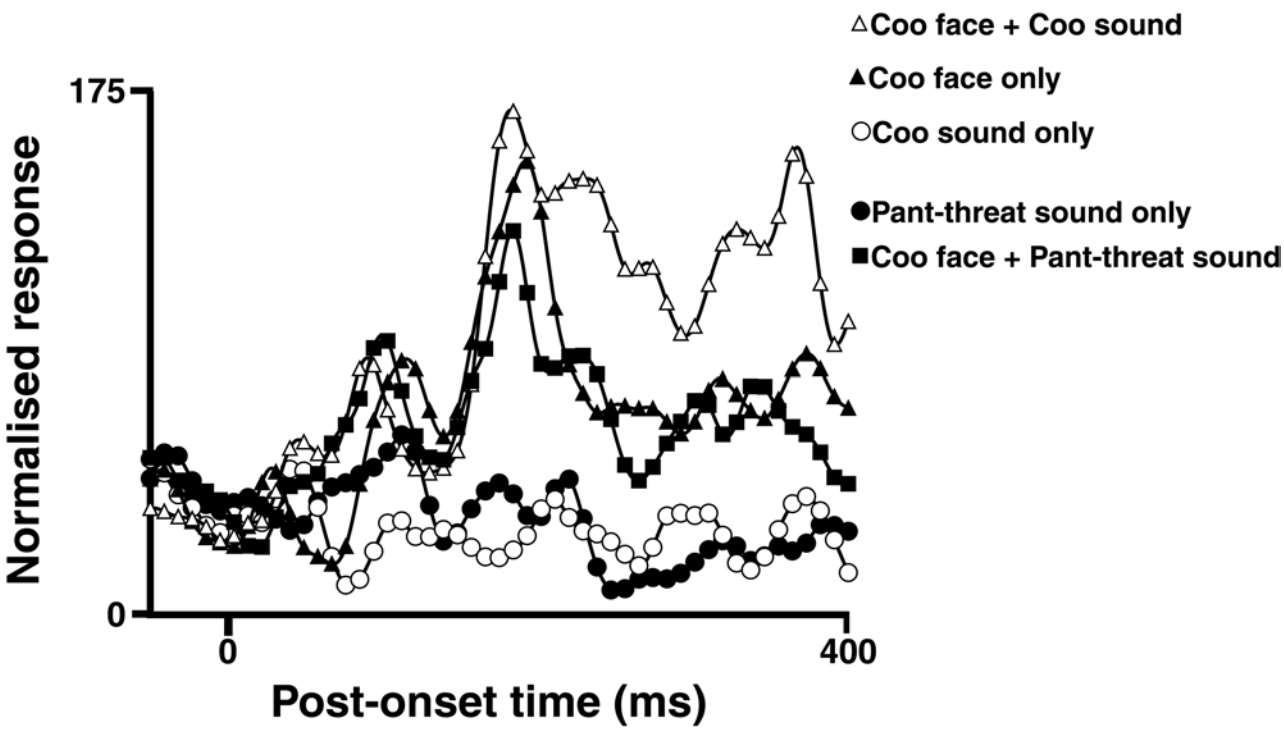

Figure 1. Responses of a neuron showing supra-additive multisensory integration of the sight and sound of coo vocalisations. The neuron responds to the sight of a coo facial gesture (solid triangles), but not to the sound of a coo vocalisation (open circles). There is an increase in the visual response to the sight of the coo facial gesture when combined with the sound of a coo vocalisation (open triangles). There is no response to the sound of a pant-threat vocalisation (solid circles), and no change in the visual response to the sight of the coo facial gesture when combined with the sound of the pant-threat vocalisation (solid squares).

Further support to this type of processing in STS neurons comes additionally from responses of neurons that do not show supra-additive multisensory integration, yet still detect the congruence between sight and sound stimuli. The neuron in Figure 2 responds best to the combined sight and sound of a pant-threat, the combined sight and sound of a coo is less effective. This indicates that the neuron is selective for a specific communicative gesture, the pant-threat. Incongruent combinations of the sight of a pant-threat facial gesture and the sound of a coo, or the sight of a coo

facial gesture and the sound of a pant-threat are less effective than the combined audio-visual pant threat stimulus. Thus even if the gesture contains auditory or visual components that the neuron 'likes', the incongruence of inputs from sight and sound results in a lower firing rate. In effect this neuron detects the mismatch of audio-visual communicative stimuli.

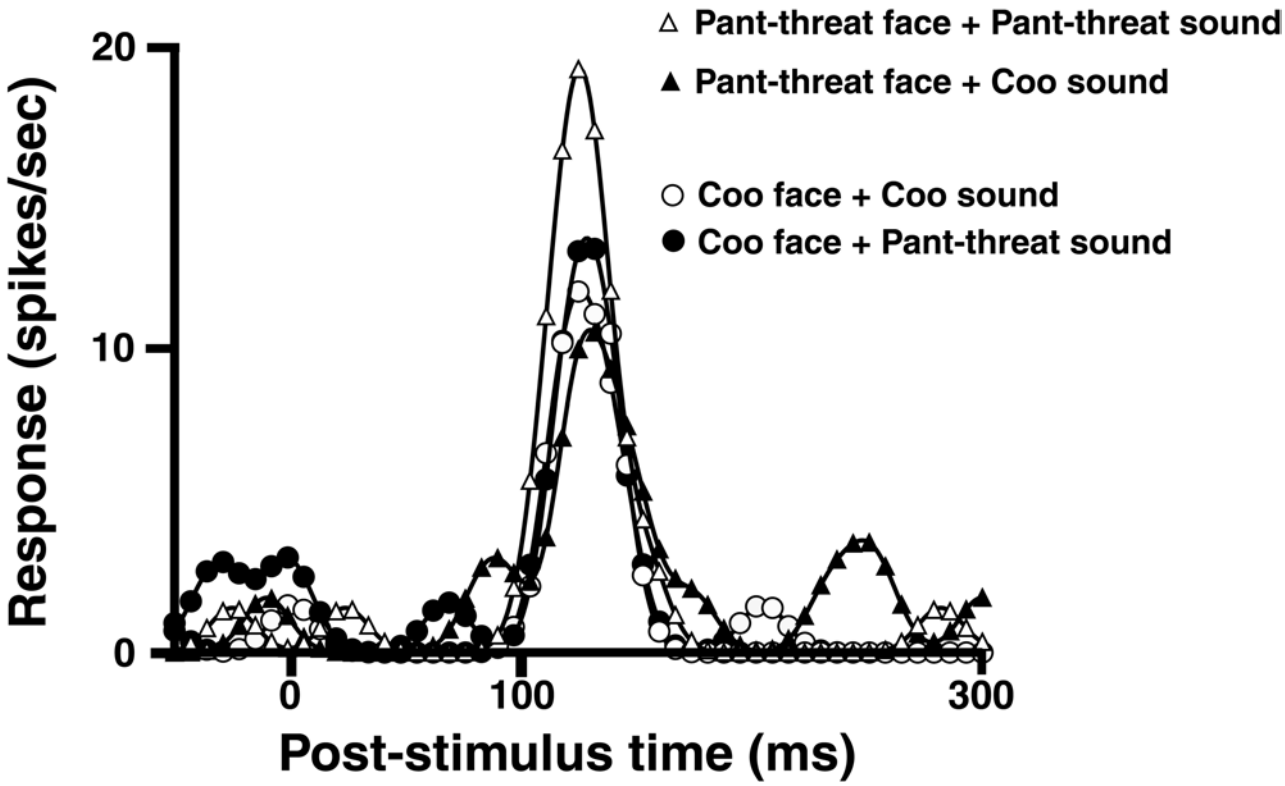

Figure 2. Responses of a neuron to the sight and sound of pant-threat gestures. The best response is to the sight of a pant-threat facial gesture combined with the sound of the pant-threat vocalisation (open triangles). The sight of a coo facial gesture combined with the sound of a coo vocalisation is less effective (open circles). The incongruent combination of the sight of the preferred pant-threat facial gesture and the sound of the coo vocalisation (solid triangles) is less effective. The incongruent combination of the sight of the coo facial gesture and the preferred sound of the pant-threat vocalisation (solid circles) is also less effective.

In summary we have found two types of neuron in the STS that may be involved in the neural processing of audio-visual communicative signals in monkeys. Firstly we find neurons showing

true multisensory integration of visually presented facial gestures and acoustically presented vocalisations. Such cells are relatively frequent, 23% of cells responsive to the sight of gestures are modulated by the sound of the same gestures [19]. Secondly we find neurons selective for specific audio-visual communicative signals that are inhibited by incongruent auditory stimuli.

In theory a network constructed from such neurons would enable the animal to differentiate between different audio-visual gestures, resolve ambiguities between sight and sound and distinguish the sources of acoustically perceived vocalisations. The origins of human language are speculative, but it is likely that the nature of processing of communicative signals in the monkey brain provides a substrate for the evolution of social communication in primates and ultimately language in humans.

References

1. Dodd, B., *The role of vision in the perception of speech.* Perception, 1977. **6**: p. 31-40.
2. McGurk, H. and J. Macdonald, *Hearing lips and seeing voices.* Nature, 1976. **264**: p. 746-748.
3. Rowell, T.E. and R.A. Hinde, *Vocal communication by the rhesus monkey (macaque mullata).* Symposium of the Zoological Society of London, 1962. **8**: p. 91-96.
4. Ghazanfar, A.A. and N.K. Logothetis, *Facial expressions linked to monkey calls.* Nature, 2003. **423**: p. 937-938.
5. Kuhl, P.K. and A.N. Meltzoff, Science, 1982. **218**: p. 1138-1141.
6. Wernicke, C., *Der aphasische symptomenkomplex*. 1874, Breslan: Cohn and Weigart.
7. Calvert, G.A., et al., *Activation of auditory cortex during silent lipreading.* Science, 1997. **276**: p. 593-596.
8. Calvert, G.A., *Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex.* Current Biology, 2000. **10**: p. 649-657.
9. Stein, B.E. and M.A. Meredith, *The merging of the senses*. 1993, Cambridge, MA: MIT Press.
10. Benevento, L.A., et al., *Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey.* Experimental Neurology, 1977. **57**: p. 849-872.
11. Bruce, C.J., R. Desimone, and C.G. Gross, *Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque.* Journal of Neurophysiology, 1981. **46**(2): p. 369-384.

12. Perrett, D.I., et al., *Neurons responsive to faces in the temporal cortex: Studies of functional organsation, sensitivity to identity and relation to perception.* Human Neurobiology, 1984. **3**: p. 197-208.

13. Hasselmo, M.E., E.T. Rolls, and G.C. Baylis, *The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey.* Behavioural Brain Research, 1989. **32**: p. 203-218.

14. Mistlin, A.J. and D.I. Perrett, *Visual and somatosensory processing in the macaque temporal cortex - The role of expectation.* Experimental Brain Research, 1990. **82**(2): p. 437-450.

15. Sugase, Y., et al., *Global and fine information coded by single neurons in the temporal visual cortex.* Nature, 1999. **400**: p. 869-873.

16. Seltzer, B. and D.N. Pandya, *Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex.* Brain Research, 1978. **149**: p. 1-24.

17. Rauschecker, J.P., B. Tian, and M. Hauser, *Processing of complex sounds in the macaque nonprimary auditory cortex.* Science, 1995. **268**: p. 111-114.

18. Barraclough, N.E., et al., *Primate superior temporal sulcus (STS) neurons integrate visual and auditory information for biological motions.* Society for Neuroscience Abstracts, 2003. **29**.

19. Barraclough, N.E., et al., *Integration of visual and auditory information by STS neurons responsive to the sight of actions.* Journal of Cognitive Neuroscience, Submitted.