Optimal time lags for linear cortical auditory attention detection: differences between speech and music listening

Simon, Adèle Maryse Danièle; Østergaard, Jan; Bech, Søren; Loquet, Gérard Sylvian Jean Marie

Published in:
Proceedings International Symposium on Hearing

DOI (link to publication from Publisher):
10.5281/zenodo.6576990

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Optimal time lags for linear cortical auditory attention detection: differences between speech and music listening

Adèle Simon[1,2,*], Jan Østergaard[1], Søren Bech[1,2], and Gérard Loquet[3,4]

[1]Department of Electronic Systems, Aalborg University, Denmark
[2]Bang & Olufsen, Struer, Denmark
[3]Bionic Institute, Merlbourne, Australia
[3]Department of Clinical Medicine, Aalborg University, Denmark
[*]Corresponding author: amds@es.aau.dk

Abstract

In recent decades, there has been a lot of interest in detecting auditory attention from brain signals. Cortical recordings have been demonstrated to be useful in determining which speaker a person is listening to a mixed variety of sounds (the cocktail party effect). Linear regression, often called the stimulus reconstruction method, shows that the envelope of the sounds heard can be reconstructed from continuous electroencephalogram recordings (EEG). The target sound, to which the listener is paying attention, can be reconstructed to a greater extent compared to other sounds present in the sound scene, which can allow attention decoding. Reconstruction can be obtained with EEG signals that are delayed compared to the audio signal, to consider the time for neural processing. It can be used to identify latencies where the reconstruction is optimal, which reflects a cortical process specific to the type of audio heard. However, most of these studies used only speech signals and did not investigate other types of auditory stimuli, such as music.

In the present study, we applied this stimulus reconstruction method to decode auditory attention in a cocktail party scenario that includes both speech and music. Participants were presented with a target sound (either speech or music) and a distractor sound (either speech or music) while continuously recording their cortical response during the listening with a 64-channels EEG system. From these recordings, we reconstructed the envelope of the stimuli, both target and distractor, by using linear ridge regression decoding models at individual time lags. Results showed different time lags for maximal reconstruction accuracies between music and speech listening, suggesting separate underlying cortical processes. Results also suggest that an attentional aspect can influence the reconstruction accuracy for middle/late time-lags.

1 Introduction

The world is composed of complex auditory scenes, where several sources of sounds coexist simultaneously, such as noise, speech, or music, and a listener can actively attend to one of the auditory streams. For instance, when sitting in a cafe, with people talking and background music, one can choose to focus on a conversation or to follow the music (Cherry, 1953). Separating and tracking individual sound streams from a complex sound scene is possible thanks to selective auditory attention.

An effect of selective attention is reflected in the cortical signal of the listener. Several studies recorded continuous neural response of listener presented with two or more sounds, with electroencephalogram or magnetoencephalogram: results showed that the cortical response track the attended sound stream better than an ignored sound stream (Ding & Simon, 2012a, 2012b; O'sullivan et al., 2015; Schäfer et al., 2018). Using this effect and an approach called stimulus reconstruction method, it has been shown that auditory attention can be decoded from continuous neural recording (O'sullivan et al., 2015). This approach uses linear filters, computed using least-squares optimization, to reconstruct the sound heard by the listener from the cortical recording (Alickovic et al., 2019). This stimulus reconstruction method have be shown to be sensitive to auditory attention for dichotic speech listening (Fuglsang, Dau, & Hjortkjær, 2017; Mirkovic et al., 2015; O'sullivan et al., 2015), and also during music listening (An et al., 2021; Cantisani, Essid, & Richard, 2019; Hausfeld et al., 2021).

When attempting to decode auditory attention with the stimulus reconstruction approach, most studies use multi-lag models to take into account cortical processing time (Di Liberto, O'Sullivan, & Lalor, 2015). In such a multi-lags model, the model is trained and evaluated on a combination of EEG recording at different time lags (e.g., 0 to 500 ms), relative to the stimulus. While using the multi-lags model can enhance the model prediction and the performance of an auditory attention decoder, it does not allow for investigating the reconstruction performance of individual time lags, which can give information on the temporal neural processing of the sound signal (Crosse et al., 2021). A single-lags model can be used, to gain insight into the reconstruction accuracy at each time lag. This can help to compare neural processes for different types of signals, or conditions (Alickovic et al., 2021; Hausfeld et al., 2018). This method can also be used to explore the effect of the attentional state of the listener on their cortical response: training such models to either reconstruct the target stimulus or the reconstructed stimulus can give information on the effect of attention (O'sullivan et al., 2015). Investigating the individual time-lags to find an optimal value that enhances stimulus reconstruction can help to gain insight into the cortical processes involved in speech and music listening. It can also provide useful information to design an auditory attention decoder, which could be fitted to either music or speech listening and enhance the performance of such an auditory attention decoder.

In the present study, we used stimulus reconstruction methods with single lags models to explore differences between cortical processing of music and cortical processing of speech. Subsequently, we compare target-trained and distractor-trained models, for both speech and music listening, to identify time lags affected by auditory attention.

2 Methods

2.1 Participants

For this study, 35 participants (14 female) were recruited, aged between 21- and 33-year-old (mean = 26,29). Participants did not report any hearing disorders or neurological disorders among the participants. Three of the participants were native English speakers, and all the others were fluent, with education or work experience in English. Written informed consent was obtained and participants were compensated for their participation in the study. Due to poor data quality, EEG recordings from two participants were excluded after recording.

2.2 Procedure and Stimuli

For each trial of one minute, the participant was exposed to two separate sound streams originating from separate loudspeakers, placed in front of her/him (+/- 30° azimuth). The direction of arrival of the target sound (left or right loudspeaker) was randomly selected for each trial The participant was instructed to pay attention to one of the sounds (target) while ignoring the other sound (distractor) the target may be either speech or music. During listening, the subject was instructed to keep their eyes fixed on a crosshair and to minimize blinks and movements. There were four categories of stimuli employed, split into two types (music and speech), with each type further subdivided into two genres.

- Piano Music: 8 excerpts of mono instrumental pieces played on a piano
- Electronic music: 8 excerpts of polyphonic pieces of instrumental electronic music
- Speech female: 8 excerpts of an audiobook read by a woman in English
- Speech male: 8 excerpts of an audiobook read by a man in English

In the same trial, the target and the distractor could have been both music, both speech, or one of each type. Each excerpt was used as a target just once. distractors were selected to obtain a balanced number of trials across conditions (Music/Speech, Music/Music, Speech/Speech, Speech/Music). Participants completed 32 one-minute trials. For each participant the experiment was conducted in a single session.

2.3 Data collection and pre-processing

A 64-channel g.HIamp-Research system was used to record continuous EEG data at 512 Hz (g.tec Medical engineering GmbH, Austria). The electrodes were placed on the scalp in accordance with the international 10-20 system. The impedance of each electrode was kept below 5 kOhms.

After data collection, pre-processing of the data was carried out using EEGLAB v2021.1 (Delorme & Makeig, 2004). The EEG data were referenced to the average of all scalp electrodes. The noise-contaminated EEG channels were visually evaluated and interpolated from neighbouring electrodes. Independent Component Analysis (ICA) was performed in EEGLAB, and the automatic identification plugin allowed the artefacts associated with eye blinks or eye movements to be removed (Pion-Tonachini, Kreutz-Delgado, & Makeig, 2019). The envelopes of the sound signal, both target and distractor, were extracted using a Hilbert transform. Both EEG data and audio envelopes were finally bandpass filtered between 1 and 8Hz and downsampled to a 64Hz sampling rate.

2.4 Stimulus reconstruction

We used a classic stimulus reconstruction approach to decode auditory attention from the EEG data (Alickovic et al., 2019; Crosse et al., 2021; O'sullivan et al., 2015). he EEG data is utilized to reconstruct an estimation of the input stimuli using a linear reconstruction l. This model relates EEG-measured brain activity to the stimulus envelope as follows:

$$s'(t) = \sum_{n} \sum_{\tau} g(1, n) R(t, n) \tag{1}$$

where s' is the reconstructed envelope, $R(t, n)$ is the EEG response at time t for electrode n, and g is the linear model, which is a function of electrode n.

The model g can be estimated by minimizing the mean squared error between the original and the reconstructed envelopes, which can be solved analytically using ridge regularization methods (Wong et al., 2018):

$$g = (R^T R + I\lambda)^{-1} R^T S \tag{2}$$

where I is the identity matrix, S is the stimulus envelope, and λ is the regularization parameter used to prevent overfitting (Alickovic et al., 2019; Wong et al., 2018). The regularization factor was set to 10^5. This value was chosen by calculating several models with different values of this regularization parameter. The value that produced the highest reconstruction accuracy (measured by the Pearson's correlation coefficient between the original and the reconstructed envelope) was used for the analysis.

We calculated the Pearson's r, or correlation coefficient, between the original target envelope and the reconstructed one (r_{target}) to assess reconstruction accuracy. To obtain r_{target}, a "Target model" was trained by using EEG signals and the original envelope of the target. To assess the processes that are encountered by the distractor stimulus, a "Distractor model" was also trained, with EEG signals and the envelope of the distractor. The obtained reconstructed distractor is then compared to the original distractor to obtain the reconstruction accuracy of the distractor, $r_{distractor}$.

2.5 Single-lag model

To explore variation across time lags, several models have been trained on each individual time lag. The models were trained using the original envelope of the sound stimulus and the corresponding EEG data from the specific time lags. For example, to compute a model g_{30} for a time lag of approximately 30 ms, which corresponds to a time lag of 2 samples at the sampling rate of 64 Hz (see Figure 1), we used original envelope S and time-lags EEG R as follows:

$$S = \begin{bmatrix} s(0) \\ s(1) \\ s(2) \\ \vdots \\ s(t) \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} r_1(2) & \cdots & r_64(2) \\ r_1(3) & \cdots & r_64(3) \\ \vdots & \ddots & \vdots \\ r_1(T) & \cdots & r_64(T) \\ r_1(T+1) & \cdots & r_64(T+1) \\ r_1(T+2) & \cdots & r_64(T+2) \end{bmatrix}$$

We computed models to covert times lags ranging from 0 ms to 500 ms, at a sample rate of 64 Hz. That corresponds to thirty-three individual single-lag models, separated by an interval of 15.625 ms.

All models were trained in a leave-one-out approach, which means that each trial was tested on a model created by averaging the parameters of the models trained on every other trial.

Four categories of single lags models were trained:

- Models optimized for music as a Target, where only trials where the target of attention was music are used for training and testing, and by using the Target envelope for training

- Models optimized for speech as a Target, where only trials where the target of attention was speech are used for training and testing, and by using the Target envelope for training
- Models optimized for music as a distractor, where only trials where the distractor was music are used for training and testing, and by using the distractor envelope for training
- Models optimized for speech as a distractor, where only trials where the distractor was speech are used for training and testing, and by using the distractor envelope for training

3 Results

3.1 Differences between speech and music listening

Figure 2 shows the reconstruction accuracy across different time lags for both trials where the target of attention was music stimulus (music listening) and trials where the target of attention was speech stimulus. Shaded areas correspond to the 95% confidence interval for the reconstruction accuracies. The reconstruction accuracies for speech were obtained similarly, but by testing all speech-target trials on models trained on speech-target trials. In Figure 2, there is a clear difference between reconstruction accuracies for speech and music. For all time lags, the reconstruction accuracy for speech is significantly higher than the reconstruction accuracies for music (permutation test speech vs. music, p¡0.05 for each time lag).

The second thing that stands out in Figure 2 is the pattern in variation of the reconstruction accuracies. For both speech and music, we can observe two peaks of increased reconstruction accuracy across the different time lags: a first peak at an early time lag and a second, larger

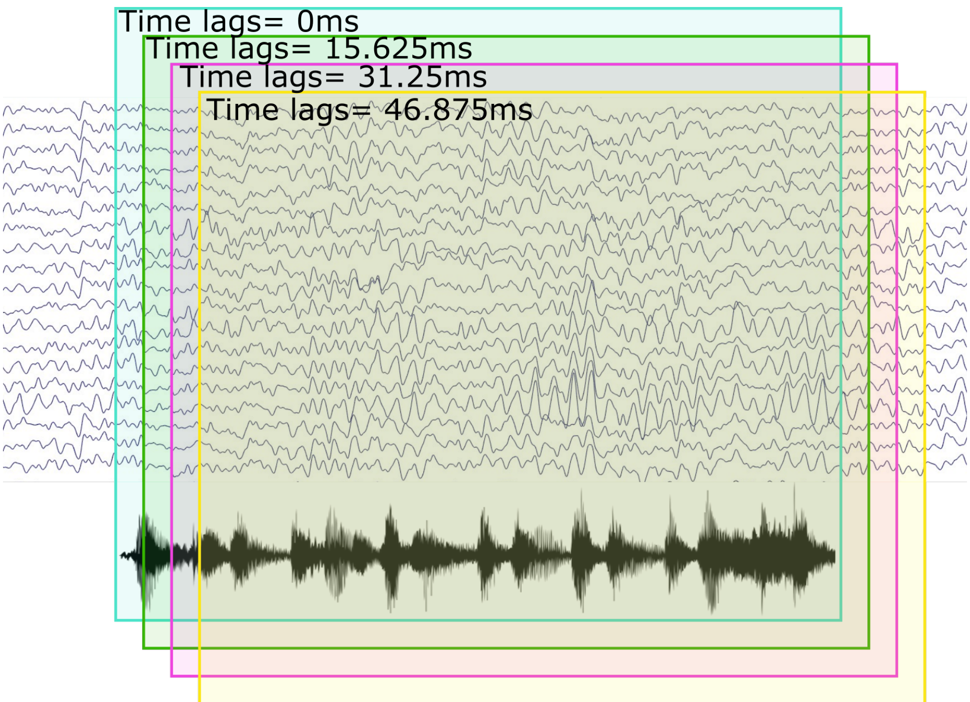

Figure 1 — Schematic of the EEG data selection used for each single lag model. Each model is trained based on the 1 minute of audio data, and 1 minute of EEG data delayed compared to the audio data.

peak at a later time lag. The first peak is located at a time lag comprised of between 30 and 50 ms, and the timing of this first peak is similar for both speech and music. However, the timing of the second peak varies between speech and music: ≈170 ms for speech and ≈265 ms for music. This difference in time lags between speech and music could indicate time process differences for speech and music sounds. aximized reconstructions for speech corroborates results preciously obtained in other studies: O'sullivan et al. (2015) describe a two peaks pattern, with increased reconstruction accuracy and increased decoding accuracy for the interval of 170-250 ms; Alickovic et al. (2021) show an increase of reconstruction accuracy when using late EEG response ; Mirkovic et al. (2015) found increased decoding accuracy for times lags between 130 to 220 ms; Wöstmann, Fiedler, and Obleser (2017)'s results showed an increase of cross-correlation between the envelope of the sound signal and M/EEG data at 80 ms, followed by a second peak of increased correlation. For music, the current results can be compared with results obtained by Hausfeld et al. (2018), where an early peak was also observed at -10 to 30 ms. Hausfeld et al. (2018) also observed a second peak of increased reconstruction accuracies at late latencies. However, they found this peak happening for time lags comprise between 460 and 500 ms, which is later than what we observe in the current study.

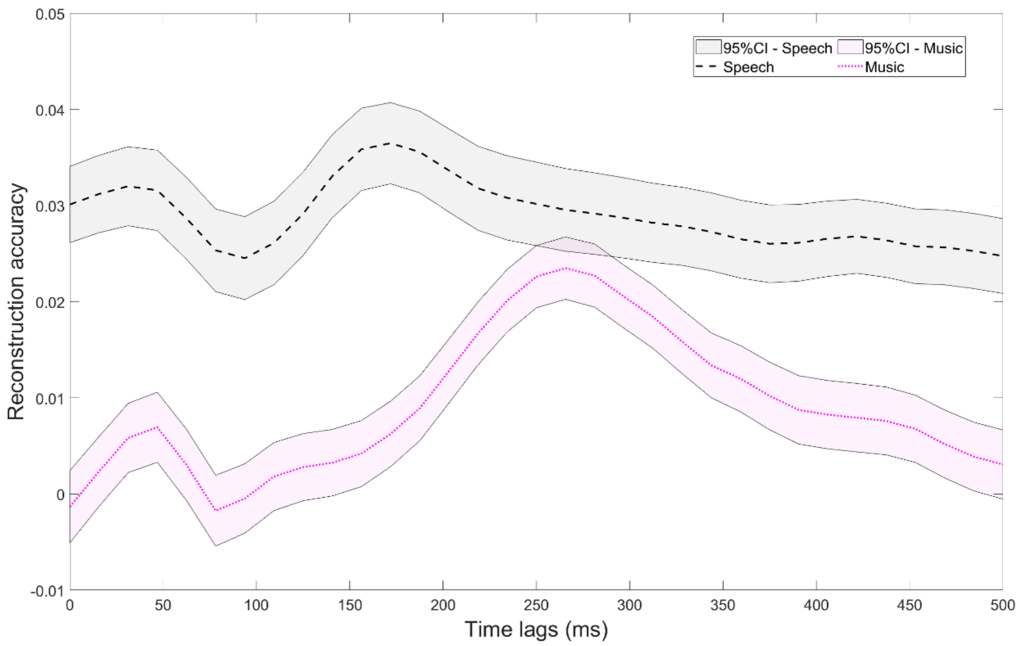

Figure 2 — Reconstruction Accuracy across all time lags for trials where the target audio is Speech and trials where the target audio is Music.

3.2 Effect of attention

To explore if the attentional processes influence the temporal pattern of reconstruction accuracy, we compared reconstruction accuracy obtained with models trained to reconstruct the target of attention and models trained to reconstruct the distractor stimulus.

To that mean, separate models have been trained to either reconstruct envelopes of the target sound or to reconstruct envelopes of the distractor sound. Target models are trained in a leave-one-out approach, by using trials where the target is of the same time as the trial under test (either music or speech). The distractor models are trained with a similar approach, but by using trials where the distractor was the same type as the distractor of the trial under test. Comparing the reconstructions accuracy of these two models can give information about the effect of attention on cortical auditory processes: increased reconstruction accuracy at a given time lag observable for the target model but not for the distractor model can indicate an attentional effect.

3.3 Music listening scenario

Figure 3 shows the reconstruction accuracies for both target models and distractor models for music listening. The two peak patterns can be seen for both models, with both peaks happening at similar time lags for both the Target model and the distractor model. Permutation tests, based on 100,000 permutations, were run to compare the reconstruction accuracy between models for each time lag. Significant differences were found for time lags between 260 and 285 ms, and also 320 and 380 ms, where the Target model results in higher reconstruction accuracies than the distractor model. The maximum of reconstruction accuracies with a difference between the performance of the Target model compared to the distractor model, which suggests that a music-specific process around 265 ms time lags may be affected by attentional processes.

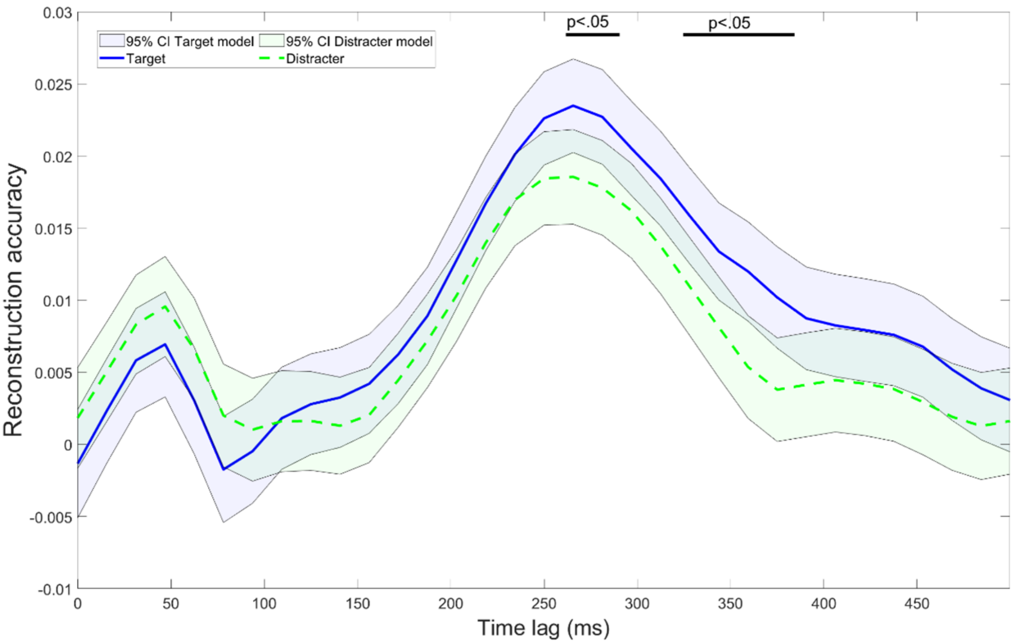

Figure 3 — Reconstruction accuracy across all time-lags, obtained with Target model, with music as a target and Decoder model,with music as a distractor

3.4 Speech listening scenario

Figure 4 shows the reconstruction accuracies for both target and distractor models for speech listening. Overall, reconstructions accuracies were obtained with the distractor models compared to the target model, for all time lags. Permutation tests, based on 100,000 permutations were run to compare the target model versus the distractor model, and indicate significant differences (p¡.05) for all time lags, except between 355 and 410 ms. Despite the difference between the target and distractor models, the variation of the reconstruction accuracies across shows different trends for the models. For both models, the first peak of increased reconstruction accuracies can be observed between 30 and 50 ms. However, while a second peak with maximal reconstruction accuracies arises at around 170 ms for the target model, there is no such increase for the distractor model. For the distractor model, an increase in reconstruction accuracy is observable at late time lags (350 to 450 ms) Taken together, it suggests that during speech listening, attentional processes affect the reconstruction accuracies level. The absence of a peak of maximal reconstruction when using the decoder model could indicate an increased attentional effect around a time lag of 170 ms, as suggested by O'sullivan et al. (2015).

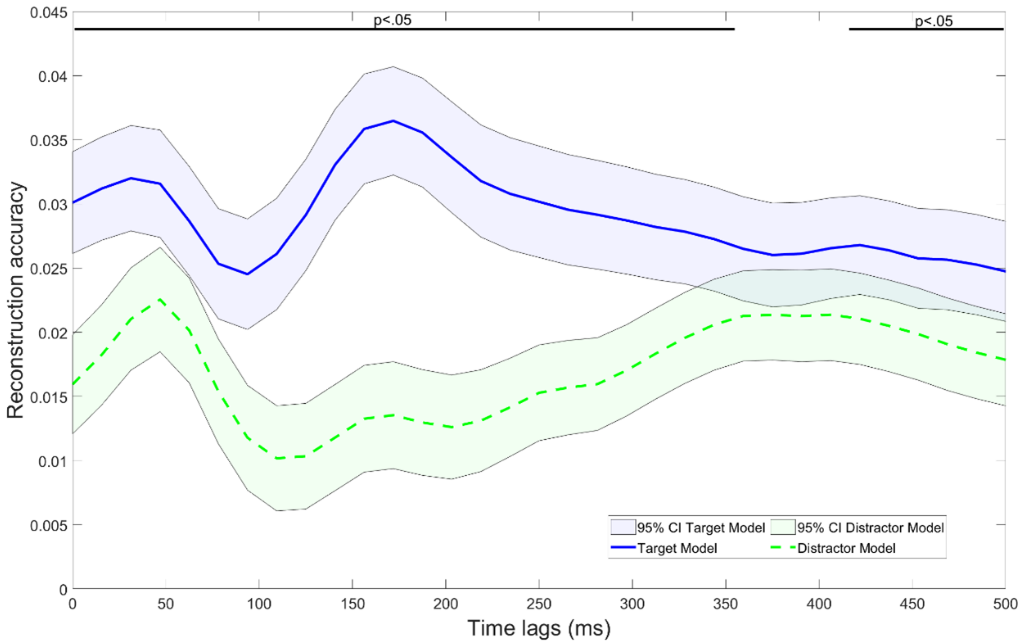

Figure 4 — Reconstruction accuracy across all time-lags, obtained with Target model, with speech as a target and Decoder model, with speech as a distractor.

4 Discussion

In this study, we used linear regression to reconstruct sound heard from EEG data. By using single-lag models we explore the effect of time lags applied to EEG data to reconstruction accuracy. We compare models trained on speech and on music to highlight temporal differences in the cortical process for speech listening and music listening.

Overall, the reconstruction accuracy is higher for speech listening compared to music listening, for all time lags. This result was expected as performance differences for reconstruction accuracy have previously been observed between speech and music (Simon et al., 2022 - Submitted; Zuk et al., 2021).

For both Speech and music listening, a two-peak pattern can be observed: an early first peak of increased reconstruction accuracies for time lags around 30 to 50 ms, and a later second peak of maximal reconstruction accuracy, where the timing differs between speech and music. This two-peak pattern has previously been observed for Speech listening (O'sullivan et al., 2015) or Music listening (Hausfeld et al., 2018).

Results suggest that the maximized reconstruction accuracies are obtained using different time lags for speech listening and music listening. For Speech, optimized reconstruction is obtained by applying a time lag of approximately 170 ms to the EEG data, relative to the audio signal. This is coherent with previous findings on optimal time lag for speech reconstruction or auditory attention decoding (Fuglsang, Dau, & Hjortkjær, 2017; Mirkovic et al., 2015; O'sullivan et al., 2015).

For music listening, the peak of optimized reconstruction is obtained for a time lag of approximately 265 ms. This optimal timing differs from previous results, where maximized reconstruction accuracies were found for music either at short time lags (Hausfeld et al., 2021), or longer time lags (Hausfeld et al., 2021; Hausfeld et al., 2018).

The early peak could suggest an early auditory process, which is coincident with both speech and music listening. The timing differences at middle/late timelags between speech and music listening suggest different cortical processes in place for speech listening and music listening.

A second question explored in the present study was to investigate the effect of selective auditory attention on reconstruction accuracies. To that mean, decoder models were trained to reconstruct the distractor stimulus, which was ignored by the participant. Comparing the reconstruction accuracies across time lags between the outputs of the target models and the distractor model can provide insight about the effect of attention.

For music, small but significant differences in reconstruction accuracies were found for time lags between 260 and 285 ms, and also 320 and 380ms. This difference is aligned with the peak of maximized reconstruction accuracies, which suggests that the process that creates this peak of maximal reconstruction might be affected by attention. On the other hand, the first peak of maximized reconstruction is similar for both target and distractor models. Taking together, these findings could suggest two separate cortical processes of music, an early one, not affected by selective auditory attention, and a late process, influenced by attentional processes.

For speech, the findings should be interpreted with more caution as a significant difference is observed for the Target and distractor model, across all time lags. These differences might be influenced by the difficulty of the task: when attending to a target sound is more challenging, an increased effort may be necessary to ignore the distractor, and the cortical tracking of the distractor might be reduced. Trials used for the speech decoder model correspond to trials where the listener had to attend to either speech signal or music signal in presence of distracting speech. However, attending to music in presence of speech is not a common task for human beings, compared for example to listening to speech in presence of music. This task might have been more challenging for the participants, which could influence the results observed in Figure 3.

Despite this offset between models, the shape of the peaks is worth considering. For both target and distractor models, an early peak of increased reconstruction accuracy is present for time lags of 30 to 50 ms. This suggests that the underlining cortical process is activated for stimuli that are inside or outside the focus of attention of the listener. For the target model, a second peak is observed at middle time lags ($\approx$170 ms), while for the distractor model this peak is almost inexistent. It could indicate that the underlining cortical process that results in this peak is activated only for attended speech. It corroborates the idea developed in (O'sullivan et al., 2015), which suggested an important attentional effect between 170 and 250 ms.

5 Conclusion

The present study explored the temporal aspect of cortical auditory processes and the effect of auditory attention by using a stimulus reconstruction approach. The results highlight two phases of auditory processes: an early process, which is concomitant for both speech and music listening and a second process, happening later for music listening than for speech listening. While the first process does not seem to be affected by attentional processes, the second might be enhanced by sounds that are actively attended by the listener. More research should be conducted to replicate, confirm, and elaborate on the current findings. Further analysis, such as using a forward model (Alickovic et al., 2019), could provide additional information on the cortical processes in places during listening in complex auditory sound scenes and explore more the similarities and differences between speech and music listening. Overall, this study highlights differences in optimal time-lags for cortical stimulus reconstruction between speech listening and music listening. This suggests temporal differences in cortical processing of speech and music. These differences could be used to fine-tune an auditory attention decoder that would be specifically tuned for music or speech.

References

Alickovic, E., Lunner, T., Gustafsson, F., & Ljung, L. (2019). A tutorial on auditory attention identification methods. *Frontiers in neuroscience*, 153.

Alickovic, E., Ng, E. H. N., Fiedler, L., Santurette, S., Innes-Brown, H., & Graversen, C. (2021). Effects of hearing aid noise reduction on early and late cortical representations of competing talkers in noise. *Frontiers in Neuroscience*, *15*.

An, W. W., Shinn-Cunningham, B., Gamper, H., Emmanouilidou, D., Johnston, D., Jalobeanu, M., Cutrell, E., Wilson, A., Chiang, K.-J., & Tashev, I. (2021). Decoding music attention from "eeg headphones": A user-friendly auditory brain-computer interface. *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 985–989.

Cantisani, G., Essid, S., & Richard, G. (2019). Eeg-based decoding of auditory attention to a target instrument in polyphonic music. *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 80–84.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, *25*(5), 975–979.

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear modeling of neurophysiological responses to speech and other continuous stimuli: Methodological considerations for applied research. *Frontiers in Neuroscience*, *15*.

Delorme, A., & Makeig, S. (2004). Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9–21.

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, *25*(19), 2457–2465.

Ding, N., & Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, *109*(29), 11854–11859.

Ding, N., & Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of neurophysiology*, *107*(1), 78–89.

Fuglsang, S. A., Dau, T., & Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *Neuroimage*, *156*, 435–444.

Hausfeld, L., Disbergen, N. R., Valente, G., Zatorre, R. J., & Formisano, E. (2021). Modulating cortical instrument representations during auditory stream segregation and integration with polyphonic music. *Frontiers in neuroscience*, *15*.

Hausfeld, L., Riecke, L., Valente, G., & Formisano, E. (2018). Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. *NeuroImage*, *181*, 617–626.

Mirkovic, B., Debener, S., Jaeger, M., & De Vos, M. (2015). Decoding the attended speech stream with multi-channel eeg: Implications for online, daily-life applications. *Journal of neural engineering*, *12*(4), 046007.

O'sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., & Lalor, E. C. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial eeg. *Cerebral cortex*, *25*(7), 1697–1706.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197.

Schäfer, P. J., Corona-Strauss, F. I., Hannemann, R., Hillyard, S. A., & Strauss, D. J. (2018). Testing the limits of the stimulus reconstruction approach: Auditory attention decoding in a four-speaker free field environment. *Trends in Hearing*, *22*, 2331216518816600.

Simon, A., Loquet, G., Østergaard, J., & Bech, S. (2022 - Submitted). Auditory attention decoding from eeg during music listening. *Frontiers in neurosciences.*

Wong, D. D., Fuglsang, S. A., Hjortkjær, J., Ceolini, E., Slaney, M., & De Cheveigne, A. (2018). A comparison of regularization methods in forward and backward models for auditory attention decoding. *Frontiers in neuroscience*, *12*, 531.

Wöstmann, M., Fiedler, L., & Obleser, J. (2017). Tracking the signal, cracking the code: Speech and speech comprehension in non-invasive human electrophysiology. *Language, Cognition and Neuroscience*, *32*(7), 855–869.

Zuk, N. J., Murphy, J. W., Reilly, R. B., & Lalor, E. C. (2021). Envelope reconstruction of speech and music highlights stronger tracking of speech at low frequencies. *PLoS computational biology*, *17*(9), e1009358.