The Neural Basis of Speech Perception – a view from functional imaging

Sophie K. Scott

1 Institute of Cognitive Neuroscience, University College London

sophie.scott@ucl.ac.uk

Abstract

Functional imaging techniques, such as Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI), have enabled neuroscientists to elaborate how the human brain solves the formidable problem of decoding the speech signal. In this paper I will outline the properties of primate auditory cortex, and use this as an anatomical framework to address the data from functional imaging studies of auditory processing and speech perception. I will outline how at least two different streams of processing can be seen in primary auditory cortex, and that this apparently maps onto two different ways in which the human brain processes speech. I will also address data suggesting that there are considerable hemispheric asymmetries in speech perception.

Index Terms: auditory cortex, speech perception, functional imaging

1. Introduction – auditory anatomy

Speech perception was one of the first cognitive functions to be localized in the human brain, and has long been associated with the left posterior temporal lobe – ‘Wernicke’s area’ [1,2] (Figure 1, A). In contrast, Broca’s area in the left inferior frontal lobe, has been identified as the neural substrate for speech production (Figure 1, B). These studies were based on neuropsychological investigations, which provide excellent profiles of disorders, but are less good as detailed anatomical localization: for example, the causes of the damage to the brain (strokes, head injuries, tumours, infectious agents etc.) do not lead to lesions with simple, clear boundaries and locations.

A complementary method, functional neuroimaging, allows us instead to investigate function in intact human brains. Both common techniques (PET and fMRI) use blood flow as an index of neural activity: where there is more neural ‘work’ being done in the brain, more blood is directed. These local changes in cerebral blood flow can be detected using a radioactive tracer (PET) or using the paramagnetic properties of oxy- and deoxygenated haemoglobin (fMRI) – we are thus working with an indirect measure of neural activity.

Recent work using functional imaging has suggested that Wernicke’s area can be usefully fractionated into different anatomical and functional streams of processing, and that this may be a helpful framework within which to think about speech perception [3] [4]. The body of this paper will address several such studies. However, this ‘streams of processing’ approach has been motivated by parallel developments in the neurophysiology of primate auditory processing, which I will briefly introduce first. Primate sensory cortex (unlike that of other mammals) is arranged in a hierarchical fashion, with input from the thalamic nuclei entering to a core, or primary area: there are then projections from this core area to a surrounding belt area, and then again to a surrounding parabelt area. In the visual system, this translated into a hierarchical system, with a retinotopic representation in primary ‘core’ visual cortex, and more complex properties of the scene (e.g. colour) being processed and represented in surrounding fields [5]. Furthermore, there is an apparent streaming of visual information, with a ventral ‘what’ pathway running along the bottom of the temporal lobes associated with the processing of visual objects (e.g. object recognition) and a dorsal ‘where/how’ pathway running up into the parietal cortex, associated with processing spatial layouts and sensory-motor interactions (e.g. picking up an object) [6].

1. Introduction – auditory anatomy

Speech perception was one of the first cognitive functions to be localized in the human brain, and has long been associated with the left posterior temporal lobe – ‘Wernicke’s area’ [1,2] (Figure 1, A). In contrast, Broca’s area in the left inferior frontal lobe, has been identified as the neural substrate for speech production (Figure 1, B). These studies were based on neuropsychological investigations, which provide excellent profiles of disorders, but are less good as detailed anatomical localization: for example, the causes of the damage to the brain (strokes, head injuries, tumours, infectious agents etc.) do not lead to lesions with simple, clear boundaries and locations.

A complementary method, functional neuroimaging, allows us instead to investigate function in intact human brains. Both common techniques (PET and fMRI) use blood flow as an index of neural activity: where there is more neural ‘work’ being done in the brain, more blood is directed. These local changes in cerebral blood flow can be detected using a radioactive tracer (PET) or using the paramagnetic properties of oxy- and deoxygenated haemoglobin (fMRI) – we are thus working with an indirect measure of neural activity.

Recent work using functional imaging has suggested that Wernicke’s area can be usefully fractionated into different anatomical and functional streams of processing, and that this may be a helpful framework within which to think about speech perception [3] [4]. The body of this paper will address several such studies. However, this ‘streams of processing’ approach has been motivated by parallel developments in the neurophysiology of primate auditory processing, which I will briefly introduce first. Primate sensory cortex (unlike that of other mammals) is arranged in a hierarchical fashion, with input from the thalamic nuclei entering to a core, or primary area: there are then projections from this core area to a surrounding belt area, and then again to a surrounding parabelt area. In the visual system, this translated into a hierarchical system, with a retinotopic representation in primary ‘core’ visual cortex, and more complex properties of the scene (e.g. colour) being processed and represented in surrounding fields [5]. Furthermore, there is an apparent streaming of visual information, with a ventral ‘what’ pathway running along the bottom of the temporal lobes associated with the processing of visual objects (e.g. object recognition) and a dorsal ‘where/how’ pathway running up into the parietal cortex, associated with processing spatial layouts and sensory-motor interactions (e.g. picking up an object) [6].

Figure 1: lateral view of the human cortex, showing the core of Wernicke’s area (A) and Broca’s area (B).

Figure 2: lateral view of the non-human primate cortex, showing the supra temporal plane (extended towards viewer), primary auditory cortex (PAC) and superior temporal sulcus.

For practical reasons, work dissociating streams of processing in the primate auditory system has lagged somewhat behind the many developments in visual neurophysiology. Primate primary auditory cortex (PAC)
is located on the supratemporal plane, making it difficult to access (figure 2), and it does not show a strong response to pure sine tones (which were typically used as simple stimuli, analogous to those used in visual experiments) [7]. Recent technical developments have permitted these problems to be overcome, along with the crucial realization that pure sine tones were not the optimal stimulus for driving auditory cortical responses [7], since most sounds encountered in the environment – especially communicative vocalisations – are highly complex.

Furthermore, unlike the visual system, the auditory system computes aspects of sounds in the ascending auditory pathway, such that the representation at the cortical level is typically of a more complex auditory object [8, 9] – this has implications for the optimal stimuli to drive neurons in PAC.

In addition to these, there is also a suggestion of further types of auditory processing: in the medial caudal belt regions, auditory neurons have been identified which respond to touch [13]. This may form a route for sensorimotor integration.

Since these developments, it has become clear that there are some general parallels between the cortical organization of sound and vision, in the sense that both hierarchical and parallel processing can be seen in auditory processing, and this has some degree of functional specialization. The core of auditory cortex receives all its input from the auditory thalamus, and is organized as three distinct core fields, in a rostral-caudal orientation (Figure 3), with tonotopic organisation [10]. These core areas are connected to surrounding belt and parabelt regions: the rostral-caudal organization is maintained within these connections, such that the most rostral core field projects to the rostral and rostral lateral belt, but not to the caudal belt, and vice versa. The same is true for the projections from the belt and core to parabelt regions. In terms of connectivity, there is thus both hierarchical and parallel processing in primate auditory cortex. The rostral-caudal organization of primate auditory cortex is also maintained in the projections to frontal cortex: there are projections from rostral belt and parabelt to ventral prefrontal regions, and adjacent but none overlapping projections to dorsal prefrontal and premotor cortex from caudal belt and parabelt [11] (Figure 4).

These two broad streams of processing are associated with different functions. Going from medial to lateral auditory fields, responses are driven by progressively wider bandwidth sounds [7]. It has been demonstrated that antero-lateral parabelt cells respond to conspecific vocalizations (i.e. noise made by other monkeys), while a greater number of more posterior lateral parabelt cells preferentially to the localization of calls (i.e. where the other monkeys are) [12]. This is strong evidence for functional streams, perhaps similar to those in the visual system, with the anterior ‘what’ stream being functionally and anatomically distinct from a posterior ‘where’ stream.

Figure 3: schematic view of auditory cortical fields, showing the core, belt and parabelt regions: the rostral parabelt (RP) and caudal parabelt (CP) are also shown [10].

Figure 4: view of left non-human primate cortex, showing the projections from the rostral belt and parabelt regions, and caudal belt and parabelt regions to frontal cortex [11].

2. Functional imaging studies of human auditory processing – sounds and speech

2.1. Auditory processing in humans

Generally, studies of human auditory processing have found similar profiles to those seen in non-human primate studies. When subjects are presented with a sound, and their neural activity is compared to another condition where they heard nothing, then activation is seen in primary auditory cortex. The activation almost always extends beyond primary auditory cortex, however. Studies addressing more complex auditory properties have shown that harmonic structure, amplitude modulation and frequency modulation are associated with peaks of activation lying outwith primary auditory cortex [14,15]. While we do not have the detailed neuroanatomical data on humans that we have on non-human primates, a meta-analysis of these studies has indicated that the peaks generally lie lateral to PAC [4]. These may correspond to belt and parabelt regions in the human auditory cortex. Interestingly, spectral modulation (relative to spectrally constant sounds) leads to peaks lying yet more laterally and extending further forward down the temporal lobe [16]. These studies provide intriguing evidence therefore that human auditory cortex shows the hierarchical processing of sound that is characteristic of the primate brain, as more complex sounds are processed further away from PAC. Furthermore, as the sounds become more speech-like, with the introduction of spectral variation, the activity runs further forward along the temporal lobe – suggesting that speech-like information may be processed along the anterior ‘what’ stream of processing [4].

2.2. Speech processing in humans

In an early PET study of speech perception [17], we compared the neural activation seen when subjects heard single words to that seen when the subjects heard signal correlated noise analogues of the speech [18]. This
revealed bilateral activation lying lateral to PAC in the left and right temporal lobes; while the activation extended posterior to PAC on the left, we were surprised to see that the activation on the left and the right also ran anterior to PAC. This was unexpected since, as indicated in the introduction, the focus of speech perception has long been on posterior temporal lobe areas. We were also intrigued that the results were so bilateral, since the evidence was so strong that speech perception is a left hemisphere property. The results of any PET or fMRI study are, however, highly dependent on the baselines used, and we were concerned therefore that our baseline, signal correlated noise might not match for all the acoustic features of speech. For example, it does not have the spectral structure of speech, nor a strong sense of pitch. We therefore ran another PET [19] study in which we used the BKB sentences as speech stimuli, and used spectral rotation as the baseline condition [20]. Spectral rotation has the advantage of matching well for the acoustic structure of speech, with preserved spectral complexity, and some degree of pitch and pitch variation. We also used a different intelligible speech condition, six-channel noise vocoded speech, and rotated noise vocoded speech as the baseline comparison. Noise vocoded speech [21] sounds like a harsh whisper, and it does not have a strong sense of pitch, but it can be well understood with little training. Examples of all the stimuli can be found here: www.phon.ucl.ac.uk/hom/brain. With this study, therefore, we wanted to be able to identify the brain areas which were sensitive to the intelligible speech conditions, but not to baseline acoustic conditions which were as well matched to the speech as possible.

The results of this study indicated that an area in the left superior temporal sulcus (STS), lateral and anterior to PAC, responded equally well to speech and noise vocoded speech, but not to either of the baseline conditions (Figure 5). With this response lying forward down the temporal lobe, our finding looked strikingly like the data from non-human primates that a ‘what’ stream of processing is involved in the early acoustic processing of conspecific vocal communication. Importantly, we also found that the right temporal lobe responded strongly to speech, the difference being that the response in the right also strong to the spectrally rotated speech: conversely, the response to both intelligible and unintelligible noise vocoded conditions was weak on the right. While there are several differences between the speech and the noise vocoded speech, it is striking that the rotated speech preserves much of the sense of pitch, and the pitch variation of the original speech (unlike noise vocoding to six channels): our tentative interpretation of this finding was that the right auditory areas were sensitive to the prosody of the speech, rather than its linguistic intelligibility per se. This result is congruous with patient studies indicating that the right temporal lobe is necessary not to detect pitch, but to compute and process pitch direction [22]: likewise, there are several studies of musical processing which suggest that pitch variation (even if extremely simple) recruits right temporal lobe mechanisms [23]. We concluded therefore that left temporal lobe areas were preferentially processing the linguistic content of the speech (both in normal and noise vocoded speech), and the right temporal lobe areas were processing the prosodic aspects of the speech, in both intelligible and unintelligible rotated speech. Furthermore, both processes were dependent upon the anterior ‘what’ processing streams in auditory cortex. In a more recent study, we investigated the neural processes underlying speech by varying intelligibility along a continuum: we used different numbers of channels of noise vocoded speech (from one to sixteen channels) and identified brain areas which correlated in activity with the associated increases in intelligibility (controlling for the corresponding increases in acoustic complexity) [24]. This revealed activation in the left and right temporal lobes. In the left temporal lobe, there was activation in lateral auditory areas which, while sensitive to the intelligibility of the speech, also reflected sensitivity to the acoustic structure. Moving forwards down the left temporal lobe, the peaks of activity showed progressively less sensitivity to the acoustic structure of the stimuli, until in the left temporal pole (the furthest peak) the activation followed the associated intelligibility of the stimuli – a logarithmic relationship between channel numbers and intelligibility, and no sensitivity to acoustic details beyond this. In contrast, the peak of activation in the right temporal lobe reflected an enhanced response to any spectral structure (i.e. more than one channel), and an enhanced response to the sixteen channel condition. This may again reflect a greater response in the right temporal lobe to sequences with pitch variation (present to some extent in any sequence with some spectral detail). The enhanced response to the sixteen channel condition, which not only can be understood, but in which the speaker identity is evident, may also reflect a greater right hemisphere involvement in speaker recognition mechanisms [25].

These findings revealed a stream of information, running down the left temporal lobe from back to front, with some initial sensitivity to speech early in the pathway, lateral to PAC, and responses down the pathway becoming more sensitive to purely linguistic, rather than acoustic features of the speech. This is consistent both with studies showing apparent tuning in early auditory cortex to the acoustic-phonetic structure of speech [26], and with studies indicating that the anterior temporal lobe represents auditory ‘word’ images in speech perception [27]. This study also indicated that again, there are differences in the ways that they left and right temporal lobes process speech – with the left temporal lobe associated with linguistic processing, and the right associated with other kinds of information carried by the voice, including prosody and speaker identity. These studies both addressed ‘bottom up’ aspects of speech perception. Of course, in normal speech perception, there are always top-down processes at work – we find it easier to understand words in sentences than in isolation, and easier to understand predictable sentences than unpredictable sentences [28]. The final study I am going to discuss was an attempt to determine the neural basis for these top down effects in speech perception. Again, we used noise vocoded sentences, this time in an fMRI study.
We had three noise vocoding conditions – 2, 8 and 32 channels. The stimuli were the SPIN sentences (with low and high predictability sentences), and the design of the study was to identify neural responses associated with the conditions where the predictability of the sentences improved the intelligibility of the sentences. From piloting, we determined that in the two channel conditions performance was poor, regardless of predictability, while in the 32 channel conditions, performance was at ceiling, regardless of intelligibility. In the 8 channel conditions, performance was dramatically improved by predictability: the low predictability sentences were recognized at around 50%, and the high predictability sentences were recognized at nearly 100%.

Our analysis therefore distinguished between brain areas associated with the increase in channels (in the left and right temporal lobes), and the brain areas where there was an interaction between channel number and predictability, such that predictability enhanced performance. This revealed a left hemisphere network of areas, specifically recruited when intelligibility was increased by predictability (and not in the 2 channel and 16 channel conditions) (Figure 6). This network comprised activation in the medial and inferior frontal lobes (areas previously linked to controlled semantic processing), the inferior parietal lobe (linked to working memory processes) and posterior cingulate (linked to memory retrieval processes). While we cannot determine what linguistic aspects of ‘predictability’ are associated with these areas, this study does enable us to see how such a distributed network is brought to bear when we passively listen to sentences, and thus how the responses seen to speech in temporal lobe auditory areas is modulated by non-auditory, non-perceptual processes.

So far, all the temporal lobe activation I have discussed has been associated with anterior ‘what’ processing streams – this is because these areas are associated with speech intelligibility processing in the human brain. However, these are not the only areas associated with speech. In previous studies, several people have suggested that posterior auditory areas are important in the linking of speech production and perception [30, 31, 32]. In the primate brain, posterior auditory fields have been linked to both spatial and auditory-tactile links [13]. In the human brain, posterior and medial auditory fields have been show to be activated in speech production, even when the speech is silently mouthed (Figure 7). The activation of auditory cortex by speech motor acts has been interpreted as a sensori-motor link (a ‘how’ stream of processing), and one which may be linked, for example, to articulatory processing in working memory [31]. Current work in my lab is attempting to determine the role of this area in phenomena like delayed auditory feedback.

Figure 6: view of left human cortex, showing the areas recruited when the predictability of sentences improves their comprehension. IFG – inferior frontal gyrus, MPFC – medial prefrontal cortex, AG – angular gyrus (in the inferior parietal lobe) and PC – posterior cingulate [29].

Figure 7: view of left human cortex, showing the posterior and medial auditory area (pAC) recruited during speech production [30].

3. Discussion and Conclusions

In this paper I have tried to demonstrate that, as in the primate brain, human auditory cortex is associated with both hierarchical processing and streams of processing. Furthermore, there appear to be different ways in which the left and right temporal lobes deal with heard speech – with the left apparently interested in the linguistic aspects of speech, and the right interested in the non-linguistic aspects of prosody, and speaker identity. In many ways the patterns we see reflect the different streams of processing seen in the primate brain. While this approach remains controversial [33], it does allow us to use anatomical frameworks and models to interpret functional imaging data, and to reduce over-reliance on Broca’s and Wernicke’s areas as explanatory constructs.

4. Acknowledgements

SKS is funded by an SRF from the Wellcome Trust.

5. References


