Systems/Circuits

Evidence for Neural Computations of Temporal Coherence in an Auditory Scene and Their Enhancement during Active Listening

James A. O'Sullivan,[1] Shihab A. Shamma,[2] and Edmund C. Lalor[1]

[1]School of Engineering, Trinity Centre for Bioengineering and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin 2, Ireland, and [2]Institute for Systems Research, University of Maryland, College Park, Maryland 20742

The human brain has evolved to operate effectively in highly complex acoustic environments, segregating multiple sound sources into perceptually distinct auditory objects. A recent theory seeks to explain this ability by arguing that stream segregation occurs primarily due to the temporal coherence of the neural populations that encode the various features of an individual acoustic source. This theory has received support from both psychoacoustic and functional magnetic resonance imaging (fMRI) studies that use stimuli which model complex acoustic environments. Termed stochastic figure–ground (SFG) stimuli, they are composed of a "figure" and background that overlap in spectrotemporal space, such that the only way to segregate the figure is by computing the coherence of its frequency components over time. Here, we extend these psychoacoustic and fMRI findings by using the greater temporal resolution of electroencephalography to investigate the neural computation of temporal coherence. We present subjects with modified SFG stimuli wherein the temporal coherence of the figure is modulated stochastically over time, which allows us to use linear regression methods to extract a signature of the neural processing of this temporal coherence. We do this under both active and passive listening conditions. Our findings show an early effect of coherence during passive listening, lasting from ~115 to 185 ms post-stimulus. When subjects are actively listening to the stimuli, these responses are larger and last longer, up to ~265 ms. These findings provide evidence for early and preattentive neural computations of temporal coherence that are enhanced by active analysis of an auditory scene.

Key words: **auditory scene analysis; denoising source separation (DSS); electroencephalography (EEG); stream segregation; temporal coherence; temporal response function (TRF)**

Introduction

In noisy and complex acoustic environments, humans and other animals can effortlessly segregate the multitude of interfering acoustic sources into perceptually distinct auditory objects (Bregman, 1990). They can also selectively attend to a particular auditory object and track it over time (Sussman et al., 2007). One theory that seeks to explain this ability, called the temporal coherence model of stream segregation, posits that responses of neural populations encoding various features of a sound source (e.g., frequency components, pitch, spatial location) tend to be temporally correlated (or coherent), but are uncorrelated with responses to features of other sources. Consequently, by grouping the coherent features together, the auditory system can segregate out one source from other (mixed but uncorrelated) sources. But to do so, there must be a (coincidence) mechanism that can detect the coherence of different channels and exploit them when they occur (Elhilali et al., 2009a; Shamma and Micheyl, 2010; Shamma et al., 2011, 2013; Krishnan et al., 2014; Wolf et al., 2014). The search for evidence of such a coincidence-measuring mechanism is the objective of this study.

A novel stimulus, known as a stochastic figure–ground (SFG) stimulus, was developed recently to explore this phenomenon (Teki et al., 2011, 2013). It consists of a sequence of inharmonic chords, each containing several pure tones, which are randomly selected from a predefined set. If a subset of these tones repeats or changes slowly in frequency over several consecutive chords, a spontaneous percept of a "figure" popping out of a random background of varying tones can emerge over time. The saliency of the figure increases as the number of temporally coherent tones increases.

In a functional magnetic resonance imaging (fMRI) experiment, Teki et al. (2011) used this stimulus to investigate preattentive stream segregation mechanisms, and showed significant activation in the intraparietal sulcus and the superior temporal sulcus. While these are important findings, alternative ap-

Received Dec. 7, 2014; revised March 10, 2015; accepted March 31, 2015.

Author contributions: J.A.O., S.A.S., and E.C.L. designed research; J.A.O. performed research; J.A.O. analyzed data; J.A.O., S.A.S., and E.C.L. wrote the paper.

This work was supported by a grant from Science Foundation Ireland (09-RFP-NES2382) and by an Irish Research Council Starter Research Project Grant (RPG2013-1). We thank R. O'Connell, G. Loughnane, M. Crosse, and G. Di Liberto for useful discussions on the experimental paradigm, and for comments on the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Edmund C. Lalor, PhD, Trinity Centre for Bioengineering, Trinity Biomedical Sciences Institute, Trinity College Dublin, 152-160 Pearse Street, Dublin 2, Ireland. E-mail: edlalor@tcd.ie.

DOI:10.1523/JNEUROSCI.4973-14.2015

 0270-6474/15/357256-08$15.00/0

proaches using techniques such as magnetoencephalography (MEG) and electroencephalography (EEG) promise further insights given their far superior temporal resolution. These methods allow us to examine the timing of what is an inherently dynamic neural process. For example, knowledge of the latency at which temporal coherence computations occur can inform theories about the neural mechanisms involved. Thus, such data could indicate whether these computations represent early, obligatory sensory responses, or alternatively whether they represent longer latency responses that depend on higher-order cognitive engagement with the stimuli. In addition, these methods can give us insight into the build-up rate of stream segregation within this SFG paradigm, which remains as yet unknown.

However, these approaches present challenges of their own in terms of disentangling temporally overlapping responses from multiple simultaneously active neural sources (Luck, 2005). Recently, the application of linear regression methods has addressed this shortcoming somewhat by facilitating the extraction of localized neural correlates of individual features of interest (Lalor and Foxe, 2010; Ding and Simon, 2012; Zion Golumbic et al., 2013; Gonçalves et al., 2014). Here, using such an approach in an EEG experiment, we show that it is possible to extract a neural signature of temporal coherence computations. Furthermore, we explore the effects of actively listening to such complex stimuli, as opposed to just passive listening as was done in the aforementioned fMRI study (Teki et al., 2011). We wondered whether active listening alters the nature of auditory processing beyond enhancing the early obligatory computations that are already occurring. In addition, how might the latency of such computations be affected by active listening? Here, we begin to answer these questions by incorporating active and passive listening conditions into our experimental paradigm.

Materials and Methods

Subjects

Ten healthy subjects (four female) between 23 and 32 years of age participated in the EEG experiment, and eight healthy subjects (one female) between 25 and 27 years of age participated in the psychoacoustic experiment (five subjects participated in both experiments). The experiment was undertaken in accordance with the Declaration of Helsinki. The Ethics Committees of the School of Health Sciences at Trinity College Dublin approved the experimental procedures and each subject provided written informed consent. Subjects reported no history of hearing impairment or neurological disorder.

Stimulus creation and delivery

We modified a recently introduced SFG stimulus that has been used successfully in both psychoacoustic and fMRI research (Teki et al., 2011, 2013). This stimulus aims to model naturally occurring complex acoustic scenes characterized by a figure and background that overlap in spectrotemporal space, and that are only distinguishable by their temporal fluctuation statistics (Fig. 1*A*). It consists of a sequence of pure tones, each 50 ms in duration, the onset and offset of which are shaped by a 10 ms raised-cosine ramp, and with a 0 ms interval between each tone. The tones are selected from a set of 128 frequencies equally spaced on a logarithmic scale between 250 and 8000 Hz, such that the separation between each frequency is approximately half a semitone (~1/24th of an octave). A chord is defined as the sum of multiple pure tones. Unlike in Teki et al. (2011), the number of tones per chord here remained unchanged at 15 tones throughout the duration of the stimulus so as to keep the broadband power and all other low-level features of the stimulus constant.

As mentioned previously, when a subset of tones repeats or changes slowly in frequency over several consecutive chords, these tones become temporally coherent with each other. This subset can therefore be perceived as a figure against the background of the remaining tones. In our experiment, the "coherence level" (CL) is defined as the number of tones that are temporally coherent from chord to chord. Relative to Teki et al. (2011), we altered the manner in which the coherent tones progressed throughout the stimulus. The figure in the aforementioned study consisted of tones that repeated in frequency for the duration of the figure. However, to make the stimuli more naturalistic, we allowed the temporally coherent tones to fluctuate randomly in frequency for the duration of the figure, but by no more than two semitones from one chord to the next. Specifically, at the start of each figure, 15 tones are randomly selected from the 128 possible frequencies to be active for the first chord. As mentioned above, a proportion of these (depending on the CL) are then randomly selected to be "temporally coherent" for the duration of the figure. On the next chord, all coherent tones change in frequency in either an upwards or downwards direction, randomly chosen to be between one and four frequency bands (~0.5–2 semitones). This pattern continues for the duration of the figure, such that all coherent tones move in a coordinated fashion (Fig. 1*A*, red). It is important to clarify that the coherent tones were only bounded between ±2 semitones between adjacent chords, and were therefore free to drift substantially in frequency over several consecutive chords.

All stimuli were created off-line using Matlab 2014 software (Mathworks) at a sampling frequency of 44.1 kHz and 16 bit resolution. All stimuli were filtered using the HUTear Matlab toolbox (Härmä and Palomäki, 2000) to simulate the frequency response of the outer and middle ear. This was done so that all frequencies were perceived with approximately the same loudness. Sounds were delivered diotically using Sennheiser HD650 headphones at a level of 60 dB SPL (the same for all participants). Stimuli were presented using Presentation software from Neurobehavioral Systems (http://www.neurobs.com).

Experimental design

Psychoacoustic experiment. A psychoacoustic evaluation of the effects of coherence was performed in Teki et al. (2011). An approximately linear relationship was found between the CL and figure detection, with detection of the figure approaching ceiling at a CL of 8. However, because of the more complex nature of our stimuli, we also conducted a psychoacoustic experiment to determine the effects of CL in our stimuli.

For the experiment, all stimuli lasted for 3 s, wherein the first and last seconds always had a CL of 0. The middle segment was randomly assigned to have a CL between 0 and 10, and in steps of two (0, 2, 4, 6, 8, 10). The subject's task was to determine whether or not this segment contained a figure (i.e., if it had a CL of ≥2).

Before the experiment, subjects were informed of the nature of the stimuli, and were shown examples of their spectrograms (similar to that seen in Fig. 1*A*). They then performed a short practice session with feedback. No feedback was provided during the actual experiment. Subjects were instructed to look at a fixation cross presented on a computer screen while performing the task. The experimental session lasted ~10 min. Ten different examples of each CL were presented, resulting in a total of 60 stimulus presentations.

EEG experiment. Subjects undertook 60 1 min trials each. The CL changed every second, and was selected randomly and uniformly from between zero and 10, and in steps of two (0, 2, 4, 6, 8, 10). The experimental session was divided into two conditions: active and passive listening. The passive condition was undertaken for the first 30 trials. This was done to ensure that subjects were naive to the content of the stimuli. They were instructed to watch a film with subtitles, which was presented on a separate computer to the one used for the presentation of the auditory stimuli. Subjects were instructed to minimize eye blinking and all other motor activity. The active listening condition was undertaken for the remaining 30 trials, where subjects maintained focus on a crosshair centered on the screen. Subjects were informed of the nature of the stimuli, and were instructed to respond to targets embedded in the stimuli. Targets consisted of a ramped figure in which six coherent tones increased in frequency on each successive chord for the duration of the segment. An example of 5 s of a stimulus with a target beginning at 2 s is shown in Figure 1*C*. The number of targets per trial was randomly selected to be between four and eight, and the interval between successive targets was

also random. Subjects responded with a button press upon hearing a target, and were told that accuracy was more important than speed. Subjects were played examples of targets before the active listening condition, and allowed a single practice session before the experiment began. After each trial, subjects were presented with feedback on the percentage of targets correctly identified, and on the number of times that they incorrectly responded.

It is important to note that the targets were merely a method to ensure the active engagement of each subject. They were not being asked to attend to temporal coherence per se (i.e., the task was not to respond when the stimulus was more coherent or less coherent). Rather, they were charged with listening out for a particular ramp that happens only at a CL of 6, where the value of 6 was chosen so as to be a midrange CL. To eliminate any confounding target related responses in the EEG data, all segments of each trial that contained a target, and the segments immediately after each target, were discarded when analyzing the EEG data. The stimuli for the active and passive listening conditions were identical, and target sections were removed from the analysis of the passive EEG data as well. To prevent discontinuities in the data, all EEG segments were windowed with a 10 ms raised-cosine ramp.

Data acquisition and preprocessing

128-channel EEG data were filtered over the range of 0 to 134 Hz and digitized at the rate of 512 Hz using a BioSemi Active Two system. Data were referenced to the average of all electrode channels. Data were digitally filtered offline between 0.1 and 5 Hz using a Chebychev type 2 filter, in both a forwards and backwards direction to remove phase distortion. To decrease the processing time required, all EEG data were then downsampled by a factor of 4 to give an equivalent sampling rate of 128 Hz. Excessively noisy EEG channels were rejected according to the criteria of Junghöfer et al. (2000), and the data on these channels were estimated using spherical spline interpolation (EEGlab; Delorme and Makeig, 2004). Independent component analysis was performed independently for each subject using the Infomax algorithm (EEGlab; Delorme and Makeig, 2004). Components constituting artifacts were removed via visual inspection of their topographical distribution and frequency content.

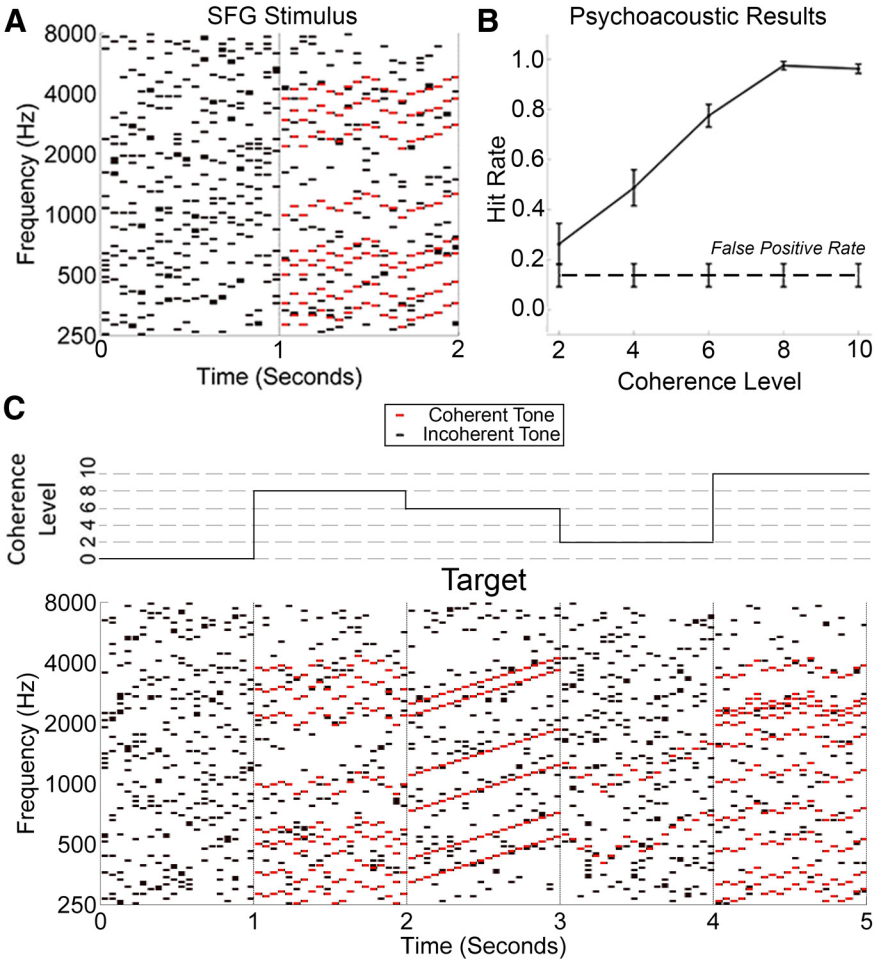

Figure 1. Examples of SFG stimuli and the psychoacoustic results. ***A***, The spectrogram of an example of an SFG stimulus used in our experiment. Each black/red dot represents a pure tone lasting for 50 ms. A chord is defined as the summation of multiple tones. Here, each chord contains exactly 15 tones, which are not harmonically related. There is no overlap between neighboring chords, resulting in 20 inharmonic chords every second. From 0 to 1 s, there is no temporal coherence between the various tones from chord to chord. However, from 1 to 2 s, a noticeable pattern emerges whereby eight groups of tones (in this example) change frequency from chord to chord in a coordinated fashion (red). These changes in frequency are random, but are limited to $\leq$2 semitones per chord. This coordinated movement (temporal coherence) between groups of tones can result in the spontaneous percept of a figure popping out of a random background of varying tones. ***B***, Results from the psychoacoustic experiment in which subjects were given a figure-detection task. The false-positive rate was calculated as the number of times that subjects reported detecting a figure when no figure was present. Error bars represent SEM. ***C***, An example of 5 s of a stimulus used for the EEG experiment. The coherence level changes every second, as illustrated by the step function above the spectrogram. In this example, a "target" occurs from 2 to 3 s, and consists of a ramped figure in which all coherent tones move upwards continuously for the duration of the figure.

Temporal response function computation

The method used here to analyze the relationship between the CL and the EEG data is known as a temporal response function (TRF; Lalor et al., 2006, 2009; Ding and Simon, 2012; Gonçalves et al., 2014). A TRF can be interpreted as a filter that describes the brain's linear transformation of an input stimulus feature to continuous EEG data, and is calculated by performing linear regression between these two variables. Intuitively, it can be thought of as being similar to a cross-correlation. We represent the EEG data at electrode channel n at time $t = 1 \ldots T$ as $r_n(t)$. In our case, the input stimulus feature $s(t)$ is a step function indicating the CL for each 1 s segment of the stimulus. To observe the effect that the CL has on the EEG data over time, a set of time lags τ is applied to $s(t)$, resulting in the lag matrix S:

$$S = \begin{pmatrix} 1 & s(1-\tau_{min}) & \cdots & s(2) & s(1) & 0 & \cdots & 0 \\ 1 & s(2-\tau_{min}) & \cdots & s(3) & s(2) & s(1) & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 & s(\mathrm{T}) & s(\mathrm{T}-1) & \cdots & s(\mathrm{T}-\tau_{max}) \end{pmatrix}$$

A column of ones is added to account for the intercept term of the regression. All time lags between -200 and 800 ms relative to the stimulus onset were used here.

The TRF is then calculated as follows: $\mathrm{TRF} = [S^{\mathrm{T}}S]^{-1}S^{\mathrm{T}}R$, where R is a matrix containing the response of each electrode at time t:

$$R = \begin{pmatrix} \mathrm{r}_1(1) & \mathrm{r}_2(1) & \cdots & \mathrm{r}_n(1) \\ \mathrm{r}_1(2) & \mathrm{r}_2(2) & \cdots & \mathrm{r}_n(2) \\ \vdots & \vdots & & \vdots \\ \mathrm{r}_1(\mathrm{T}) & \mathrm{r}_2(\mathrm{T}) & \cdots & \mathrm{r}_n(\mathrm{T}) \end{pmatrix}$$

To prevent overfitting, ridge regression is performed whereby a bias term λM is added to the autocovariance matrix S^TS, resulting in the following modified equation: $\mathrm{TRF} = [S^{\mathrm{T}}S + \lambda M]^{-1}S^{\mathrm{T}}R$, where

$$M = \begin{bmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 1 & \end{bmatrix}$$

and λ is a constant parameter, selected to optimize the correlation between actual and predicted EEG data. See Lalor et al. (2006) for further details.

Time-locked responses

While the TRF represents a summary measure of how the brain responds to changes in CL, we also wished to investigate time-locked responses to each individual CL. To do this, all segments of the data containing a CL N were concatenated into a single matrix, along with the subsequent and preceding 1 s segments of data. To improve the signal-to-noise ratio of the obtained time-locked responses, denoising-source-separation (DSS) was applied to the data using the Noise Tools toolbox (http://audition.ens.fr/adc/NoiseTools). DSS is a blind-source separation technique that optimally extracts neural activity according to some bias criterion (de Cheveigné and Parra, 2014). This is implemented by performing joint diagonalization between two covariance matrices, the first of which (c0) is the covariance matrix of the raw data, and the second (c1) is the covariance matrix of the same data but filtered in such a way as to emphasize a particular feature of interest. In our case, the bias criterion was the average of the data over trials (de Cheveigné and Simon, 2008; Ding et al., 2013). Specifically, c1 was the sum of the covariance matrices of the mean of each CL:

$$c1 = cov(\overline{CL_0}) + cov(\overline{CL_2}) + \ldots + cov(\overline{CL_{10}})$$

where $\overline{CL_N}$ is the mean of the EEG data in response to CL N. DSS was performed on each subject independently, with the first 10 DSS components projected back to sensor space to obtain the denoised EEG data.

Due to the fact that the CL changed every second, there was no refractory period after each change to allow the EEG data to return to a resting state. As such, for each subject, we subtracted the average responses to each CL from the average response obtained to CL 0. Finally, all epochs for each CL were baseline corrected by subtracting the mean of the response in the interval from −500 to 0 ms.

Global field power

To examine our TRFs and time-locked responses for evidence of signal, we first calculated a measure known as global field power (GFP; Lehmann and Skrandies, 1980). This is a single, reference-independent measure of response strength over the entire scalp. It is simply a measure of the standard deviation (SD) of the response across channels calculated at each point in time.

Results

Behavioral results

Psychoacoustic experiment

The results of the psychoacoustic experiment are presented in Figure 1*B*. Despite the more complex nature of our stimuli compared with previous studies (Teki et al., 2011, 2013), these data demonstrate that listeners are capable of detecting the figures embedded in the stimuli. Furthermore, they reveal that the hit rate increases as an approximately linear function of CL. The false-positive rate was calculated as the number of times that subjects reported detecting a figure when no figure was present. The hit rate for CL 2 was the only one that was not significantly greater than the false-positive rate ($p = 0.125$, Wilcoxon signed-rank test).

EEG experiment

In the active listening condition subjects were compliant with the task, responding correctly with a median of 72% per trial (25$^{\mathrm{th}}$ percentile, 69%; 75$^{\mathrm{th}}$ percentile, 83%), and incorrectly responding when no target was present with a median of once per trial (25$^{\mathrm{th}}$ percentile, 0; 75$^{\mathrm{th}}$ percentile, 1).

EEG results

TRF analysis

GFP plots of the grand-average TRFs obtained for the active and passive listening conditions are shown in Figure 2*A*. Robust responses are clearly visible for both conditions at poststimulus time lags. To quantitatively establish the time points at which these responses were significantly different than zero activity, we carried out two-tailed t tests for each electrode and at each time point between −200 and 800 ms. Because all electrodes and time points were being assessed simultaneously, multiple comparisons were corrected for via the Benjamini and Hochberg (1995) false detection rate (FDR) algorithm, with an FDR of $q = 0.05$; a procedure recommended by Lage-Castellanos et al. (2010). Two time intervals of significant activation were identified for the active condition and one interval for the passive condition. The timing of these intervals is summarized in Table 1, and can be visualized in Figure 2*B*,*C*. All reported values are mean ± SD (milliseconds), where the mean was calculated across subjects.

Based on this analysis, we tested to see whether there were any significant differences between the active and passive responses from 150 to 400 ms and from 500 to 700 ms after stimulus (two-tailed paired t test). Separate corrections were performed on these time frames using the Benjamini et al. (2006) FDR algorithm, a method that is encouraged when statistical tests are performed with a priori hypotheses (Groppe et al., 2011). The FDR q was again set to be 0.05. For the first time frame, significant differences were found in the interval from 257 ± 16 to 328 ± 40 ms (mean ± SD), with an effect size (Cohen's d) spanning from 1.56 ($t = 257$ ms) to 0.56 ($t = 328$ ms). No significant differences were found for the second time frame (Cohen's $d < 0.49$). In both cases, the effect size was calculated using the GFP. These results can be visualized in Figure 2*D*.

To get a better sense of the source of the responses, scalp voltage and current-source density (CSD; Perrin et al., 1989; Kayser and Tenke, 2006) topographical maps of the grand-average TRFs were computed and plotted (Fig. 2*E*,*F*). These analyses suggest bilateral sources originating from temporal regions. Using a nonparametric test to detect differences in the topographic distribution of voltage responses across the scalp (T-ANOVA; Brunet et al., 2011), no significant differences were found between the topographies of the active and passive conditions when comparing the peak of their respective responses (Active: $t = 200$ ms; Passive: $t = 150$ ms; $p = 0.15$).

Time-locked responses

GFP plots of the grand-average time-locked responses are shown for the active (Fig. 3*A*) and passive (Fig. 3*B*) listening conditions. Each CL is represented by a different color. For both listening conditions, there is a slow rise in activation, which plateaus at ~500 ms after stimulus onset. The topographical distributions of the responses are similar to those of the TRF analyses, again suggesting a bilateral source originating from temporal regions.

Similarly to the TRF analysis described previously, we wished to quantitatively establish the time points at which these responses were significantly different from zero activity. To do so we carried out two-tailed t tests for each CL, for each electrode, and at each time point between −500 and 1500 ms. Again, multiple comparisons were corrected for via the Benjamini and Hochberg (1995) FDR algorithm, with an FDR of $q = 0.05$. For

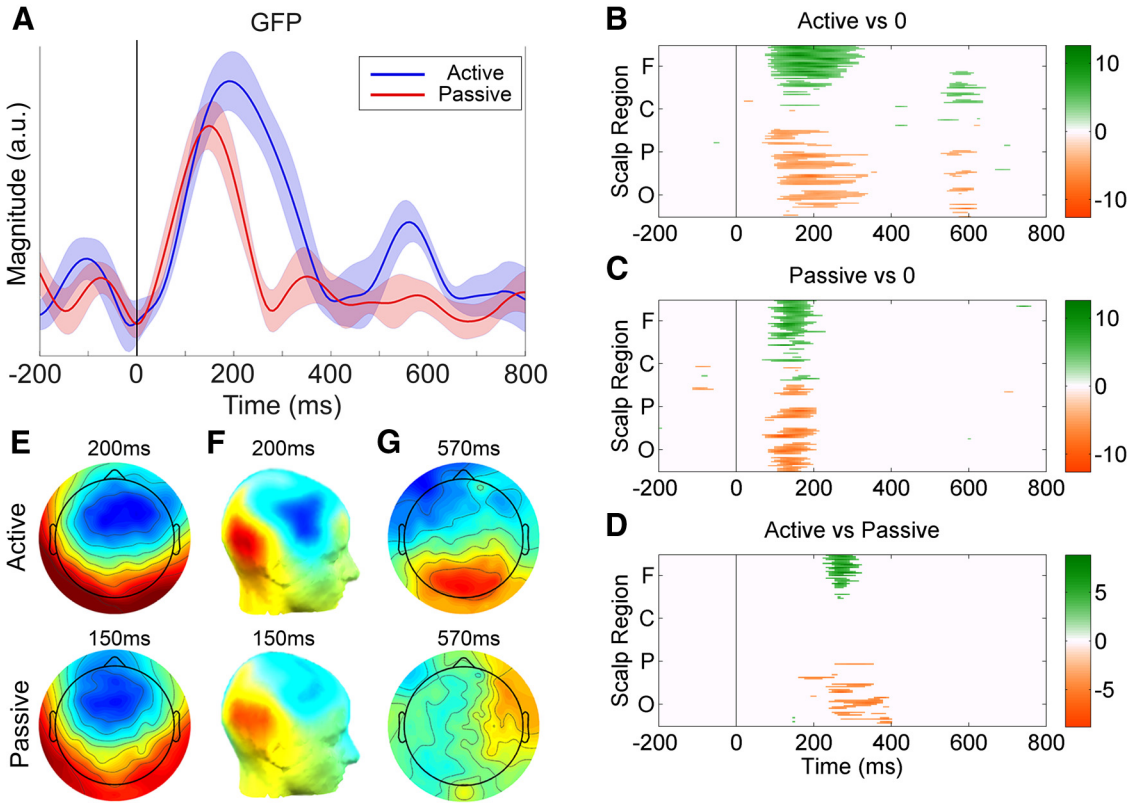

Figure 2. Results of the TRF analysis. ***A***, GFP plots of the grand-average TRFs for the active (blue) and passive (red) conditions. The solid black line indicates zero time lag between EEG and stimulus. The shaded areas indicate SEM. ***B***, ***C***, Statistical cluster plots marking the time points for all electrodes at which the TRF response differed significantly from zero on the basis of two-tailed *t* tests, with multiple comparisons corrected for via FDR ($q = 0.05$). White denotes nonsignificance, whereas positive *t* values (TRF > 0) are marked on a green scale and negative *t* values (TRF < 0) are marked in gold. Electrodes are ordered from the bottom: occipital (O), parietal (P), central (C), and frontal (F) proceeding in the anterior direction in rows from left to right. ***D***, The time points at which the active response differed significantly from the passive response (2-tailed paired *t* test, FDR corrected, $q < 0.05$). ***E***, ***F***, Scalp voltage topographies (***E***) and CSD topographies (***F***) of the TRFs for the Active ($t = 200$ ms) and Passive ($t = 150$ ms) conditions. ***G***, Scalp voltage topographies of the TRFs for the Active and Passive conditions at 570 ms. All topographies are displayed on the same scale.

Table 1. The times at which the active and passive TRFs were significantly different from zero activity

Condition	Onset 1	Peak 1	Offset 1	Onset 2	Peak 2	Offset 2
Active	117 ± 21 ms	210 ± 47 ms	265 ± 56 ms	555 ± 50 ms	570 ± 93 ms	617 ± 50 ms
Passive	117 ± 19 ms	156 ± 18 ms	187 ± 16 ms	—	—	—

Results are displayed as mean ± SD (2-tailed *t* test, FDR corrected). The onset, peak, and offset of the responses are shown.

the active condition, responses to CLs of 6, 8, and 10 exhibited activity significantly different from zero activity. However, for the passive condition, this was only true for CLs of 8 and 10. We also observed a clear gradation in terms of the onset of this activity, with successively lower CLs eliciting responses at longer latencies. Passive responses also occurred substantially later than active responses. The time points at which this activity began for each CL is summarized in Table 2.

A clear separation in magnitude can be seen for many of the CLs, particularly for the active listening condition. To quantify any significant differences, the mean of the responses in the interval from 500 to 1000 ms was selected for each subject, and a Friedman test was implemented. A significant effect of CL was found for both the active and passive listening conditions ($\chi^2_{(4)} = 32.64$, $p = 1 \times 10^{-6}$ and $\chi^2_{(4)} = 21.44$, $p = 2.6 \times 10^{-4}$; respectively). Post hoc analyses were performed using right-tailed Wilcoxon signed-rank tests, and multiple comparisons were corrected for via FDR ($q = 0.05$). The effect size was evaluated using Cohen's *d*. First, we tested to see whether the response to each CL was significantly greater than the response to the lower adjacent CL within each listening condition (e.g., is the response to CL 10 during the active listening condition greater than the response to CL 8 during the active listening condition). Second, we tested to see whether there were significant differences between the active and passive listening conditions for each CL; i.e., is the response to CL *N* during the active listening condition greater than the response to the same CL during the passive listening condition. These results can be visualized in Table 3.

Discussion

Using linear regression methods, we have shown that it is possible to obtain a highly temporally resolved neural signature of the computation of temporal coherence in a complex acoustic scene (Fig. 2*A*). The observed response has an onset at ~115 ms for both active and passive listening conditions. This response is likely to be preattentive, which is in accord with previous research that has observed preattentive neural computations of temporal coherence (Teki et al., 2011). Under active listening, this response persists significantly longer and peaks later with a greater amplitude, which suggests that these computations are enhanced and prolonged during active engagement in analyzing an auditory scene. The fact that no significant differences were found between

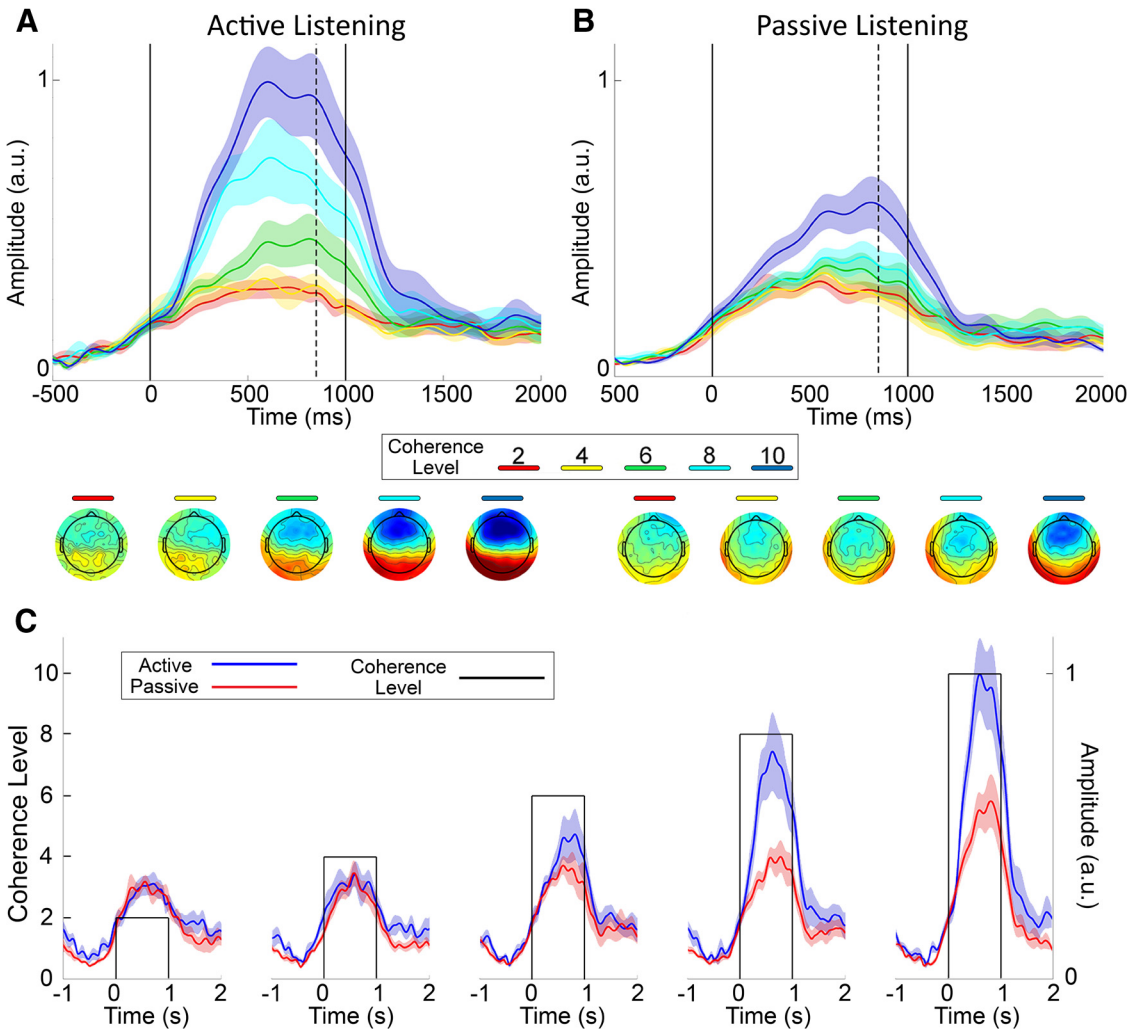

Figure 3. Results of the time-locked average responses. ***A***, ***B***, GFP plots of the time-locked grand-average responses to the five different CLs during the Active (***A***) and Passive (***B***) listening conditions. Each CL is represented by a different color. The shaded areas indicate SEM. The dotted lines indicate the time point at which the topographies are shown (850 ms). The solid lines illustrate the duration for which each CL lasted (1 s). ***C***, The same information as in ***A*** and ***B***, except plotted so as to directly compare the Active (blue) and Passive (red) listening conditions with respect to each individual CL (solid lines). All data are normalized relative to the Active condition in response to a CL of 10.

Table 2. The times at which the time-locked responses are first significantly different from zero activity

CL	Active	Passive
10	206 ± 67 ms	386 ± 100 ms
8	261 ± 49 ms	433 ± 103 ms
6	394 ± 45 ms	—
4	—	—
2	—	—

Results are displayed as mean ± SD (2-tailed *t* test, FDR corrected, $q = 0.05$).

the topographies of the active and passive responses at the peak of their time courses suggests that similar neural mechanisms are performing these computations under both conditions.

There are two factors that suggest that the response seen here is likely an explicit measure of temporal coherence computations. First, due to the design of our stimulus, every chord contains exactly 15 tones, meaning that the broadband power and all low-level stimulus features remain constant for the duration of its presentation. Second, our analysis approach explicitly relies on a relationship between the EEG data and the modulating CL. Therefore, although our stimulus certainly activates multiple stages of the auditory processing hierarchy, which are unrelated to temporal coherence computations, it should not activate these areas in a way that correlates with our regression signal (Gonçalves et al., 2014).

As to the latencies at which a response can be considered "low level," previous experiments using target detection tasks in a stream segregation paradigm have shown that neural responses earlier than 75 ms are similar for both detected and undetected targets, whereas later activity is not (Gutschalk et al., 2008; Königs and Gutschalk, 2012). These authors therefore suggested that such early activity reflects purely sensory stimulus processing (Gutschalk and Dykstra, 2014). Therefore, the fact that we see no significant activation in the TRFs obtained in this experiment before ~115 ms suggests that low-level stimulus features are indeed absent from this response.

One peculiar result from our analysis was the fact that we observed significant activation at ~570 ms in the active listening condition only (Fig. 2*B*), but found no significant differences between the active and passive conditions at this latency (Fig. 2*D*). However, there was a significant difference in terms of their topographies (T-ANOVA, $p = 0.03$; Fig. 2*G*), which suggests that the lack of any differences in terms of amplitude could simply be

Table 3. Statistical tests comparing CLs for the time-locked responses

Comparing amplitudes of the neural responses to successive CLs for the Active and Passive conditions			Comparing Active versus Passive responses to each individual CL	
CL	Active	Passive	CL	Active > Passive
4 > 2	$p = 0.423$; $d = 0.124$	$p = 0.500$; $d = 0.043$	2	$p = 0.539$; $d = 0.083$
6 > 4	*$p = 0.014$; $d = 0.825$	$p = 0.097$; $d = 0.286$	4	$p = 0.385$; $d = 0.186$
8 > 6	*$p = 0.002$; $d = 0.765$	$p = 0.065$; $d = 0.464$	6	*$p = 0.010$; $d = 0.748$
10 > 8	*$p = 0.002$; $d = 0.760$	*$p = 0.010$; $d = 1.115$	8	*$p = 0.002$; $d = 1.253$
—	—	—	10	*$p = 0.014$; $d = 1.097$

All tests were performed using right-tailed Wilcoxon signed-rank tests. Multiple comparisons were corrected for via FDR with a false detection rate of $q = 0.05$. Asterisk indicates significance. Effect size was calculated using Cohen's *d*.

an issue of statistical power. Further work is required to determine the functional significance of the activity seen at this latency.

TRF interpretation

The TRF should be considered as a measure of the average brain response to a unit increment in CL, the value of which changes once per second in this particular experiment. Given that each chord lasts for 50 ms, the fact that we first observe significant activity at ~115 ms in our TRFs seems to suggest that just two chords are sufficient to begin the stream segregation process. This may be the case for large CLs, but due to the fact that the TRF is an average brain response, the same activity would likely not be observed for lower CLs.

When comparing the active and passive listening conditions, the cause of the larger and longer-lasting TRFs for the active condition is likely due to more accurate and consistent neural tracking of the varying CL, a difference that is in accord with the behavioral advantages conferred by attention in segregating an auditory scene (Fritz et al., 2007; Spielmann et al., 2014).

Time-locked responses

The larger TRFs observed for the active listening condition suggest that we should see larger responses to CL changes in our time-locked responses. This was indeed the case (Fig. 3). With regards to the latency of the responses, during the active listening condition we first see significant activity in response to a CL of 10 at ~200 ms. The latency of this first response then occurs progressively later for each successive (lower) CL. These responses were substantially later in the passive listening condition, beginning only at ~385 ms in response to a CL of 10. We also observed significant separation between the magnitudes of the responses to each individual CL. Under active listening, this separation is facilitated, with CLs of 6, 8, and 10 significantly different from each other. However, under passive listening, only a CL of 10 is separable from the rest. We found no significant differences between CLs 2 or 4, neither within nor between listening conditions. Furthermore, neither of these CLs elicited activity significantly different from zero in either listening condition. That said, given the morphology of the GFP plots for these lower CLs and the similarities in the topographical distribution of the responses across all CLs, we contend that the lack of statistically significant activity at these lower CLs is likely to be simply an issue of response power. The signal-to-noise ratio for these CLs may simply be too low for the responses to reach significance with the amount of data available. In support of this conjecture, the psychoacoustic results demonstrate that a CL of 4 is indeed detectable (Fig. 1*B*).

Psychoacoustic results

There is a largely linear relationship between hit rate and CL, with hit rate approaching ceiling at a CL of 8 (Fig. 1*B*). This rising behavioral performance as a function of increasing CL mirrors the growing amplitudes of the GFP plots observed in response to increasing CLs (Fig. 3*C*). Compared with Teki et al. (2011), the hit rate for each CL appears to be lower in our experiment. For example, Teki et al. (2011) report a hit rate of ~0.95, ~0.9, and ~0.65 for CLs of 6, 4, and 2, respectively. We, instead, report a hit rate of ~0.8, ~0.5, and ~0.2, respectively. Furthermore, contrary to the findings of Teki et al. (2011), the hit rate for CL 2 in our experiment was not significantly greater than the false-positive rate. This difference can probably be attributed to the more complex nature of our stimuli.

Neural sources

With regards to the neural sources of these responses, previous research using fMRI has shown activation bilaterally in the posterior intraparietal sulcus in response to varying CLs (Teki et al., 2011). For our data, the CSD analyses performed on the topographic distributions of the observed TRFs also suggest bilateral sources originating from temporal regions during both active and passive listening (Fig. 2*F*).

Ecological validity

An important characteristic of SFG stimuli is their rapid build-up rate (the time required to segregate the figure from the background). Contrary to many brain-imaging and electrophysiological experiments on stream segregation that report a build-up time on the order of several seconds (Micheyl et al., 2007; Gutschalk et al., 2008; Pressnitzer et al., 2008; Elhilali et al., 2009b), the stimuli used here can easily be segregated in hundreds of milliseconds (Teki et al., 2011, 2013). This makes them more suitable for studying the real-time computations that the brain must routinely make in naturally complex acoustic environments. For instance, our results here correspond well with previous experiments using natural speech in the classic "cocktail party" paradigm. One such study (Power et al., 2012; cf. O'Sullivan et al., 2014) presented subjects with two speakers, and instructed them to attend to one speaker while ignoring the other. Their results showed that both the attended and unattended speech streams elicited similar neural activity in the interval of 50–150 ms. However, the attended speech exhibited a subsequent peak at ~200 ms, whereas the unattended speech elicited much reduced activation. Interestingly, in the current study, we observed a peak in the response of the active listening condition at ~200 ms, whereas activity in the passive condition peaked at ~150 ms and subsided by ~190 ms. The similarity between the latencies of the responses in these two experiments is striking given the dissimilarity of the stimuli. More work is required to determine whether temporal coherence computations are causally linked to the responses seen in such experiments using natural speech, or whether two different neural systems are at play with coincidentally similar latencies.

Notes

Supplemental material for this article is available at http://www.mee.tcd.ie/lalorlab/resources/temporalCoherenceExample.wav, where an example of an SFG stimulus used in the EEG experiment can be found. This supplemental material has not been peer reviewed.

References

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc B 57:289–300.

Benjamini Y, Krieger AM, Yekutieli D (2006) Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93:491–507. CrossRef

Bregman A (1990) Auditory scene analysis: the perceptual organization of sound. Cambridge, MA: MIT.

Brunet D, Murray MM, Michel CM (2011) Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput Intell Neurosci 2011:813870. CrossRef Medline

de Cheveigné A, Parra LC (2014) Joint decorrelation, a versatile tool for multichannel data analysis. Neuroimage 98:487–505. CrossRef Medline

de Cheveigné A, Simon JZ (2008) Denoising based on spatial filtering. J Neurosci Methods 171:331–339. CrossRef Medline

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21. CrossRef Medline

Ding N, Simon JZ (2012) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 107:78–89. CrossRef Medline

Ding N, Chatterjee M, Simon JZ (2013) Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. Neuroimage 88C:41–46. CrossRef Medline

Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA (2009a) Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron 61:317–329. CrossRef Medline

Elhilali M, Xiang J, Shamma SA, Simon JZ (2009b) Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. PLoS Biol 7:e1000129. CrossRef Medline

Fritz JB, Elhilali M, David SV, Shamma SA (2007) Auditory attention—focusing the searchlight on sound. Curr Opin Neurobiol 17:437–455. CrossRef Medline

Gonçalves NR, Whelan R, Foxe JJ, Lalor EC (2014) Towards obtaining spatiotemporally precise responses to continuous sensory stimuli in humans: a general linear modeling approach to EEG. Neuroimage 97:196–205. CrossRef Medline

Groppe DM, Urbach TP, Kutas M (2011) Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology 48:1711–1725. CrossRef Medline

Gutschalk A, Dykstra AR (2014) Functional imaging of auditory scene analysis. Hear Res 307:98–110. CrossRef Medline

Gutschalk A, Micheyl C, Oxenham AJ (2008) Neural correlates of auditory perceptual awareness under informational masking. PLoS Biol 6:e138. CrossRef Medline

Härmä A, Palomäki K (2000) HUTear—a free Matlab toolbox for modeling of human auditory system. In: Proc. 1999 Matlab DSP Conference, pp 96–99.

Junghöfer M, Elbert T, Tucker DM, Rockstroh B (2000) Statistical control of artifacts in dense array EEG/MEG studies. Psychophysiology 37:523–532. CrossRef Medline

Kayser J, Tenke CE (2006) Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. Clin Neurophysiol 117:348–368. CrossRef Medline

Königs L, Gutschalk A (2012) Functional lateralization in auditory cortex under informational masking and in silence. Eur J Neurosci 36:3283–3290. CrossRef Medline

Krishnan L, Elhilali M, Shamma S (2014) Segregating complex sound sources through temporal coherence. PLoS Comput Biol 10:e1003985. CrossRef Medline

Lage-Castellanos A, Martínez-Montes E, Hernández-Cabrera JA, Galán L (2010) False discovery rate and permutation test: an evaluation in ERP data analysis. Stat Med 29:63–74. CrossRef Medline

Lalor EC, Foxe JJ (2010) Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. Eur J Neurosci 31:189–193. CrossRef Medline

Lalor EC, Pearlmutter BA, Reilly RB, McDarby G, Foxe JJ (2006) The VESPA: a method for the rapid estimation of a visual evoked potential. Neuroimage 32:1549–1561. CrossRef Medline

Lalor EC, Power AJ, Reilly RB, Foxe JJ (2009) Resolving precise temporal processing properties of the auditory system using continuous stimuli. J Neurophysiol 102:349–359. CrossRef Medline

Lehmann D, Skrandies W (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr Clin Neurophysiol 48:609–621. CrossRef Medline

Luck SJ (2005) An introduction to the event-related potential technique. Cambridge, MA: MIT.

Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Courtenay Wilson E (2007) The role of auditory cortex in the formation of auditory streams. Hear Res 229:116–131. CrossRef Medline

O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, Slaney M, Shamma SA, Lalor EC (2014) Attentional selection in a cocktail party environment can be decoded from single-trial EEG. Cereb Cortex. Advance online publication. Retrieved April 9, 2015. Medline

Perrin F, Pernier J, Bertrand O, Echallier JF (1989) Spherical splines for scalp potential and current density mapping. Electroencephalogr Clin Neurophysiol 72:184–187. CrossRef Medline

Power AJ, Foxe JJ, Forde EJ, Reilly RB, Lalor EC (2012) At what time is the cocktail party? A late locus of selective attention to natural speech. Eur J Neurosci 35:1497–1503. CrossRef Medline

Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. Curr Biol 18:1124–1128. CrossRef Medline

Shamma SA, Micheyl C (2010) Behind the scenes of auditory perception. Curr Opin Neurobiol 20:361–366. CrossRef Medline

Shamma SA, Elhilali M, Micheyl C (2011) Temporal coherence and attention in auditory scene analysis. Trends Neurosci 34:114–123. CrossRef Medline

Shamma S, Elhilali M, Ma L, Micheyl C, Oxenham AJ, Pressnitzer D, Yin P, Xu Y (2013) Temporal coherence and the streaming of complex sounds. In: Basic aspects of hearing (Moore BCJ, Patterson RD, Winter IM, Carlyon RP, Gockel HE, eds), pp 535–543. Heidelberg: Springer.

Spielmann MI, Schröger E, Kotz SA, Bendixen A (2014) Attention effects on auditory scene analysis: insights from event-related brain potentials. Psychol Res 78:361–378. CrossRef Medline

Sussman ES, Horváth J, Winkler I, Orr M (2007) The role of attention in the formation of auditory streams. Percept Psychophys 69:136–152. CrossRef Medline

Teki S, Chait M, Kumar S, von Kriegstein K, Griffiths TD (2011) Brain bases for auditory stimulus-driven figure–ground segregation. J Neurosci 31:164–171. CrossRef Medline

Teki S, Chait M, Kumar S, Shamma S, Griffiths TD (2013) Segregation of complex acoustic scenes based on temporal coherence. Elife 2:e00699. CrossRef Medline

Wolf G, Mallat S, Shamma S (2014) Audio source separation with time-frequency velocities. In: IEEE international workshop on machine learning for signal processing (MLSP), pp 1–6.

Zion Golumbic E, Cogan GB, Schroeder CE, Poeppel D (2013) Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party." J Neurosci 33:1417–1426. CrossRef Medline