New fMRI methods for the study of language

Roel M. Willems[1,2,3], Marcel A. J. van Gerven[2]

1. Centre for Language Studies, Radboud University, Nijmegen, The Netherlands
2. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
3. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Chapter 50 in *Oxford Handbook of Psycholinguistics*, Rueschemeyer & Gaskell (eds). Oxford University Press

Draft version

Abstract

The use of various techniques for measuring brain activation has led to a dramatic increase in knowledge about how the brain is involved in language. One of these techniques is functional Magnetic Resonance Imaging (fMRI). Here we describe novel ways of analyzing data that take away some of the classical limitations of fMRI. One important message from the chapter is that new analysis techniques allow for the use of more naturalistic and continuously presented stimuli like spoken narratives or movies, than was considered possible before. In part 1 we describe how some traditional limitations of fMRI for language research can relatively easily be overcome. In part 2 we describe state of the art approaches combining fMRI data analysis with computational modeling. We hope that description of these techniques will be inspirational for those who want to perform cognitive neuroscience studies of language, most notably at the level of discourse.

Keywords: fMRI; experimental design; language; computational modeling; naturalistic

Introduction

Functional Magnetic Resonance Imaging (fMRI) has become a standard part of the toolkit of psycholinguists. Ever since Ogawa and colleagues described how to measure a signal from the human brain which is a correlate of local brain activation (the Blood Oxygenation Level Dependent, or BOLD signal) (Ogawa et al., 1990), the first fMRI studies on language comprehension were performed[1]. A PubMed search with keywords Language AND fMRI shows that the number of publications has increased about threefold since the turn of the century (from on average 250 publications per year between 1998-2002, to around 750 publications per year in the 2011-2015 period). With its increased popularity also knowledge about the possibilities and impossibilities of research using fMRI has greatly increased. Some of the 'what you cannot do with fMRI'-knowledge is outdated, and the first part of this chapter is devoted to challenging a few popular beliefs about the limits of using fMRI in language research. Most of these issues were valid in the early days of fMRI research, but new developments in hardware and analysis techniques render that they can be overcome. An important advance is that novel fMRI data analysis techniques allow for the use of more naturalistic language stimuli, e.g. the presentation of continuous language recordings or movies. The second part of the chapter builds on this by introducing and discusses several analysis techniques which use computational modeling to get better insight into the neurocognitive underpinnings of language. These are advanced and relatively novel techniques and here we aim to illustrate their potential for the use of more naturalistic stimuli in fMRI studies of language.

Part 1: Common beliefs about fMRI in language research

Here are 3 common beliefs about fMRI that researchers take into account when designing a psycholinguistic fMRI experiment:

- Stimuli should be separated by relatively long intertrial intervals (in the order of 5 to 6 seconds), so presentation of continuous language stimuli is not possible (section 1.1)
- Stimuli cannot be presented auditorily because of the loud noise produced by the MR scanner (section 1.2)
- Participants cannot speak in the MR environment so studying speech production is impossible (section 1.3)

[1] Note that earlier neuroimaging studies of language comprehension were done using Positron Emission Tomography (PET), (e.g. Mazoyer et al., 1993; Petersen et al., 1990; Price et al., 1994). The focus of this chapter is on fMRI and we will not discuss PET and its applications, except for noting that the popularity of PET has decreased given its invasive nature (a contrast agent has to be inserted), and refinement of fMRI techniques.

In the remainder of part 1 we will show that these common beliefs are outdated. They can be relatively easily overcome.

1.1. fMRI can be used with continuous stimuli

Presenting continuous stimuli (e.g. spoken narratives, movies, recordings of conversations) is typically avoided in fMRI research. The reason has to do with the slowness of the BOLD response. Figure 1 shows an idealized BOLD response. A stimulus is presented at time point 0, and as can be seen in the figure the BOLD response only peaks about 6 seconds later (one time point in the figure is one Time to Repetition (TR) which is 2 seconds in this case). This poses a problem for rapid presentation of stimuli. If we present stimuli in rapid succession, the BOLD curves to each stimulus will start to overlap, and it will be impossible to assess which stimulus generated which response. Even worse, when we present stimuli very rapidly after one another the estimated BOLD response will start to plateau, i.e. it will have no variance left. Figure 2A illustrates this. What we see here is the *estimated* BOLD response (so not an actually measured BOLD response) in a hypothetical experiment in which a word was presented every 2 seconds, for a duration of 500 ms per word (intertrial interval = 1.5 seconds). What we see in the figure is that the estimated BOLD response rises quickly at the start of the experiment, and then plateaus to become a flat line. The standard way of doing fMRI data analysis is to fit for each and every voxel the *estimated* BOLD response to the *actual* BOLD response, and see how good the fit is. This is linear regression, and trying to regress an (almost) flat line onto a signal will not give a good fit. It becomes impossible to assess whether a given voxel was sensitive to presentation of the words.

One way to solve this issue of overlapping (plateauing) BOLD curves is to present a stimulus, and then wait for the BOLD response to go back to baseline before presenting the next stimulus. This is called slow event-related fMRI. With intertrial intervals of around 16-20 seconds it is very inefficient in that it takes a long time to collect data for a reasonable amount of trials. A better solution is rapid event-related fMRI in which stimuli are presented relatively fast after one another (~4 seconds ITI), but with *variable* intertrial intervals. That is, sometimes the ITI is 4 seconds, sometimes it is 2.5 seconds, sometimes it is 6.5 seconds, etc. In this way we create variance in the estimated BOLD signal, as is illustrated in figure 2B. In the figure we see the estimated BOLD curve (in blue) for stimuli presented with an average ITI of 5 seconds, but this ITI varies from trial to trial (the stimulus onsets are represented in red). We see that there is variance in the estimated BOLD signal, and linear regression of the estimated onto the actual BOLD signal can be used. The introduction of rapid event-related fMRI was a breakthrough for the field (see Miezin et al., 2000 for an excellent explanation),

but it still places a limit on the speed with which we can present stimuli. The ITI cannot be too short (how short it can be is a matter of debate). An extra handicap is the slow sampling rate in a typical fMRI experiment. The brain is usually sampled every 2 seconds, and hence we measure the BOLD curve only very sparsely.

Intertrial intervals of 5 seconds or more are still too slow when using continuous stimuli such as speech spoken at a normal rate. Yarkoni and colleagues introduced a clever trick which allows stimuli to be presented much more rapidly (Yarkoni et al., 2008). They had participants read short narratives one word at a time. Each word was on the screen for 200 ms, with an intertrial interval (inter-word interval) of 150 ms. Estimating the BOLD response to the single words leads to a plateaued response (this is the situation as in Fig. 2A). Instead of estimating the BOLD curve to the words, Yarkoni and colleagues estimated the BOLD curves to several word *characteristics.* For instance, they asked which brain regions were sensitive to differences in lexical frequency between words. The lexical frequencies naturally differed considerably between the words, and this creates the necessary variance in the estimated BOLD curve for lexical frequency. The approach is illustrated in Figures 2C and 2D, which show estimated BOLD responses from an experiment in which we applied this approach (Willems et al., 2016). In the experiment participants listened to short narratives spoken at a natural rate. The sampling rate (TR) of the experiment was higher than usual. Brain activity was measured every 880 ms (TR=0.88 sec). Since words have different durations (in the spoken modality), and the pauses between words are not constant, there is a naturally occurring jitter between words. First note that this means that the estimated BOLD response to *words* does not plateau (Fig. 2C), but shows some variance. This means that we can fit brain responses to words with very high efficiency: in this example 1291 words were presented and only 8 minutes of data collection per subject was sufficient to collect all data for the experiment (note that the information in Figure 2C was only a subset of the design of the original experiment). Second, and the main point of the technique pioneered by Yarkoni and colleagues, we asked which brain regions were sensitive to differences in surprisal value between the words. Surprisal value is a measure derived from information theory and is related to the expectancy of a given word. The estimated BOLD response to surprisal value is shown in Figure 2D and as we can see, the variance in the estimated signal is considerable, despite the rapid presentation of words. Note that this approach also works for natural reading, provided that eye movements are recorded during the scanning session (eyetracking combined with fMRI) (Choi et al., 2014; Schuster et al., 2015).

Here we end the more technical exposition (there is a large literature on design efficiency and estimability in fMRI, we mention (Liu and Frank, 2004; Liu, 2004)), and formulate the take home message of this section. The take home message is that one can present stimuli very rapidly

('continuously') in an fMRI experiment, as long as there is enough variance in the stimuli as regards the characteristic of interest. This will ensure that the estimated BOLD response does not plateau, but has enough variance in it to be regressed onto the actual measured BOLD response. One advantage of this approach is that stimuli can be presented at a more natural pace then with standard fast event-related fMRI. Another advantage is a dramatic increase in efficiency: many trials can be presented within a short amount of scanning time. Finally, an advantage is that the approach allows for post-hoc characterization of the stimuli. If a researcher wants to investigate a different characteristic than the main reason of performing the study, this is possible (with sufficient variance in the characteristic of interest in the stimulus). This opens up the possibility of re-using existing data sets, further increasing experimental efficiency (e.g. (Hanke et al., 2014)).

In the remainder of this section we briefly mention a few other ways of analyzing fMRI data that are acquired while participants were presented with continuous language stimuli. The first is a straightforward variant on the Yarkoni et al. approach. Instead of creating an estimated BOLD response including all stimuli presented in rapid succession, one can also focus on only few time points within a continuous stream of information. Zacks and colleagues for instance showed brief movie clips to participants and created an estimated BOLD response to the points in time in which an event change occurred in the movie clip (Zacks et al., 2001). In this way the BOLD events were separated far enough from each other in non-regular intervals to lead to an estimated BOLD response which does not plateau (see also Speer et al., 2009). In a similar vein we modeled the BOLD response to action and mentalizing events occurring within a narrative presented at a normal speech rate (Nijhof and Willems, 2015).

Another way of analyzing fMRI data that are acquired while participants engage in viewing or listening to continuous stimuli was applied by Lerner and colleagues (Lerner et al., 2011). They presented a short narrative which was scrambled in time at different time scales. Participants would listen to the original narrative (no scrambling), to a version in which paragraphs were scrambled (breaking continuity at that particular time scale), to a version in which sentence order was scrambled, or to a version in which words were scrambled. They used inter-subject correlation analysis to assess which brain regions show a similar time course across participants for the original story, and compared this to brain areas which show the same time course across participants for the versions in which paragraphs, sentences, or words were scrambled. Other analysis techniques exist for analyzing data from continuous language stimuli. Some of these are introduced in Part 2 below. For an excellent overview the interested reader is referred to Andric and Small (2015).

1.2. fMRI can be used with auditory stimuli

While collecting images the MR machine produces very loud noises. These are caused by switching of the magnetic gradients and by the very large force applied to for instance cables in the MR machine. Participants for instance need to wear ear plugs protection to avoid hearing damage. There are, however, dedicated inner ear phones that allow for presenting auditory stimuli and at the same time minimizing disturbance from the scanner noise. Next to this, modern MRI machines tend to be equipped with hardware which reduces scanner noise. Together this means that auditory presentation works well for single word presentation and up (single sentences / extended pieces of discourse). For presentation of for instance phonemes the interference of the scanner noise is sometimes considered too disturbing and sparse scanning sequences can be used (see Peelle, 2014 for overview). These are scanning sequences in which the machine is not collecting images *during* presentation of the stimuli (and hence no loud noise is emitted), but only *after* presentation of the stimuli. This approach takes advantage of the slowness of the BOLD response. The reasoning is that since the BOLD response lags behind neural activation anyway, it is not necessary to sample BOLD at the time of presentation of the stimulus. After all, the signal that we measure will only peak several seconds after presentation of the stimulus. By sampling only when activation is expected to be measurable, a silent stimulus presentation can be combined with measurement of the neural response to the stimulus. The price that is paid with this approach is that the brain activation is not sampled continuously, and that it relies on having a reliable estimate of the time course of the BOLD curve. These are relatively minor shortcomings, and the approach has shown its merits in various studies. Like we said above, from our experience, recent advances in scanner hardware as well as presentation equipment (e.g. headphones) render silent scanning not necessary for the bulk of auditory language experiments.

1.3. Studying speech production with fMRI works

Movements of the participant can be detrimental to fMRI data. Movements of the head are most problematic, but it should be noted that large movements of for instance the arms can hardly be performed without moving the head, albeit even slightly. Precautions are taken to avoid head movements: Participants are asked to lie as still as possible, and the head is fixated to further reduce head motion. Cushions can for instance be placed in the empty spaces between the head and the head coil to constrain movement of the head. Other options to stabilize the head are head casts and bite bars. One reason why head motion is bad for fMRI is that it makes voxels 'move' artificially. In the data analysis we assume that a voxel which is at a given location at the beginning of the

experiment, will still be in that location at the end of the experiment. If there is a displacement of the head let's say halfway the experiment, the time course of that voxel will be contaminated: It will have the time course of voxel X for the first half of the experiment and the time course of voxel Y for the second half of the experiment. While this is very bad in the case of large displacements, in typical group analyses the data are spatially smoothed (or 'blurred'), rendering the effect of small head movements manageable (except for cases in which spatial smoothing is unwanted, see part 2). A second – and related – reason why motion is bad for fMRI data is that it can lead to *edge artifacts*. These can occur at the edges of the brain, where neural tissue borders other kinds of tissue or fluids. Examples are the ventricles which are filled with cerebro-spinal fluid (CSF), or the tissue and CSF which surrounds the cortex at the outer part of the brain, just beneath the skull. These parts of the brain are most likely not involved in cognitive processes, and the BOLD response of CSF will be low and (most likely) be unrelated to language comprehension. Now suppose there is an area whose BOLD response is very sensitive to language comprehension, and there is a displacement of the head around halfway the experiment, just when the participant heard a stimulus of condition A. Due to the displacement, voxels which were silent in the first part of the experiment (voxels in the CSF) are replaced by cortical voxels which are sensitive to the stimulus. In our statistical analysis there will be a very large artificial 'response' to condition A at the time point of the displacement, which will not be present for condition B. Proper stimulus randomization will take care of such influences to a large degree when the motion is not systematically correlated with the stimulus. When studying speech production, there is a more severe problem with head motion. In speech production, the event of interest is by definition when participants move their head (i.e. when they speak). Now if for some reason the movement is slightly more during one condition compared to the other, the contrast between the two conditions will show a lot of edge artifacts. The statistical map will show bright colors (high statistical values) at the edges of the cortex and the skull, and around the ventricles. If motion is strongly correlated with the events of interest (the trials), it potentially becomes impossible to separate artificial 'activations' (edge artifacts) from real activations, related to speaking. Because of the harmful effect of stimulus-correlated motion speech production studies were traditionally avoided in fMRI (instead, Positron Emission Tomography (PET) was used as a preferred method). Interestingly speech production studies show that it is possible to have a reliable signal while participants speak in the scanner (e.g. Segaert et al., 2012). So while it is important to consider the points raised above, experience shows that in practice it is possible to study neural activity related to speech production with fMRI.

As a final note we want to point out that verbal responses can be recorded in a scanner environment. Although these recordings will be noisy (due to the scanner noises), on-line and off-line filtering will

make them eligible and suited for scoring, both during as well as after the experiment (de Boer et al., 2013; Willems et al., 2010).

Part 2: Computational modeling and its application to fMRI

Next to the use of experimental designs that approach increasingly realistic conditions of speech perception and production, new developments in computational modeling are paving the way towards a more fine-grained understanding of how the human brain processes linguistic stimuli. We here discuss computational modeling approaches that have gained prominence in recent years. These approaches provide increased sensitivity compared to the contrast-based general linear model (GLM) approach that is conventionally used in cognitive neuroscience. Furthermore, they allow for a more detailed modeling of how human brain activity is perturbed by high-dimensional and semantically rich sensory input.

2.1 Multivariate pattern analysis

Multivariate pattern analysis (MVPA) refers to the decoding of task regressors from *multivariate patterns* of brain activity (Cox and Savoy, 2003; Haynes and Rees, 2006; Heinzle et al., 2012; Norman et al., 2006; Tong and Pratte, 2012). This can be contrasted with the conventional GLM approach, where task and nuisance regressors are used to predict *univariate* single-voxel responses. A main advantage of MVPA over the GLM approach is that it uses multivariate response patterns rather than responses in individual voxels, thereby increasing sensitivity (but see (Allefeld and Haynes, 2014) for a multivariate extension of the GLM).

Prediction of task regressors from patterns of brain activity is achieved by estimating a predictive model on training data and evaluating its predictive performance on test data which has not been used for model estimation, see (Bzdok, 2016) for a review of such models. Model evaluation is typically repeated in multiple regions of interest (ROI) or using a searchlight approach (Kriegeskorte et al., 2006) which uses small spherical ROIs centered at each voxel in the brain to obtain a measure of predictive performance across the brain proper.

As a case in point consider the study by (Abrams et al., 2013) which showed that using MVPA a more extensive brain network was identified which discriminates between intelligible and unintelligible speech compared to the use of a univariate general linear model. In the same vein, (Staeren et al., 2009) showed that information related to presented sound categories could be detected from fMRI responses using MVPA but remained undetectable using conventional contrast-based approaches.

MVPA has not only been used to isolate regions involved in low-level linguistic processing but also to isolate regions pertaining to more abstract semantic processing. For example, (Simanova et al., 2014) have used MVPA to show that conceptual information independent from input modality is represented in frontal areas as well as left inferior temporal cortex. This was achieved by training a classifier on one input modality and testing it on another input modality (cf. Figure 3A). In related work, an MVPA searchlight approach was used to show that fMRI-based decoding of spoken words in bilinguals revealed language-independent semantic representations in anterior temporal lobe (Correia et al., 2014). As another example of the decoding of high-level semantic information, (Frankland and Greene, 2015) have used MVPA to demonstrate the involvement of left mid-superior temporal cortex in the encoding of sentence meaning.

2.2 Encoding models

Encoding models (Naselaris et al., 2011; van Gerven, 2016) are a family of models that, similar to the GLM, predict voxel-specific responses from a set of regressors. However, in case of encoding models, the objective is not to define a small set of task regressors that are hypothesized to modulate brain activity, but rather to come up with a computational model that explains as much of the variance in the data as possible. In other words, an important difference between the conventional GLM approach and the present one is that a predictive model is generated on the basis of the *data*. In the conventional GLM approach a predictive model is generated by the *researcher*, who tests the hypothesis that voxels will respond to the stimuli in a particular way by introducing suitable task regressors. By comparing different computational models in terms of explanatory power we can select among competing hypotheses about brain function.

Encoding models consist of a feature model and a forward model. The feature model maps input (e.g. visual images or presented words) to a feature space whose features are thought to modulate brain activity. The forward model is used to predict brain measurements from feature values. Encoding models can, for example, be used to elucidate how primary auditory cortex responds to naturalistic input in terms of low-level sensory features (Santoro et al., 2014). Results show that neuronal populations in posterior auditory regions prefer coarse spectral information at high temporal precision whereas neuronal populations in anterior auditory regions prefer fine-grained spectral information at low temporal precision. In the language domain, the use of encoding models was pioneered by (Mitchell et al., 2008), who have shown that fMRI activation patterns induced by nouns that have not been seen before by the encoding model can be predicted by mapping nouns to their associated verbs using Wordnet (Miller, 1995) and subsequently mapping verbs to fMRI voxel responses using a linear regression model. In a more recent variant of this approach, (Huth et al.,

2016) constructed a stimulus representation based on word co-occurrences which was used to map semantic selectivity across cortex using fMRI data collected while subjects listened to hours of narrative stories. Results showed that selectivity was relatively symmetric between hemispheres and largely reproducible across subjects. Encoding models can be used to transform complex naturalistic stimuli into more meaningful representations that can be shown to drive neural responses. The task of the computational modeler is to isolate which stimulus representation most accurately drives the responses. These representations are becoming increasingly intricate. For example, (Wehbe et al., 2014) have used low-level, syntactic, semantic, and discourse features to examine which brain regions are involved in story reading subprocesses.

Another closely related computational approach is representational similarity analysis (RSA) (Kriegeskorte et al., 2008). As in encoding models a feature model is used to transform input stimuli to feature values. However, rather than mapping these feature values to observed brain measurements using a forward model, RSA bypasses the construction of a forward model and uses the feature values to evaluate whether the representational structure of the employed feature model matches that of the response patterns in a particular brain region. This is realized by evaluating the (dis)similarities between trials according to the feature model as well as according to the brain region at hand. If the (dis)similarity patterns match, the feature model and the brain region are considered to reflect similar computational properties. In the language domain, RSA has been used, for instance, to reveal commonalities and differences in the semantic processing of words and objects (Devereux et al., 2013) as well as to reveal a hierarchical organization of auditory and motor representations in speech perception (Evans and Davis, 2015).

Recently, researchers have started to explore the use of artificial neural networks (ANNs) as models of human brain function both in the context of the encoding and the RSA approach (Kriegeskorte, 2015; van Gerven, 2016; Yamins and Dicarlo, 2016). ANNs consist of loosely coupled elements called artificial neurons that are able to learn complex non-linear mappings from input vectors to output vectors. They are particularly attractive as computational models in cognitive neuroscience. First, due to new advances in deep and recurrent neural network learning (Cox and Dean, 2014; LeCun et al., 2015), their performance is starting to approach that of humans in cognitively challenging tasks. Second, as ANNs are loosely modeled after their biological counterparts, they provide an opportunity to gain new insights into the nature of representation learning in biological neural networks (Chung et al., 2016; Saxe et al., 2013).

In the past, neural networks have been used to learn representations of semantic information that can account for behavioral data (Joanisse and McClelland, 2015; McClelland, 2003). For example.

(Hinton and Shallice, 1991) showed that a recurrent neural network which was trained to output semantic feature vectors when presented with letter strings exhibited characteristics that are associated with dyslexia when the network was lesioned. More recently, ANNs have started to become used to predict neural response patterns induced by subjects' processing of sensory input. As an example, ANNs are able to learn representations of distributional semantics by mapping single words to dense output vectors that capture the semantic content of individual words (Mikolov et al., 2013). These distributed codes have been used in the context of encoding models to indirectly map single-word representations of visual images to BOLD activation patterns (Umut Güçlü and van Gerven, 2015a; S. Nishida, A. G. Huth, J. L. Gallant, 2015). Expanding on this work, by using recurrent neural networks that capture temporal dynamics, researchers have demonstrated improved predictions of fMRI responses to conceptual information (Güçlü and van Gerven, 2016) and have produced natural language descriptions of experienced stimuli from fMRI activation patterns (E Matsuo, I Kobayashi, S Nishimoto, S Nishida, 2016).

New advances in deep learning (LeCun et al., 2015), in which ANNs are employed that consist of many layers of increasingly complex representations of a stimulus are also starting to have an impact in the language domain. It has, for example, transformed research in computational linguistics (Manning, 2015). In cognitive neuroscience, deep neural networks have been shown to provide state-of-the-art predictions of neural responses to naturalistic stimuli, with deep layers projecting to increasingly downstream areas of the ventral and dorsal visual pathways (U Güçlü and van Gerven, 2015; Umut Güçlü and van Gerven, 2015b). Deep neural networks that have been trained to represent speech (Kell et al., 2016) or music (Güçlü et al., 2016) stimuli are now starting to shed light on the functional organization of auditory cortex, showing that areas distal from primary auditory cortex capture increasingly complex stimulus features (Güçlü et al., 2016). See Figure 3B for an example.

One may ask to what extent neural networks can be considered accurate computational models of human brain function. Our response to this question is that, as Box famously quipped, *all models are wrong but some are useful* (Box, 1979). We may add to this that usefulness should always be evaluated relative to our objectives. If the goal is to identify which brain regions are modulated by a particular task regressor then the GLM can be the model of choice. If, on the other hand, the goal is to come up with a model which best explains the behavioral and neural data obtained as subjects engage in cognitively challenging tasks, an artificial neural network would be the preferred model. In the language domain such an ANN would have to account for all the intricacies associated with processing of linguistic information. This includes explaining how raw sensory signals (e.g. speech or written words) are transformed by neural computations to yield highly abstract and semantically

meaningful representations as well as how predictive and attentional processes shape the processing of linguistic information (Clark, 2016).

Concluding, we consider the use of advanced computational models such as contemporary neural networks to be an exciting and important development as it shifts the research focus from human brain *mapping* to human brain *modelling*. That is, rather than demonstrating the involvement of a particular brain region in a certain restricted task, we are in the business of developing computational models that aim to predict as accurately as possible how the whole brain responds to naturalistic and ecologically valid input (Einhäuser and König, 2010) as measured using increasingly sophisticated experimental and acquisition protocols. We expect the endeavor to build explicit models of human brain function that are able to solve cognitively challenging tasks to become increasingly important in the neuroscience of language.

Summary and concluding remarks

In this chapter we described recent developments in fMRI research which extend the possibilities of using fMRI for language research beyond what was typically considered possible. We illustrated that with relatively straightforward adaptations continuous stimuli can be used in fMRI research. Examples are narratives or other discourse spoken at a natural rate, or read with a natural reading pace. We additionally described that fMRI can be used to study auditory stimuli (despite the loud noise from the scanner), and that speech production can be studied despite head motion inherent to speaking. We then illustrated new possibilities for fMRI research that employ computational modeling. Examples are approaches that generate a model of relevant features of stimuli using one part of the data, and test the fit of this model on another part of the data, and approaches in which the fit of predictions from – for instance – competing cognitive models onto neural data is assessed. These approaches are computationally demanding, and require an expertise which is not typically part of the skill set of psycholinguists. Close collaboration will therefore be required between those who know which are the relevant exciting new questions that need to be asked, and those who are capable of designing the sophisticated computational models that allow these questions to be addressed.

References

Abrams, D.A., Ryali, S., Chen, T., Balaban, E., Levitin, D.J., Menon, V., 2013. Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's areas. Cereb. Cortex 23, 1703–1714. doi:10.1093/cercor/bhs165

Allefeld, C., Haynes, J.D., 2014. Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. Neuroimage 89, 345–357. doi:10.1016/j.neuroimage.2013.11.043

Andric, M., Small, S.L., 2015. fMRI methods for studying the neurobiology of language under naturalistic conditions, in: Willems, R.M. (Ed.), Cognitive Neuroscience of Natural Language Use. Cambridge University Press, Cambridge, UK.

Box, G.E.P., 1979. Robustness in the strategy of scientific model building, in: Launer, R.L., Wilkinson, G.N. (Eds.), Robustness in Statistics. Academic Press, pp. 201–236.

Bzdok, D., 2016. Classical Statistics and Statistical Learning in Imaging Neuroscience. arXiv Prepr. 1603.01857, 1–50.

Choi, W., Desai, R.H., Henderson, J.M., 2014. The neural substrates of natural reading: a comparison of normal and nonword text using eyetracking and fMRI. Front. Hum. Neurosci. 8, 1024. doi:10.3389/fnhum.2014.01024

Chung, S., Lee, D.D., Sompolinsky, H., 2016. Linear readout of object manifolds. Phys. Rev. E 93, 060301. doi:10.1103/PhysRevE.93.060301

Clark, A., 2016. Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Oxford University Press.

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., Bonte, M., Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., Bonte, M., 2014. Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. J. Neurosci. 34, 332–8. doi:10.1523/JNEUROSCI.1302-13.2014

Cox, D.D., Dean, T., 2014. Neural networks and neuroscience-inspired computer vision. Curr. Biol. 24, R921–R929. doi:10.1016/j.cub.2014.08.026

Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI) ``brain reading'': detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19, 261–270.

de Boer, M., Toni, I., Willems, R.M., 2013. What drives successful verbal communication? Front. Hum. Neurosci. 7, 622. doi:10.3389/fnhum.2013.00622

Devereux, B.J., Clarke, A., Marouchos, A., Tyler, L.K., 2013. Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. J. Neurosci. 33, 18906–16. doi:10.1523/JNEUROSCI.3809-13.2013

E Matsuo, I Kobayashi, S Nishimoto, S Nishida, H.A., 2016. Generating natural language descriptions for semantic representations of human brain activity, in: Association for Computational Linguistics. pp. 22–27.

Einhäuser, W., König, P., 2010. Getting real - Sensory processing of natural stimuli. Curr. Opin. Neurobiol. 20, 389–395. doi:10.1016/j.conb.2010.03.010

Evans, S., Davis, M.H., 2015. Hierarchical organization of auditory and motor representations in speech perception: Evidence from searchlight similarity analysis. Cereb. Cortex 25, 4772–4788. doi:10.1093/cercor/bhv136

Frankland, S.M., Greene, J.D., 2015. An architecture for encoding sentence meaning in left mid-superior temporal cortex. Pnas 112, 11732–11737. doi:10.1073/pnas.1421236112

Güçlü, U., Thielen, J., Hanke, M., van Gerven, M.A.J., 2016. Brains on Beats, in: Neural Information Processing Systems. pp. 1–12.

Güçlü, U., van Gerven, M., 2015. Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. Neuroimage In Press.

Güçlü, U., van Gerven, M.A.J., 2016. Modeling the dynamics of human brain activity with recurrent neural networks. Front. Syst. Neurosci. 1–19.

Güçlü, U., van Gerven, M.A.J., 2015a. Semantic vector space models predict neural responses to complex visual stimuli. arXiv Prepr. 1510.04738, 1–7.

Güçlü, U., van Gerven, M.A.J., 2015b. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. J. Neurosci. 35, 10005–10014. doi:10.1523/JNEUROSCI.5023-14.2015

Hanke, M., Baumgartner, F.J., Ibe, P., Kaule, F.R., Pollmann, S., Speck, O., Zinke, W., Stadler, J., 2014. A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. Sci. Data 1, 1–18. doi:10.1038/sdata.2014.3

Haynes, J.-D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 7, 523–534.

Heinzle, J., Anders, S., Bode, S., Bogler, C., Chen, Y., Cichy, R.M., Hackmack, K., Kahnt, T., Kalberlah, C., Reverberi, C., Soon, C.S., Tusche, A., Weygandt, M., Haynes, J.-D., 2012. Multivariate decoding of fMRI data. e-Neuroforum 3, 1–16.

Hinton, G.E., Shallice, T., 1991. Lesioning an attractor network: investigations of acquired dyslexia. Psychol. Rev. 98, 74–95. doi:10.1037/0033-295X.98.1.74

Huth, A.G., Heer, W.A. De, Griffiths, T.L., Theunissen, F.E., Jack, L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532, 453–458. doi:10.1038/nature17637

Joanisse, M.F., McClelland, J.L., 2015. Connectionist perspectives on language learning, representation and processing. Wiley Interdiscip. Rev. Cogn. Sci. 6, 235–247. doi:10.1002/wcs.1340

Kell, A., Yamins, D., Norman-Haignere, S., McDermott, J., 2016. Speech-trained neural networks behave like human listeners and reveal a hierarchy in auditory cortex, in: CoSyne 2016. pp. 109–110.

Kriegeskorte, N., 2015. Deep neural networks: A new framework for modeling biological vision and brain information processing. Annu. Rev. Vis. Sci. 1, 417–446. doi:10.1146/annurev-vision-082114-035447

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103, 3863–3868.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. doi:10.1038/nature14539

Lerner, Y., Honey, C.J., Silbert, L.J., Hasson, U., 2011. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J. Neurosci. 31, 2906–2915.

Liu, T.T., 2004. Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part II: design of experiments. Neuroimage 21, 401–413.

Liu, T.T., Frank, L.R., 2004. Efficiency, power, and entropy in event-related FMRI with multiple trial

types. Part I: theory. Neuroimage 21, 387–400.

Manning, C.D., 2015. Computational Linguistics and Deep Learning. Comput. Linguist. 41, 701–707. doi:10.1162/COLI

Mazoyer, B.M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., Mehler, J., 1993. The cortical representation of speech. J. Cogn. Neurosci. 5, 467–479. doi:10.1162/jocn.1993.5.4.467

McClelland, J.L., 2003. The parallel distributed processing approach to semantic cognition. Nat. Rev. Neurosci.

Miezin, F.M., Maccotta, L., Ollinger, J.M., Petersen, S.E., Buckner, R.L., 2000. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. Neuroimage 11, 735–59.

Mikolov, T., Yih, W., Zweig, G., 2013. Linguistic regularities in continuous space word representations. Proc. NAACL-HLT 746–751.

Miller, G. a., 1995. WordNet: a lexical database for English. Commun. ACM 38, 39–41. doi:10.1145/219717.219748

Mitchell, T.M., Shinkareva, S. V, Carlson, A., Chang, K.-M., Malave, V.L., Mason, R.A., Just, M.A., 2008. Predicting human brain activity associated with the meanings of nouns. Science (80-.). 320, 1191–1195.

Naselaris, T., Kay, K.N., Nishimoto, S., Gallant, J.L., 2011. Encoding and decoding in fMRI. Neuroimage 56, 400–410.

Nijhof, A.D., Willems, R.M., 2015. Simulating Fiction: Individual Differences in Literature Comprehension Revealed with fMRI. PLoS One 10, e0116492. doi:10.1371/journal.pone.0116492

Norman, K.A., Polyn, S.M.M., Detra, G.J., Haxby, J. V, 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10, 424–430.

Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. U. S. A. 87, 9868–9872.

Peelle, J.E., 2014. Methodological challenges and solutions in auditory functional magnetic resonance imaging. Front. Neurosci. 8, 253. doi:10.3389/fnins.2014.00253

Petersen, S.E., Fox, P.T., Snyder, A.Z., Raichle, M.E., 1990. Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. Science 249, 1041–1044.

Price, C.J., Wise, R.J., Watson, J.D., Patterson, K., Howard, D., Frackowiak, R.S., 1994. Brain activity during reading. The effects of exposure duration and task. Brain A J. Neurol. 117 (Pt 6), 1255–1269.

S. Nishida, A. G. Huth, J. L. Gallant, S.N., 2015. Word statistics in large-scale texts explain the human cortical semantic representation of objects, actions, and impressions.

Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., Formisano, E., 2014. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. PLoS Comput. Biol. 10, e1003412. doi:10.1371/journal.pcbi.1003412

Saxe, A., McClelland, J., Ganguli, S., 2013. Dynamics of learning in deep linear neural networks. Adv. Neural Inf. Process. Syst. 1–9.

Schuster, S., Hawelka, S., Richlan, F., Ludersdorfer, P., Hutzler, F., 2015. Eyes on words: A fixation-related fMRI study of the left occipito-temporal cortex during self-paced silent reading of words and pseudowords. Sci. Rep. 5, 12686. doi:10.1038/srep12686

Segaert, K., Menenti, L., Weber, K., Petersson, K.M., Hagoort, P., 2012. Shared syntax in language production and language comprehension--an FMRI study. Cereb. Cortex (New York, N.Y. 1991) 22, 1662–1670. doi:10.1093/cercor/bhr249

Simanova, I., Hagoort, P., Oostenveld, R., van Gerven, M., 2014. Modality-independent decoding of semantic information from the human brain. Cereb. Cortex 24, 426–434.

Speer, N.K., Reynolds, J.R., Swallow, K.M., Zacks, J.M., 2009. Reading stories activates neural representations of visual and motor experiences. Psychol. Sci. 20, 989–999. doi:10.1111/j.1467-9280.2009.02397.x

Staeren, N., Renvall, H., De Martino, F., Goebel, R., Formisano, E., 2009. Sound categories are represented as distributed patterns in the human auditory cortex. Curr. Biol. 19, 498–502.

Tong, F., Pratte, M.M.S., 2012. Decoding patterns of human brain activity. Annu. Rev. Psychol. 63, 483–509.

van Gerven, M.A.J., 2016. A primer on encoding models in sensory neuroscience. J. Math. Psychol. In Press. doi:10.1016/j.jmp.2016.06.009

Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., 2014. Simultaneously Uncovering the Patterns of Brain Regions Involved in Different Story Reading Subprocesses 1–19. doi:10.1371/journal.pone.0112575

Willems, R.M., de Boer, M., de Ruiter, J.P., Noordzij, M.L., Hagoort, P., Toni, I., 2010. A cerebral dissociation between linguistic and communicative abilities in humans. Psychol. Sci. 21, 8–14.

Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., van den Bosch, A., 2016. Prediction During Natural Language Comprehension. Cereb. Cortex (New York, N.Y. 1991) 26, 2506–2516. doi:10.1093/cercor/bhv075

Yamins, D.L.K., Dicarlo, J.J., 2016. Using goal-driven deep learning models to understand sensory cortex. doi:10.1038/nn.4244

Yarkoni, T., Speer, N.K., Balota, D. a, McAvoy, M.P., Zacks, J.M., 2008. Pictures of a thousand words: investigating the neural mechanisms of reading with extremely rapid event-related fMRI. Neuroimage 42, 973–987. doi:10.1016/j.neuroimage.2008.04.258

Zacks, J.M., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., Buckner, R.L., Raichle, M.E., 2001. Human brain activity time-locked to perceptual event boundaries. Nat. Neurosci. 4, 651–655.

Figure captions

Fig. 1. Idealized BOLD curve, sometimes called the hemodynamic response function (HRF). The curve peaks around 6-8 seconds after stimulus onset (stimulus onset is point 0). The shape of the idealized BOLD curve is based on measurement of the actual BOLD curve in the human cortex. The time axis (x-axis) is in TRs, with one TR being 2 seconds. The y-axis expresses signal intensity in arbitrary units.

Fig. 2. Estimated BOLD time courses for different experimental designs. **A)** The estimated BOLD time course for an experiment in which a stimulus is presented every two seconds. Each stimulus takes 500 milliseconds. Because of the slowness of the BOLD response, the BOLD curves to the stimuli quickly start to overlap and the overall response plateaus. This makes it impossible to regress the estimated time course onto the actually measured time courses (the data). **B)** In this scenario the stimuli are presented with an average intertrial interval of 5 seconds (stimulus duration again is 500 ms). Importantly, the ITI is variable: It is sometimes longer or shorter than 5 seconds. This leads to the necessary variation in the estimated BOLD time course (blue). The red line illustrates presentation of the stimuli. **C)** Estimated BOLD time course from an experiment in which participants listened to a continuous spoken narrative. Every word in the narrative was modeled as separate trial. Due to the inherent variation in onsets and durations of the words, the response does not plateau as in Fig. 2A. **D)** Estimated BOLD time course for surprisal values of the words presented in the experiment (same experiment as in 2C). Because words had considerable variation in surprisal values (see text), the estimated BOLD curve has enough variance to be useful in the analysis. This is the approach described by Yarkoni et al. (2008). Examples 2C and 2D are based on the experimental design in Willems et al. (2016).

Fig. 3. Examples of computational approaches in cognitive neuroscience. A) A classifier is used to predict from a sphere of voxels whether subjects are viewing animal or tool stimuli. By consecutively repositioning the sphere an accuracy value a_i is obtained for each voxel *i* across the brain volume. By training the classifier on one modality (pictures, spoken words, written words, natural sounds) and testing it on another modality, accuracy maps are obtained that inform about those brain regions that respond to conceptual information in an amodal manner (after (Simanova et al., 2014)). B) A deep neural network trained for tagging of music fragments is used to transform raw sensory input into auditory features that become more complex in deeper layers of the neural network. Using a representational similarity analysis, it is shown that activity patterns in different brain regions correspond to different deep neural network layers, revealing a representational gradient.

Fig. 1

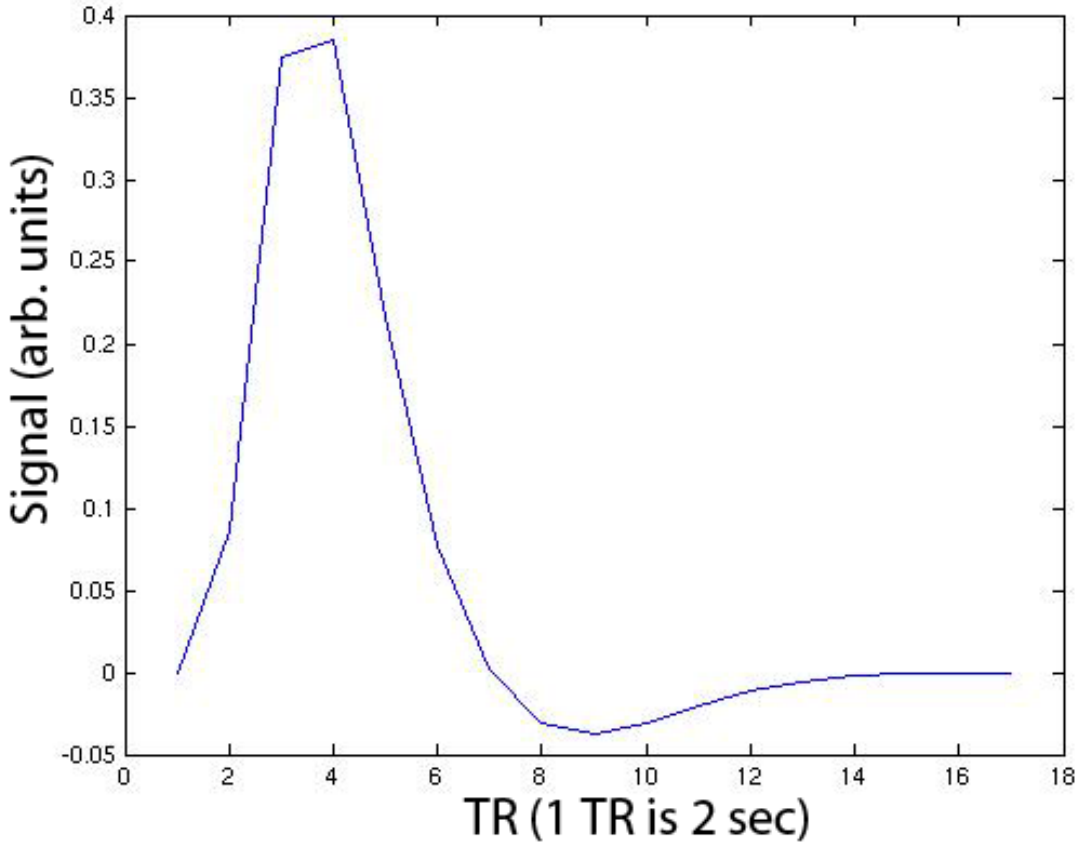

Signal (arb. units)
TR (1 TR is 2 sec)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
0
2
4
6
8
10
12
14
16
18

Fig. 2

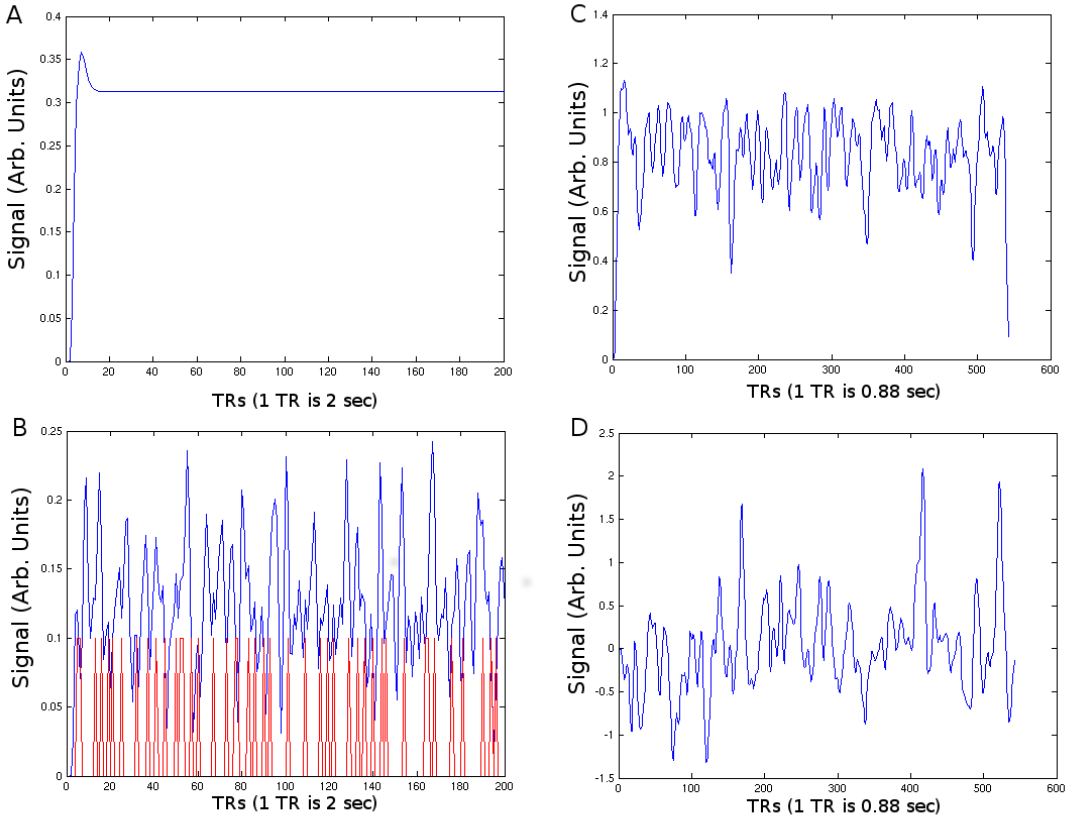
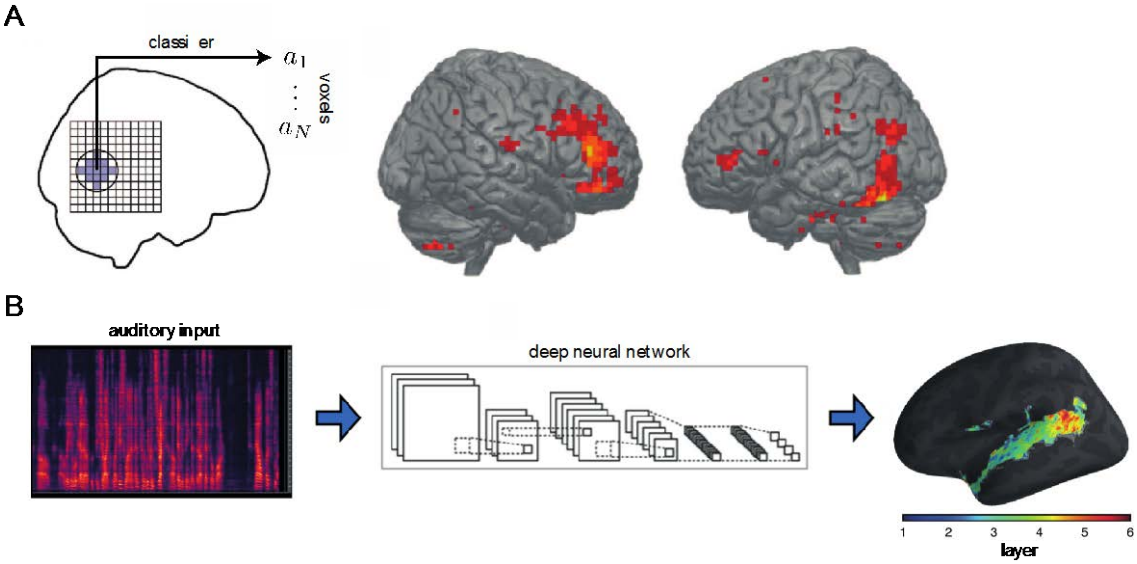

A
Signal (Arb. Units)
TRs (1 TR is 2 sec)
C
Signal (Arb. Units)
TRs (1 TR is 0.88 sec)
B
Signal (Arb. Units)
TRs (1 TR is 2 sec)
D
Signal (Arb. Units)
TRs (1 TR is 0.88 sec)

Fig. 3.


A
classi er
a_1
a_N
voxels
B
auditory input
deep neural network
1
2
3
4
5
6
layer