Opinion

Pre-lexical abstraction of speech in the auditory cortex

Jonas Obleser[1] **and Frank Eisner**[2]

[1] Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany
[2] Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK

Speech perception requires the decoding of complex acoustic patterns. According to most cognitive models of spoken word recognition, this complexity is dealt with before lexical access via a process of abstraction from the acoustic signal to pre-lexical categories. It is currently unclear how these categories are implemented in the auditory cortex. Recent advances in animal neurophysiology and human functional imaging have made it possible to investigate the processing of speech in terms of probabilistic cortical maps rather than simple cognitive subtraction, which will enable us to relate neurometric data more directly to behavioural studies. We suggest that integration of insights from cognitive science, neurophysiology and functional imaging is necessary for furthering our understanding of pre-lexical abstraction in the cortex.

Introduction

Spoken language comprehension fundamentally involves extracting words from a continuous stream of speech. Although this process is entirely effortless most of the time, listeners are faced with considerable variation in the speech signal: the physical properties of the waveform corresponding to a given speech segment can be radically different, depending on a variety of factors that include talker characteristics, speech rate or ambient noise. Conversely, a given sound can be interpreted by listeners as a different speech segment, depending on phonological and acoustic context [1,2].

To deal with this variability, the auditory cortex derives increasingly abstract representations from the speech signal. Most cognitive models of spoken word recognition implement abstract pre-lexical units of representation such as phonemes. These units have the function of integrating and categorising acoustic-phonetic information along multiple dimensions. They also form the constituent parts of how words are stored in the mental lexicon, the repository of words that the listener knows.

Phonemic categories are established in the first months of life [3] and become less flexible with age, but even the adult perceptual system allows for talker- and context-specific adjustment [4]. An adjustment that is made to one particular phonemic category, for example in response to the accent of a talker, can generalize to any word in the mental lexicon that contains this phoneme, and therefore does not need to be made again for future encounters of that talker or speakers with a similar accent. Acoustic variation can thereby be normalised at an early stage of processing.

Pre-lexical abstraction is currently only loosely defined in terms of its neural implementation. Neurophysiological work in animals has revealed fundamentals of how acoustic features are processed in the auditory cortex, including complex signals such as speech sounds. In humans, functional neuroimaging research has identified a set of candidate regions in the superior temporal lobe which have a crucial role in the different stages of processing of the speech signal [5], but few of these have targeted pre-lexical abstraction directly. Recent developments in functional imaging techniques enable us to move beyond broad comparisons of activation strength ('cognitive subtraction', in which two experimental conditions are contrasted with each other to single out a specific processing step) and instead allow observation of emerging abstract categories in neuroimaging data in a fashion that requires fewer prior assumptions about the choice of baseline condition ('multivariate pattern analysis') [6–8].

Understanding the process of pre-lexical abstraction in the auditory cortex depends on the integration of findings from cognitive psychology, neurophysiology and functional neuroimaging. Here, we examine the current evidence and suggest avenues for future research in which neuroscientific investigations into human speech perception are informed by cognitive models.

The view from cognitive psychology and speech science

Recent studies have revealed specific mechanisms that contribute to achieving perceptual constancy in the light of variability in the speech signal. For example, listeners are able to adjust to a talker idiosyncrasy (e.g. as caused by identity, emotional state or regional accent) by re-tuning pre-lexical category representations in a talker-specific manner [4,9,10]. Pre-lexical phonemic categories are a highly effective locus for this type of adjustment: adjustments at this level can generalize to all future instances of the phonemic categories in the speech of this talker, no matter in which word they occur [11,12]. Hence, only a relatively small set of representations need to be retuned because every language is made up of a very limited set of phonemic categories, but contains many thousands of words.

This advantage of pre-lexical representations is reflected in the architecture of most cognitive models of spoken word recognition (see Ref. [13] for a review). These

---

Corresponding authors: Obleser, J. (obleser@cbs.mpg.de); Eisner, F. (f.eisner@ucl.ac.uk).

1364-6613/$ – see front matter © 2008 Elsevier Ltd. All rights reserved. doi:10.1016/j.tics.2008.09.005 Available online 11 December 2008

models are organized hierarchically into consecutive levels of processing that deal with increasingly abstract units of representation. The TRACE model [14], for instance, consists of three layers – corresponding to acoustic-phonetic features, phonemes and words – and the primary flow of information is from the feature level to the lexical level, in which words are represented abstractly.

However, the focus of these models is on word recognition, and not on pre-lexical abstraction. Although the purpose of pre-lexical categories has been to mediate between the speech signal and the mental lexicon, the input to most models is not a psychologically realistic representation of the speech signal, but instead consists of unambiguous, already abstracted, strings of phonemes (other units of pre-lexical representation have been proposed, including features, diphones, triphones or syllables [15]).

Neither human speech science nor automatic speech recognition approaches [16] have been able to identify invariant acoustic features that correspond to the perception of speech sounds reliably across different contexts and talkers. It seems that pre-lexical speech processing depends on both an acoustic-phonetic analysis and other factors such as talker, speech rate or phonological context. These sources of information are combined into an output that is probabilistic rather than discrete (Figure 1a).

A probabilistic view of pre-lexical abstraction has recently been implemented by Norris and McQueen [17] in the Shortlist B model. In contrast to models in which word recognition is achieved on the basis of unambiguous strings of discrete phonemes, the input to Shortlist B takes the form of phoneme likelihoods which vary over time. These phoneme likelihoods were derived from a behavioural gating experiment, and thus reflect actual listening ambiguities that arise for phonemes that are similar on a perceptual dimension at a given point in time. Phoneme likelihoods are noisier than discrete phoneme strings, but, in addition to being a more realistic representation of what listeners are faced with, they also contain more information about the acoustic context. Although not currently implemented in Shortlist B, a probabilistic type of input to a computational model has the potential of taking into account adjustments to talker idiosyncrasies by altering phoneme likelihoods in a talker-and context-specific manner (Figure 1a).

Insights on the abstraction of speech sounds from animal physiology

The human auditory system shares basic mechanisms of spectro-temporal encoding and cortical map formation with other mammals. In the processing of simple to increasingly complex sounds, a hierarchical division of auditory cortex into 'core', 'belt' and 'parabelt' areas has been established in non-human primates [18,19]. Although there is a debate about the number and exact function of these subfields in different species (e.g. see Ref. [5]), especially 'belt' and 'parabelt' regions are thought to support the

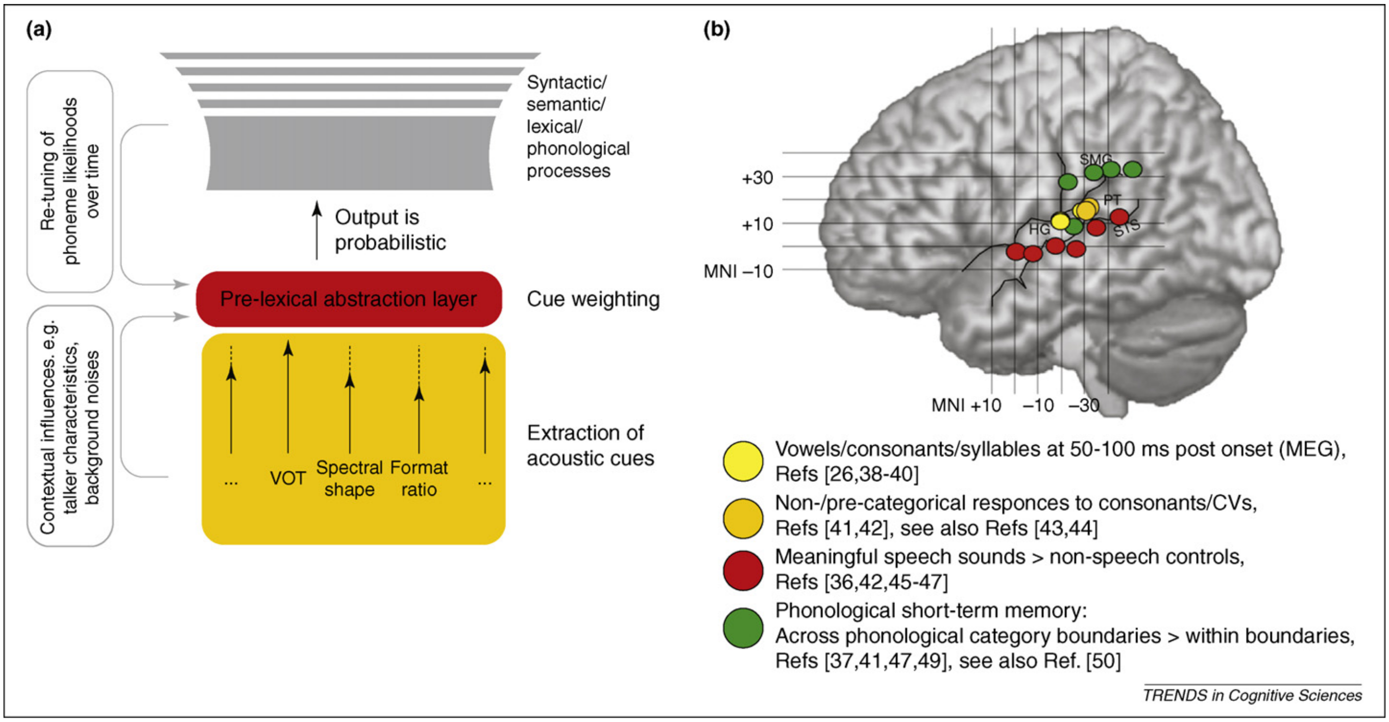

Figure 1. Functional **(a)** and macroanatomical **(b)** architecture of the pre-lexical abstraction process. (a) Acoustic cues such as burst intensity, voice onset time and spectral centre of gravity are extracted in primary auditory cortex [28,31,32] and then integrated into language-specific abstract units which can be used for word recognition. We suggest that the output of the abstraction process is a probabilistic result of a cue weighting process. Left upward arrow in (a): contextual influences such as talker characteristics or noise in the listening situation require re-adjustment of the relative influence of acoustic-phonetic cues on the probabilistic output of abstraction. Left downward arrow in (a): perceptual learning mechanisms can utilize higher-level (e.g. lexical) information for re-tuning phoneme likelihoods over time [4,17,59]. Phonological memory is thought to operate on units that result from this pre-lexical abstraction process. (b) A set of brain activations from MEG and fMRI studies which have directly addressed pre-lexical processing of speech are mapped onto a sagittal view of a standardized brain (Montreal Neurological Institute [MNI] coordinate system). The colour coding represents a distinction of early components in the MEG signal (P50m, N100m; yellow), non- or pre-categorical responses to meaningful speech sounds and spectro-temporally matched complex sounds (orange), meaningful speech sounds evoking a larger response compared with non-speech sounds (red) and detection of a phonological change, but not merely an acoustic change (green).

analysis of conspecific vocalisations and other auditory signals which are behaviourally relevant [5,19–21].

Within this hierarchy, spectro-temporal properties of the acoustic signal are processed selectively. Orthogonal to a gradient of preferred frequency, subfields in auditory cortex exhibit various gradients of selective responsiveness to spectral bandwidths, periodicities and spectrotemporal motion parameters (e.g. Refs [20,22–24]). This layout enables the formation of orderly maps from spectrally complex sounds. Acoustic dimensions that are relevant to human speech sounds (e.g. formants, characteristic peaks in the frequency spectrum of a speech sound) have been shown to be topographically organized in the auditory cortex [25,26].

But how do abstract categories of sounds emerge? Currently, the best evidence on mechanisms of auditory map formation comes from species that are rather remote from human anatomy. In gerbils, Ohl *et al.* [27] demonstrated the emergence of distinct patterns of auditory cortex activity over the time course of a discrimination learning experiment. The animals were trained on multiple tokens taken from two broad categories of inherently meaningless sounds (frequency-modulated sweeps of either rising or falling pitch). This was interpreted to reflect acquisition of an abstract category. The authors detected category formation using a classification algorithm in a higher-dimensional search space (i.e. a space of more than just two or three dimensions, being defined by the numbers of signal detectors [18 in this study]). Interestingly, the newly formed categories were not observed as differences in activation strength but as differing higher-dimensional maps of activation patterns (for comparable multidimensional approaches to human neuroimaging, see Box 1).

More recently, Engineer and colleagues [28] trained rats to discriminate a large set of speech sounds and obtained neural confusion matrices (i.e. pairwise measures of neural pattern similarity) for a comprehensive set of English consonants. The discrimination performances of the rats correlated with spike timing patterns elicited by these sounds in primary auditory cortex. Although this study did not demonstrate abstraction of speech sounds, it did show that distinguishable patterns of cortical activation emerge from acoustic dimensions that are relevant for speech.

Studying the formation of phonemic categories with human speech in animals is largely limited to bottom-up processing. Although the categories can gain behavioural relevance in the non-human brain (e.g. by means of conditioning), interfacing with the lexical level is impossible to study in animals. Schreiner and Winer [29], in a review on auditory cortical map formation, note that the 'existence of topographic organization in auditory cortex is indisputable, yet their functional implications for neural processing in the generation of perception and behaviour remain obscure'. The key finding from the neurophysiology literature for pre-lexical processing is that acoustic-phonetic features of the speech signal such as voicing, spectral shape, formants or amplitude modulation [30–32] are

Box 1. Abstraction: a case for multivariate pattern recognition methods

If a given brain region processes two types of sounds in an equally vigorous manner, then cognitive subtraction logic (e.g. voiced > voiceless sounds, consonants > vowels) will fail – despite the fact that crucial steps of analysis, pattern formation and abstraction could take place within this region. Because the subtraction of BOLD activation potentially misses crucial steps of acoustic-phonetic processing, research into speech sound recognition and abstraction is a prime candidate for the family of fMRI analysis techniques which are now usually labelled 'multivariate pattern analysis'.

All algorithms of pattern classification currently in use for fMRI data (for reviews see Refs [56,57]) have in common the use of sub-threshold activation differences within a cluster of voxels to detect (i.e. optimally classify) different states of activation across this cluster of voxels. Non-smoothed and non-averaged data points from all voxels in a given brain region are considered by the classifier and can potentially contribute to optimally distinguishing different activation patterns, whereas some voxels will prove more informative to the classifier decisions than others (RFE, Recursive Feature Elimination; [8]). The algorithms are often trained on a given set of trials from all conditions to be classified and then used to test (on separate test trials which are kept aside from the fMRI time series) whether a multi-voxel pattern is more likely to reflect one or the other cognitive state or stimulus condition.

This means a fundamental change in experimental logic. Formisano and colleagues [8,60] recently demonstrated how such pattern recognition analyses might be used for understanding speech sound abstraction by using linear support-vector-machine learning. Their study presented three vowel categories by three speakers (i.e. included acoustic variance and used more than just binary stimulus conditions). Instead of confining the voxel search space to a small region of interest via a localiser experiment, the classifier was given data points from the whole fMRI volume. The results can be visualized by mapping the most informative voxels in classifying either vowel category or speaker identity from all subjects in a cortically aligned brain space (Figure I). Importantly, Formisano *et al.*'s [8] approach permits testing the classifier on trials in which the stimuli were physically different from the training trials: their findings show that the classification can generalize, for example to the same vowel uttered by a new talker, and thus are able to demonstrate learning of an 'abstract' representation [60].

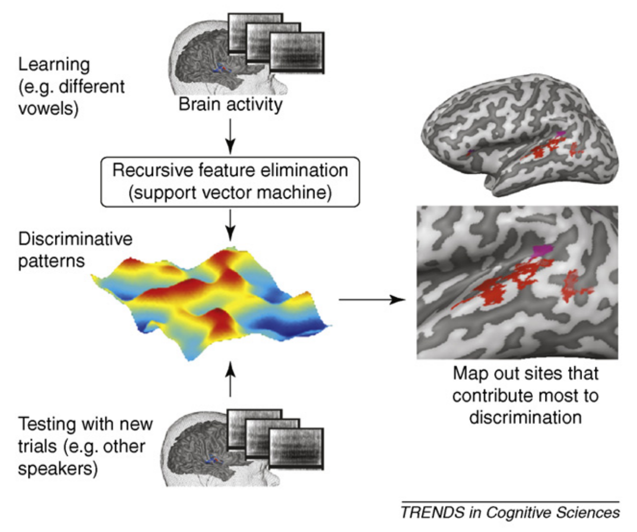

Figure I. Outline of Formisano *et al.*'s [8] approach of using a linear support-vector-machine learning combined with recursive feature elimination, applied to speech sound abstraction (under review; see Ref. [60] for a detailed description of the methodology). Materials reproduced with kind permission of Elia Formisano, see Ref. [60].

made accessible by the computations of the ascending auditory pathway and primary auditory cortex.

In search for the neuroanatomy of pre-lexical abstraction

Recent advances in functional magnetic resonance imaging (fMRI) methodology have made it possible to study the functional neuroanatomy of speech perception non-invasively in humans. Through neuroimaging we have gained insights on the neuroanatomical organization of the speech comprehension system as a whole (e.g. Refs [33–36]). However, relatively few studies have directly addressed pre-lexical processing and categorisation.

Figure 1b shows peak activations from a selection of studies that have targeted pre-lexical processing of speech with various strategies. They all used native sub-lexical speech stimuli, such as isolated vowels, consonants or syllables (except Ref. [37], which used pseudowords). The peak activations illustrate a functional hierarchy in the processing of these sounds.

First, relatively early components in the magnetocencephalographic (MEG) signal (P50m, N100m) have been localized to the lateral part of Heschl's gyrus. This response has been observed with many types of auditory stimuli and is unlikely to be unique to speech, although its precise latency and topography seem to be modulated by meaningful speech sounds [38–40].

Second, posterior to Heschl's gyrus, the planum temporale (PT) often does not show stronger responses to speech compared to non-speech sounds that are matched in terms of acoustic complexity. Rather, the PT exhibits sensitivity to acoustic-phonetic variation [41,42] and to spectrotemporal complexity in general [43]. On the basis of functional connectivity analyses, this region has been suggested to be at an intermediate stage of processing between primary auditory cortex and the superior temporal sulcus (STS) [44] (see also Box 2).

Third, it is only in the STS that responses to meaningful speech sounds differ consistently from non-speech baseline conditions (e.g. Refs [36,45–47]).

Fourth, detection of a phonological change [48], but not merely acoustic change, is associated with activity in the supramarginal gyrus (SMG). This indicates that the supramarginal gyrus has access to already-abstracted phonological units (Box 2), and these activations have often been interpreted as evidence that the SMG has an important role in phonological working memory [37,49–51].

The distribution of the peak activations in Figure 1b clearly shows that a vast area of the peri-Sylvian cortex is recruited for pre-lexical processing. Note in particular that there are prominent peaks along the STS. This area has also been shown to process lexical, semantic and syntactic information, talker variation and functions which are not necessarily linguistic, such as social interaction or multisensory integration [52–55]. Pre-lexical processing, therefore, does not seem to be confined to relatively low-level auditory cortex such as lateral Heschl's gyrus (Brodmann's Area 42 [BA 42]); most likely corresponding to human parabelt areas) or the planum temporale (BA 42/22). On a macroanatomical scale, there is considerable overlap in the sites along the STS that exhibit sensitivity to pre-lexical, lexical and possibly higher levels of linguistic processing.

Although there is evidence for a functional and neuroanatomical hierarchy in the cortical processing of speech

Box 2. Where is the locus of pre-lexical abstraction?

Figure I shows a schematic aggregate result across studies and brain regions: by applying cognitive subtraction logic, only the superior temporal sulcus (STS) shows a significant preference for speech sounds over meaningless noises or tones (also clearly reflected in Figure 1b in the main text). The planum temporale (PT) has been implied to be a hierarchically up-stream structure to STS [42,44]. It is often reported to show no such preference as long as a certain spectrotemporal complexity is attained (cf. the pattern for Heschl's gyrus [HG] or more primary auditory cortex responses), although its precise role in the functional hierarchy of abstraction is unresolved.

The problem faced with cognitive subtraction designs is a general one, which bears relevance when we aim to localize a transitory processing stage like pre-lexical abstraction:

- Can the output of a brain region A (for example the PT), which itself exhibits a non-categorical response but is activated to the same extent and strength by two stimulus types, be indicative of the categorization process nevertheless?
- Can we infer from equal activation by these complex stimuli (but less so by very simple stimuli) that it is treating these stimuli the same?

Maybe region A transmits distinct output for these two stimulus types and this output is in the shape of an abstract phonological category signalled to brain region B (for example the STS). In other words, is the categorical response observed in brain region B indicative of a successful categorization process having taken place at an hierarchically lower stage – or does it reflect the abstraction and categorization process itself (as suggested by our tentative interpretation of Figure 1b in main text)?

This is a fundamentally unsolved problem. It will be tackled best by moving beyond activation strength measures (such as the relative strength of a hemodynamic response or the amplitude of an event-related potential) and analysing the actual functional behaviour of a brain area (e.g. 'neurometric curves' [41]) or its complex multivariate response patterns from a set of voxels rather than an averaged response (Box 1).

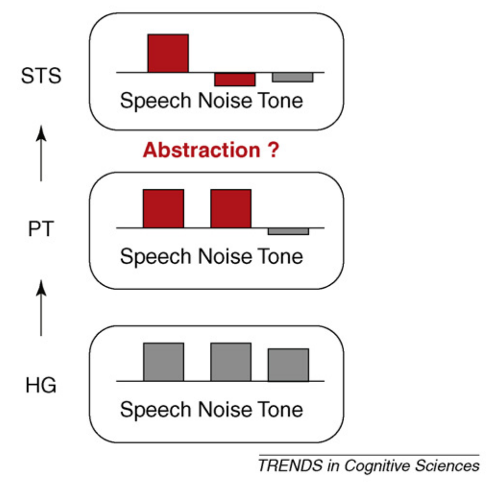

Figure I. A schematic aggregated result of cognitive-subtraction fMRI studies, illustrating the difficulty establishing a processing hierarchy for different stimuli based on the overall activation patterns in various regions. Are we to conclude that no difference in BOLD activation strength means no difference in processing? Is a difference in a region like the STS indicative of the abstraction process itself or rather the result of this process? Abbreviations: HG, Heschl's gyrus; PT, planum temporale; STS, superior temporal sulcus.

sounds, at present there are few human neuroimaging studies with sufficient neuroanatomical detail and adequate experimental manipulations to allow for direct inference on the pre-lexical transformation of the speech signal into phonological units in structures of auditory cortex. Furthermore, there are limitations to cognitive subtraction experiments. Because this method depends on comparisons of relative activation, the choice of baseline condition is crucial: processing steps of interest can be cancelled out when the experimental condition (e.g. different phonemes) and baseline condition (e.g. complex non-speech sounds) share acoustic features (see also Box 2).

As described in Box 1, a promising alternative approach to subtraction designs is multivariate pattern analysis: this subsumes a family of methods that has recently been applied to neuroimaging [56,57]. The fundamental leap forward compared to conventional fMRI studies is the simultaneous analysis of a whole array of data points from a given region, allowing for the detection of subtle and abstract neural response pattern differences (i.e. differences invisible if only looking at single data points at a time).

Concluding remarks

Behavioural investigations in speech sciences and computational modelling have led to a detailed understanding of how the speech perception system can be conceptualised. Although this type of research alone cannot produce a neuroanatomical model of speech processing, it should guide neuroscientific investigations by providing a theoretical framework. Using the cognitive subtraction method, functional neuroimaging studies have broadly defined the neuroanatomy of pre-lexical processing. Multivariate analysis techniques of neuroimaging data have the potential to study spectro-temporal encoding and abstraction of speech in more detail, and crucially, in a manner that can be related to results from other fields (Box 3). A study by Mitchell and colleagues [58] has already shown that it is feasible to train a computational model of language processing using neural data as input. We suggest that the output of these multivariate methods (Figure I in Box 1) can serve as input to cognitive models of speech perception, in parallel to behaviour-based likelihoods that have been used in speech science [17], waveform-based likelihoods that can be extracted with automatic speech recognition techniques [16] or spike-timing patterns that have been observed in animal studies [28]. The integration of findings from all of these areas, and the latest technological developments within each of them, can lead to a testable, neuroanatomical model of pre-lexical abstraction.

Box 3. Questions for future research

Clinical relevance of pre-lexical abstraction

There has been an increased interest from a clinical perspective in the neural mechanisms of pre-lexical abstraction, as certain clinical populations exhibit ineffective processing of speech. It is likely that an impairment of the abstraction process contributes to some congenital and acquired disorders such as specific language impairment, dyslexia or auditory processing disorder, but it is unclear at present which neural systems underlie this deficiency.

Feed-back versus feed-forward in auditory cortex

There is an ongoing debate and an abundance of behavioural evidence for or against online feedback in speech comprehension [17,61]. Although human neuroimaging data might indicate a feed-forward flow of information in speech sound abstraction, animal research has indicated that early auditory fields might contribute dynamically [24,62] to higher-level processing via local feedback loops and re-entrant processing. This is difficult to investigate in humans as BOLD fMRI provides a temporally integrated picture of neural activity, whereas MEG does not have sufficient spatial resolution to disentangle small auditory fields. Although intrinsically limited to patient studies, invasive electrophysiological measurements of neural activity could provide the missing link between animal studies and human neuroimaging [63].

Role of oscillatory networks

The electrophysiology of neural oscillations in perceptual object formation in the visual domain is increasingly well understood [64], and there is evidence to indicate that synchronization changes in oscillatory networks are also a signature of auditory and speech perception [65–67]. A functional neuroimaging method that relies on metabolic measures such as fMRI does not permit investigation of oscillatory networks. It is an important and timely goal to reconcile results from fMRI studies in the spatial-anatomical domain with these more time-based measures from EEG and MEG research.

Acknowledgements

J.O. is employed by the Max Planck Society, Germany, and received additional funding from the Landesstiftung Baden-Württemberg gGmbH during initial phases of this project. F.E. is funded through a Wellcome Trust Senior Fellowship awarded to Sophie Scott. We are grateful to D. Sauter, C. Cantona and three anonymous reviewers for very helpful comments.

References

1 Repp, B.H. and Liberman, A.M. (1987) Phonetic category boundaries are flexible. In *Categorical Perception* (Harnad, S., ed.), pp. 89–112, Cambridge University Press

2 Holt, L.L. (2006) Speech categorization in context: joint effects of nonspeech and speech precursors. *J. Acoust. Soc. Am.* 119, 4016–4026

3 Werker, J.F. and Tees, R.C. (2002) Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 25, 121–133

4 Norris, D. *et al.* (2003) Perceptual learning in speech. *Cognit. Psychol.* 47, 204–238

5 Scott, S.K. and Johnsrude, I.S. (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26, 100–107

6 Haxby, J.V. *et al.* (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430

7 De Martino, F. *et al.* (2007) Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers. *Neuroimage* 34, 177–194

8 De Martino, F. *et al.* (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fmri spatial patterns. *Neuroimage* 43, 44–58

9 Eisner, F. and McQueen, J.M. (2005) The specificity of perceptual learning in speech processing. *Percept. Psychophys.* 67, 224–238

10 Kraljic, T. and Samuel, A.G. (2007) Perceptual adjustments to multiple speakers. *J. Mem. Lang.* 56, 1–15

11 Cutler, A. *et al.* How abstract phonemic categories are necessary for coping with speaker-related variation. In *Laboratory Phonology* (Vol. 10) (Fougeron, C., ed.), Mouton de Gruyter (in press)

12 McQueen, J.M. *et al.* (2006) Phonological abstraction in the mental lexicon. *Cogn. Sci.* 30, 1113–1126

13 McQueen, J.M. (2005) Speech perception. In *The Handbook of Cognition* (Lamberts, K. and Goldstone, R., eds), pp. 255–275, Sage Publishers

14 McClelland, J.L. and Elman, J.L. (1986) The TRACE model of speech perception. *Cognit. Psychol.* 18, 1–86

15 McQueen, J.M. and Cutler, A. (2001) Spoken word access processes: an introduction. *Lang. Cogn. Process.* 16, 469–490

16 Scharenborg, O. *et al.* (2005) How should a speech recognizer work? *Cogn. Sci.* 29, 867–918

17 Norris, D. and McQueen, J.M. (2008) Shortlist B: a bayesian model of continuous speech recognition. *Psychol. Rev.* 115, 357–395

18 Kaas, J.H. and Hackett, T.A. (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11793–11799

19 Rauschecker, J.P. and Tian, B. (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11800–11806

20 Petkov, C.I. *et al.* (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol.* 4, e215

21 Petkov, C.I. *et al.* (2008) A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–374

22 Shamma, S. (2001) On the role of space and time in auditory processing. *Trends Cogn. Sci.* 5, 340–348

23 Read, H.L. *et al.* (2002) Functional architecture of auditory cortex. *Curr. Opin. Neurobiol.* 12, 433–440

24 Fritz, J. *et al.* (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223

25 Ohl, F.W. and Scheich, H. (1997) Orderly cortical representation of vowels based on formant interaction. *Proc. Natl. Acad. Sci. U. S. A.* 94, 9440–9444

26 Obleser, J. *et al.* (2004) Magnetic brain response mirrors extraction of phonological features from spoken words. *J. Cogn. Neurosci.* 16, 31–39

27 Ohl, F.W. *et al.* (2001) Change in pattern of ongoing cortical activity with auditory category learning. *Nature* 412, 733–736

28 Engineer, C.T. *et al.* (2008) Cortical activity patterns predict speech discrimination ability. *Nat. Neurosci.* 11, 603–608

29 Schreiner, C.E. and Winer, J.A. (2007) Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron* 56, 356–365

30 Eggermont, J.J. (2001) Between sound and perception: reviewing the search for a neural code. *Hear. Res.* 157, 1–42

31 Steinschneider, M. *et al.* (1995) Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain Res.* 674, 147–152

32 Steinschneider, M. *et al.* (2005) Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cereb. Cortex* 15, 170–186

33 Binder, J.R. *et al.* (1997) Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* 17, 353–362

34 Scott, S.K. *et al.* (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406

35 Davis, M.H. and Johnsrude, I.S. (2003) Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431

36 Liebenthal, E. *et al.* (2005) Neural substrates of phonemic perception. *Cereb. Cortex* 15, 1621–1631

37 Jacquemot, C. *et al.* (2003) Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J. Neurosci.* 23, 9541–9546

38 Obleser, J. *et al.* (2003) Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *Neuroimage* 20, 1839–1847

39 Obleser, J. *et al.* (2006) 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

40 Tavabi, K. *et al.* (2007) Auditory evoked fields differentially encode speech features: an MEG investigation of the p50m and n100m time courses during syllable processing. *Eur. J. Neurosci.* 25, 3155–3162

41 Raizada, R.D.S. and Poldrack, R.A. (2007) Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56, 726–740

42 Obleser, J. *et al.* (2007) Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb. Cortex* 17, 2251–2257

43 Griffiths, T.D. and Warren, J.D. (2002) The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353

44 Kumar, S. *et al.* (2007) Hierarchical processing of auditory objects in humans. *PLoS Comput. Biol.* 3, e100

45 Rimol, L.M. *et al.* (2005) Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study. *Neuroimage* 26, 1059–1067

46 Obleser, J. *et al.* (2006) Vowel sound extraction in anterior superior temporal cortex. *Hum. Brain Mapp.* 27, 562–571

47 Dehaene-Lambertz, G. *et al.* (2005) Neural correlates of switching from auditory to speech perception. *Neuroimage* 24, 21–33

48 Phillips, C. (2001) Levels of representation in the electrophysiology of speech perception. *Cogn. Sci.* 25, 711–731

49 Celsis, P. *et al.* (1999) Diferential differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage* 9, 135–144

50 Jacquemot, C. and Scott, S.K. (2006) What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci.* 10, 480–485

51 Buchsbaum, B.R. and D'Esposito, M. (2008) The search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* 20, 762–778

52 Hickok, G. and Poeppel, D. (2007) The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402

53 Wong, P.C.M. *et al.* (2004) Neural bases of talker normalization. *J. Cogn. Neurosci.* 16, 1173–1184

54 Beauchamp, M.S. *et al.* (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat. Neurosci.* 7, 1190–1192

55 Van Atteveldt, N. *et al.* (2004) Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282

56 Norman, K.A. *et al.* (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430

57 O'Toole, A.J. *et al.* (2007) Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *J. Cogn. Neurosci.* 19, 1735–1752

58 Mitchell, T.M. *et al.* (2008) Predicting human brain activity associated with the meaning of nouns. *Science* 320, 1191–1195

59 Eisner, F. and McQueen, J.M. (2006) Perceptual learning in speech: stability over time. *J. Acoust. Soc. Am.* 119, 1950–1953

60 Formisano, E. *et al.* (2008) 'Who's Saying 'What'? Brain-Based Decoding of Human Voice and Speech. *Science* (In press)

61 McClelland, J.L. *et al.* (2006) Are there interactive processes in speech perception? *Trends Cogn. Sci.* 10, 363–369

62 Polley, D.B. *et al.* (2006) Perceptual learning directs auditory cortical map reorganization through top-down influences. *J. Neurosci.* 26, 4970–4982

63 Boatman, D.F. and Miglioretti, D.L. (2005) Cortical sites critical for speech discrimination in normal and impaired listeners. *J. Neurosci.* 25, 5475–5480

64 Engel, A.K. and Singer, W. (2001) Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* 5, 16–25

65 Luo, H. and Poeppel, D. (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010

66 Supp, G.G. *et al.* (2004) Lexical memory search during N400: cortical couplings in auditory comprehension. *Neuroreport* 15, 1209–1213

67 Hannemann, R. *et al.* (2007) Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Res.* 1153, 134–143