EJN fens EUROPEAN JOURNAL OF NEUROSCIENCE

European Journal of Neuroscience, Vol. 31, pp. 189–193, 2010 doi:10.1111/j.1460-9568.2009.07055.x

COGNITIVE NEUROSCIENCE

TECHNICAL SPOTLIGHT
Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution

Edmund C. Lalor[1,2] and John J. Foxe[1,3,4]
[1]Trinity College Institute of Neuroscience and
[2]Centre for Bioengineering, Trinity College Dublin, Dublin 2, Ireland
[3]The Cognitive Neurophysiology Laboratory, Children's Evaluation and Rehabilitation Center, Departments of Pediatrics and Neuroscience, Albert Einstein College of Medicine, 1410 Pelham Parkway South, Bronx, New York 10461, USA
[4]Cognitive Neurophysiology Laboratory, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA

Keywords: auditory evoked potential, EEG, impulse response, speech

Abstract

The human auditory system has evolved to efficiently process individual streams of speech. However, obtaining temporally detailed responses to distinct continuous natural speech streams has hitherto been impracticable using standard neurophysiological techniques. Here a method is described which provides for the estimation of a temporally precise electrophysiological response to uninterrupted natural speech. We have termed this response AESPA (Auditory Evoked Spread Spectrum Analysis) and it represents an estimate of the impulse response of the auditory system. It is obtained by assuming that the recorded electrophysiological function represents a convolution of the amplitude envelope of a continuous speech stream with the to-be-estimated impulse response. We present examples of these responses using both scalp and intracranially recorded human EEG, which were obtained while subjects listened to a binaurally presented recording of a male speaker reading naturally from a classic work of fiction. This method expands the arsenal of stimulation types that can now be effectively used to derive auditory evoked responses and allows for the use of considerably more ecologically valid stimulation parameters. Some implications for future research efforts are presented.

Introduction

The event-related potential (ERP) technique has been central to investigations of both normal and clinical processing in the auditory system of humans for the past several decades (e.g. Burkard *et al.*, 2007; Leavitt *et al.*, 2007). Its major advantage is the superb temporal resolution of the auditory evoked potential (AEP), which facilitates the precise measurement of the timing of sensory and cognitive processing of auditory stimuli (Picton *et al.*, 1974). The AEP is typically obtained by averaging responses to numerous, repeated presentations of temporally discrete stimuli. Such a stimulus sequence can be considered a train of impulses and, thus, the AEP can be considered an impulse response of the auditory system. The averaging process often necessitates the presentation of stimuli with interstimulus intervals in excess of several hundred milliseconds to avoid overlapping responses. This procedure is inefficient and, perhaps more importantly, severely constrains environmental validity considering that the human auditory system has evolved to process stimuli with an enormously broad range of temporal characteristics.

Correspondence: Dr Edmund C. Lalor, [1]Trinity College Institute of Neuroscience, as above. Dr John J. Foxe, [3]The Cognitive Neurophysiology Laboratory, as above.
E-mail: edlalor@tcd.ie or foxe@nki.rfmh.org

Received 21 May 2009, accepted 11 November 2009

A method that circumvents these constraints, allowing acquisition of temporally detailed electrophysiological responses to continuous auditory stimuli, was recently described by Lalor *et al.* (2009). It has been termed the AESPA technique (for Auditory Evoked Spread Spectrum Analysis) and is adapted from the visual-based VESPA (Visual Evoked Spread Spectrum Analysis) method (Lalor *et al.*, 2006). The estimation of the AESPA response is based on the assumption that the neural data recorded while a subject listens to a continuous auditory stimulus with varying intensity consists of a convolution of that varying intensity with an unknown impulse response plus noise. The impulse response, i.e. the AESPA, is then determined by performing least-squares estimation using a sliding window of auditory amplitude values and the measured neural data.

Given their importance to human behaviour, the ability to assess the electrophysiological processing of continuous streams of speech represents a particularly important challenge. The extremely complex temporal statistics of such streams has placed a stranglehold on the design of paradigms for examining the physiological timing of speech using the AEP method, with researchers restricted to averaging EEG epochs time-locked to very short speech segments (Wood *et al.*, 1971; Saint-Amour *et al.*, 2007; Aiken & Picton, 2008a), 'speech-like' stimuli (Sanders *et al.*, 2002), individual words from speech streams (Kutas & Hillyard, 1984; Hagoort *et al.*, 2004), or single sentences

(Steinhauer *et al.*, 1999). Averaged responses to such stimuli, treated as discrete impulses, ignore the statistics of the presented audio signal and preclude the short-time measurement of continuous speech processing dynamics.

Here we describe an extension to the AESPA method that facilitates the measurement of temporally detailed electrophysiological responses from subjects as they listen to natural, continuous speech. The extension involves obtaining an AESPA response by utilizing the amplitude envelope of a natural, continuous speech stream. The experimental setup and steps involved in estimating the AESPA are illustrated in Fig. 1 and are further described in the Appendix.

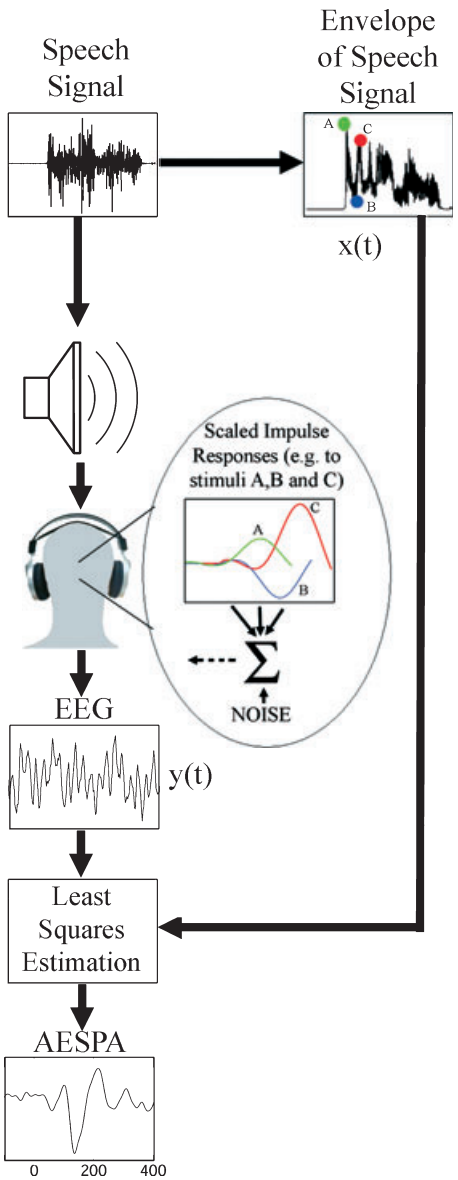

FIG. 1. Method for generating the AESPA response to natural, continuous speech. The speech signal is presented to the subject concurrent to electrophysiological (e.g. EEG) data being recorded. The amplitude envelope of the speech signal is calculated by determining the RMS of the audio signal values occurring in the time frame of each sample of the neural data. It is assumed that the neural data consists of a convolution of the amplitude envelope of the speech signal with an unknown impulse response function, plus noise. Given the recorded data, knowledge of the speech signal and accurate synchronization between the two, this impulse response function, known as the AESPA, can be estimated using least-squares estimation.

Responses to natural speech in three subjects with scalp recordings and one subject with intracranially recorded data

Figure 2a shows the EEG voltage timecourse of the AESPA response to the speech streams averaged over the three participating subjects at the medial–frontal location Fz, re-referenced to the average of the mastoids. Clearly discriminable components are evident with positive peaks at ~40 and 170 ms and a prominent negativity at ~80 ms. In order to get a better sense of the source of the response, current-source density (CSD) maps of the averaged EEG responses were also computed and plotted (Fig. 2b). Each of the CSD waveforms shown was averaged over three virtual electrode locations (FT7, FC5 and C5, upper left; FT8, FC6 and C6, upper right; T9, TP7 and TP9, lower left; and T10, TP8 and TP10, lower right) and the locations indicated on the maps were at the centroids of the triangles formed by these triplets of virtual electrode locations. Clear foci of current flow are seen over superior temporal areas, suggesting that the source of the AESPA is in early auditory cortex. The facts that CSD waveforms from these two foci are inverted in polarity relative to each other but the temporal progression of their respective components is so similar suggests that the underlying current flow is tangential to the cortical surface, implying that the source of the neural activity is a single source located in a fissure; possibly in Heschl's gyrus (Yvert *et al.*, 2005).

Figure 3 shows the timecourse of the AESPA at five intracranial electrodes and the location of those electrodes in the right temporal cortex of the participating patient. The Talairach coordinates of the electrode locations were determined by visual inspection of an MRI scan of the patient taken while the electrode grid was in place. For the purposes of illustration, and owing to imaging artifacts caused by the electrode grid, these coordinates are highlighted on a 3-D reconstruction of the mirror image of the patient's left hemisphere. The responses have a similar morphology to those found in the EEG data. A reversal in polarity is apparent when comparing electrodes on opposite banks of the lateral sulcus, mirroring the inversion of the CSD waveforms shown in Fig. 2, providing further evidence of the early cortical origin of the AESPA response. The standard AEP obtained by averaging responses to repeated presentations of a 100-ms (including 5 ms rise–fall), 2 kHz tone stimulus showed a similar inversion for the early components. For a more detailed discussion of the relationship between the AEP and the AESPA, see Lalor *et al.* (2009).

Discussion

The ability to obtain AESPA responses to completely natural, continuous speech with high temporal resolution has clear and potent utility. This is true, not just in terms of gaining insight into the physiological processing of this most important class of auditory signal, but also for investigating the many neurological disorders in which speech processing is compromised. For example, studies of neurophysiological responses to speech in aphasia, autism, schizophrenia and coma patients could all be enhanced through use of natural speech rather than brief, 'speech-like' stimuli. Improved environmental validity would not only render the results of speech experiments more compelling but it would also almost certainly enhance subject engagement in auditory research paradigms.

Another area in which the AESPA response could yield important new results is in research on the processing of multiple concurrent audio streams. The aforementioned stimulation limitations that have applied to paradigms for examining the electrophysiological timing of speech processing have also been very restrictive in terms of auditory streams in general, particularly where more than one stream is presented at the same time. This stems from the difficulty in resolving

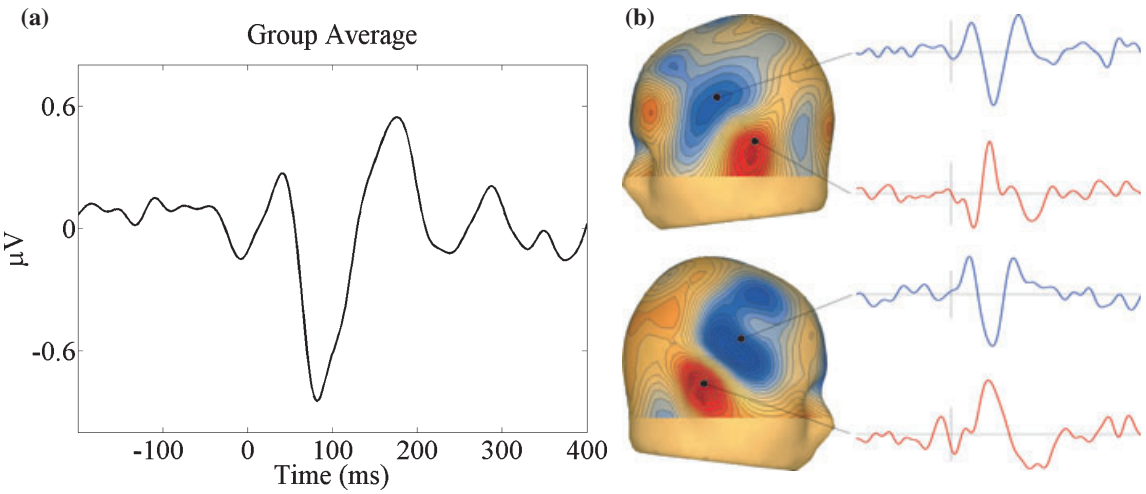

FIG. 2. (a) AESPA responses at electrode location Fz (frontocentral) averaged over three EEG subjects who listened to 143, 140 and 48 min of speech divided into 47, 46 and 16 blocks respectively. (b) CSD maps of the group average AESPA at 85 ms, indicating that this component originates in auditory cortex. The CSD waveforms shown were averaged over three virtual electrode locations (FT7, FC5 and C5, upper left; FT8, FC6 and C6, upper right; T9, TP7 and TP9, lower left; and T10, TP8 and TP10, lower right). The locations indicated on the maps were at the centroids of the triangles formed by these triplets of virtual electrode locations. Vertical axes indicate ±20 $\mu V/cm^2$.

AEP responses to more than one simultaneously presented discrete stimulus. Researchers investigating auditory scene analysis have been forced to go to great lengths to create the perception of auditory streaming using discrete stimuli (e.g. Sussman *et al.*, 1999; Ritter *et al.*, 2006; De Sanctis *et al.*, 2008) and those studying auditory attention to one of several simultaneously presented stimuli have had to sacrifice the AEP's excellent temporal resolution by using rapid, sinusoidal modulations in order to generate the auditory steady-state response (Bidet-Caulet *et al.*, 2007). Another method that has met with some success in this regard is that of using auditory stimuli controlled by binary maximum length sequences (m-sequences; Picton *et al.*, 1992; Lasky *et al.*, 1993). This method takes advantage of the fact that one can obtain an AEP by cross-correlating the recorded neurophysiological data with the input m-sequence. This is possible because m-sequences have the property of being uncorrelated with shifted versions of themselves (Lee & Schetzen, 1965). This is not true of speech streams or of naturally occurring auditory streams in general. The AESPA method does not employ cross-correlation and is thus not restricted to using such artificially created stimulus sequences. Furthermore, obtaining separate AESPA responses using simultaneous audio streams is straightforward provided those streams are not correlated with each other.

Accordingly, the AESPA method seems ideally suited to research in the fields of auditory scene analysis and auditory attention, with research into the effect of attention to one of several concurrently presented speech streams being particularly apposite. This crucially important situation is one that occurs on an everyday basis in real life and which humans have evolved to manage with ease. Despite this, since the early behavioural investigations (Cherry, 1953; Rand, 1974; Cutting, 1976), neurophysiological research has struggled to make inroads, with only a small number of AEP-based studies seeking to address the question (Hink & Hillyard, 1976). Confidence in the utility of the AESPA for research on attention can be garnered from a recent report demonstrating the effect of directing attention to one of two simultaneously presented visual stimuli using the visual analog of the AESPA, the VESPA (Lalor *et al.*, 2007).

It should be noted that the AESPA, as described in the current study, is derived using only the envelope of the presented speech stream. This is clearly a limitation of the technique as it is currently implemented, as this ignores the fine temporal structure of the stimulus signal.

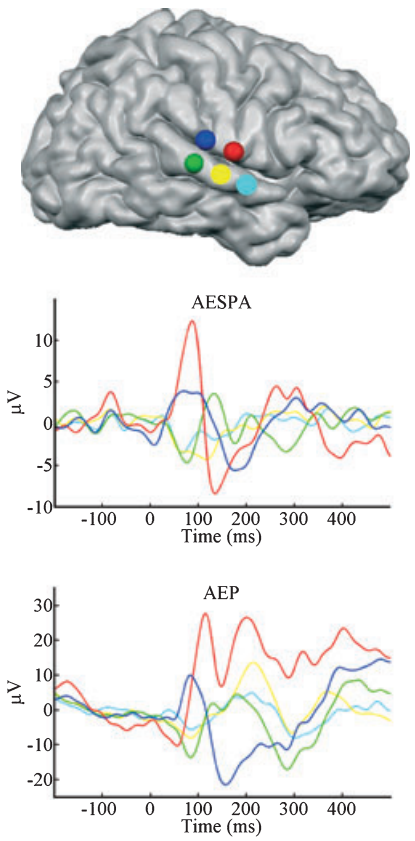

FIG. 3. AESPAs obtained using an intracranial electrode grid placed over the right temporal lobe of one subject who listened to < 11 min of speech. Also plotted are the electrode locations (Talairach coordinates: Red: x = 64 mm, y = −8 mm, z = 11 mm; Blue: x = 63 mm, y = −20 mm, z = 16 mm; Green: x = 64 mm, y = −25 mm, z = 6 mm; Yellow and Cyan could not be accurately located using visual inspection of the images). A reversal in the polarity of the AESPA and the AEP can be seen for electrodes on either side of the lateral sulcus.

Nonetheless, it has been shown that it is the envelope which is of overriding importance when it comes to speech recognition (Shannon *et al.*, 1995; Smith *et al.*, 2002), with reports suggesting that envelope frequencies between 4 and 16 Hz contribute most to speech intelligibility (Aiken & Picton, 2008b). Recognition is enhanced considerably when envelopes corresponding to sounds filtered into several complementary frequency bands are combined (Shannon *et al.*, 1995; Smith *et al.*, 2002). The current method could potentially be extended by adjusting the bandwidth of the filter used to determine the envelope and deriving AESPA responses corresponding to each filter setting in parallel, although it must be noted that the number of usable filters is limited by the frequency content of the acquired neural signal.

Because of its importance, a number of previous EEG studies have focused on the speech envelope. Many of these differ from the method presented here in that they are focused on following responses to specific single frequencies (e.g. Purcell *et al.*, 2004; Aiken & Picton, 2008a) rather than the full broadband envelope of natural speech. Aiken & Picton (2008b) implemented a similar model to that proposed here; however, rather than estimating the impulse response of the auditory system using EEG and the speech envelope, they assumed an impulse response that consisted of averaged EEG responses to speech onset. Using repeated presentation of sentences, they then compared the result of the linear convolution of the speech envelope and the impulse response to the averaged EEG response to the sentence. Their results indicate a high degree of correlation, which emphasizes the validity of using EEG to measure responses to the envelope of speech; however, their assumption that the auditory system's linear impulse response can be faithfully represented by the averaged AEP has been shown to be inaccurate (Lalor *et al.*, 2009).

No hemispheric differences are apparent in CSD maps from the three EEG subjects, despite the well-known bias of language processing to the left hemisphere. This is not hugely surprising as the AESPA responses here have been estimated using a linear assumption between audio signal intensity and measured EEG. Thus, the acquired AESPA responses probably reflect largely activity of those cells whose outputs change somewhat linearly with respect to the intensity of the audio signal, with cells involved in higher level processing contributing to a lesser degree. While the assumption of linearity is clearly sufficient to gain a useful sensory response (Fig. 2), it is undoubtedly an oversimplification. However, by simply including products of sound signal amplitudes at different time lags as inputs, the model can easily be extended to higher orders (Lalor *et al.*, 2008). Data were recorded only from the right temporal lobe of the patient who participated in the intracranial recording, so an interhemispheric comparison could not be made.

In summary, temporally detailed evoked responses to natural, continuous speech have been obtained using a novel method known as the AESPA. The flexibility of this new technique gives it a number of advantages over the traditional AEP method and suggests its applicability to fundamental research on speech processing as well as studies of attention, auditory scene analysis and a number of neurological disorders.

Acknowledgements

This work was supported by US National Science Foundation grant BCS0642584 to J.J.F. and a Government of Ireland Postdoctoral Research Fellowship to E.C.L. Sincere appreciation to Dr Theodore Schwartz for patient surgery, care and electrode placement, to Dr Jonathan Dyke for his assistance with the radiology, to Emma-Jane Forde and Manuel Gomez-Ramirez for help with data collection and to Dr Pejman Sehatpour for assistance with the MRI 3-D reconstruction. Thanks to Professor Barak A. Pearlmutter for useful discussions. We are ever grateful to the participating patients at New York Presbyterian Hospital who donated their time and energy with dignity and grace to this project at what could only be described as a very challenging time for them.

Abbreviations

AEP, auditory evoked potential; AESPA, auditory evoked spread spectrum analysis; CSD, current-source density; ERP, event-related potential; RMS, root mean square.

References

Aiken, S.J. & Picton, T.W. (2008a) Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.*, **245**, 35–47.

Aiken, S.J. & Picton, T.W. (2008b) Human cortical responses to the speech envelope. *Ear Hear.*, **29**, 139–157.

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.E., Giard, M.H. & Bertrand, O. (2007) Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J. Neurosci.*, **27**, 9252–9261.

Burkard, R.F., Don, M. & Eggermont, J.J. (2007) *Auditory Evoked Potentials: Basic Principles and Clinical Application*. Lippincott, Williams and Wilkins, Philadelphia, PA.

Cherry, E.C. (1953) Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.*, **25**, 975–979.

Cutting, J.E. (1976) Auditory and linguistic processes in speech perception: inferences from six fusions in dichotic listening. *Psychol. Rev.*, **83**, 114–140.

De Sanctis, P., Ritter, W., Molholm, S., Kelly, S.P. & Foxe, J.J. (2008) Auditory scene analysis: the interaction of stimulation rate and frequency separation on pre-attentive grouping. *Eur. J. Neurosci.*, **27**, 1271–1276.

Hagoort, P., Hald, L., Bastiaansen, M. & Petersson, K.M. (2004) Integration of word meaning and world knowledge in language comprehension. *Science*, **304**, 438–441.

Hink, R.F. & Hillyard, S.A. (1976) Auditory evoked potentials during selective listening to dichotic speech messages. *Percept. Psychophys.*, **20**, 236–242.

Kutas, M. & Hillyard, S.A. (1984) Brain potentials during reading reflect word expectancy and semantic association. *Nature*, **307**, 161–163.

Lalor, E.C., Pearlmutter, B.A., Reilly, R.B., McDarby, G. & Foxe, J.J. (2006) The VESPA: a method for the rapid estimation of a visual evoked potential. *Neuroimage*, **32**, 1549–1561.

Lalor, E.C., Kelly, S.P., Pearlmutter, B.A., Reilly, R.B. & Foxe, J.J. (2007) Isolating endogenous visuo-spatial attentional effects using the novel visual-evoked spread spectrum analysis (VESPA) technique. *Eur. J. Neurosci.*, **26**, 3536–3542.

Lalor, E.C., Yeap, S., Reilly, R.B., Pearlmutter, B.A. & Foxe, J.J. (2008) Dissecting the cellular contributions to early visual sensory processing deficits in schizophrenia using the VESPA evoked response. *Schizophr. Res.*, **98**, 256–264.

Lalor, E.C., Power, A.J., Reilly, R.B. & Foxe, J.J. (2009) Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J. Neurophysiol.*, **102**, 349–359.

Lasky, R.E., Shi, Y. & Hecox, K.E. (1993) Temporal masking of human auditory evoked brain stem responses using two simultaneously presented maximum length sequences. *Ear Hear.*, **14**, 183–189.

Leavitt, V.M., Molholm, S., Ritter, W., Shpaner, M. & Foxe, J.J. (2007) Auditory processing in schizophrenia during the middle latency period (10-50 ms): high-density electrical mapping and source analysis reveal subcortical antecedents to early cortical deficits. *J. Psychiatry Neurosci.*, **32**, 339–353.

Lee, Y.W. & Schetzen, M. (1965) Measurement of the Wiener kernels of a non-linear system by cross-correlation. *Int. J. Control*, **2**, 237–254.

Picton, T.W., Hillyard, S.A., Krausz, H.I. & Galambos, R. (1974) Human auditory evoked potentials. I. Evaluation of components. *Electroencephalogr. Clin. Neurophysiol.*, **36**, 179–190.

Picton, T.W., Champagne, S.C. & Kellett, A.J. (1992) Human auditory evoked potentials recorded using maximum length sequences. *Electroencephalogr. Clin. Neurophysiol.*, **84**, 90–100. (Erratum. *Electroencephalogr. Clin. Neurophysiol.*, **84**, 300. (1992)).

Purcell, D.W., John, S.M., Schneider, B.A. & Picton, T.W. (2004) Human temporal auditory acuity as assessed by envelope following responses. *J. Acoust. Soc. Am.*, **116**, 3581–3593.

Rand, T.C. (1974) Letter: dichotic release from masking for speech. *J. Acoust. Soc. Am.*, **55**, 678–680.

Ritter, W., De Sanctis, P., Molholm, S., Javitt, D.C. & Foxe, J.J. (2006) Preattentively grouped tones do not elicit MMN with respect to each other. *Psychophysiology*, **43**, 423–430.

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W. & Foxe, J.J. (2007) Seeing voices: high-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, **45**, 587–597.
Sanders, L.D., Newport, E.L. & Neville, H.J. (2002) Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nat. Neurosci.*, **5**, 700–703.
Shannon, R.V., Fan-Gang, Z., Kamath, V., Wygonski, J. & Ekelid, M. (1995) Speech recognition with primarily temporal cues. *Science*, **207**, 303–304.
Smith, Z.M., Delgutte, B. & Oxenham, A.J. (2002) Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, **416**, 87–90.

Appendix: data recording and analysis

Continuous EEG was recorded from three subjects (two right-handed males and one left-handed female) aged 20–38 years, all with normal hearing, as they listened to an unaltered recording of a male voice reading from a fiction novel. Audio segments averaged 181 s in duration and the three subjects listened to 47, 46 and 16 segments respectively. Triggers indicating the start and end of each audio segment were recorded along with the EEG. The EEG data were recorded from 168 electrodes, filtered over the range 0–134 Hz and digitized at a rate of 512 Hz using the BioSemi Active Two system (BioSemi). Subsequently, the EEG was digitally filtered with a high-pass filter with passband > 2 Hz and −60 dB response at 1 Hz, and a low-pass filter with 0–35 Hz passband and −50 dB response at 45 Hz, and re-referenced to the average of the data recorded from the mastoids. Each subject provided informed written consent and the experiment was undertaken in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of the Nathan Kline Institute. The audio stream was presented binaurally using an NVIDIA GeForce4 Ti4200 sound card (NVIDIA) and Sennheiser HD600 headphones (Sennheiser).

The intracranial data were recorded from a patient who suffered from intractable epilepsy which pharmacological approaches had proven insufficient to control. The patient (right-handed female, age 39 years) listened to four segments totalling 652 s of the same audio stimulus while data were recorded from 100 intracranial electrodes placed subdurally over the right temporal lobe. Written informed consent was provided by the patient and the procedures were approved by the Institutional Review Board of Weill Cornell Medical College. The patient was on anti-seizure medications consisting of a combination of levetiracetam (keppra) and zonisamide (zonegran) at the time of recording. The audio stream was presented binaurally using the PCMCIA Sound Blaster® Audigy 2 ZS Notebook sounds card and Sennheiser HD600 headphones, and triggers indicating the start and end of each audio segment were recorded along with the data. The data were recorded using BrainAmp™ (Brain Products GmbH, München, Germany) amplifiers, bandpass-filtered online from 0.01 to 250 Hz and digitized at a rate of 1000 Hz. Subsequently, the data were digitally filtered with a high-pass filter with passband > 2 Hz and −60 dB response at 1 Hz, and a low-pass filter with 0–35 Hz passband and −50 dB response at 45 Hz.

In order to estimate the AESPA it was necessary to determine the amplitude envelope of the speech signal. Given that the sampling rate of the audio signal was 44 100 Hz and that of the EEG was 512 Hz, the root mean square (RMS) of an average of 86.133 audio signal samples, ranging from 0 to 1, was calculated for each sample of EEG. Similarly, for the intracranial data, which was sampled at 1000 Hz, the RMS of an average of 44.1 audio signal samples was calculated for each data sample. This represents a simple resampling of the audio signal.

To perform the least-squares fit we form the n-dimensional column vector x_t consisting of the sample values of the resampled audio signal

$$(x(t-t_0), x(t-(t_0+1)), \ldots, x(t-(t_0+n-1)))^T, \quad (1)$$

where n is the number of sampled points of the impulse response function, w, that we wish to estimate and t_0 is the estimation window offset. The values for t_0 and n used in this study, for example, were −198 and 598 ms respectively. We can then estimate the n-dimensional vector, w, consisting of the sampled points of the response function.

$$(w(t_0), w(t_0+1), \ldots, w(t_0+n-1))^T, \quad (2)$$

by minimizing,

$$E = \left\langle ||w^T x_t - y_t||^2 \right\rangle = w^T \langle x_t x_t^T \rangle w - 2w^T \langle x_t y_t \rangle + \langle y_t y_t \rangle,$$

where $\langle \ldots \rangle$ indicates an average over t.

Expanding $dE/dw = 0$ gives

$$w = \left\langle x_t x_t^T \right\rangle^{-1} \langle x_t y_t \rangle. \quad (3)$$

w can be solved for straightforwardly by first constructing the $n \times n$ matrix $x_t x_t^T$ at each sample point of our resampled audio signal, determining a running sum across all time points and dividing the sum by the number of time points. Second, the n-dimensional vector $x_t y_t$ is calculated, again at each time point, and the mean is determined across all time points, again using a running sum. The final step involves a simple matrix inversion and multiplication.

We can further improve the quality of our estimate by adding a regularization term. This serves to increase the bias but reduce the variance of the estimate, resulting in a net reduction in estimation error. Adding a term that quadratically penalizes the difference between each two neighboring terms of w we obtain the equation

$$w, = \left\langle x_t x_t^T + \lambda M \right\rangle^{-1} \langle x_t y_t \rangle, \text{ where } M = \begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \cdots & \cdots & \cdots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix}. \quad (4)$$

An empirically determined value of $\lambda = 4.4 \times 10^{-3}$resultsin reduced estimation error without overly penalizing the heights of actual components. See Lalor *et al.* (2006) for further details.

The sliding window of −198 to 598 ms was chosen in order to display a 'pre-stimulus' baseline period and to allow the AESPA to return to baseline. It is important to note that because the AESPA represents a function that produces an estimate of the EEG output based on a broad window of input stimulus values, the meaning of the interval is slightly different from that used with standard AEPs, which represents a response to an assumed ideal impulse.

Unlike the AEP, the AESPA does not correspond to a specific discrete event occurring at time 0. Instead, each time-point on the time axis can be interpreted as being the relative time between the EEG and the input intensity signal. Therefore, the AESPA at −100 ms, for example, indexes the contribution of the input intensity signal to the EEG 100 ms earlier; obviously this should be zero, i.e. our 'pre-stimulus' baseline period. As another example, the AESPA at +100 ms indexes how the contribution of the input intensity signal to the EEG 100 ms later. Unsurprisingly, from what we know from human electrophysiology, this is unlike to be zero.

The CSD signals and maps were calculated using BESA (MEGIS Software GmbH). The signals represent the volume current flow out of the brain through the skull into the skin and is determined using the surface Laplacian operator, i.e. by computing the second spatial derivative of the voltage distribution in tissue.

The CSD montage shows a large signal if the cortical surface (convexity) is active. This corresponds to a radial superficial dipole. If the brain activity is in a fissure (tangential dipole) the CSD signal is much weaker. The current flows in a tangential direction, resulting in two peaks of opposite polarity in the CSD maps, corresponding to where the volume current is leaving (+) and entering (−) the skull. The two zones of maximum current are closer together than in voltage maps because voltage is an integral over current source density. In determining the CSD signals, BESA translates the 168-channel voltage data to 81 virtual electrode locations according to the standardized 10-10 system.

Steinhauer, K., Alter, K. & Friederici, A.D. (1999) Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.*, **2**, 191–196.
Sussman, E., Ritter, W. & Vaughan, H.G. Jr (1999) An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, **36**, 22–34.
Wood, C.C., Goff, W.R. & Day, R.S. (1971) Auditory evoked potentials during speech perception. *Science*, **173**, 1248–1251.
Yvert, B., Fischer, C., Bertrand, O. & Pernier, J. (2005) Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. *Neuroimage*, **28**, 140–153.