

A System Identification Approach to Determining Listening Attention from EEG Signals

Emina Alickovic, Thomas Lunner and Fredrik Gustafsson

Conference Publication



N.B.: When citing this work, cite the original article.

Original Publication:

Emina Alickovic, Thomas Lunner and Fredrik Gustafsson, A System Identification Approach to Determining Listening Attention from EEG Signals, 2016, 2016 24th European Signal Processing Conference (EUSIPCO 2016).

Copyright: www.ieee.org

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-130732>



A System Identification Approach to Determining Listening Attention from EEG Signals

Emina Alickovic*, Thomas Lunner^{† ‡ §}, Fredrik Gustafsson*

*Department of Electrical Engineering, Linköping University, Linköping, SE-58183, Sweden
Email: emina.alickovic@liu.se, fredrik@isy.liu.se

[†]Eriksholm Research Centre, Oticon A/S, 20 Rortangvej, Snekkersten, Denmark
Email: thlu@eriksholm.com

[‡]Technical Audiology, Department of Experimental and Clinical Research, Linköping University, Linköping, SE-58183, Sweden

[§]Swedish Institute for Disability Research, Linnaeus Centre HEAD, Linköping University, Linköping, SE-58183, Sweden

Abstract—We still have very little knowledge about how our brains decouple different sound sources, which is known as solving the cocktail party problem. Several approaches; including ERP, time-frequency analysis and, more recently, regression and stimulus reconstruction approaches; have been suggested for solving this problem. In this work, we study the problem of correlating of EEG signals to different sets of sound sources with the goal of identifying the single source to which the listener is attending. Here, we propose a method for finding the number of parameters needed in a regression model to avoid overlearning, which is necessary for determining the attended sound source with high confidence in order to solve the cocktail party problem.

Index Terms—attention, cocktail party, linear regression (LR), finite impulse response (FIR), multivariable model, sound, EEG.

I. INTRODUCTION

It is well known that humans can effortlessly separate different sound sources and attend to only one of these sources in a complex environment, a so-called cocktail party, as first described in 1953 [1]; however, still very little is known about how this is achieved. Different sound sources excite different responses in human brain activity, but it is highly challenging to determine these responses and the durations of their effects on the brain in electroencephalographic (EEG) and magnetoencephalographic (MEG) signals. In this study, we introduce a model-based learning approach to study this problem. A forward (causal FIR) model from sound to EEG is proposed.

Recently, it has been shown that EEG data can be used to determine attention in a two-speaker environment [2]. Following the realization that brain activity follows the amplitude of the sound envelope, various regression approaches were developed to gain a better understanding of the auditory system [3], [4]. This understanding could also contribute to advances in cognitive hearing aids (HAs). Whereas modern HAs use beam-forming to emphasize sounds coming from one direction while ignoring the other sounds and have no knowledge of the sound source attended to by the listener, next-generation cognitive HAs may have the ability to find the sound source to which a listener is directing his or her attention.

Regression methods and stimulus reconstruction approaches have already been successfully applied to intracranial EEG or electroencephalographic (ECoG) data [4], [5], [6], MEG data [3], [7] and EEG data [8], [2]. Although impressive results can be obtained using ECoG data, ECoG measurements are invasive and can only be used with listeners under medical care; as such, these approaches are not plausible for everyday applications. Studies performed on MEG data have shown that the envelope of the input sound can be estimated and that the resulting estimate is generally more strongly correlated with attended sound than with unattended sound. However, MEG instruments are costly, not portable and not readily available, which makes them difficult to use in real-time devices. By contrast, EEG instruments are relatively cheap, more widely available and portable, making it possible to incorporate them into many real-time devices. For portable devices such as HAs, EEG measurements are a potentially realistic tool because they are non-invasive and, recently, considerable progress has been made in the development of wearable EEG devices [9], [10], [11], [12].

Various approaches that can solve the cocktail party problem to some extent have been suggested. In [13], EEG features (a cross-correlation measure, the power in the α band and a measure of the auditory steady-state response (ASSR)) were extracted and fed into a linear discriminant classifier. In [14], canonical correlation analysis (CCA) and mapping from EEG signals to sound with both single-channel inversion (AESPA) and all-channel inversion (optimal) were tested, and the obtained decoding accuracies were 65 – 80 % for CCA and AESPA and 75 – 95 % for the optimal approach for 60-s-long samples. Recently, non-linear approaches based on deep neural networks (DNNs) were evaluated in [15], [16] because DNNs offers a wide range of different possibilities, although at higher computational costs. In addition, one study has also investigated whether it is possible to further improve encoding/decoding performance by incorporating knowledge about the peripheral auditory system into the process of extracting the sound envelope [17]. A recent study [2] demonstrated that neural processing at approximately 200 ms is critical for solving the cocktail party problem when mapping from EEG

signals to sound.

A common feature of these studies is that they have all proposed inverse models for proceeding from EEG signals to sound, that is, anti-causal FIR models. The disadvantages of such models include a lack of understanding of dynamic effects and difficulties in real-time implementation. Here, we consider only real-time-implementable solutions. Therefore, in this study, we consider only forward models (causal FIR models) because such models can be implemented in real time and integrated into many everyday devices. In the literature, FIR models with reverse causality are typically used. Incoming sound leaves traces in EEG signals within several tens of milliseconds, but it is not known exactly when. To solve this problem, we propose a technique based on linear regression and the F distribution to select an appropriate number of parameters with minimal losses.

II. PROBLEM FORMULATION

The cocktail party problem arises when a number of persons $i = 1, 2, \dots, n_u$ emit sound signals simultaneously and a listener receives the sum of these signals, i.e.,

$$u_{total} = \sum_{i=1}^{n_u} u_i(t) \quad (1)$$

The assumption is made that the listener is attempting to focus on only one speaker at a time, and the technical challenge is to determine which of the speech signals $u_i(t)$ is the subject of focus based on external sensor measurements, here, EEG signals $y_j(t)$, with $j = 1, 2, \dots, n_y$.

Usually, the problem includes the following:

- n_u different sound sources $u_i(t)$. In the experiment considered in this study, $n_u = 2$.
- n_y different EEG signals $y_j(t)$. In our experiment, we have access to full-scalp EEG signals, with $n_y = 128$.

The primary goal of this study is to determine to which of n_u sound sources the listener is attending. When the sound sources are available digitally, it is straightforward to amplify one and attenuate the others.

III. SIGNAL PRE-PROCESSING

Our experiment can be described as follows:

- A subject was asked to attend to a sound source on the right, u_2 , in all 30 trials.
- Each trial was approximately one minute long.
- In each trial, the subject was presented with 2 works of classic fiction in the left and right ears concurrently.
- EEG data were collected at a sampling frequency of 512 Hz.

These data were previously analyzed in [2] using different analysis tools.

Because the sampling rates of the sound and EEG signals were different and EEG signals measure energy rather than amplitude, as in the case of speech signals, some pre-processing was required:

- The EEG signals were filtered with a band-pass Butterworth filter between 1 and 8 Hz, i.e., the delta and theta bands, and down-sampled to 64 Hz.
- The envelopes of the sound sources were obtained by using the Hilbert transform of the fast sampled $u(t)$ signals and down-sampled to the same sample rate as that of the EEG signals.

IV. MODEL FORMULATION

In the literature, linear models are often used to solve the cocktail party problem. A causal FIR model ($FIR(k)$) can be described in terms of the following difference equation:

$$y(t) = b_0 u(t) + b_1 u(t-1) + \dots + b_k y(t-k) + e(t) \quad (2)$$

where $e(t)$ is the disturbance and k is the order of the model. Generally, $e(t)$ is considered to be white noise, $E(e(t)) = 0$ and $Var(e(t)) = \lambda$. Using the shift operator q in the time domain, $qu(k) = u(k+1)$, we can rewrite (2) as:

$$y(t) = (b_0 + b_1 q^{-1} + b_2 q^{-2} + \dots + b_k q^{-k}) u(t) + e(t) \quad (3)$$

This leads to the *transfer operator* $B(q) = b_0 + b_1 q^{-1} + b_2 q^{-2} + \dots + b_k q^{-k}$. Now, we can express (3) as:

$$y(t) = B(q)u(t) + e(t) \quad (4)$$

where $B(q)$ is the causal polynomial.

In the literature, an FIR model with reverse causality is often used, with the following form:

$$u(t) = A(q)y(t) + v(t) \quad (5)$$

where $A(q)$ is the anti-causal polynomial.

A causal FIR single input multiple output (SIMO) models based on (4) for each sound source can be expressed as a linear regression (LR):

$$Y = U_i \cdot B_i + E, \quad i = 1, 2, \dots, n_u \quad (6)$$

where B_i is $n_b \times n_y$ Henkel matrix, Y and E are $N \times n_y$, U_i is $N \times n_b$ and $n_b = k + 1$. In (6), the goal is to estimate B_i .

V. MODEL ORDER SELECTION

Let ε be defined as $\varepsilon = y(t) - B^T u_i(t)$. A least square (LS) estimation problem can be written as

$$\hat{B} = \arg \min_B \sum_{t=1}^N [\varepsilon \Lambda^{-1} \varepsilon] \quad (7)$$

where $\Lambda = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \lambda_{n_y} \end{bmatrix}$ which allows us to come

$$\hat{B}_i = \left[\frac{1}{N} \sum_{t_1}^N u_i(t) u_i(t)^T \right]^{-1} \left[\frac{1}{N} \sum_{t_1}^N u_i(t) y(t)^T \right] \quad (8a)$$

$$\hat{B}_i = [U_i^T \cdot U_i]^{-1} \cdot [U_i^T \cdot Y] \quad (8b)$$

The LS criterion is now:

$$\hat{V}_N^i(k) = \min_{B_j} \sum_{j=1}^{n_y} \lambda_j \left[\sum_{t=1}^N (y_j(t) - B_j^T u_i(t))^2 \right] \quad (9)$$

Since all B s are independent, this sum is minimized by individually minimizing the inner summation terms, independently of Λ . Now, loss function becomes

$$\hat{V}_N^i(k) = \frac{1}{N} \min_{B_i} \sum_{t=1}^N \|y(t) - B_i^T u_i(t)\|^2 \quad (10)$$

One idea is to select the sound source which minimizes loss function $V_N^i(k)$ [18]:

$$\hat{V}_N^i(k) = \arg \min_B V_N^i(k) \quad (11)$$

$$V_N^i(k) = \|Y - U_i B_i\|_F^2 \quad (12)$$

Frobenius norm $\|W\|_F^2 = \text{trace}(W^T W)$ is used for matrix-valued residuals $V_N^i(k)$.

First, the quadratic form of the minimizing loss function $\hat{V}_N^i(k)$ can be calculated as

$$\hat{V}_N^i(k) = Y^T (I - U_i (U_i^T U_i)^{-1} U_i^T) Y \quad (13)$$

Now, a non-standard problem in this application compared to examples in the system identification literature is the extremely small decrease in the loss function, meaning that degree of explanation (in percent) $100(1 - V_N^i(k)/V_N^i(0))$ is rather close to zero, compared to applications where the model explains most of the data where this degree is close to one.

The question that arises is how over-learning can be avoided. When the number of measurements N is small relative to the number of parameters n_b , LR can possibly result in over-learning. In this study, we attempt to find the best n_b for avoiding over-learning because such models typically yield relatively low predictive performance.

It is easy to formulate an LR problem, but even if the problem is correctly posed, it is necessary to determine what the expected value for $V_N^i(k)/V_N^i(0)$ is if U_i contains pure noise. The answer to this question is partially given in [19] for single-output models and here we extend it to SIMO models as stated in Lemma 1.

Lemma 1. Assume that U_i and Y can be described by a model structure Υ_1 :

$$Y = U_{i1} B_{i1} + E_1 \quad (14)$$

where E_1 is white Gaussian noise with variance matrix Λ defined in (7). Let

$$V_N^{(1)} = \min_{B_1} \|\Lambda^{-1/2} (Y - U_{i1} \cdot B_{i1})\|_F^2 \quad (15)$$

Suppose that we have another model structure Υ_2 with another regressor,

$$Y = U_{i1} B_{i1} + U_{i2} B_{i2} \quad (16)$$

and its loss function is

$$V_N^{(2)} = \min_{B_1, B_2} \|\Lambda^{-1/2} (Y - U_{i1} \cdot B_{i1} - U_{i2} \cdot B_{i2})\|_F^2 \quad (17)$$

Let $k_1 = \text{size}(B_{i1}, 1)$, $k_2 = \text{size}(B_{i2}, 1)$ and $M = N - k_1 - k_2$. Then, we have:

$$t(k_1, k_2, N) = \frac{M}{V_N^{(2)}} \cdot \frac{V_N^{(1)} - V_N^{(2)}}{k_2} \in F(k_2, M) \quad (18)$$

Proof. We start with proving

$$V_N^{(2)} \in \chi^2(n_y \cdot M) \quad (19)$$

Let $U_i B_i = [U_{i1} B_{i1} \quad U_{i2} B_{i2}] = [U_{i1} \quad U_{i2}] [B_{i1} \quad B_{i2}]^T$ with dimensions $N \times (k_1 + k_2)$. Replacing it in (17), we have:

$$\begin{aligned} Y - U_i B_i &= Y - U_i (U_i^T U_i)^{-1} U_i^T Y = G_N Y \\ &= G_N (U_i B_i + W_N) = G_N W_N \end{aligned}$$

where $G_N = I - U_i (U_i^T U_i)^{-1} U_i^T$ is idempotent matrix, $G_N G_N = G_N$ and all its eigenvalues are either zero or one. We also have the identity $\text{trace } G_N = \text{trace } I - \text{trace } ((U_i^T U_i)^{-1} U_i^T U_i) = N - (k_1 + k_2) = M$. Thus, M is the number of eigenvalues that are one and the remaining $k_1 + k_2$ eigenvalues are zero. Due to its symmetry property, G_N can be diagonalized by orthogonal matrix Q as $\Gamma = Q G_N Q^T$ where Γ is diagonal matrix with M ones and $k_1 + k_2$ zeros. Now we have:

$$\begin{aligned} V_N^{(2)} &= \text{trace}([\Lambda^{-1/2} (Y - U_i B_i)]^T [\Lambda^{-1/2} (Y - U_i B_i)]) \\ &= \text{trace}(W_N^T G_N^T \Lambda^{-1} G_N W_N) = \text{trace}(\Lambda^{-1} W_N^T G_N W_N) \\ &= \text{trace}(\Lambda^{-1} (Q W_N)^T \Gamma (Q W_N)) = \sum_{j=1}^{n_y} \lambda_j^{-1} \sum_{t=1}^M \hat{w}_j^2(t) \end{aligned}$$

Because the elements of W_N are independent and normal with variances in Λ and Q is an orthogonal matrix, the elements of $Q W_N$ are also independent and normal, and by the definition of the chi-squared distribution, we have $\lambda_j^{-1} \sum_{t=1}^M \hat{w}_j^2(t) \in \chi^2(M)$ and total sum follows a χ^2 distribution with $n_y \cdot M$ degrees of freedom, that is, $V_N^{(2)} \in \chi^2(n_y \cdot M)$.

In similar manner, it is easy to prove that $(V_N^{(1)} - V_N^{(2)}) \in \chi^2(n_y \cdot k_2)$ and $V_N^{(1)} - V_N^{(2)}$ and $V_N^{(2)}$ are independent. Now, together with the definition of the F distribution, we have:

$$t(k_1, k_2, N) = \frac{M}{V_N^{(2)}} \cdot \frac{V_N^{(1)} - V_N^{(2)}}{k_2} \in F(k_2, M) \quad q.e.d.$$

□

Lemma (1) tells us whether it is useful to include B_{i2} , from which we can obtain a suitable size for the model set and model Υ_2 as a suitable model structure. It follows that if the model Υ_1 given in (14) is sufficient and it is unnecessary to consider Υ_2 , then the normalized decrease in the loss functions has the distribution given in (18). Model Υ_1 does not hold, and therefore, model Υ_2 should be considered, when the computed decrease is significantly greater, i.e., if $t(k_1, k_2, N) > F_\alpha(k_2, M)$, where α is the critical value.

If we assume that the true model Υ_1 is only noise, i.e., $Y = E_1$, then we can take $k_1 = 0$. We can use k_2 parameters in the model that attempts to model the noise to achieve a high confidence in our classification. The results will be F

distributed.

Lemma 1 is very much in spirit like Akaike information criteria (AIC). If we have a range of possible model orders and we wish to see which order is the best suited to our purposes, we can apply Lemma 1 over a range of different k_2 values while keeping the k_1 value fixed. It may occur that the desired confidence is obtained for more than one model order. The typical choice in this case would be to select a relatively low or the lowest order. However, it may be the case that this is not the best choice because it is possible that $t(k_i, k_j, N) > F_\alpha(k_j, M)$ where k_i and k_j are the parameters for which the desired confidence is obtained. In this study, we attempt to address this issue.

Corollary V.1. *To determine whether the inclusion of a higher-order model Υ_3 , where $\Upsilon_1 \subset \Upsilon_2 \subset \Upsilon_3$, is "necessary" to solve the cocktail party problem, Eq. 17 becomes Eq. 14, $k_1^{new} = k_2^{old}$, and the new model Υ_3 will be*

$$Y = U_{i1}B_{i1} + U_{i2}B_{i2} + U_{i3}B_{i3} \quad (20)$$

with $k_2^{new} = k_3 = \text{size}(B_{i3}, 1)$ and $M^{new} = N - k_1 - k_2 - k_3$. Then, Lemma 1 can be used to determine whether it is necessary to include the regressor $U_{i3}B_{i3}$ in the regression. If $t(k_1^{new}, k_2^{new}, N) > F_\alpha(k_2^{new}, M^{new})$, then the answer is "Yes". This process can be repeated depending on the critical value α and the desired performance.

VI. CLASSIFICATION

If the appropriate order of the model can be found, it should be easy to perform simple LS estimation (8a) - (8b). The cocktail party problem is now to see which input signal gives the smallest cost. That is, we do not need estimate B_i itself, just the cost function at minimum $\hat{V}_N^i(k)$ given in (11).

Then, as the first possibility, we can determine the attended sound source by

$$\hat{i} = \arg \min_i \hat{V}_N^i(k) \quad (21)$$

VII. EXPERIMENTAL RESULTS

For simple LS estimate to be *consistent*, i.e. \hat{B}_i to converge to the true value, it is necessary that $U_i^T U_i$ is *non-singular*. Otherwise, regularization is needed. This will be the case if $N > n_b$ and columns in U_i are linearly independent. Then, \hat{B}_i can be computed as in (7). The experimental results show that once the number k_2 becomes 90 or higher, one of the matrices $U_i^T U_i$ becomes singular leading to *inconsistency* and *over-fitting*.

In models Υ_1 and Υ_2 , the model estimate \hat{B}_i will be incorrect, and there is a trade-off between bias (meaning that the model is not sufficiently flexible) and variance (meaning that the disturbance E influences \hat{B}_i and causes variations among repeated observations). We start with N corresponding to one trial (estimation for one batch of one minute).

Next, we seek to determine the number of parameters required to achieve a high confidence in classification. First, we wish to see how many parameters are required to achieve a classification confidence of $\alpha = 0.6$.

As the first step, we assume that model Υ_1 , i.e., the true model, is only noise such that $B = 0$ and $k_1 = 0$; therefore, the loss function becomes

$$V_N^{(i1)} = \|Y\|_F^2 \quad (22)$$

In model Υ_2 , we distinguish the two different cases of attended and unattended sound sources, i.e.,

$$V_N^{(i2)} = \|Y - U_i \cdot B_i\|_F^2, \quad i = 1, 2 \quad (23)$$

First, we fix $k_1 = 0$ and select k_2 values from the range [1 90] to find all model orders, if any, for which confidence of α or higher is achieved. The results are shown in Fig. 1 where confidence levels averaged over 30 trials are shown for $k_2 \in [1 90]$. As seen from this figure, the average confidence is greater than α for multiple model orders, and the highest confidence is obtained for $k_2 < 30$. However, because Fig. 1 shows only the means across 30 trials for $k = 90$ parameters, this plot may not be reliable because there might be large deviations in the data. One suitable tool for further visualizing the consistency of the data is a box plot. The model of order k with the highest 25th percentile for both sound sources will be regarded as having the "most necessary" parameters. Because we wish to allow for some flexibility in the model, Fig. 2 shows the results obtained for $k_2 \in [10 29]$.

Here, we are seeking the number of parameters for which the 25th percentile is above α for both sound sources. This is seen to be the case for $k_2 = [10 11 12 17]$. To decide which k_2 value yields the best-fit LR, we apply Corollary V.1. The confidences averaged across 30 trials are less than .1 and .3 when we set $k_1^{new} = 10$ and test whether the inclusions of regressors with 11 and 12 parameters are "necessary". From this, we can see that if we need to choose a model with a number of parameters 10,11 or 12, a model with 10 parameters is appropriate. When we set $k_1^{new} = 10$ and $k_2^{new} = 7$, we find that the average confidence for sound source 2 is greater than 0.5, which tells us that the inclusion of a model of order 16 is necessary if we set $\alpha = 0.5$. From this, we can conclude that the FIR(16) model should be used to solve the cocktail party problem for this subject.

To illustrate the pragmatic value of Lemma (1), we then study the difference in quadratic loss values, i.e., $\hat{V}_N^1(16) - \hat{V}_N^2(16)$. From Fig. 1 - 2, we can see that this difference averaged across 30 trials is positive, i.e., $\hat{V}_N^1(16) > \hat{V}_N^2(16)$. From (21), we have:

$$\hat{i} = \arg \min_i \hat{V}_N^i(16) = 2 \quad (24)$$

demonstrating that the listener, in average, was attending to the sound source 2.

In addition, we also select a subset of the electrodes, i.e., those at the temporal lobe close to the ears, to investigate the extent of model loss for this subset and whether results similar to those for the full-scalp EEG data can be obtained. Ear-EEG refers to in-ear devices/hearables that contain electrodes and are placed in or near the ear to serve as practical devices for all-day use. We know from previous studies that the signals

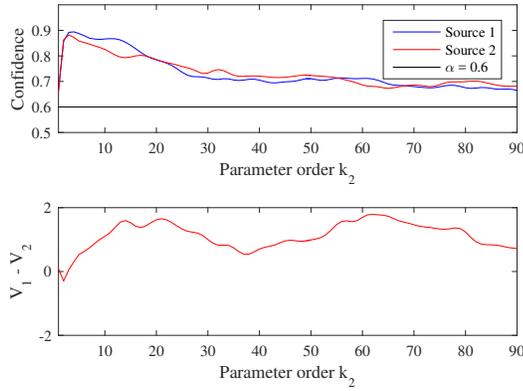


Fig. 1. (a) Confidence in classification for $k_2 \in [1 \ 90]$ (b) $V_N^1 - V_N^2$

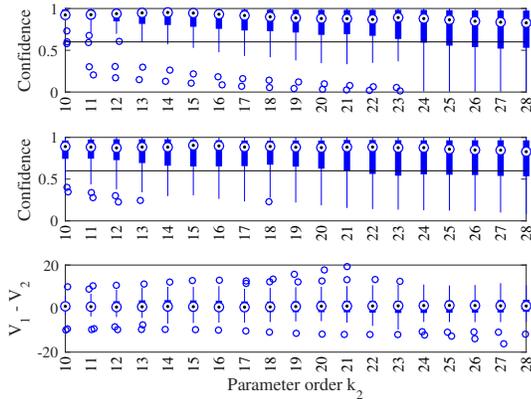


Fig. 2. Box-plot results for $k_2 \in [10 \ 29]$ for (a) sound source 1 and (b) sound source 2. (c) $V_N^1 - V_N^2$

from Ear-EEG electrodes are highly correlated with those from temporal-lobe electrodes [20]. A subset consisting of 24 electrodes (12 electrodes on each side) was selected. Similar results were found, i.e., FIR(16) should be used in LR.

VIII. CONCLUSION

In this study, we studied the problem of over-learning and how it can be avoided. The method suggested here can be used to determine the optimal order for any FIR model; therefore, it can be used as a tool for solving the cocktail party problem. The advantages of this model are that it is linear in all parameters and the least-squares criterion provides an analytical solution for all parameters. Additionally, this model can be implemented in real time. Using this approach, we found that the FIR(16) model is needed to avoid over-learning in LR and demonstrated how large a difference in losses between 2 sources is necessary to achieve significance in sound source classification. To summarize, a contribution of the present article is a real-time model for determining order of FIR filter in context of regression model learning, tailored to solving the cocktail party problem.

ACKNOWLEDGMENTS

The authors would like to thank Edmund Lalor for providing the data used to evaluate the proposed technique. The authors

also thank James O’Sullivan and Sina Khoshfetrat Pakazad for their technical assistance.

REFERENCES

- [1] E. C. Cherry, “Some experiments on the recognition of speech, with one and with two ears,” *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953.
- [2] J. A. O’Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, “Attentional selection in a cocktail party environment can be decoded from single-trial eeg,” *Cerebral Cortex*, vol. 25, no. 7, pp. 1697–1706, 2015.
- [3] N. Ding and J. Z. Simon, “Emergence of neural encoding of auditory objects while listening to competing speakers,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 29, pp. 11854–11859, 2012.
- [4] N. Mesgarani and C. E. F., “Selective cortical representation of attended speaker in multi-talker speech perception,” *Nature*, vol. 485, 2012.
- [5] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang, “Reconstructing speech from human auditory cortex,” *PLoS Biol.*, vol. 10, p. e1001251, 01 2012.
- [6] E. M. Z. Golumbic, N. Ding, S. Bickel, P. Lakatos, C. A. Schevon, G. M. McKhann, R. R. Goodman, R. Emerson, A. D. Mehta, J. Z. Simon, D. Poeppel, and C. E. Schroeder, “Mechanisms underlying selective neuronal tracking of attended speech at a cocktail party,” *Neuron*, vol. 77, no. 5, pp. 980 – 991, 2013.
- [7] S. Akram, A. Presacco, J. Z. Simon, S. A. Shamma, and B. Babadi, “Robust decoding of selective auditory attention from {MEG} in a competing-speaker environment via state-space modeling,” *NeuroImage*, vol. 124, Part A, pp. 906 – 917, 2016.
- [8] A. J. Power, J. J. Foxe, E.-J. Forde, R. B. Reilly, and E. C. Lalor, “At what time is the cocktail party? a late locus of selective attention to natural speech,” *European Journal of Neuroscience*, vol. 35, no. 9, pp. 1497–1503, 2012.
- [9] T. Lunner and F. Gustafsson, “Hearing device with brainwave dependent audio processing,” Apr. 10 2014. US Patent App. 14/048,883.
- [10] N. Pontoppidan, T. Lunner, M. Pedersen, L. Hauschultz, P. Koch, G. Naylor, and E. Petersen, “Hearing assistance device with brain computer interface,” Dec. 18 2014. US Patent App. 14/303,844.
- [11] T. Lunner and N. Pontoppidan, “Configurable hearing instrument,” June 19 2014. US Patent App. 14/103,399.
- [12] T. Lunner, “Hearing device with external electrode,” Mar. 3 2015. US Patent 8,971,558.
- [13] C. Horton, R. Srinivasan, and M. D’Zmura, “Envelope responses in single-trial eeg indicate attended speaker in a ‘cocktail party’,” *Journal of Neural Engineering*, vol. 11, no. 4, p. 046015, 2014.
- [14] E. Lalor, N. Mesgarani, S. Rajaram, A. O’Donovan, J. Wright, I. Choi, J. Brumberg, N. Ding, A. K. Lee, N. Peters, S. Ramenahalli, J. Pompe, B. Shinn-Cunningham, M. Slaney, and S. Shamma, “Decoding auditory attention (in real time) with eeg,” in *Proceedings of the 37th ARO MidWinter Meeting*, Association for Research in Otolaryngology (ARO), February 2013.
- [15] M. Yang, S. A. Sheth, C. A. Schevon, G. M. McKhann II, and N. Mesgarani, “Speech reconstruction from human auditory cortex with deep neural networks,” in *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, (Dresden, Germany), February 2015.
- [16] D. D. Wong, U. Pomper, E. Alickovic, J. Hjortkaer, M. Slaney, S. Shamma, and A. de Cheveigne, “Decoding speech sound source direction from electroencephalography data,” *ARO winter meeting (abstract)*, 2016.
- [17] W. Biesmansy, N. Das, T. Francart, and A. Bertrand, “Auditory-inspired speech envelope extraction methods for improved eeg-based auditory attention detection in a cocktail party scenario,” tech. rep., 2015.
- [18] F. Gustafsson, L. Ljung, and M. Millnert, *Signal processing*. Lund: Studentlitteratur, english ed., 2010.
- [19] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ 07458: Prentice Hall PTR, 2nd ed., 1999.
- [20] L. Fiedler, T. Lunner, A. Brandmeyer, M. Wostmann, C. Graversen, and J. Obleser, “In-ear-eeg indicates neural signatures of effortful auditory processing,” in *Proceedings of Advances and Perspectives in Auditory Neuroscience (APAN)*, (Chicago, Illinois, USA), October 2015.