# Writing Procedure

\begin{enumerate}

\item -> Background research (Kmeans, HAC studies in AI and Neuro)

\item -> Background

\item -> Methods, proposed solution

\item -> Results/Key Insights/Findings

\item -> What was unknown?, RQ

\item -> Discussion: Interpretation, implications, comparison with existing literature, and potential applications of findings. Discuss the significance of results in the broader context of the field.

\item -> Motivation

\item -> Additional Motivations

\item -> Key findings in Introduction & Study limitations

\item -> Introduction paragraph

\item -> Structure

\item -> Discussion: Conclusion, future work

\item -> Abstract

\end{enumerate}

# Introduction

## Representational geometries, and clustering methods for analyzing representational geometries

And in particular, the representations of \textit{visual and auditory stimuli}. These have been subject of particular study due to their proximity between the representation and physical space. What's the plan? I know I need to just start. But what else. I will not do it

The aim of representational theory is to find the mapping between input and representational space.

These three spaces can be measured and then modelled. Input spaces can sometimes be measured and modeled by science, but these ostentatiously objective models are of little interest for the study of representations.

Representational similarity analysis (RSA) measures the correlates between models of stimuli, representations, and meanings. The physical proximity of representation (e.g., V1) to content (e.g., a color stimulus) makes it possible to validate any representation models. We cannot expect the relational structures to be isomorphic across the input and representational domains.

## Representations of symbols in the brain, Speech/phonemic data, TIMIT dataset, representational geometry study on TIMIT

- Define symbols, concepts

- Introduce the TIMIT dataset briefly (its purpose, structure, and significance).

- Discuss previous research that has utilized the TIMIT dataset. Highlight key findings, methodologies employed, and any gaps or limitations identified in these studies.

- Brains and ANNs are commonly thought of as consisting representational cascades.

- Meaning: In the brain it has sensory and a semantic representation. (..2 spaces: semantic space and the representational space.)

Motor-perceptual features are through interaction:

- Articulation, pitch…

Phone conceptual semantic features define its meaning through:

(1) Summation of motor-perceptual features

(2) Explicit combination of other symbols

Symbols like phonemes refer to concepts. In this case the concept

- A phone is a symbol referring to a phoneme. Allophones have different sensory representations but the same semantic representation. A phoneme is the concept distinguished by the semantic representation.

- Symbols refer to referents

- They exist at multiple levels of processing; at a low level, we have sensory representations, at a high level are representations of abstract concepts that can only remotely be tied back to sensory stimuli.

- In ESNs (and brains) symbols are patterns of neural activity

- The referents are pieces of contents / sets or classes of things (sets of sensory inputs, concepts, objects, and other relevant entities)

- Their relationship is representations

RNNs represent concepts within their dynamics. These concepts can be ... (discrete sets entities). The representation of a concept we shall refer to as symbols (that may be symbol structures; composed of other symbols). The concept corresponding to a symbol is its referent. This view coincides with a PSS perspective and makes several assumptions. Nevertheless, consider an example; concepts like phonemes (stimuli sets) in an ESN. Theories of the nature of symbols have been proposed; attractors, conceptors.

Traditional cognitive science frames the nature of cognitive representations as fundamentally symbolic. "Symbols lie at the root of intelligent action"; The hypothesis of \citet{newell2007computer} that humans, computers, and any other system capable of intelligence must be a physical symbol system, one that performs operations on symbols, has been widely adopted. Symbols are entities that represent other entities. Human speech may exemplify the symbolic nature of our cognitive representations.\label{sec:phonemes} Phonemes are the mental representations of phones, the smallest still distinguishable units of speech sound. For example, the phoneme /t/ is a symbol of the English-proficient mind and the mental representation of the phones [t\super h], [t], and [\textipa{R}]\footnote{Phonemes are broadly transcribed using a slanted bracket. Phones are narrowly transcribed by a square bracket. Phonic transcriptions will be written using the International Phonetic Alphabet (IPA).}, each of which encompasses a subset of speech sounds. The phones that a phoneme represents are its \textit{allophones}. Communication through speech is possible, among many other factors, for the speech representations of the communicators coincide; The representations of English speakers generally feature the same 44 phonemes representing the same set of phones with some variability across dialects. Note that phonemes are language specific; Every language uses a different set of symbols to represent its particular speech sounds. Thus, cognition seems to rely on symbolic representations.

Cognitive neuroscience attempts to explain how these cognitive symbols arise in sub-symbolic brains. From a sub-symbolic perspective, the brain is a network of neurons. When this dynamical system is stimulated, for example, by sensory inputs, patterns of neural firing result. These firing patterns may be considered symbols if they are associated with a specific entity. Symbol acquisition, then, is the process of reinforcing these firing patterns (through Hebbian learning) face to the repeated exposure to instances of the referenced entity. For example, when a child is first exposed to speech, every sound signal will trigger an arbitrary firing pattern. As the child repeatedly hears the phones [t\super h], [t], and [\textipa{R}] in similar linguistic contexts, a single pattern of neural firing becomes associated with them. This pattern corresponds to the cognitive symbol, the phoneme, /t/. After symbol acquisition, the pattern is no longer tied to any specific instance but represents a whole subset among the possible speech sounds and sensory experiences in general). Notably, children learn to distinguish phonemes without external supervision; No explicit instructions are needed let alone possible since only later in development do phonemes become the building blocks of words and enable the understanding and production of language \citet{maye2000learning}. So, symbols correspond to sub-symbolic firing patterns acquired through unsupervised learning.

Beyond individual stimuli, \textit{classical symbols} are a much studied class of representational content. Symbols are entities that refer other entities, or referents. The classical definition of symbols by semiotics \citep{peirce1902logic} and later, computer science and AI \citet{newell2007computer} frames them as arbitrary in form; the physical form of the utterance [dog] bears no relation to the symbol's semantics, the referenced (set of) dog(s). Given their ubiquity in spoken, written, and visual language, their representation has been subject of study. \citet{borghesani2017neuro} provides a summary: of that study.

In particular, the TIMIT acoustic

\textbf{Study of stimuli representations in AI, XAI (without conceptors)}

Representation plays a central role in understanding the cognition of neural networks.

In AI, however, few methods for identifying how ANNs represent stimuli have been developed. A large group is available in the area of Explainable AI (XAI)...

The nature of neural representation is commonly thought to

## Study of representation of symbols in AI (using conceptors)

- Symbols:

There seem to be discrete, distinguishable units within the representational space.

The field of artificial intelligence (AI) has sought to bridge between its symbolic and sub-symbolic representations. One such bridge is being built between Echo State Networks (ESNs) and Conceptors.

The field of Neuro-Symbolic AI that attempts to bridge between AI's neural-dynamic and symbolic-logic perspectives. Previously referred to Numeric-dynamic and Logic-symbolic. Both AI and cognitive neuroscience are interested in bridging the dynamic-neural level There seem to be two realities to neural networks; one reality is the continuous neural activity and the discrete relevant entities of our realities using \textit{symbols} \--- units that represent entities other than themselves.

- Recognition

a. It has been found how these symbols can be identified given the set of referents (a symbol-referent link).

b. Moreover, there are methods for recognizing when a symbol is present in an RNN's dynamics ( method for symbol recognition).

With these ingredients a. and b. the symbolic perspective on ESN activity can be used for recognition and classification of stimuli using the activity in the ESN(...).

-> classification of new examples as in japanese speaker recognition

compare this to baseline on small set with perfect classes.

From this we have symbols. But symbols can also reference abstract entities. So what about symbols that contain symbols. hierarchical clustering.

- Conceptors: Conceptors are a means for a user to represent, characterize, store, recognize and relate these symbols. Conceptors were proposed to bridge between the numeric-dynamic- perspectives.

- Physical symbol system.

there to be recurring patterns, symbols, discrete units, with specific meanings for the realized behavior or thought or application domain. Much like the brain that represents objects and concepts as symbols. These units are called symbols. They are related (juxtapposed, merged [creativity], linked) to form cognition. We thus conceive a neural network as a physical symbol system with the following characteristics:

- it contains symbols: physical patterns

- these are recombined to form expressions (symbol structures) that are specific to the domain of application, loosely correspond to "thoughts", that is combinations of multiple symbols

- processes that entail the rules of change

- I define symbols as non-arbitrary forms. We will \textit{not} adopt the semiotic or PSS definition of symbols as arbitrary units whose meaning is assigned by some external mechanism. Instead, we adopt a \textit{dynamic symbol} definition of symbols as intrinsically meaningful. In the current context, symbols are patterns within neural activity whose meaning (i.e., the set of entities they refer to, the referents) lies within their form (captured in the shape of conceptors) and the connections to its sub-symbols.

- Dynamic symbols.

- Recognition or classification tasks refer to matching a referent with the pattern of neural acitivity (i.e. a symbol) that refers to it.

- Artificial neural networks are applied to domains like language where symbols occur.

- In the continuous dynamics of RNNs, dinstinct symbols arise. Quite surprisingly, some discrete the continuous neural activity in RNNs \textit{represents} discrete concepts as they for instance appear in language, logics, and other PSSs.

Conceptors previously: Given X, what are the corresponding patterns of neural firing? In this paper I ask the opposite: Given patterns of neural firing, what are the symbols???

The field of artificial intelligence (AI) has sought to bridge between its symbolic and sub-symbolic representations. One such bridge is being built between Echo State Networks (ESNs) that are, like the human brain, networks of sub-symbolic units (neurons) and \textit{conceptors} that offer a symbolic representation of the sub-symbolic firing patterns. I will proceed by introducing ESNs and then conceptors taking inspiration from \citet{jaeger2014controlling}. On the sub-symbolic end, ESNs are a type of recurrent neural network. As such, they implement a dynamical system; Their internal states are a function of time. This time-extended activity relies on cyclic connections to introduce memory into the system and enable the processing of time series like speech. However, ESNs distinguish themselves from other RNNs by their large size (many internal neurons), low density (low connections-to-neurons ratio), and random and untrained internal- and input weights \citep{yildiz2012re}. Given these features, ESNs can have a capacity for performing a high-dimensional non-linear expansion of their input signals, that, under the right circumstances, uniquely \textit{echo} of these inputs, paralleling how a brain may transform sensory inputs into characteristic neural firing patterns. Figure \ref{fig:ESN} depicts a typical ESN as it is driven by an input sequence $u$, possibly a speech sample, and elicits a high-dimensional response $x$, the sequence of internal states of the ESN's reservoir. An output layer is sometimes added to ESNs for (re)productive purposes but omitted here since only the internal states of the network will be of interest for the current study as will become apparent shortly.

\begin{figure}

\centering

\includegraphics[width=5cm]{img/esn.pdf}

\caption{Sketch of an ESN driven with an input signal. State vector $x$ is a high-dimensional response to the input signal $u$.}

\label{fig:ESN}

\end{figure}

On the symbolic end, conceptors offer a symbolic representation of the sub-symbolic activity patterns within neural networks like ESNs. A conceptor is a positive semi-definite matrix that can be derived from any set of network states like the state sequence $x$. Their mathematical properties (Section \ref{sec:conceptors}) allow conceptors to, i.e., recognize, control, compare, combine, and logically evaluate patterns of network activities. Thereby, using conceptors, groups of ESN activities are considered as entities \textbf{[(possibly discrete and semantic)]} similar to how neural firing patterns of the brain were considered symbols by cognitive neuroscience. Thus, conceptors can be used as lenses on the dynamics of ESNs; lenses that open up a variety of symbolic mechanisms.

A common use-case of conceptors with ESNs has been supervised time series classification \citep{jaeger2014controlling}. An instance of this method will be outlined in Experiment \ref{sec:exp1}, but the guiding idea is the following. Under the echo state property (ESP), an ESN responds uniquely to its inputs; i.e., there is a functional relationship between inputs and responses \textbf{[Why is it not bijective?]} \textbf{[Can I leave this out?]}. More concretely, with this functional relationship between inputs and responses, the responses can be seen as high-dimensional embedding of the input. Hence, any label assigned to a reservoir response is taken to equally apply to the input sequence that induced that response. To train the classifier, one captures the reservoir responses of each class using a conceptor. To classify, the new response is assigned the class whose conceptor indicates the highest similarity \textbf{[or correlation?]} to the previously seen responses of the class. \citet{jaeger2007optimization} demonstrated this method in speaker recognition on the \textit{Japanese Vowels} dataset to a perfect accuracy, a result that promisingly exceeded several previous attempts [cite...]. [Describe the advantages of this classification method.] Building on this experiment, \citet{vlegels2022multivariate} showed that competitive accuracies in the classification of non-stationary time-series may be achieved using several variations of this methods. Moreover, \citet{Chat2022} adapted the method to time-series \textit{recognition}. The difference between classification and recognition is that the former estimate the class of pre-segmented signals (one label per segment), whereas the latter estimates the class present within unsegmented signals (multiple labels per signal over time) \citep{Perdigao11}. Concretely, \citet{Chat2022} demonstrated their method on the recognition of phonemes on the TIMIT Acoustic-Phonetic Continuous Speech Corpus (TIMIT), the dataset also used in the current study. Concluding, in the supervised tasks of time series classification and recognition, the classes of interest are given, but they may not always be.

Unsupervised learning tasks typically involve analyzing the structure of data without given classes or labels. For example, one may seek to identify important clusters (groups), hierarchies of clusters, axes, or relationships among a set of data points. In the following cases, conceptors were used to identify initially unknown structures or relationships among neural network state clouds for explainability and data analysis. \citet{bricman\_nested\_nodate} used conceptors for neural network explainability. The developed technique, Nested State Cloud, extracts a knowledge graph from the latent space encodings of a set of 100 symbols (e.g., "apple"). Each symbol encompassed multiple instances, contexts in which they appear (e.g., "The apple fell."), all of which had different embeddings. Conceptors served as a compact representation of each symbol's embeddings and allowed them to be semantically related in the graph. Similarly, \citet{MossakowskiDG19} used conceptors to represent and then organize classes of network states. This time, the symbols (i.e., classes) were Japanese speakers, and the network states were an ESN's responses to their speech samples. Using the fuzzy generalization of the Löwner ordering as a dissimilarity function on the speaker-specific conceptors, the symbols were organized in a tree by hierarchical agglomerative clustering. The authors mention the potential application in classification, but whether the resulting hierarchy truly coincided with any features of speakers like dialectal groups remained unknown. [One more example] These studies analyze structures between classes of reservoir states, but this means that they, like the supervised classification algorithm, start from pre-established symbols. I shall refer to these as \textit{above-symbol structures}, for they are at a level of abstraction above that of symbols.

## Motivation

- Interpretability: Understand the presence of classes within the blackbox (not only based on outputs, accuracy). In most applications, we require networks to adopt certain external symbols (classes cat, dog, sounds [a], [b], letters "a", "b). They are man-made structures of the network's inputs and outputs. The network learns a function that relate these symbols to each other. To some degree its internal representation will feature these symbols. But what symbols does the network come up with to do so? \textit{How} does it fulfill its design purpose, i.e. minimize its cost function? Its internal symbols might be very different to the symbols of in and outputs! .... Explainability.

- Global Explainability

- Pragmatics: BPTT, Unavailability of classes may indicate need for further training

- Robustness and Adversarial Attacks: Understanding the geometry can also be crucial for identifying vulnerabilities to adversarial attacks. If representations of different classes are too close geometrically, small perturbations can lead to misclassifications.

- Neuroscience Implications: For researchers in neuroscience, understanding the representational geometry in artificial networks can provide hypotheses or models for studying the brain. It can help in understanding how neurons and neural circuits represent and categorize sensory inputs.

- Transfer learning, generalizability, recombination: Move to more general and creative AI systems via the tools of conceptors

- Higher cognitive functions: Computational creativity, in particular, may require the unsupervised generation of new symbols. According to \citet{boden2004creative}, the most advanced type of creativity involves transforming an existing conceptual space. For example, some groundbreaking scientific discoveries may not be reachable via the mere exploration of existing theories (the current conceptual space) but may require a shift in the problem representation. Similarly, Douglas Hofstadter models creativity as an ability to form and alter symbols. Thus, without the capability to discover new symbolic representations, the creative agent might find itself quite limited, capable of optimizing , but unable to redefine the problem space itself. Regarding analogy-making, a core feature of human creativity, Hofstadter's Copycat model thereof relies on the fluidity of symbolic representations. Therefore, potential improvements in the available symbols could enhance analogical reasoning potential and, by extension, creative potential.

When using symbolic(-conceptor)-based mechanisms to implement cognitive functions, the representational choice of how to split and group a continuous domain into a set of symbolic items is crucial to the performance. It influences any higher-level processes. For example, items with similar representations are easier to be abstracted or generalized from (Tenenbaum, since less synaptic weights are needed to instantiate a reference to both items). Choosing a proper set of symbols is thus crucial... What about cases where symbols are not available or the provided are not enough to solve a task? A child figures out which symbols make sense to operate in this messy world without supervision. I am looking for a computational mechanism to identify symbols (meaningful units) within inputs. What symbols can be used to represent some data? ESNs make reflections of their inputs, so patterns among their reflections can inform about patterns of the inputs.

- If the clusters align with real categories, there are grounds to interpret them as symbols.

This paper aims to identify these symbols within the unlabeled dynamics of Echo State Networks.

- Old:

A significant challenge and the main aim of this study lies in unsupervised symbol identification \footnote{Symbol identification is distinct from acquisition since no learning is involved.}. The motivation for analyzing such \textit{below-symbol structures} is the move toward more autonomous cognitive agency with applications in, for instance, computational creativity. Concretely, in the context of ESNs, the question arises of how to determine which network state instances should be included within each conceptor, that is, which patterns of network activity should be treated as the same symbol. It is relevant for several reasons. Firstly, unsupervised symbol identification is a step toward more general cognitive agency. Representations are often fed with a priori information on their task to improve performance; for instance, the dataset for training a speech recognition model may feature additional phonemic transcriptions. These will be used to form the model's task representation. However, there often is little basis to presume that, for any given task, the known representation is optimal. Moreover, for yet unsolved tasks, the inability to autonomously identify relevant symbols may form a ceiling to machine cognition. This need for autonomy is analogous to the earlier example of how children need to learn to distinguish phonemes due to, or rather in spite of, the lack of explicit instruction or supervision. [Find good representations of this infinite, continuous, and dynamic environment. Thus, symbols may also be seen as constant, discrete representations of a subspace of the continuous, dynamic system that is the world.]

The identification of both below- and above-symbol symbol structures resembles classical clustering tasks. Clustering refers to the unsupervised task of organizing a data set into groups, or \textit{clusters}, such that instances within the same cluster are more similar to each other than to those of other clusters. This similarity may be determined by a dissimilarity measure like the Euclidean distance. Clustering could play a significant role in finding below-symbol structures relevant to unsupervised symbol identification. By feeding the states of a neural network into a clustering algorithm, they can be grouped into clusters, each of which can then be treated as a symbol with the cluster members being the representations of that symbol. The analogy holds, for similarity is expected among states of the same symbol and dissimilarity between states of different symbols. Conceptors may serve to represent and evaluate the clusters. Conceptors can compactly represent and compare sets of network states and thereby act as centroids, i.e., exemplar-based representations of the symbols \citet{bricman\_nested\_nodate}. Hence, the process of unsupervised symbol identification will be seen as a form of clustering, where each cluster corresponds to a symbol and is represented by a conceptor. Similarly, clustering algorithms seem equally applicable to the analysis of above-symbol structures by pre-grouping the network states by symbol according to a priori memberships.

## What is unknown?

RNNs (like ESNs) exhibit neural dynamics, but their representational geometry of stimuli has not been studied. Still largely blackboxes.

- The following is \textbf{unknowns}:

a. which set of symbols a RNN uses to represent a set of unclassified/unstructured stimuli.

b. the representational geometry is unknown. Which is the symbolic structure of the dynamics of an ESN at any given processing level?

c. which abstract symbols might be present in the Network above the pre-established symbols. Given that symbols can be composed of other symbols. There are various possible levels of symbolic abstraction to consider when searching for the symbols in a network. Which symbols are present in the neural dynamics at a given processing level?

- The representational geometry at a given processing level. We see a cognitive agent as multi-level structure. These levels are not necessary hierarchically ordered, but follow some configuration that we consider constant for the sake of this study that does not involve learning. The perceptual system, widely studied in both artificial and biological systems, is considered to consist, for instance, of a cascade or hierarchy of processing levels. It is unknown how some stimuli are represented and how these representations are relate. [litterature showing the necessity of it finding the representational geometry]. In AI this has implications for interpretability and perhaps the development of more complex cognitive systems (that represent co-dependent achievements).

**Alternative RQs**

\textbf{How to identify distinct and meaningful clusters within the an RNN's representation of stimuli?}

- How to identify and interpret clusters within an RNN's representation of stimuli?

- Can a set of relevant symbols be extracted from the neural dynamics of RNNs?

- Can a relevant symbolic representation be extracted from unstructured neural dynamics?

- What geometry underlies a neural representation of a set of stimuli?

- Old:

Concluding, this study explores conceptor-based, unsupervised, options for identifying symbols within the sub-symbolic dynamics of ESNs (the responses to input time series). Therefore, it investigates the question of whether meaningful clusters and hierarchies can be identified among ESN state sequences. These structures are considered meaningful if they coincide with domain-relevant features of the corresponding inputs.

## Proposed Solution

- Study the representational geometry of phonemes inside ESNs

We set up a network such that we can expect the representational geometry / its states to form clusters in alignment with classes of the stimuli.

Focus on \textit{segments}

Then, Kmeans and HAC clustering

- Evaluation is made by implicitly drawing the parallel to humans; symbolic classes

\textbf{Phonemes, TIMIT}

- Emphasize how your research builds upon, differs from, or addresses gaps in these previous studies.

- Quality evaluation

- the resulting conceptor matrix acts as a recognizer for that symbols. 3 methods. the one with conceptors wins.

- Old text: An unsupervised method for identifying relevant symbols within a ESN activity. This requires definitions. A symbol is an activity pattern captured using a conceptor. The domain of ESN activity is a \textit{set} of ESN states. Their order is not considered as it falls away using conceptors. \textit{Relevance} is established by comparing to the intrinsic and extrinsic (human-made) features of the symbols of domain of application. Distinctiveness is assumed as a symbol-intrinsic property. There is no natural law to my knowledge that maintains this principles since symbols are human/brain-made structures. However, it seems to be assumable that instances of the symbol are similar to each other (cohesion) and distinct from the instances of other symbols (separation). First, when learning occurs, weight resources in a network are limited. It becomes necessary to differentiate between symbols. If instances of one symbol were too similar to instances of another, it would be computationally costly and often impractical for the cognitive system to distinguish between them.

Second, share perceptual physical features among the associated referents (especiali stimuli).

Third, by analogy to the human categories / semiotics. [... chat gpt] It also presupposes non-overlapping symbols, which is made for we consider only a single processing level. Definition of symbols as patterns or regularities/constants.

We consider a symbol a pattern consistently arises when its referents are active (active means currently presented or another symbol occurring) but rarely arises in presence to other stimuli.

I take inspiration from two classical hard clustering algorithms, K-means and hierarchical agglomerative clustering (HAC). Hard clustering assigns each data point to one, and only one, cluster, enforcing a clear separation between clusters. In contrast, soft clustering, like that of \citet{bricman\_nested\_nodate}, assigns data points a level of membership for each cluster, allowing for overlap between clusters. I assumed, in the spirit of traditional symbolic representation theory, hard clustering to be the better fit for the endeavor of automated symbol identification using symbols to represent discrete, separated entities instead of multiple entities to varying degrees. ESN dynamics will be hard-clustered by adapting K-means and HAC, two well-known hard clustering algorithms with guaranteed convergence. K-means partitions data into a number of non-overlapping clusters and will be used to group below-symbol instances into symbolic groups. Importantly, traditional K-means can cluster $N$ data points within $O(N)$ enabling its use on larger data sets. Meanwhile, HAC returns a hierarchy of clusters and will be used to analyze above-symbol structures among pre-existing symbols. HAC requires $O(N^2)$ with optimal implementation. Both algorithms have previously been used for time series clustering by [...]. However, while ESN state sequences are time series, it is unclear whether the previous methods are suited to the clustering of high-dimensional non-linear dynamics. Moreover, previous time series clustering algorithms have scaled badly with long, potentially infinite, time series, were often offline (fail to adapt as new data points), or struggled to handle variable length time series, a limitation mediocrely solved using Dynamic Time Warping. These shortcomings were equally menacing time series \textit{classification} algorithms but were resolved by use of conceptors; the similarity in methods gives hope to similarly resolve them in conceptor-based time series \textit{clustering}.

The developed methods will be demonstrated on phonemic speech recordings. As \hyperref[sec:phonemes]{mentioned}, Phonetics \--- the branch of linguistics concerned with speech \--- groups speech sounds into phones (sound units) and phones into phonemes (symbols); yet, further above-symbol organizations exist. For example, Figure \ref{fig:TIMITaxonomy} depicts a possible taxonomy of the phonemes present in the TIMIT corpus which contains clusters like vowels and nasals\footnote{The TIMIT labels, also shown in Figure \ref{fig:TIMITaxonomy} and listed in Table \ref{tab:dataset}, are from IPA, although not written in Greek letters. The translation keys are given in the documentation.}.

\begin{figure}

\centering

\includegraphics[width=7.47cm]{img/TIMITaxonomy.png}

\caption{A taxonomy of all phones present in TIMIT \citet{Pfei2011}}

\label{fig:TIMITaxonomy}

\end{figure}

The availability of these phonetic systems makes phoneme recordings a good domain for evaluating clustering algorithms. On the one hand, these phonemic classes and taxonomies can be used as ground truth clusterings to determine the performance of the developed clustering algorithms. [maybe connect the following sentence with analogy to child] On the other hand, the results from this study may provide an empirical verification or falsification of phonetic theory; whereas the groups and taxonomies of phones were largely developed based on their articulatory features, the following clustering Experiments will investigate whether these groups can be found on a purely acoustic basis. [Phonemes clustering literature on TIMIT (reference to methods to avoid redundancy)] [transition]

For clustering ESN responses to phoneme recordings, the ESN must be suited for the task. The following assumption affects the experimental design; Given an ESN, the more accurately a conceptor-based classifier can distinguish activity groups, the better these classes can be retrieved by an unsupervised clustering algorithm. The classification and clustering performances seem correlated because, for many hyperparameters like the aperture and spectral radius, a sweet spot seems to exist, where the ESN's signal-to-noise ratio is maximized and where the network is neither over- nor under-excited, that can be determined independent of the application in clustering or classification. Thus, I assume that tailoring the ESN to the TIMIT data and classification task will improve clustering on the same data. The hyperparameters were optimized in Experiment 1 on time series classification. Moreover, performing the classification task in Experiment 1 has the positive side-effect of demonstrating the application of conceptors to the phoneme classification on the TIMIT dataset. The following classification method builds on that of \citet{jaeger2014controlling}. To my knowledge, the currently highest accuracy of 78.4\% for phoneme classification on TIMIT's test set was achieved using fixed-sized kernel logistic regression [\citet{karsmakers2007fixed} mentioned in \citet{Perdigao11}].

**- Old part from methods about solution:**

Reiterating, a symbol was defined as a pattern of network (ESN) activity that represents some entity. Such activity patterns, or symbols, can be represented, identified, compared, or induced using conceptors. In the previous Experiment, each positive conceptor represented the ESN activity patterns characteristic for a set of states. phoneme (the symbol) that, in turn, was the ESN. From these clusters, conceptors can be derived that reflect a new set of symbols.

To compute conceptor for any symbol, a set of network states (a state cloud) in which the pattern is present is needed. These states are responses to signal instances of an abstract class whose defining concept (the entity) is represented by the symbol.

When deriving a conceptor for any symbol from a set of states, it is preferable for that symbol to be the main commonality between the states. Geometrically, this translates to having lots of variance along principal components and low variance

In the following two experiments, the generalized centroid-based hard clustering and HAC algorithms were adapted to cluster the ESN responses to the phonemic speech recordings. The general theme of both experiments was to (A) embed time series (the pre-processed phoneme recordings) as reverberations of an ESN (the response), (B) use conceptors or other centroids to represent these ESN reverberations and (C) group them into a set of meaningful clusters. Figure \ref{fig:pipeline} depicts the steps from the original signals to ESN states to conceptors to clusters. Although the conceptors were subjected to the clustering algorithms, the tight relationship between signals, ESN responses, and conceptors presumably allowed the resulting clusters to generalize from conceptors to the other two representations. This generalization was exploited during cluster evaluation when comparing the found clusters against phonetic groups of the TIMIT speech samples.

\begin{figure\*}[htbp]

\centering

\includegraphics[width=\textwidth]{img/pipeline.pdf}

\caption{The steps from the MFC space of pre-processed speech signals to ESN state space by driving the ESN (A), followed by a transfer to conceptor space by computing conceptors (B), and finally the partitioning of the data into clusters (C). Note, while Experiment 2 aims to find partitions, Experiment 3 identifies hierarchies; thus, the right illustration would differ for Experiment 3. [Double check representation is truthful of apertures.]}

\label{fig:pipeline}

\end{figure\*}

## Additional motivations

A final motivation is to extend the use cases of conceptors to time series clustering. Reiterating, conceptor-based time series classification was achieved by classifying an ESN's responses to the time series. Similarly, clusters found among ESN dynamics may coincide with relevant clusters among the time series that underlie these dynamics. This relationship will prove useful in extrinsically evaluating the clusters. Moreover, the formalism of conceptors would thus be extended to time series clustering.

## Study limitations

1. What I'm not studying.

- Representational topology

- Time-continuous

- Symbols like in PSS

- The linguistic meaning of phonemes beyond their "sensory meaning" -> only perceptual features

- Assumption of corresponding clusters with same K

- Some generalizations are left to future research. Now we considered independent RNN runs (in response to pre-segmented speech). At each processing level, RNN activity can be seen as a flow from symbols (or symbol structures). Each symbol appears and disappears like an event. Identifying symbols in this perspective requires distinguishing when symbols and space (what pattern is activity). actually as events extended in space (its ) and time (the information). Complete unsupervised symbol extraction would thus require

- Segmentation

- Association of segments

- Autonomous applications (e.g., the task of identifying symbols in unsegmented speech)...

Clustering only in space, not in time

- Segmentation, although exploration of topic is appended

# Discussion

\section{Discussion}\label{sec:discussion}

Concluding, this study aimed to identify distinct and meaningful clusters within an ESN's representation of phonemic utterances.

\paragraph{Exp 1: Phoneme Classification} Experiment 1 demonstrated phoneme classification using conceptors. [Comparison to phoneme classification benchmark...] [Comparison to Jaeger's experiments...]

\paragraph{Aperture normalization} The methods of conceptor based time series classification used in Experiment 1 were largely inspired by \citep{jaeger2014controlling} with the exception of the aperture adaptation. The first deviation here was to adapt explicitly only the apertures of the positive conceptors $C^+$......

\textbf{[normalizing/optimizing the aperture of the negative conceptors does not help]}

It was made to avoid to give conceptors with larger singular values an advantage during classification. To understand this effect, consider the example of a $2\times 2$ conceptor matrix $A$. \textbf{[...]} To avoid such classification errors, the apertures were iteratively adapted until approaching a target singular value sum using the following procedure. Indeed this normalization further improved classification.

It must be remarked that, by that process, the final apertures will deviate from the one originally chosen aperture via the $\nabla$-criterion. Figure \ref{fig:normalization} depicts this normalization process for the classification experiment.

\paragraph{Dataset} Without normalization in time, performance decreases. How to handle different utterance speeds? Add leakiness to ESN to integrate time scales? Non-stationary signals

\paragraph{Z-Condition} Worse due to overfitting: Exponential reason (AI2). The size of $z$ and the associated computational costs are larger than for the former method, and it tends to lead to better $training$ classification accuracy, but much worse test accuracy indicating overfitting.

\paragraph{Exp 2: Advantages of clustering method} Conceptors do not [...]. As constant-sized matrices, they do not scale with the number of time steps or modeled states. This is a leap for analyzing time series since previous algorithms typically scale poorly with longer samples; for example, computing the distance between time series in dynamic time warping (DTW) scales quadratically with sample length dynamic. Notwithstanding, the quality of conceptors depends on the number of states used for their computation [...]. [Comparing responses of different lengths would require workarounds like Dynamic Time Warping (DTW).]

Third, ESNs deal with time-extended inputs causing several difficulties for preexisting clustering algorithms. For example, the euclidian distance can only be used to compare samples of different durations (for clustering algorithms that require distance functions) and have time-.

[Time complexity of clustering algorithms: especially Experiment 2]

[Mention esp might not be satisfied]

[Effects of spectral radius on ESP]

It should be stressed that, in the current work, \textit{ESN activities} are classified and clustered using conceptors; Not speech signals, but an ESN's responses to these signals are used as input to classification and clustering algorithms.

\paragraph{Applications} BPTT: To evaluate how well a classifier can distinguish between a set of classes, one may resort to performance metrics or investigate the degree of certainty in its predictions. However, such an evaluation relies only on the outputs and disregards the internal states that led to the predictions. For complex models or whole pipelines, it may be of interest to find groups or clusters present in the RNN's activations or within a subset thereof, e.g., corresponding to only one part of a pipeline. This study ought to be an exploratory step toward answering the following questions: How clustered are the activities within a NN? When, during training, do class differences in network activities arise?

Nonetheless, a phonetic interpretation about the speech signals will be made which, however, requires an assumption. Commonly, the echo state property is used to ensure a functional relationship between driver signal (phoneme recording) and RNN response \citet{yildiz2012re}. However, for the echo state to be reached requires left-infinite or at least considerably long input sequences, for the starting state to be washed out. In our case, inputs will be downsampled to a length of 10 rendering the claim of the ESP impossible based on current theory. The analyzed states will most likely still correspond to the network's initial transient period. Thus, functionality between input and response cannot be guaranteed but only assumed. Hence, we shall nonetheless use the results of clustering and classification of ESN activities from the transient period for an interpretation about the input signals.

The current algorithm is not guaranteed to converge since it is an adaptation of K-Means that uses a non-Bregman divergence function (more on this in the comment under Experiment 2).

Adv of conceptor-based clustering to classical: Moreover, they compress one or more network states into a mathematical object (a matrix). These

Another advantage, This hurdle can m and one advantage applying conceptors to time series is, as will be further argued, that they capture their information in a static mathematical compressing the temporal dimension and

Assumptions of generalized clustering: \citet{banerjee2005clustering} bregman divergence, zero dist function when coinciding points and centroids.

To summarize, the three main motivations of the following clustering methods are to provide additional.

\begin{enumerate}

\item Explainability: Clustering RNN dynamics could help explain what groups of activity are present in the network. For example, both algorithms could be helpful tools for interpreting BPTT-trained RNNs. For example, Experiment 2 could show how classes become more distinct or numerous within the RNN (form activity clusters) while training.

\item Data analysis: Clustering time series with traditional clustering algorithms can be challenging. For instance, due to varying input lengths, methods for embedding time-series data are rather limited (e.g., latent-space encodings). This paper shows that the clusters present in RNN dynamics can be tied back to clusters within the inputs used to produce these dynamics.

\item Conceptor Classification: To the end of successful clustering, a phoneme classification experiment (Experiment 1) will be performed. In this experiment, the current methods for conceptor-based classification are applied and extended.

\paragraph{Conceptors as state Embeddings}

The temporal and spatial compression realized by conceptors, allows them to be used as embeddings of ESN responses. Unifying responses in conceptor-space, independently of their length and maintaining only their most important information, allows state sequences to be related and compared. [Examples of the use of conceptors as embeddings]

\end{enumerate}

Classically, these algorithms are used on data in Euclidian space and their components are well-defined for that space (mean, Euclidian distance, manhattan distance, etc.). However, the activities within the ESN are subjected to non-linear transformations like $tanh$ at each time step as is the case for most neural networks. I hypothesize that Euclidian distance functions may not be well-suited to capturing the distances between network states. Moving to conceptor space will be an attempt to improve the clustering of ESN activities through a more adequate means for cluster representation, comparison, and assignment.

Attempted explanation of swallowing problem

Finally, one may object that, for unsupervised symbol acquisition, the use of hard clustering is very artificial and something akin to Hebbian learning should be preferred, something that resembles how the brain acquires symbols. However, often in computing, and especially reservoir computing for material constraints, learning may not be possible. Yet, one may want to identify the symbols prevalent in the "artificial mind".

... We attempt to solve this mismatch by mapping the network states to be clustered to conceptors since conceptors offer several tools for comparing and analyzing network states. By moving to conceptor space, the following adaptations were made.

\begin{enumerate}

\item plus becomes or: Firstly, the operations of Euclidian vector space were changed.

\item division become aperture adaptation

\item Distance becomes the one proposed by \citet{jaeger2014controlling} OR...

\end{enumerate}

Problems. Distance metric is rather unclear, lack of semantics.

Average linkage: Linkage algorithms: Pros and Cons slide of video

Categories may be due to coarticulation / correlations

Importantly, this is Not a Bregman divergence, because cosine similarity does not fulfill the triangle inequality. Therefore, this algorithm is not guaranteed to converge.