

BACHELOR THESIS
ARTIFICIAL INTELLIGENCE

Radboud University



**Subclass-Aware Decoding of Auditory
Event-Related Potential Data**

Author:
Ella Has
s1051869
ella.has@ru.nl

First supervisor:
dr. M.W. Tangermann
Dept. of Artificial
Intelligence
michael.tangermann@donders.ru.nl

Second reader:
dr. J. Thielen
Dept. of Artificial
Intelligence
jordy.thielen@donders.ru.nl



June 22, 2023

Abstract

A language training task using Brain-Computer Interfacing (BCI) for patients with aphasia [1] has had promising results. The task consists of listening to six words of which one is the target word. The BCI classification model classifies the participant's brain response to each word as either target or non-target. If it successfully classifies one word as target and the others as non-target, and the classified 'target word' matches the actual target word, the participant receives positive feedback.

As correct feedback is key for successful training, the classification of the words can always be improved upon. One possible avenue of improving the performance is using the identity of the words as subclasses in a classification model. The aim of this paper is to explore different ways of exploiting the subclass information. The subclass-aware decoding methods relevance subclass LDA (RSLDA) [2], which uses multi-target shrinkage of the subclass means, and Riemannian classification in tangent space after parallel transport [3], as well as a novel Riemannian sub-prototype model, were tested on the word-evoked auditory event-related potential (ERP) data from the above-mentioned language training task.

None of the subclass-aware models performed significantly better than their global counterpart. However, for some words, the sub-prototype model performed quite well. Additionally, it is shown again that multi-target shrinkage of the mean [2] is an improvement over estimating the mean from subclass data only. Unfortunately, RSLDA could not be applied on the parallel transport features (as is done by Kolkhorst and colleagues [3]). Given the benefit of RSLDA over separate classifiers in Euclidean space, the performance of this approach on the language training data set remains an interesting question.

Contents

1	Introduction	3
2	Background Riemannian Geometry	5
2.1	Riemannian Manifolds and Distance	5
2.2	BCI Classification with Riemannian Geometry	5
2.2.1	Sample Covariance for ERP	5
2.2.2	Tangent Space	6
3	Methods	7
3.1	Data Set	7
3.2	Preprocessing	7
3.3	Classifiers	8
3.3.1	Euclidean Approaches	8
3.3.2	Riemannian Approaches	10
3.4	Evaluation	11
4	Results	13
4.1	Average ERP Responses	13
4.2	Euclidean Approaches	13
4.2.1	Overall Performance	13
4.2.2	Performance per Subclass	15
4.3	Riemannian Approaches	15
4.3.1	Overall Performance	15
4.3.2	Performance per Subclass	15
4.3.3	Parallel Transport and RSLDA	16
5	Discussion	25
5.1	Euclidean Approaches	25
5.2	Riemannian Approaches	25
5.2.1	Parallel Transport	25
5.2.2	Sub-Prototypes	26
5.3	Limitations	26
5.4	Conclusion	27
6	Acknowledgments	28
7	Source Code	29
References		30

A Statistical Tests	32
B Additional Figures	33
B.1 Average EEG Responses Channel Cz	33
B.2 Euclidean ROC Curves	42
B.2.1 Block-Toeplitz	42
B.2.2 Ledoit-Wolf	57
B.3 Riemannian ROC Curves	66
B.3.1 Parallel Transport	66
B.3.2 Sub-Prototype	75

Chapter 1

Introduction

Brain-Computer Interfacing (BCI) paradigms are used for a variety of medical purposes, from assistive technologies that help patients communicate, like spelling paradigms [4], to rehabilitation of cognitive function, facilitating cortical plasticity [5]. Especially motor rehabilitation through BCI training is well researched, but also cognitive impairments such as attention or memory deficits can be potentially rehabilitated with the help of BCI [6]. Patients with aphasia after stroke may benefit from BCI rehabilitation paradigms as well. In a pilot study, motor aphasia patients participated in a BCI spelling paradigm [7]. At first, the EEG signals were not classifiable, but after training with the aid of a piece of cardboard covering the non-target letters, participants were able to spell freely without the aid. Beside providing a tool for communication, this BCI paradigm aims to facilitate rehabilitation. Three of the four participants showed an increase in the P300 response to target letters. This suggests that the participants can learn to elicit classifiable responses.

Another language training task to help patients with aphasia recover was developed by Musso and colleagues [1]. Based on the multi-speaker auditory oddball task [8], where participants listen to different tones of which one is the target tone, the patients listen to words and are asked to attend to the target word. A cueing sentence is given at the start of a trial. Then six different words are played in rapid succession, each 15 times, each from a different speaker. The target word is the one that properly finishes the sentence. A BCI decodes whether the participant attended to the correct stimulus. If the BCI decodes that the target word was attended to, the participant gets positive feedback. If the BCI cannot distinguish which word was attended to, or it classifies a non-target word as the word attended to, the participant gets neutral feedback, telling them which word was correct. The participants improved substantially over the course of the training.

However, different words may elicit different responses. For example, the onset of the P300 could be somewhat shifted [8]. The different words could be seen as subclasses of the main classes 'target' and 'non-target'. These subclass labels may be valuable information, but are left unused in most classifiers, like Linear Discriminant Analysis (LDA), which is commonly used in BCI applications [2]. Since correct feedback is key to successful training [1], it is worth exploring whether the subclass information may improve classification performance. This paper explores several ways of incorporating subclasses into the design of classification models.

One such subclass-aware model is relevance subclass LDA (RSLDA) [2]. A classifier is trained for each subclass, and the mean of each subclass is regularized with the data of the other subclasses to make a better estimation. The idea is that the other subclasses hold some information about the true mean of the subclass since the subclasses are quite similar, and so the estimate of the mean can be improved by using the other subclasses.

A different subclass-aware method was proposed by [3]. They compare separate classifiers, subclass-regularized classification, and global classification on visual data. Global classification refers to approaches where the subclass labels are not taken into account. They found that separate classifiers do not perform well with limited data, as each subclass is too small. Subclass-regularized classification did outperform global classification on the visual data. The authors use Riemannian geometry to create the features and bring the means of the subclasses together. The features are the covariance matrix of each trial. To keep the temporal information, the cross-covariance of the trial with a prototypical ERP is also included. The covariance matrices lie on a Riemannian manifold, but they are projected onto their tangent space. This makes it possible to apply classifiers in euclidean space on the covariance data points. However, each subclass mean is located on a different tangent space. A mean cannot be regularised towards a mean in a different space. To solve this, parallel transport is used to center the subclasses before projection, so all means are in the same tangent space.

In this paper, different approaches to using subclass information are trained and evaluated on data from the aphasia rehabilitation task [1]. The subclass-aware models are compared to their global counterpart to isolate the effect of the way subclass information is used. To that end, a pair of a global and a subclass-aware model shares the same preprocessing pipeline, as well as any other implementational details that do not involve subclass information, such as the shrinkage approach for the covariance.

The research question is: Can subclass-aware decoding methods improve performance of the classification of word-evoked auditory Event-Related Potential (ERP) data from EEG recordings compared to global classification?

First follows a brief introduction to the intuition and mathematics of Riemannian classification. The different models will be detailed in Chapter 3, the Methods. Then the performance evaluations are presented in Chapter 4, the Results, and discussed in Chapter 5, the Discussion.

Chapter 2

Background Riemannian Geometry

Before the models are described, we first give a brief introduction to BCI classification using Riemannian geometry. The intuition of Riemannian manifolds and Riemannian distance are explained, as well as how BCI classification can be done with Riemannian space.

2.1 Riemannian Manifolds and Distance

In Riemannian classification, the covariance matrix of each epoch is used for features. These sample covariance matrices exist on a manifold, a smooth space that is locally Euclidean [9]. The distances between the covariance matrices are not measured with the length of the straight line connecting them, like in Euclidean space, but instead with the length of the geodesic, which is the shortest curve along the manifold that connects two points. We will not dive into the mathematics of this, since that is beyond the scope of this paper. The mean of the covariance matrices is determined by minimizing the sum of the squared lengths of the geodesics that connect the mean to each of the points. This problem has no closed form solution [10].

2.2 BCI Classification with Riemannian Geometry

Riemannian classification is useful for BCI because of its computational efficiency and the ease with which transfer learning can be done due to the invariance to congruent transformation [10]. It was at first mainly used for motor imagery, as sample covariance matrices do not account for the temporal information in the time series data. It takes a trick to apply Riemannian classification to event-related potential (ERP) data. Another trick is required if you want to use one of the classification models that are commonly used in Euclidean space.

2.2.1 Sample Covariance for ERP

For ERP paradigms, the temporal information cannot be missed. The standard sample covariance matrices are thus unsuitable as data points. However, this problem can be solved by appending a prototypical ERP response for each class to a trial to get a so-called super trial. Equation 2.1 shows what the super trial looks like, with **P1** and **P2**

being the prototypes of class 1 and 2 respectively. The prototype can be obtained by averaging all epochs of a class.

$$\tilde{\mathbf{X}}_i = \begin{bmatrix} \mathbf{P1} \\ \mathbf{P2} \\ \mathbf{X}_i \end{bmatrix} \quad (2.1)$$

When the covariance of this super trial $\tilde{\Sigma}_i$ is then calculated, part of the covariance matrix represents the cross-covariance between the trial and the prototypical responses, \mathbf{C}_{P1,X_i} and \mathbf{C}_{P2,X_i} (see Equation 2.2). The cross-covariance contains information about whether the trial is in phase with a prototype, and so it preserves the temporal information.

$$\tilde{\Sigma}_i = \begin{bmatrix} \Sigma_{P1} & \mathbf{C}_{P1,P2}^\top & \mathbf{C}_{P1,X_i}^\top \\ \mathbf{C}_{P1,P2} & \Sigma_{P2} & \mathbf{C}_{P2,X_i}^\top \\ \mathbf{C}_{P1,X_i} & \mathbf{C}_{P2,X_i} & \Sigma_{X_i} \end{bmatrix} \quad (2.2)$$

2.2.2 Tangent Space

The usual classification methods like Linear Discriminant Analysis (LDA) cannot be directly applied in Riemannian space. However, the data points can be projected onto the tangent space in a way that preserves the relative distances [11]. The tangent space is a euclidean space defined by all the tangent vectors of a tangent point on the manifold. The tangent point is generally taken to be the Riemannian mean of the data set. In the tangent space, all the usual classification methods can be applied, since this is a euclidean space.

Chapter 3

Methods

This chapter details the methods and data set used. First, I describe the data set. Then the methods of preprocessing are outlined. Finally, the different classifiers are detailed. Each of the classifiers is either global or subclass-aware, and it either works in Euclidean space or Riemannian space.

Global classifiers do not use subclass information. The subclass-aware classifiers do use the subclass information, in this case which word the stimulus is. Each subclass contains both target and non-target epochs. Subclass information is known, while the class label, target/non-target, needs to be decoded. Three different subclass-aware approaches are tested, relevance subclass LDA [2], a Riemannian parallel transport approach [3], and a novel Riemannian approach.

3.1 Data Set

The classification approaches are tested on data from the aphasia training by Musso and colleagues [1]. The data was supplied by Michael Tangermann. It is not publicly available. The participants listen to an unfinished sentence and then hear a series of words. They are asked to pay attention to the word that finishes the sentence while ignoring all other words.

Only data from normally aged controls is used, since the set of words does not differ between these sessions. Additionally, there was no feedback nor dynamic stopping in these sessions. The data from three participants is used.

Per participant, there is one session. The sessions consist of six runs of each of the three conditions: single speaker, six speakers, and headphones. In a run, each word is the target word in one trial, so there are six trials per run and 36 trials for each condition per participant. In each trial, all words are played fifteen times. In total, there are 540 stimuli per run, 3240 stimuli per condition, and 9720 stimuli per session. The data contains 69 channels, of which 63 are EEG channels. The sampling rate is 1000 Hz. Per session there are 1.319.760 time points recorded for each condition. The first four runs of the same condition within a session form the training data. The remaining two runs are the holdout set. Each model is trained and tested separately on each of the conditions.

3.2 Preprocessing

The timeseries data is filtered to [0.5, 16] Hz with a zero-phase Butterworth filter. The highpass is the same as [1], while the lowpass is set to a higher frequency to potentially

capture differences between subclasses at frequencies up to 16 Hz. The EEG data is then downsampled to 100 Hz. Epochs are cut from -0.2 s to 1.2 s relative to the stimulus onset. The training set of each condition consists of 360 target epochs and 1800 non-target epochs. Since the models are compared on the same preprocessed data, it does not matter for the comparison whether outliers are removed, assuming that neither is better able to handle outliers than the other. Hence, neither outlier channels nor outlier epochs are removed from the training data.

3.3 Classifiers

Several subclass-aware classifiers are compared to their global counterpart. Preprocessing is equal in each pair of classifiers.

3.3.1 Euclidean Approaches

The models working in Euclidean space with time window features are described below.

Feature Extraction

The features for the Euclidean space LDA approaches are determined by the average of each of ten time windows per channel, of which the boundaries are equally spaced between 0 s and 1.2 s relative to the stimulus onset, [0. 0.12 0.24 0.36 0.48 0.6 0.72 0.84 0.96 1.08 1.2]. For each channel, the average activity of each of the time windows is calculated, giving a total of $63 \times 10 = 630$ features.

Global Classifiers

The Euclidean subclass approaches are compared to Shrinkage LDA (sLDA) and Block-Toeplitz LDA (ToeplitzLDA). These models are chosen as the baseline because they differ from relevance subclass LDA only in whether or not the subclass information is used. These two flavours of covariance shrinkage are chosen because sLDA is commonly used and ToeplitzLDA performs better on EEG data.

Shrinkage LDA shrinks the covariance matrix with Ledoit-Wolf shrinkage ([12]). The implementation is offered by Scikit-Learn ([13]). Block-Toeplitz LDA is designed and implemented by [14]. Block-Toeplitz covariance shrinkage enforces stationarity on the non-event related signals. The assumption is that noise is stationary, while the event-related potential is time-locked. The enforced stationarity leads to a block-wise symmetric covariance matrix. The symmetry greatly reduces the number of free parameters that need to be estimated. That makes the approach less prone to overfitting.

Subclass-Aware Classifiers

The investigated subclass-aware Euclidean approaches are separate LDA classifiers for each subclass (sep. LDA), and relevance subclass LDA (RSLDA), proposed by [2].

Separate LDA simply separates the data by subclass and trains an LDA model for each of them. The means and covariance matrices are estimated separately for each subclass. A test data point gets classified with the model corresponding to the data point's subclass.

In RSLDA, each subclass gets a separate LDA model as well, but the difference lies in the means of the subclasses, which are estimated by shrinking it towards the means of the other subclasses.

Training the RSLDA classifier follows the following steps. First, the mean of each subclass is estimated. Then the appropriate mean is removed from each data point in the full set. The covariance is estimated on the full mean-free data set. Finally, an LDA model is trained for each subclass using the estimated subclass mean and the global covariance.

The means are estimated using multi-target shrinkage ([15]). Multi-target shrinkage (MTS) of the mean can be understood as a weighted average of the full data set. If the subclass distributions only differ in mean, then the data points of the current subclass are weighted highest and data points from all other subclasses, the targets, are weighted less or equal. First, the mean of each subclass is calculated. The estimated mean of the current subclass is denoted with $\hat{\mu}$, with dimensionality $1 \times N$, where N is the number of features. \hat{T} holds the estimated target means and has dimensionality $s - 1 \times N$, where s is the number of subclasses. Then, the parameters of the optimization problem are calculated. \hat{A} in equation 3.1 represents the estimated quality of the targets, or how close the targets are to the to-be-estimated mean. Its dimensionality is $s - 1 \times s - 1$. \hat{b} in equation 3.2 contains the estimated variance \widehat{Var} of the targets and has a length of $s - 1$.

$$\hat{A}_{kl} = \sum_{i=1}^p (\hat{T}_i^k - \hat{\mu}_i)(\hat{T}_i^l - \hat{\mu}_i) \quad (3.1)$$

$$\hat{b}_k = \sum_{i=1}^p \widehat{Var}(\hat{T}_i^k) \quad (3.2)$$

We then get the optimization problem in equation 3.3, where λ is the optimal shrinkage intensity for each target subclass. The dimensionality of λ is $s - 1$. Nearby targets with low variance are the most informative for the subclass mean, and so these have the highest weights. Two constraints are imposed on the solution: $\sum_k \lambda_k \leq 1$, as the total weights including of the current subclass must add up to one, and each element of λ is non-negative, since target means cannot contain negative information on the subclass mean. The optimization with the constraints form a quadratic program.

$$\hat{\lambda} = \arg \min_{\lambda} \frac{1}{2} \lambda^\top \hat{A} \lambda - \hat{b}^\top \lambda \quad (3.3)$$

The shrunk subclass mean is then the weighted average of all subclass means, where λ contains the weights of the targets, and the weight of the current mean is determined by the sum of the target weights.

$$\hat{\mu}^{MTS}(\lambda) = (1 - \sum_{k=1}^K \lambda_k) \hat{\mu} + \sum_{k=1}^K \lambda_k \hat{T}_k \quad (3.4)$$

Once the data is made mean-free with the estimated subclass means, the global covariance is estimated with either Ledoit-Wolf or Block-Toeplitz shrinkage. The model is compared to the global model with the same type of covariance shrinkage.

It is also tested how the Block-Toeplitz relevance subclass LDA model performs if one class mean is estimated per subclass and the other is estimated globally. This is referred to in this paper as a hybrid subclass-aware model because it has global and subclass-aware aspects.

See Chapter 7 for the Python implementation of RSLDA and Sep. LDA. The implementations use the libraries Scikit-Learn 1.2.2 [13], qpsolvers 3.4.0 [16], and Numpy 1.24.2 [17], as well as the Block-Toeplitz covariance implementation from [14] from the module toeplitzlda 0.2.6.

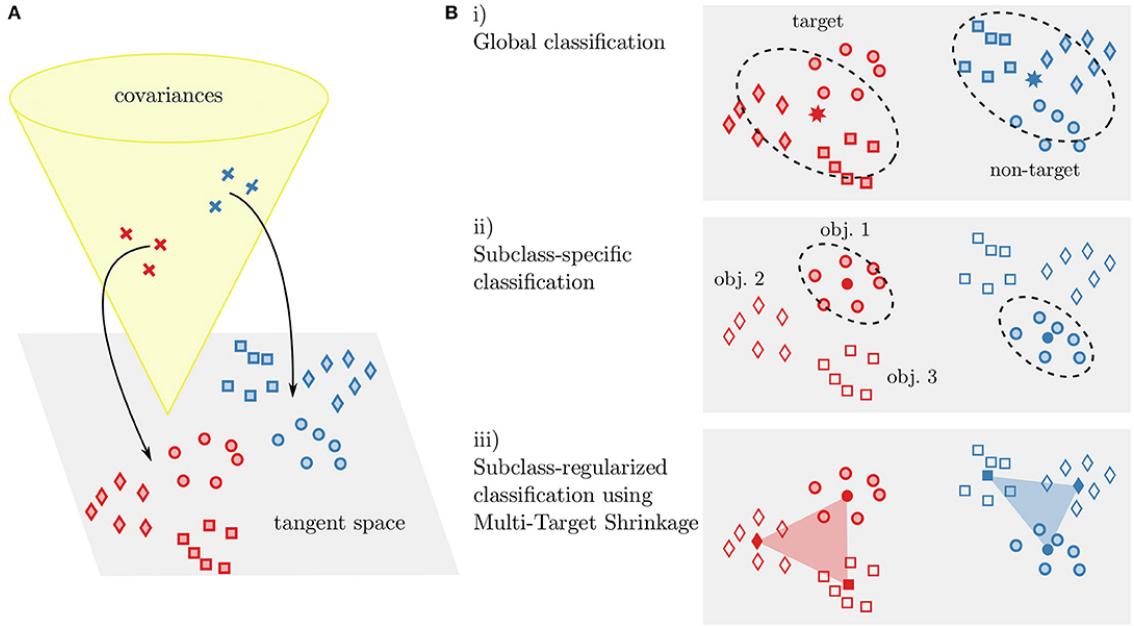


Figure 3.1: Illustration of Riemannian classification approaches [3].

- A) The covariance matrices in Riemannian space are projected onto tangent space.
- B) i: Global classification (TS+LDA), ii: subclass specific classification (Sep. TS+LDA), iii: subclass-regularized classification (Sep. TS+RSLDA) are applied in tangent space

3.3.2 Riemannian Approaches

The following models work in a Euclidean tangent space of the Riemannian space of sample covariance matrices. A sample covariance matrix is the covariance matrix of an epoch, as opposed to the full data set. They are used as the features of an epoch. Although the models work in Euclidean space, they are here referred to as Riemannian approaches because they use features that are obtained through Riemannian space. See Figure 3.1 for an illustration of the global model and two of the subclass-aware models [3].

Feature Extraction

The preprocessed data (see section 3.2) is filtered with two xDAWN [18] spatial filters per class, so four in total, to reduce dimensionality. Then the sample covariance matrices are estimated for each epoch. Since covariance matrices do not contain temporal information, a prototypical ERP for each class is appended to the epoch to form a super trial [10]. Part of the covariance matrix then consists of the cross-covariance of the epoch with each of the prototypical responses. The cross-covariance contains information about whether the epoch's response is in phase with the prototype, and thus preserves the temporal information, which would otherwise have been lost in a spatial covariance matrix. The dimensionality of a sample covariance matrix is $(p + 1) \cdot 4 \times (p + 1) \cdot 4$, where p is the number of prototypes, which is 2 in the basic case, but is increased for sub-prototype LDA. Sub-prototype LDA is explained below.

The obtained covariance matrices, which are in Riemannian space, are projected down onto the tangent space. There they can be classified with any Euclidean-space method, like LDA.

Global Classifiers

The Riemannian subclass-aware classifiers are compared to a global approach (TS+LDA) [3]. For this classifier, the features are extracted and projected onto the tangent space as described above, where an LDA model is trained on the projected data points.

One variant is tested to check the benefit of fitting the tangent point before projection. The tangent point, which is the global Riemannian mean, is computationally expensive to find. However, it is also possible not to fit the tangent point and simply take the identity matrix \mathbf{I} to be the tangent point. This classifier is referred to as unfit-TS+LDA.

Subclass-Aware Classifiers

Besides changing the classification approach, the subclass-aware classifiers make some modifications to the feature extraction process.

The separate classifier approach (Sep. TS+LDA) [3] performs parallel transport on the sample covariance matrices before projecting them onto the tangent space. This is to bring the means of each subclass together. Each sample covariance is transported along the geodesic $\Gamma_{\mathbf{C}_j \rightarrow \mathbf{I}}$ that connects the subclass mean \mathbf{C}_j of subclass j and the identity matrix \mathbf{I} (see Equation 3.5), such that the subclass mean of each subclass ends up being \mathbf{I} after transportation of all covariance matrices. After projecting the transported data onto tangent space, three different models are trained: a global LDA model, which uses the subclass information only in the parallel transport stage, a separate LDA model for each subclass, and finally a relevance subclass LDA model. The covariance matrix of the projected data is estimated globally for the separate LDA models for a better estimate.

The attempt to combine RSLDA with parallel transport (TS+RSLDA) was unsuccessful, since the quadratic program solver could not solve the optimization problem of multi-target shrinkage with the transported and projected data. See Chapter 5 for a discussion of the potential problems.

$$\Gamma_{\mathbf{C}_j \rightarrow \mathbf{I}}(C) = (\mathbf{C}_j)^{-1/2} \mathbf{C} (\mathbf{C}_j)^{-1/2} \quad (3.5)$$

Another change to the feature extraction can be made to incorporate subclass information: a novel approach called sub-prototype TS+LDA. Instead of appending a prototype for each class to create the super trial, the average responses per subclass for each class are appended. In the case of the aphasia training data [1], this means a total of twelve prototypes, two for each of the six subclasses. The cross-covariances now show whether a trial is in phase with each of the prototypical subclass responses.

Several combinations of the three subclass-aware components with Riemannian features (separate classifiers, parallel transport, and subclass prototypes) are investigated. See Table 3.1 for all combinations, including the Euclidean approaches.

See Chapter 7 for the implementation of parallel transport. The libraries Numpy 1.24.2 [17], Scipy 1.10.1 [19], and pyRiemann 0.5 [11] are used.

3.4 Evaluation

The models are evaluated on the holdout data set of the two last runs of the session, which contains 180 target epochs and 900 non-target epochs. They are compared based on the area under the ROC curve (AUC). Statistical tests are only performed in cases where a subclass-aware model outperforms the global model. The difference in performance of two models is evaluated with a one-sided Wilcoxon signed-rank test, comparing the AUC

Model Name	R	BT	SC	MS	SP	PT
Global sLDA						
Global ToeplitzLDA		x				
Sep. sLDA			x			
Sep. ToeplitzLDA		x	x			
sRSLDA			x	x		
ToeplitzRSLDA		x	x	x		
TS+LDA	x					
Sub-prototype TS+LDA	x				x	
PT TS+LDA	x					x
Sep. TS+LDA	x		x			x
Sub-prototype PT TS+LDA	x				x	x
Sep. sub-prototype TS+LDA	x		x		x	x
TS+RSLDA	x		x	x	x	x

Table 3.1: Model overview. Legend: R = Riemannian, BT = Block-Toeplitz covariance estimation, SC = Separate classifier per subclass, MS = Multi-target mean shrinkage, SP = Subclass prototypes, PT = Parallel transport.

for each word. This test is chosen because it is a paired test and does not assume normal distributions [20]. It is chosen to be one-sided, because only when a subclass-aware model performs better than the global counterpart, there is a statistical test, so it is already assumed to perform better, not worse. The significance level α is 0.05. With Bonferroni correction [21], that becomes $\alpha = 0.05/n$ where n is the number of tests. The number of tests turns out to be 13, and so $\alpha = 0.05/13 = 0.0038$. No significance testing was done on the performance at word level due to the limited sample size.

Chapter 4

Results

This chapter details characteristics of the data sets and the results of the comparison between subclass-aware and global approaches. For additional figures, see the Appendix.

None of the subclass-aware models outperformed the global baseline performance consistently, but the different approaches vary in how close they get to the baseline. For example, parallel transport (PT TS+LDA) generally performs similarly to its global equivalent, while separate LDA classifiers in several conditions barely outperforms chance level. Performance is generally worse for the conditions with one speaker or headphones than with six speakers. The condition affects all methods similarly. However, some subclass-aware models seem to deal better with certain conditions than global methods, although the differences are non-significant.

4.1 Average ERP Responses

As can be seen in Figure ??, the average ERP responses in channel Cz per class of each of the words differs from the global class means. The difference is more noticeable for the target means than for the non-target means. This could be due to the fact that there are five times as many non-target epochs as target epochs, so the noise is better averaged out for the non-target means.

4.2 Euclidean Approaches

The Euclidean baseline approaches usually perform better than the subclass-aware models, but some models come close in certain conditions.

4.2.1 Overall Performance

Block-Toeplitz relevance subclass LDA (ToeplitzRSLDA), where the subclass means are shrunk towards each other, performs better than separate classifiers for Block-Toeplitz LDA (ToeplitzLDA), but does not exceed baseline performance in most cases. Only for one participant under one condition, participant 1 with one speaker (1D), RSLDA has a slightly larger area under the ROC curve (AUC ROC) than the global ToeplitzLDA, as can be seen in Figure 4.2. However, this difference is not significant. This same data set, participant 1 under condition 1D, also has shrinkage RSLDA (sRSLDA) outperform global shrinkage LDA (sLDA) (see Figure 4.3). In general though, sRSLDA does not outperform sLDA.

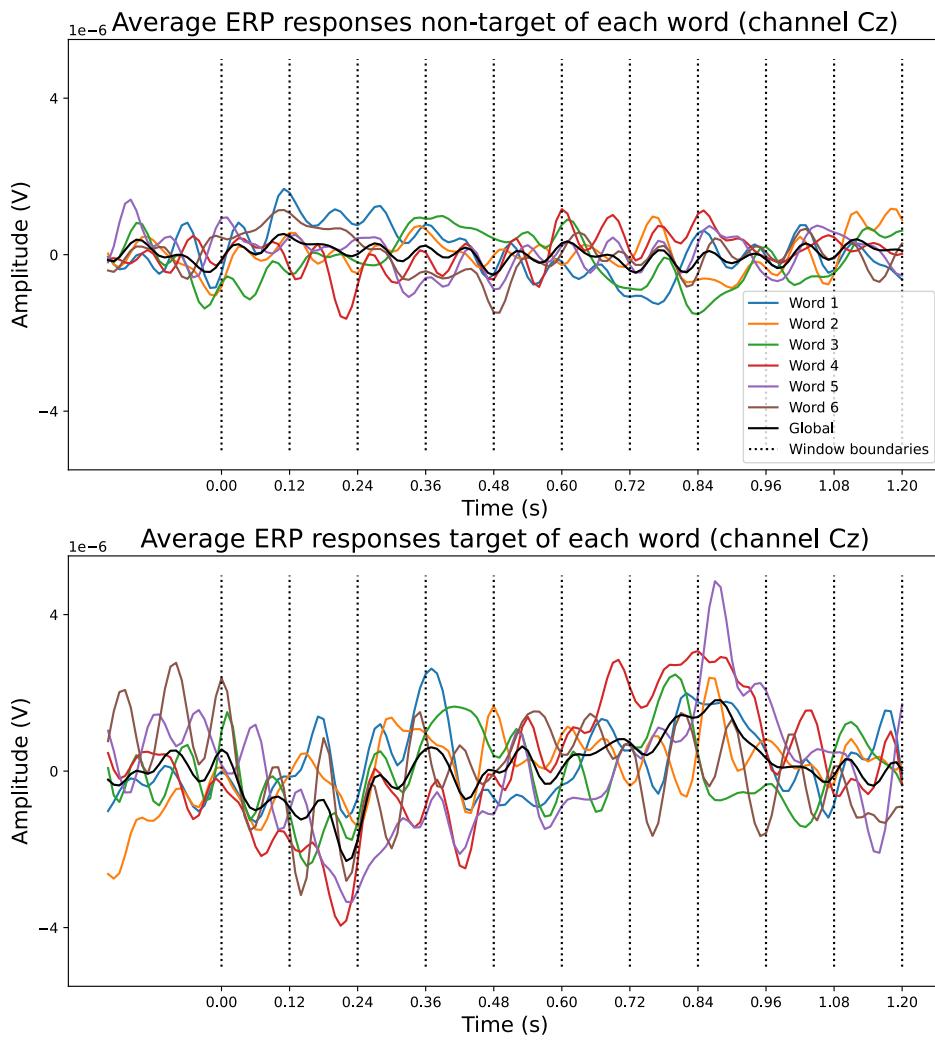


Figure 4.1: Average ERP response to non-target and target words (channel Cz), participant 2, condition 6D.

Notably, the differences between relevance subclass LDA (RSLDA) and separate LDA are smaller when using Ledoit-Wolf shrinkage than Block-Toeplitz shrinkage. In some cases, separate LDA even performs better than RSLDA.

If only the non-target mean is estimated per subclass, but the target mean is estimated globally, performance is better than when both are estimated per subclass or only the target mean is estimated per subclass, in all but one case (participant 2, headphone (HP) condition).

4.2.2 Performance per Subclass

There are some differences in classification performance of the Euclidean approaches between words. Information about which word is denoted with which number was not available and word numbering likely differs per participant, so no comparisons of word difficulty can be made between participants. No significance testing was done on the word level, so no observations can be reliably generalized.

For some words, relevance subclass LDA (RSLDA) and separate classifiers work reasonably similar, like for participant 3, condition 6D, with Block-Toeplitz shrinkage in Figure 4.5, words 1, 3, and 4 have an equal or higher AUC score for RSLDA than global ToeplitzLDA. However, for the other words RSLDA performs worse than global ToeplitzLDA.

4.3 Riemannian Approaches

In general, Riemannian classification performs more poorly than the Euclidean models. Under the one speaker (1D) and headphone (HP) conditions, all models, global and subclass-aware, perform around chance level. For the six speakers condition, some subclass-aware models perform near or over the level of global Riemannian classification (TS+LDA).

4.3.1 Overall Performance

There is no significant difference between the performance of parallel transport of subclass means and global classification. These models perform approximately equally well in most cases. See for example Figure 4.6.

Separate classifiers perform worse in any combination, whether combined with parallel transport or sub-prototypes. There are only two cases of separate parallel transport and one of sub-prototypes with separate LDA models performing any better than chance. See Figure 4.7 for a typical separate LDA performance for participant 3 and a untypical performance for participant 2.

Sub-prototype LDA performs similar to global classification for two out of three participants under the six speaker condition (see Figure 4.7). Combining sub-prototype LDA and parallel transport with global classification performs similarly to doing only one or the other.

4.3.2 Performance per Subclass

While parallel transport performs more consistently than sub-prototype LDA, sub-prototype LDA outperforms parallel transport and global LDA for certain words. For most words,

the Riemannian subclass-aware approaches perform worse than global Riemannian classification, but for some words a subclass-aware approach works much better. Keep in mind that no significance testing was done on the word level performance.

Parallel transport performs similarly to global performance across words, just like it does overall. Separate classifiers after parallel transport do differ in performance across words. For participant 1 under condition 6D, separate classifiers perform similarly to global performance for most words, but it has a near-chance performance for word 1, as can be seen in Figure 4.8.

The sub-prototype approaches perform better than global approaches for specific words under specific conditions. As you can see in Figure 4.9, for participant 1 under condition 6D, sub-prototypes perform better than global LDA for word 5 and similar to global LDA for word 2. Participant 2, who has a relatively high performance of sub-prototype approaches in general, word 3 is classified well with the sub-prototypes. All other words are also classified with similar performance to the global approaches.

4.3.3 Parallel Transport and RSLDA

As described in Chapter 3, the solver of the quadratic program of multi-target shrinkage of the mean fails to give a result on parallel transported data. This is true for all participants and conditions. Some details are described here about the failure to solve the optimization.

First of all, the solver terminates in similar time no matter the maximum number of iterations. The maximum number of iterations was set to the largest number that could be represented, and still the program ended within 5 seconds, which is similar to the time taken with the default number of iterations of 1000. Whether the data is whitened or not also did not make a difference. Puzzlingly, the solver gives no errors about the problem being invalid nor about the solver failing during execution, and yet it returns `None` to indicate that no solution was found. Finally, although the bounds on the solution of a maximum sum of weights of 1 and non-negative values for each weight are important, it was attempted to obtain a solution by removing these bounds. However, this also results in no optimal solution found.

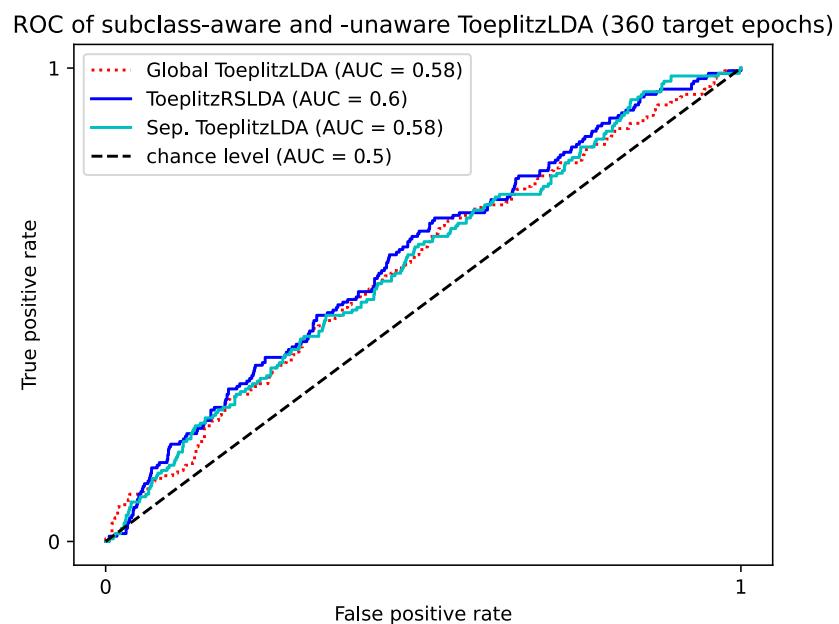
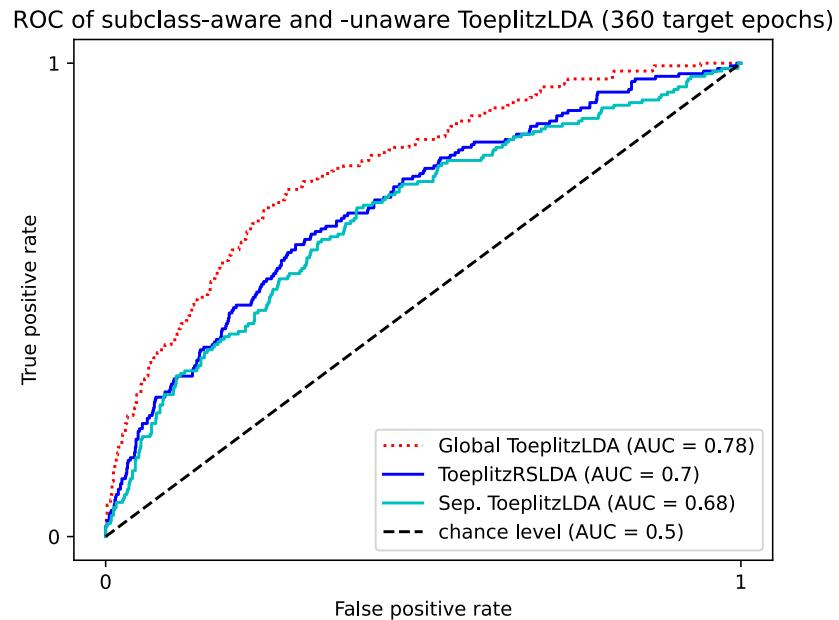


Figure 4.2: Receiver Operator Characteristic curves of Block-Toeplitz approaches for participant 1, condition 6D (top) and 1D (bottom).

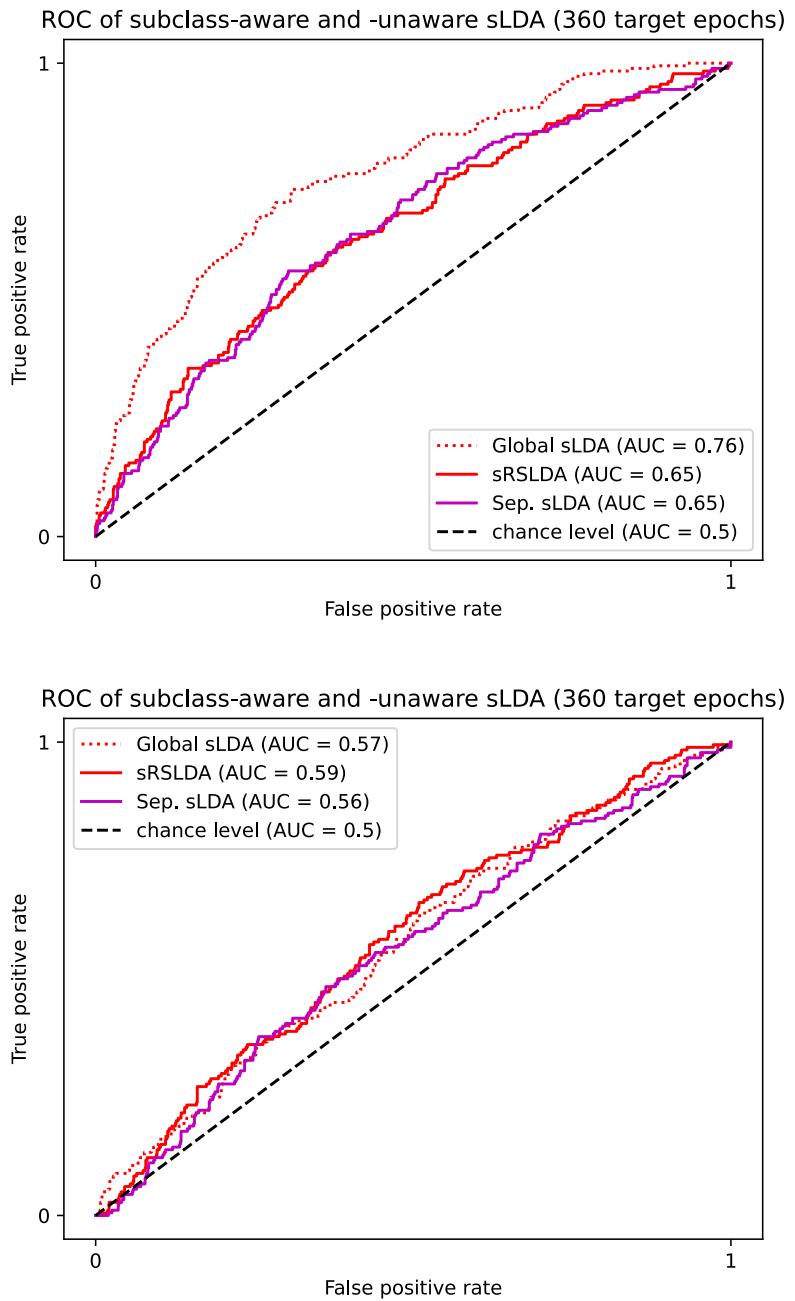


Figure 4.3: Receiver Operator Characteristic curves of shrinkage LDA approaches for participant 1, condition 6D (top) and 1D (bottom).

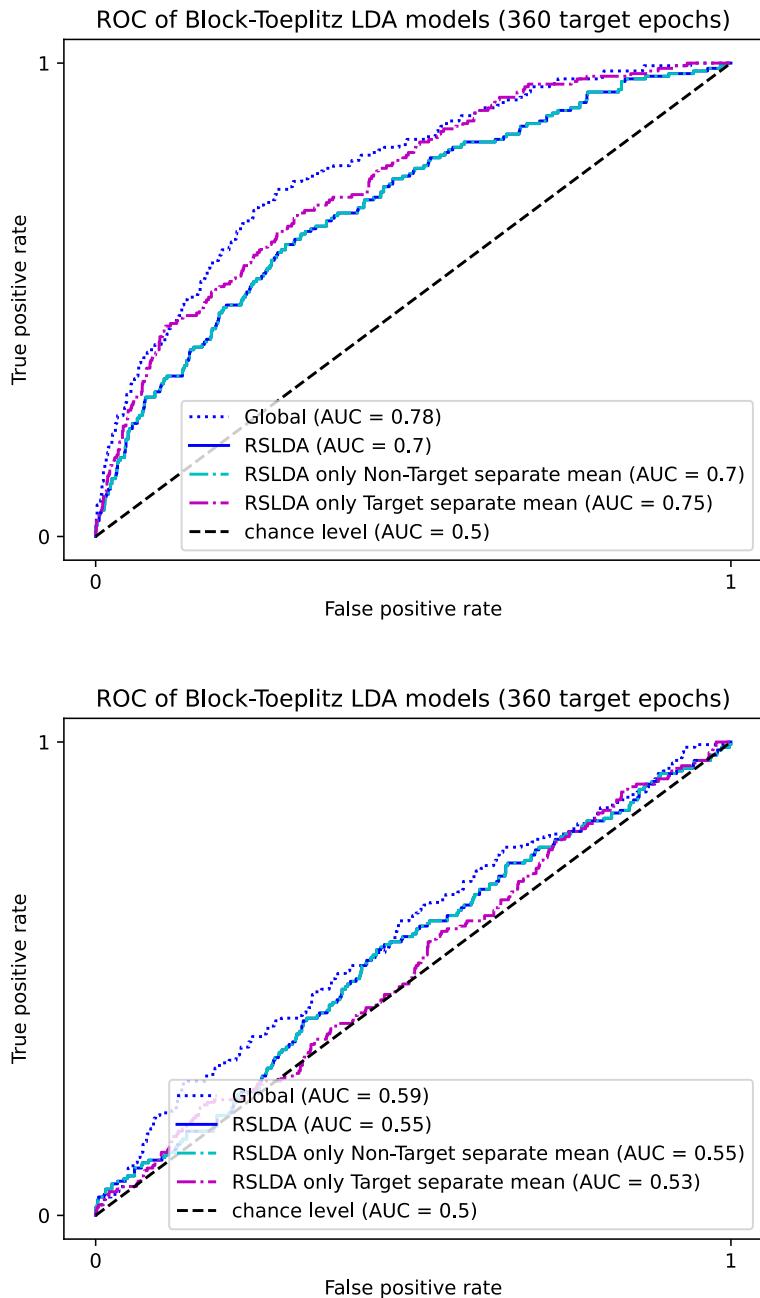


Figure 4.4: Receiver Operator Characteristic curves of hybrid Block-Toeplitz relevance subclass LDA with global target mean or global non-target mean for participant 1, condition 6D (top) and participant 3, condition HP (bottom).

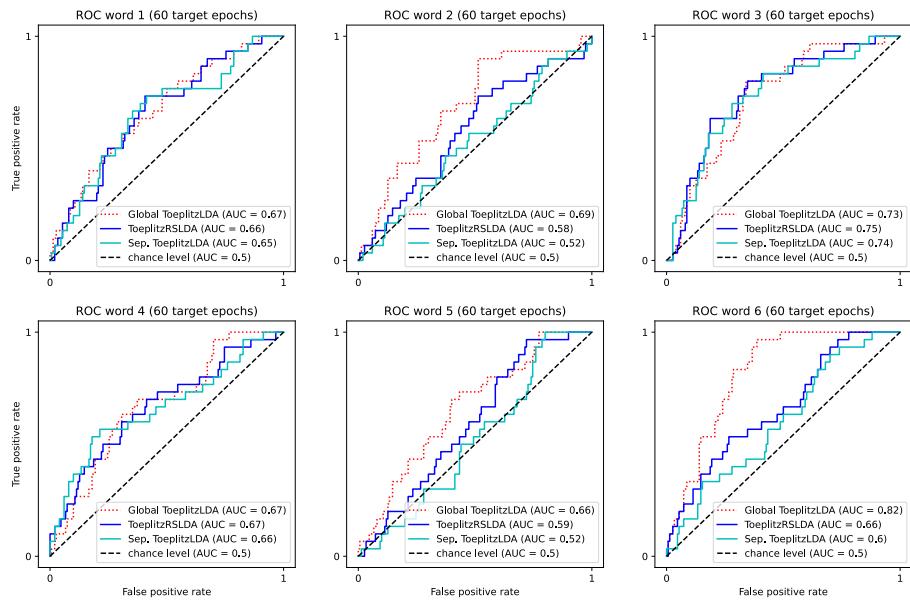
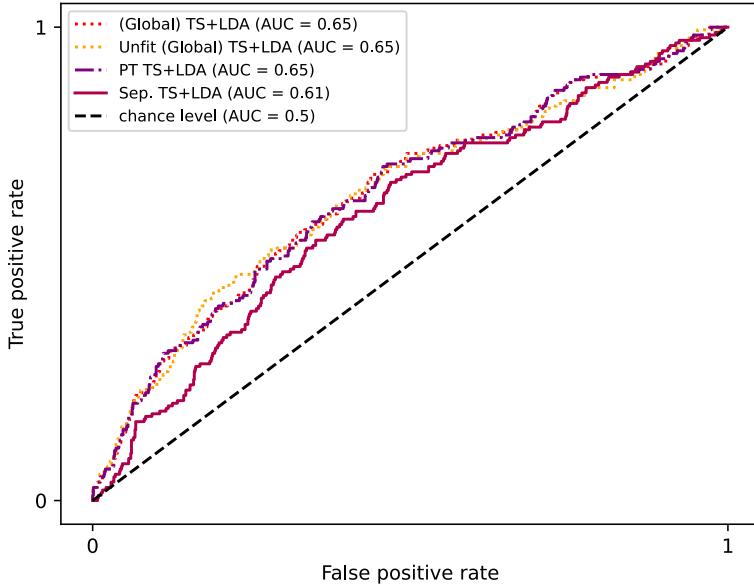


Figure 4.5: Receiver Operator Characteristic curves of Block-Toeplitz approaches per word for participant 3, condition 6D.

ROC of Riemannian parallel transport and global LDA (360 target epochs)



ROC of Riemannian parallel transport and global LDA (360 target epochs)

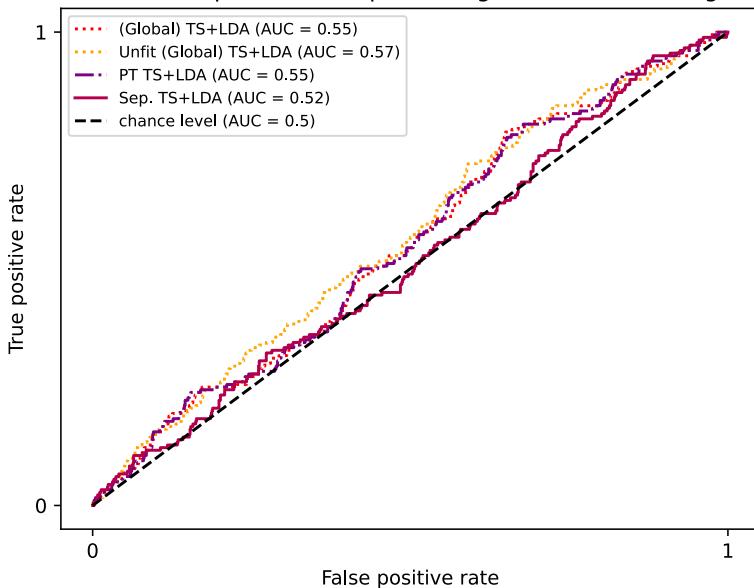
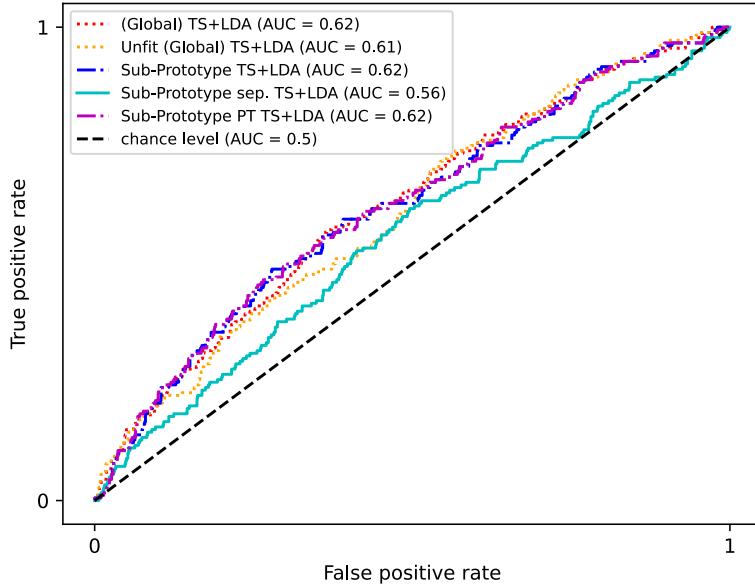


Figure 4.6: ROC curves of parallel transport and global approaches for participant 1, condition 6D (top) and participant 2, condition 1D (bottom).

ROC of Riemannian sub-prototype and global LDA (360 target epochs)



ROC of Riemannian sub-prototype and global LDA (360 target epochs)

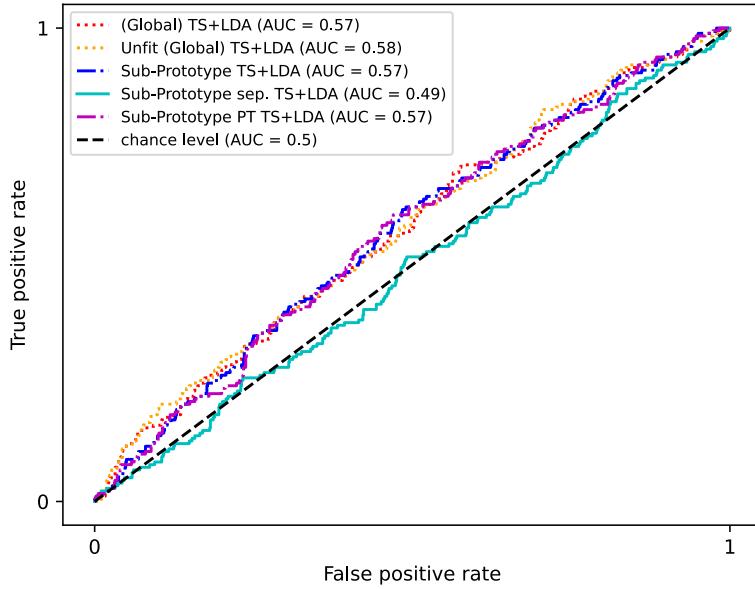


Figure 4.7: ROC curves of sub-prototype and global approaches for participant 2, condition 6D (top) and participant 3, condition 6D (bottom).

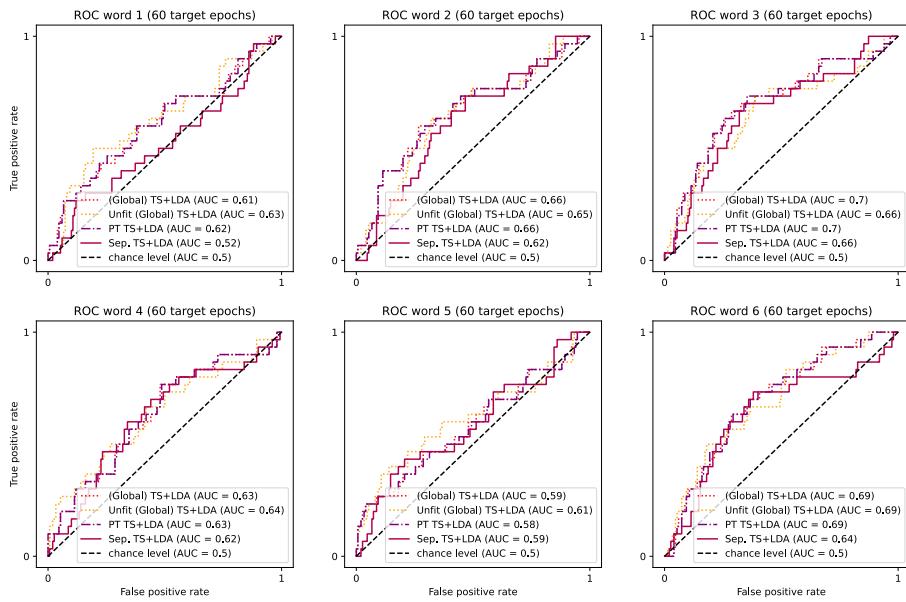


Figure 4.8: ROC curves of parallel transport and global approaches per word for participant 1, condition 6D.

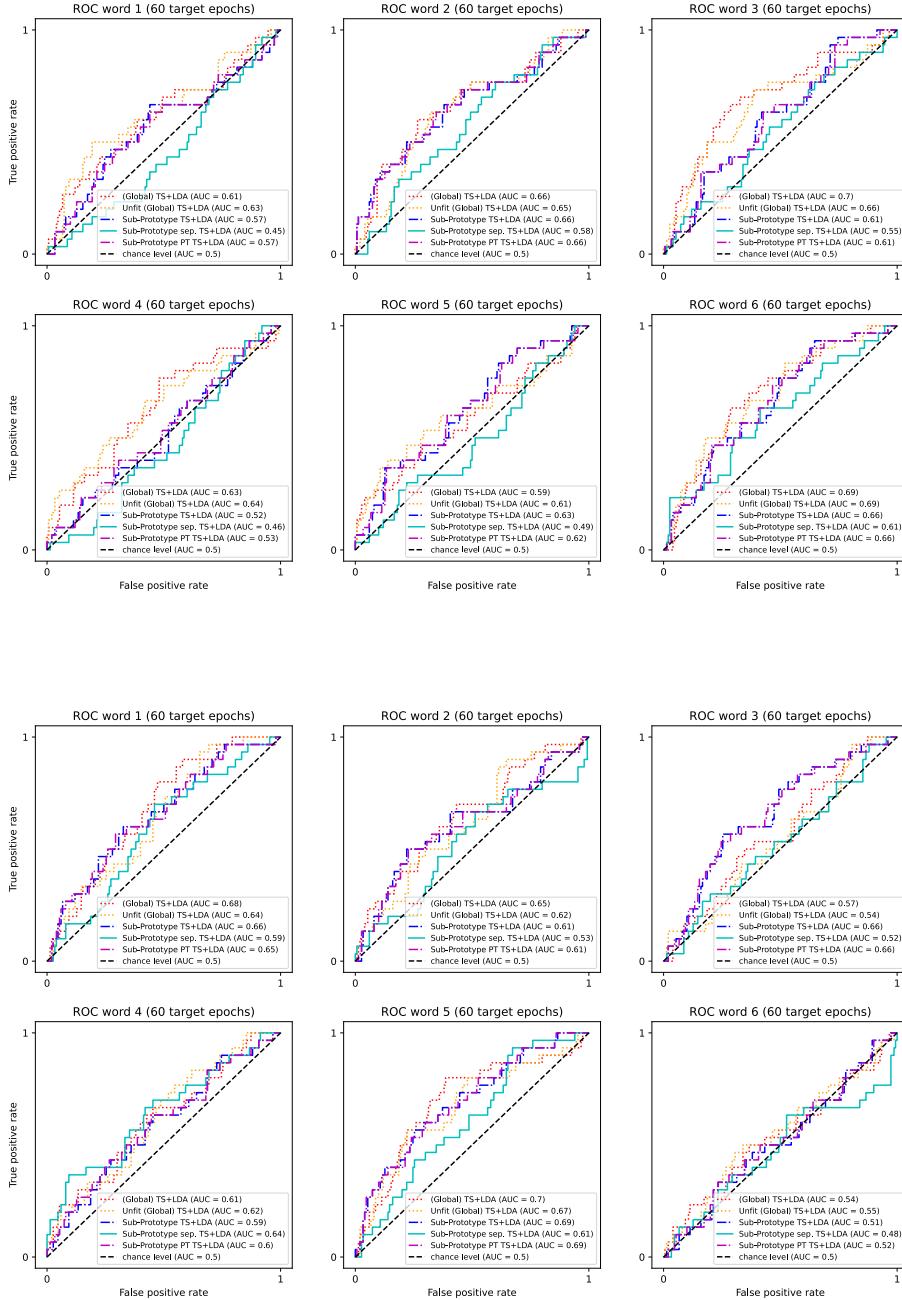


Figure 4.9: ROC curves of sub-prototype and global approaches per word for participant 1, condition 6D (top) and participant 2, condition 6D (bottom).

Chapter 5

Discussion

In this chapter, the conclusions and limitations of this study are discussed. First, the implications of the results of the Euclidean and Riemannian approaches are presented. Then the limitations are explained.

5.1 Euclidean Approaches

For the Euclidean approaches, relevance subclass LDA (RSLDA), separate LDA classifiers, and a hybrid RSLDA approach were compared to a global LDA model. Two types of covariance shrinkage were tried for each of the models. ToeplitzLDA is preferred over SLDA for all approaches. This is consistent with the findings of Sosulski and Tangermann (2022) [14].

Relevance subclass LDA (RSLDA) usually performs better than separate classifiers. The multi-target mean shrinkage and the global estimate of the covariance improve the performance. However, it appears that the best estimate of the mean for this data is still the global mean, since global classification consistently outperformed the subclass-aware approaches.

The results of the hybrid relevance subclass LDA models suggests that the estimate of the non-target subclass means is better than of the target subclass means, since using a mean per subclass only for the non-target words but using the global target mean for the target words gives a better performance than using the subclass mean for both target and non-target classes. The better estimate for non-target means is expected, because there are five times as many epochs available as there are target epochs. It may be interesting to find out how many epochs per subclass are needed for the subclass mean estimates to rival the global mean estimate.

5.2 Riemannian Approaches

The Riemannian approaches compared to a global Riemannian tangent space model used parallel transport, sub-prototypes, and separate classifiers in several combinations (see Table 3.1 in Methods for an overview).

5.2.1 Parallel Transport

Parallel transport without separate classifiers performs equally to global classification. In this model, all points are transported in Riemannian space based on their subclass to let the subclass means fall together in one point. After that, the points are projected

onto tangent space and classified with a global LDA model. The part that differentiates this from the global approach, parallel transport, does not appear to add to classification performance, nor does it reduce performance. While parallel transport is needed to allow relevance subclass LDA to be used on the projected data because all subclass means should have the same tangent point for this, it seems to have no influence on classification performance in and of itself.

Applying relevance subclass LDA after parallel transport did not work, since the quadratic program of multi-target shrinkage of the mean could not be solved. The solver throws no errors, so it is unclear what the issue is. Needing a large number of iterations has been ruled out as a cause, as the solver terminates in approximately equal time no matter the maximum number of iterations. Something must go wrong early. The multi-target shrinkage problem is supposed to be convex, so the optimal solution should exist [15]. Since RSLDA performs better than separate classifiers for the Euclidean approaches and in combination with parallel transport performs better than a global Riemannian approach in a visual paradigm [3], it may be worth investigating how to make it work for this data set with Riemannian features.

5.2.2 Sub-Prototypes

Sub-prototypes work well for certain words, so perhaps a mixed model that applies sub-prototypes only for those words that benefit would give an improvement. In practice, it would be a challenge to find out which words would benefit, especially since this is likely to differ between participants.

Combining sub-prototypes with parallel transport performs approximately equal to sub-prototypes alone, so parallel transport seems to not add to performance, just like in the global LDA case. It is probably not useful to apply parallel transport when using subclass prototypes. Additionally, when combining sub-prototypes and parallel transport, it is not recommended to use separate classifiers. Separate classifiers perform quite poorly in this combination.

5.3 Limitations

There are some limitations to the models and the conclusions.

A limitation to the validity of the conclusions is the amount of participants that the approaches are tested on. Only the data of three participants was used. This limited the possibility for significance testing. Also, considerable differences could be seen between them. All in all, no conclusions can be directly generalized to all participants.

Additionally, the subclass-aware approaches have limitations. The approaches each assume a fully known set of subclasses of which all have examples of both classes in the training data [3]. This restriction is met for the aphasia training paradigm [1], but may not be in all paradigms. If new words are introduced after the training phase, the models need to train on those to be able to classify them. Some workarounds may be possible, like grouping new words with old ones. However, this would require knowledge of how words can be effectively grouped. This would require research into what properties of a word are important for the differences in response.

Furthermore, the main downside of sub-prototypes is the large amount of features. Apart from reducing the training time, a smaller amount of features could potentially improve performance by limiting overfitting. One simple way is reducing the amount of spatial filters to one per class. Another idea is having only one sub-prototype set, as well

as the global prototypes, for each epoch. Each epoch gets the global prototypes and the subclass prototypes of its own subclass appended. This reduces the amount of prototypes from twelve to four in our case with six subclasses.

5.4 Conclusion

Several subclass-aware methods were tested on an aphasia training paradigm. The methods include separate classifiers, relevance subclass LDA, parallel transport, and a novel approach called sub-prototypes, as well as combinations of some of these. None of the subclass-aware models consistently performed better than their global counterpart. This is possibly due to the amount of epochs available to make good estimates of the subclass means. Relevance subclass LDA could not sufficiently counteract the smaller amount of data that a subclass has compared to the full data set.

However, our novel approach, sub-prototypes, may be promising since they performed well for some words. Some improvements could be made to the model to enhance the performance. Some ideas to reduce the dimensionality of the features were proposed for future research.

Also potentially promising is a combination of parallel transport and relevance subclass LDA (RSLDA). Parallel transport could unfortunately not be combined with RSLDA due to technical problems, but if the gain in performance of RSLDA compared to separate classifiers in the Euclidean approaches is the same with parallel transport, parallel transport with RSLDA may perform better than global Riemannian classification.

In conclusion, no clear improvements on global classification were found that could improve the classification performance for the aphasia training of Musso and colleagues [1], but sub-prototypes warrant further investigation because of its small successes. Perhaps future improvements of sub-prototypes will lead to enhanced classification performance of the auditory word-evoked ERP data, and in turn to better feedback for aphasia patients.

Chapter 6

Acknowledgments

I would like to thank my supervisor, Michael Tangermann, for guidance, help and suggestions. I also thank the other members of my thesis group, Laurenz Schindler, Jesse Manders, Mats Robben, and Ognyan Bachvarov. Our code review sessions helped to confirm I went about my coding correctly, the presentation practice gave me valuable feedback on my explanations, and our weekly meetings formed a source of inspiration.

Chapter 7

Source Code

The source code can be accessed at <https://github.com/ellahas/Subclass-Aware-Decoding>

References

- [1] M. Musso, D. Hübner, S. Schwarzkopf, *et al.*, “Aphasia recovery by language training using a brain–computer interface: A proof-of-concept study,” *Brain communications*, vol. 4, no. 1, fcac008, 2022.
- [2] J. Höhne, D. Bartz, M. N. Hebart, K.-R. Müller, and B. Blankertz, “Analyzing neuroimaging data with subclasses: A shrinkage approach,” *NeuroImage*, vol. 124, pp. 740–751, 2016.
- [3] H. Kolkhorst, J. Veit, W. Burgard, and M. Tangermann, “A robust screen-free brain-computer interface for robotic object selection,” *Frontiers in Robotics and AI*, vol. 7, p. 38, 2020.
- [4] L. Acqualagna and B. Blankertz, “Gaze-independent bci-spelling using rapid serial visual presentation (rsvp),” *Clinical Neurophysiology*, vol. 124, no. 5, pp. 901–908, 2013.
- [5] N. Mrachacz-Kersting, N. Jiang, A. J. T. Stevenson, *et al.*, “Efficient neuroplasticity induction in chronic stroke patients by an associative brain-computer interface,” *Journal of neurophysiology*, vol. 115, no. 3, pp. 1410–1421, 2016.
- [6] R. Mane, T. Chouhan, and C. Guan, “Bci for stroke rehabilitation: Motor and beyond,” *Journal of neural engineering*, vol. 17, no. 4, p. 041001, 2020.
- [7] S. C. Kleih, L. Gottschalt, E. Teichlein, and F. X. Weilbach, “Toward a p300 based brain-computer interface for aphasia rehabilitation after stroke: Presentation of theoretical considerations and a pilot feasibility study,” *Frontiers in human neuroscience*, vol. 10, p. 547, 2016.
- [8] M. Schreuder, B. Blankertz, and M. Tangermann, “A new auditory multi-class brain-computer interface paradigm: Spatial hearing as an informative cue,” *PLoS one*, vol. 5, no. 4, e9813, 2010.
- [9] P. Petersen, *Riemannian geometry*. Springer, 2006, vol. 171.
- [10] A. Barachant and M. Congedo, “A plug&play p300 bci using information geometry,” *arXiv preprint arXiv:1409.0107*, 2014.
- [11] A. Barachant, “Pyriemann 0.5,” 2023.
- [12] O. Ledoit and M. Wolf, “A well-conditioned estimator for large-dimensional covariance matrices,” *Journal of multivariate analysis*, vol. 88, no. 2, pp. 365–411, 2004.
- [13] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [14] J. Sosulski and M. Tangermann, “Introducing block-toeplitz covariance matrices to remaster linear discriminant analysis for event-related potential brain–computer interfaces,” *Journal of Neural Engineering*, vol. 19, no. 6, p. 066001, 2022.

- [15] D. Bartz, J. Höhne, and K.-R. Müller, “Multi-target shrinkage,” *arXiv preprint arXiv:1412.2041*, 2014.
- [16] S. Caron, D. Arnström, S. Bonagiri, *et al.*, *qpsolvers: Quadratic Programming Solvers in Python*, version 3.3.1, Apr. 2023.
- [17] C. R. Harris, K. J. Millman, S. J. van der Walt, *et al.*, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2).
- [18] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, “Xdawn algorithm to enhance evoked potentials: Application to brain–computer interface,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2035–2043, 2009.
- [19] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [20] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *The Journal of Machine learning research*, vol. 7, pp. 1–30, 2006.
- [21] E. W. Weisstein, “Bonferroni correction,” *MathWorld*, 2004. [Online]. Available: <https://mathworld.wolfram.com/BonferroniCorrection.html>.

Appendix A

Statistical Tests

The p-values of the statistical tests are reported here.

Subclass-Aware Model	Global Counterpart	Participant	Condition	p-value
ToeplitzRSLDA	Global ToeplitzLDA	1	1D	0.34375
sRSLDA	Global sLDA	1	1D	0.34375
Sep. ToeplitzLDA	Global ToeplitzLDA	1	1D	0.421875
PT TS+LDA	(Global) TS+LDA	2	6D	0.078125
PT TS+LDA	(Global) TS+LDA	3	6D	0.921875
Sub-Prototype TS+LDA	(Global) TS+LDA	1	1D	0.078125
Sub-Prototype PT TS+LDA	(Global) TS+LDA	1	1D	0.078125
PT TS+LDA	(Global) TS+LDA	3	1D	0.34375
PT TS+LDA	(Global) TS+LDA	1	HP	0.421875
PT TS+LDA	(Global) TS+LDA	2	HP	0.28125
Sub-Prototype TS+LDA	(Global) TS+LDA	3	HP	0.5
PT TS+LDA	(Global) TS+LDA	3	HP	0.28125
Sub-Prototype PT TS+LDA	(Global) TS+LDA	3	HP	0.5

Table A.1: The p-values of the statistical tests done to compare models with their global counterpart. The Wilcoxon signed-rank test is only performed for models with a higher AUC score than their global counterpart.

Appendix B

Additional Figures

B.1 Average EEG Responses Channel Cz

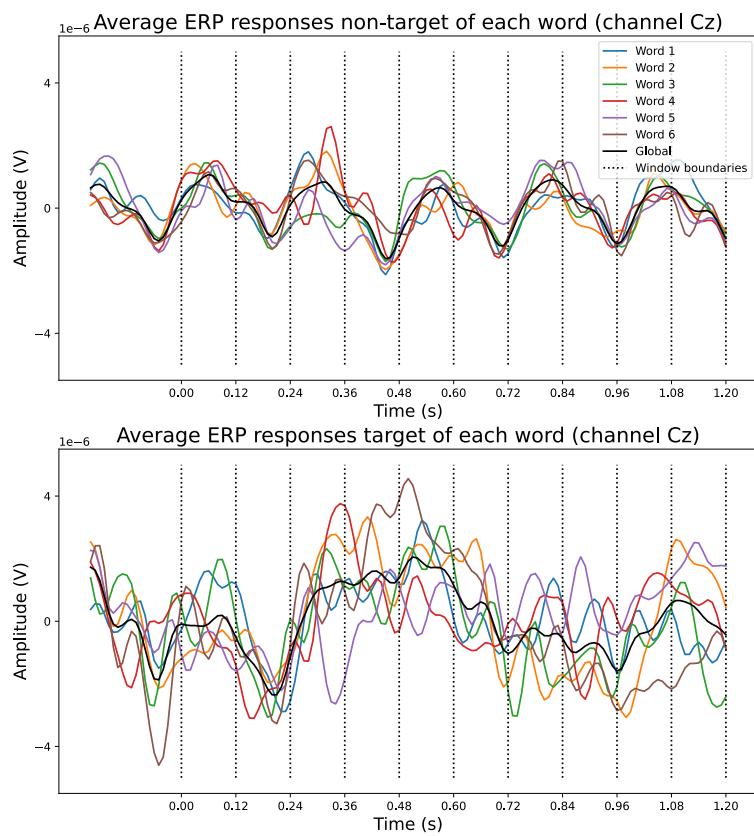


Figure B.1: Average event-related potential to target and non-target words in channel Cz for Participant 1, condition 6D

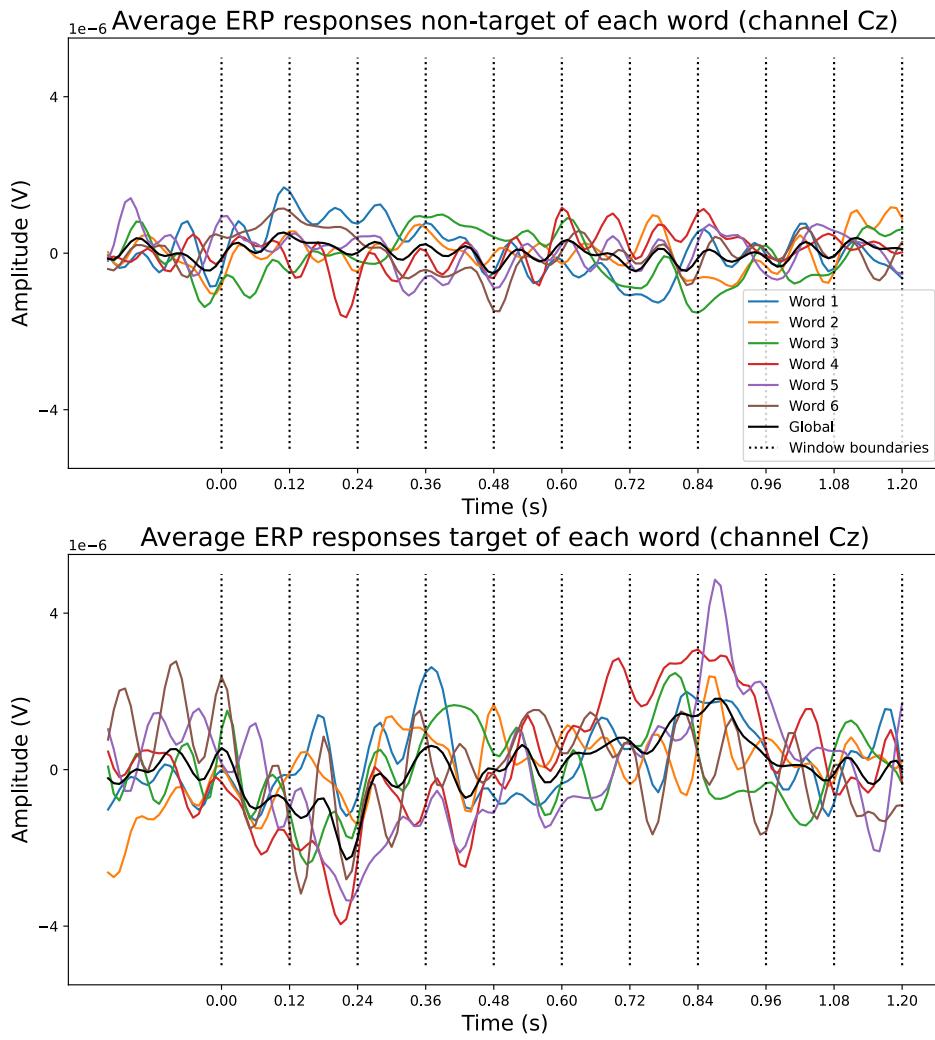


Figure B.2: Average event-related potential to target and non-target words in channel Cz for Participant 2, condition 6D

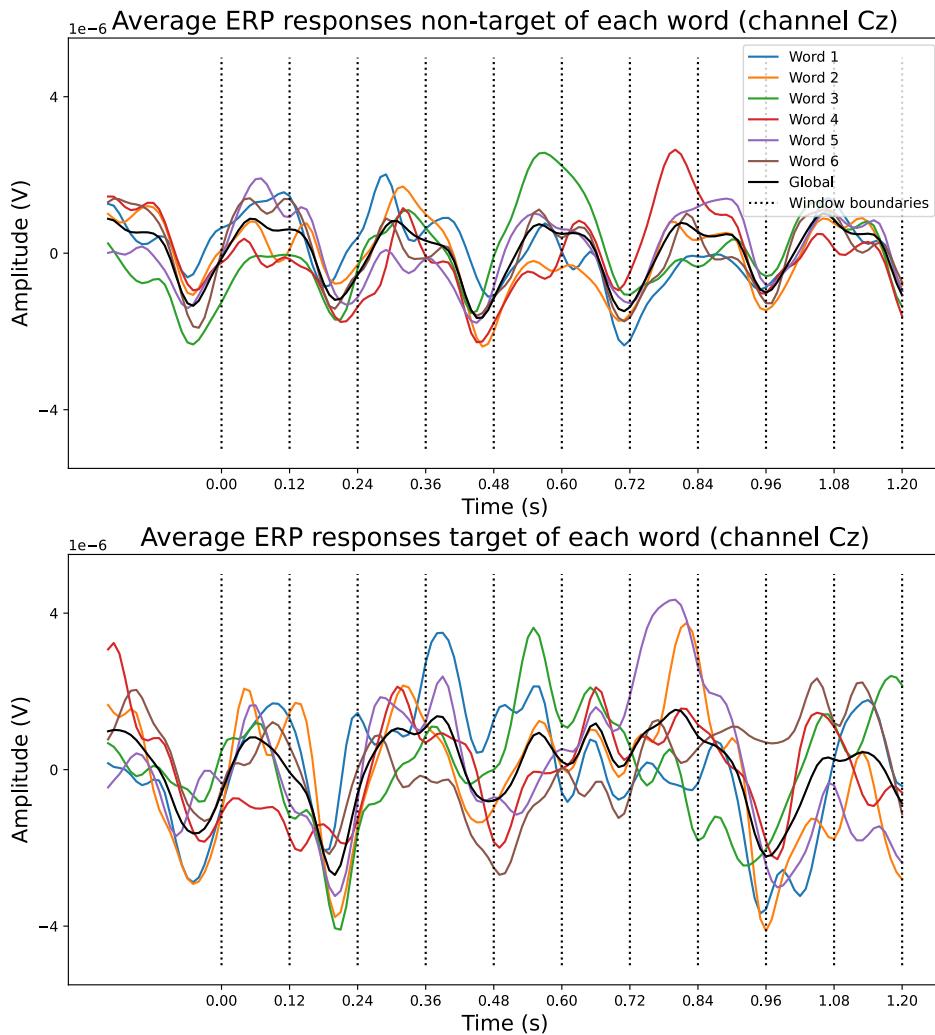


Figure B.3: Average event-related potential to target and non-target words in channel Cz for Participant 3, condition 6D

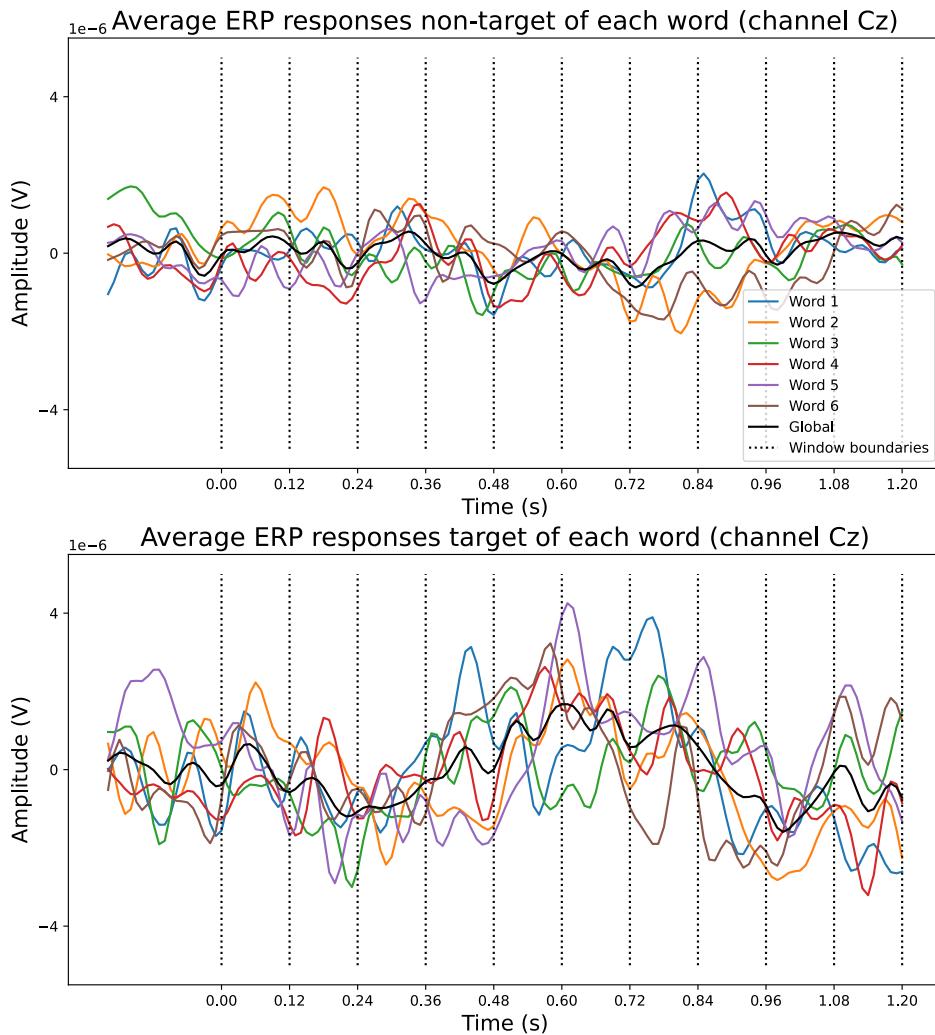


Figure B.4: Average event-related potential to target and non-target words in channel Cz for Participant 1, condition 1D

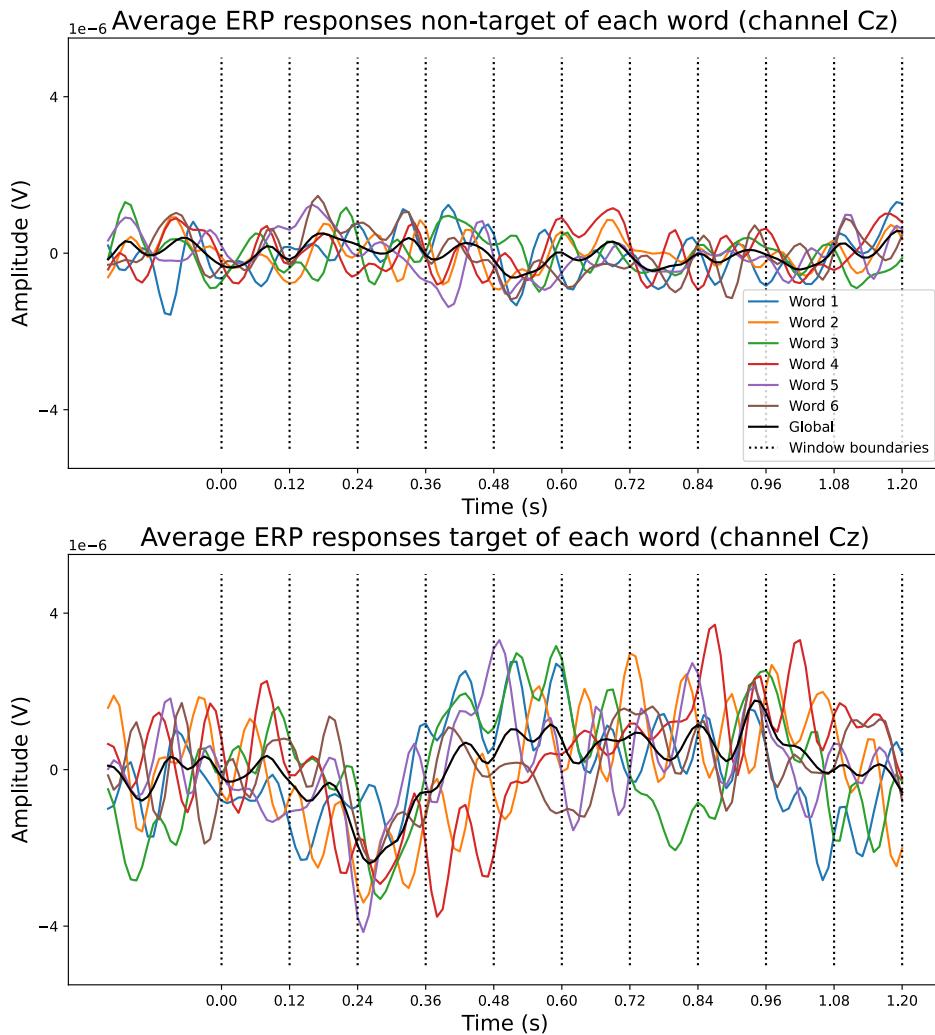


Figure B.5: Average event-related potential to target and non-target words in channel Cz for Participant 2, condition 1D

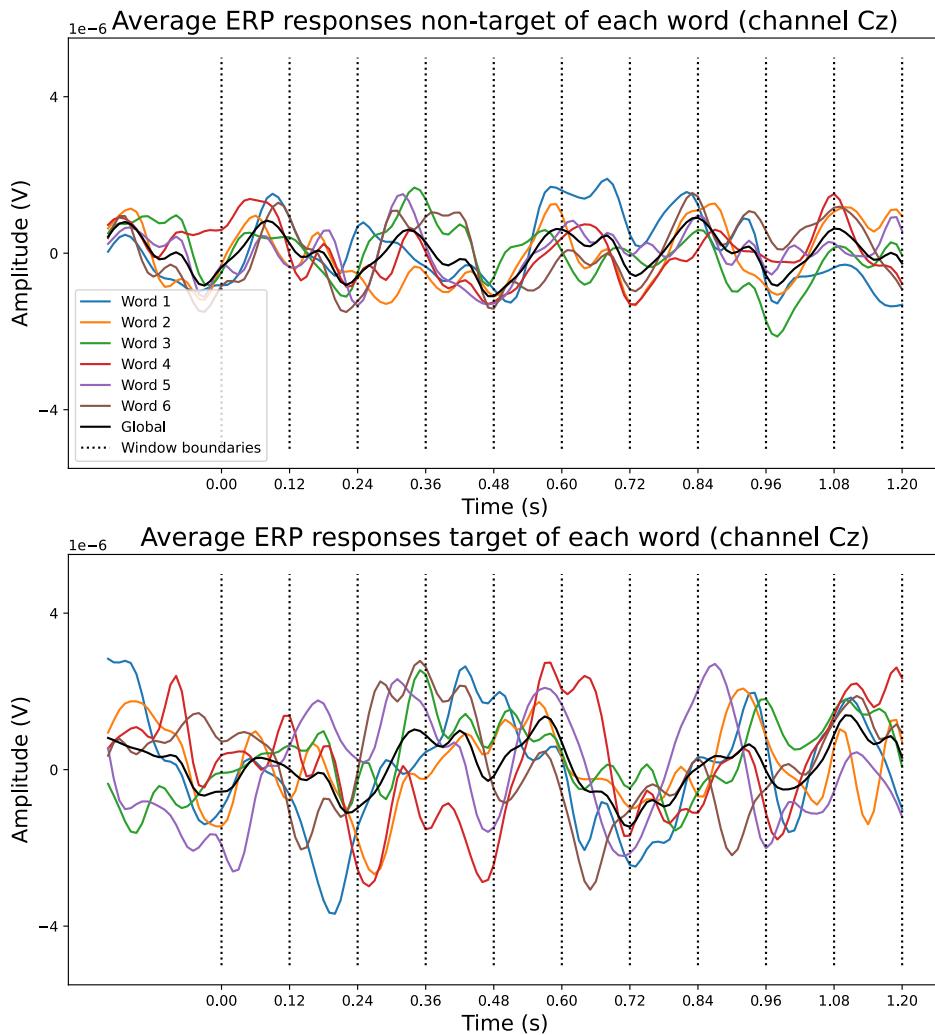


Figure B.6: Average event-related potential to target and non-target words in channel Cz for Participant 3, condition 1D

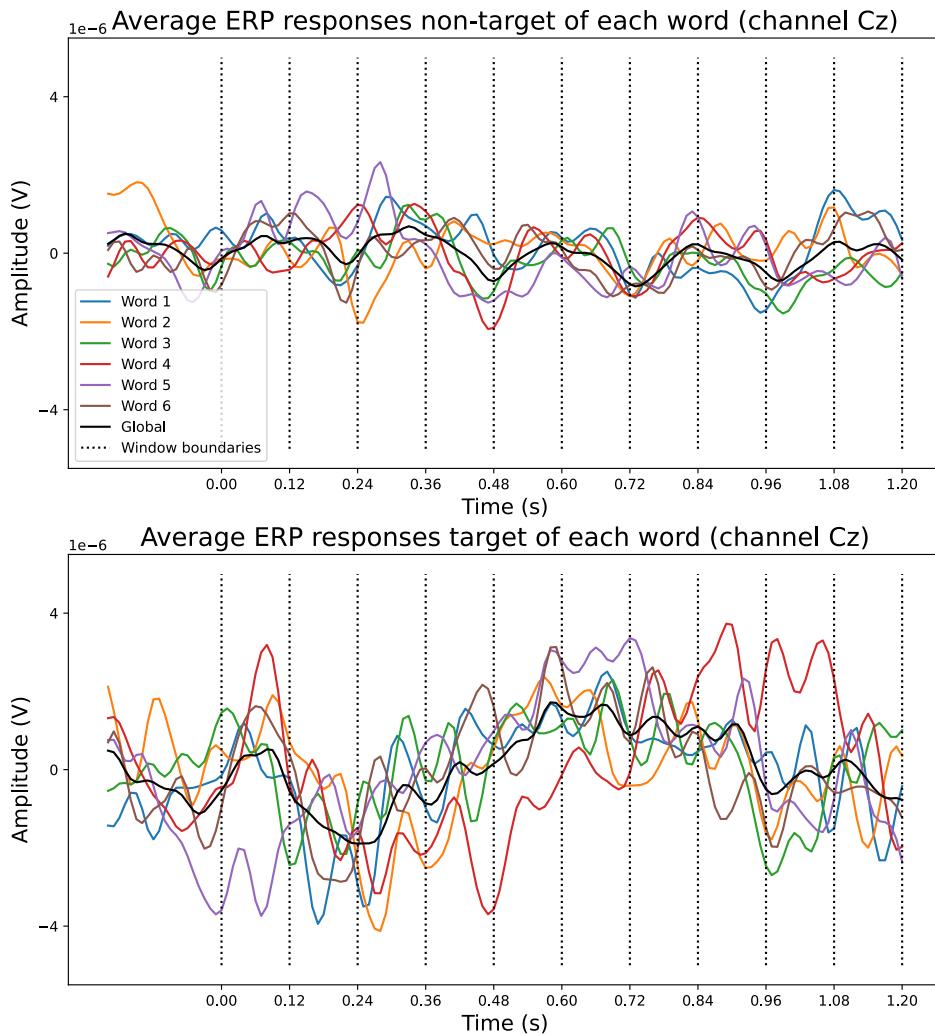


Figure B.7: Average event-related potential to target and non-target words in channel Cz for Participant 1, condition HP

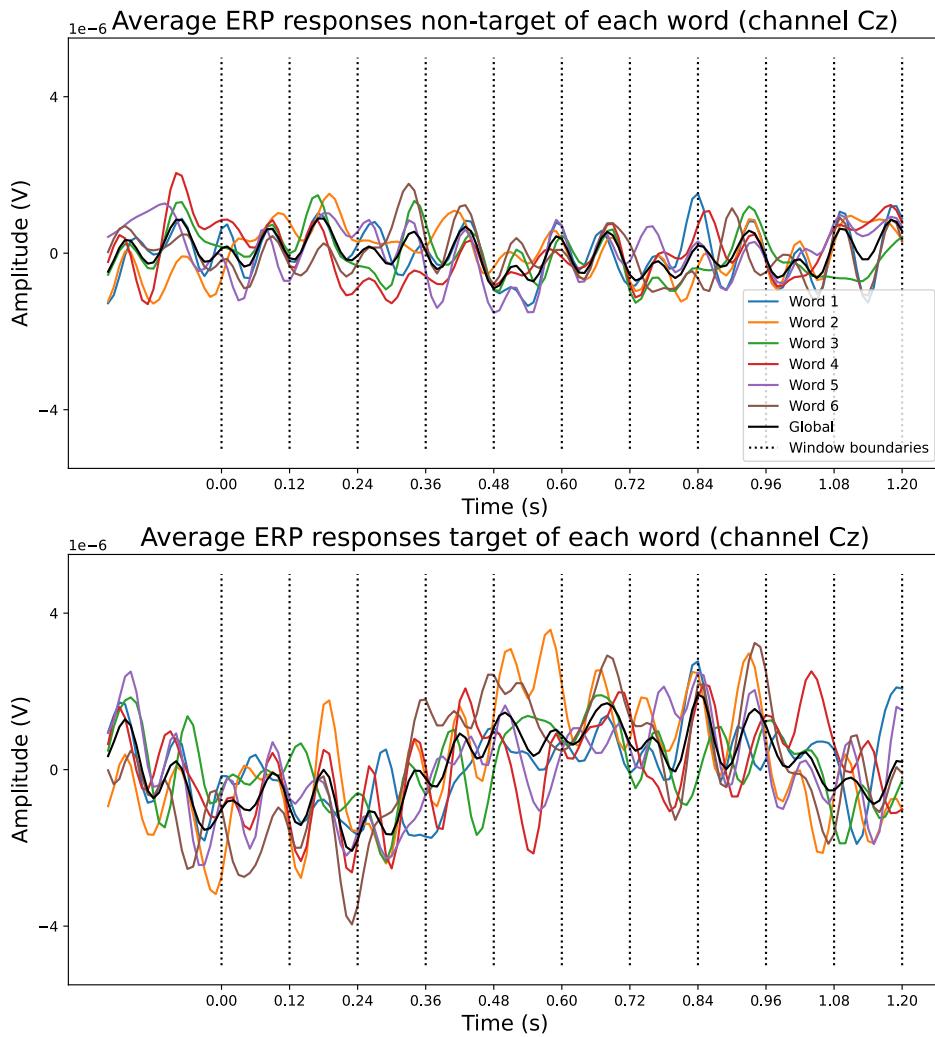


Figure B.8: Average event-related potential to target and non-target words in channel Cz for Participant 2, condition HP

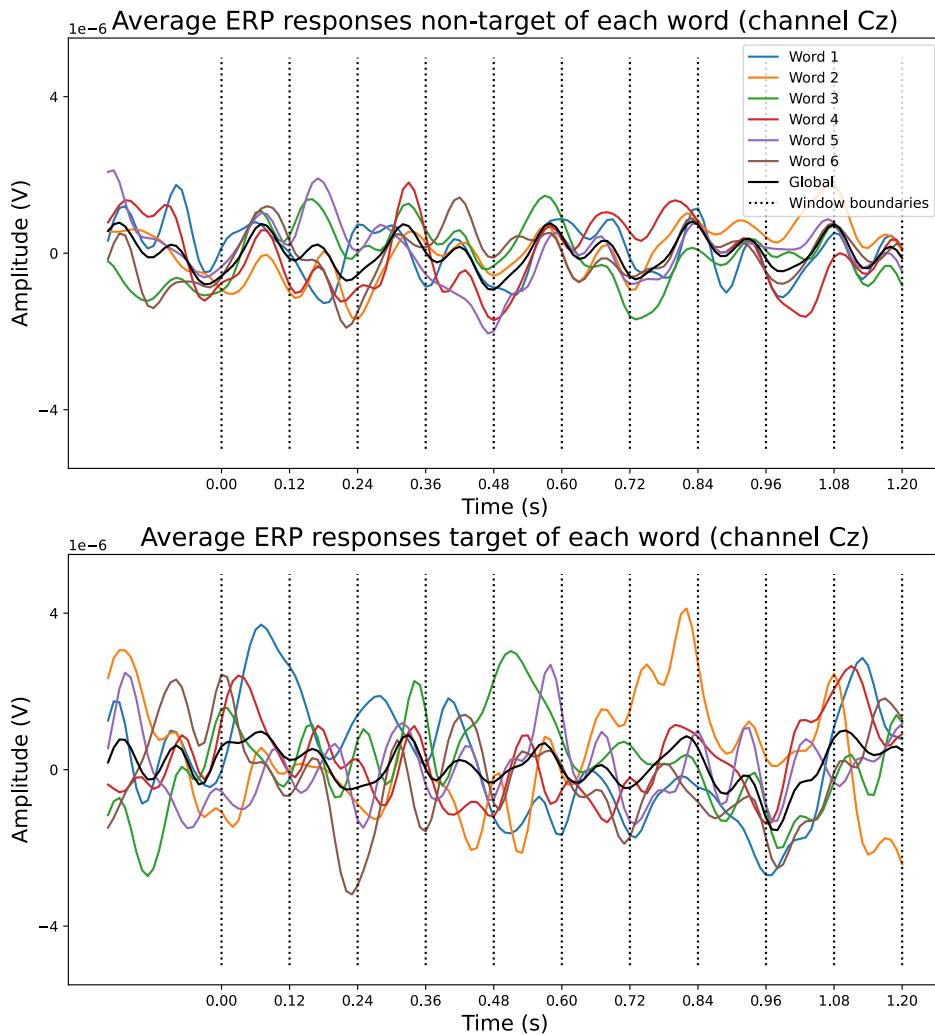


Figure B.9: Average event-related potential to target and non-target words in channel Cz for Participant 3, condition HP

B.2 Euclidean ROC Curves

Below are all the ROC curves of the Euclidean approaches.

B.2.1 Block-Toeplitz

All Words

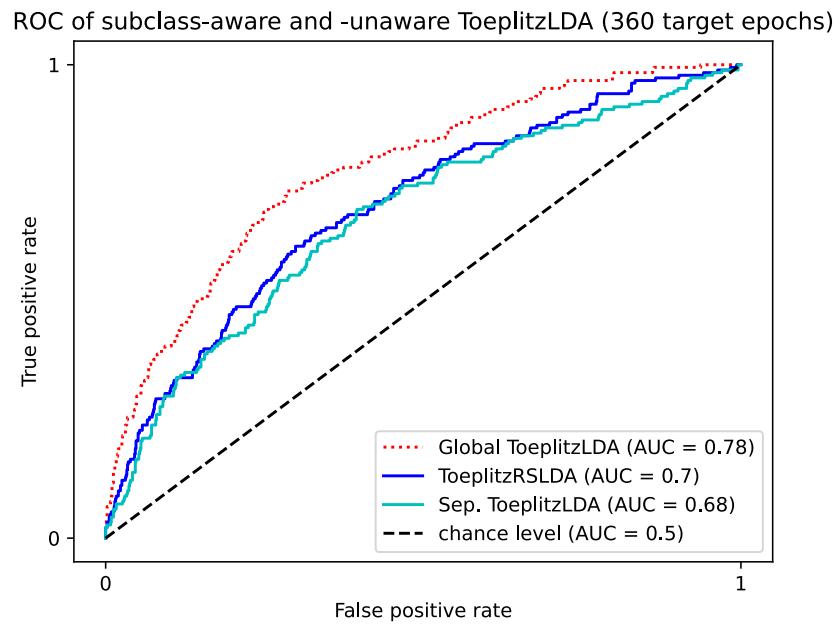


Figure B.1: ROC curves of Block-Toeplitz approaches for Participant 1, condition 6D

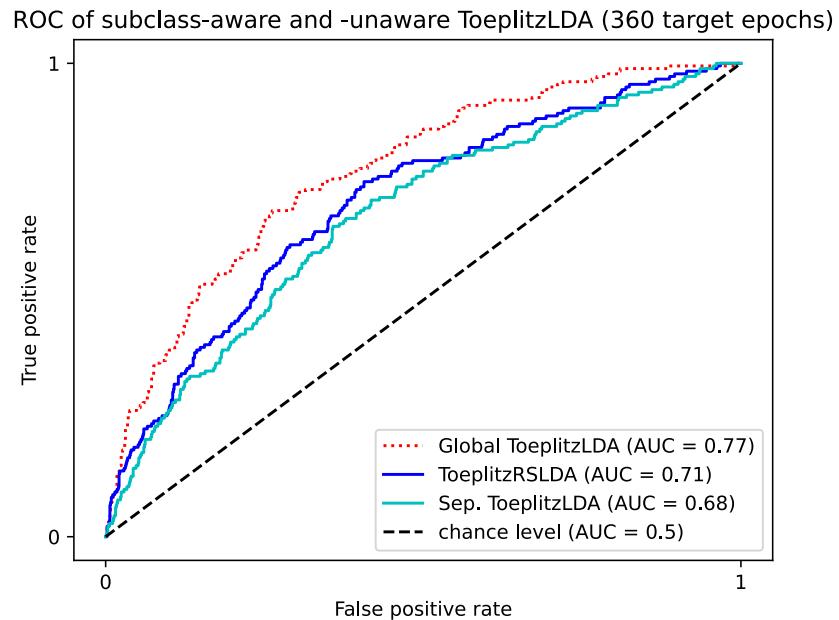


Figure B.2: ROC curves of Block-Toeplitz approaches for Participant 2, condition 6D

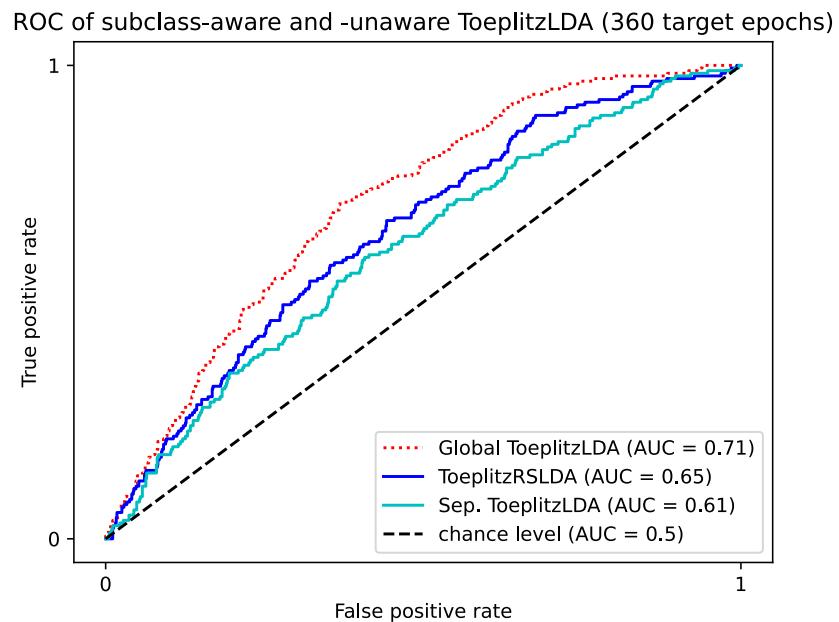


Figure B.3: ROC curves of Block-Toeplitz approaches for Participant 3, condition 6D

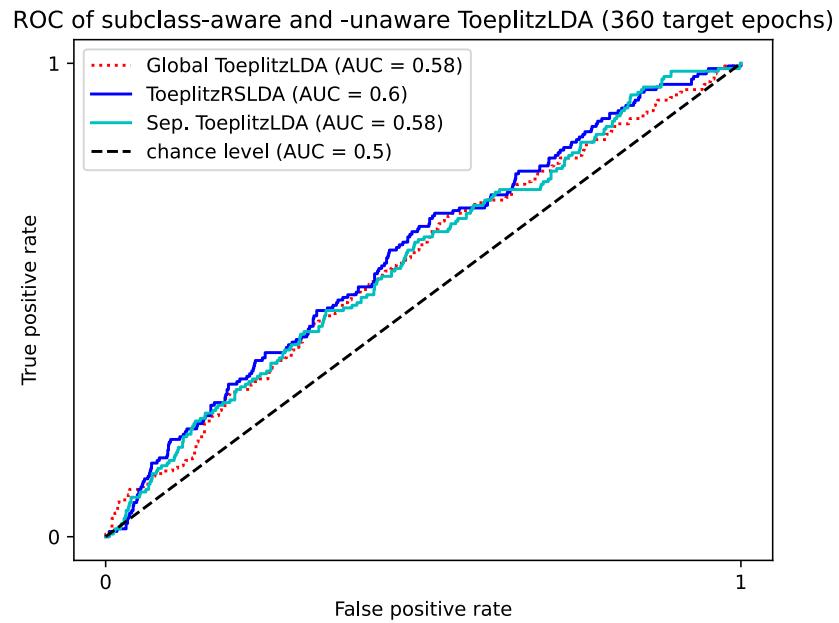


Figure B.4: ROC curves of Block-Toeplitz approaches for Participant 1, condition 1D

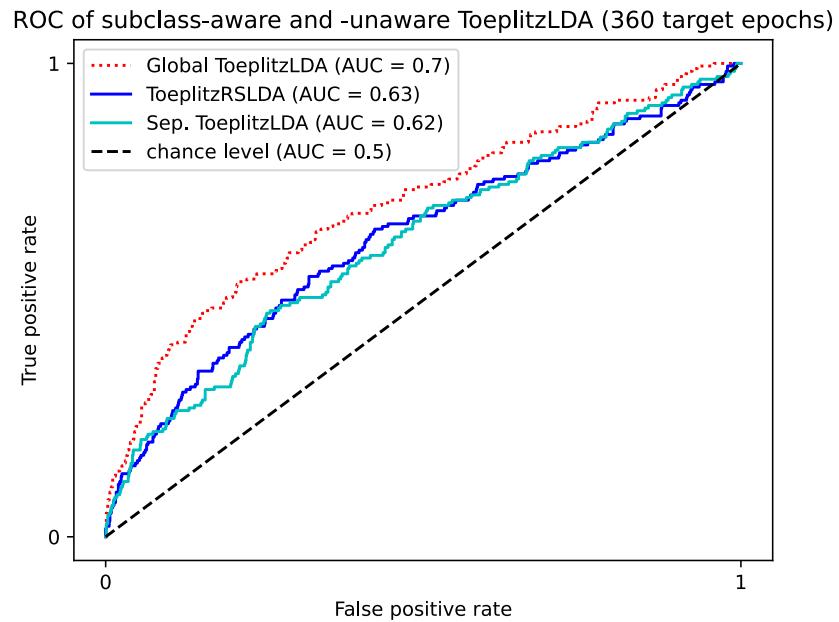


Figure B.5: ROC curves of Block-Toeplitz approaches for Participant 2, condition 1D

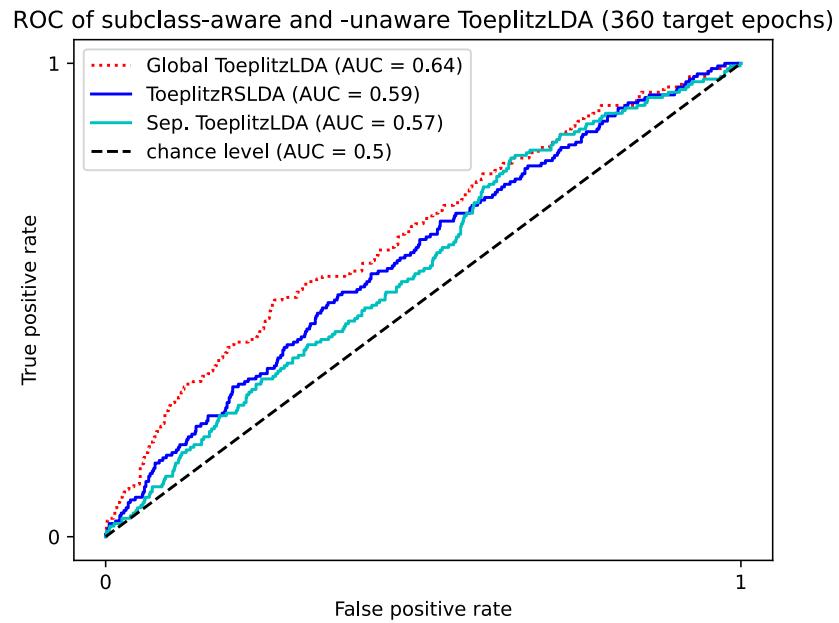


Figure B.6: ROC curves of Block-Toeplitz approaches for Participant 3, condition 1D

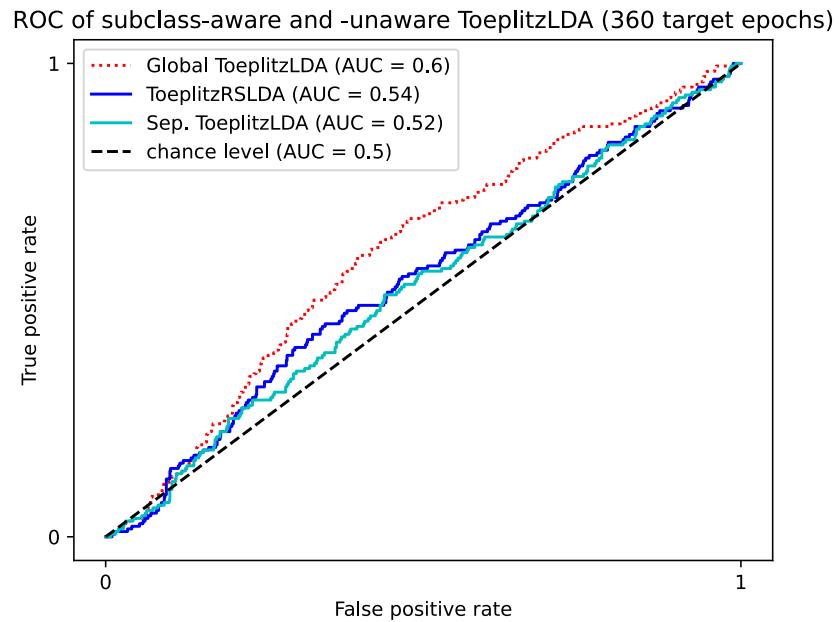


Figure B.7: ROC curves of Block-Toeplitz approaches for Participant 1, condition HP

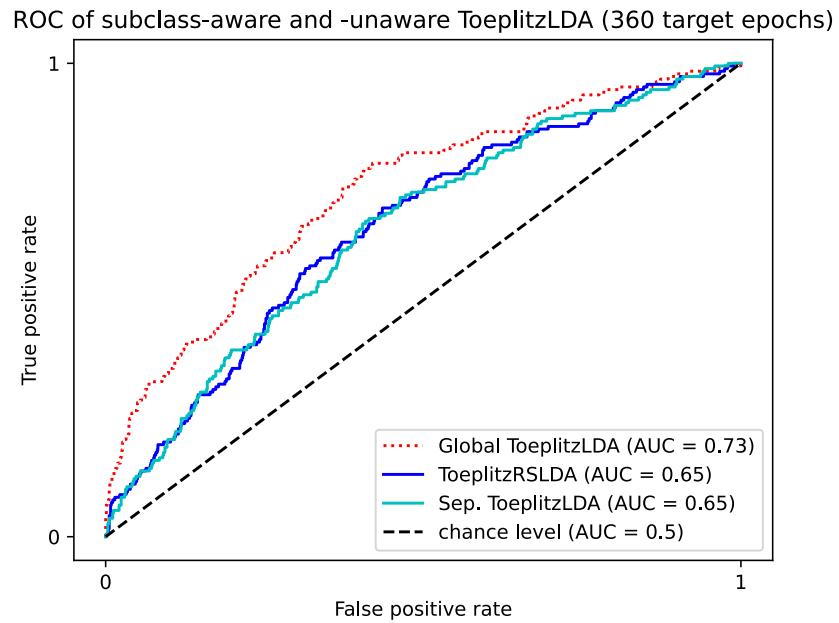


Figure B.8: ROC curves of Block-Toeplitz approaches for Participant 2, condition HP

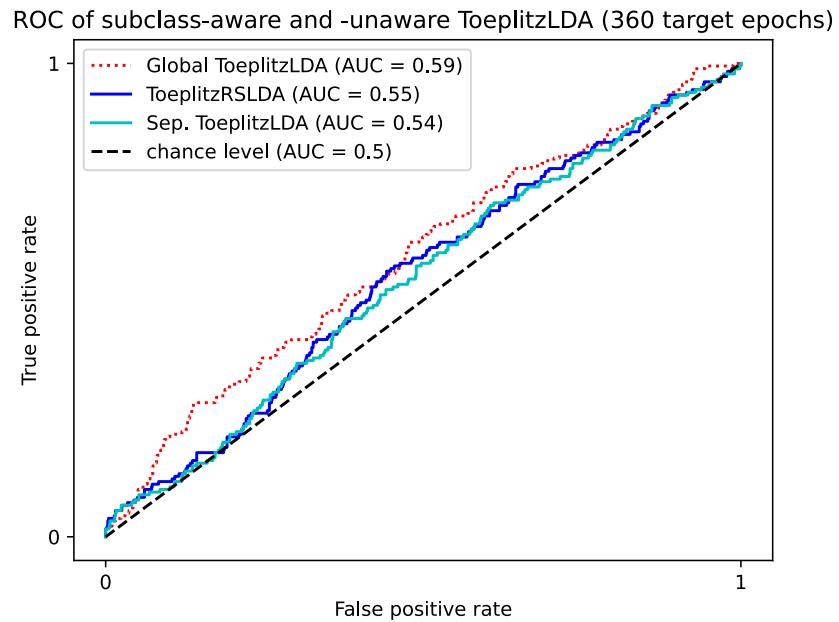


Figure B.9: ROC curves of Block-Toeplitz approaches for Participant 3, condition HP

Per Word

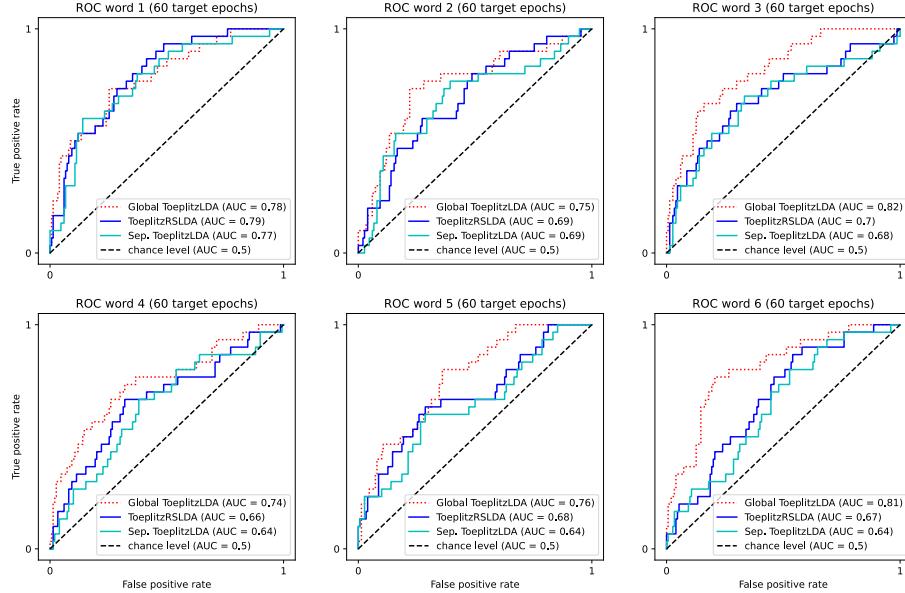


Figure B.10: ROC curves of Block-Toeplitz approaches per word for Participant 1, condition 6D

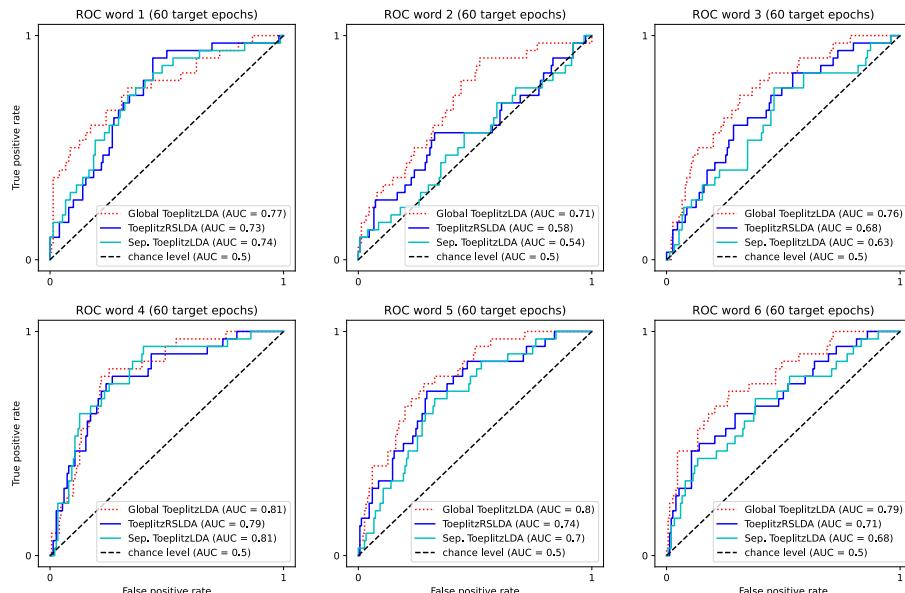


Figure B.11: ROC curves of Block-Toeplitz approaches per word for Participant 2, condition 6D

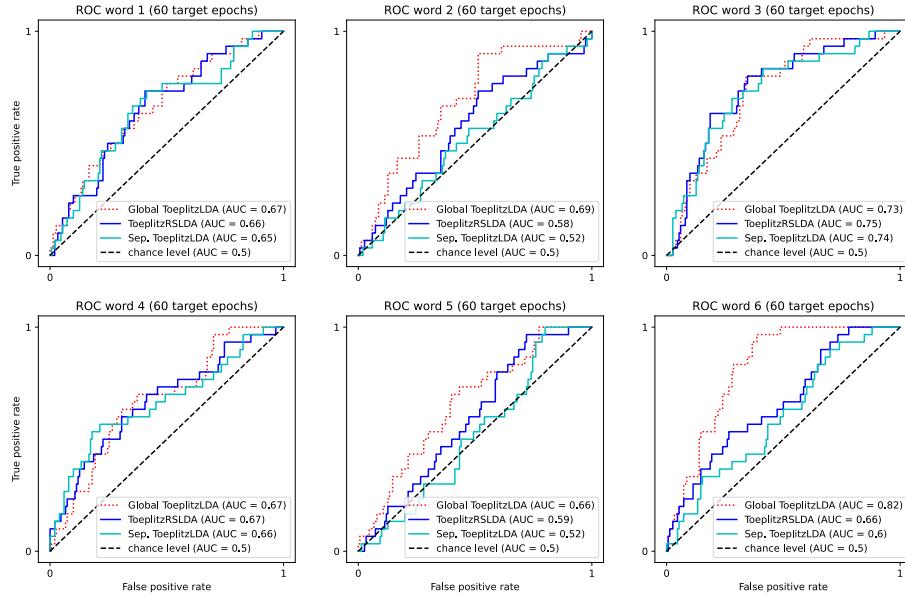


Figure B.12: ROC curves of Block-Toeplitz approaches per word for Participant 3, condition 6D

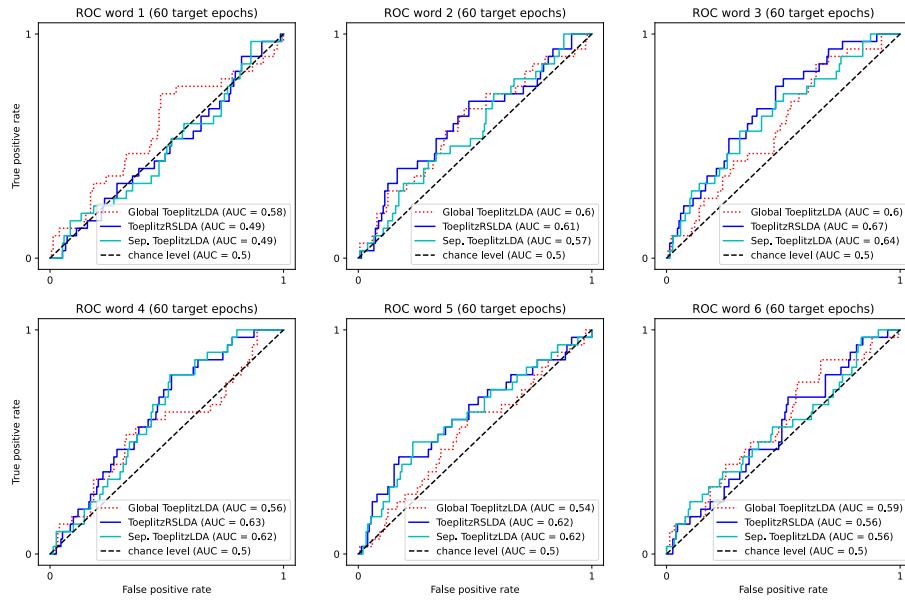


Figure B.13: ROC curves of Block-Toeplitz approaches per word for Participant 1, condition 1D

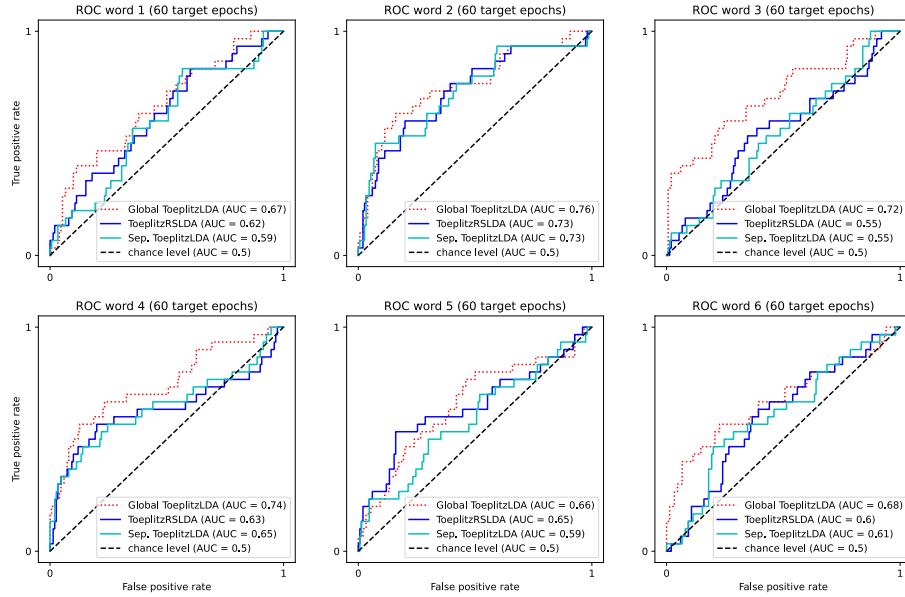


Figure B.14: ROC curves of Block-Toeplitz approaches per word for Participant 2, condition 1D

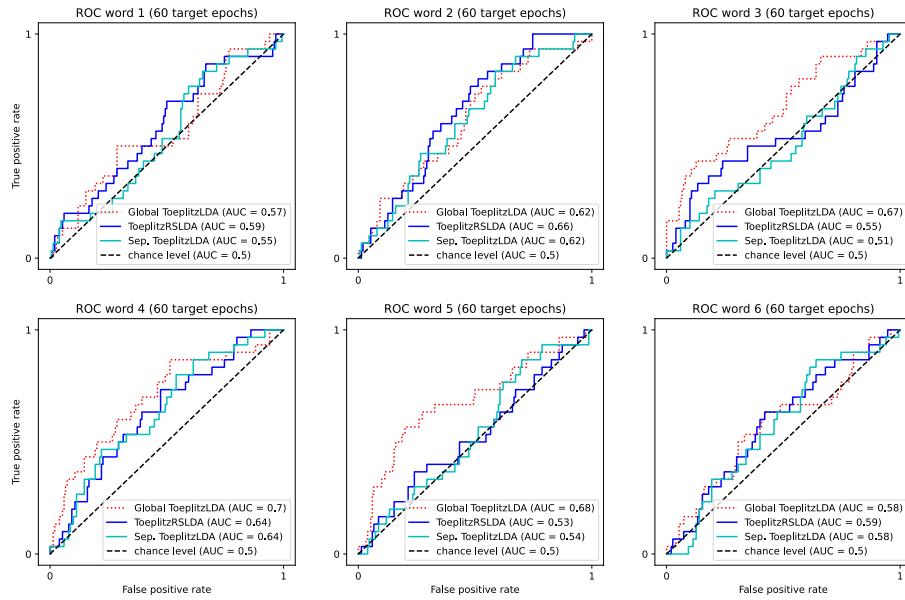


Figure B.15: ROC curves of Block-Toeplitz approaches per word for Participant 3, condition 1D

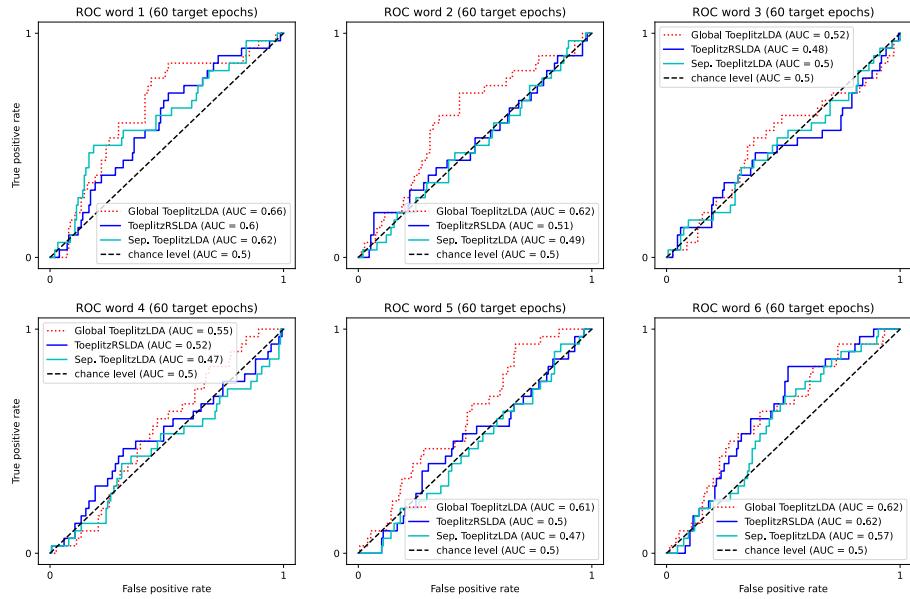


Figure B.16: ROC curves of Block-Toeplitz approaches per word for Participant 1, condition HP

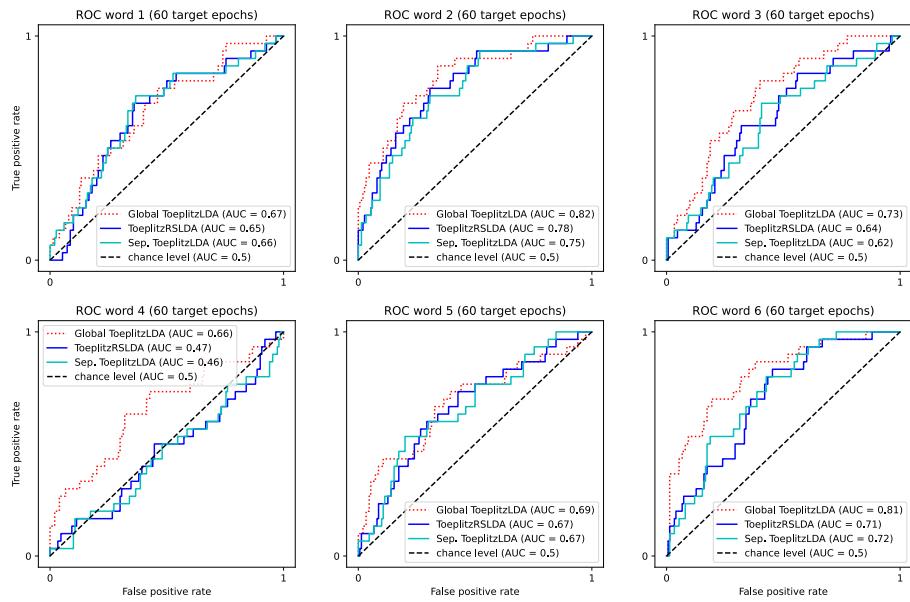


Figure B.17: ROC curves of Block-Toeplitz approaches per word for Participant 2, condition HP

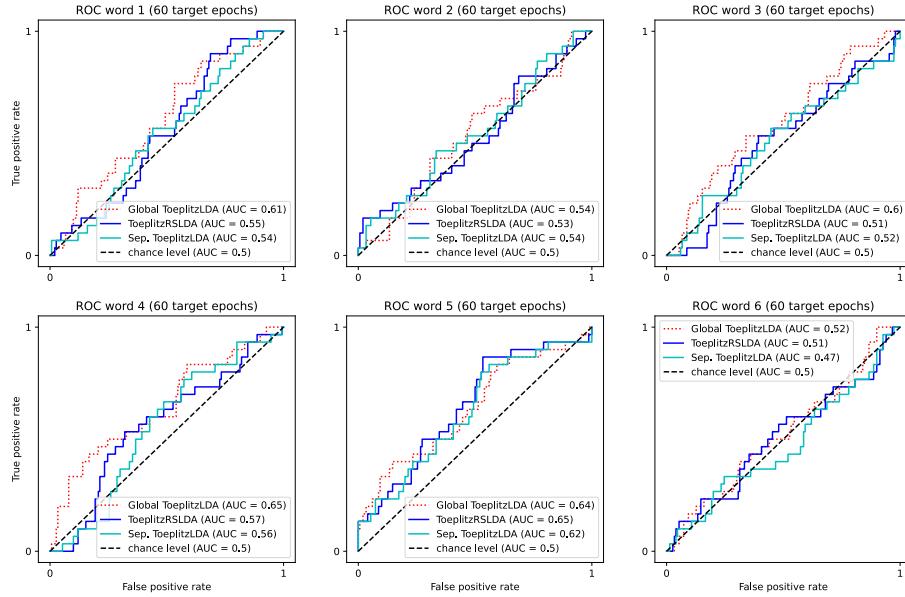


Figure B.18: ROC curves of Block-Toeplitz approaches per word for Participant 3, condition HP

Hybrid Subclass-Aware

Below are the ROC curves for the hybrid subclass-aware approaches, where either the target or the non-target mean is global and the other is estimated per subclass using multi-target shrinkage.

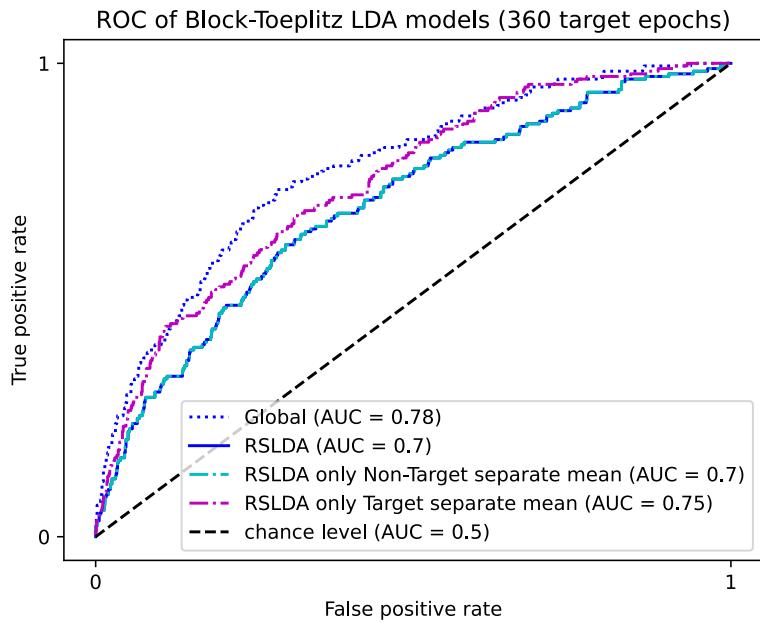


Figure B.19: ROC curves of Block-Toeplitz hybrid approaches per word for Participant 1, condition 6D

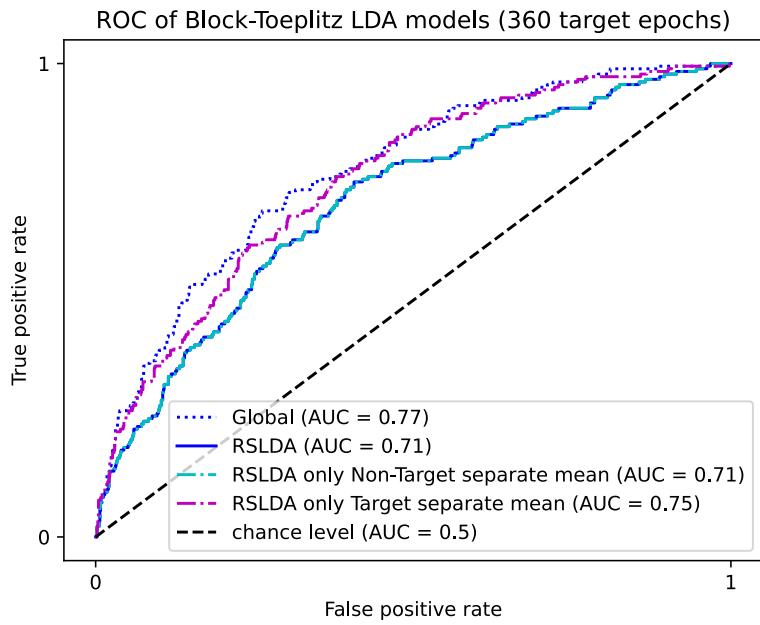


Figure B.20: ROC curves of Block-Toeplitz hybrid approaches per word for Participant 2, condition 6D

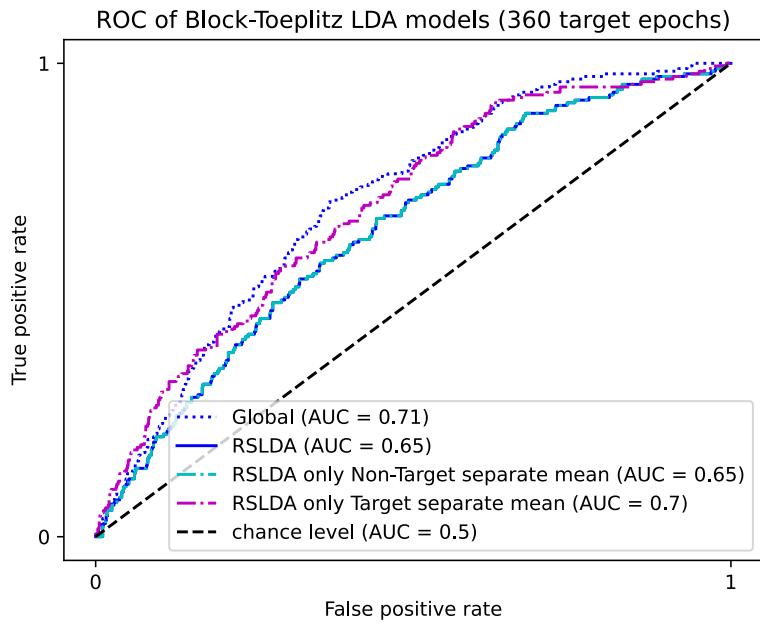


Figure B.21: ROC curves of Block-Toeplitz hybrid approaches per word for Participant 3, condition 6D

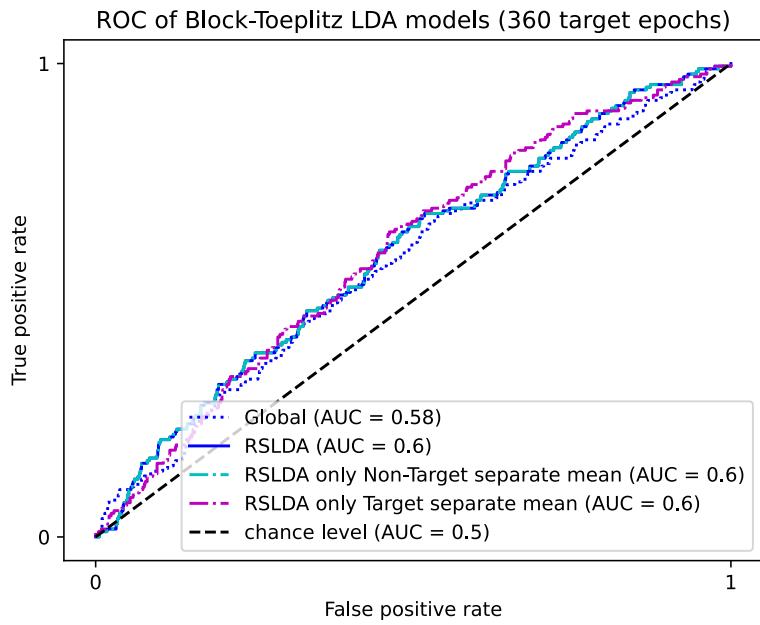


Figure B.22: ROC curves of Block-Toeplitz hybrid approaches per word for Participant 1, condition 1D

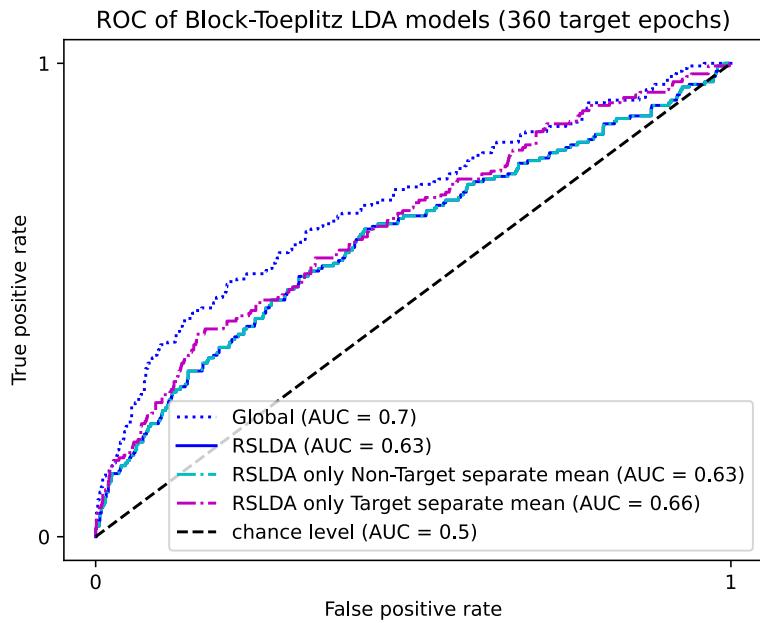


Figure B.23: ROC curves of Block-Toeplitz hybrid approaches per word for Participant 2, condition 1D

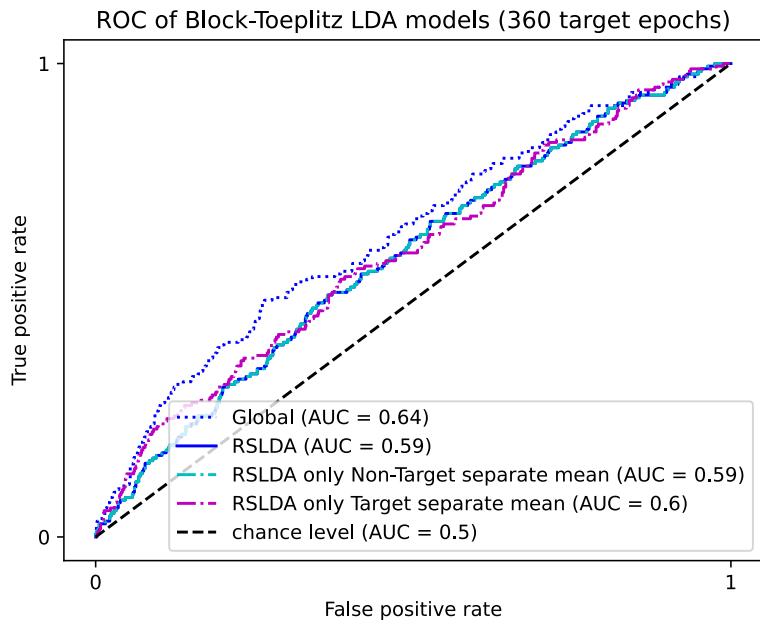


Figure B.24: ROC curves of Block-Toeplitz hybrid approaches per word for Participant 3, condition 1D

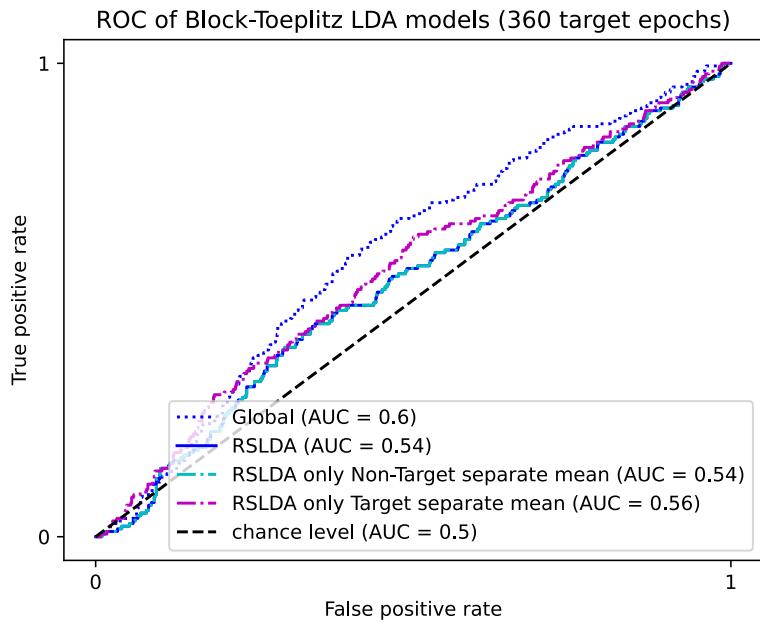


Figure B.25: ROC curves of Block-Toeplitz hybrid approaches per word for Participant 1, condition HP

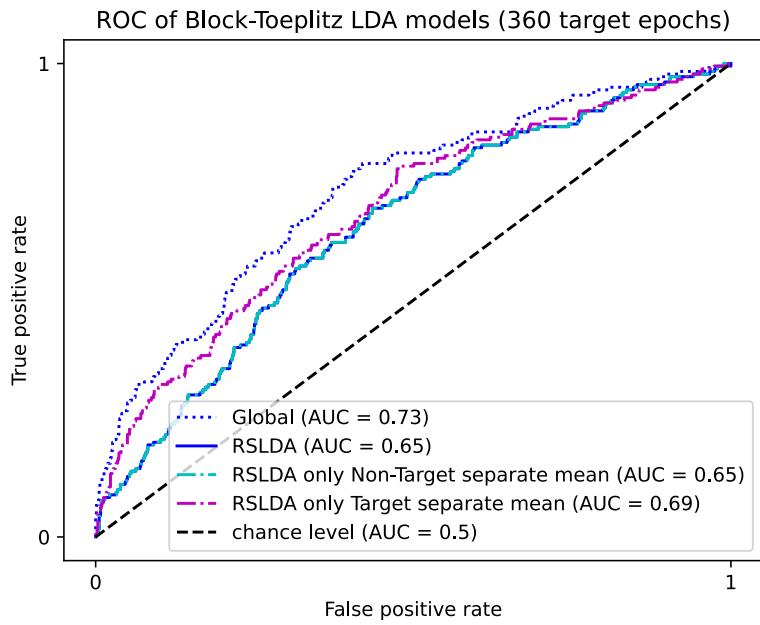


Figure B.26: ROC curves of Block-Toeplitz hybrid approaches per word for Participant 2, condition HP

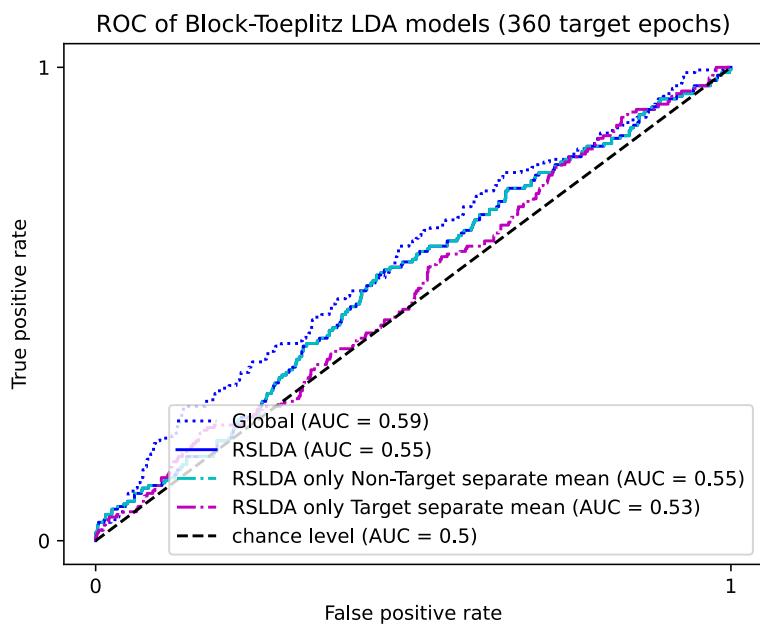


Figure B.27: ROC curves of Block-Toeplitz hybrid approaches per word for Participant 3, condition HP

B.2.1 Ledoit-Wolf

All Words

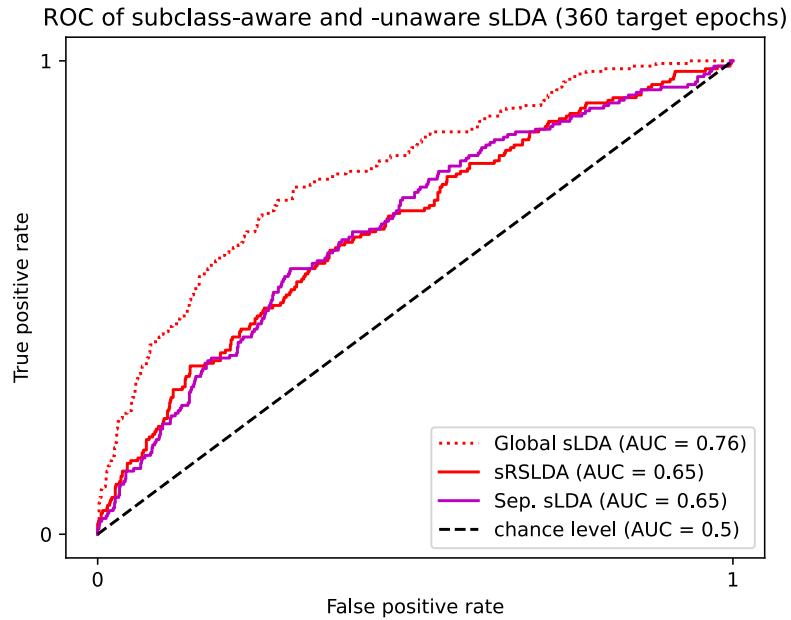


Figure B.1: ROC curves of Ledoit-Wolf shrinkage approaches for Participant 1, condition 6D

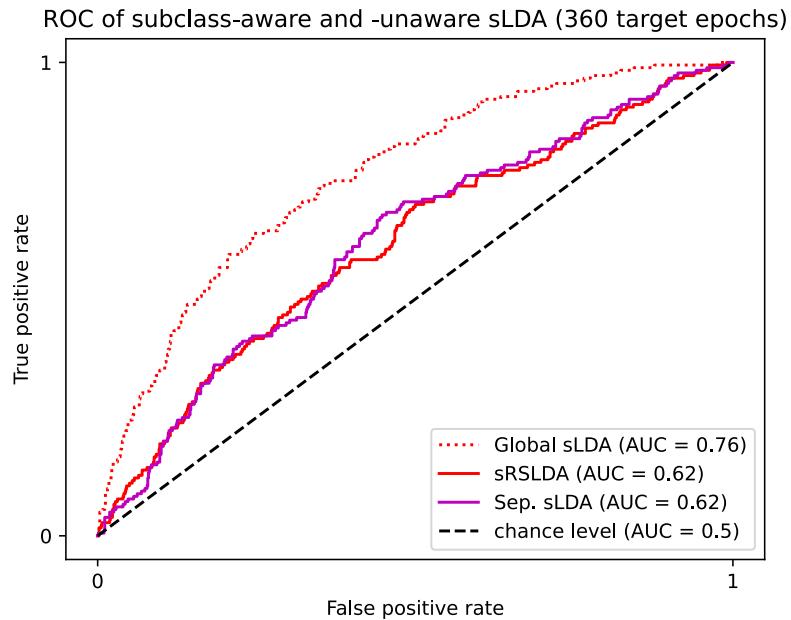


Figure B.2: ROC curves of Ledoit-Wolf shrinkage approaches for Participant 2, condition 6D

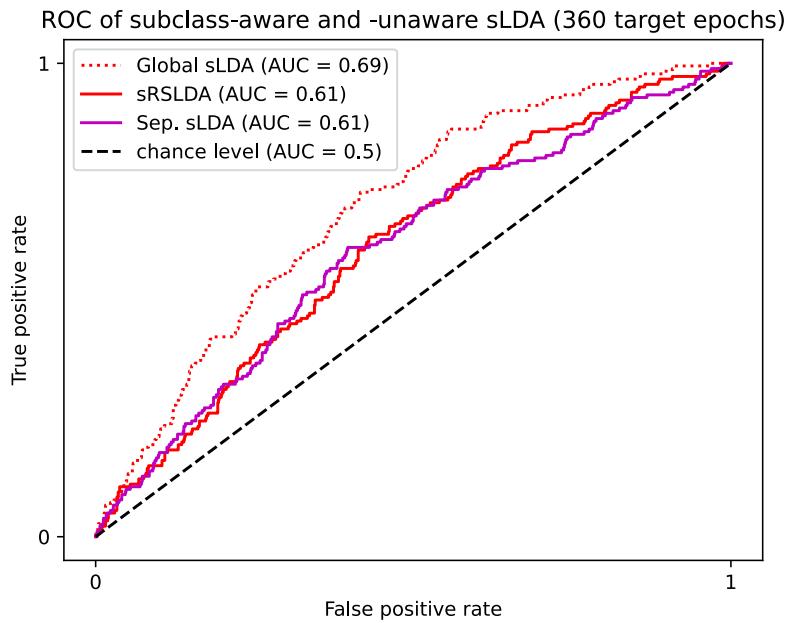


Figure B.3: ROC curves of Ledoit-Wolf shrinkage approaches for Participant 3, condition 6D

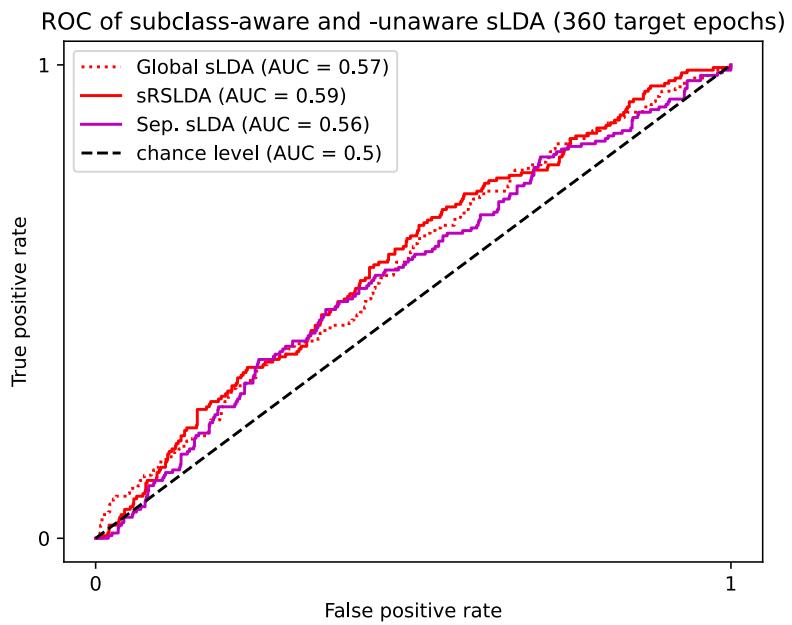


Figure B.4: ROC curves of Ledoit-Wolf shrinkage approaches for Participant 1, condition 1D

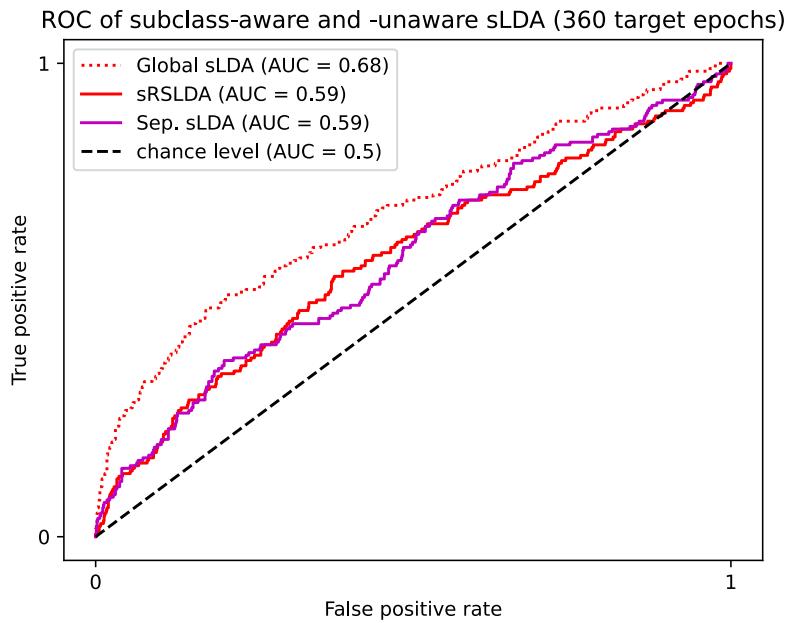


Figure B.5: ROC curves of Ledoit-Wolf shrinkage approaches for Participant 2, condition 1D

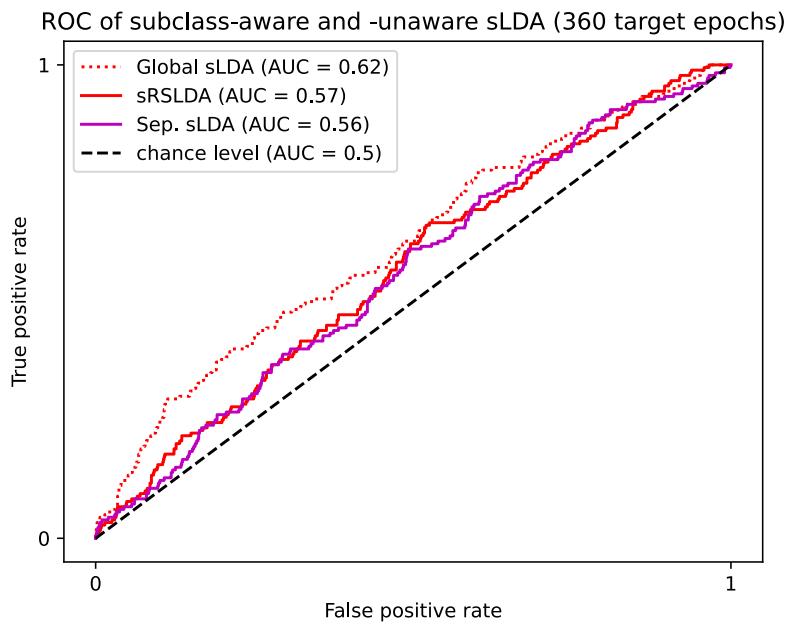


Figure B.6: ROC curves of Ledoit-Wolf shrinkage approaches for Participant 3, condition 1D

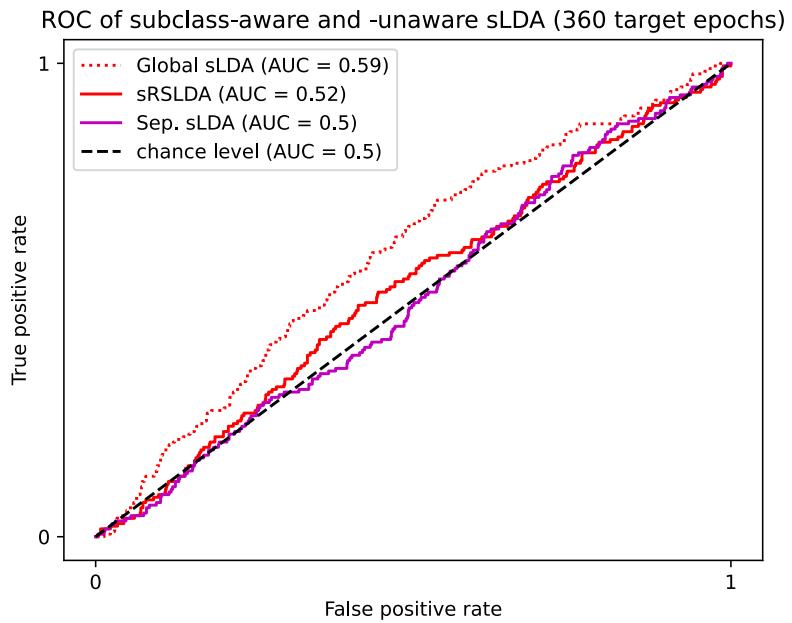


Figure B.7: ROC curves of Ledoit-Wolf shrinkage approaches for Participant 1, condition HP

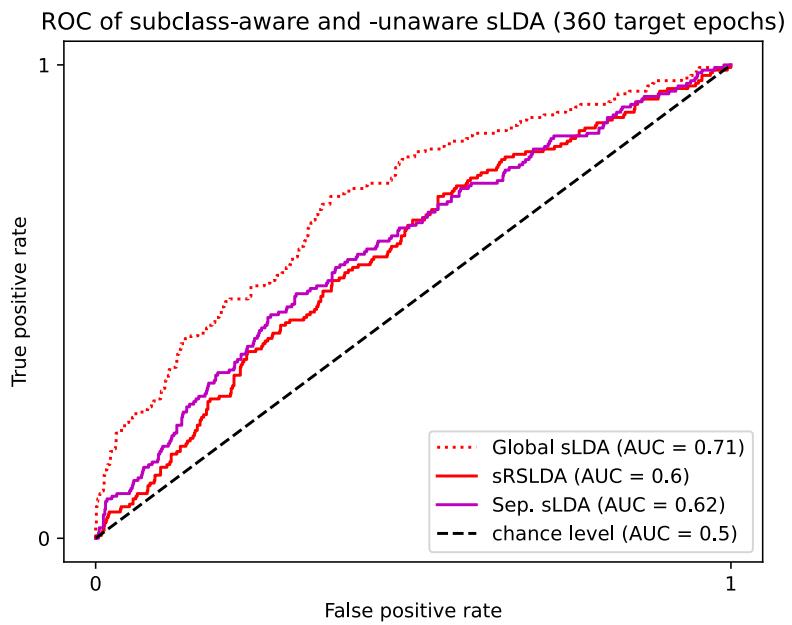


Figure B.8: ROC curves of Ledoit-Wolf shrinkage approaches for Participant 2, condition HP

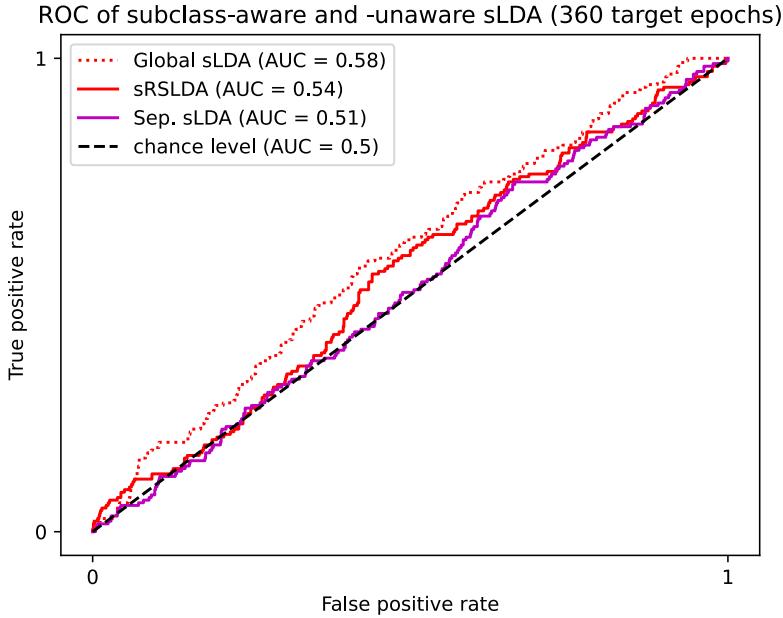


Figure B.9: ROC curves of Ledoit-Wolf shrinkage approaches for Participant 3, condition HP

Per Word

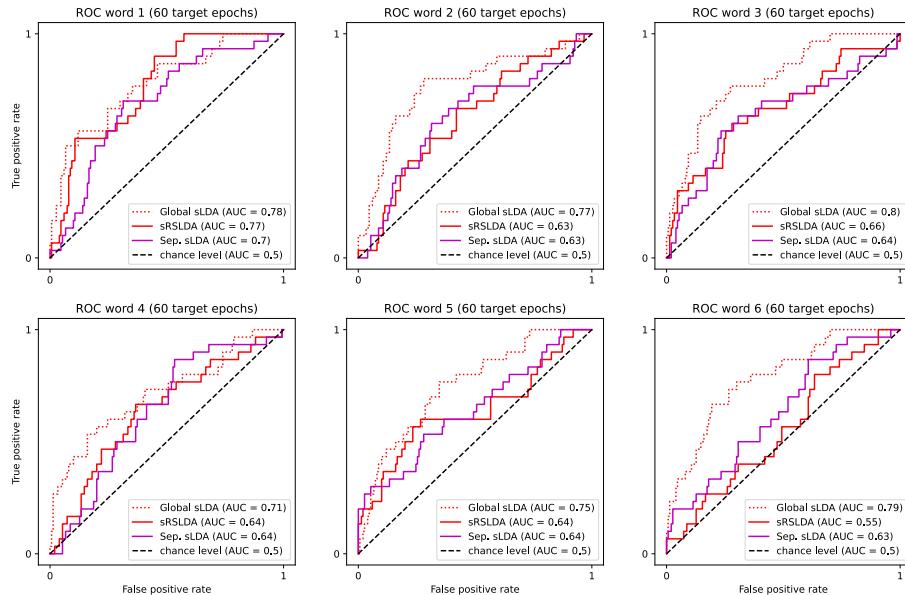


Figure B.10: ROC curves of Ledoit-Wolf approaches per word for Participant 1, condition 6D

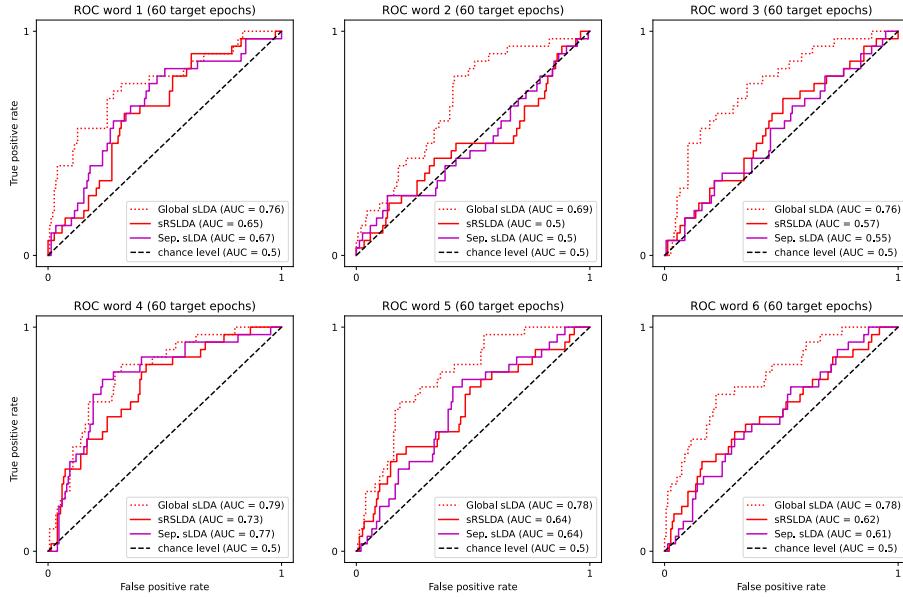


Figure B.11: ROC curves of Ledoit-Wolf approaches per word for Participant 2, condition 6D

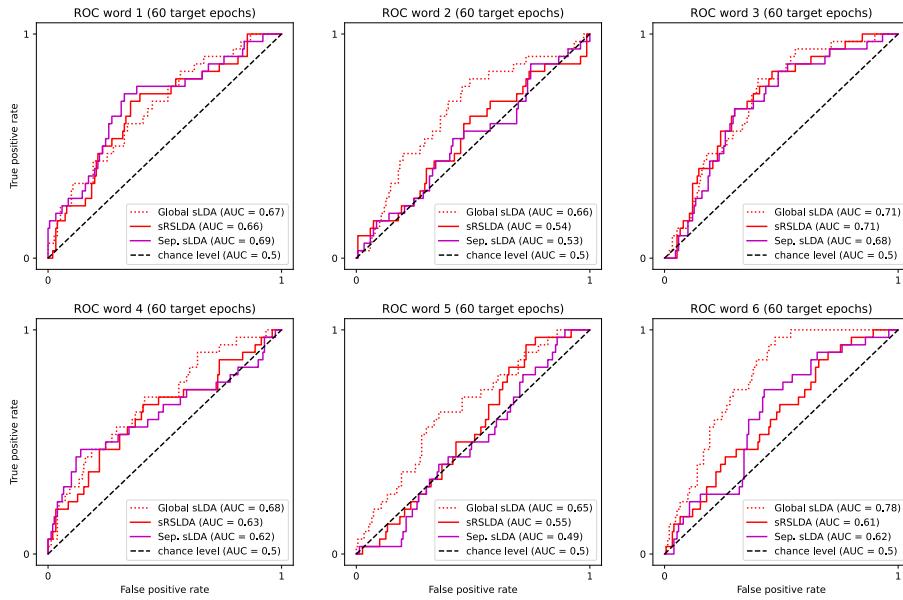


Figure B.12: ROC curves of Ledoit-Wolf approaches per word for Participant 3, condition 6D

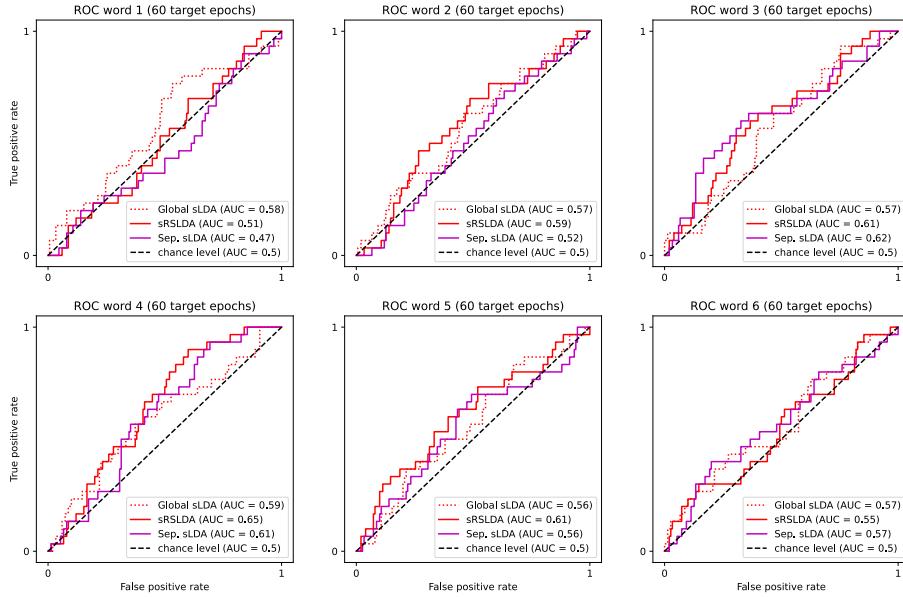


Figure B.13: ROC curves of Ledoit-Wolf approaches per word for Participant 1, condition 1D

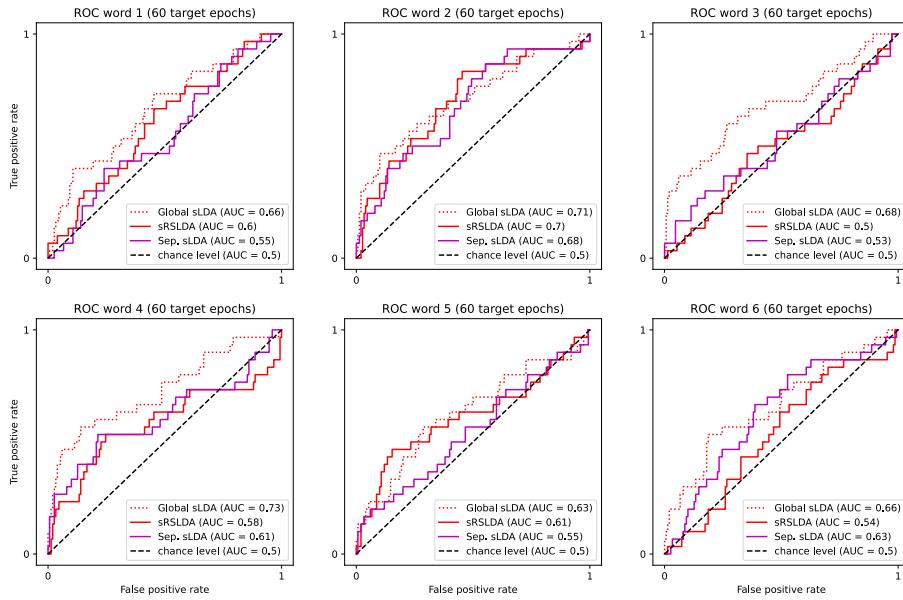


Figure B.14: ROC curves of Ledoit-Wolf approaches per word for Participant 2, condition 1D

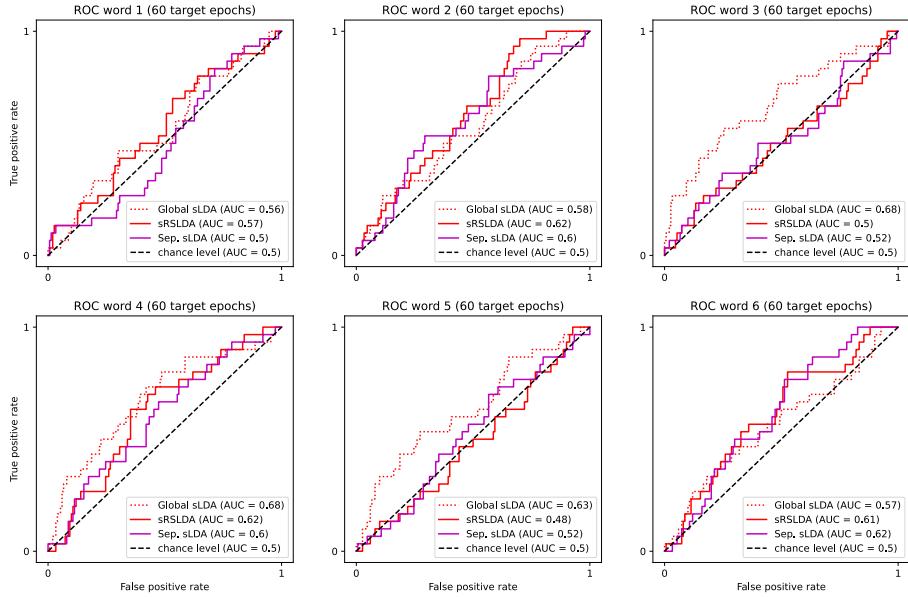


Figure B.15: ROC curves of Ledoit-Wolf approaches per word for Participant 3, condition 1D

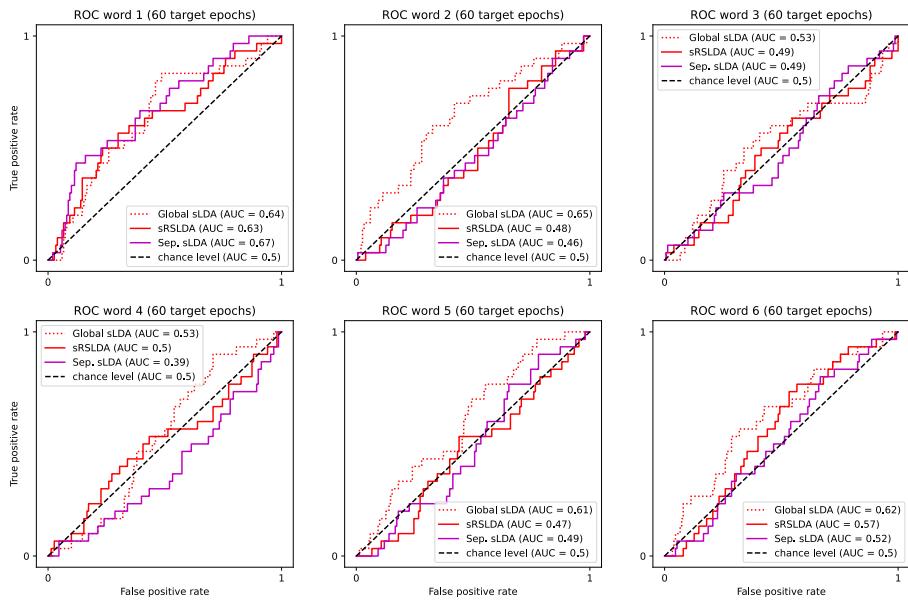


Figure B.16: ROC curves of Ledoit-Wolf approaches per word for Participant 1, condition HP

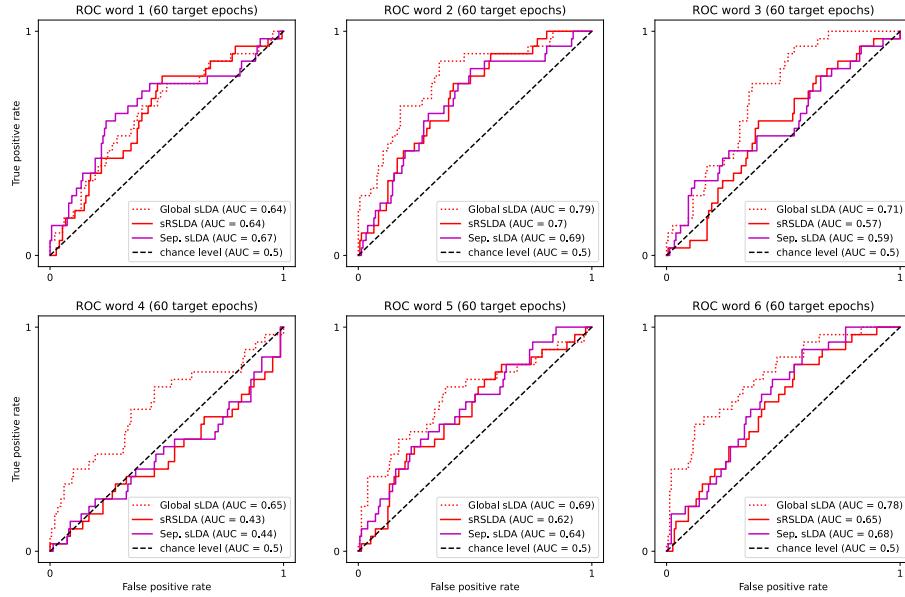


Figure B.17: ROC curves of Ledoit-Wolf approaches per word for Participant 2, condition HP

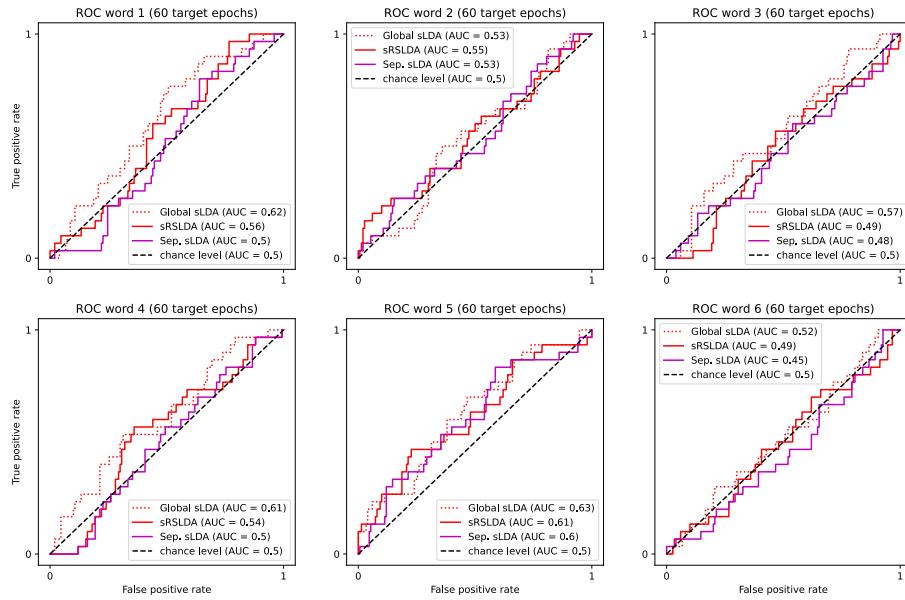


Figure B.18: ROC curves of Ledoit-Wolf approaches per word for Participant 3, condition HP

B.3 Riemannian ROC Curves

Below are all the ROC curves of the Riemannian approaches.

B.3.1 Parallel Transport

All Words

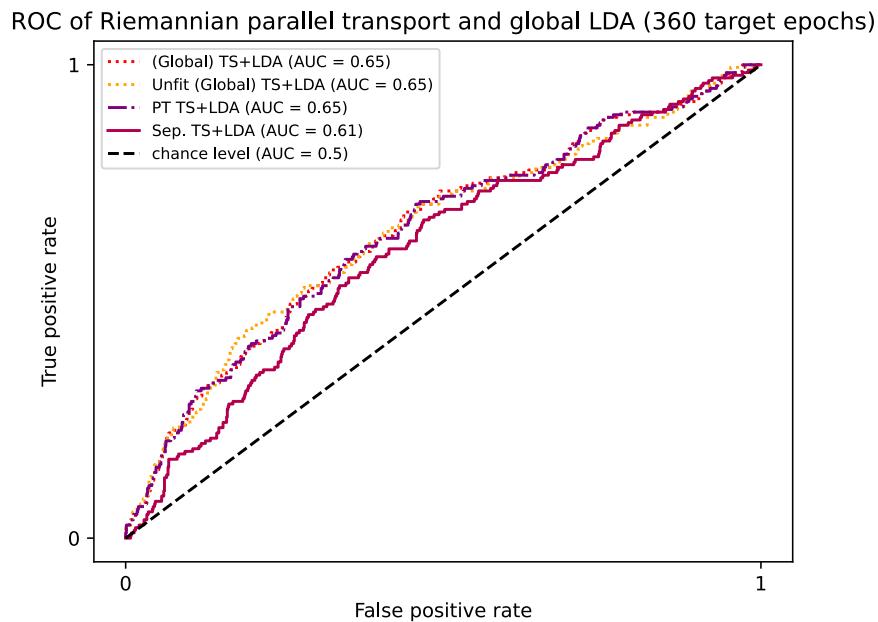


Figure B.1: ROC curves of parallel transport and global approaches for Participant 1, condition 6D

ROC of Riemannian parallel transport and global LDA (360 target epochs)

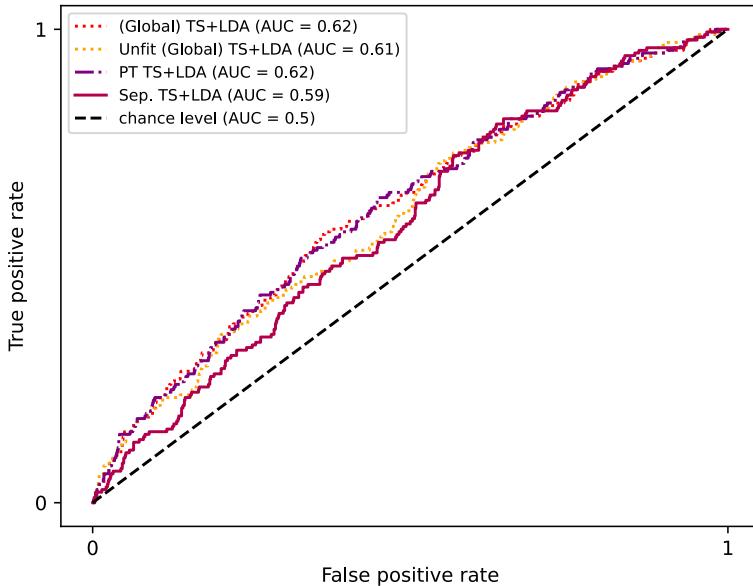


Figure B.2: ROC curves of parallel transport and global approaches for Participant 2, condition 6D

ROC of Riemannian parallel transport and global LDA (360 target epochs)

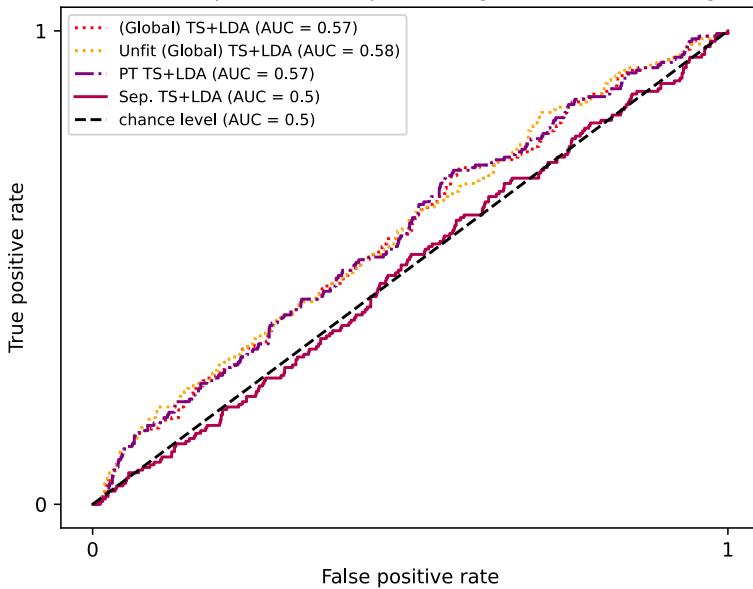


Figure B.3: ROC curves of parallel transport and global approaches for Participant 3, condition 6D

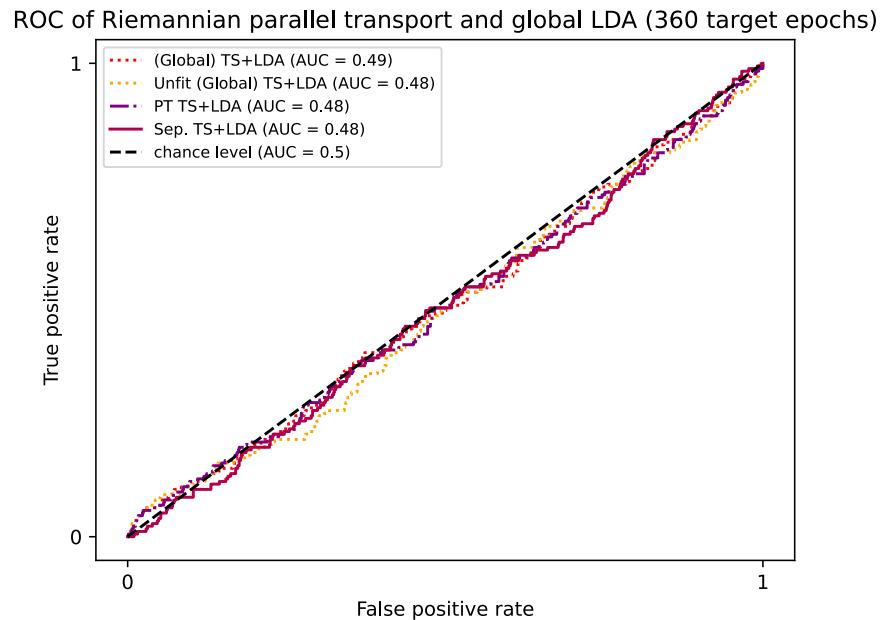


Figure B.4: ROC curves of parallel transport and global approaches for Participant 1, condition 1D

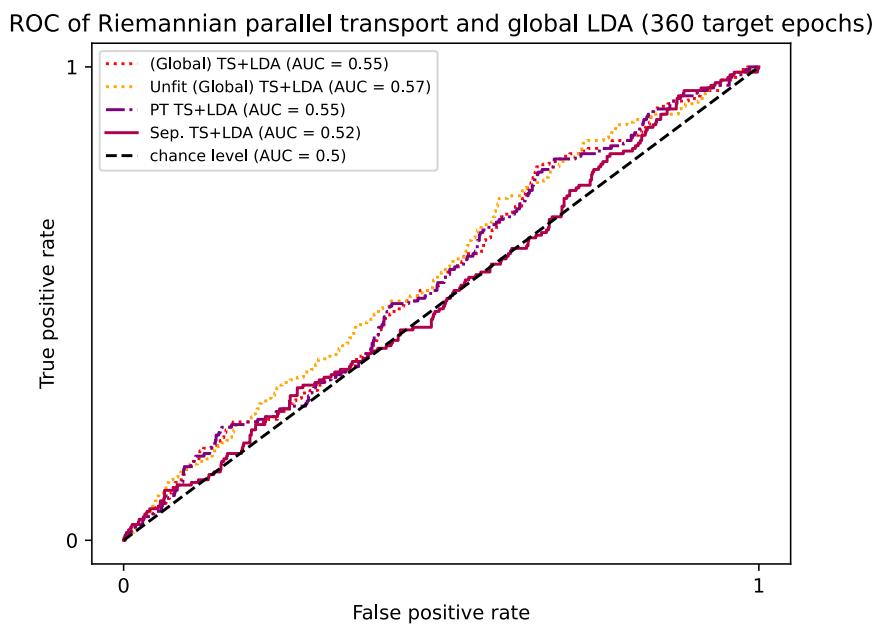


Figure B.5: ROC curves of parallel transport and global approaches for Participant 2, condition 1D

ROC of Riemannian parallel transport and global LDA (360 target epochs)

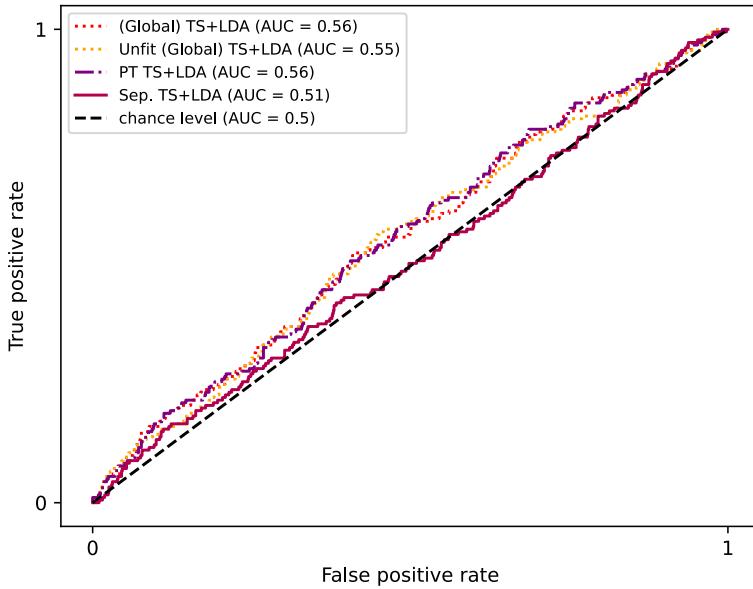


Figure B.6: ROC curves of parallel transport and global approaches for Participant 3, condition 1D

ROC of Riemannian parallel transport and global LDA (360 target epochs)

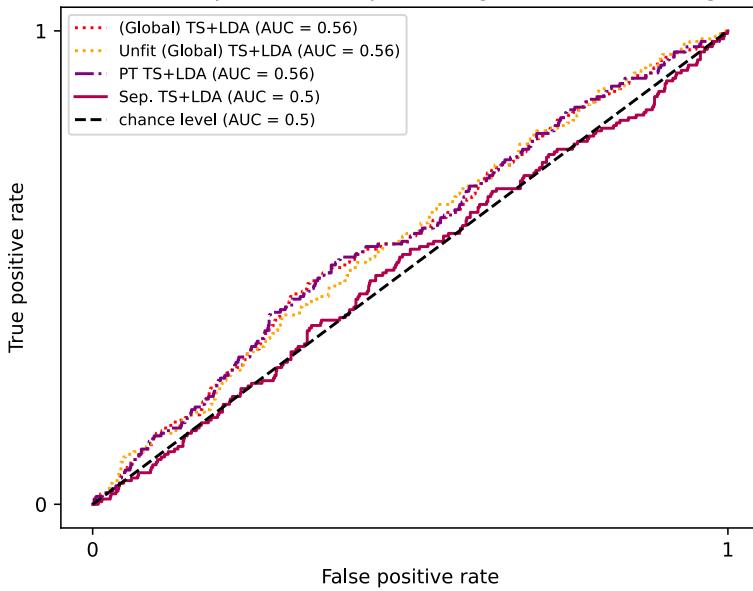


Figure B.7: ROC curves of parallel transport and global approaches for Participant 1, condition HP

ROC of Riemannian parallel transport and global LDA (360 target epochs)

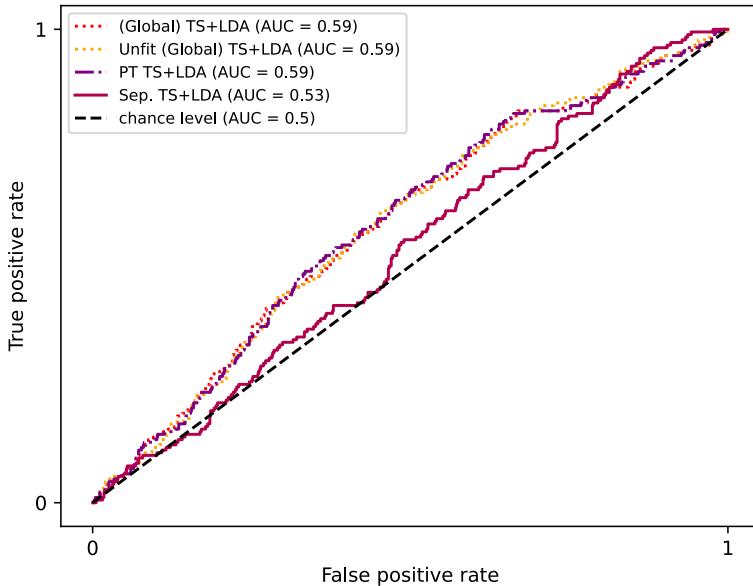


Figure B.8: ROC curves of parallel transport and global approaches for Participant 2, condition HP

ROC of Riemannian parallel transport and global LDA (360 target epochs)

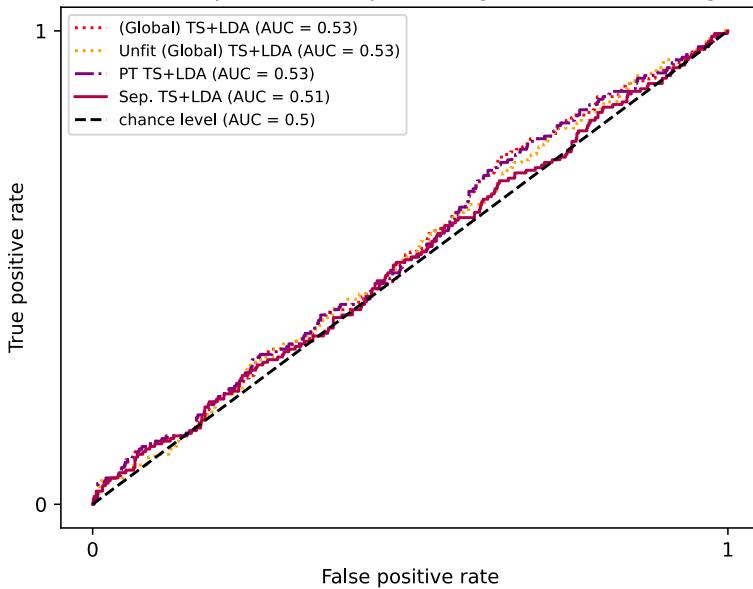


Figure B.9: ROC curves of parallel transport and global approaches for Participant 3, condition HP

Per Word

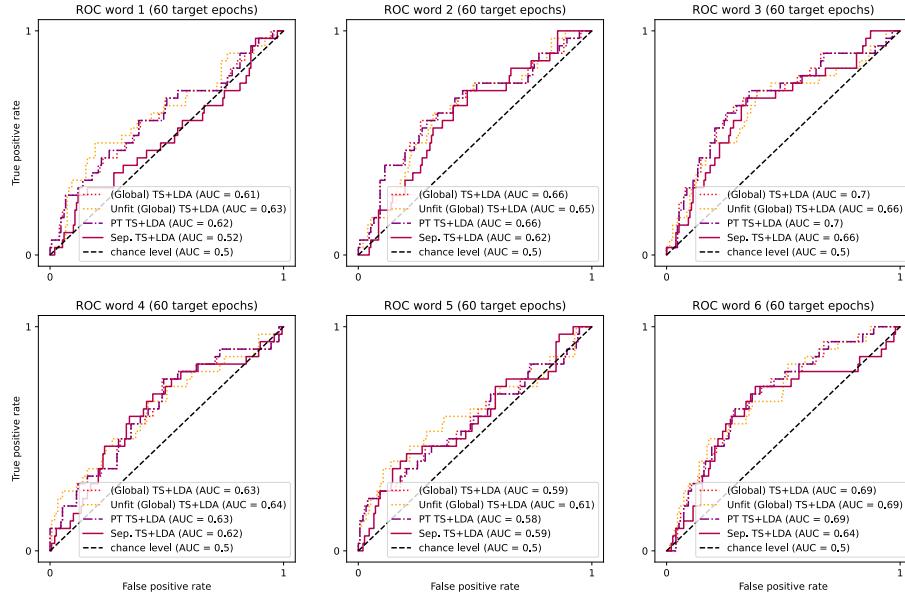


Figure B.10: ROC curves of parallel transport and global approaches per word for Participant 1, condition 6D

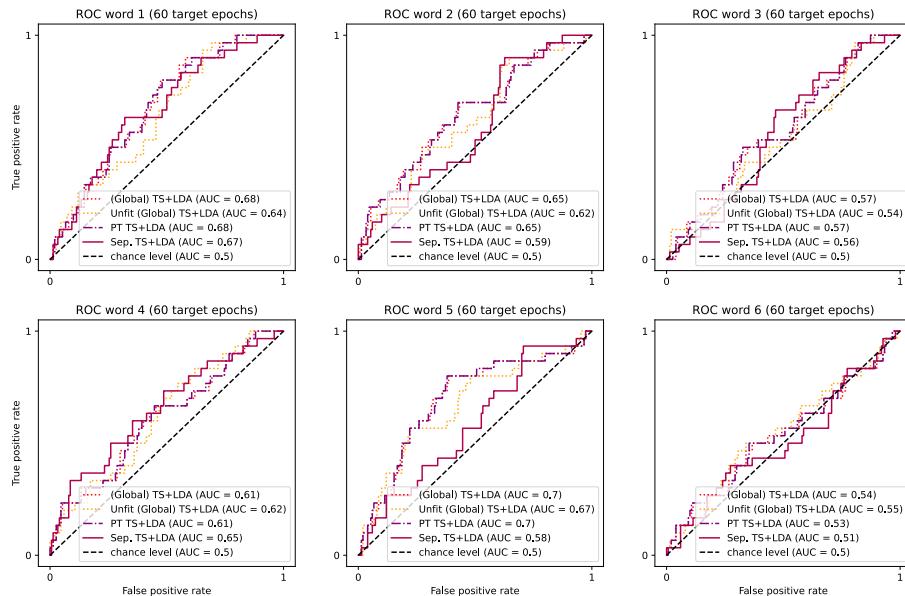


Figure B.11: ROC curves of parallel transport and global approaches per word for Participant 2, condition 6D

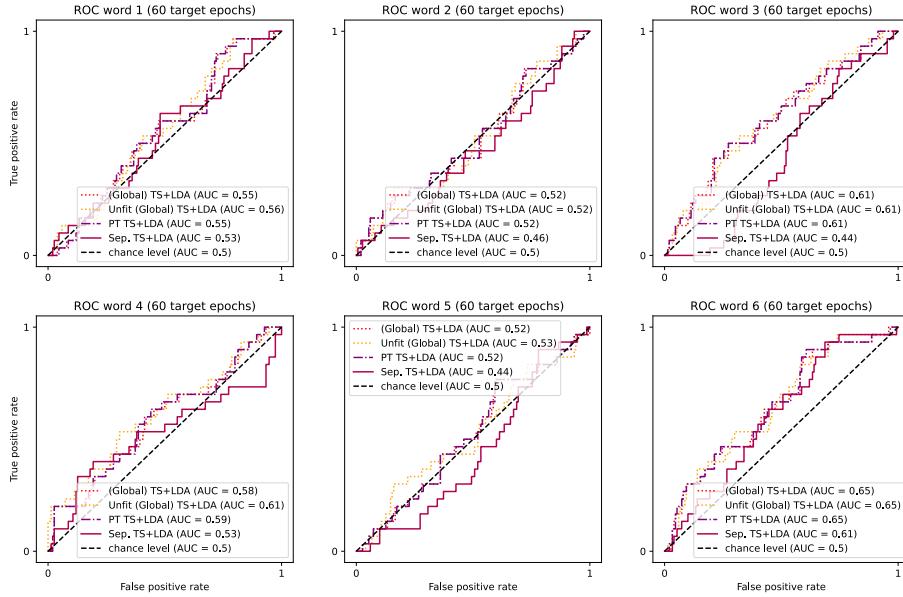


Figure B.12: ROC curves of parallel transport and global approaches per word for Participant 3, condition 6D

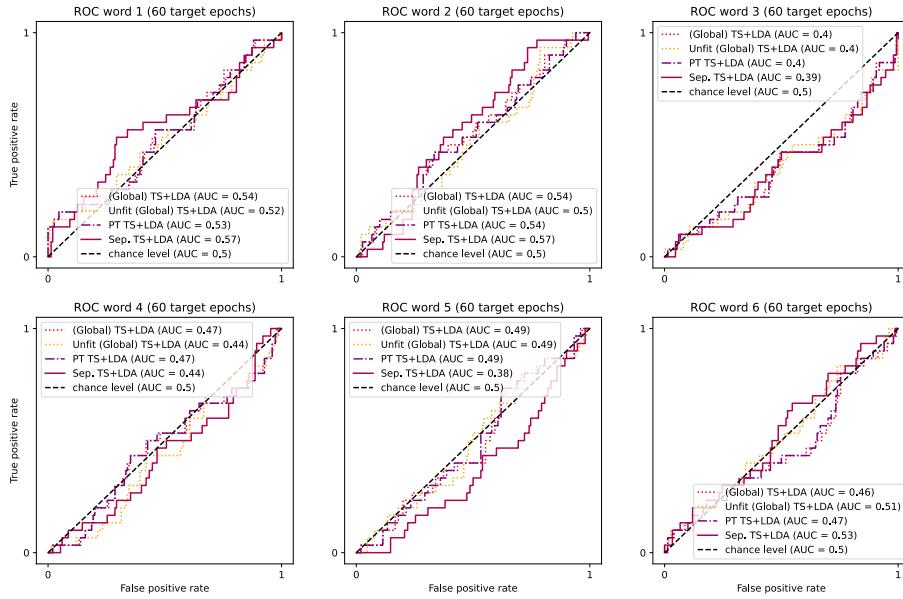


Figure B.13: ROC curves of parallel transport and global approaches per word for Participant 1, condition 1D

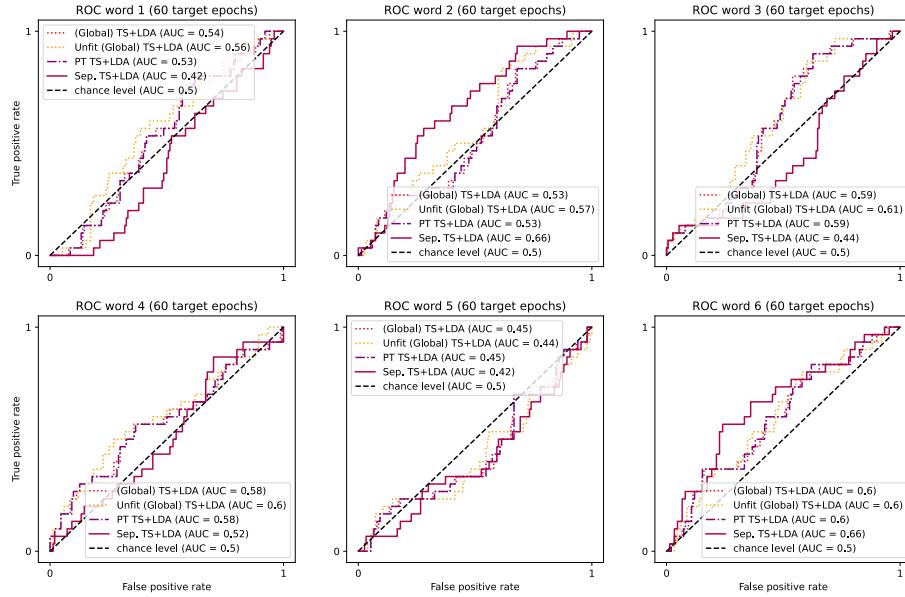


Figure B.14: ROC curves of parallel transport and global approaches per word for Participant 2, condition 1D

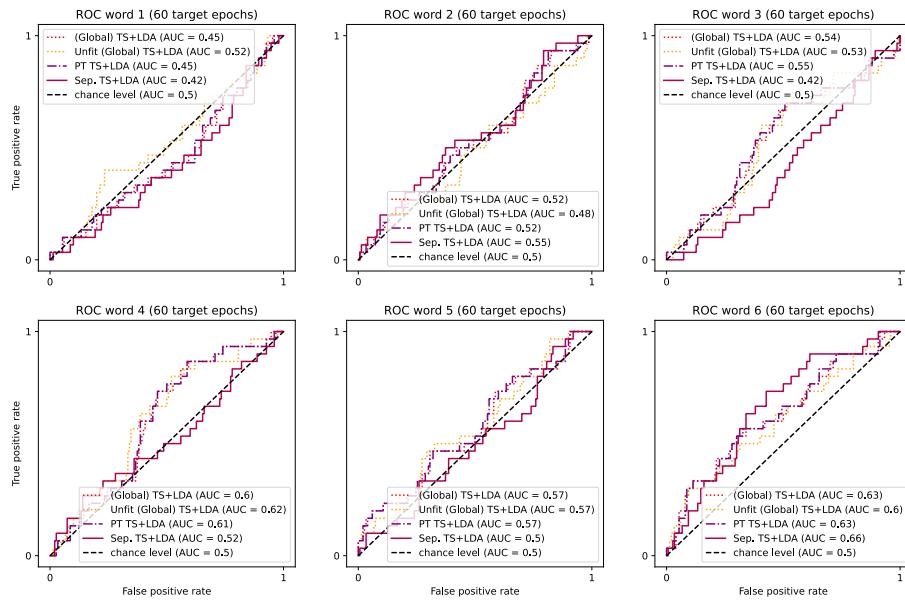


Figure B.15: ROC curves of parallel transport and global approaches per word for Participant 3, condition 1D

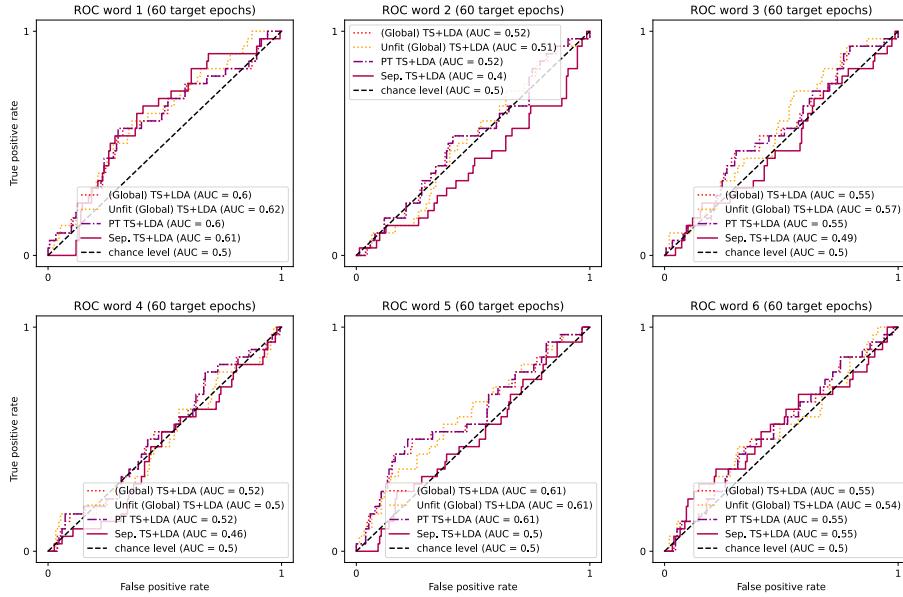


Figure B.16: ROC curves of parallel transport and global approaches per word for Participant 1, condition HP

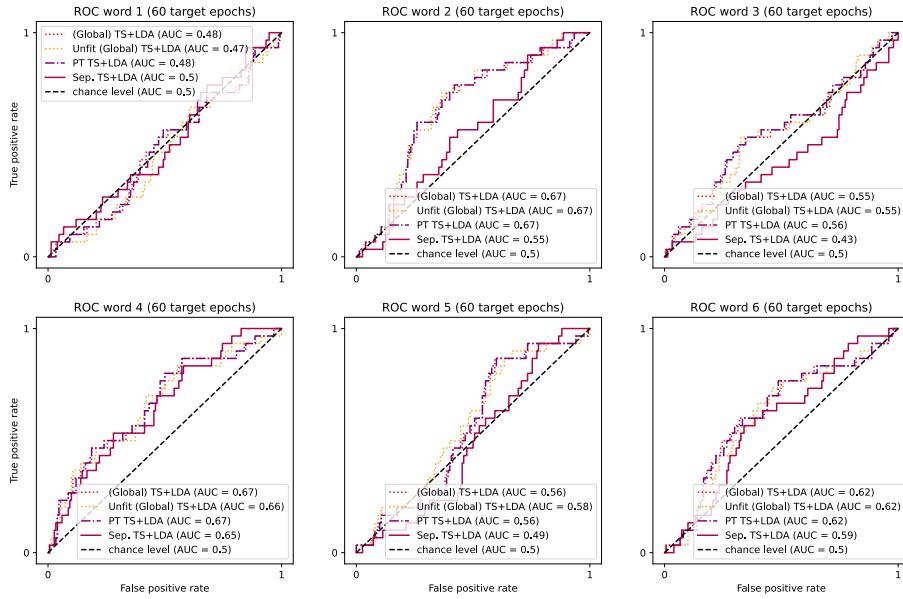


Figure B.17: ROC curves of parallel transport and global approaches per word for Participant 2, condition HP

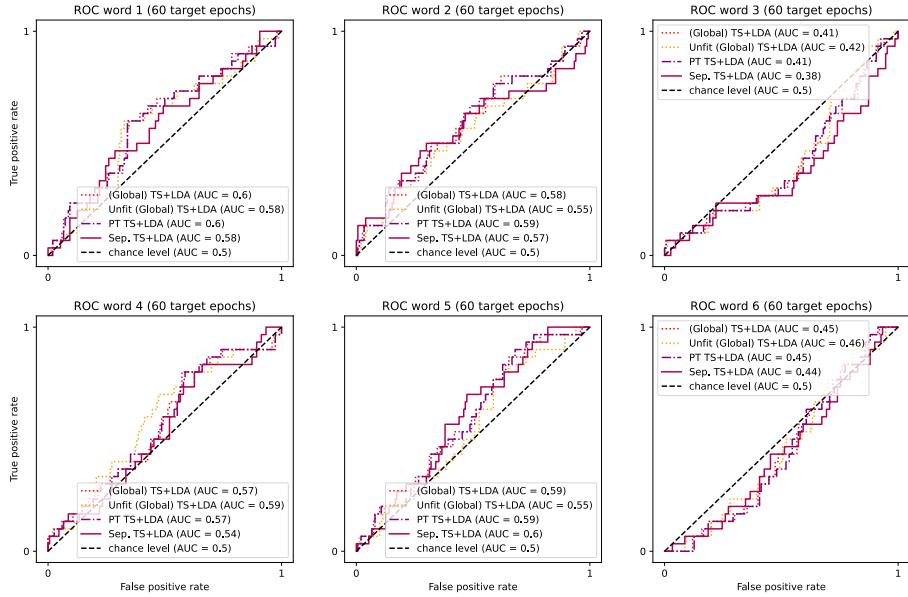


Figure B.18: ROC curves of parallel transport and global approaches per word for Participant 3, condition HP

B.3.2 Sub-Prototype

All Words

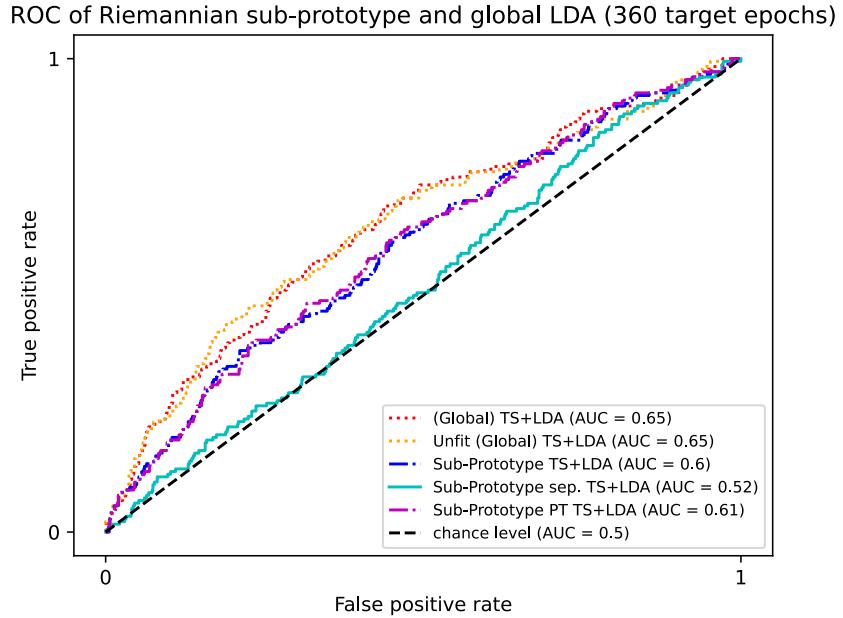


Figure B.19: ROC curves of sub-prototype and global approaches for Participant 1, condition 6D

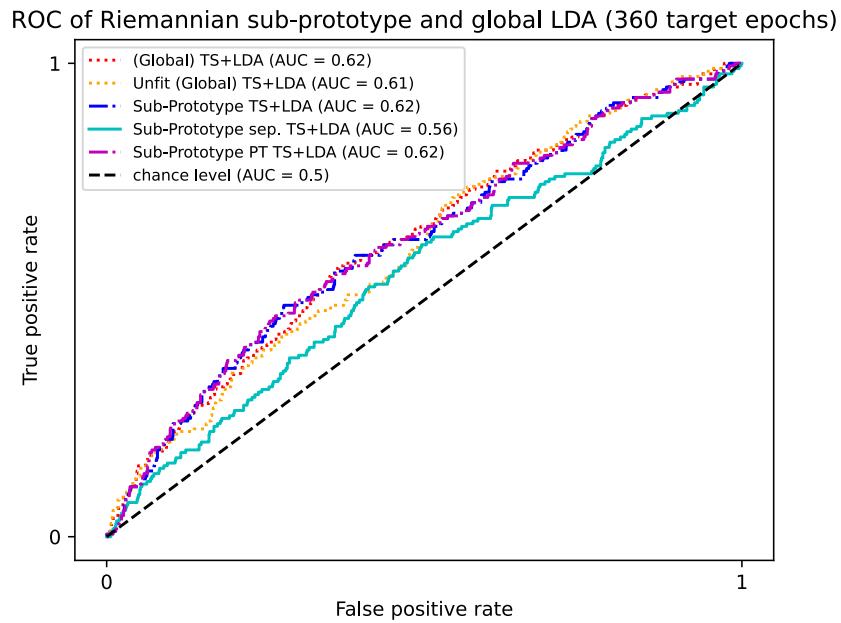


Figure B.20: ROC curves of sub-prototype and global approaches for Participant 2, condition 6D

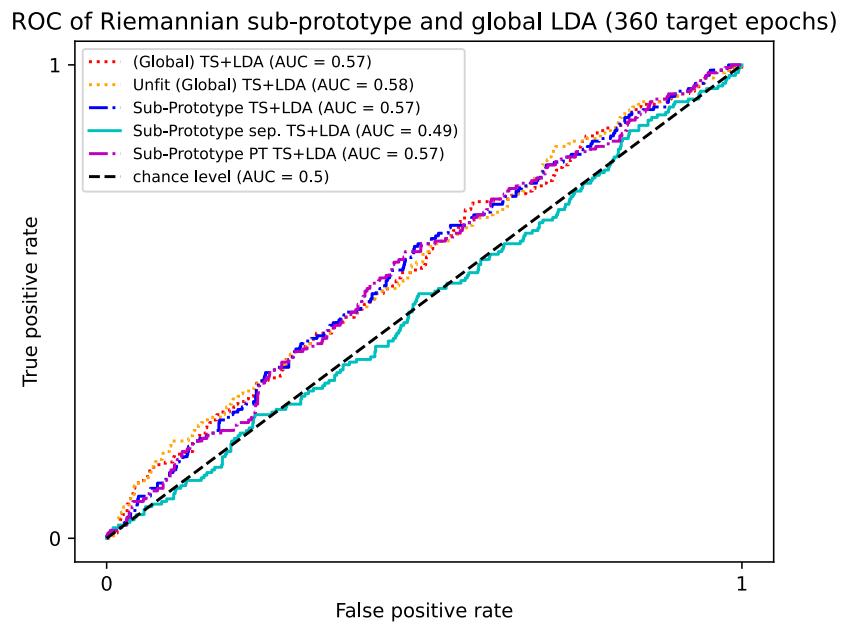


Figure B.21: ROC curves of sub-prototype and global approaches for Participant 3, condition 6D

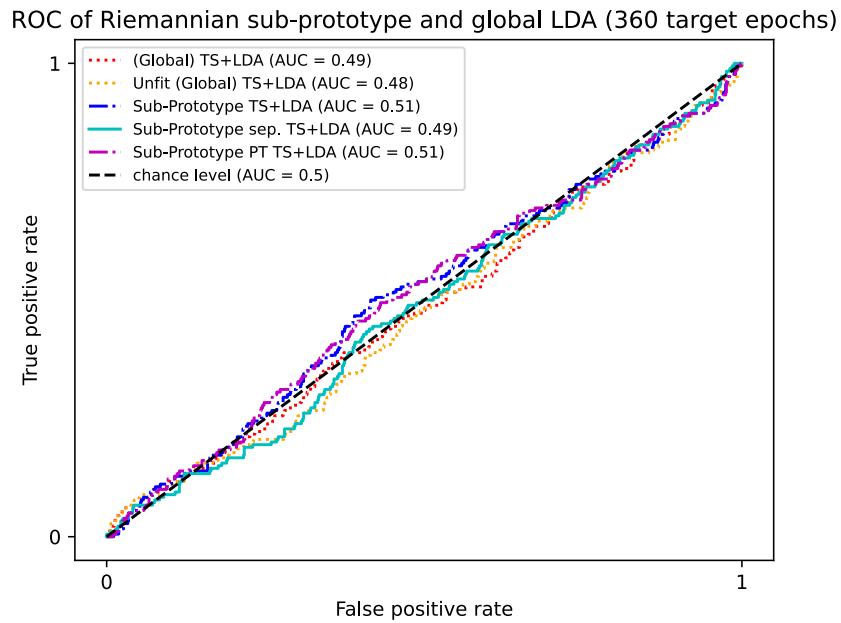


Figure B.22: ROC curves of sub-prototype and global approaches for Participant 1, condition 1D

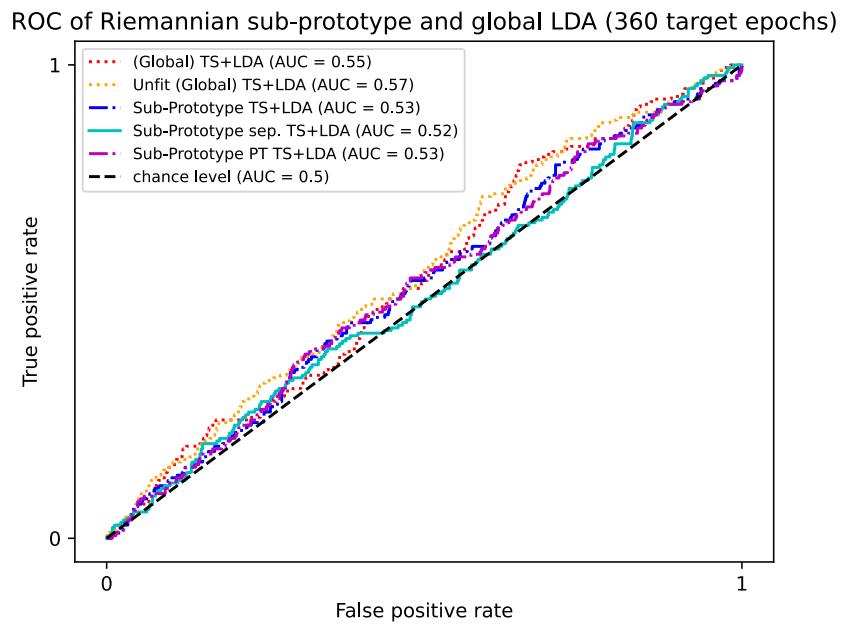


Figure B.23: ROC curves of sub-prototype and global approaches for Participant 2, condition 1D

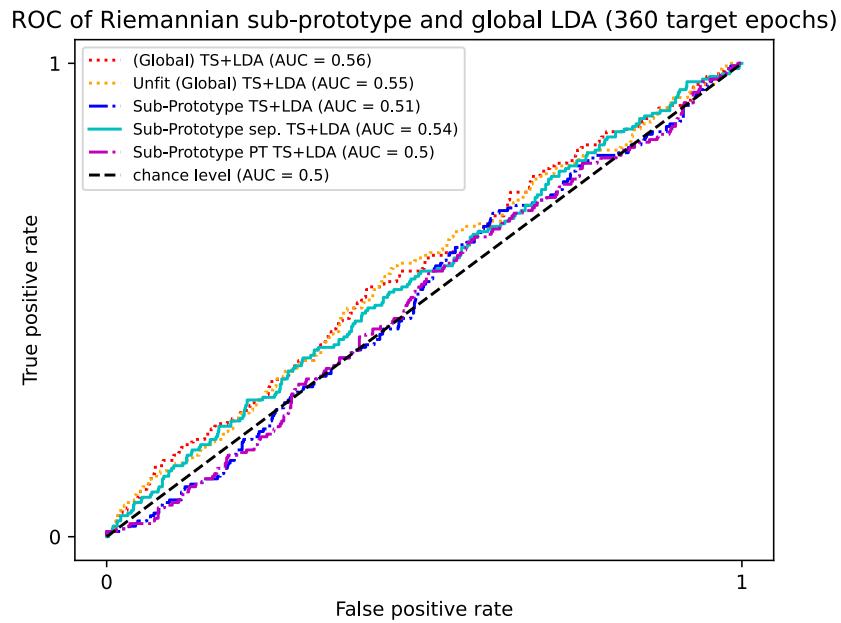


Figure B.24: ROC curves of sub-prototype and global approaches for Participant 3, condition 1D

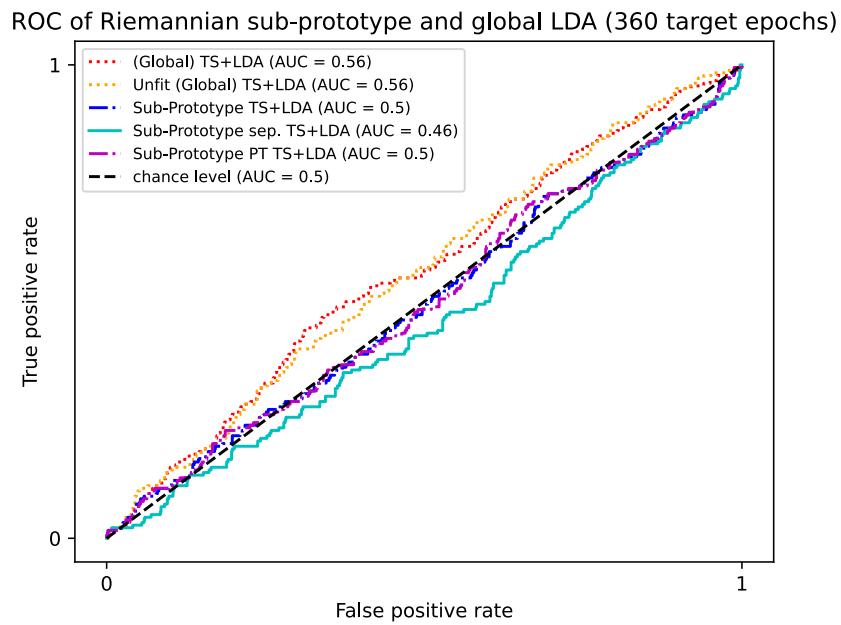


Figure B.25: ROC curves of sub-prototype and global approaches for Participant 1, condition HP

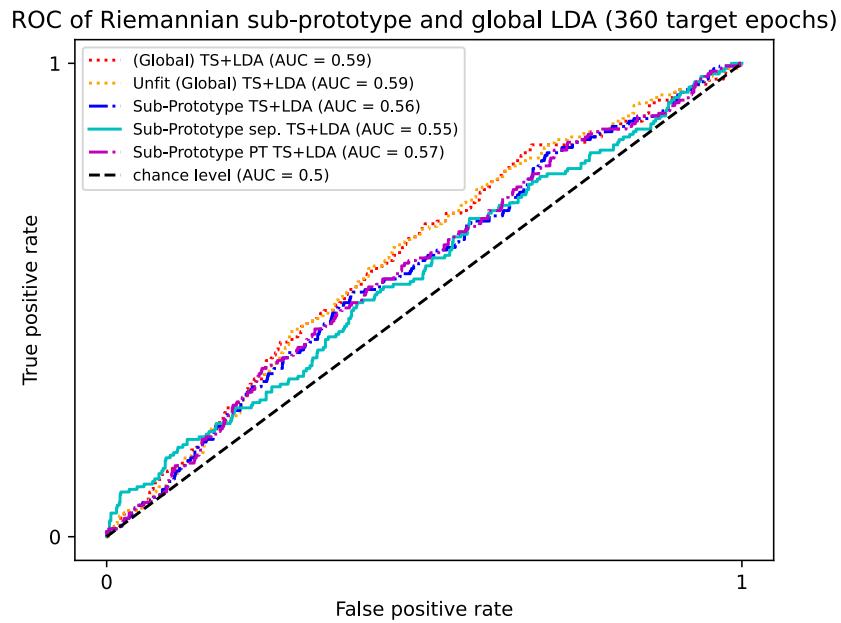


Figure B.26: ROC curves of sub-prototype and global approaches for Participant 2, condition HP

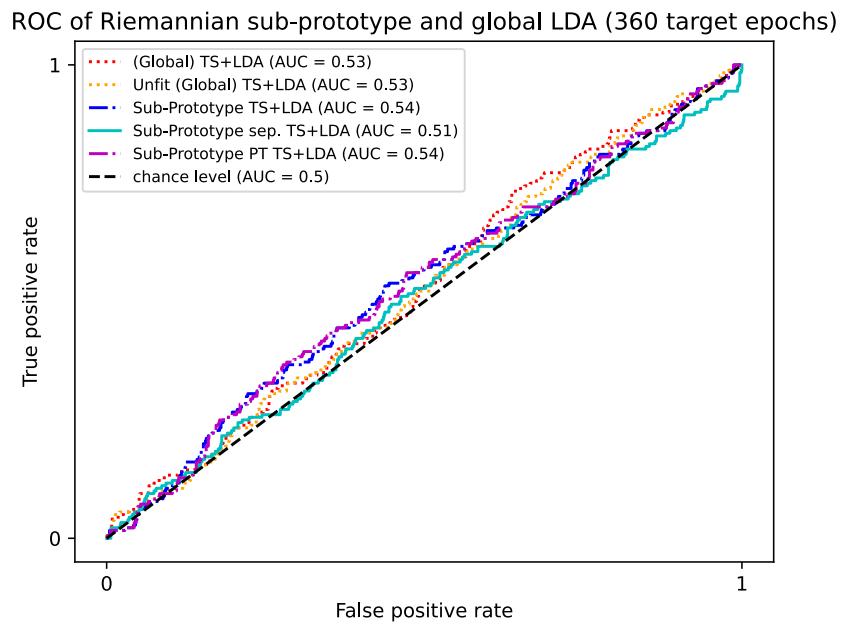


Figure B.27: ROC curves of sub-prototype and global approaches for Participant 3, condition HP

Per Word

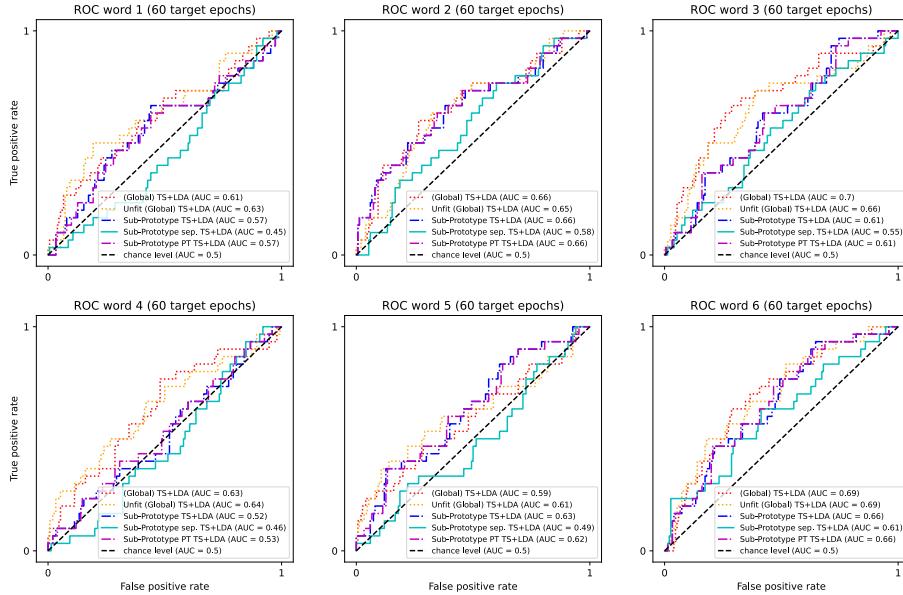


Figure B.28: ROC curves of sub-prototype and global approaches per word for Participant 1, condition 6D

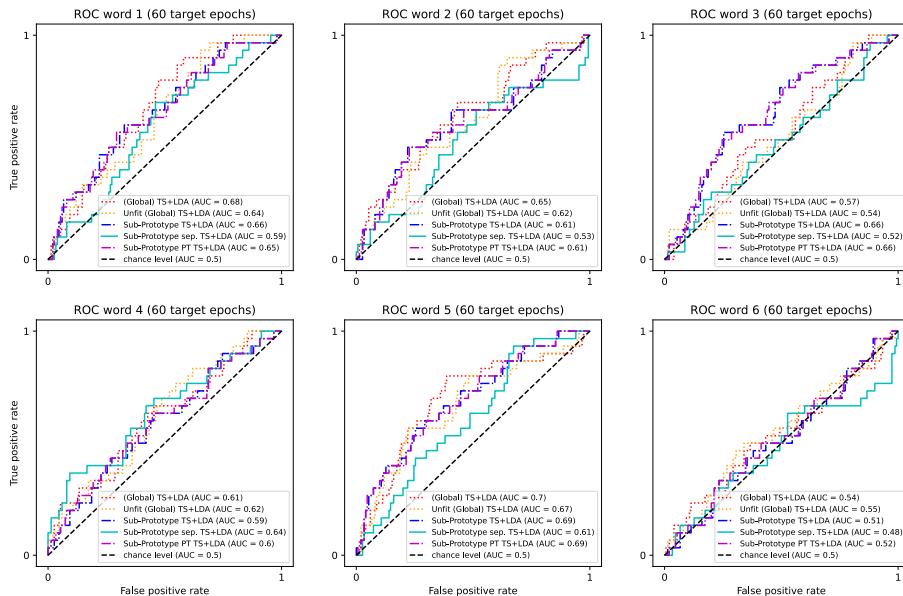


Figure B.29: ROC curves of sub-prototype and global approaches per word for Participant 2, condition 6D

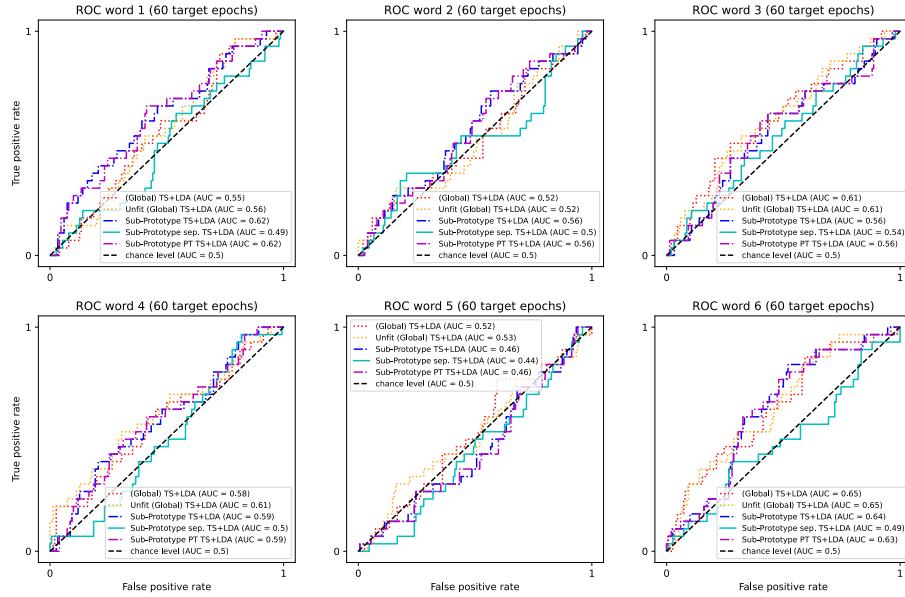


Figure B.30: ROC curves of sub-prototype and global approaches per word for Participant 3, condition 6D

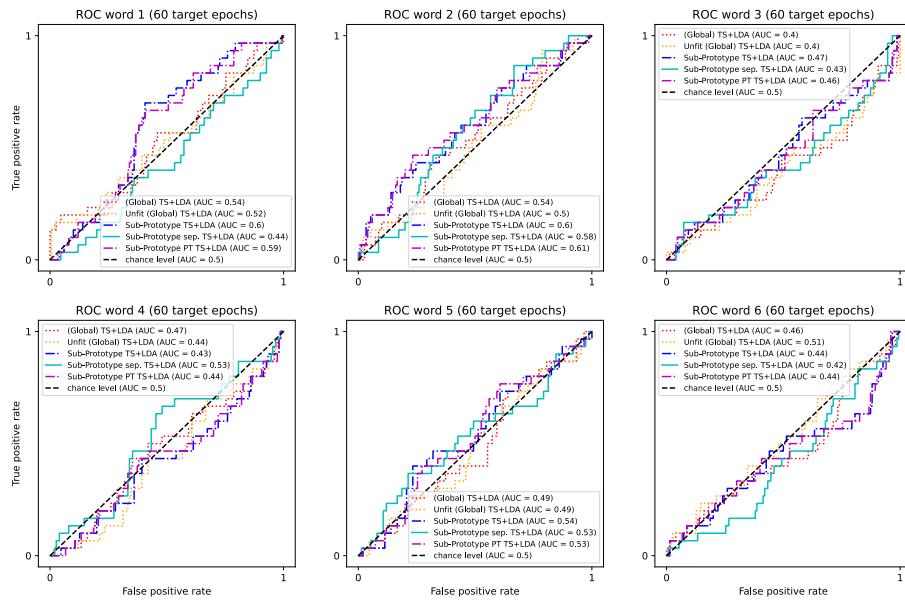


Figure B.31: ROC curves of sub-prototype and global approaches per word for Participant 1, condition 1D

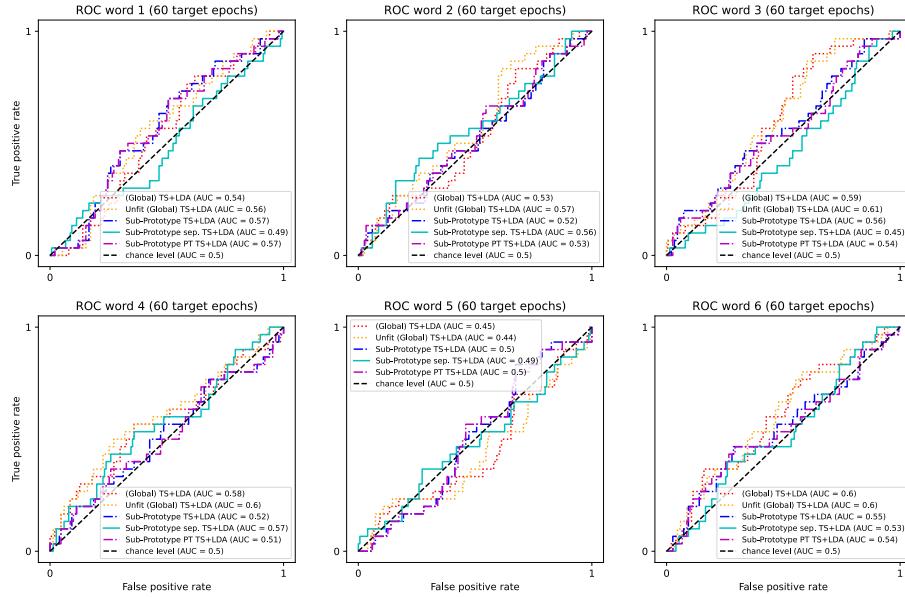


Figure B.32: ROC curves of sub-prototype and global approaches per word for Participant 2, condition 1D

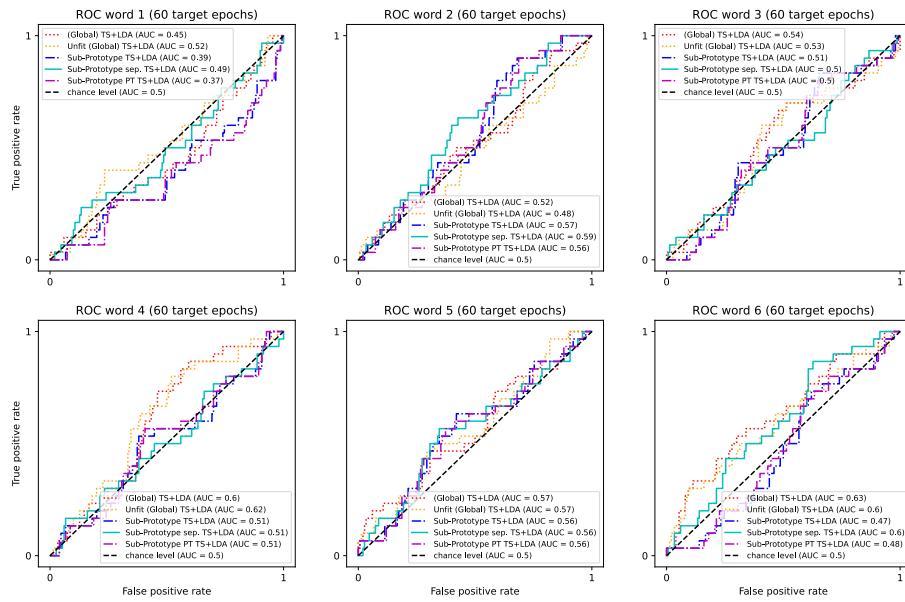


Figure B.33: ROC curves of sub-prototype and global approaches per word for Participant 3, condition 1D

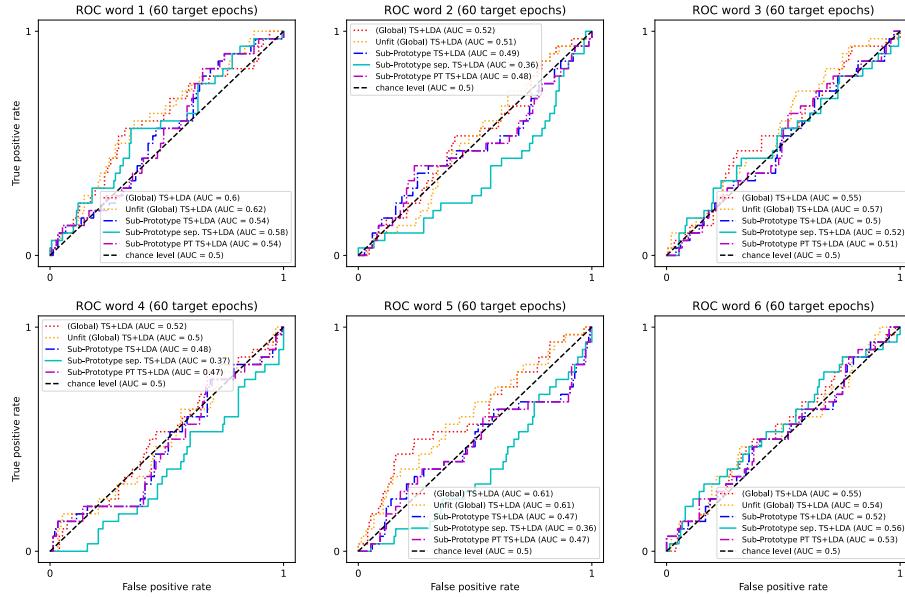


Figure B.34: ROC curves of sub-prototype and global approaches per word for Participant 1, condition HP

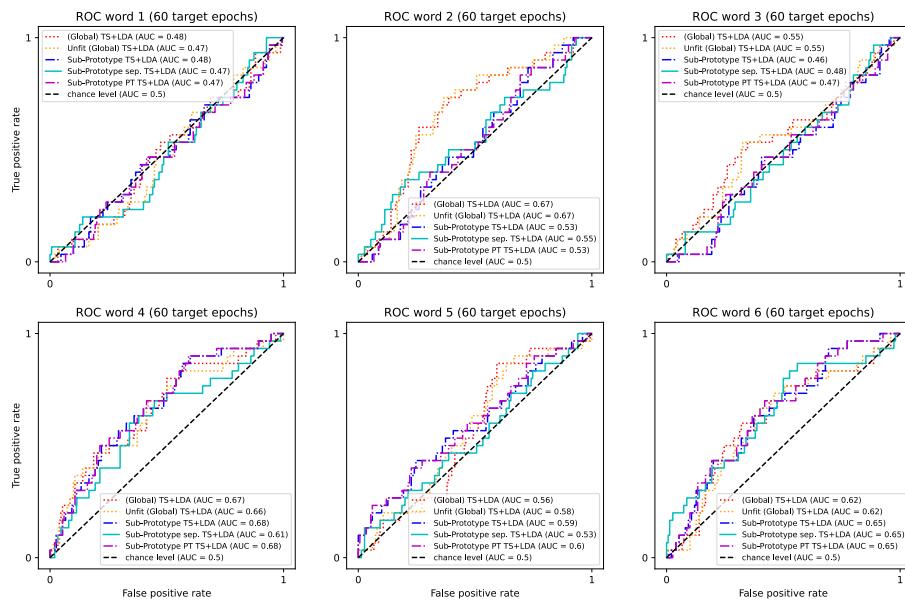


Figure B.35: ROC curves of sub-prototype and global approaches per word for Participant 2, condition HP

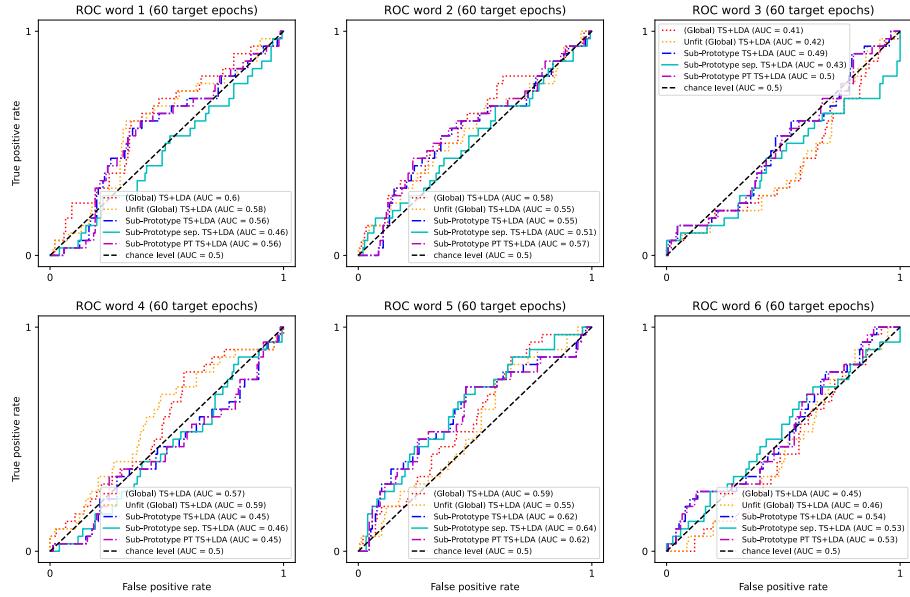


Figure B.36: ROC curves of sub-prototype and global approaches per word for Participant 3, condition HP