

# An Interpretable Performance Metric for Auditory Attention Decoding Algorithms in a Context of Neuro-Steered Gain Control: Supplementary Material

Simon Geirnaert, Tom Francart, and Alexander Bertrand, *Senior Member, IEEE*

In the supplementary material, related to the paper *An Interpretable Performance Metric for Auditory Attention Decoding Algorithms in a Context of Neuro-Steered Gain Control*, we describe a subjective listening test to validate the choice for the comfort level  $c = 0.65$  (Section I) and elaborate on the influence of the hyperparameters  $P_0$  (the confidence level) and  $c$  (comfort level) on the MESD metric (Section II). Furthermore, we investigate in Section III how the ESD and the number of states of the optimized Markov chain depend on the decision window length and accuracy, for the MMSE-based decoder with averaging of autocorrelation matrices.

## I. VALIDATION OF THE COMFORT LEVEL $c$

To validate the chosen  $c$ -value ( $c = 0.65$ ) of Section III-A in case of a (more relevant) connected discourse stimulus instead of standard sentences (as used in Section III-A), we conducted a subjective listening experiment to determine  $\text{SNR}_c$ . Eight normal hearing participants, aged between 24 and 29 and with Dutch as their mother tongue, were asked to listen to a mixture of two non-standardized, commercial recordings of stories, 6 min and 34 s long. The stimuli were biologically calibrated. The participants were allowed to adapt the SNR with a slider between 0 and 50 dB and were instructed to select the minimal SNR (between the dominantly amplified speaker and the competing speaker) that still allowed them to comfortably listen to the dominantly amplified speaker for a duration of, e.g., 30 min. When they selected a value for  $\text{SNR}_c$ , they were instructed to listen to the dominantly amplified speaker for three more minutes at their selected  $\text{SNR}_c$ , where now the previously suppressed speaker is the dominantly amplified speaker. As a validation procedure, the participants self-reported their listening effort, probing the amount of effort required to understand the loudest speaker. A review on the self-reported listening effort and other methods to assess listening effort can be found in [1]. The minimal reported, maximal reported and median  $\text{SNR}_c$  is equal to 4.56 dB, 23.55 dB and 10.89 dB. All reported listening efforts were below 25%.

To obtain the SRT, we used the results from [2], where they performed a similar experiment (using similar conditions) in an age-matched, normal hearing group to determine the SRT of connected discourse using the self-assessed Békésy procedure. We use the median  $\text{SRT} = -16.27$  dB as a value for  $\text{SNR}_{\max}$

$= 16.27$  dB. Note that this SRT differs from the one reported in Section III-A, as we are now dealing with a connected discourse instead of standard sentences, while also a different procedure for assessing speech intelligibility has been used.

The resulting  $c$ -value (12) is equal to  $c = 0.727$ . Given the large variability on the reported comfort level, we consider this value to be reasonably close to the proposed value  $c = 0.65$ , which was calculated based on data from the literature.

## II. THE RELATION BETWEEN THE MESD AND THE HYPERPARAMETERS

Fig. 1 shows how the MESD metric depends on the hyperparameters  $P_0$  (the confidence level) and  $c$  (the comfort level). The MESD's are based on the results of an MMSE-based decoder with averaging of autocorrelation matrices, described in Section III and Fig. 4 of the paper. When varying one hyperparameter, the other hyperparameters are kept constant at their default values ( $P_0 = 0.8, c = 0.65, N_{\min} = 5$ ). The black diamonds indicate the chosen hyperparameter value in the paper. Fig. 1a shows that  $P_0 = 0.8$  yields a good trade-off between a high confidence level and a small enough MESD. As the MESD has a positive second-order derivative in function of  $P_0$ , an extra amount of confidence results in an even larger increase in MESD, which is why it is important to choose its value as low as possible, without giving too much in on the reliability of the gain control system.

The MESD is a discrete function of the comfort level  $c$  (Fig. 1b). As the lower bound of the  $P_0$ -confidence interval needs to be above comfort level  $c$ , a higher comfort level results in more states and thus in a higher MESD. Note that because of the flooring operation in (4), this a discrete function. Again, higher comfort levels result in a steeper increase in switch duration. The comfort level  $c = 0.65$  that resulted from the analysis and experiments in Section III-A of the paper and Section I of the supplementary material seems to avoid this high cost of extra comfort while assuring, by design, enough comfort for the user.

## III. THE ESD AND NUMBER OF STATES IN FUNCTION OF THE DECISION WINDOW LENGTH

In Section III-B, the MESD has been applied to the performance curve of the MMSE-based decoder with averaging of autocorrelation matrices versus averaging decoders (Fig. 4).

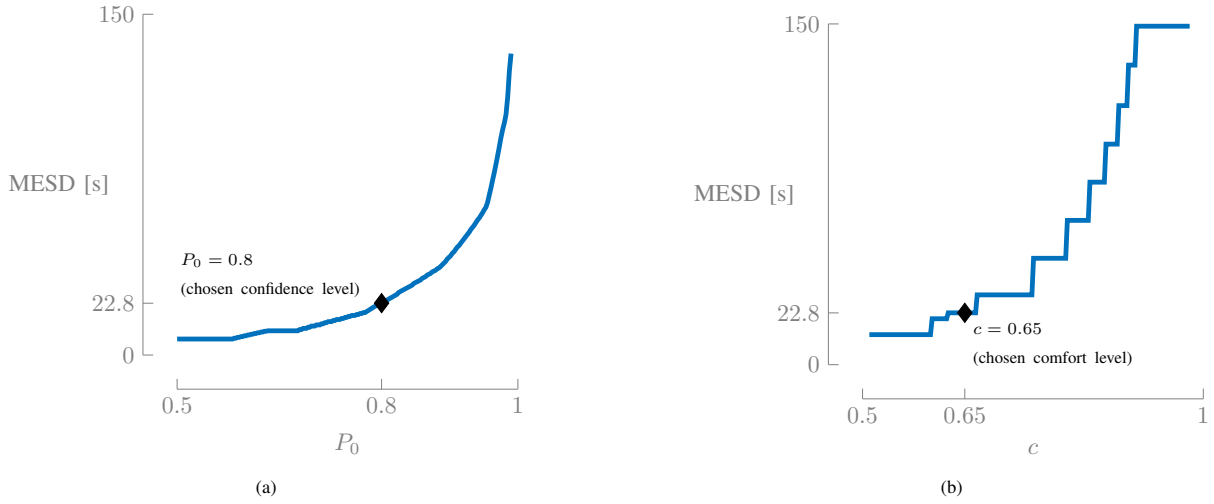


Fig. 1: The MESD increases in function of (a) the confidence level  $P_0$ , with a positive second-order derivative and (b) the comfort level  $c$ , in a discrete way, also with an increasing slope. The MESD's are shown for the performance curve of the MMSE-based decoder with averaging of autocorrelation matrices. The chosen confidence level and comfort level are indicated by a diamond ( $\blacklozenge$ ). When varying a hyperparameter, the other hyperparameter is kept constant at the default value ( $c = 0.65$ ,  $P_0 = 0.8$ ).

We mentioned that the optimal MESD for averaging of auto-correlation matrices is obtained at a Markov chain of seven states. Fig. 2 shows the optimal number of states  $\hat{N}_\tau$  and target state  $k_c$  per decision window length (see Section II-E and Algorithm 1) and the ESD per decision window length, at the optimal number of states  $\hat{N}_\tau$ . It is over this curve that the ESD is minimized to obtain the MESD (Section II-E and Algorithm 1).

In Fig. 2, it can be seen that when  $\hat{N}_\tau$  remains constant, the ESD increases almost linear with decision window length  $\tau$ . In (10), when the number of states  $N$  and thus target state  $k_c$ , remains constant, it appears that the step time  $\tau$  is the dominant factor over the variation in transition probability  $p$ . This implies that the interesting decision window lengths coincide with changes in the number of states. Relative to  $\hat{N}_\tau = 7$  at the MESD, an increase in decision window length results in a decrease of  $\hat{N}_\tau$  to five. However, the target state  $k_c$  only decreases from five to four, such that the drop in ESD around  $\approx 6$  s is not large enough to decrease below the minimal ESD for seven states. When decreasing  $\tau$ ,  $\hat{N}_\tau$  and  $k_c$  increase steeply because of the steep decrease in accuracy (Fig. 4), which is not sufficiently compensated by the small decrease in step time  $\tau$ . The AAD accuracy  $p$  (depending on decision window length  $\tau$ ) thus mainly plays a role in determining the optimal number of states  $\hat{N}_\tau$  via the design constraints (Section II-C), which is the first step in optimizing the ESD (Section II-E and Algorithm 1), while the transition points of  $\hat{N}_\tau$  are most interesting for minimizing the ESD to obtain the MESD, as the ESD almost linearly increases with  $\tau$  for a constant  $\hat{N}_\tau$ .

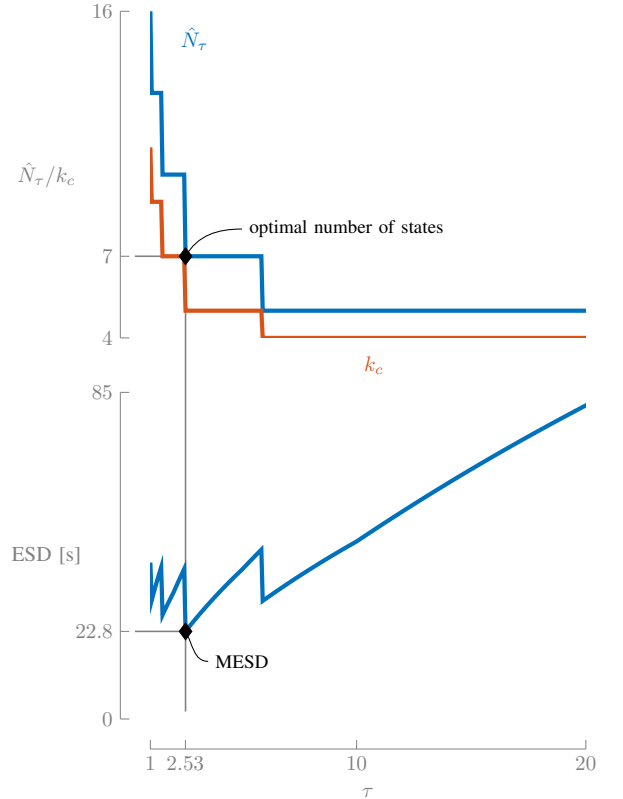


Fig. 2: The optimal number of states  $\hat{N}_\tau$  and corresponding target state  $k_c$  decrease in function of the decision window length  $\tau$ . The minimal ESD (MESD) depends both on the optimal number of states (via the AAD accuracy) and the decision window length.

## REFERENCES

- [1] R. McGarrigle, K. J. Munro, P. Dawes, A. J. Stewart, D. R. Moore, J. G. Barry, and S. Amitay, "Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'." *Int. J. Audiol.*, vol. 53, no. 7, pp. 433–445, 2014.
- [2] L. Decruy, N. Das, E. Verschuere, and T. Francart, "The Self-Assessed Békésy Procedure: Validation of a Method to Measure Intelligibility of Connected Discourse," *Trends Hear.*, vol. 22, pp. 1–13, 2018.