



# Effect of Task and Attention on Neural Tracking of Speech

Jonas Vanthornhout\*, Lien Decruy and Tom Francart

Department of Neurosciences, ExpORL, KU Leuven, Leuven, Belgium

EEG-based measures of neural tracking of natural running speech are becoming increasingly popular to investigate neural processing of speech and have applications in audiology. When the stimulus is a single speaker, it is usually assumed that the listener actively attends to and understands the stimulus. However, as the level of attention of the listener is inherently variable, we investigated how this affected neural envelope tracking. Using a movie as a distractor, we varied the level of attention while we estimated neural envelope tracking. We varied the intelligibility level by adding stationary noise. We found a significant difference in neural envelope tracking between the condition with maximal attention and the movie condition. This difference was most pronounced in the right-frontal region of the brain. The degree of neural envelope tracking was highly correlated with the stimulus signal-to-noise ratio, even in the movie condition. This could be due to residual neural resources to passively attend to the stimulus. When envelope tracking is used to measure speech understanding objectively, this means that the procedure can be made more enjoyable and feasible by letting participants watch a movie during stimulus presentation.

**Keywords:** auditory processing, envelope tracking, decoding, EEG, attention

## OPEN ACCESS

### Edited by:

Claude Alain,  
Rotman Research Institute (RRI),  
Canada

### Reviewed by:

Christoph Kayser,  
Bielefeld University, Germany  
Samira Anderson,  
University of Maryland, College Park,  
United States

### \*Correspondence:

Jonas Vanthornhout  
jonas.vanthornhout@kuleuven.be

### Specialty section:

This article was submitted to  
Auditory Cognitive Neuroscience,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 24 May 2019

**Accepted:** 30 August 2019

**Published:** 16 September 2019

### Citation:

Vanthornhout J, Decruy L and  
Francart T (2019) Effect of Task and  
Attention on Neural Tracking of  
Speech. *Front. Neurosci.* 13:977.  
doi: 10.3389/fnins.2019.00977

## 1. INTRODUCTION

EEG-based measures of neural tracking of natural running speech are becoming increasingly popular to investigate neural processing of speech and for applications in domains such as audiology (Vanthornhout et al., 2018; Lesenfants et al., 2019) and disorders of consciousness (Braiman et al., 2018). They are also useful for the basic scientific investigation of speech processing, which has traditionally been conducted using simple, repetitive stimuli. The benefit of using a natural continuous speech stimulus is that there is no influence of neural repetition suppression and the ecological validity is greater (Brodbeck et al., 2018; Summerfield et al., 2008). With the stimulus reconstruction method, natural running speech is presented to a listener. A feature of the speech signal, such as the envelope, is then reconstructed from the brain responses. The correlation between the original and reconstructed signal is a measure of neural coding of speech (Ding and Simon, 2012). Inversely, the brain responses can be predicted from features of the speech signal. Lalor et al. (2009), for example, found that the transfer function from the acoustic envelope to brain responses also called the temporal response function (TRF), exhibits interpretable peaks, such as the N1 and P2 peaks (see below).

These methods, reconstructing the stimulus or predicting the brain responses, are often used with either a single speaker or in a competing talker paradigm. With a single speaker, it is usually assumed that the listener actively attends to and understands the stimulus. In the competing talker paradigm, the listener's task is usually to attend to one of several speakers. However, in either paradigm, a listener may tune out once in a while.

In this study, we investigated the effect of attention on neural tracking as a function of the signal-to-noise ratio. Participants paid full attention, or we distracted them by letting them watch a movie. We did this for two types of stimuli, a story and repeated sentences.

Attention is a conscious, active process and has a robust top-down effect on processes along most of the auditory pathway (Forte et al., 2017; Fritz et al., 2007). The concept of auditory attention can be illustrated by the cocktail party problem described by Cherry (1953). Although it is possible to listen to a particular speaker while ignoring other speakers, some critical information, such as your name, can still be recognized when uttered by one of the ignored speakers. The combination of bottom-up salience and top-down attention leads to speech understanding (Shinn-Cunningham, 2008).

The effect of attention has already been investigated in objective measures of hearing, such as auditory brainstem responses (ABRs), auditory steady-state responses (ASSRs) and late auditory evoked potentials. It is often assumed that attention has a minimal effect on the ABR, as it can even be measured when the subject is asleep. However, the results are inconclusive (Brix, 1984; Hoormann et al., 2000; Galbraith et al., 2003; Varghese et al., 2015; Lehmann and Schönwiesner, 2014).

The effect of attention has also been investigated for ASSRs (Ross et al., 2004). Roth et al. (2013) found a significant reduction in ASSR SNR when the difficulty of a video game played by the participants increased. This is consistent with the theory that irrelevant stimuli are suppressed when neural resources are required for another task (Hein et al., 2007). However, Linden et al. (1987) did not find an effect of selective attention at 40 Hz. Moreover, for 20 Hz ASSRs, Müller et al. (2009) found some differences while Skosnik et al. (2007) did not.

Studies using late auditory evoked potentials, have shown that responses in cortical regions are modulated by attention (Picton and Hillyard, 1974; Fritz et al., 2007; Astheimer and Sanders, 2009; Näätänen et al., 2011; Regenbogen et al., 2012). From this, we expect that mainly later responses are modulated by attention as they have a cortical origin.

In general, even with similar stimuli, different studies obtained different results. A reason can be that the participants did not do what was expected of them. Furthermore, these differences could be due to differences in the attentional tasks used. As the human brain is finely attuned to speech perception, it is interesting to investigate the effect of attention on neural measures of speech tracking, which can be used with natural speech as the stimulus, in contrast to the artificial repeated stimuli used in the paradigms described above. There is a wide range of data available on neural tracking of speech in a competing talker paradigm. Two speakers or more are presented simultaneously to a listener, whose task it is to attend to one of them. A stimulus feature is then decoded from the EEG and correlated with the features of each speaker. A decoder transforms the multi-channel neural signal into a single-channel feature, by linearly combining amplitude samples across sensors and a post-stimulus temporal integration window. Comparison of these correlations yields information on which speaker was attended. It has been shown that the attended speaker can be decoded

with high accuracy (in the order of 80–90% for trials of 20–30 s) (O'Sullivan et al., 2015; Mirkovic et al., 2015; Das et al., 2016; Das et al., 2018).

By changing the number of EEG samples included in the decoder, i.e., the latencies of EEG relative to the stimulus, it can be investigated at which time lags, and therefore neural sources, the effect of attention appears. Not unexpectedly, attention mostly affects longer latencies, ranging from 70 ms to 400 ms (Hillyard et al., 1973). Ding and Simon (2012) investigated the neural responses in a dual-talker scenario and found that peaks in the TRF at 100 ms were more influenced by attention than peaks at 50 ms. Other experiments show effects at a later stage, beyond 150 ms (Snyder et al., 2006; Ross et al., 2009). O'Sullivan et al. (2015) found the best attention decoding performance with latencies around 180 ms. Moreover, Puvvada and Simon (2017) found that early responses (< 75 ms) are minimally affected by attention. However, even at the brainstem level, around 9 ms latency, modulation by selective attention can already be found, showing the top-down effects (Forte et al., 2017). As a conclusion, attention prominently affects the later responses but can also affect very early responses.

In a competing talker paradigm, it is also interesting to investigate the relative level of neural speech tracking: is the attended speaker represented more strongly, the unattended one suppressed, or a combination? Bidet-Caulet et al. (2010) shows that the representation of the attended stimulus is not enhanced, but responses to the unattended stimulus are attenuated. Irrelevant stimuli are actively suppressed. Melara et al. (2002), on the other hand, shows that excitatory (enhancing the attended stimulus) and inhibitory (attenuating the unattended stimulus) processes work interactively to maintain sensitivity to environmental change, without being disrupted by irrelevant events. Many researchers have found enhancement of the cortical envelope tracking of attended sounds, relative to unattended sounds. This attentional effect can be found as early as 100 ms post-stimulus (Melara et al., 2002; Choi et al., 2014), corresponding with lags associated with the auditory cortex. Kong et al. (2014), however, demonstrated that top-down attention could both enhance the neural response to the attended sound stream and attenuate the neural responses to an unwanted sound stream.

On the other hand, the literature on the effect of attention on the neural tracking of a single speaker is more sparse. Kong et al. (2014) measured neural tracking of a single talker in quiet, while the participant either actively listened to a stimulus or watched a movie and ignored the auditory stimulus. They found that neural tracking, measured as the peak cross-correlation between the EEG signal and stimulus envelope, was not significantly affected by the listening condition. However, the shape of the cross-correlation function showed stronger N1 and P2 responses in the active listening condition and weaker P1 response compared to the movie condition. However, they did not include more intensive distractors, nor did they investigate the effect of speech intelligibility. Moreover, their analysis was limited to a cross-correlation while decoding the envelope from the neural responses would be more powerful as it uses information from all channels.

We investigated the effect of attention on neural envelope tracking. Attention was manipulated by letting the participants actively attend to the stimulus and answer comprehension questions, or watch a silent movie and ignore the acoustic stimulus. This was done at multiple levels of speech understanding by adding stationary background noise.

We hypothesized that envelope tracking would be strongest in the attended condition. However, watching a movie may stabilize the level of attention, as maintaining focus on the movie might be more comfortable than maintaining focus on (uninteresting) sentences, therefore reducing intra-subject variability. We also hypothesized that this effect would be most apparent for decoders with a temporal integration window beyond 100 ms.

## 2. METHODS

### 2.1. Participants

We recruited 19 young normal-hearing participants between 18 and 35 years old. Every subject reported normal hearing (thresholds lower than 20 dB HL for all audiometric octave frequencies), which was verified by pure tone audiometry. They had Dutch (Flemish) as their mother tongue and were unpaid volunteers. Before each experiment, the subjects signed an informed consent form approved by the local Medical Ethics Committee (reference no. S57102).

### 2.2. Apparatus

The experiments were conducted using APEX 3 (Francart et al., 2008), an RME Multiface II sound card (RME, Haimhausen, Germany) and Etymotic ER-3A insert phones (Etymotic Research, Inc., Illinois, USA) which were electromagnetically shielded using CFL2 boxes from Perancea Ltd. (London, United Kingdom). The speech was always presented at 60 dBA, and the set-up was calibrated with a 2 cm<sup>3</sup> coupler (Brüel & Kjær 4152, Nærum, Denmark) using the speech weighted noise of the corresponding speech material. The experiments took place in an electromagnetically shielded and soundproofed room. To measure EEG, we used an ActiveTwo EEG set-up with 64 electrodes from BioSemi (BioSemi, Amsterdam, Netherlands).

### 2.3. Behavioral Experiments

Behavioral speech understanding was measured using the Flemish Matrix test (Luts et al., 2015). Each Matrix sentence consisted of 5 words spoken by a female speaker and was presented to the right ear. Right ear stimulation was chosen as this is how the Matrix test has been standardized. The sentences of the Flemish Matrix test have a fixed structure of “name verb numeral adjective object,” e.g., “Sofie ziet zes grijze pennen” (“Sophie sees six gray pens”). Each category of words has 10 possibilities. The result of the Flemish Matrix test is a word score.

For 8 participants, several lists of sentences were presented at a fixed SNR ranging from -12 dB SNR to -3 dB SNR<sup>1</sup>. For the constant procedure, we estimated the speech reception threshold

(SRT) of the participants by fitting a psychometric curve through their word scores, using the formula  $score = \frac{1}{1 + e^{-\frac{SNR - \alpha}{\beta}}}$ , with  $\alpha$  the SRT and  $\beta$  the slope. For the other 11 participants, to speed up the recordings, we used 3 runs of an adaptive procedure (Brand and Kollmeier, 2002) in which we changed the SNR until we obtained 29, 50, or 71% speech understanding. We considered the SNR after the last trial as the SRT at the desired level.

## 2.4. EEG Experiments

After the behavioral experiment, we conducted the EEG experiment. The stimuli were presented with the same set-up as the behavioral experiments, with the exception that we used diotic stimulation for the EEG experiment to make the experiments more comfortable.

### 2.4.1. Speech Material

We presented stimuli created by concatenating two lists (20 sentences per list) of Flemish Matrix sentences (2 s per sentence) with a random gap of minimum 0.8 s and maximum 1.2 s between the sentences. The total duration of this stimulus was around 120 s, with 80 s of speech. The stimulus was presented at different SNRs (-12.5 dB SNR up to +2.5 dB SNR and without noise). Each stimulus was presented 3 times. The total duration of the experiment was 2 h, excluding breaks. To keep the subjects attentive, they could take a break when desired. To keep the participants attentive, we asked questions during the trials without a movie. These questions about the stimuli were presented before and after the presentation of the stimulus. The questions were typically counting tasks, e.g., “How many times did you hear “red balls?” The answers were noted but were not used for further processing.

The participants also listened to the children’s story “Milan,” written and narrated in Flemish by Stijn Vranken (Flemish male speaker). It was approximately 15 min long and was presented without any noise. The purpose was to have a continuous, attended stimulus to train the linear decoder (see below).

### 2.4.2. Attention Conditions

While listening to the Matrix sentences, the participants’ attention was manipulated using two different tasks. The participants were either instructed to (1) attentively listen to the sentences, and respond to questions or (2) watch a silent movie. We instructed the participants not to listen to the auditory stimulus when watching the movie.

In the movie condition, we used a subtitled cartoon movie of choice by the participant to aim for a similar level of distraction across subjects. The text dialogue effectively captures attention while not interfering with auditory processing. A cartoon movie was chosen to avoid realistic lip movements, which may activate auditory brain areas (O’Sullivan et al., 2017). Navarra (2003) investigated if incongruent linguistic lip movements interfere with the perception of auditory sentences, and found that visual, linguistic stimuli produced a greater interference than non-linguistic stimuli. For some subjects  $n = 9$ , we also presented a movie during a story stimulus. This story stimulus was a 15 min excerpt from “De Wilde Zwanen.” It was different from the previously mentioned story stimulus. We used a visual distractor instead of an auditory distractor to avoid potential confounds due

<sup>1</sup>For practical reasons, for 5 participants, we used -9.5 dB SNR, -8.5 dB SNR, -6.5 dB SNR, -5.5 dB SNR and -3.5 dB SNR. For 1 participant, we used -8 dB SNR and -12 dB SNR. For 2 participants, we used -9 dB SNR, -6 dB SNR and -3 dB SNR, and also -4 dB SNR for the last participant.

to the participant's ability to segregate auditory streams. This also limited other confounds of common auditory distractors, such as differences in attention that may occur between male vs. female speakers or speakers in another language.

To further distract the subjects, and direct their full attention to a non-listening task, we also included a condition in which the participants played a computer game while ignoring the auditory stimulus. They were instructed to conduct a visuospatial task, namely to play the computer game Tetris, with the difficulty level set such that they were just able to play the game and had to allocate all of their mental resources to do the task, i.e., the game required maximal effort. They controlled the game using a numeric computer keyboard. This condition was similar to the difficult visuospatial task in Roth et al. (2013).

## 2.5. Signal Processing

All signal processing was implemented in MATLAB R2016b (MathWorks, Natick, USA).

### 2.5.1. Speech

We measured neural envelope tracking by calculating the correlation between the stimulus speech envelope and the envelope reconstructed using a linear decoder.

The speech envelope was extracted from the stimulus according to Biesmans et al. (2017), who investigated the effect of envelope extraction method on auditory attention detection and found the best performance using a gammatone filterbank followed by a power law. In more detail, we used a gammatone filterbank (Søndergaard and Majdak, 2013; Søndergaard et al., 2012) with 28 channels spaced by 1 equivalent rectangular bandwidth, with center frequencies from 50 to 5,000 Hz. From each subband, we extracted the envelope by taking the absolute value of each sample and raising it to the power of 0.6. The resulting 28 subband envelopes were averaged to obtain one single envelope.

To remove glitches that were mostly present in the Tetris condition, we blanked each sample having a higher amplitude than 500  $\mu$ V. These blanked portions were linearly interpolated. To decrease processing time, the EEG data and the envelope were downsampled to 256 Hz from their respective original sample rate 8,192 Hz and 44,100 Hz.

To reduce the influence of ocular artifacts, we used a multi-channel Wiener filter (Somers et al., 2018). This spatial filter estimates the artifacts and subtracts them from the EEG. To build the spatial filter, we applied thresholding on channels Fp1, AF7, AF3, Fpz, Fp2, AF8, AF4, AFz. Samples with power of 5 times higher than the time-averaged power were considered as artifact samples. After artifact suppression, the EEG data were re-referenced to the average of the 64 channels.

The speech envelope and EEG signal were band-pass filtered. We investigated the delta band (0.5–4 Hz) and theta band (4–8 Hz). These two bands encompass the word rate (2.5 Hz) and syllable rate (4.1 Hz) of the Matrix sentences, which are essential for speech understanding (Woodfield and Akeroyd, 2010). The same filter, a zero-phase Chebyshev type 2 filter with 80 dB attenuation at 10% outside the passband, was applied to the EEG and speech envelope. After filtering, the data were further downsampled to 128 Hz.

The decoder linearly combines EEG electrode signals and their time-shifted versions to reconstruct the speech envelope optimally. In the training phase, the weights to be applied to each signal in this linear combination are determined. The decoder was calculated using the mTRF toolbox (version 1.1) (Lalor et al., 2006, 2009) and applied as follows. As the stimulus evokes neural responses at different delays along the auditory pathway, we define a matrix  $R$  containing the shifted neural responses of each channel. With  $g$  the linear decoder and  $R$  the shifted neural data, the reconstruction of the speech envelope  $\hat{s}(t)$  was calculated as

$$\hat{s}(t) = \sum_{n=1}^N \sum_{\tau} g(n, \tau) R(t + \tau, n)$$

with  $t$  the time ranging from 0 to  $T$ ,  $n$  the index of the recording electrode and  $\tau$  the post-stimulus integration window length used to reconstruct the envelope. The matrix  $g$  can be determined by minimizing a least-squares objective function

$$g = \arg \min E(|\hat{s}(t) - s(t)|^2)$$

with  $E$  the expected value,  $s(t)$  the real speech envelope and  $\hat{s}(t)$  the reconstructed envelope. In practice, we calculated the decoder by solving

$$g = (RR^T)^{-1}(RS^T)$$

with  $S$  a vector of stimulus envelope samples. The decoder was calculated using ridge regression on the inverse autocorrelation matrix, the regularization parameter  $\lambda$  was chosen as the maximal absolute value of the autocorrelation matrix. We used post-stimulus lags of 0–75 ms or 0–500 ms. We choose the 0–75 ms as it is less influenced by attention compared to the later responses (Snyder et al., 2006; Ross et al., 2009; Choi et al., 2014; Das et al., 2016; Puvvada and Simon, 2017) and because it yielded best results to estimate speech intelligibility in our earlier study (Vanthornhout et al., 2018). Because other research found attentional modulation at later responses, we also included a 0–500 ms integration window. We trained a new decoder for each subject on the story stimulus, which was 15 min long. After training, the decoder was applied to the EEG responses of the Flemish Matrix material. To measure the correspondence between the speech envelope and its reconstruction, we calculated the bootstrapped Spearman correlation between the real and reconstructed envelope. Bootstrapping was applied by Monte Carlo sampling of the two envelopes. The shown correlations are the median of the bootstraps, the 95% confidence interval always starts 0.018 under the median and ends 0.018 above the median.

We also estimated the temporal response function on the stories and the Matrix sentences in the speech in the quiet condition. This is similar to the calculation of the decoder, but we now predict EEG instead of reconstructing the envelope.

## 3. RESULTS

### 3.1. Behavioral Speech Understanding

For each subject, we fitted a psychometric curve on the SNR vs. percentage correct score and estimated the corresponding SRT

and slope. The average SRT is  $-8.03$  dB SNR (standard deviation:  $1.27$  dB) with an average slope of  $14.1\%/dB$  (standard deviation:  $3.4\%/dB$ ).

## 3.2. Neural Responses to Speech

### 3.2.1. Effect of Attention

**Figure 1** show neural envelope tracking as a function of the SNR using respectively a  $0\text{--}75$  ms and a  $0\text{--}500$  ms integration window in the delta band ( $0.5\text{--}4$  Hz) for the 2 conditions: maximal attention and watching a movie. As we tested a wide variety of SNRs, we clustered the SNRs using k-means clustering in 7 clusters<sup>2</sup>. When multiple correlations per SNR were present for a subject, the average was taken. Visual assessment shows a systematic increase in neural tracking with SNR (related to intelligibility). Indeed, the Spearman correlations between SNR and neural envelope tracking for the attention and movie condition using a  $0\text{--}75$  ms integration window in the decoder are respectively:  $0.38$  ( $p < 0.001$ ) and  $0.52$  ( $p < 0.001$ ). These correlations are significantly different from each other [ $p = 0.02$  (Hittner et al., 2003; Diedenhofen and Musch, 2015)]. Using a  $0\text{--}500$  ms integration window, they are respectively:  $0.36$  ( $p < 0.001$ ),  $0.48$  ( $p < 0.001$ ) and also significantly different ( $p = 0.04$ ). For the theta band ( $4\text{--}8$  Hz), we found a significant correlation between SNR and neural envelope tracking in both windows. However, there was no significant difference between the attention and movie conditions.

### 3.2.2. Effect of SNR

In the delta band, using a  $0\text{--}75$  ms integration window, we found a significant difference in neural envelope tracking between the attention and movie condition at  $-6.9$  dB SNR using a paired permutation test ( $p = 0.007$  after Holm-Bonferroni correction). Using a  $0\text{--}500$  ms integration window, we found a significant difference using a paired permutation test between the attention and movie condition at  $-11.3$  dB SNR,  $-9.2$  dB SNR,  $-6.9$  dB SNR,  $-3.8$  dB SNR and in speech in quiet ( $p = 0.007$ ,  $p = 0.024$ ,  $p = 0.024$ ,  $p = 0.024$ , and  $p = 0.024$  after Holm-Bonferroni correction). In the theta band, we did not find a significant difference at any SNR for both integration windows. As we hypothesized that the movie condition would have a lower variance than the attention condition, we compared the variance of the neural envelope tracking per SNR. However, while such a trend seems to be present when visually assessing the size of the whiskers in the box plots, using a Brown-Forsythe test, we found no significant differences in the spread between attention and movie.

As the attention and movie condition show the expected increase of neural envelope tracking with SNR, we conducted a similar analysis as the behavioral data and attempted to estimate the SRT by fitting a psychometric function on the envelope tracking vs. SNR data across subjects. We used the same psychometric curve as for the behavioral data with the exception that the guess-rate and lapse-rate were not fixed. Across subjects, for the attention condition, we found an SRT of  $-9.15$  dB SNR (95% confidence interval [ $-9.34$ ;  $-8.96$ ] dB SNR). For the watching

condition, we found an SRT of  $-6.96$  dB SNR (95% confidence interval [ $-8.57$ ;  $-5.35$ ] dB SNR). Compare to the behaviorally measured SRT of  $-8.03$  dB SNR.

### 3.2.3. Difference in Topographies

Using a cluster-based analysis (Maris and Oostenveld, 2007), we investigated whether the TRFs of the attention and movie condition were significantly different. It also indicated which latencies and channels contributed most to this difference. A cluster-based approach solves the multiple comparison problem by clustering adjacent lags and channels. No significant difference was found for the story data. The Matrix data, however, yielded a significant difference between the attended and movie condition. **Figure 2** shows the difference in TRFs between the attention and movie condition in speech in quiet for 17 subjects. This difference is most pronounced from  $141$  to  $188$  ms in the right-frontal region. The actual TRFs for this interval are shown in **Figures 3, 4**. For visualization purposes, we averaged the TRFs across participants.

### 3.2.4. Difference Between TRFs

To better understand why the topographies are different, we averaged the channels contributing most to the difference and did a cluster-based analysis on this TRF. **Figure 5** shows the TRF for the attention and the movie condition for the Matrix sentences using a  $0\text{--}500$  ms integration window and unfiltered data. 9 participants listened attentively to a story, while in another condition they listened passively to a story while watching a movie. Seventeen participants did the same for the Matrix sentences. We did a cluster-based analysis to find differences between the TRFs of the two attention conditions (Maris and Oostenveld, 2007). We did not find a significant difference for the story, but for the Matrix sentences, a cluster from  $141\text{--}188$  ms showed a significant difference. The attended TRF shows a P1 peak at  $50$  ms, N1 peak at  $80$  ms and a P2 peak at  $160$  ms. The movie TRF is similar with the exception that the P2 peak is less pronounced.

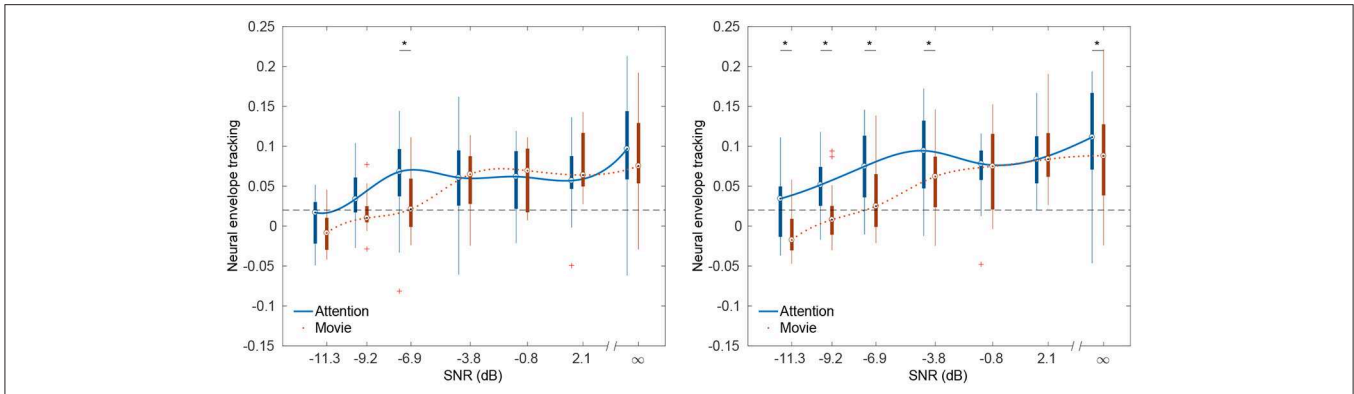
## 3.3. Tetris

While we found some significant neural envelope tracking in the Tetris condition, correlations between actual and reconstructed envelope were very low. For example, in the condition without noise, the median envelope tracking was  $0.044$  with an interquartile range of  $0.10$ . This makes it impossible to draw further conclusions or conduct meaningful statistics. This may be because the resources of the participant were exhausted or because of motion artifacts introduced by playing Tetris, reducing the EEG signal quality.

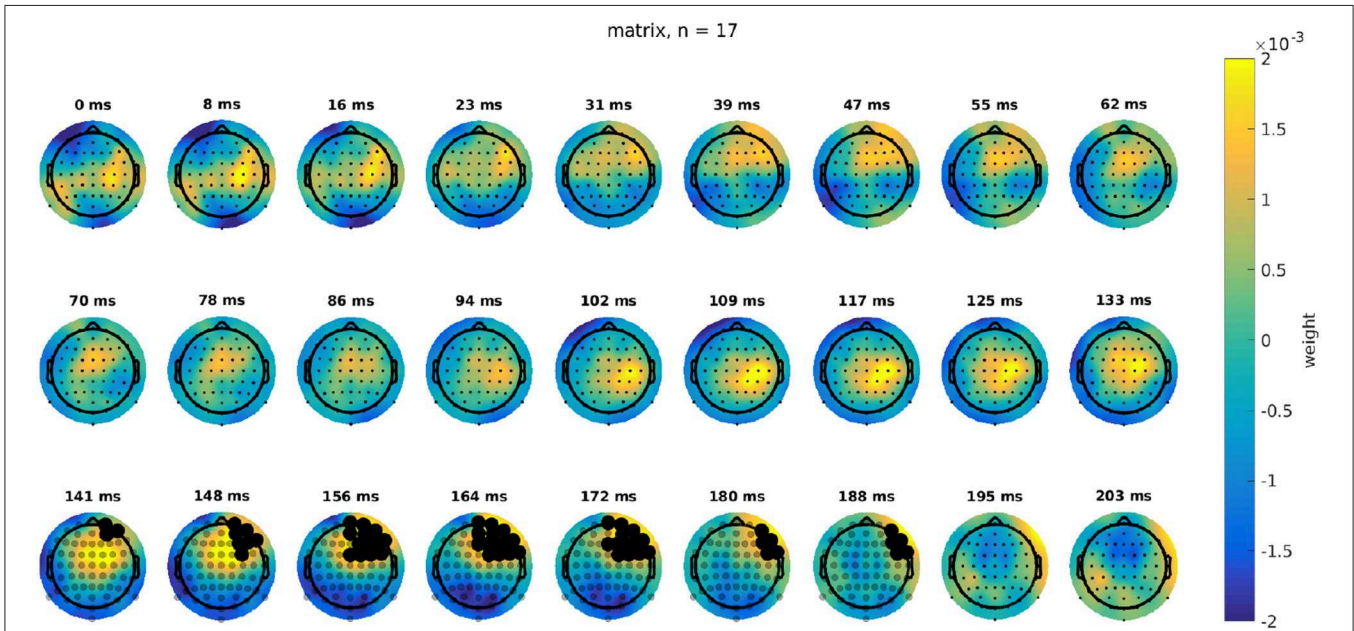
## 4. DISCUSSION

We investigated the effect of attention on neural envelope tracking of a single talker, in conditions of maximal attention and while watching a silent cartoon movie. To gain insight into the neural origin of the differences, we estimated the TRFs and investigated their time course and topography.

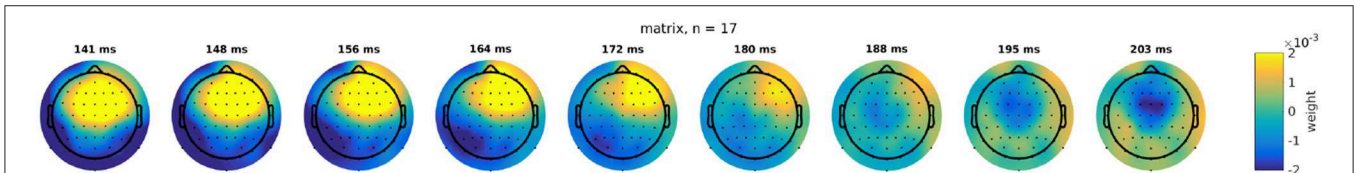
<sup>2</sup>The maximal deviation between the shown SNR and the actual SNR was  $2.5$  dB.



**FIGURE 1 |** Neural envelope tracking as a function of SNR for multiple integration windows. Significant differences in envelope tracking between the attention and movie condition are indicated with a star. The lines are a cubic spline interpolating the medians of each SNR. **(Left)** Neural envelope tracking as a function of SNR in the delta band (0.5–4 Hz) using a 0–75 ms integration window. **(Right)** Neural envelope tracking as a function of SNR in the delta band (0.5–4 Hz) using a 0–500 ms integration window.



**FIGURE 2 |** The difference in topography of the attention condition and the movie condition for the matrix data at different latencies. This was obtained by subtracting the TRFs for the movie and attention conditions. Channels and latencies that were found to contribute significantly to this difference in a cluster-based analysis are indicated with black dots.

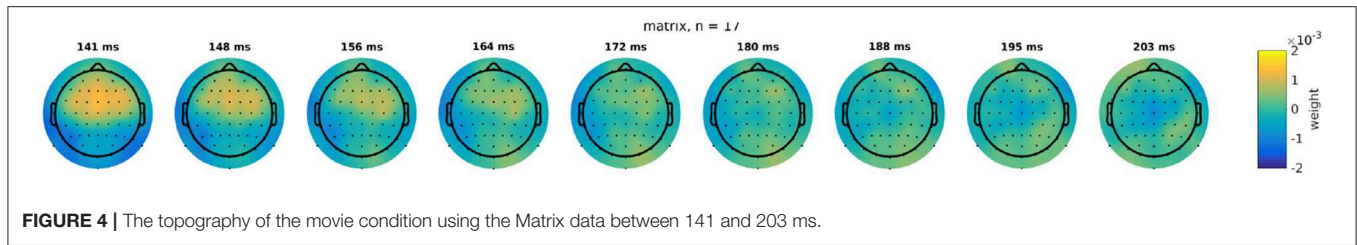


**FIGURE 3 |** The topography of the attention condition using the Matrix data between 141 and 203 ms.

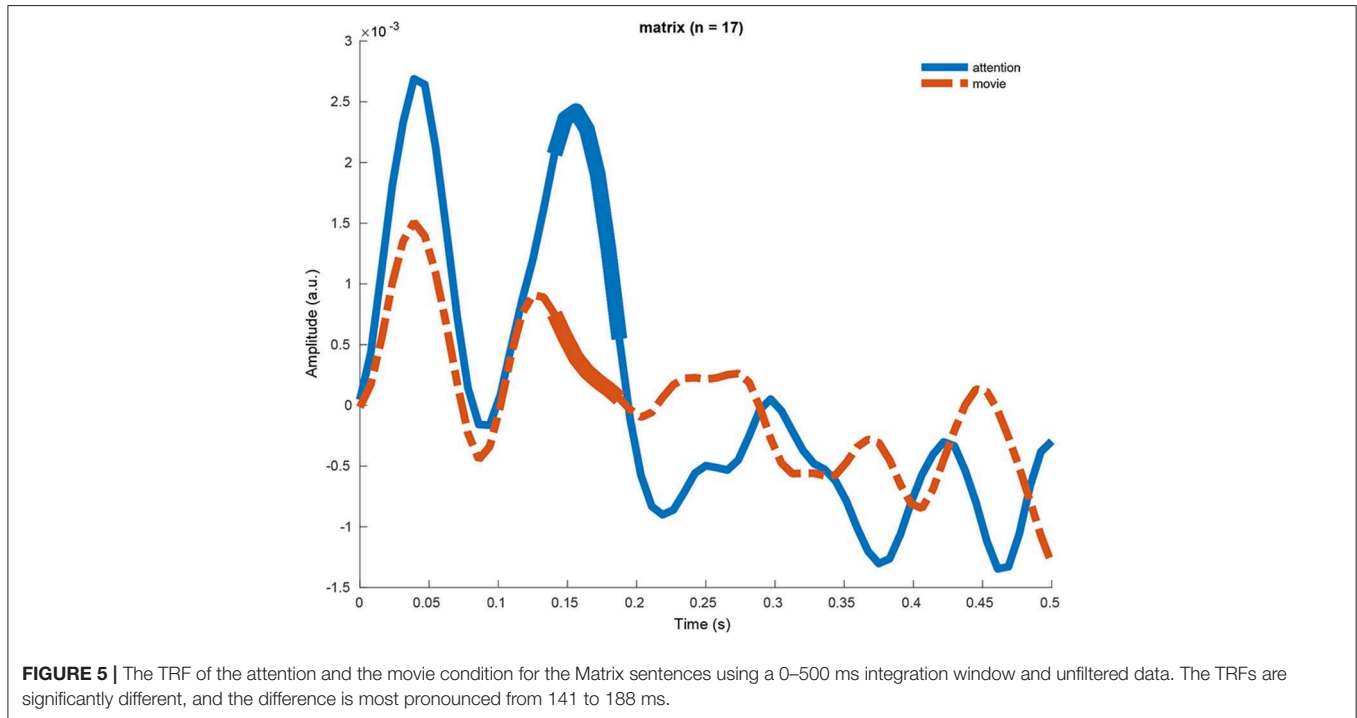
### 4.1. Effect of Task and SNR

At low SNRs, and therefore, low levels of speech intelligibility, we found higher levels of neural envelope tracking in the attention condition than in the movie condition. This is consistent with the

notion that attention requires additional neural resources. The difference in envelope tracking between attention and movie was not visible at higher SNRs, possibly because participants could easily resort to passive listening in these conditions. At low SNRs,



**FIGURE 4 |** The topography of the movie condition using the Matrix data between 141 and 203 ms.



**FIGURE 5 |** The TRF of the attention and the movie condition for the Matrix sentences using a 0–500 ms integration window and unfiltered data. The TRFs are significantly different, and the difference is most pronounced from 141 to 188 ms.

passive listening was not possible, and the extensive resources needed for active listening were not available due to engagement with the movie.

In contrast with our results, Kong et al. (2014) found no significant difference between active and passive listening. This can be explained by the number of participants, the stimulus type and the chosen SNRs. Kong et al. (2014) used 8 participants and a story stimulus, while we have 17 participants for the Matrix condition. The higher amount of subjects gives us more statistical power and Matrix sentences have a very rigid structure compared to a story which may also reduce variance across participants. Also, we measured the attentional effect at more SNRs than Kong et al. (2014). We found that at some SNRs levels, the attentional effect is less prominent. Similar to their results, we did not find a significant difference using the story where we used only 9 subjects.

Apart from the level of neural envelope tracking, its relation with SNR and thus speech intelligibility is also important. We saw a consistent increase in neural envelope tracking with SNR for both the attended condition and the movie condition. Unexpectedly, in the movie condition, the correlation between SNR and neural level tracking was higher than in the attended

condition. We hypothesize that while attention yields higher levels of envelope tracking, it is hard for the subjects to remain fully concentrated during the entire experiment. Therefore, while the movie condition yielded lower envelope tracking levels, it might also have yielded more stable attention and less variable level of envelope tracking. If the level of envelope tracking reflects the stimulus SNR, then a more stable level of envelope tracking would yield a higher correlation between SNR and envelope tracking.

However, we have to acknowledge some confounds in the comparison between the attention and movie conditions. First, the perceptual load was different. In the attention condition, participants have to perform a task while there is no movie playing. In the movie condition, the participants have to watch the movie passively. Second, as the movie was subtitled, some language processing was required, tapping into similar neural resources as for perceiving the speech stimulus. Third, while we used signal processing to remove ocular artifacts, the movie condition may have yielded more ocular activity and thus more ocular artifacts. One could devise an intermediate condition in which the participants have to pay attention to an auditory stimulus while ignoring a visual stimulus or the other way

around. However, in such a scenario, the inhibitory system of the subject is also tested.

We did not show the results of the Tetris experiment as the data were of poor quality, probably due to movement artifacts. However, if a viable way can be found to deal with artifacts, a similar experiment as in Roth et al. (2013) would be very interesting. Compared to the other two conditions, in Tetris, all mental resources can be exhausted, and it is possible that speech understanding is not possible anymore. If neural envelope tracking is still present, it should merely be a bottom-up, passive representation of the stimulus. However, it is entirely possible that no neural envelope tracking is present as all mental resources are exhausted (Hein et al., 2007). This could be another reason why our Tetris data seemed of poor quality.

## 4.2. Neural Sources Related to Attention

The literature shows that attentional modulation has been found in the parietal cortex and prefrontal cortex (Yantis, 2008; Lavie, 2010). To investigate this, we assessed the spatial maps of the TRFs for both the attention and movie condition. We found that the difference between the attended TRF and the movie TRF was most pronounced in the right frontal region, which is supported by other research (Alho, 2007).

## 4.3. Latency

In the literature, using the competing talker paradigm, responses earlier than 75 ms are minimally affected by attention while later responses much more so (Snyder et al., 2006; Ross et al., 2009; Choi et al., 2014; Das et al., 2016; Puvvada and Simon, 2017). Using the forward model, we found a significant difference between the attention and movie TRF. This difference was most pronounced from 141 ms to 188 ms. This is in line with the competing talker literature. Moreover, the attended TRF morphology is similar to LAEP literature, showing a P1 peak at 50 ms, N1 peak at 80 ms and a P2 peak at 160 ms. The movie TRF is similar with the exception that the P2 peak is less pronounced. Although the TRFs in quiet at earlier latencies (< 75 ms) does not significantly differ between attention and movie TRF, we still find a significant difference in correlation at some low SNRs.

## 4.4. Distractors

When studying the effect of attention on neural envelope tracking, the difficulty is that there is no reference for the level of attention. Therefore one needs to resort to, on the one hand, ways to motivate the subject to focus solely on the task at hand (attentively listening to sentences), or on the other hand, distract the participants from the task in a controlled way. The latter can be achieved by instructing the participants to ignore the stimulus and providing another task to distract them.

There is a body of research on attention and distraction in auditory and visual tasks. However, most of the literature is focused on how various secondary tasks can distract from a primary task (Murphy et al., 2017; Lavie, 2010, 2005).

We chose one distractor task: a subtitled cartoon movie. The benefit is that it is engaging and relaxing for the participant. A potential downside is that reading the subtitles requires language

processing, which is a resource also required for processing the auditory stimulus. Subtitles are a distractor, but maybe they only have a small perceptual load. Another downside is that a movie does not exhaust the available attentional resources of the participants. However, this is difficult to enforce and to quantify how well the participants did.

Murphy et al. (2017) reviewed the literature regarding perceptual load in the auditory domain. A key point of perceptual load theory is that it proposes that our perceptual system has limited processing capacity and that it is beyond our volitional control as to how much of that capacity will be engaged at any given time. Instead, all of the available information is automatically processed until an individual's perceptual capacity is exhausted. While this theory is confirmed by many experiments in the purely visual domain, it is less clear in the auditory domain. It seems that the auditory system, with its complicated auditory stream segregation (Shinn-Cunningham, 2008), allows for a less strong selection mechanism. While generally, unattended information receives less or no processing, this is not consistently the case. The envelope tracking results with the competing talker paradigm (see above) confirm this, and this is also consistent with our results. In the movie TRF, we see that the auditory stimulus is still processed by the brain. However, later processing (from 140 ms on) is reduced.

A concern with using concurrent auditory stimuli to distract from the stimulus of interest is that auditory stream segregation might fail, especially if the listener has an unknown hearing deficit. Therefore, a distractor in a non-auditory domain may be desirable. There is some evidence that perceptual load theory also holds across modalities. Macdonald and Lavie (2011), Raveh and Lavie (2015) demonstrated inattentive deafness (failure to notice an auditory stimulus) under visual load, thereby extending the load theory of attention across the auditory and visual modalities, making it clear that vision and audition share a common processing resource, which is consistent with our results. In the movie condition, we saw no P2 peak in the TRF. Also at low SNRs (< -6.6 dB SNR), we found higher neural envelope tracking in the attention condition compared with the movie condition.

## 4.5. Implications for Applied Research—Neural Envelope Tracking as a Measure of Speech Understanding

The stimulus reconstruction method is promising to obtain an objective measure of speech understanding for applications in diagnostics of hearing. Ding and Simon (2013) found a significant correlation between intelligibility and reconstruction accuracy at one signal stimulus SNR. Vanthornhout et al. (2018) and Lesenfants et al. (2019) developed a clinically applicable method to objectively estimate the SRT based on reconstruction accuracy and found a significant correlation between predicted and actual speech reception threshold. In the current study, the average SRT using a behavioral test was -8.0 dB SNR, using the stimulus reconstruction method we found an SRT of -9.2 dB SNR in the attention condition and an SRT of -7.0 dB SNR in the



movie condition. Both objective methods are thus close to the behavioral method. The rank order of these SRT values also makes intuitive sense: when paying maximal attention, and estimating the SRT based on envelope tracking, the lowest value (best performance) is obtained. In the behavioral experiment, the subjects' neural resources are taxed more because they need to repeat each sentence, leading to a slightly higher SRT. Finally, when not paying attention to the stimulus, the SRT reaches its highest value.

We found that neural envelope tracking was better correlated with SNR in the movie condition, which is important for the estimation of the SRT. While the level of attention influenced the level of neural tracking, to obtain an objective measure of speech understanding it may be more important to have stable attention and thus lower variability than to have high levels of neural envelope tracking. Equalizing the level of attention across subjects and conditions by giving them an unrelated and/or easy task, such as watching a movie might, therefore, yield the best results.

In addition to procedural benefits, an additional benefit of letting subjects watch a movie instead of attending to low-content sentences is that it is much more pleasant for the participants and therefore easier to implement in the clinic, especially in populations such as children.

## 5. CONCLUSION

Using a movie, we varied the level of attention while we estimated the neural envelope tracking. Neural envelope tracking was significantly different between the condition with maximal attention and the movie condition. This difference was most pronounced in the right-frontal region of the brain. However, this does not seem to be a problem for estimating speech understanding as neural envelope tracking was still highly correlated with the stimulus SNR, even more in the movie condition.

## REFERENCES

- Alho, K. (2007). Brain activity during selective listening to natural speech. *Front. Biosci* 12:3167. doi: 10.2741/2304
- Astheimer, L. B., and Sanders, L. D. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biol. Psychol.* 80, 23–34. doi: 10.1016/j.biopsycho.2008.01.015
- Bidet-Caulet, A., Mikyska, C., and Knight, R. T. (2010). Load effects in auditory selective attention: evidence for distinct facilitation and inhibition mechanisms. *Neuroimage* 50, 277–284. doi: 10.1016/j.neuroimage.2009.12.039
- Biesmans, W., Das, N., Francart, T., and Bertrand, A. (2017). Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 402–412. doi: 10.1109/TNSRE.2016.2571900
- Braiman, C., Fridman, E. A., Conte, M. M., Voss, H. U., Reichenbach, C. S., Reichenbach, T., et al. (2018). Cortical response to the natural speech envelope correlates with neuroimaging evidence of cognition in severe brain injury. *Curr. Biol.* 28, 3833–3839.e3. doi: 10.1016/j.cub.2018.10.057

## DATA AVAILABILITY

The datasets for this manuscript are not publicly available because this is not allowed by the local medical ethics committee. Requests to access the datasets should be directed to TF (tom.francart@med.kuleuven.be).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Medical Ethics Committee UZ/KU Leuven. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JV, LD, and TF designed the experiments and analyzed the neural data. JV and LD recorded the neural data. JV and TF wrote the manuscript. All authors commented on the manuscript.

## FUNDING

Financial support was provided by the KU Leuven Special Research Fund under grant OT/14/119 to TF. Research funded by a Ph.D. grant of the Research Foundation Flanders (FWO). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 637424), from the YouReCa Junior Mobility Programme under grant JUMO/15/034 to JV.

## ACKNOWLEDGMENTS

The authors thank Lise Goris, Lauren Commers, Paulien Vangompel, and Lies Van Dorpe for their help with the data acquisition.

- Brand, T. and Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J. Acoust. Soc. Am.* 111, 2801–2810. doi: 10.1121/1.1479152
- Brix, R. (1984). The influence of attention on the auditory brain stem evoked responses preliminary report. *Acta Oto-Laryngol.* 98, 89–92. doi: 10.3109/00016488409107538
- Brodbeck, C., Hong, L. E., and Simon, J. Z. (2018). Transformation from auditory to linguistic representations across auditory cortex is rapid and attention dependent for continuous speech. 28, 3976–3983. doi: 10.1101/326785
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229
- Choi, I., Wang, L., Bharadwaj, H., and Shinn-Cunningham, B. (2014). Individual differences in attentional modulation of cortical responses correlate with selective attention performance. *Hear. Res.* 314, 10–19. doi: 10.1016/j.heares.2014.04.008
- Das, N., Bertrand, A., and Francart, T. (2018). EEG-based auditory attention detection: boundary conditions for background noise and speaker positions. *J. Neural Eng.* 15:066017. doi: 10.1088/1741-2552/aae0a6

- Das, N., Biesmans, W., Bertrand, A., and Francart, T. (2016). The effect of head-related filtering and ear-specific decoding bias on auditory attention detection. *J. Neural Eng.* 13:056014. doi: 10.1088/1741-2560/13/5/056014
- Diedenhofen, B., and Musch, J. (2015). Cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS ONE* 10:e0121945. doi: 10.1371/journal.pone.0121945
- Ding, N., and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854–11859. doi: 10.1073/pnas.1205381109
- Ding, N., and Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33, 5728–5735. doi: 10.1523/JNEUROSCI.5297-12.2013
- Forste, A. E., Etard, O., and Reichenbach, T. (2017). The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *eLife* 6:e27203. doi: 10.7554/eLife.27203
- Francart, T., van Wieringen, A., and Wouters, J. (2008). APEX 3: a multi-purpose test platform for auditory psychophysical experiments. *J. Neurosci. Methods* 172, 283–293. doi: 10.1016/j.jneumeth.2008.04.020
- Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Auditory attention—focusing the searchlight on sound. *Curr. Opin. Neurobiol.* 17, 437–455. doi: 10.1016/j.conb.2007.07.011
- Galbraith, G. C., Olfman, D. M., and Hu, T. M. (2003). Selective attention affects human brain stem frequency-following response. *Neuroreport* 4, 735–738. doi: 10.1097/00001756-200304150-00015
- Hein, G., Alink, A., Kleinschmidt, A., and Müller, N. G. (2007). Competing neural responses for auditory and visual decisions. *PLoS ONE* 2:e320. doi: 10.1371/journal.pone.0000320
- Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 177–180. doi: 10.1126/science.182.4108.177
- Hittner, J. B., May, K., and Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *J. Gen. Psychol.* 130, 149–168. doi: 10.1080/00221300309601282
- Hoormann, J., Falkenstein, M., and Hohnsbein, J. (2000). Early attention effects in human auditory-evoked potentials. *Psychophysiology* 37:14. doi: 10.1111/1469-8986.3710029
- Kong, Y.-Y., Mullangi, A., and Ding, N. (2014). Differential modulation of auditory responses to attended and unattended speech in different listening conditions. *Hear. Res.* 316, 73–81. doi: 10.1016/j.heares.2014.07.009
- Lalor, E. C., Pearlmutter, B. A., Reilly, R. B., McDarby, G., and Foxe, J. J. (2006). The vespa: a method for the rapid estimation of a visual evoked potential. *Neuroimage* 32, 1549–1561. doi: 10.1016/j.neuroimage.2006.05.054
- Lalor, E. C., Power, A. J., Reilly, R. B., and Foxe, J. J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J. Neurophysiol.* 102, 349–359. doi: 10.1152/jn.90896.2008
- Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends Cogn. Sci.* 9, 75–82. doi: 10.1016/j.tics.2004.12.004
- Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Curr. Dir. Psychol. Sci.* 19, 143–148. doi: 10.1177/0963721410370295
- Lehmann, A., and Schönwiesner, M. (2014). Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. *PLoS ONE* 9:e85442. doi: 10.1371/journal.pone.0085442
- Lesenfants, D., Vanthornhout, J., Verschuere, E., Decruy, L., and Francart, T. (2019). Predicting individual speech intelligibility from the neural tracking of acoustic- and phonetic-level speech representations. *Hear. Res.* 380, 1–9. doi: 10.1101/471367
- Linden, R. D., Picton, T. W., Hamel, G., and Campbell, K. B. (1987). Human auditory steady-state evoked potentials during selective attention. *Electroencephalogr. Clin. Neurophysiol.* 66, 145–159. doi: 10.1016/0013-4694(87)90184-2
- Luts, H., Jansen, S., Dreschler, W., and Wouters, J. (2015). *Development and Normative Data for the Flemish/Dutch Matrix Test*. Technical report.
- Macdonald, J. S. P., and Lavie, N. (2011). Visual perceptual load induces inattentive deafness. *Attent. Percept. Psychophys.* 73, 1780–1789. doi: 10.3758/s13414-011-0144-4
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Melara, R. D., Rao, A., and Tong, Y. (2002). The duality of selection: excitatory and inhibitory processes in auditory selective attention. *J. Exp. Psychol.* 28:279. doi: 10.1037/0096-1523.28.2.279
- Mirkovic, B., Debener, S., Jaeger, M., and De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: Implications for online, daily-life applications. *J. Neural Eng.* 12:046007. doi: 10.1088/1741-2560/12/4/046007
- Müller, N., Schlee, W., Hartmann, T., Lorenz, I., and Weisz, N. (2009). Top-down modulation of the auditory steady-state response in a task-switch paradigm. *Front. Hum. Neurosci.* 3:1. doi: 10.3389/fnhum.09.001.2009
- Murphy, S., Spence, C., and Dalton, P. (2017). Auditory perceptual load: a review. *Hear. Res.* 352, 40–48. doi: 10.1016/j.heares.2017.02.005
- Näätänen, R., Kujala, T., and Winkler, I. (2011). Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology* 48, 4–22. doi: 10.1111/j.1469-8986.2010.01114.x
- Navarra, J. (2003). “Visual speech interference in an auditory shadowing task: the dubbed movie effect,” in *Proceedings of the 15th International Congress of Phonetic Sciences* (Barcelona), 1–4.
- O’Sullivan, A. E., Crosse, M. J., Di Liberto, G. M., and Lalor, E. C. (2017). Visual cortical entrainment to motion and categorical speech features during silent lipreading. *Front. Hum. Neurosci.* 10:679. doi: 10.3389/fnhum.2016.00679
- O’Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355
- Picton, T. W., and Hillyard, S. A. (1974). Human auditory evoked potentials. II: effects of attention. *Electroencephalogr. Clin. Neurophysiol.* 36, 191–200. doi: 10.1016/0013-4694(74)90156-4
- Puvvada, K. C., and Simon, J. Z. (2017). Cortical representations of speech in a multitalker auditory scene. *J. Neurosci.* 37, 9189–9196. doi: 10.1523/JNEUROSCI.0938-17.2017
- Raveh, D., and Lavie, N. (2015). Load-induced inattentive deafness. *Attent. Percept. Psychophys.* 77, 483–492. doi: 10.3758/s13414-014-0776-2
- Regenbogen, C., Vos, M. D., Debener, S., Turetsky, B. I., Mößnang, C., Finkelmeyer, A., et al. (2012). Auditory processing under cross-modal visual load investigated with simultaneous EEG-fMRI. *PLoS ONE* 7:e52267. doi: 10.1371/journal.pone.0052267
- Ross, B., Hillyard, S. A., and Picton, T. W. (2009). Temporal dynamics of selective attention during dichotic listening. *Cereb. Cortex* 20, 1360–1371. doi: 10.1093/cercor/bhp201
- Ross, B., Picton, T. W., Herdman, A. T., and Pantev, C. (2004). The effect of attention on the auditory steady-state response. *Neurol. Clin. Neurophysiol.* 2004:22.
- Roth, C., Gupta, C. N., Plis, S. M., Damaraju, E., Khullar, S., Calhoun, V., et al. (2013). The influence of visuospatial attention on unattended auditory 40 Hz responses. *Front. Hum. Neurosci.* 7:370. doi: 10.3389/fnhum.2013.00370
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends Cogn. Sci.* 12, 182–186. doi: 10.1016/j.tics.2008.02.003
- Skosnik, P. D., Krishnan, G. P., and O’Donnell, B. F. (2007). The effect of selective attention on the gamma-band auditory steady-state response. *Neurosci. Lett.* 420, 223–228. doi: 10.1016/j.neulet.2007.04.072
- Snyder, J. S., Alain, C., and Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* 18, 1–13. doi: 10.1162/089892906775250021
- Somers, B., Francart, T., and Bertrand, A. (2018). A generic EEG artifact removal algorithm based on the multi-channel Wiener filter. *J. Neural Eng.* 15:036007. doi: 10.1088/1741-2552/aaac92

- Søndergaard, P., and Majdak, P. (2013). "The auditory modeling toolbox," in *The Technology of Binaural Listening*, ed J. Blauert (Berlin; Heidelberg: Springer), 33–56.
- Søndergaard, P. L., Torrésani, B., and Balazs, P. (2012). "The linear time frequency analysis toolbox" in *International Journal of Wavelets, Multiresolution Analysis and Information Processing, Vol. 10*, ed J. Blauert (Berlin; Heidelberg: Springer), 33–56.
- Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11, 1004–1006. doi: 10.1038/nn.2163
- Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., and Francart, T. (2018). Speech intelligibility predicted from neural entrainment of the speech envelope. *J. Assoc. Res. Otolaryngol.* 19, 181–191. doi: 10.1101/246660
- Varghese, L., Bharadwaj, H. M., and Shinn-Cunningham, B. G. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Res.* 1626, 146–164. doi: 10.1016/j.brainres.2015.06.038
- Woodfield, A., and Akeroyd, M. A. (2010). The role of segmentation difficulties in speech-in-speech understanding in older and hearing-impaired adults. *J. Acoust. Soc. Am.* 128, EL26–EL31. doi: 10.1121/1.3443570
- Yantis, S. (2008). The neural basis of selective attention. *Curr. Dir. Psychol. Sci.* 17, 86–90. doi: 10.1111/j.1467-8721.2008.00554.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Vanthornhout, Decruy and Francart. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.