

OPINION

A little more conversation, a little less action — candidate roles for the motor cortex in speech perception

Sophie K. Scott, Carolyn McGettigan and Frank Eisner

Abstract | The motor theory of speech perception assumes that activation of the motor system is essential in the perception of speech. However, deficits in speech perception and comprehension do not arise from damage that is restricted to the motor cortex, few functional imaging studies reveal activity in the motor cortex during speech perception, and the motor cortex is strongly activated by many different sound categories. Here, we evaluate alternative roles for the motor cortex in spoken communication and suggest a specific role in sensorimotor processing in conversation. We argue that motor cortex activation is essential in joint speech, particularly for the timing of turn taking.

Spoken communication relies on the listener rapidly decoding the signal that is produced by the speaker. The apparent ease of speech production and perception processes belies the complexity of the motor acts that are necessary to produce speech and of the resultant acoustic signal that the listener processes. Indeed, specific speech sounds (phones) can be hard to separate and identify from the speech signal. This is because individual speech sounds can be produced in a variety of ways. The way that sounds are produced varies with their position in a word: in British English, the phone /p/ in 'port' is quite unlike the /p/ sound in 'sport' — for example, in the former it is aspirated (produced with a puff of air) and in the latter it is not. Speech sounds also vary according to the surrounding phonemes, so the /s/ at the start of 'sue' is acoustically different from the /s/ at the start of 'see', as the position of the lips anticipates the following vowel. In continuous speech, we can consider the sounds of speech to run into each other and influence each other, similar to letters in cursive handwriting as compared with printed letters¹.

Experimental evidence of the variability in speech sounds was very striking to the first researchers who were able to investigate

the structure of speech in spectrograms² (BOX 1), and it was proposed that the listener tracks the articulatory gestures (that is, the movements of the articulators — the larynx, jaw, soft palate, lips and tongue — that are needed to produce the words that are being spoken) that the speaker aims to make when speaking, even if these gestures are not fully realized. One central feature of this 'motor theory of speech perception' is that speech is special in the sense that it is processed differently to other acoustic signals. Specifically, motor theorists state that speech is perceived as the gestures of the articulators, and that this process recruits the motor system³⁻⁵. The history of the motor theory is well described by Galantucci, Fowler and Turvey⁶. They note that it has proved difficult to show that the motor acts of speech and the resulting sounds can ever really dissociate, because speech sounds are intrinsically linked to the actions that produce them⁶. Nonetheless, the motor theory has remained an important approach to our understanding of speech processing.

An alternative view is that speech perception involves acoustic processing of the signal⁷. According to this 'acoustic' perspective, speech is perceived by analysing its spectrotemporal properties. Constellations of

acoustic cues, such as voicing, spectral centre of gravity and amplitude, are mapped onto perceptual categories, such as phoneme, diphone and syllable. No one acoustic feature is dominant in this process: even a simple phonetic contrast such as the difference between /aba/ and /apa/ is distinguished by at least sixteen different acoustic cues⁸. Speech perception is thus viewed as the recognition of complex acoustic patterns, which occurs entirely in the auditory system and does not involve the motor system. This would translate in anatomical terms into speech processing being dependent on auditory areas and auditory association areas in the temporal lobes rather than the motor and premotor cortex^{9,10}.

In this Perspective, we critically evaluate the role of the motor cortex in speech processing. Two recent papers have specifically criticized the evidence for roles for the motor systems in human perception¹¹, with particular reference to speech¹². Here, we not only assess evidence for and against the motor theory of speech perception, but also attempt to integrate other possible roles for motor processes in speech, and discuss the extent to which motor processes can be considered essential¹³ to speech processing.

Motor responses to speech

Evidence for motor cortex involvement in speech perception. The motor theory of speech perception has received a lot of support recently from studies of the motor system and its response to heard speech⁶. Here we define the motor system as the primary motor cortex (Brodmann's area (BA) 4) and the premotor cortex (BA 6 and BA 44), including the supplementary motor area but excluding Broca's area in its widest sense (that is, BA 45 as well as BA 44) because it is not consistently activated by speech production¹⁴. These regions form part of the 'mirror' or action-execution system — that is, cortical fields that are activated during both the performance and the perception of a motor act. Several studies have shown that hearing speech increases activation in the motor and premotor cortices, both in terms of enhanced electromyography measurements from the muscles around the mouth in transcranial magnetic stimulation (TMS) studies¹⁵ and in terms of peaks of activation

in the motor cortex in functional MRI (fMRI) studies^{16,17}. Disruption of premotor cortex activity with TMS has been shown to detrimentally affect the discrimination of stop consonants in syllables that were masked with white noise¹³. Motor responses to speech are specific to the ways in which speech sounds are produced: speech sounds that are made in different ways (for example, /p/ and /t/) differentially activate the motor areas that are associated with different articulators¹⁸. For example, the pattern of motor cortex response to hearing the bilabial plosive /p/ is different from the motor cortex response to hearing the alveolar plosive /t/⁶⁹. Together, these studies support the hypothesis that the motor system has a role in speech perception.

Anatomical links between audition and the motor cortex. In terms of neuroanatomy, there is evidence of a role for both acoustic

and motoric processing of speech and sound⁹. In the primate temporal lobe there are at least two streams of processing from the primary auditory cortex (FIG. 1): one stream runs down the temporal lobe and is sensitive to aspects of the acoustic information in sounds and speech (a ‘what’ stream); another runs posterior from the primary auditory cortex and shows properties of sensorimotor interactions for speech and other sounds (a ‘how’ pathway)^{9,10}. Although the two auditory streams interact functionally¹⁹, they also show different response profiles in functional imaging experiments. For example, the anterior stream is sensitive to the intelligibility of speech²⁰, whereas the posterior route is sensitive to the perceived difficulty of production of non-native phonemes¹⁷ and is activated by silent articulation²¹. Speech has thus been argued to be processed both as a sound (‘what’) and as an action (‘how’)⁹. Evidence from non-human

primate studies indicates that these anterior and posterior streams project to adjacent but non-overlapping regions in the frontal cortex, and that the posterior stream specifically projects to the premotor cortex²². The posterior pathway is thus an anatomical pathway for the gestural processing of speech — as predicted by the motor theory³ — and for the motor sensitivity to speech sounds, which has been reported in studies^{16,17}.

There is evidence that the ‘how’ pathway is less available for conscious processing than the ‘what’ pathway. For example, listeners can accurately tap along to sound sequences, and accurately change the timing of their taps to track small timing changes in the sequence (a sensorimotor task). However, in a purely perceptual listening task with the same stimuli (and no tapping), the same listeners need larger timing changes in order to report hearing a difference²³. Likewise, in speech production speakers can correct their articulations to compensate for online acoustic distortions of the sounds they are making without being aware that these distortions are occurring²⁴. Speaking when one’s speech is distorted in real time strongly activates posterior auditory areas²⁴, which implicates the ‘how’ pathway in both detecting and compensating for distortion of speech output. By contrast, perceptual identification and recognition of the sounds of speech primarily involves temporal lobe areas lateral and anterior to the primary auditory cortex that are part of the anterior ‘what’ pathway^{20,25}.

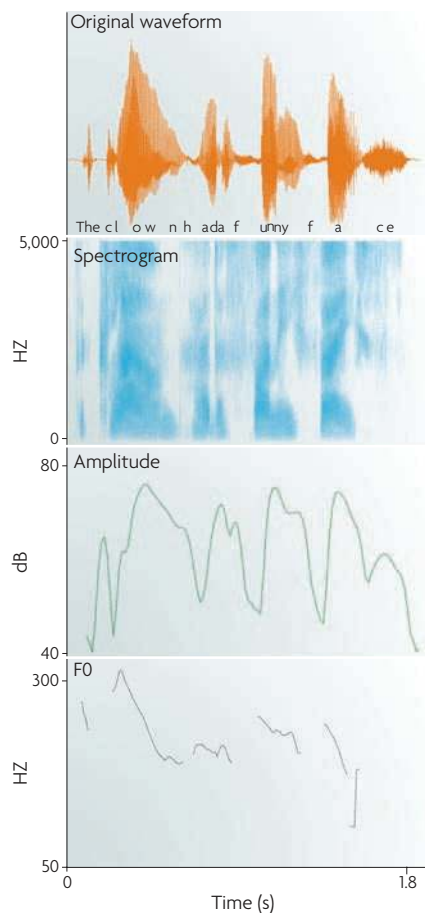
Dissociation of speech input and output systems in neuropsychology. The sensorimotor (how) stream of auditory processing clearly provides an anatomical route for linking speech perception and speech production and for motor representations to play a part in the perception of speech. However, what remains unclear is what that role might be — are motor processes truly essential for speech perception¹³ or is their involvement less central to the perception of speech and more important in other linguistic and non-linguistic computations (for example, sequence processing or semantic representations)? Certainly, there is considerable clinical evidence (see below) that impairments in speech comprehension can be dissociated from impairments in speech production. Assuming that speech comprehension is preceded by perception, this dissociation would argue against a central role of motor processes in speech perception. For example, neuropsychological evidence suggests that disorders of the motor control of speech output are associated with anterior brain

Box 1 | **Structure in sound**

Every sound informs the listener about the actions and objects that made it (‘action to sound’). Striking a bell produces a characteristic sound with a sharp ‘attack’ time (which reflects the striking action) and the simultaneous onset of many harmonics (reflecting the metal with which the bell is made and the resonance characteristics of its shape). We are sensitive to both physical and dynamic information in sounds — for example, we can hear the acoustic differences between bouncing and breaking objects⁸².

The way a sound is made also affects how we interact with it (‘sound to action’): people move in time to musical sequences based on when they hear the ‘beat’ of the sounds rather than the sound’s physical onset^{83,84}. In turn, these perceptual beats depend on how the sound is made: the sharp attack of a drum leads to an earlier beat than that of a bowed note on a violin. Thus, different orchestral instruments are physically sounded at varying times in order to seem ‘in time’ with each other in an ensemble⁸⁵.

In speech there are different kinds of structural information, which are to some degree independent (see the figure). The complex phonetic information is created by the laryngeal sounds being filtered by the positioning and movement of the articulators (shown in the spectrogram). Pitch variation (represented by the fundamental frequency, F0) is used to express meaning and emphasis, and carries linguistic meaning in some languages. As in music, speech also has a rhythm, in which the perceptual beats correspond to the properties and timing of the syllables^{81,86}. Acoustically, as in music, these beats correspond broadly to aspects of the amplitude envelope, especially aspects associated with vowel onsets^{81,86}. Brain responses to speech could be differentially tracking any or all of these factors — for example, the left superior temporal sulcus preferentially processes the phonetic information in speech^{20,25,39,40}.



lesions²⁶ and do not compromise speech comprehension, which implies that the perceptual systems that underlie comprehension are unimpaired in patients with such disorders. Thus, patients with anterior brain lesions, leading to expressive aphasia, perform normally in tests of auditory speech perception and comprehension — for example, auditory single word–picture matching^{27,28}. Conversely, patients with receptive aphasia following a stroke show deficits in speech perception and comprehension (for example, poor scores on auditory single word–picture matching²⁸) whereas their speech production is fluent, if nonsensical²⁹ (unlike that of patients with an expressive aphasia). Both receptive and expressive aphasic patients show abnormalities in the explicit perceptual categorization of speech sounds (for example, labelling a sound as /da/ or /ga/), which would, for the expressive aphasic patients, suggest a role for motor control areas in speech perception. However, these deficits do not correlate with speech-comprehension measures in either patient group³⁰, and it has been suggested that the categorization task is not measuring the speech-perception processes that ordinarily contribute to speech understanding^{10,11,17}. Thus, from the clinical literature it seems that motor processes that are involved in speech production do not directly contribute to speech perception and comprehension^{10,11}.

Dissociation of speech input and output systems in development. There is also scant evidence that motor processes have a central role in the development of speech perception and comprehension. A lack of auditory stimulation during development has detrimental effects on speech production, but a lack of speech motor output does not necessarily affect speech perception and comprehension. Thus, children born with hearing loss can learn to speak because they have somatosensory feedback, but it is difficult (even with relatively moderate hearing loss) and can take longer than in hearing children³¹. By contrast, individuals who have grown up with severe dysarthria or anarthria — that is, with very impaired or no speech production ability — can have intact speech comprehension³². These clinical findings suggest that speech comprehension and production dissociate in development. In normally developing infants, speech-perception skills are in place well before speech-production skills³³. Vowel-perception performance at 6 months predicts later vocabulary acquisition³⁴, suggesting that the detailed perceptual processing of speech underlies early language development,

before skilled speech-production abilities are acquired. Studies have shown that although there are links between speech-production development and other manual motor skills³⁵, variability in speech perception and comprehension (at 21 months) correlates with neither speech-production measures nor measures of oro-facial or manual motor skills. Instead, speech-perception skills vary with cognitive measures and socio-economic status³⁶. If there are interactions between perceptual processes and motor representations in speech perception, they are not factors that drive variation in the development of spoken language.

Functional neuroimaging studies of speech perception. Speech perception was one of the earliest topics to be addressed with functional brain imaging, possibly owing to the successful delineation of cortical areas associated with speech in neuropsychological studies³⁰. Early speech-perception studies revealed extensive bilateral activation of the dorsolateral temporal lobes, including the primary auditory cortex, in response to hearing speech as compared with a silent baseline condition³⁷. More recent studies have narrowed this activity down to regions in the left (and sometimes the right) superior temporal sulcus (STS)^{20,21,25,38–42}. It is notable that few of these studies of speech perception showed motor cortex activity at a whole-brain level of analysis (such responses were not commonly investigated with region-of-interest analyses), and the discussions correspondingly tend not to focus on motor theories of speech perception. As noted in a recent meta-analysis of the pre-lexical processing of speech⁴³, speech perception is mainly associated with activation in dorsolateral temporal lobe regions, with an emphasis on the STS.

However, the pattern of activation to heard speech can extend beyond the dorsolateral temporal lobes — for example, into anterior and ventral temporal lobe regions²⁵ that are associated with semantic knowledge⁴⁴. Moreover, the cortical activation in response to speech can vary with aspects of the task — for example, frontal activation extending into the premotor cortex is seen if participants actively try to detect or classify speech sounds¹⁰ or if the speech is distorted in some way^{41,45}. Some authors have expressly addressed the issue of how task structure affects the involvement of the motor cortex in speech perception^{10,46}, and we consider this issue in the discussion of candidate roles for motor systems in language processing below.

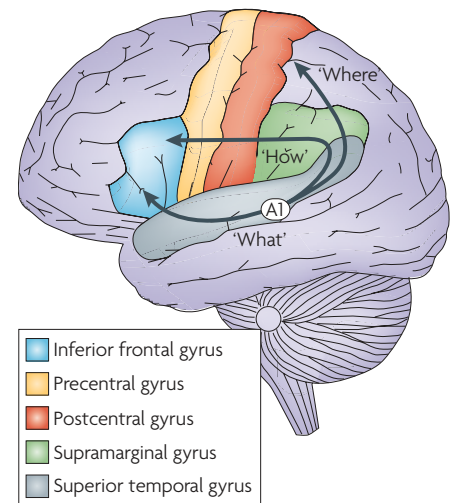


Figure 1 | The anatomy of sound perception. The posterior and anterior auditory streams. The main anatomical structures that are recruited for speech perception are shown in colour on a lateral view of the human brain. From the primary auditory cortex (A1), the arrows depict the putative directions of anterior and posterior streams of processing for speech and other sounds. The anterior stream decodes meaning in sounds ('what') and encompasses parts of the auditory association cortex in the superior temporal gyrus (dark green) and the inferior frontal gyrus (blue); the posterior pathway has been suggested to be engaged in sensorimotor integration and spatial processing ('how' and 'where') and includes parts of the posterior superior temporal gyrus, the inferior parietal cortex (supramarginal gyrus, green), motor and sensory areas (yellow and red, respectively) and the inferior frontal gyrus^{9,22,93}.

From the studies we have described so far there are two, largely non-overlapping, sets of data regarding the neural basis of speech perception. One set reveals activity mainly in the dorsolateral temporal lobes; the other shows an involvement of the motor cortex (along with dorsolateral temporal lobe activation) in speech perception. Why is there such a difference in emphasis and interpretation? Apart from the use (or not) of the motor theory of speech as an explanatory framework, this difference could come down to the use of control conditions in functional imaging experiments. Speech is a complex sound as well as a linguistic signal, and hence to isolate the neural response to the phonetic and linguistic information, functional imaging studies since the early 1990s have tended to use acoustic controls that possess similar spectrotemporal properties to speech, including spectrally rotated speech^{20,25,39,40,42}, 'musical rain' (REF. 41), reversed speech⁴⁷ and signal-correlated noise^{38,45}. However, reports of

motor activation in response to speech at a whole-brain level of analysis do not typically use contrasts against a complex acoustic control to reveal the motor activation. This makes it more difficult to determine whether the effects reflect phonetic or linguistic processing of the speech, or more general acoustic processing of complex sounds. This point is illustrated by a summary plot of motor responses to a range of sound categories (including speech, degraded speech, emotional vocalizations, tool sounds and music) (FIG. 2). TMS studies of speech perception^{15,18} have also failed to provide convincing evidence that motor activation in response to speech sounds is necessarily different from activation in response to the control sounds¹². For example, one study¹⁵ revealed no significant difference between the motor activation due to speech and that due to environmental, non-verbal sounds, such as car engines and breaking glass.

It is possible that other fMRI techniques, such as multivariate pattern analysis and functional connectivity, may reveal different patterns of activity in brain regions in response to speech and other sounds⁴³. However, to date, the evidence from functional imaging strongly implies that the motor cortex activity associated with speech perception,

although undoubtedly present, is not driven by the phonetic or linguistic content of the speech. This activity could instead be linked to aspects of auditory events that are common to a wide variety of sounds, including speech (BOX 1; FIG. 2). This does not weaken the findings of motor cortex activation in response to speech or query their support of the motor theory of speech perception, but it does call into question what the motor cortex responses to speech actually mean.

Candidate functions of motor responses

Linguistic functions. There are several linguistic functions that may be served by the motor cortex. Motor activity could reflect a role for the motor cortex (although not necessarily the same subregions) in one, some or all of the following. First, the primary and premotor cortices could be important in phonemic processing, as per the motor theory of speech perception^{6,16,17}. However, as discussed in the previous section, the variability of motor and premotor cortex responses to speech sounds, in contrast with the specificity of responses to speech (relative to acoustic baselines) in the temporal lobes, suggests that the motor cortex response to speech does not reflect a primary, perceptual route to the comprehension of speech (see also

REF. 12). This is supported by the difference in lesions that cause speech-perception and -production deficits, respectively, in clinical groups, and the different profiles of speech perception and production in development.

Second, the motor cortex could have a role in syntactic processing: it has been suggested that syntax is represented and processed in systems that regulate motor output, including speech production⁴⁸. In this account, syntax (as a rule-based system) is considered to have evolved from brain mechanisms for processing intention-based action sequences⁴⁹. Thus, syntax could be mediated by motor representations of actions in the perception of spoken language⁴⁸. Consistent with this possibility, several authors have linked activity in ventral premotor areas with aspects of local structure computation in syntactic processing^{50,51}. Activations in these more posterior premotor areas, which lie behind the classic Broca's area (BA 44 and BA 45), have been proposed to interact with higher-order syntactic processing in Broca's area. Notably, this processing may not be restricted to speech, as activation of the premotor cortex has also been reported in response to aspects of non-speech sequences⁵².

Third, several authors have suggested a role for motor cortex representations of action in semantic knowledge⁵³. There is ample evidence that aspects of semantic knowledge involve interactions with the motor system — both behaviourally and in terms of neuroanatomy. Behaviourally, motor tasks influence semantic processing and semantic processing influences motor task performance. For example, when participants are asked to pick up tools while performing a spatial task, they tend to pick the objects up by their handles, which they are less likely to do if they are simultaneously performing a distracting semantic task⁵⁴. This was interpreted as showing that the functional grasping of objects requires input from the semantic system. This interaction is consistent with aspects of embodied semantic representations — that is, semantic representations that are not completely abstract in form⁵³. These semantic-action links extend into cortical activity profiles: studies on the neural representation of action words have shown that such words somatotopically activate the motor cortex, so that hearing words like 'kick' activates the most dorsal part of the motor strip (controlling the leg), whereas words like 'pick' activate an area ventral to this (controlling the arm and hand) and words like 'lick' activate the most ventral part of the motor cortex (controlling the articulators)⁵⁵. These activations are fast and automatic⁵⁵. We argue that these effects are probably not

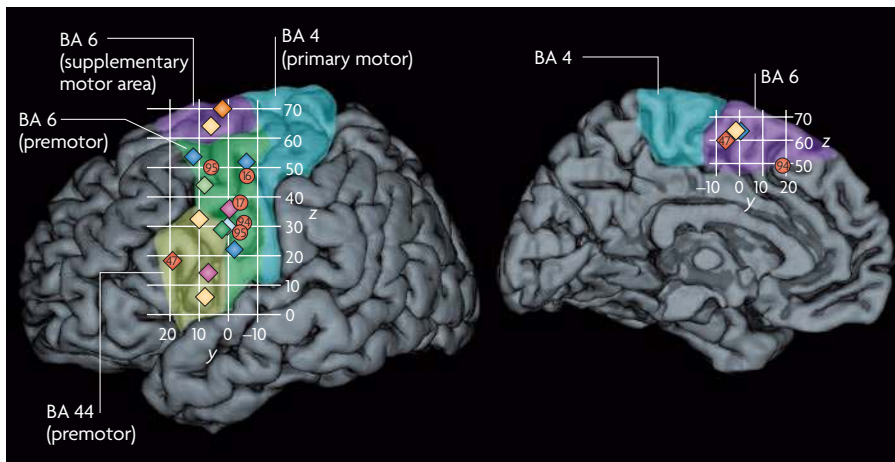


Figure 2 | Responses to sound in the motor cortex. A comparison of peak activations in the left motor cortex obtained from 12 functional imaging studies of auditory perception. The peaks were selected based on the following criteria: auditory presentation of stimuli, the absence of a motor task (that is, button press or covert rehearsal) on contrast-relevant trials, and a whole-brain-level analysis. Peak voxel locations are plotted, in MNI space, for responses to speech (red)^{16,17,47,94,95}, degraded speech (orange)⁹⁶, human emotional vocalizations (mid blue)⁸⁰, human song (light blue)⁹⁷, animal vocalizations (light green)⁹⁸, tool sounds (dark green)⁹⁹, action sounds (that is, sounds derived from human actions, such as kissing or ripping paper) (yellow)⁸⁸, and music (pink)¹⁰⁰. Diamond symbols indicate the activation peaks that were obtained from direct contrasts of the sound category of interest with a suitably complex auditory control. Circles indicate no direct comparison with an acoustic control. Together, these data suggest limited evidence in the literature to date that motor activity during passive listening to speech exceeds that shown in response to other complex sounds. The numbers inside these symbols indicate the relevant papers in the reference list. The SPM SurfRend Toolbox and FreeSurfer (CorTechs Labs, inc.)^{101,102} were used to generate the brain surface images. BA, Brodmann's area.

Box 2 | Sound and speech, action and emotion

Sounds activate the motor cortex in primates⁸⁷, which is evidence for an auditory mirror system, comprising neurons that are active in both the production and the perception of actions (however, see REF. 11). Functional MRI studies in humans have shown that noises that are made with either the hand or the mouth activate the premotor cortex somatotopically, with more dorsal responses to hand sounds and more ventral responses to mouth sounds⁸⁸. Thus, the neural response to sound seems to reflect the recruitment of motor processes that are physically linked to the effectors or actions that produce the sound. Furthermore, the processing of sound in the motor cortex is plastic and can change with training. Non-musicians who were trained to play a particular piece of music showed activation in the premotor cortex and action-observation regions in the parietal lobe when they passively listened to the piece, compared with a control condition in which the same notes were played in a different order⁸⁹.

Studies have found that non-verbal expressions of emotion — for example, laughs and screams — also activate the primary motor and premotor cortices^{80,90}. This effect is modulated by properties of valence and arousal⁹⁰, with more positive expressions recruiting the motor and premotor cortices more strongly than more negative expressions. These findings have been linked to the behavioural contagion of positive emotions like laughter⁹¹ — when we hear or see someone else smiling or laughing, it can be difficult to resist laughing along. This motor response to laughter (and also to cheering) is consistent with findings which show that, generally, when we like other people we tend to mirror their posture and even use the same words as them⁶⁴. This effect can be manipulated by motor behaviour: requiring people to perform coordinated acts together leads later to more cooperative behaviour between them than if they are asked to perform non-coordinated joint actions⁹².

restricted to the motor cortex: for example, across both visual and auditory presentations, the perception of words that are highly imageable (such as ‘glove’) activates the visual association cortex in a way that less imageable words (such as ‘love’) do not^{56,57}. These findings are elegant demonstrations of the recruitment of modality-specific cortex in the representation of semantic knowledge, and show that semantic information is embodied in the physical aspects of the stimulus.

Task-related functions of the motor cortex.

The three alternatives above reflect automatic processing of the auditory speech signal. However, it is also possible that the motor cortex is specifically involved in speech processing when the particular task or listening context demands it — that is, that motor knowledge is used to support other speech-processing systems⁴⁶. Thus, stronger responses are seen in both the left and the right premotor cortex when participants listen to degraded speech than to clear speech⁴⁵ or to non-native speech sounds¹⁷. Likewise, responses to seeing someone speak are enhanced in the motor cortex when the visual signal is degraded but not when non-speech signals (such as sticking the tongue out) are degraded³⁸. There is also evidence that there are greater motor responses to rare words than to frequently occurring words⁵⁹, which is consistent with a role for the motor system in supporting semantic processing. This ‘lexical-frequency effect’ occurs later than phonological influences on motor cortex activation, suggesting that the motor system interacts

with lexical systems in the temporal lobes to assist comprehension when we encounter an unfamiliar word⁵⁹. These task effects might explain why motor responses are inconsistently seen in functional imaging studies — there might be motor processes that are associated with perceptual difficulty or with specific things that the participants are asked to do when they hear the speech, rather than with the basic aspects of speech perception and comprehension^{10,46}.

Sound-to-action functions of the motor cortex.

A third possibility is that the motor cortex processes the actions that are reflected by sounds — for example, the difference between a plucked and a bowed note on a stringed instrument (BOX 1) — and that in turn are conveyed in relatively simple properties of the sounds (for example, the onsets). In this account, motor cortex responses to sound might depend on very basic ways in which a sound is produced by an action (BOX 1), and hence motor cortex responses to various sound categories are seen (BOX 2; FIG. 2). Importantly, these responses do not require the identification of that sound and could be mediated entirely through the posterior ‘how’ pathway. For example, tracking the rhythm of a repeating sound would involve the ‘how’ pathway, whereas distinguishing whether that sound is a drum or a footstep would require the ‘what’ pathway (BOX 1). Below, we suggest that these sensation–action links have evolved into something that is important in complex motor coordination, and that

this coordination may have a particular significance for human language use.

Motor cortex involvement in joint speech.

We argue that motor cortex responses are important in speech processing because they underlie sensorimotor processes that are essential in conversational speech. Speech perception and production are usually studied separately, using simple words or sentences, but the primary use of spoken language is conversation, which has been described as a true linguistic universal⁶⁰. Where people have language, they use it to talk to each other in conversations — arguably, we have evolved to speak in dialogue rather than monologue. People learn their first language first and foremost as a joint behaviour, and the turn-taking aspects of conversation are learnt before infants have many words with which to talk⁶¹. Thus, we argue that spoken language is relevant to the motor system because it is a behaviour with which people can expertly and smoothly coordinate their own actions⁶².

Convergence in conversation

There is ample evidence that we do indeed coordinate various aspects of our speech with that of fellow speakers. Classic studies of dialogue found the phenomenon of interactional synchrony⁶³: talkers unintentionally coordinating their actions and postures. Such mirroring of each other’s behaviour allows us to express our affiliation with those with whom we speak⁶⁴ and also helps the process of conversation itself: in conversation we align our conceptual and syntactic structures to those of our co-talkers⁶⁵. This linguistic alignment has been argued to be central to making conversation possible⁶⁶. Conversational utterances are often elliptical and incomplete — for example, people rarely speak in proper sentences that could be understood without the context of the conversation. We therefore need to continually adjust what we say to accommodate the specific knowledge of the other talkers⁶⁶. This convergence can be extended to motor aspects of speech control in conversation: when we hold a conversation we coordinate our breathing⁶⁷, and our pronunciations begin to converge with those of our co-talkers⁶⁸. The dynamic properties of the control of conversation have been linked to a perception–action pathway in the brain^{65,66}, which maps onto the ‘how’ pathway, linking the human auditory cortex and the motor system⁹ (FIG. 1). From this perspective, the tracking of specific articulations, which has been demonstrated in the motor cortex⁶⁹,

Glossary

Convergence

In this context, the way that different aspects of joint speech (both motoric and linguistic) become united, or coordinated, between speakers.

Diphone

A cluster of two phones that can be legally combined in a language (for example, /sk/ is legal at the start of a syllable in English, but /ks/ is not); diphones thus contain transitional information between the two phones, and are more information-rich than single phones.

Embodied semantic representations

In this context, theories of semantic representations that link the more abstract elements of the representations to more concrete elements of their material properties; for example, part of the meaning of 'a football' is represented by how one might kick it.

Expressive aphasia

A speech-production deficit in which people have reduced fluency, grammatical errors and problems in articulating accurately.

Linguistics

In this context, the phonemic, semantic or syntactic processing of heard speech, which is distinct from the processing of the basic acoustic properties of speech (for example, loudness).

Local structure computation

The sequential analysis of heard speech (for example, 'the sandwich was eaten'), as opposed to higher-order, hierarchical computations across longer timescales (for example, 'the sandwiches were eaten by the children at the party').

Phone

A single speech sound (which is always a variant of a phoneme); for example, the aspirated /p/ at the start of 'port' is a different phone from the /p/ of 'sport', but these are both examples (allophones) of the phoneme /p/.

Phoneme

An elemental sound of speech (such as /p/ or /t/) that can be used in the explicit transcription and classification of the sounds of a language.

Phonemic

Pertaining to the representation and processing of phonemes.

Phonetic

Pertaining to speech sounds (phones).

Pre-lexical processing

In this context, the neural processing of speech sounds before the representation of word identity and meaning.

Receptive aphasia

A speech-perception and -comprehension deficit in which the patient has great difficulty in following what is being said to them. Speech production is unimpaired in terms of fluency but speech content can be meaningless, and many patients are unaware that they have a problem.

Semantic

Relating to the meaning of things, in this case words and language.

Spectral centre of gravity

The average value of the spectral components of a sound, which captures how the sound is weighted across low to high frequencies; for example, 's' has a higher spectral centre of gravity than 'sh'.

Speech comprehension

In this context, post-perceptual, lexical, semantic and linguistic processing of speech. Although speech comprehension does require good speech perception, comprehension can also be enhanced by higher-order syntactic and semantic features (for example, sentence predictability).

Speech perception

In this context, the pre-lexical perceptual processing of the speech signal.

Syllable

Like a diphone, a syllable typically contains information about the organization of speech at a level higher than the phoneme. A single-syllable word, like 'start', can be broken down into an onset and a rhyme (for example, st-art), and may consist of only the rhyme (for example, 'art'): the rhyme may be further broken down into a nucleus and coda (for example, ar-t).

Syntax

The rules that determine the correct arrangement and inflection of words in spoken or written language.

Voicing

The sound made by vibrations of the vocal folds; for example, the sound at the start of 'zoo' is voiced, whereas that at the start of 'sue' is unvoiced.

could be facilitating the convergence on pronunciations of the words, rather than performing phonetic analysis.

Coordinating conversation — smooth turn taking. Conversation is found in all human cultures; furthermore, the tight temporal coordination of the act of conversation is also universal. Conversation is characterized by several properties⁷⁰: a speaker change occurs; one person speaks at a time; simultaneous speech is common but brief; transitions with no gap and no overlap are common; the order, size and duration of turns are not fixed; the number of talkers can vary; and talk can be continuous or discontinuous. Of these properties, the most striking is that turn taking in conversation typically occurs without pauses or overlapping speech. As has been pointed out previously⁶², conversation represents a "...considerable achievement. Conversations unfold in real time, and yet parties to the conversation synchronize their turns — usually highly coherent and consistent with respect to the topic — in a matter of milliseconds." A recent study showed that 45% of turn transitions in over 1,500 examples from a corpus of telephone calls occurred within a window of -250 to $+250$ ms, and 85% fell within a window of

-750 to $+750$ ms⁷¹. This represents an astonishing level of coordination, especially when one considers that it can take place without any visual input (for example, on the telephone) and between complete strangers⁷². Smooth and well-coordinated turn taking is a hallmark of successful conversations, in all cultures and across modalities (for example, in signing)⁷³. Disrupting the timing of turn taking seriously disrupts the fluency and ease of an interaction, as anyone who has tried to have a telephone conversation through a bad satellite link with lots of echoes can attest⁷⁴.

Along with Iacoboni⁷⁵, we argue that the tight temporal coordination of turn taking relies heavily on the motor system. We propose that, during conversation, the temporal lobes and associated regions that are important in acoustic and linguistic representation and processing track the meaning of what is being said, and the motor system concurrently tracks the speech rate and rhythm of the current talker so that picking up the next turn will be a seamless process (FIG. 3). In this view, the motor system is not only crucial to organizing the act of speaking, it is also

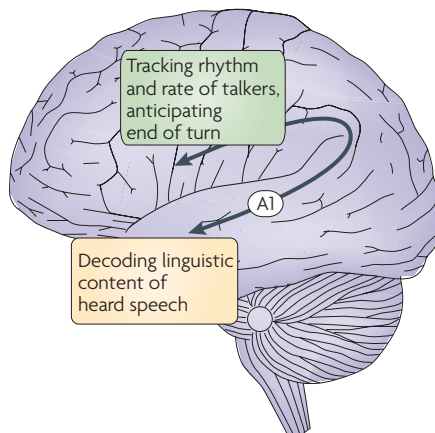


Figure 3 | Candidate roles for auditory streams of processing during conversation.

The arrows originating from the primary auditory cortex (A1) illustrate a functional division in how sounds are processed (following from the streams of processing shown in FIG. 1). The meaning of sound, especially linguistic meaning in speech, is decoded primarily in the superior temporal lobe and anterior to the primary auditory cortex. We suggest that aspects of coordinating conversation, in particular turn taking, are mediated by the posterior 'how' pathway. This perception-action pathway subserves monitoring of the speech signal for rhythm and rate, which enables the listener to anticipate the end of a speaker's turn, and thereby facilitates smooth turn taking.

essential in facilitating the conversational dance of turn taking and coordinating the other factors, such as interactional synchrony and convergence, that make conversation possible^{65,66}. The motor system may also facilitate an alignment of interactions even when the meaning and content of what is said are highly constrained — for example, complete strangers can read a novel piece of text in synchrony with one another, apparently by coordinating the rhythm and melody of their speech⁷⁶ (BOX 1). Here, the acoustic properties of speech facilitate the coordination of actions, enabling the accurate entrainment of speech timing by the two talkers⁷⁷.

A specific model of turn taking has used the concept of entrained oscillations as a way of accounting for the smooth patterns of turn taking in conversation⁷³. In this model, the listener entrains to the speech rate of another talker at the level of the syllable, and uses this entrainment to accurately time their speech output as that talker comes to the end of a turn. We suggest that the computation of these rate and rhythm factors depends primarily on the motor system; this is consistent with interpretations of motor activation in perception tasks as being essential in anticipatory responses⁷⁸.

Motor activation and cohesive behaviour — some predictions. There is little evidence in the literature that directly bears on our hypothesis. Going back to the clinical findings, it is clear that holding a conversation can be a real problem for people with aphasic syndromes, and that this can lead to considerable social isolation. Although patients with expressive aphasia are good at conveying their needs, their conversational skills are impaired. By contrast, there is some evidence that in patients with receptive aphasia many of the basic rules of conversation are intact⁷⁹. However, these are highly indirect sources of evidence, and our hypothesis needs to be directly tested. According to our hypothesis, greater motor activation would occur in response to heard speech when the participant believes that they are in a conversation than when they are hearing a recorded monologue — that is, there would be greater motor system involvement when the subject believes that someone is speaking to them and that they will at some point reply⁷⁰.

There are currently no neuroscientific studies on motor activations during conversation. Techniques with good temporal resolution, such as TMS, electroencephalography and magnetoencephalography, would be well placed to capture the dynamic aspects of conversational turn taking. Studies using

fMRI would also be informative once the procedural challenges of using fMRI to scan someone having realistic conversation had been addressed.

We would also like to suggest that roles for the motor system in the orchestration of conversation are almost certainly not limited to speech. In fMRI studies, peaks of activation occur in response to heard speech relative to silence^{16,17}, but strong motor responses also occur to non-verbal (although vocal) expressions of emotion (relative to an acoustic baseline), especially expressions that typically occur in social contexts, such as laughter⁸⁰ (BOX 2; FIG. 2). Indeed, this link between the sensorimotor processing of action information in sound and speech suggests that such tightly coordinated motor acts would not be limited to vocal expressions, and we suggest that this role could also be extended to synchronous activity and timing in playing music and song in polyphonic settings. The acoustic factors that have been proposed to underlie synchrony and rhythm (for example, the way the sound starts) have been argued to be similar in speech and music⁸¹ (BOX 1).

Conclusions

Acoustic speech signals can activate areas of the motor cortex. In contrast to the predictions of the motor theory of speech perception, this does not seem to reflect phonetic processing of the speech signal. We argue here that the motor cortex is not essential for perceiving spoken language; if it was, it should be more common to see motor cortex activation to speech in fMRI studies, it should be easier to distinguish unambiguous speech-perception problems following anterior brain lesions, and there should be a clearer relationship between production and perception in development. Instead the motor activation can be linked to speech in several ways: as an embodiment of aspects of semantic information, as a feature that is important in processing both linguistic and non-linguistic syntactic information, as a meta-linguistic process that can be invoked depending on task difficulty or specific task requirements, and as a response to the motor information that is expressed in speech and other sounds. In addition, we argue that a basic role for motor responses to heard speech is a glimpse of a much more important role for motor systems in joint speech, or conversation, in which the motor cortex is essential for phenomena such as convergence^{65–68}, interactional synchrony^{63,64} and ensuring smooth turn transition^{62,73,75}. We argue that speech is processed and represented differently in the temporal lobes and the motor cortex, and that the motor system

is involved in a (non-speech-specific) action pathway. According to our hypothesis this sound-to-action pathway is highly refined in human language to allow us to talk smoothly in turn with one another, even if we are strangers and if we cannot see one another. Thus, we propose that there is a central role for motor representations and processes in conversation^{65,66}, an essential aspect of human language in its most basic use.

Sophie K. Scott, Carolyn McGettigan and Frank Eisner are at the Institute for Cognitive Neuroscience, University College London, 17 Queen Square, London, WC1N 3AR, UK.

Correspondence to S.K.S.

e-mail: sophie.scott@ucl.ac.uk

doi:10.1038/nrn2603

Published online 11 March 2009

- Kluender, K. R. & Alexander, J. M. in *The Senses, a Comprehensive Reference* vol. 3 (eds Basbaum, A. I. et al.) 829–860 (Academic, San Diego, 2008).
- Lieberman, A. M., Delattre, P. & Cooper, F. S. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Am. J. Psychol.* **65**, 497–516 (1952).
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. Perception of the speech code. *Psychol. Rev.* **74**, 431–461 (1967).
- Lieberman, A. M. & Mattingly, I. G. The motor theory of speech-perception revised. *Cognition* **21**, 1–36 (1985).
- Fowler, C. A. An event approach to the study of speech-perception from a direct realist perspective. *J. Phon.* **14**, 3–28 (1986).
- Galantucci, B., Fowler, C. A. & Turvey, M. T. The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* **13**, 361–377 (2005).
- Diehl, R. L. & Kluender, K. R. On the objects of speech perception. *Ecol. Psychol.* **1**, 121–144 (1989).
- Lisker, L. Rapid vs rabad: a catalogue of acoustical features that may cue the distinction. *Haskins Laboratories Status Report on Speech Research* **54**, 127–132 (1978).
- Scott, S. K. & Johnsrude, I. S. The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* **26**, 100–107 (2003).
- Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nature Rev. Neurosci.* **8**, 393–402 (2007).
- Hickok, G. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J. Cogn. Neurosci.* 13 Jan 2009 (doi:10.1162/jocn.2009.21189).
- Lotto, A. J., Hickok, G. & Holt, L. L. Reflections on mirror neurons and speech perception. *Trends Cogn. Sci.* 17 Feb 2009 (doi:10.1016/j.tics.2008.11.008).
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D. & Iacoboni, M. The essential role of premotor cortex in speech perception. *Curr. Biol.* **17**, 1692–1696 (2007).
- Wise, R. J. S., Greene, J., Büchel, C. & Scott, S. K. Brain systems for word perception and articulation. *Lancet* **353**, 1057–1061 (1999).
- Watkins, K. E., Strafella, A. P. & Paus, T. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* **41**, 989–994 (2003).
- Wilson, S. M., Saygin, A. P., Sereno, M. I. & Iacoboni, M. Listening to speech activates motor areas involved in speech production. *Nature Neurosci.* **7**, 701–702 (2004).
- Wilson, S. M. & Iacoboni, M. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage* **33**, 316–325 (2006).
- Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* **15**, 399–402 (2002).
- Tardif, E., Spierer, L., Clarke, S. & Murray, M. M. Interactions between auditory 'what' and 'where' pathways revealed by enhanced near-threshold discrimination of frequency and position. *Neuropsychologia* **46**, 958–966 (2008).

20. Scott, S. K., Blank, C. C., Rosen, S. & Wise, R. J. S. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* **123**, 2400–2406 (2000).
21. Wise, R. J. S., Scott, S. K., Blank, C. C., Mummery, C. J. & Warburton, E. Identifying separate neural sub-systems within 'Wernicke's area'. *Brain* **124**, 83–95 (2001).
22. Romanski, L. M. *et al.* Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neurosci.* **2**, 1131–1136 (1999).
23. Repp, B. H. Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 600–621 (2001).
24. Tourville, J. A., Reilly, K. J. & Guenther, F. H. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* **39**, 1429–1443 (2008).
25. Scott, S. K., Rosen, S., Lang, H. & Wise, R. J. S. Neural correlates of intelligibility in speech investigated with noise-vocoded speech – a positron emission tomography study. *J. Acoust. Soc. Am.* **120**, 1075–1083 (2006).
26. Mohr, J. P. *et al.* Broca aphasia – pathologic and clinical. *Neurology* **28**, 311–324 (1978).
27. Blank, S. C., Bird, H., Turkheimer, F. & Wise, R. J. Speech production after stroke: the role of the right pars opercularis. *Ann. Neurol.* **54**, 310–320 (2003).
28. Crinion, J. T. *et al.* Listening to narrative speech after aphasic stroke: the role of the left anterior temporal lobe. *Cereb. Cortex* **16**, 1116–1125 (2006).
29. Bogen, J. E. & Bogen, G. M. Wernicke's region – where is it? *Ann. NY Acad. Sci.* **280**, 834–843 (1976).
30. Basso, A., Casati, G. & Vignolo, L. A. Phonemic identification defect in aphasia. *Cortex* **13**, 85–95 (1977).
31. Mogford, K. in *Language Development in Exceptional Circumstances* (eds Bishop, D. V. M. & Mogford, K.) 110–131 (Churchill Livingstone, New York, 1988).
32. Bishop, D. V. M. in *Language Development in Exceptional Circumstances* (eds Bishop, D. V. M. & Mogford, K.) 220–238 (Churchill Livingstone, New York, 1988).
33. Werker, J. F. & Yeung, H. H. Infant speech perception bootstraps word learning. *Trends Cogn. Sci.* **9**, 519–527 (2005).
34. Tsao, F.-M., Liu, H. M. & Kuhl, P. K. Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child. Dev.* **75**, 1067–1084 (2004).
35. Bates, E. & Dick, F. Language, gesture, and the developing brain. *Dev. Psychobiol.* **40**, 293–310 (2002).
36. Alcock, K. J. & Krawczyk, K. *Motor skills and the vocabulary burst*. (International Conference for the Study of Child Language, Berlin, 2005).
37. Wise, R. *et al.* Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain* **114**, 1803–1817 (1991).
38. Mummery, C. J., Ashburner, J., Scott, S. K. & Wise, R. J. S. Functional neuroimaging of speech perception in six normal and two aphasic subjects. *J. Acoust. Soc. Am.* **106**, 449–457 (1999).
39. Narain, C. *et al.* Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb. Cortex* **13**, 1362–1368 (2003).
40. Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T. & Medler, D. A. Neural substrates of phonemic perception. *Cereb. Cortex* **15**, 1621–1631 (2005).
41. Uppenkamp, S., Johnsrude, I. S., Marslen-Wilson, W. & Patterson, R. D. Locating the initial stages of speech-sound processing in human temporal cortex. *Neuroimage* **31**, 1284–1296 (2006).
42. Obleser, J., Scott, S. K. & Eulitz, C. Now you hear it, now you don't: transient traces of consonants and their nonspeech analogues in the human brain. *Cereb. Cortex* **16**, 1069–1076 (2006).
43. Obleser, J. & Eisner, F. Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn. Sci.* **13**, 14–19 (2009).
44. Patterson, K., Nestor, P. J. & Rogers, T. T. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Rev. Neurosci.* **8**, 976–987 (2007).
45. Davis, M. H. & Johnsrude, I. S. Hierarchical processing in spoken language comprehension. *J. Neurosci.* **23**, 3423–3431 (2003).
46. Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G. & Rogers, J. C. Motor regions contribute to speech perception: awareness, adaptation and categorization. *J. Acoust. Soc. Am.* **123**, 3580 (2008).
47. Jardri, R. *et al.* Self awareness and speech processing: an fMRI study. *Neuroimage* **35**, 1645–1653 (2007).
48. Fogassi, L. & Ferrari, P. F. Mirror neurons and the evolution of embodied language. *Curr. Dir. Psychol. Sci.* **16**, 136–141 (2007).
49. Greenfield, P. M. Language, tools and brain: the ontology and phylogeny of hierarchically organized sequential behaviour. *Behav. Brain Sci.* **14**, 531–595 (1991).
50. Friederici, A. D. Broca's area and the ventral premotor cortex in language: functional differentiation and specificity. *Cortex* **42**, 472–475 (2006).
51. Fiebach, C. J. & Schubotz, R. I. Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains? *Cortex* **42**, 499–502 (2006).
52. Schubotz, R. I. & von Cramon, D. Y. Functional-anatomical concepts of human premotor cortex: evidence from fMRI and PET studies. *Neuroimage* **20**, S120–S131 (2003).
53. Fischer, M. H. & Zwaan, R. A. Embodied language: a review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol.* **61**, 825–850 (2008).
54. Creem, S. H. & Proffitt, D. R. Grasping objects by their handles: a necessary interaction between cognition and action. *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 218–228 (2001).
55. Pulvermüller, F. Brain mechanisms linking language and action. *Nature Rev. Neurosci.* **6**, 576–582 (2005).
56. Wise, R. J. *et al.* Noun imageability and the temporal lobes. *Neuropsychologia* **38**, 985–994 (2000).
57. Fiebach, C. J. & Friederici, A. D. Processing concrete words: fMRI evidence against a specific right-hemisphere involvement. *Neuropsychologia* **42**, 62–70 (2004).
58. Fridriksson, J. *et al.* Motor speech perception modulates the cortical language areas. *Neuroimage*, **41**, 605–613 (2008).
59. Roy, A. C., Craighero, L., Fabbri-Destro, M. & Fadiga, L. Phonological and lexical motor facilitation during speech listening: a transcranial magnetic stimulation study. *J. Physiol. Paris* **102**, 101–105 (2009).
60. Levelt, W. J. M. *Speaking: from Intention to Articulation* (MIT press, Cambridge, Massachusetts, 1989).
61. Beebe, B., Alson, D., Jaffe, J., Feldstein, S. & Crown, C. Vocal congruence in mother-infant play. *J. Psycholinguist. Res.* **17**, 245–259 (1988).
62. Beattie, G. *Talk: an Analysis of Speech and Non-Verbal Behaviour in Conversation* (Open Univ. Press, Milton Keynes, 1983).
63. Condon, W. S. & Ogston, W. D. A segmentation of behaviour. *J. Psychiatr. Res.* **5**, 221–235 (1967).
64. Chartrand, T. L. & Bargh, J. A. The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* **76**, 893–910 (1999).
65. Garrod, S. & Pickering, M. J. Why is conversation so easy? *Trends Cogn. Sci.* **8**, 8–11 (2004).
66. Pickering, M. J. & Garrod, S. Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* **11**, 105–110 (2007).
67. McFarland, D. H. Respiratory markers of conversational interaction. *J. Speech Lang. Hear. Res.* **44**, 128–143 (2001).
68. Pardo, J. S. On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am.* **119**, 2382–2393 (2006).
69. Pulvermüller, F. *et al.* Motor cortex maps articulatory features of speech sounds. *Proc. Natl Acad. Sci. USA* **103**, 7865–7870 (2006).
70. Sacks, H., Schegloff, E. A. & Jefferson, G. A. A simplest systematics for the organization of turn-taking in conversation. *Language* **50**, 697–735 (1974).
71. De Ruiter, J. P., Mitterer, H. & Enfield, N. J. Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* **82**, 515–535 (2006).
72. Beattie, G. W. & Barnard, P. J. The temporal structure of natural telephone conversations (Directory Inquiry calls). *Linguistics* **17**, 213–229 (1979).
73. Wilson, M. & Wilson, T. P. An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* **12**, 957–968 (2005).
74. Nobuhiko, K. & Kenzo, I. Pure delay effects on speech quality in telecommunications. *IEEE J. Sel. Areas Commun.* **9**, 586–593 (1991).
75. Iacoboni, M. in *Perspectives on Imitation: from Neuroscience to Social Science* (eds Hurley, S. & Chater, N.) 77–100 (MIT Press, 2005).
76. Cummins, F. Practice and performance in speech produced synchronously. *J. Phon.* **31**, 139–148 (2003).
77. Cummins, F. Rhythm as entrainment: the case of synchronous speech. *J. Phon.* **6** Oct 2008 (doi:10.1016/j.wocn.2008.08.003).
78. Prinz, W. What re-enactment earns us. *Cortex* **42**, 515–517 (2006).
79. Schienberg, S. & Holland, A. L. in *Clinical Aphasiology: Conference Proceedings* (ed. Brookshire, R. H.) 106–110 (BRK Publishers, Minneapolis, 1980).
80. Warren, J. E. *et al.* Positive emotions preferentially engage an auditory-motor "mirror" system. *J. Neurosci.* **26**, 13067–13075 (2006).
81. Scott, S. K. The point of P-centres. *Psychol. Res.* **61**, 4–11 (1998).
82. Warren, W. H. Jr. & Verbrugge, R. R. Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 704–712 (1984).
83. Hove, M. J., Keller, P. E. & Krumhansl, C. L. Sensorimotor synchronization with chords containing tone-onset asynchronies. *Percept. Psychophys.* **69**, 699–708 (2007).
84. Gordon, J. W. The perceptual attack time of musical tones. *J. Acoust. Soc. Am.* **82**, 88–105 (1987).
85. Rasch, R. Synchronization in performed ensemble music. *Acustica* **43**, 121–131 (1979).
86. Marcus, S. M. Acoustic determinants of perceptual centre (P-center) location. *Percept. Psychophys.* **30**, 247–256 (1981).
87. Kohler, E. *et al.* Hearing sounds, understanding actions: action representation in mirror neurons. *Science* **297**, 846–848 (2002).
88. Gazzola, V., Aziz-Zadeh, L. & Keysers, C. Empathy and the somatotopic auditory mirror system in humans. *Curr. Biol.* **16**, 1824–1829 (2006).
89. Lahav, A., Saltzman, E. & Schlaug, G. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.* **27**, 308–314 (2007).
90. Meyer, M., Zysset, S., von Cramon, D. Y. & Alter, K. Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. *Brain Res. Cogn. Brain Res.* **24**, 291–306 (2005).
91. Provine, R. R. Contagious laughter – laughter is a sufficient stimulus for laughs and smiles. *Bull. Psychon. Soc.* **30**, 1–4 (1992).
92. Wiltermuth, S. S. & Heath, C. Synchrony and cooperation. *Psychol. Sci.* **20**, 1–5 (2009).
93. Rauschecker, J. P. & Tian, B. Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl Acad. Sci. USA* **97**, 11800–11806 (2000).
94. Binder, J. R., Swanson, S. J., Hammeke, T. A. & Sabsevitz, D. S. A comparison of five fMRI protocols for mapping speech comprehension systems. *Epilepsia* **49**, 1980–1997 (2008).
95. Wilson, S. M., Molnar-Szakacs, I. & Iacoboni, M. Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cereb. Cortex* **18**, 230–242 (2008).
96. Scott, S. K., Rosen, S., Wickham, L. & Wise, R. J. S. A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *J. Acoust. Soc. Am.* **115**, 815–821 (2004).
97. Callan, D. E. *et al.* Song and speech: brain regions involved with perception and covert production. *Neuroimage* **31**, 1327–1342 (2006).
98. Doehrmann, O., Naumer, M. J., Volz, S., Kaiser, J. & Altmann, C. F. Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia* **46**, 2776–2786 (2008).
99. Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J. & DeYoe, E. A. Distinct cortical pathways for processing tool versus animal sounds. *J. Neurosci.* **25**, 5148–5158 (2005).
100. Bangert, M. *et al.* Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage* **30**, 917–926 (2006).
101. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I: Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).
102. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195–207 (1999).

Acknowledgements

S.K.S., C.M. and F.E. are funded by Wellcome Trust Grant WT074414MA. We would like to thank K. Kluender, H. Mitterer, L. Bernstein, T. Manly and M. Davis for very helpful discussions on many of these issues.

FURTHER INFORMATION

Sophie Scott's homepage: <http://www.icn.ucl.ac.uk/Research-Groups/Speech-Communication-Group/index.php>
SPM SurfRender Toolbox: <http://spm.surfrender.sourceforge.net>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF