Univerzita Karlova

Filozofická fakulta

Ústav anglického jazyka a didaktiky

DIPLOMOVÁ PRÁCE

Bc. Kateřina Kynčlová

Perception of Silent Speech in L1 Users of Sign and Spoken Languages

Percepce tiché řeči u rodilých uživatelů znakového a mluveného jazyka

Praha, 2024

vedoucí práce: Kateřina Chládková, M.A., Ph.D.

PODĚKOVÁNÍ

Ráda bych touto cestou poděkovala všem, kteří přispěli k realizaci této diplomové práce. Zejména srdečně děkuji vedoucí mé diplomové práce, Dr. Kateřině Chládkové, jejíž odborné vedení, obdivuhodná ochota a podpora mi pomohly překonat mnohé výzvy. Dále děkuji všem respondentům za jejich účast na této studii, Haně Peštové za tlumočení do ČZJ, Viktorovi Čermákovi a Jakubovi Kreuzigerovi za tvorbu informačních materiálů v ČZJ, Natálii Kikoťové za nahrání video stimulů použitých v této studii, členům výzkumného týmu SPEAKin Lab za jejich účast na pilotním testování a Psychologickému ústavu Akademie věd ČR za poskytnutí aparátu a prostoru k testování. V neposlední řadě děkuji rodině, přátelům a kolegům, kteří mi byli oporou po celou dobu studia.

PROHLÁŠENÍ

Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 12.8.2024

Kateřina Kynčlová

ABSTRACT

The present study focused on the neural processes underlying the perception of silent visual speech and aimed to investigate whether silent articulatory cues could facilitate statistical learning, a mechanism shown to aid the perception and segmentation of continuous speech. Additionally, this thesis investigated whether this process could be affected by the participants' primary mode of communication. The neural activity of adult participants with normal and impaired hearing was measured with EEG as they were exposed to two silently mouthed syllable streams. The structured stream consisted of four repeating trisyllabic pseudo-words whose boundaries were cued merely by the transitional probabilities between the syllables. On the other hand, the random stream consisted of twelve syllables in pseudo-random order and did not contain any covert statistical structure. Inter-trial phase coherence at a syllable rate (3.3 Hz) and a word rate (1.1 Hz) was computed to assess the phase synchronisation of the recorded neural activity and the silent speech stimulus. A phase-locked activity at the word rate would indicate that the covert statistical words were detected as a result of statistical learning. The participants also took a post-exposure forced-choice task to assess their explicit knowledge of the learnt structures and thereby assess the effects of statistical learning behaviourally, and a lip-reading test to assess their ability to decode articulatory cues in a known language. The EEG results of adults with normal hearing showed that the phase synchronisation at the word rate increased throughout the exposure and was the largest during the exposure to the second block of the structured stream, indicating sensitivity to the statistical regularities in the observed silent speech. Although the EEG results showed effects of statistical learning, the forced-choice task failed to detect such effects, potentially highlighting the shortcomings of using only the behavioural assessment in some studies. Furthermore, the localisation of the increased neural activity indicated that there might be different processing strategies based on the participants' primary mode of communication and native language. The results of this study suggest that statistical learning during speech perception can be successfully measured online regardless of the language modality and could be affected by familiarity with the input, native phonotactics, and the brain's plasticity.

Keywords: *silent speech, statistical learning, speech perception, EEG*

ABSTRAKT

Tato práce se zabývá neurálními procesy, které probíhají při percepci tiché vizuální řeči, s cílem zjistit, zda mohou tiché artikulační pohyby zprostředkovávat statistické učení, což je proces, který prokazatelně pomáhá percepci a segmentaci nepřerušované řeči. Dále se práce zaměřuje na to, zda může být proces statistického učení při percepci tiché řeči ovlivněn primárním způsobem komunikace účastníků. Dospělí s normálním sluchem a s poruchami sluchu byli vystaveni dvěma proudům tiché řeči, přičemž byla měřena jejich neurální aktivita. Strukturovaný proud tiché řeči obsahoval čtyři opakující se tříslabičná pseudo-slova, jejichž hranice bylo možné odhalit pouze na základě pravděpodobností souvýskytu slabik. Naproti tomu náhodný proud obsahoval dvanáct slabik v pseudo-náhodném pořadí a neobsahoval žádnou skrytou statistickou strukturu. Synchronizace nahrané neurální aktivity a řečového stimulu byla vypočítána na frekvenci slabik (3,3 Hz) a slov (1,1 Hz) pomocí tzv. *inter-trial phase coherence*. Neurální aktivita synchronizovaná s frekvencí výskytu skrytých tříslabičných slov by naznačovala, že účastníci byli schopni detekovat hranice těchto slov pomocí statistického učení. Účastníkům byl také administrován hodnotící test, v němž měli ohodnotit familiaritu jednotlivých slov, s cílem otestovat výsledky statistického učení behaviorálně, a test odezírání, který měřil jejich schopnost odezírat slova a věty ve známém jazyce. U účastníků s normálním sluchem se synchronizace neurální aktivity na frekvenci slov zvyšovala během experimentu a dosáhla nejvyšších jednotek ve druhém bloku strukturované podmínky, což naznačuje, že účastníci byli schopni odhalit statistické pravidelnosti v proudu tiché řeči. Přestože EEG výsledky ukazují výsledky statistického učení, hodnotící test statistické učení neodhalil, což může naznačovat nedostatky behaviorálních metod testování použitých v některých studiích. Lokalizace zvýšené neurální aktivity naznačuje možné rozdíly ve strategiích zpracování řeči mezi účastníky s různými způsoby primární komunikace a rodným jazykem. Výsledky této studie naznačují, že statistické učení během percepce řeči může být měřeno online pomocí neurozobrazovacích metod, a to bez ohledu na modalitu daného jazyka, a že výsledky statistického učení mohou být ovlivněny familiaritou s formou inputu, fonotaktikou rodného jazyka a plasticitou mozku.

Keywords: *tichá řeč, statistické učení, percepce řeči, EEG*

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BSL - British sign language
CSL - Czech sign language
CV (syllable) - consonant-vowel (syllable)
EEG - electroencephalography
ERP - event-related potential
fMRI - functional magnetic resonance imaging
ITPC - inter-trial phase coherence
L1 - first language
L2 - second language
P2V - phoneme-to-viseme mapping
p-STS - posterior superior temporal sulcus
R-S (order) - random-structured (order)
S-R (order) - structured-random (order)
SL - statistical learning
TP - transitional probability
VS - visual speech
WLI - word-learning index

TABLE OF CONTENTS

1. INTRODUCTION

Segmenting words from continuous speech not bounded by distinct pauses is a crucial ability for the acquisition of a new language. Previous research explored the mechanisms and cues that aid language learners in segmenting novel continuous auditory speech, allowing them to acquire the language gradually. It has been demonstrated that during auditory speech segmentation, infants use distributional cues (namely transitional probabilities of syllable co-occurrence) and prosodic cues (Saffran et al., 1996). The need for distributional cues, even in the presence of prosodic cues, speaks in favour of the statistical learning theory, which describes a phenomenon believed to aid word segmentation during novel speech perception by discovering and acquiring the underlying repeating patterns in one's environment.

Several studies over the past decades have provided evidence of statistical learning (SL) in speech (see Christiansen, 2019; and Frost, Armstrong and, Christiansen, 2019, for review); however, to our knowledge, these studies have primarily used auditory speech stimuli. Therefore, an intriguing question is raised: does the recorded effect truly reflect the mechanisms aiding speech segmentation in general, or does it primarily record the process of parsing auditory speech using the regular auditory points as anchors? This perspective opens up avenues for further research and invites further exploration into the topic. One such new direction will be investigated in the present thesis.

Speech devoid of all auditory cues, i.e. silent visual speech, is an input often used in research focusing on the deaf and hearing-impaired community (Calvert et al., 1997; Hall et al., 2005, MacSweeney et al., 2000; Park et al., 2016). Silent speech allows researchers to investigate neural adaptation and plasticity in deaf individuals and simultaneously represents a medium to compare the hearing and hearing-impaired populations. Indeed, neural activation during silent speech has been investigated frequently (Bernstein et al., 2011; Bourguignon et al., 2020; Calvert, 1997; Crosse et al., 2015; Hall, 2005; MacSweeney, 2000; Park, 2016). These studies demonstrated that visual-only speech elicits a similar neural response to auditory speech (Calvert, 1997; Hall, 2005; MacSweeney, 2000; Park, 2016). However, researchers predominantly focused on processing of existing languages, providing the observer with critical contextual cues, and on the activated brain areas rather than on the process of perception and segmentation. So far, there has been little discussion about the processes underlying the perception of silent visual speech despite its potential to uncover the adaptation of a deaf individual's brain to the absence of auditory input. Investigating this adaptive process can help address which learning and processing strategies are learned based on our language experience and which are, for lack of a better term, universal. One significant issue in the research into statistical learning during speech perception is the lack of variety in the language modality and the participant's primary mode of communication. Addressing these issues is vital for discrimination the auditory speech perception from speech perception in general.

The primary objective of the present research is to investigate mechanisms underlying the perception of novel silent visual speech, with particular attention paid to segmenting and learning strategies taking place during speech processing. Furthermore, we aim to investigate the effect of primary modality and native language on the processing of visual speech by using 19-channel electroencephalography (EEG) to record the neural activity of hearing and deaf and hard-of-hearing participants during exposure to silent speech streams with and without an underlying statistical structure organised into trisyllabic words. We chose to adapt a design used frequently in studies investigating statistical learning in auditory speech; more specifically, we altered the dataset used by Batterink and Paller in their 2017 study into online statistical learning in adults. Using the altered stimuli, we predict that we will be able to record the effects of statistical learning. We hypothesise that (H1) visual cues will be sufficient to facilitate statistical learning in visual silent speech, that (H2) greater effects of statistical learning will be found in the novel stream of visual speech with the covert statistical structure, and that (H3) the primary modality and language background will affect the participant's neural processing and the ability to learn during the perception of silent visual speech. We believe that the present study can contribute to understanding the general processes underlying speech perception regardless of its modality, and highlight the importance of visual information during speech perception.

2. THEORETICAL BACKGROUND

2.1 Inattentive Learning During Speech Perception

2.1.1 Word Segmentation as a Result of Implicit and Statistical Learning

Word segmentation, i.e. the detection of word boundaries, is typically one of the first tasks a language learner must face during the acquisition of the particular language. Unlike written language, spoken language is rarely bounded by pauses in the speech stream (Cole, Jakimik, and Cooper, 1980) and thus word segmentation must rely on other cues. In an auditory or audiovisual speech stream, word segmentation is aided by prosodic as well as distributional cues. Saffran et al. 1996 demonstrated that even if prosodic cues are available and used by infants during language acquisition, distributional information is still required possibly because infants use prosody to isolate sections of speech upon which they conduct distributional analysis using innate statistical mechanisms designed to extract regularities from natural language input (Morgan & Saffran, 1995; Saffran et al., 1996a, 1996b). The process leading to the individual segmented units is described slightly differently in different schools of thought.

Word segmentation is a phenomenon widely examined in the studies concerned with implicit (IL) and statistical learning (SL) (Perruchet & Pacton, 2006). Though out of the two it was only the latter that initially aimed at studying word segmentation (opposed to IL's focus on rule abstraction), it can be argued that as the scope of both approaches broadened, they both began to explore the same general incidental learning processes (Perruchet & Pacton, 2006). Both implicit and statistical learning refer to an adaptation to the regularities in the environment without intention to learn (Perruchet & Pacton, 2006) and are a key process in information processing (Frost et al., 2019). Implicit learning, an approach older than SL, is focused on the formation of chunks within the stimuli. According to this approach, participants memorise fragments of strings in alignment with their memory capacity. On the other hand, statistical learning favours the interpretation that participants perform statistical computations on the stimuli they are exposed to based on transitional probabilities within the individual elements (Perruchet & Pacton, 2006).

The idea of transitional probabilities being used in learning dates back to Harris (1955) who proposed that linguists can possibly discover individual word units in an unfamiliar language by counting which and how many phones can occur in a sequence (Saffran et al., 1996a). In a pioneering study on this topic, Hayes and Clark (1970) used an artificial stream of synthesised non-linguistic sounds (glides and warbles) which represented phonemes for the exposure task in which participants passively listened to the continuous stream (Saffran 1996a: 609). Their findings demonstrate that the participants were able to successfully segment the continuous speech stream as suggested by their behavioural test results (Hayes & Clark, 1970). Nearly three decades later, Saffran et al.'s seminal study further investigated this learning mechanism on English-like stimuli (Saffran et al., 1996b). Four consonants (p, t, b, d) and three vowels (a, i, u) combined in 12 CV syllables formed an artificial language of four trisyllabic words with the transitional probability between syllables being higher within the word boundary than outside of it. Saffran's dataset and experiment design was later adapted for a number of further studies with both infants (Aslin, Saffran, & Newport, 1998; Choi et al., 2020) and adults (Batterink & Paller, 2015; Batterink & Paller, 2017).

Studies exploring the phenomenon of statistical learning in the auditory domain demonstrated that preverbal infants can learn the transitional probabilities of individual syllables in a continuous stream of pseudospeech after 2 minutes of exposure (Saffran et al., 1996b; Turk-Browne et al., 2009). Although assessment methods of statistical learning for infants and adults differ, studies concerned with statistical learning in adults demonstrated that adults use similar trajectories during word segmentation as infants to segment meaningful units (Saffran et al., 1996b).

The two approaches to incidental learning are viewed differently in existing literature. Some view them as independent processes (Van Der Linden), some see them as synonymous labels for one phenomenon (Conway and Christiansen), and some lean towards combining the two into a more complex model (Perruchet & Pacton, 2006). A model prevalent in statistical learning literature describes the computation of statistical probabilities and chunk formation as successive steps in the word segmentation process with statistical computations being used to form chunks. Another model describes chunks being formed randomly based on the limited attentional capacity. The chunks then evolve as the likeliness of remembering such strings increases or decreases as a result of associative learning principles (Perruchet & Pacton, 2006).

Although there are several approaches to incidental learning and its processes, the basic definition stays clear of contrasting views. By one means or another, the learner is able to segment a continuous stream into meaningful units without an intent or conscious knowledge of the segmented elements. For the purpose of clarity, this study shall proceed to address this phenomenon as statistical learning in line with the secondary sources used, however, it does not disregard any of the proposed models.

2.1.2 Assessment of Word Segmentation and Statistical Learning During a Continuous Stimuli Stream

Over the past several decades, statistical learning was traditionally assessed behaviorally in an exposure-posttest design, with post-exposure tests such as looking paradigms for infants and forced-choice identification tasks for adults, always with just a limited number of trials (Batterink & Paller, 2017). However, recent studies raise the question of validity of such assessment and argue against the use of offline post-exposure tests as the only form of assessment of SL, since results obtained that way do not reflect all components of statistical learning and are heavily affected by cognitive load, lack of attention, or fatigue (Batterink & Paller, 2017).

The two supposed components of statistical learning are (1) a *word identification component* described as the perceptual binding of stimulus units into larger items and (2) a *memory storage component*, arguably a peripheral process of storing the integrated items in long-term memory (Batterink & Paller, 2017). Researchers have voiced their concerns about limitations of post-exposure tasks which test mainly the second component of statistical learning especially when used on their own. Furthermore, different post-exposure tasks might engage different processing strategies. For example, a target detection task might reflect contributions from implicit as well as explicit memory and therefore might be less prone to long-term memory limitations than the rating task which possibly engages explicit memory only (Batterink & Paller, 2017). It is therefore strongly advised to use assessments effectively monitoring both components of SL and not focusing merely on testing the successful storing and retrieval of the statistical structures, which might be highly affected by behavioural biases.

Biases associated with behavioural decisions and task complexity can be avoided by using **online assessment during the exposure task** which allows for monitoring of the gradual acquisition of knowledge during the participant's exposure to the stimuli. Researchers came up with several methods of assessing statistical learning online. Studies using online assessment methods of statistical learning typically use a recording of the neural steady-state response or event-related brain potentials (see e.g. Buiatti, Pena, & Dehaene-Lambertz, 2009, or Abla, Katahira, & Okanoya, 2008). These methods examine the first component of statistical learning, in other words, the gradual binding of stimuli units into more complex items. According to Batterink and Paller (2017), recording a neural steady-state response is the ideal method to measure response to rapidly changing continuous stimuli streams which are typical for statistical learning experiments.

Neural oscillations are an imminent part of the brain's electrophysiological signal. The intrinsic neural activity is sustained at rest but becomes temporally aligned with the temporal characteristics of a periodic stimulus, such as speech, in a process referred to as neural tracking or neural entrainment[1] (Giraud & Poeppel, 2012). The neural oscillations

[1] The term *neural entrainment* is deemed problematic by Obleser and Kayser for cases when the stimulus is merely quasi-periodic (Obleser & Kayser, 2019). For these cases, Obleser and Kayser suggest the terms *neural tracking* or *neural entrainment in a broader sense*. Other secondary sources are often not concerned with this issue and use the term neural entrainment more broadly without further specification of its scope. To refrain from confusion, the present study uses the term *neural tracking*.

interact with the stimulus generated activity and phase reset with the salient points in the input (Giraud & Poeppel, 2012). However, the temporally aligned activity may not cease immediately with the end of the stimulus and may linger for several cycles (Obleser & Kayser, 2019).

In order to be successfully tracked, the units marked by the salient points must be reasonably periodic (Tune & Obleser, 2022). For speech, such salient points are marked by the boundaries of relevant linguistic chunks, such as syllables and words (Myers, Lense, & Gordon, 2019). Although natural speech is not strictly periodic, its quasi-periodicity, stemming from the fluctuating linguistic units, upon which a strict regularity can be imposed if needed (Obleser & Kayser, 2019), allows for tracking of the stimulus properties (Giraud & Poeppel, 2012). Due to the combinatory nature of the linguistic units, the tracking can occur simultaneously on multiple temporal levels at distinct hierarchically organised frequency bands (Ding et al., 2016).

It is important to note that there is no one-to-one mapping of frequency bands and a distinct function; on the contrary, a particular frequency band has a plethora of functions and is involved in many cognitive processes (Tune & Obleser, 2022). Neural oscillations and their synchronisation to the stimulus support speech perception; however, they are also vital for learning and memory processes due to their role in synaptic plasticity (Tune & Obleser, 2022).

It has been argued that the delta, theta, and the gamma band are the most significant frequency bands for speech perception (Giraud & Poeppel, 2012). The delta band (1-4 Hz) corresponds to speech processing at the word frequency, the theta band (4-8 Hz) corresponds to the syllable frequency, and finally, the gamma band (>30 Hz) corresponds to the semantic, syntactic, and phonological integration (Tune & Obleser, 2022). Due to the mentioned combinatorial nature of speech, the signal can be processed and tracked simultaneously at all the bands mentioned above, as the smaller units (e.g. syllables) are combined into larger units (e.g. words) in real-time during speech perception.

The process of **neural tracking** can be successfully captured with non-invasive neuro-imagining methods, such as electroencephalography (EEG), which records the electrical signal of the brain (Giraud & Poeppel, 2012). The strength of the phase-locking of the neural signal and the stimulus at different frequencies and over several trials is demonstrated by the inter-trial phase coherence (ITPC) value (for more details on the computation of ITPC, see *Chapter 3.*). Not only can we see the synchronisation at the particular frequency bands by computing the ITPC, but research also shows that it is possible to decode the speech stimulus from the recorded neural signal by cause of the phase synchronisation (Obleser & Kayser, 2019). Finally, since the neural tracking of speech shows the phase synchronisation of the neural activity and the salient points in the speech stimulus, the neural tracking can be used to assess the effects of segmentation and inattentive learning during speech perception.

Several researchers chose to use the monitoring of **neural steady-state response** as a means of assessing the learning process during speech perception. Buiatti, Pena, and Dehaene-Lambertz (2009) used electroencephalography (EEG) to monitor the learning process during continuous auditory speech, by assessing the correlations between the elicit steady-state response at the word rate and the behavioural detection of words. Kabdebon et al. (2015) showed that phase-locking of the neural activity to the artificial speech stream at syllable and word frequencies can be used to test statistical learning in 8-month old infants. Batterink and Paller (2017) used a similar phase-locking paradigm to assess statistical learning in adults through exposure to auditory presented syllable streams.

During Batterink and Paller's experiment, the neural response of participants was recorded while they were being exposed to a structured auditory stream of four trisyllabic words and a random unstructured auditory stream. Their findings demonstrate that online assessment during the exposure task gives evidence of statistical learning in the gradual shift of participants' neural tracking of the stimuli from a syllable rate to a word rate over the course of the exposure task, in other words, a shift in the recorded steady-state neural response from individual stimulus units to more complex integrated items. These results were achieved by time-locking the collected data to the onset of each word or the third syllable and computing the phase-locking value of the neural activity and the stimulus for each participant. A phase-locking value (inter-trial phase coherence, or ITPC) at the word rate which grew significantly with the exposure indicated participants' sensitivity to the trisyllabic structure within the structured stream but not in the random unstructured stream. The results demonstrate that it is possible to effectively measure the gradual acquisition of knowledge during the first component of statistical learning. The validity of the online assessment is reflected not only in the shift of neural tracking from higher to lower frequency, but also in the correlation of the online assessment results with the post-exposure task results (Batterink & Paller, 2017).

Although the online assessment of statistical learning showed significant results, the authors point out the possibility that the increased entrainment to trisyllabic structure in the structured condition can reflect the general tendency to see patterns in the environment and the tendency of processing input in bundles (Batterink & Paller, 2017). Individuals who process the stimuli in groupings of three would therefore show a higher entrainment to trisyllabic structure in the structured stream than individuals who process the stimuli in groupings of two or four. Furthermore, patterns in the stimuli might be discovered by other means than the computation of transitional probabilities. The processing of the stimuli might be affected by an interplay of the auditory sensory memory which stores recently heard syllables (Batterink & Paller, 2017). If the syllable is heard again while still active in the sensory memory, the processing of said syllable would differ from the rest and the participants might perceive such syllable or a combination of syllables as an underlying integrated unit independent from the transitional probabilities computation (Batterink & Paller, 2017). This raises the question whether such strategies would be independent from statistical learning or whether they would represent another route to extracting regularities in the environment as a part of the statistical learning process (Batterink & Paller, 2017).

Several other studies chose to inspect the online processes of speech segmentation and statistical learning, mostly using the recording and subsequent computation of **event-related brain potentials** (ERPs). Abla et al. examined ERPs recorded during a continuous auditory speech stream organised into tritone words and demonstrated that the word onset elicited larger N100 and N400 amplitudes, thus confirming Sanders et al.'s results (Sanders, Newport, & Neville, 2002) which showed larger negative potentials (N100, N400) elicited by the word onset in a structured speech stream of trisyllabic nonwords (Abla, Katahira, & Okanoya, 2008).

Both the N100 and the N400 are components associated with language processing. The auditory N100 is associated with auditory segmentation and phonological perception and may signalise apprehension of word boundaries in continuous speech streams (Payne, Shantz, & Federmeier, 2020). The N400 response is most frequently associated with predictability from the context, semantic expectancy and semantic-pragmatic violations, e.g. in sentences with semantically anomalous final words, as demonstrated in Kutas and Hillyard's study which showed that semantically anomalous and semantically congruent final sentence words elicit a different N400 response (cited in Payne, Shantz, & Federmeier, 2020). However, the N400 response can be associated with the processing of a word or a meaningful stimulus in general without the need for semantic anomalies (Payne, Shantz, & Federmeier, 2020).

Similar results to those of Abla et al. and Sanders et al. were demonstrated in Cunillera et al.'s study which used an event-related brain potential experiment to explore how statistical cues and stress cues aid speech segmentation (Cunillera et al., 2006). Apart from a stream of artificial speech marked by statistical cues only, a stream analogical to the ones used by Sanders et al. or Saffran et al., a stream with a combination of statistical and stress cues was presented to the participants. Their findings gave further evidence that the N400 component is involved in the process of learning the nonsense words and could be identified as an indicator of speech segmentation, indicating lexical search (Cunillera et al., 2006).

Although Batterink and Paller argue that the computation of event-related brain potentials to individual syllables in a rapidly presenting continuous stream is not ideal and that the neural steady-state response is a better suited method for such stream, they computed ERPs to word onsets in an additional analysis to align with the prior literature (Batterink & Paller, 2017). However, the measured N400 response in their study did not significantly correlate with the word-learning index or the phase-locking value at the word frequency.

Arguably, this might be so due to different aspects of SL and word segmentation being reflected by the two different measures (Batterink & Paller, 2017). Furthermore, a particular type of assessment might be favoured based on the type of input or experimental task.

2.1.3 Statistical Learning Across Languages and Modalities

There is a plethora of evidence suggesting that statistical learning is widely available to learners across modalities, domains, and species (Fiser & Aslin, 2001), however, there are certain differences in the process of statistical learning based on the nature of the input and the participant's familiarity with it (Bulf, Johnson, & Valenza, 2011). The previous research shows that one's native language background affects the timing and likelihood of a successful speech segmentation in auditory statistical learning studies (Caldwell-Harris et al., 2015, cited in Frost et al., 2019). The statistics of one's native language lead to biases towards probable unit combinations affecting the pattern detection in the stimuli stream (Dal Ben, Souza, & Hay, 2021). The performance of a participant who has been exposed to language prior to the auditory statistical learning experiment reflects the influence of their native language patterns on the artificial language they are exposed to (Frost et al., 2019).

The results of Siegelman et al.'s study, which explored how prior knowledge and modality affect the processes of statistical learning, suggest that there are common SL principles across modalities (Siegelman et al., 2018). Indeed, studies with visual and even tactile stimuli suggest that statistical learning is a domain-general principle. Fiser et al. explored statistical learning in the visual domain and demonstrated that passive viewing of complex visual scenes results in SL (Fiser et al., 2001). In Fiser et al.'s experiment, the participants watched visual scenes of 12 shapes arranged on a fixed grid, similar to artificial word structures in auditory experiments. The learned statistics included single-shaped frequency, absolute shape position, and shape-pair arrangement (Fiser et al., 2001) similar to learning syllable co-occurrences and individual words in the auditory speech stream. Participants' results in post-exposure tasks demonstrated that they were able to extract independent components from complex visual scenes automatically and spontaneously (Fiser et al., 2001: 503), similar to participants in studies that focused on auditory statistical learning.

Visual statistical learning was further explored by Turk-Browne et al. who used fMRI and subsequent post-exposure tasks to compare the participants's processing of statistically structured and unstructured sequences of shapes (Turk-Browne et al., 2009). Turk-Browne et al.'s study is one of few that explores neural foundations of visual statistical learning. The observed activation of striatum and medial temporal lobe (hippocampus) in their study reflect a relationship between SL and other forms of associative learning and memory (Turk-Browne et al. 2009). Their results suggest that visual statistical learning can occur implicitly and quickly and in a similar manner to the SL present in the auditory domain. Furthermore, the results of this study prove that neuroimaging methods are a technique suitable to explore the visual processes of SL (Turk-Browne et al., 2009).

Interestingly, Frost et al.'s study showed that successful visual statistical learning predicts ability in literacy acquisition in a second language when L2 differs in terms of statistical properties from their native language (L1) (Frost et al., 2013). Results of this study demonstrated that higher scores obtained by computing transitional probabilities in a continuous stream of non-linguistic visual shapes correlated with higher scores for learning Hebrew in adult American Hebrew students (Frost et al., 2013). Furthermore, the score obtained by computing transitional probabilities did not correlate with working memory (Frost et al. 2013), thus excluding its significant effect on both scores. These findings seem to prove a significant link between visual statistical learning and the acquisition of reading ability in L2, perhaps suggesting a common mechanism (Frost et al. 2013), which is arguably poorer in individuals with visual primary language modality.

2.2 Silent Visual Speech

Even though several studies explored statistical learning in the visual domain and proved its effects in the said domain, the visual aspects of language were not thoroughly explored in relation to statistical learning. The evidence showing that prior knowledge and familiarity with the input affect the results of statistical learning suggest that familiarity and experience with visual speech (VS), i.e. silent articulatory cues to linguistic content decodable by lip reading, will significantly influence word segmentation. It is possible that one’s native modality will affect statistical learning and word segmentation to a similar extent as one’s native language. Experience with visual speech and the extraction of linguistic information from it through the act of lip reading should therefore be taken into account.

2.2.1. Linguistic Information Carried By Silent Visual Speech

Lip reading or *speech reading* consists of decoding speech based on the visual information extracted from lip movements (Bourguignon et al., 2020). Lip movements are one of the features that shape the identity of individual phonemes (Bourguignon et al., 2020) and thus visually differentiating two sounds with different lip configuration. On the other hand, visual cues cannot distinguish phonemes with the same lip configuration, such as /p/, /b/, and /m/. Recognition of phonemes based on the visual cues resulted in the coinage of the working term *viseme* which refers to the visual equivalent of a phoneme, a term most frequently used in computer speech recognition studies (Bear et al., 2014). Although the term *viseme* lacks a precise definition, it is used to refer to a set of phonemes with an identical visual lip configuration. Therefore, a phoneme corresponds to precisely one viseme, however, a viseme can map to several phonemes in a phoneme-to-viseme mapping (P2V) (Bear et al., 2014). Over the time P2V resulted in several viseme phoneme sets such as {/p/ /b/ /m/} {/f/ /v/} {/θ/ /ð/} {/ʃ/ /ʒ/} {/k/ /g/} {/w/} {/r/} {/l/ /n/} {/t/ /d/ /s/ /z/} (Binnie, Jackson, & Montgomery, 1976) (with the sets differing slightly based on the author of the mapping and their native language). Phonemes in the same viseme sets cannot be recognised in silent speech based on visual cues only. Furthermore, apart from the identification of individual phonemes, lip reading also facilitates speech parsing by indicating phrase/sentence boundaries since lip movements provide significant information about the speech envelope (Bourguignon et al., 2020).

Evidence that visual speech undoubtedly carries linguistic information was given by Muthukumaraswamy et al. who proved not only that biological and non-biological movement is distinguished in neural processing, but also that silent non-linguistic and linguistic lip movements are processed differently by the hearing population which suggests that the brain uses different encoding strategies based on the function of the observed movement (Muthukumaraswamy et al., 2006). The idea that linguistic movement is processed differently was supported by Calvert et al. (1997). Calvert et al.'s study proves that, similarly to silent speech, observed phonologically plausible silent pseudospeech activates auditory cortices showing that both silent speech and silent pseudospeech are differentiated from non-linguistic facial movements regardless of the presence of a semantic context and that the modulation of auditory speech by lip reading appears to be at a prelexical level at the stage of phonetic classification (Calvert et al., 1997).

The amount of reliance on the visual modality during speech perception is highly individual, though it can be assumed that certain populations will rely on visual cues in speech more than others. In the hearing population, lip reading[2] is a crucial ability for understanding speech in deprived auditory conditions since it immensely enhances the perception of audiovisual speech outside of the hearer's awareness. The influence of the seen lip movements on the auditory perception of speech apparent in incongruent audiovisual speech is demonstrated among other evidence by the McGurk effect, in which auditory and visual stimuli synthesise the perceived sound (McGurk & MacDonald, 1976). Furthermore, the presence of neural entrainment of lip movements to low-frequency brain oscillations has been demonstrated in speech processing brain areas and it has been discovered that this entrainment is modulated by congruence of the auditory and visual speech stimuli and contributes to speech comprehension (Park et al., 2016). At the same time, non-manual features, such as facial expressions or body posture, are one of the main parts of sign

[2] The term *lip reading* used in the present study refers to the same phenomenon as *silent lip reading, speechreading*, and *silent speechreading*. It refers to observing and extracting linguistic information from silent linguistic lip movements (referred to as *silent speech, visual speech, silent visual speech,* and *visual-only speech*).

languages alongside manual features, such as gestures, and finger spelling (Cooper et al., 2011). In sign languages, lip reading is crucial, since it is often the lip movement only that differentiates two signs (e.g. '*sklo'* (*glass*) and '*okno'* (*window*) in Czech sign language). The role of visual speech in language perception is hence pivotal regardless of the modality of the perceived language and its examination can help us understand how the integration of the visual and auditory modalities produces unified perception of speech (Calvert et al., 1997).

The lip reading skills and the general reading skills are closely related since they both require extracting information from a visual language input (Mohammed et al., 2006). The lip reading skill can therefore be affected not only by one's native modality and experience with lip reading but also by reading ability as well as by dyslexia and other similar learning difficulties (Mohammed et al., 2006). The results of Mohammed et al.'s study show that lip reading skill correlates with reading ability in the adult deaf and adult dyslexic participants (Mohammed et al., 2006).

2.2.2 Neural Processing of Silent Visual Speech in Hearing and Hearing-Impaired Adults

Several studies (Calvert et al., 1997; Hall et al., 2005, MacSweeney et al., 2000; Park et al., 2016) demonstrated that observing silent speech evokes a similar neural response as listening to auditory speech. It has been shown that observing silent speech activates the auditory cortices in the hearing population albeit with an additional delay compared to the one present at auditory speech processing (Bourguignon et al., 2020). Bourguignon et al. explains the greater visual-to-auditory delay by mapping out the complex encoding process of extracting the linguistic information from silent speech and feeding it to the auditory cortices. The study suggests that the brain of a hearing adult observing visual speech first synchronises to the visual properties of the lip movements in the visual cortex at frequencies >1 Hz (Bourguignon et al., 2020). Slower frequency information is then extracted from the lip movements and mapped onto the corresponding sounds by the right angular gyrus, more precisely by an area close to it called temporal visual speech area (Bernstein et al., 2011). This information then proceeds further to the auditory cortices. In their study, this leads to the entrainment to silent speech at frequencies <1 Hz, which match with phrasal, stress, and sentential rhythmicity, and the activation of the auditory cortices (Bourguignon et al., 2020).

It has been shown that the auditory cortices are activated by seeing silent speech even when the observer does not know its content (Bourguignon et al., 2020). However, data collected in previous studies show that the knowledge of the content of silent visual speech significantly aids successful parsing and tracking of silent speech at syllable rate. Crosse et al. were able to confirm entrainment to silent speech with known content in the hearing population at frequencies 4-8 Hz (Crosse et al., 2015), however silent speech with unknown content lead only to entrainment at <1 Hz frequencies in Bourguignon et al.'s study

(Bourguignon et al., 2020). This might be caused by the extraction relying on a prediction of mouth movements, a fact supported by the short lip-to-brain delay (~40 ms) (Bourguignon et al., 2020), by the observer's insufficient lip reading skill and unfamiliarity with the stimuli's modality, or by the nature of the chosen stimuli.

Cerebral localisation of audiovisual and visual silent speech processing in deaf and hearing adults has been documented in numerous fMRI studies. In MacSweeney et al.'s study, observing silent speech lead to an increased activation in the posterior cingulate cortex and the lingual/hippocampal gyri in deaf participants, in contrast to the consistent activation in the left temporal lobe in the hearing population (MacSweeney et al., 2002). Activation in the posterior regions was not found in the hearing participants. The right hippocampus which was activated in the deaf participants activates during memory retrieval, suggesting a role of memory during speechreading in the deaf population (MacSweeney et al., 2002).

MacSweeney et al.'s data suggest that acoustic experience modulates the **functional circuits for speech processing** since speech processing is different in congenital deaf people. Their study suggests that if hearing is absent from birth, the speech processing system in the left temporal lobe develops idiosyncratically. On the other hand, the speech processing areas in the right temporal lobe might be less affected by congenital deafness since the right temporal lobe is activated in deaf participants during visual speech processing extending to Herschl's gyrus, an area traditionally classed as secondary auditory cortex, which is activated during auditory speech in the hearing population as well as by sign language in native deaf signers. This region might also be significant for processing of the visual aspects of language (MacSweeney et al., 2002). It is important to note that the demonstrated neural activation in both deaf and hearing participants is specific to linguistic mouth movements since non-linguistic mouth movements were processed analogously in the occipito-temporal regions by both groups (MacSweeney et al., 2002).

Capek et al.'s study supports the idea that the hearing status as well as speech reading skill modulate the activation in the left superior temporal regions during silent speechreading (Capek et al., 2008). Their data showed activation in left middle and posterior superior temporal cortex during speechreading in both deaf and hearing participants, however the activation was greater in deaf participants when speechreading skill was used as a covariate. Capek explains this contrast to previous findings by pointing out that previous studies used a closed stimulus set (numbers 1-9, MacSweeney, 2002) as opposed to English sentences (Capek et al., 2008). When the speechreading skill was used as covariate, greater activation was shown in left posterior superior temporal sulcus (p-STS), an area that might be responsible for cross-modal processing of audiovisual speech in hearing people (Capek et al., 2008). Possible explanation for the greater activation in deaf participants might therefore be the sensitivity to dominant speech modality within p-STS. It is plausible that p-STS develops a greater sensitivity to visual speech in deaf people as opposed to greater sensitivity to auditory speech as a primary and visual speech as a secondary function in the hearing population (Capek et al., 2008). The greater activation in the deaf participants might also reflect a more general plasticity of the brain in deaf people as demonstrated by regions specialised for auditory stimuli being recruited for different modality processing in the deaf population (Capek et al., 2008).

In MacSweeney et al.'s 2004 study, visual gestural language activated frontal-posterior network and superior temporal cortex including planum temporale in both deaf and hearing participants, however the activation was greater in deaf participants. Their results support the previous findings which demonstrated that when hearing is absent in early development, the visual processing role of certain brain regions, such as planum temporale (an area processing visual movement in both deaf and hearing populations), is enhanced. In their study, the native language of the deaf participants (BSL as opposed to Tic Tac) evoked greater activation in the left posterior superior temporal sulcus and gyrus, suggesting the linguistic specificity of the activated regions and the importance of the left posterior perisylvian cortex for language processing regardless of modality (MacSweeney et al., 2004).

These findings demonstrated that silent speech is processed in brain areas specialised for language processing in contrast to non-linguistic lip movements and that hearing status as well as speechreading skill (which is tightly associated with the modality of one’s native language) modulate the processing of silent speech. Furthermore, these findings demonstrate the extensive degree of brain plasticity in the deaf population. The observed extent of brain plasticity and idiosyncratic neurophysiology of congenital deaf people can be profoundly helpful in the research of what parts of auditory speech are innate or learnt. However, it is important to note that each brain region does not have only one function and the regions noted above are not the only regions activated by silent speech. Observing silent speech and the simultaneous speechreading activates an extensive network of linguistic and sensory-motor areas (MacSweeney et al., 2004).

3. THE PRESENT STUDY

3.1 The Research Question and Hypotheses

As has been outlined in the previous chapter, the aim of the present study is to investigate the mechanisms involved in the perception of novel silent visual speech and examine the segmenting and learning strategies taking place during speech processing, in order to better understand the general processes underlying speech perception regardless of its modality. The present study test the following three working hypotheses: H1 that visual cues will suffice in facilitating statistical learning in visual silent speech; H2 that greater effects of statistical learning will be found in the novel stream of visual speech with the covert statistical structure; and H3 that the primary modality and language background will affect the participant's neural processing and the ability to learn during the perception of silent visual speech.

3.2 Method

3.2.1 Participants

25 adult participants were included in the present study (five additional participants were tested but excluded due exceedingly noisy data). Participants were divided into three groups based on their hearing status and their native language.

Group A consisted of seventeen adult participants (3 of them were excluded due to a high number of artefacts, final *n*=14, 13 women, mean age=30 y, SD=9.88) with normal hearing and no significant prior knowledge of any sign language. **Group B** consisted of seven hearing impaired adult participants (1 excluded due to artefacts, final *n*=6, 5 women, mean age=33 y, SD=6.66), four of whom were adults with a complete hearing loss at birth (*n*=3) or in adulthood (*n*=1), and three of whom were hard-of-hearing adults. The L1s of the participants were Czech and Czech sign language, respectively in the two groups, with the exception of the Group B participant with a hearing loss in adulthood and the hard-of-hearing adults whose L1 was Czech. **Group C** consisted of six adult participants (1 excluded due to artefacts, final *n*=5, 5 women, mean age=30 y, SD=7.64) with normal hearing and no significant prior knowledge of any sign language, whose L1 was English. All participants reported normal or corrected to normal vision, no neurological or psychiatric disorders, and were not taking any medication or substances affecting the nervous system. Furthermore, none of the participants reported being diagnosed with dyslexia or having a familial risk of dyslexia.

The order of the presented stimuli was not determined by the participant's belonging to the group A, B, or C. In order to achieve exposure of all groups to both stimuli orders, All groups were further divided into two subgroups (A1, A2, B1, B2, C1, C2) as shown in *Table 1*. Both hearing and hearing impaired participants were exposed to the same stimuli and underwent the same pre-exposure procedure and post-exposure tasks. All participants were informed about the procedure and possible risks prior to the experiment. The instructions and the administration of the experiment were in the participants' native language.

Subgroup	**A1**	**A2**	**B1**	**B2**	**C1**	**C2**
Hearing status	normal hearing	normal hearing	impaired hearing	impaired hearing	normal hearing	normal hearing
L1	Czech	Czech	CSL	CSL	English	English
Condition order	S-R	R-S	S-R	R-S	S-R	R-S
Task order	S-rt-R-lip	R-S-rt-lip	S-rt-R-lip	R-S-rt-lip	S-rt-R-lip	R-S-rt-lip

Table 1: Subgroups of participants and their respective hearing status and condition order. Abbreviations in Task order refer to the structured condition (S), random condition (R), rating task (rt), and the lip-reading task (lip).

3.2.2 Stimuli

The **transcript** for the stimuli in the present study mirrors the datasets typically used in auditory word segmentation studies with several changes made in order to achieve visually distinguishable stimuli. The present set of stimuli was modelled mainly after the stimuli of Aslin et al.'s study and its modification in Podlipský et al.'s study which was altered in order to obtain stimuli that would allow for an effortless acoustic segmentation. Given the present language environment and the nature of the present study, the present stimuli were created such as to avoid visemes that would be hard to perceive for a Czech, CSL, or an English L1 speaker and aimed to facilitate visually distinguishable speech streams. Phonemes that represented the same or visually similar visemes were replaced (bi>be, do>fe, ro>fo). These changes resulted in a stimulus set shown in *Figure 1*.

Transcripts for the **structured speech stream** consisted of four repeating trisyllabic nonsense words consisting of 12 CV syllables (*pa, be, ku, da, fo, pi, ti, bu, fe, go, la, tu*) in a pseudorandom order (see *Figure 1*) with no longer pauses indicating the word boundary. Boundaries between the words were cued merely based on transitional probabilities between the syllables, other cues were avoided. The transitional probability between neighbouring syllables was higher within the word boundary (1.0) than outside of it (.33). For example, every *pa* in the structured stream was followed by *be* (1.0) and every *be* was followed by *ku* (1.0), however, *ku* was equally likely to be followed by *da, go,* and *ti* (.33). Consecutively repeated words were not allowed and the frequency of each word in the structured stream was approximately the same.

The **random stream** consisted of the same 12 CV syllables in a pseudorandom order (with occasional occurrences of mispronounced syllables). A set of 1080 syllables was randomised and all consecutive repeating syllables were subsequently replaced. The frequency of each syllable in the random stream was approximately the same. The order of the syllables in the random stream could not be predicted based on the transitional probabilities or any other cues. Excerpts of the transcripts for both conditions can be found in the *Appendix (8.1)*.

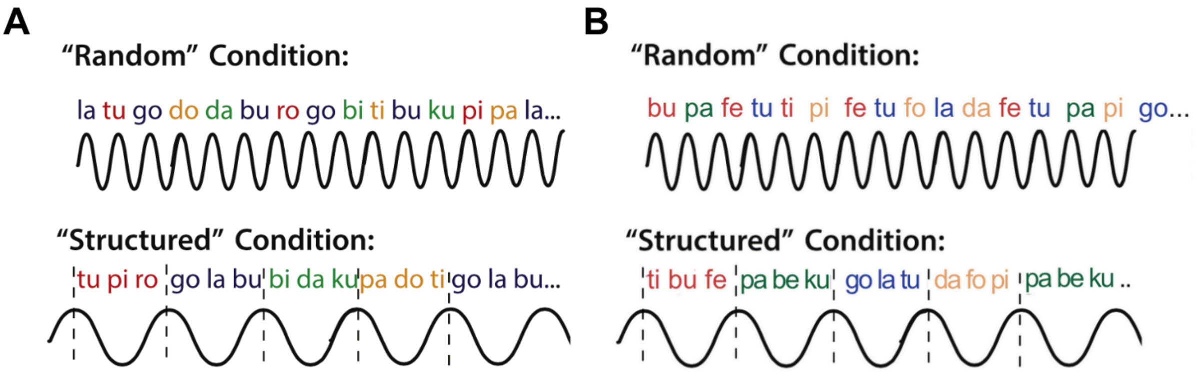
A
"Random" Condition:
la tu go do da bu ro go bi ti bu ku pi pa la...
"Structured" Condition:
tu pi ro go la bu bi da ku pa do ti go la bu...
B
"Random" Condition:
bu pa fe tu ti pi fe tu fo la da fe tu pa pi go...
"Structured" Condition:
ti bu fe pa be ku go la tu da fo pi pa be ku ..

Figure 1 ***A:*** *An excerpt from the random and structured auditory speech datasets used in Batterink and Paller's 2017 study and the respective EEG-based entrainment measure of learning for both the random condition at a syllable frequency and the structured condition at a word frequency (figure from Batterink and Paller 2017: 33).* ***B:*** *An excerpt from the transcripts of the random and structured visual speech datasets used in the present study.*

The **final recorded stimuli** were presented in a video-only format showing the head and shoulders of a speaker reading the transcripts on a computer screen (see *Figure 2*). The speaker in the presented video was a native speaker of Czech with a C2 level English who used audiovisual speech while recording the stimuli in order to preserve all properties of natural speech. The speaker was instructed to read the presented text with natural articulation and in alignment with the rhythm set by a metronome (200 bpm) which corresponded to 3.33 Hz per syllable. This rate mirrored the syllable rate used in previous literature (Batterink & Paller, 2017). The stimuli for the random speech stream, the structured speech stream, the rating task, and the speechreading test were recorded separately with brief breaks between the recordings. The recorded videos were subsequently edited in order to eliminate any errors made by the speaker and to create four separate blocks of continuous streams as well as stimulus sets for both post-exposure tasks. The audio was subsequently removed from the videos and the intelligibility of the resulting videos was tested by a test group of native Czech speakers prior to the experiment. All videos showed the same speaker.

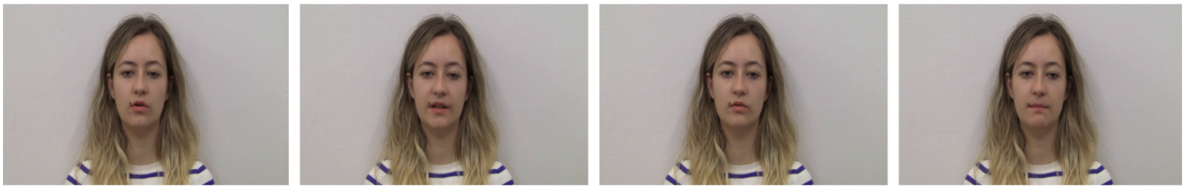

Figure 2: An excerpt from the visual stimulus used for the random condition.

The onsets of every first syllable of the trisyllabic words in the structured stream and every first syllable of a random trisyllabic structure in the random stream were marked in *ELAN* (version 6.5, Max Planck Institute for Psycholinguistics). Onsets of the individual syllables were then calculated based on the onset and the duration of the trisyllabic words and structures. The marked onsets were used in the subsequent phase-locking analysis (see *Chapter 3.3.1*).

3.2.3 Procedure

Both hearing and hearing impaired participants underwent the same procedure. For deaf participants, a Czech sign language interpreter was present during the entire experiment to give instructions and answer questions. Demographic and language background data from all participant groups were obtained in the form of a short questionnaire distributed prior to the experiment. The experiment presented to the participants was designed in *PsychoPy* (Peirce et al., 2019) and consisted of four segments separated by a short break. The exposure task was separated into two segments, a structured stream and a random stream. The remaining two segments of the experiment design were the post-exposure tasks.

EEG data were recorded over the course of the entire experiment with a sampling rate of 200 Hz from 19 electrodes attached to an electrode cap placed on the participant's head prior to the experiment. Additional five electrodes were placed on the participant's nose, on the outer canthus of the right eye, under the right eye, and on the mastoid bones behind the right and the left ear. Conductive gel was placed in the gap between the participant's skin and the electrode sensors to ensure sufficient conductivity. Recordings were made with the TruScan software (Deymed diagnostic).

In the **exposure task**, participants were presented with videos of both the structured stream and the random stream in a consecutive condition-dependent order. The order of the two conditions was counterbalanced across participants (see *Table 1*). Each condition consisted of 6 minutes of a continuous silent speech stream broken up into two 3-minute blocks separated by a short break. The post-exposure rating task was administered right after the structured condition since it referred only to the structured stream. No post-exposure task tested the knowledge gained from the random stream (Batterink & Paller, 2017).

Each of the exposure task segments showed a set of instructions followed by a continuous visual-only speech stream. Participants were instructed to attend to the silent video, limit their movement and focus on the speaker's mouth. During the brief break between the blocks, participants were advised to close their eyes or change the centre of their focus to avoid strained eyes. The two conditions were further separated by a longer 4 minute break (i.e. after the rating task in the *S-R* condition or after the random stream in the *R-S* condition) in which the participants were instructed to disassemble and reassemble a wooden puzzle, with the intention to further rest their eyes from viewing the computer screen and maintain their attention.

The **post-exposure rating task** which followed immediately after the structured speech stream was modelled after an analogical task in Batterink and Paller's 2017 study, although it was slightly altered in order to better fit the scope and the aim of the present study. On each trial, the participants were presented with a video of a silently mouthed trisyllabic pseudoword, trisyllabic non-word, disyllabic part-word, or a disyllabic non-word. Both the pseudowords and the part-words, but not the non-words, repeatedly appeared in the structured speech stream. The aim of the rating task was to assess participants' explicit memory of the pseudowords (Batterink & Paller, 2017) present in the structured speech stream and thus testing the knowledge gained as a result of the statistical learning process.

Participants were instructed prior to observing the short videos to evaluate these excerpts based on their familiarity. A familiarity prompt appeared on the screen after each video and the participants were asked to mark the excerpt as familiar or unfamiliar by pressing a corresponding key. This simplified design was chosen over the familiarity scale used in Batterink and Paller's study, the use of which could result in misinterpretation of the scale and inaccurate data. The rating task consisted of 12 trials. Four pseudo-words, six non-words, and two part-words were presented to all participants in the same fixed order (see *Table 2*).

Words	*pabeku, tibufe, golatu, dafopi*
Part-words	*pabe, bufe*
Non-words	*pitugo, lafobe, gotifo, bepafe, tuda, fego*
Presented order	*tibufe, pitugo, tuda, pabeku, bufe, lafobe, gotifo, pabe, golatu, fego, dafopi, bepafe*

Table 2: Rating task dataset.

After the exposure and the post-exposure assessment, each participant was given a short visual **test of speechreading** presented in their L1. The format of the presented test was loosely based on the British *Test of Adult Speechreading* created by (Mohammed, MacSweeney, & Campbell, 2003) and consisted of a video-only recording of a speaker silently mouthing words and sentences and a picture grid to choose the correct answer from.

Since the dataset for the British *Test of Adult Speechreading* was not available, an original dataset was created for the purpose of the present study based on the format of the British test described in Mohammed et al.'s study (Mohammed, MacSweeney, & Campbell, 2003).

In the first part of the test, participants were presented with a short video of a speaker silently mouthing **a disyllabic word** followed by a grid with nine pictures (1 target word, 8 distractors). For each set the participant had to identify the target word mouthed by the speaker by pressing the corresponding number (1-9) on a keyboard in front of them. The picture grid was revealed after the presentation of the video was over. Each picture grid contained nine pictures of visually distinguishable Czech words, such as *kuře, rohlík, osel, komín,* for groups A and B or nine pictures of visually distinguishable English words, such as *chicken, pancake, dolphin, airplane,* for Group C (see *Figure 3 A*). The Czech and English datasets differed slightly in order to remain visually distinguishable disyllabic target words and accommodate the participant's cultural background. However, the difficulty of the test remained the same. The first part consisted of 10 trials.

In the second part of the test, participants were presented with a short video of a speaker silently mouthing **a sentence** followed by a grid with six pictures (1 target sentence, 5 distractors) depicting a scene described by the sentence. Participants had to identify the target sentence by pressing the corresponding number (1-6) among distractors with a similar topic (see *Figure 3 B*). In order to identify the correct target sentence, the participant had to correctly lip read more than one word. The picture grid was revealed after the presentation of the video was over. The content of the Czech and the English dataset was the same. The second part consisted of 7 trials. The silent videos presented to the participant were filmed and edited in a similar manner as the videos for the exposure task.

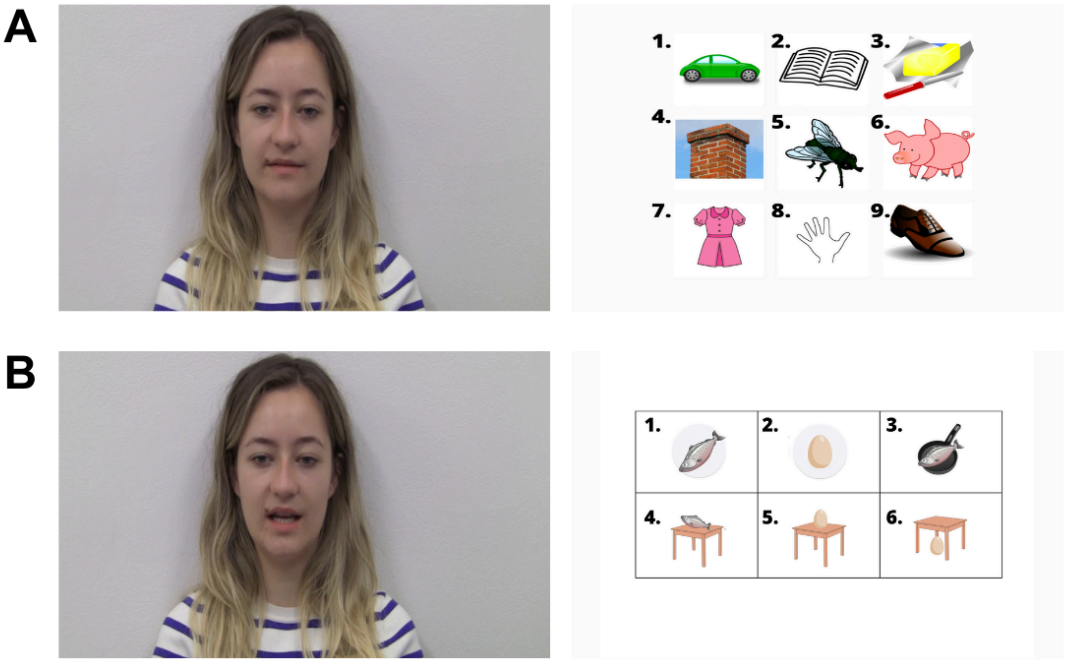

Figure 3 ***A:*** *An excerpt from the first part of the lip reading task - a video of a silently mouthed word and a picture grid screen.* ***B:*** *An excerpt from the second part of the lip reading task - a video of a silently mouthed sentence and a picture grid screen. Full datasets for both parts of the test can be found in the Appendix (8.3).*

3.3 Analysis

3.3.1 EEG Data Analysis

The EEG data analysis was carried out in the *EEGLAB Matlab toolbox* (Delorme & Makeig, 2004). The analysis followed the procedure administered by Batterink and Paller in their 2017 study. The collected EEG data was band-pass filtered from .1 to 30 Hz and time-locked to the onset of each word in the structured condition or every first syllable of the trisyllabic structure in the random condition. Data were then extracted into epochs whose length corresponded to the duration of 36 syllables (10.8 sec). Data containing large artifacts (threshold value = +/- 210 μV) were removed using an automatic artifact rejection procedure (Batterink & Paller, 2017).

Inter-trial phase coherence (ITPC), also referred to as *phase-locking value*, was used to quantify neural entrainment at syllabic and word frequencies in both conditions. ITPC value 0 indicates non-phase-locked activity, whereas ITPC value 1 indicates phase-locked activity. Significant ITPC indicates that the EEG activity in the given trial is phase-locked and not phase-random (Batterink & Paller, 2017). A continuous Morlet wavelet transformation from 0.2 to 20.2 Hz was used to compute intertrial coherence via the *newtimeif* function in *EEGLAB*. Sensitivity to the underlying trisyllabic structure of the structured speech stream is believed to be reflected in a relatively higher ITPC at the word frequency and lower ITPC at the syllable frequency in the structured condition (Batterink & Paller, 2017). The word frequency corresponded to approximately 1.1 Hz, whereas the syllable frequency corresponded to 3.3 Hz (Batterink & Paller, 2017). Sensitivity to the trisyllabic structure, referred to as *word learning index* (WLI), was quantified according to the following formula used by Batterink and Paller (Batterink & Paller, 2017):

$$WLI = \frac{ITC\ word\ frequency}{ITC\ syllable\ frequency}$$

A resulting higher WLI value indicates greater neural entrainment on the trisyllabic word frequency relative to the raw syllable frequency, which indicates the presence of statistical learning (Batterink & Paller, 2017). The aim of the present study was to examine whether the WLI value in the structured condition would be higher compared to the random condition and whether this value would increase with exposure in the structured condition but not in the random condition.

3.3.2 Behavioural Data Analysis

Data from the post-exposure rating task were used to compute participant's rating accuracy and rating score. Part-word data were removed from the analysis due to a suspected possible misinterpretation of the part-word familiarity prompt. Computed rating accuracy referred to the percentage of trials that the participant rated correctly and the rating score referred to the result of subtracting the average score given for non-words from the average score given for words. The presence of statistical learning was demonstrated by values above 0 (Batterink & Paller, 2017).

Data obtained in the lip reading test were used to calculate participant's lip reading skill. Lip reading skill was based on the percentage of successful trials in both parts of the lip reading test.

3.3.3 Statistical Analyses

R (version 4.3.1, R Core Team, 2021) was used to run the statistical models using the packages lme4 (Bates et al., 2015), and lmerTest (Kuznetsova et al., 2017). In addition, the package ggeffects (Lüdecke, 2018) was used for the estimation of means and confidence intervals, and ggplot2 (Wickham, 2016) to create plots.

ITPC and WLI were analysed using linear mixed-effects models (function *lmer*), which were used to analyse the effects of condition, block, and first presentation. In the ITPC model, condition (random vs structured, sum-coded as -1 vs +1), chunking (representing the rate of syllables vs words, sum-coded as -1 vs +1), and block (first vs second, sum-coded as -1 vs +1) as well as their interactions were entered as fixed effects. In the WLI model, the condition (random vs structured, sum-coded as -1 vs +1), block (1 vs 2, sum-coded as -1 vs +1), and their interactions were modelled as fixed effects. Per-participant and per-channel random intercepts were included in both models.

The post-exposure rating and lip-reading data were analysed with t-tests that compared participants' average performance in each of the tasks against chance. Finally, the EEG and behavioural data were compared using Spearman correlation tests on the continuous WLI values against the proportional lip-reading accuracies from the lip-reading task with isolated words, and on the WLI values against the word rating scores from the statistical learning post-test. For assessing significance in all models, we assume the threshold value of alpha = 0.05.

3.4 Results

3.4.1 EEG results

3.4.1.1 Group A Results: Hearing L1 Czech Participants

The linear mixed-effects model for the predicted ITPC values for Group A yielded a significant intercept which means that the ITPC averaged across all conditions and blocks is reliably larger than 0. The model also detected a significant effect of block, showing that ITPC was larger in the second block, and a significant effect of condition, showing that ITPC was larger for the structured than for the random condition. There was a significant triple interaction of condition, chunking, and block. Pairwise comparisons of the means (plotted in *Figure 5 Left*) show that it was the ITPC for words in the structured condition that rose significantly between the first and the second block. The model summary is shown in *Table 3*.

The model for WLI in Group A yielded a significant intercept, indicating that WLI across all conditions and blocks was larger than 0. There was a significant effect of block, and a significant interaction between condition and block. Pairwise comparisons, plotted in *Figure 5 Right*, revealed that WLI was larger in the second than in the first block, and this between-block difference was driven by the structured condition. The model summary is shown in *Table 4*.

	Est	SE	df	t	p
(Intercept)	0.08	0.005	13	15.974	<.001*
Condition (-rnd+str)	0.003	0.001	763	2.88	0.004*
Chunking (-syll+word)	0.002	0.001	763	1.487	0.137
Block (-1+2)	0.004	0.001	763	3.433	0.001*
condition:chunking	0.001	0.001	763	1.284	0.199
condition:block	0.001	0.001	763	1.144	0.253
chunking:block	0.002	0.001	763	1.532	0.126
condition:chunking:block	0.003	0.001	763	2.786	0.005*

Table 3: ITPC fixed-effect model summary for Group A.

	Est	SE	df	t	p
(Intercept)	1.133	0.06	14.575	18.982	<.001*
Condition (-rnd+str)	0.017	0.02	761	0.858	0.391
Block (-1+2)	0.044	0.02	761	2.192	0.029*
condition:block	0.066	0.02	761	3.319	0.001*

Table 4: WLI fixed-effect model summary for Group A.

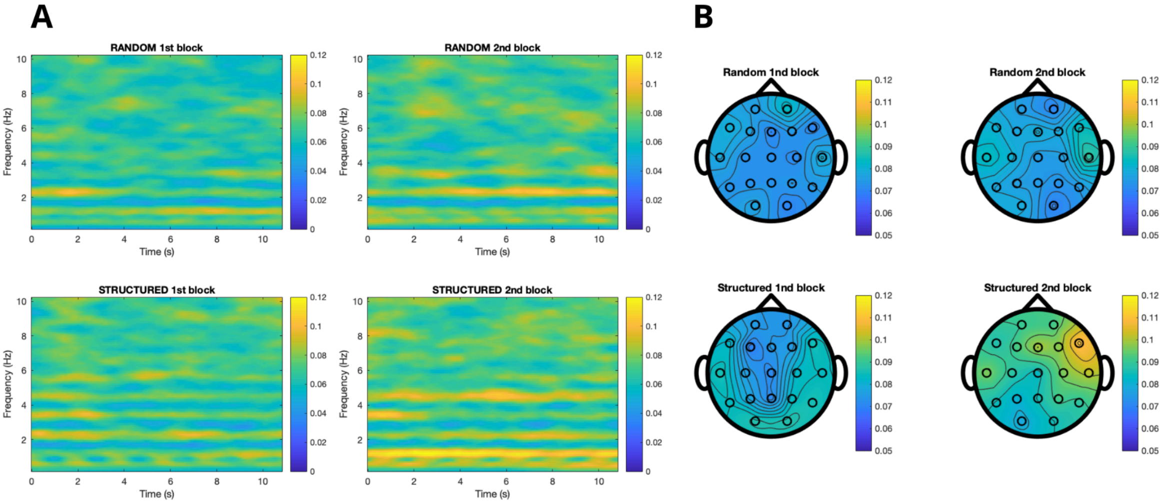

Figure 4 ***A:*** *ITPC values for Group A in the first and the second exposure block of each condition, averaged across the 7 analysed channels (Pz, O1, O2, T3, T4, T5, T6) and across all epochs.* ***B:*** *Localisation of the increased neural phase-locking activity in Group A at the word frequency (1.1 Hz).*

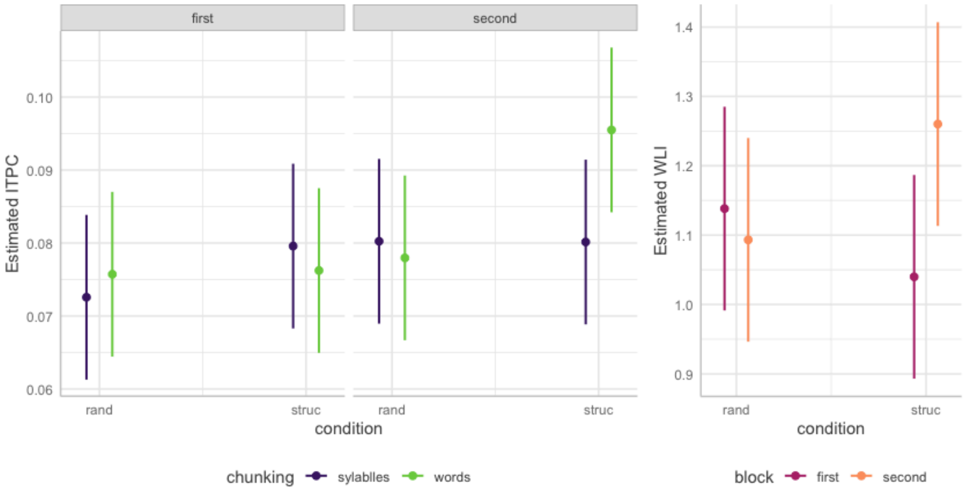

Figure 5 ***Left:*** *Estimated ITPC per condition, exposure block, and chunking rate for Group A.* ***Right:*** *Estimated WLI per condition and exposure block (means and 95% confidence intervals) for Group A.*

3.4.1.2 Group B Results: Deaf or Hard-of-Hearing Czech/Czech Sign Language Participants

The linear mixed-effects model for the predicted ITPC values for Group B detected a significant intercept showing that overall, ITPC was larger than 0. There was a main effect of block and as seen in *Figure 6*, ITPC was larger in the first block than in the second block (i.e. in the opposite direction than in Group A). No other main or interaction effects were significant. The grand average ITPC per condition, per chunking rate, and per exposure block is plotted in *Figure 7 Left*. The model summaries are shown in *Table 5*.

	Est	**SE**	**df**	**t**	**p**
(Intercept)	0.08	0.008	5	11.081	0.001*
Condition (-rnd+str)	<0.001	0.002	323	0.138	0.89
Chunking (-syll+word)	0.003	0.002	323	1.785	0.075
Block (-1+2)	-0.008	0.002	323	-5.097	<0.001*
condition:chunking	-0.003	0.002	323	-1.772	0.077
condition:block	<0.001	0.002	323	0.047	0.963
chunking:block	0.003	0.002	323	1.669	0.096
condition:chunking:block	0.002	0.002	323	1.227	0.221

Table 5: ITPC fixed-effect model summary for Group B.

The linear mixed effects model for the predicted WLI values in Group B shows a significant intercept, a significant main effect of condition, and a significant main effect of block. Overall, WLI was larger than 0, and as shown in *Figure 7 Right*, it was larger in the random than in the structured condition, and larger in the second than in the first block. The model summary is shown in *Table 6.*

	Est	SE	df	t	p
(Intercept)	1.145	0.098	5.02	11.722	<0.001*
Condition (-rnd+str)	-0.105	0.025	322.957	-4.278	<0.001*
Block(-1+2)	0.089	0.025	326.541	3.612	0.004*
condition:block	0.017	0.025	322.957	0.712	0.477

Table 6: WLI fixed-effect model summary for Group B.

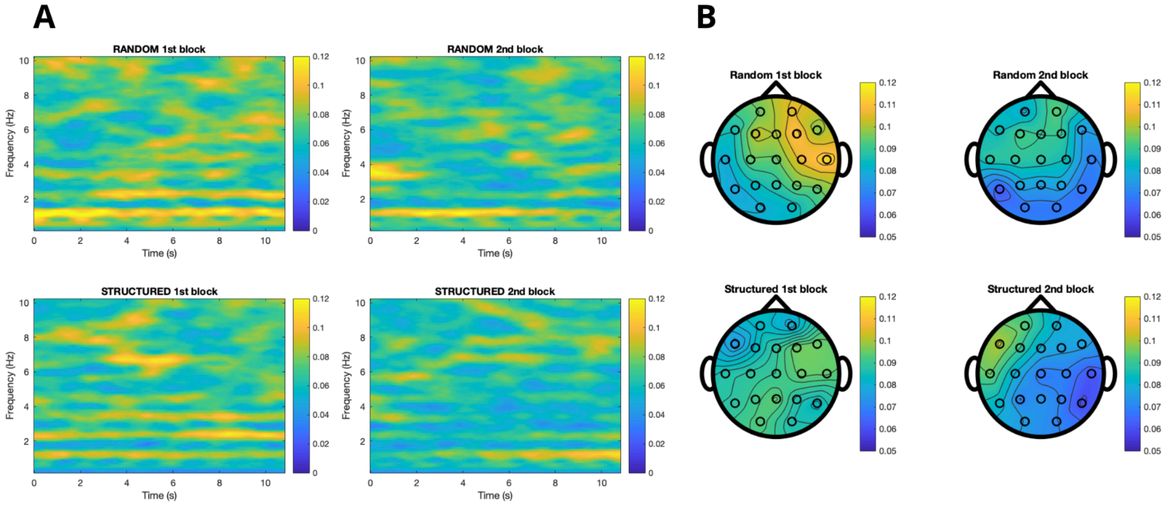

Figure 6 ***A:*** *ITPC values for Group B in the first and the second exposure block of each condition, averaged across the 7 analysed channels (Pz, O1, O2, T3, T4, T5, T6) and across all epochs.* ***B:*** *Localisation of the increased neural phase-locking activity in Group B at the word frequency (1.1 Hz).*

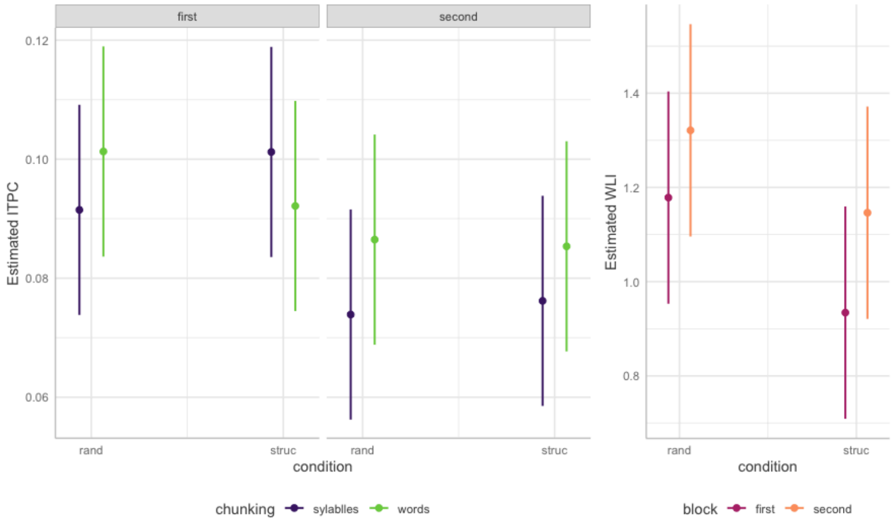

Figure 7 ***Left:*** *Estimated ITPC per condition, exposure block, and chunking rate for Group B.* ***Right:*** *Estimated WLI per condition and exposure block (means and 95% confidence intervals) for Group B.*

3.4.1.3 Group C Results: Hearing L1 English Participants

The linear mixed-effects model for the predicted ITPC values for Group C detected a significant main effect of chunking and a significant interaction between chunking and block. The grand average ITPC per condition, per chunking rate, and per exposure block is plotted in *Figure 9 Left*. The model summary is shown in *Table 7*.

	Est	**SE**	**df**	**t**	**p**
(Intercept)	0.082	0.007	4	11.135	0.004*
Condition (-rnd+str)	<0.001	0.002	268	0.293	0.77
Chunking (-syll+word)	0.006	0.002	268	3.81	0.002*
Block (-1+2)	<0.001	0.002	268	0.055	0.956
condition:chunking	-0.003	0.002	268	-1.819	0.07
condition:block	-0.001	0.002	268	-0.51	0.61
chunking:block	0.004	0.002	268	2.298	0.022*
condition:chunking:block	<0.001	0.002	268	0.118	0.906

Table 7: ITPC fixed-effect model summary for Group C.

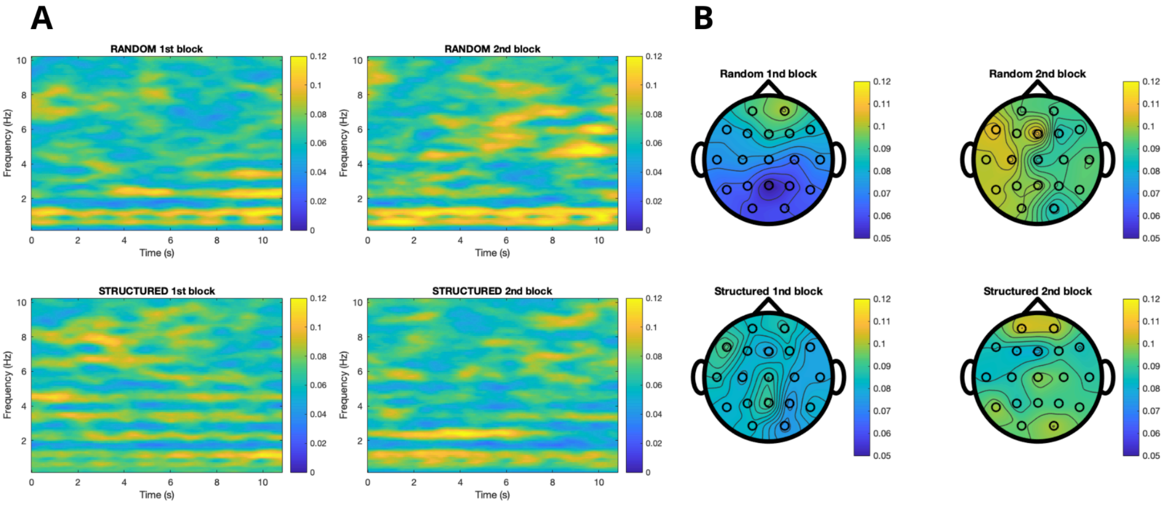

Figure 8 ***A:*** *ITPC values for Group C in the first and the second exposure block of each condition, averaged across the 7 analysed channels (Pz, O1, O2, T3, T4, T5, T6) and across all epochs.* ***B:*** *Localisation of the increased neural phase-locking activity in Group C at the word frequency (1.1 Hz).*

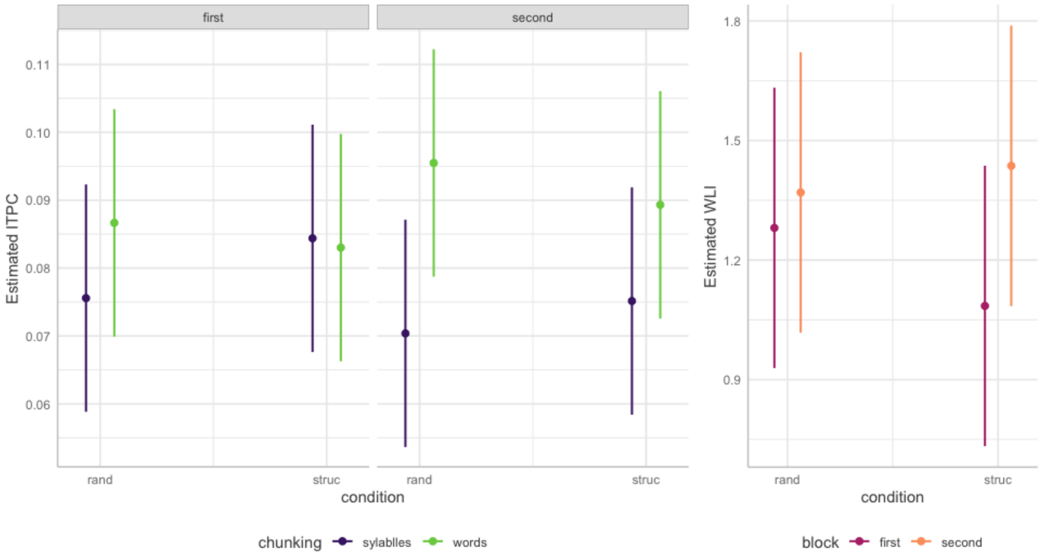

Figure 9 ***Left:*** *Estimated ITPC per condition, exposure block, and chunking rate for Group C.* ***Right:*** *Estimated WLI per condition and exposure block (means and 95% confidence intervals) for Group C.*

	Est	**SE**	**df**	**t**	**p**
(Intercept)	1.293	0.159	5.039	8.149	<0.001*
Condition (-rnd+str)	-0.032	0.037	266	-0.877	0.382
Block(-1+2)	0.11	0.037	266	2.995	0.003*
condition:block	0.066	0.037	266	1.787	0.075

Table 8: WLI fixed-effect model summary for Group C.

The linear mixed-effects model for the predicted WLI values show a significant intercept and a main effect of block (see *Figure 9 Right*). The model summary is shown in *Table 8*.

3.4.2 Behavioural Results

3.4.2.1 Rating Task

Table 10 shows the mean rating accuracy and mean rating score values across all participant groups. The mean rating accuracy of **Group A** was 44.86% (SD=15.30) and did not differ from chance with alpha 0.1 (t = -1.258, df = 13, p = 0.231, 95% conf. int. = 47.6–73.4%). In line with that, the mean rating score of -0.46 did not indicate evidence of statistical learning. Evidence of statistical learning would be suggested by *rating score ≥ 1*, with rating score = *3* indicating perfect sensitivity to the embedded structures (Batterink & Paller, 2017). Only one participant's rating score (equal to 3) indicated possible evidence of statistical learning which was supported by their above chance rating accuracy (88%). Running a separate test for non-words and for words showed that the mean familiarity score of the non-words (0.54) was numerically but not statistically higher than the mean familiarity score for words (0.43), and both were not different from chance. It is important to note the possible effect of the fixed trial order which might negatively affect the familiarity score of the word which is presented first (pseudo-word *tibufe*) due to participant's unfamiliarity with the task design and possible unreadiness (see *Table 9*).

The mean accuracy of **Group B** was at 60.5% (SD=12.29), and trended as different from chance, not reaching significance with our alpha of 0.05 (t = 2.090, df = 5, p = 0.091, 95% conf. int. = 36.1–53.7%). Numerically, the mean rating accuracy in Group B was higher than the mean rating accuracy in Group A. Three participants in Group B showed evidence of statistical learning (rating score ≥ 1) and rating accuracy above the chance level.

The mean rating accuracy of **Group C** was 47.8% (SD=16.30), not different from chance (t = -0.299, df = 4, p = 0.780, 95% conf. int. = 27.4–68.2%). Two participants in Group C showed evidence of statistical learning (rating score ≥ 1) and rating accuracy above the chance level. The fixed trial order did not appear to affect the results of Group B and C.

	Tibufe (W)	Pitugo (N)	Pabeku (W)	Lafobe (N)	Gotifo (N)	Golatu (W)	Dafopi (W)	Bepafe (N)
Group A	7.1%	42.9%	78.6%	42.9%	78.6%	64.3%	21.4%	21.4%
Group B	50%	66.7%	83.3%	33.3%	83.3%	66.7%	66.7%	33.3%
Group C	20%	80%	100%	20%	60%	20%	60%	20%

Table 9: Familiarity rating accuracy for pseudo-words (W) and non-words (N) across all participants.

	Participants showing effects of SL	Mean rating accuracy (%)	Mean rating score
Group A	1/14 (7.14%)	44.86 (SD=15.30, 95% CI=36.02-53.69)	-0.46 (SD=1.22)
Group B	3/7 (42.86%)	60.5 (SD=12.29, 95% CI=47.58-73.42)	0.83 (SD=0.98)
Group C	2/6 (33.3%)	47.8 (SD=16.30, 95% CI=27.37-68.23)	-0.17 (SD=1.30)

Table 10: Mean results of the post-exposure rating task for Group A, B, and C. Rating accuracy above 50% is regarded as an effect of SL, following Batterink and Paller's analysis (Batterink and Paller, 2017).

3.4.2.2 Lip Reading Skill

Participants' lip reading skill was assessed using the percentage of correctly answered trials. The number of response options in the lip reading test for isolated words was 9, which means that the chance level is 100/9 = 11.11%. The number of response options in the lip reading test for sentences was 6, which means that the chance level is 100/6 = 16.67%.

Participants in **Group A** were slightly more successful in the first part of the lip reading task. The mean success rate for lip-reading words was 48.57% and significantly above chance level (t = 7.825, df = 13, p < 0.001, 95% conf. int. = 38.23–58.91%). The mean lip-reading score for sentences was 30.61% and also significantly above chance (t = 2.705, df = 13, p = 0.018, 95% conf. int. = 19.47–41.75%).

Participants in **Group B** performed better than Group A, however, their performance was significantly better in the first part of the assessment. The mean success rate for lip read words was 71.67% and significantly above chance (t = 10.077, df = 5, p < 0.001, 95% conf. int. = 56.22–87.11%), whereas the lip-reading score for sentences was 30.95% and not different from chance (t = 1.664, df = 5, p = 0.157, 95% conf. int. = 8.89–53.02%).

In **Group C**, the mean success rate for lip read words was 62% and significantly above chance (t = 4.990, df = 4, p = 0.008, 95% conf. int. = 33.69–90.32%). The mean success rate for lip read sentences was 40% and was also significantly above chance (t = 3.335, df = 4, p = 0.029, 95% conf. int. = 20.57–59.43%). The mean accuracy of the lip read words and sentences for all the groups as well as the rating accuracy for the individual items can be seen in *Table 11* and *Table 12*.

	ryba	*slunce*	*prase*	*komín*	*osel*	*hrnec*	*hruška*	*koláč*	*balón*	*okno*	mean accuracy
Group A	42.86	28.57	50	64.29	42.86	14.29	14.29	85.71	92.86	57.14	48.57 (SD= 17.91)
Group B	66.67	33.33	66.67	100	66.67	33.33	66.67	100	100	83.33	71.67 (SD= 14.72)
	fish	*airplane*	*dolphin*	*chimney*	*donkey*	*saucepan*	*cherry*	*pancake*	*balloon*	*window*	mean accuracy
Group C	80	100	80	60	20	40	80	60	40	60	62 (SD= 22.80)

Table 11: Success rate (%) for lip read words and mean accuracy across all items and all participants.

	cs_s1	*cs_s2*	*cs_s3*	*cs_s4*	*cs_s5*	*cs_s6*	*cs_s7*	mean accuracy
Group A	14.29	14.29	42.86	50	28.57	21.43	42.86	30.61 (SD= 19.29)
Group B	50	0	33.33	66.67	33.33	16.67	16.67	30.95 (SD= 21.03)
	en_s1	*en_s2*	*en_s3*	*en_s4*	*en_s5*	*en_s6*	*en_s7*	mean accuracy
Group C	0	60	0	60	80	40	40	40 (SD= 15.65)

Table 12: Success rate (%) for lip read sentences and mean accuracy across all items and all participants.

3.4.3 Correlations Between the EEG and Behavioural Results

The Spearman's correlations between WLI and rating accuracy and WLI and lip-reading accuracy for **Group A** were ρ = 0.287 and ρ = -0.145 respectively. For **Group B**, the Spearman’s correlation between WLI and rating accuracy was ρ = -0.123 and the correlation between WLI and lip-reading accuracy was ρ = -0.551. For **Group C**, the Spearman's correlations between WLI and rating accuracy and WLI and lip-reading accuracy were ρ = 0.564 and ρ = -0.359 respectively.

4. DISCUSSION

The present research aimed to investigate the online process of statistical learning during visual silent speech processing, and evaluate the influence of one's language background on the learning process that occurs during speech perception. Statistical learning (SL) can be understood as a two-step process where regularities in one's environment are first uncovered and gradually acquired, then stored in one's memory. While there is evidence of statistical learning across modalities, SL in speech has typically been tested on auditory speech, with auditory cues potentially aiding the segmenting of the novel speech. The extent to which language learners exploit SL in speech without auditory cues, i.e. in silent visual speech (VS), thus remains unexplored. Using silent VS not only lets us investigate SL during perception of speech without the aid of auditory cues but also lets us compare processes underlying speech perception in different populations whose primary mode of communication differs. Furthermore, since familiarity with the input was shown to significantly improve the SL outcomes (Bulf, Johnson, & Valenza, 2011), the present research aimed to explore how the primary mode of communication and its modality affect the perception and potential learning during the silent VS.

The present study adapted the traditional auditory SL design to suit visual speech stimuli. During the experiment, participants were exposed to silent videos of two streams of continuous visual speech: a structured stream with repeating trisyllabic nonsense words and a random stream without structure. Word boundaries in the structured stream were cued by transitional probabilities (TPs) of syllable co-occurrence only, and no cues indicated word boundaries in the random stream. EEG monitored participants' neural activity, and ITPC was subsequently computed at the syllabic and word rate frequencies. The word-learning index (WLI) was calculated as ITPC at word frequency divided by ITPC at syllable frequency. A forced-choice behavioural task followed the exposure to compare EEG and behavioural results. Based on prior research, we **predicted** greater phase-locked activity and higher WLI in the structured stream, reflecting statistical learning.

We **hypothesised** that visual articulatory cues would suffice for neural synchronisation despite the absence of auditory cues since visual information from the articulatory lip movements is one of the factors shaping a phoneme's identity, known to aid significantly during speech perception and comprehension, and we **predicted** that detecting word structures would be influenced by participants' language background, namely their primary mode of communication and their native language. To test this hypothesis, we compared native speakers of Czech with normal hearing (Group A), native speakers of English with normal hearing (Group C), and deaf and hard-of-hearing adults who primarily use Czech sign language or rely significantly on lip reading during speech comprehension (Group B).

Our **results** demonstrated that statistical learning is a mechanism underlying speech segmentation, even in the absence of auditory and contextual cues, and that this process can be effectively captured online during the exposure phase using neuroimaging methods. Our results further support the idea that the language background of the learner, namely their primary mode of communication and their native language, influences both the neural processing of visual silent speech and the associated learning. The following sections will address the present findings in greater detail.

4.1 Neural Evidence of Statistical Learning in Silent Visual Speech

4.1.1 Inter-trial Phase Coherence and Word-learning Index as Markers of Statistical Learning

The present study investigated whether statistical learning during visual speech perception can be successfully measured online using the inter-trial phase coherence (ITPC) and the word learning index (WLI) as markers of the perceptual shift and increasing phase-locking at the word rate. Our analysis demonstrated an **increased phase-locking (ITPC) value at the word frequency** (i.e. the frequency at which the word-like structures appeared in the speech stream) in the structured condition compared to the random condition in Czech participants with normal hearing (**Group A**), which additionally significantly increased with exposure (see *Figure 5 Left*). This result suggests that the neural activity phase-aligned with the rate of the underlying word structures as the brain gradually discovered them. The increase of the phase-locked activity with exposure is crucial as it demonstrates the process of acquisition of the underlying structures on a neural level, a process that cannot be captured with post-test assessment. In other words, our results suggest that we were able to capture the process of learning, at least on a neural level, during the exposure phase. Apart from the increase of phase-locked activity at the word frequency, we observed a slight decrease in tracking at syllable frequency in the second block of the structured stream, which, together with the increase of ITCP at word frequency, illustrates the perceptual shift in processing as a result of statistical learning. The perceptual shift in processing observed when comparing the ITPC at the word and the syllable frequency can be further illustrated by the word learning index (WLI), which reflects the sensitivity to the underlying word structures. Using the visual articulatory cues only, the participants were able to subconsciously detect the word structures in the structured stream and their neural activity successfully aligned to their frequency. Since the phase-locking to the trisyllabic structures is significantly greater in the structured condition, the results cannot be explained by a general trisyllabic chunking strategy, which would lead to processing both streams similarly (see *Chapter 4.1.2*). **Our prediction (H1)** that the visual cues will be sufficient and the effects of statistical learning will be reflected during visual speech processing was thus borne out.

The results of Group A further show that the phase-locked activity at the word rate in the structured condition differs significantly from the phase-locked activity at the word rate in the random condition. Furthermore, a significant triple interaction between the condition, block, and frequency of the phase-locked activity supports the claim that the neural activity evoked by the structured stream shows gradual alignment to the word structures, which the neural activity evoked by the random stream does not. The **word-learning index** (WLI), which relates the ITPC value at the syllable rate to the ITPC value at the word rate and shows the likelihood of learning, was significantly greater in the structured condition, as shown in *Figure 5 Right*. The WLI value increased significantly with exposure over the course of the experiment, arguably as a result of the gradual process of detecting and acquiring the statistical structures, seemingly making WLI a potential neural marker of statistical learning. Our **second hypothesis (H2)** predicting greater effects of SL in the structured condition compared to the random condition was thus borne out as well. The increased effects of neural learning in the structured compared to the random condition mirror the results found in neural processing of known and unknown language, respectively. This indicates that the neural processing of auditory and visual speech differs based on its content and predictability.

4.1.2 Individual Differences in Processing Strategies

Previous literature on statistical and implicit learning shows that acquiring trisyllabic structures in a trisyllabic speech stream might be affected by individual processing strategies, referred to as chunking, which are constrained by the individual's cognitive capacity and reflect individual processing preferences rather than the computation of transitional probabilities. It is important to note that neither of these strategies confutes the existence of the other. On the contrary, these strategies may constitute a two-step speech processing strategy.

In the present data, the phase-locking values at the word rate frequency in the first block of the structured condition mirrored the values in the first block of the random condition more closely than in the second block. The significant differences in the frequency of the phase-locked activity in the second block of the structured stream and the second block of the random stream suggest that other processing strategies other than arbitrary chunking preferences are taking place. This assumption is further supported by the posttest evidence discussed in *Chapter 4.1.3*. In light of these results, it is possible to argue that the participants first use chunking strategies to process smaller chunks of the continuous speech stream, resulting in a similar neural response to the first blocks of the streams, upon which they then conduct the statistical analysis, as reflected in the neural response of the second exposure blocks. It is thus possible that chunking represents a vital stepping stone in the theory of statistical learning, perhaps due to cognitive load.

Furthermore, it is essential to note that additional individual differences, other than chunking preference, can affect the results of statistical learning. Cognitive load, limited cognitive capacity, fatigue, and lack of attention can all influence the SL outcome. Previous literature also shows that the synchronisation of neural oscillations and the stimuli does not stop simultaneously with the end of exposure, and could transcend to the other trial, or in this case, condition. If presented last, the random stream could thus be affected by the previous alignment of the neural oscillation and the structured stream. However, the order of conditions was not found significant in our data, therefore, we cannot make any conclusions about the possible carry-on effects across conditions.

4.2 Behavioural Assessment of Statistical Learning and Lip Reading Skills

4.2.1 Forced-Choice Word-Recognition Task

As discussed in *Chapter 4.1.1*, the present significant neuroimaging results for Group A suggest that statistical learning occurred during the exposure to silent visual speech. Despite the proposed evidence of successful statistical learning in the structured condition demonstrated by the ITPC and WLI values, the behavioural post-exposure word-recognition task alone **did not reveal any statistical learning effects**, as the reported rating accuracy did not exceed the chance level. Given the ability of previous studies to capture results of SL in an auditory speech in behavioural posttest assessment, it is possible that the modality in which the stimuli are presented affects the difficulty of storage and subsequent retrieval of the resulting statistical items. Since evidence shows that familiarity with the input affects the effectiveness of SL, it is not surprising that processing less familiar input, such as visual speech, might be more demanding than processing auditory speech, albeit novel.

However, surprisingly, even though the behavioural assessment alone did not show any effects of SL, the statistical analysis results show a positive correlation between the rating accuracy and the increasing WLI values. In other words, participants with the highest increase in the word-learning index throughout the experiment also show the highest rating accuracy. It is thus possible that the poor performance in a behavioural task of some participants who show neural results of SL is caused by cognitive limitations, such as memory limitations or cognitive load, resulting in chance-level results. Although our results show significantly greater effects of learning at the neural level in the structured condition compared to the random condition, it is essential to note that, as of yet, we cannot pinpoint how and when the effects are reflected in behaviour.

The present results indicate that the post-exposure behavioural methods, namely the forced-choice word-recognition task, might not be entirely suited for statistical learning assessment since they do not record the online process of gradual knowledge acquiring and instead assess only the successful storing and retrieving of the resulting items. Using behavioural tasks as the only method to assess the results of SL might lead to premature conclusions about absence of the SL effects. Furthermore, it is plausible that other factors, apart from cognitive abilities, affect the success rate of behavioural tasks. The effect of language background on the performance in the present behavioural task shall be discussed in *Chapter 4.3.2*.

4.2.2 Lip Reading Assessment

We hypothesised that visual articulatory cues would be sufficient to facilitate statistical learning and, as discussed in *Chapter 4.1.1*, our hypothesis proved correct. At the same time, we assumed that the lip reading skill of the participant would affect their readiness to attend to the articulatory cues and the ease with which the cues are processed. A higher lip reading skill could thus lead to a faster onset of the phase-locked neural activity or greater phase-locking. To test this assumption, we implemented a lip reading picture task to test the lip reading skill of the participants. Participants were asked to lipread individual words as well as sentences. Their performance in both parts of the test was judged separately.

The lip reading task accuracy of the participants with normal hearing (**Group A**) exceeded the chance level, even though they expressed their frustration over the difficulty of the task. Participants performed better at lip reading the individual words than at lip reading sentences, which is significantly more difficult. Furthermore, participants showed the best results at lip reading words with strong articulatory cues, such as /o/ in *komín, koláč,* and *balón*, and performed poorly at lip reading words with consonant clusters, such as *slunce, hrnec,* and *hruška*. The results of both parts of the assessment are shown in *Table 11* and *Table 12*.

The statistical analysis found a negative correlation between the lip reading accuracy and the increase of the word-learning index over the course of the experiment. The statistical results suggest that the participants who showed the highest lip reading skill also showed the lowest increase of the word-learning index. These results could be caused by the word-learning index increasing earlier, i.e. in the first block of the structured condition, and thus not increasing as dramatically in the second block of the experiment compared to the other participants. Alternatively, the conscious effort to decode the unknown novel speech based on one's lip reading experience in a different language might negatively affect the subconscious decoding of the statistical probabilities. Therefore, participants with less experience with lip reading would be less prone to implementing their past experience during the exposure to the novel silent speech. We assumed that participants with impaired hearing would have more experience with lip reading and therefore demonstrate a higher lip reading skill. These results shall be discussed in *Chapter 4.3.3*.

4.3 Language Background as a Factor Potentially Affecting SL in Silent Visual Speech

4.3.1 Effect of the Hearing Status

As discussed in the chapters above, Group A showed evidence of statistical learning facilitated by visual articulatory cues only. Given the evidence from previous literature suggesting that the familiarity with the input significantly affects neural tracking of speech (Bulf, Johnson, & Valenza, 2011), our aim was to investigate whether the primary mode of communication and its modality influence the tracking of silent VS and the potential learning. To investigate the possible effect of the primary language modality, the present chapter presents a comparison of the results between the participants with normal hearing (**Group A**) and impaired hearing (**Group B**).

4.3.1.1 EEG Results: Inter-Trial Phase Coherence and Word-Learning Index

Our results show that Group A demonstrated the highest **inter-trial phase coherence (ITPC) values at the word frequency** in the second block of the structured condition with the ITPC at the word rate increasing with exposure. In contrast, Group B shows the highest neural tracking at the word frequency in the first block of the experiment, with phase-locking scores in the random condition succeeding the scores in the structured condition. For Group B, the ITPC at the word rate subsequently decreases with exposure, however, it is important to note that even though the the $\text{ITPC}_{\text{words}}$ decreases slightly, the $\text{ITPC}_{\text{syllables}}$ decreases more rapidly. In the second block of both conditions, the phase-locking values are greater at the word frequency compared to the syllable frequency, as illustrated in *Figure 7 Left*. Even though we see a decrease of the ITPC values in the Group B data, it is important to note that the phase-locking in the first blocks of both conditions was significantly larger compared to Group A. Therefore, even though the values decrease, the resulting ITPC values in the second block align with the values measured in the second block of the structured condition for Group A. Ultimately, the ITPC model shows that Group A demonstrated a significant difference between both conditions and both blocks, however, Group B only showed significant difference between the blocks.

As for the **word-learning index** (WLI), Group A showed the highest measured WLI score in the second block of the structured condition and a decrease in the WLI score in the second block of the random condition. On the other hand, Group B showed a higher WLI score in the second block of both conditions compared to the first blocks.

Considering the observed effect of familiarity with the input on neural tracking, we assumed that hearing-impaired participants, whose primary mode of communication is either in the visual modality entirely or who highly depend on visual cues in communication, will show more robust neural tracking of the silent visual speech due to their familiarity with the silent articulatory cues. The phase-locking of the neural activity is indeed more robust in comparison to the participants with normal hearing, however, we do not observe a significant difference between the two conditions. The results thus seem to suggest that the hearing-impaired participants, unlike the hearing participants, were not able to uncover the covert statistical words, or to process them in real-time.

The present findings might result from the systematic differences between Czech and Czech Sign Language (CSL). Spoken Czech consists of hierarchically organised compositional units. Phonemes, syllables, words, and phrases can be processed simultaneously in real time by the listener. However, this is not the case of CSL, in which the smallest unit is not a phoneme, but a sign, which typically cannot be disassembled into smaller units. The different nature of both languages might thus lead to different processing strategies and the potential absence of chunking.

It is important to note that in comparison to Group A, Group B was profoundly more heterogeneous, comprising both deaf and hard-of-hearing individuals, and significantly smaller. We acknowledge that the heterogeneity and the size of the group poses a limitation to our interpretation of the present data, however, the present state of the group was inevitable due to the level of difficulty encountered when recruiting the deaf participants.

4.3.1.2 Behavioural Assessment Results

Group B performed better at the **forced-choice rating task** than Group A. However, similarly to the hearing participants, the performance of the hearing-impaired participants was at chance level and no indication of statistical learning was observed. As shown in *Table 9*, an interesting difference can be seen in the rating accuracy of the first presented item. The rating accuracy of the hearing participants was remarkably low, which was not the case for the hearing-impaired participants. This result could be caused by the immediate readiness of the hearing-impaired participants to process a language in the visual modality.

The statistical analysis revealed a negative correlation between the rating accuracy and the increase of the word-learning index, suggesting that participants with the highest rating accuracy demonstrated the smallest change in the WLI value. However, Group B's robust neural tracking in the first block of the experiment caused the majority of the WLI increase data to be negative. The smallest change therefore refers to the smallest decrease in WLI value. The participants with the smallest decrease in WLI scores also demonstrated the highest rating accuracy.

Our third and final hypothesis (H3) predicted that the language background of the participants will affect their processing of VS. We assumed that the language background of the participants, in this case their primary mode of communication, will affect the extent to which they are able to use the articulatory **lip-reading cues** to discriminate the linguistic units.

Performance of Group A participants in both parts of the lip reading test (lip reading of words and sentences) was above the chance level. On the other hand, the performance of the participants from Group B was only above chance at lip reading words and at chance level at lip reading sentences. The lack of above chance performance is presumably due to the CSL having a different syntactic and grammatical structure from Czech, which was used for the lip reading assessment. For this reason, the lip reading of sentences should not be used too strictly to compare the performance of normal-hearing and hearing-impaired adults due to the slight disadvantage the hearing-impaired participants had. Future studies focusing on lip reading skill assessment as a control should therefore take the difference between the two languages into consideration. Comparison of the two groups's performance can be seen in *Table 11* and *Table 12*.

Correlation between the lip reading accuracy and the WLI score increase showed a weak negative effect in Group A and a moderate negative effect in Group B, suggesting the same trend in both groups. In other words, the participants with lower lip reading skill demonstrated greater increase of the word-learning index, perhaps due to their subconscious visual speech processing not being influenced by their conscious effort to decipher the content of the speech stream. Furthermore, though the negative correlation demonstrates a relationship between poorer lip reading and the ability to learn more, the mode of learning is not specified. It can thus be learning of either verbal or non-verbal structures.

4.3.1.3 Localisation of the Increased Neural Activity

Our neuroimaging results demonstrated that, at least on neural level, learning took place during the perception of visual speech. However, the correlation with lip-reading ability prompted a discussion on how VS is processed and whether both groups process VS in the same way, namely as a linguistic stimulus without sound. Despite the limitations caused by the poor EEG localisation, we examined the brain areas that showed the greatest activity during speech exposure. We assumed that these active areas might be influenced by the participant's hearing and, by extent, by the brain's cross-modal plasticity. The activated brain areas cannot be identified with absolute certitude due to the limitations of the apparatus whose localisation precision cannot be compared to that of magnetic resonance imaging. For that reason, this work shall more confidently identify only larger areas of activation, such as entire lobes, rather than smaller brain regions.

Our results indicate that in **hearing participants,** the first block of the structured condition evoked slightly increased activity in the occipital lobe, known to be activated during visual stimulation. However, in the second block of the structured condition, the area of activation shifted towards the **right anterior temporal lobe**. Anterior temporal lobe can be responsible for semantic processing and semantic memory (Wong and Gallate, 2012), however, the right anterior temporal lobe is also strongly associated with processing of verbal versus non-verbal input (Rice, Lambon Ralph, & Hoffman, 2015). Therefore, while there is evidence of learning, it is unknown whether the learnt structures are processed as non-verbal or verbal linguistic stimuli. Furthermore, the temporal lobe is engaged in the facial recognition and processing, memory engagement, or processing of complex visual scenes (Kriegeskorte et al., 2007), which could lead to an increased activation of this area. Finally, we shall not disregard the poor localisation of the apparatus, and therefore should not exclude the possibility that the increased activity is localised in the right angular gyrus referred to as temporal visual speech area known for supporting the extraction of linguistic information from the lip movements (Bernstein et al., 2011). Activation of the right angular gyrus would be in line with the significant ITPC values in the second block of the structured condition.

Interestingly, although the greatest activation can be seen in the right temporal lobe, an increased activation can be seen in the left temporal lobe as well, suggesting that some participants demonstrated an increased activation of this area and processes the input bilaterally, in areas known for language processing. The increased activation of the temporal lobes in the second block of the condition and an increased activation of the occipital lobe in the first block of the condition, possibly indicates the processing of the visual information through the ventral stream (Goodale & Milner, 1992). Furthermore, given the differences between the activated areas in the first block of the random and the structured condition, it is possible that even in the first block of the structured condition, the brain detected certain patterns that required greater engagement of the occipital lobe. The activated brain regions are illustrated in *Figure 4 B.*

The areas of activation of the hearing participants and the hearing-impaired participants differ in several aspects. **Hearing-impaired participants** processed the first block of the structured condition in the occipital lobe similarly to the hearing participants, and in the right temporal lobe. However, in the second block of the structured condition, the increased neural activity shifts to the **left temporal lobe**, in contrast to the increased activity in the right temporal lobe observed in hearing participants. The increased activation of the right temporal lobe in the first block of the structured condition may suggest that learning and memory engagement occurred earlier in the hearing-impaired participants, which would be in line with the increased ITPC values in the first blocks. Furthermore, it might suggest that by the second block, the input was processed as verbal linguistic stimuli (Rice, Lambon Ralph, & Hoffman, 2015). The difference could be caused by the higher familiarity with the input modality, which is closer to the primary mode of communication of hearing-impaired participants, by the ability to tune into the visual modality of language more easily and quickly, or by the cross-modal brain plasticity affecting the primary functions of certain brain regions which otherwise would not be utilised due to an impairment.

Interestingly, the greatest increase in the neural activity could be observed in the first block of the random stream in the frontal and right temporal lobe. The activation was thus the same as in the hearing participants, only amplified. However, in the second block of both conditions we can see a shift of the increased activity to the left hemisphere, which was not observed in hearing participants. The activated brain regions are illustrated in *Figure 6 B.*

As discussed in *Chapter 2.2.2*, several prior studies described the brain activation during silent speech perception in hearing and deaf participants. The results differed based on the nature of the stimuli and the task. MacSweeney et al.'s 2002 study reported increased activation in the posterior regions and the right hippocampus in the deaf participants and increased activation in the left temporal lobe in the hearing participants. These results potentially mirror the results found for the first block of the structured condition. However, although MacSweeney et al. suggested activation of the right temporal lobe for the deaf participants and left temporal love for the hearing participants, our results demonstrate an opposite trend. The authors specified that the mentioned areas are activated by linguistic mouth movements, whereas the non-linguistic mouth movements are processed analogously by both groups in the occipital-temporal regions (MacSweeney et al., 2002), which reflects what we can see in the first block of the structured condition for both Group A and Group B.

On the other hand, Capek et al.'s 2008 study which implemented more complex linguistic stimuli (in comparison to numbers used in MacSweeney's study) reports greater activation of the left middle and posterior superior temporal cortex in the deaf participants, potentially due to the adaptability of the p-STS to the primary language modality. The left-lateralisation of the neural activity evoked by complex visual speech stimuli in deaf adults mentioned by Capek et al. aligns with the present data.

4.3.2 Effect of the Native Language

The present results suggest that the mechanisms involved in processing of visual silent speech might be affected by one's hearing status and their exposure to auditory input throughout their life. These results thus show that the primary mode of communication and the familiarity with the input can affect the neural tracking of VS. Our study further speculated that the native language can also affect neural tracking, albeit being in the same modality. **Group A**, whose native language was Czech, was therefore compared with **Group C**, whose native language was English. Despite sharing the primary mode of communication with Group A, Group C, i.e. hearing native speakers of English, was exposed to non-native articulation during exposure, which decreased their familiarity with the input. The following results thus describe the observed effect of non-native articulation on the neural perception of silent VS.

4.3.2.1 EEG Results

As discussed in the previous chapters, **Group A** demonstrated the highest ITPC values at the word frequency in the second block of the structured condition. The ITPC values at the word frequency increased with exposure in the structured condition, arguably as a result of SL. On the other hand, **Group C** showed that the ITPC at the word frequency increased with exposure in both conditions. There was no significant difference between the phase-locking frequency in the structured and the random block. Surprisingly, the highest ITPC values at the word rate were demonstrated in the second block of the random condition. The stronger phase-locking in the random condition can possibly result from the random stream mirroring the English articulation and syntax better than the structured condition. Alternatively, given the smaller size of the sample, condition order and individual processing preferences can also cause the present results. Future research would require a bigger sample to obtain unequivocal results. Exposing the participants to additional streams with and without their native articulation and the covert statistical structure could also shed light on the differences between the two groups.

Although the ITPC at the word frequency was the highest in the second block of the random condition, the WLI scores for Group C proved to be the highest in the second block of the structured condition, mirroring the results of Group A. However, yet again, it was only the main effect of the block that proved to be significant in the WLI model.

4.3.2.2 Behavioural Assessment Results

Group C outperformed Group A at the **forced-choice rating task**. Interestingly, it seems that similarly to Group B, Group C was not taken aback by the first presented item, even though their primary modality differs from the modality of the stimulus. It appears that the phonotactic probabilities of the native language affect the performance in the forced-choice task. The results show substantially lower mean rating accuracy for items *golatu* (Group A - 64.3%, Group C - 20%) and *lafobe* (Group A - 42.9%, Group C - 20%), but higher rating accuracy for the item *pitugo* (Group A - 42.9%, Group C - 80%). Although Group C outperformed Group A in the rating task, their rating accuracy was nonetheless at chance level. However, the statistical analyses revealed a moderate positive effect between the rating accuracy and the increase of the WLI score, i.e. a stronger effect than in the native Czech speakers. Participants with the highest increase of the WLI score also demonstrated the highest rating accuracy.

Similarly to Group A, the performance of Group C was above the chancel level in both parts of the **lip-reading test** (lip reading of words and sentences). Although the list of the presented items differed based on the native language of the group, comparison of the results show that Group C outperformed Group A slightly in both parts of the test. The performance at lip reading words proved to be overall balanced, with only the lexical item *donkey* showing substantially lower rating accuracy. The performance at lip reading sentences shows greater variability, however, the mean rating accuracy is still greater for the English native speakers. Despite the slightly better performance, we can conclude that the lip reading skill of participants from Group A and Group C is similar. The statistical analysis shows a moderate negative correlation between the lip reading accuracy and the WLI score increase, suggesting that the effect is stronger in Group C than in Group A. Participants with the highest lip reading accuracy thus demonstrated the lowest increase in the WLI score.

4.3.2.3 Localisation of the Increased Neural Activity

Our results demonstrated that the neural activation during VS perception differed in participants with different primary modes of communication. We further wanted to explore the possible differences in language background that could cause different neural activation and processing strategies. To achieve this, we examined two groups of participants with the same primary mode of communication but with a different native language (Czech and English).

As illustrated in *Figure 8 B*, the greatest phase-locking activity at the word frequency for **Group C** could be found in the **left temporal lobe** for the second block of the random condition, and in the **frontal lobe and occipital lobe** for the second block of the structured condition. Together with the ITPC data, these results might suggest that the random condition possibly visually simulated English natural speech better than the structured condition, since the neural activity is the greatest in the second block of the random condition and is located in the left hemisphere where arguably verbal structures are processed, and where silent speech is processed in hearing participants according to MacSweeney et al. (Rice, Lambon Ralph, & Hoffman, 2015).

The first block of the random condition shows increased activation on the right side of the frontal lobe, a result similar to that of Group A. On the other hand, the first block of the structured condition resulted in a slightly increased activity in the left temporal lobe, frontal lobe and the occipital lobe, however, the activity seemed to be less localised across the group resulting in less compelling evidence of increased activation. Futhermore, it is possible that the first block of the structured condition activated the **hippocampus**, however, due to the poor localisation of the EEG, this assumption cannot be deemed conclusive.

The present results suggest that participants exposed to non-native articulation processed the stimuli differently, however, it is important to note that the sample of English-speaking participants was significantly smaller than the sample of Czech-speaking participants. A larger sample of English-speaking participants would be needed to ensure full representativeness, and unfortunately, based on this sample, we cannot interpret the differences between Czech and English participants with unwavering confidence. The data show that increased activity in hearing Czech participants is more localised, while in English participants, it appears more arbitrary (i.e., increased activity could occur in different parts of the brain for each participant). This effect might be caused by the individual differences in speech processing, which are more pronounced in a smaller sample.

Based on the results obtained from the comparisons between Group A and B, and A and C, we can conclude that our **third hypothesis (H3)** predicting differences in processing of the silent visual speech due to one's language background was correct. Differences in processing can be seen in the frequency of phase-locking in the individual blocks and conditions but also in the brain regions activated by the exposure to visual speech.

5. CONCLUSION

The present study aimed to investigate the neural processes involved in the perception of silent visual speech. The neural activity of the participants was measured as they were exposed to a structured and unstructured (random) stream of novel visual speech. The aim of this work was to investigate whether the visual articulatory cues are sufficient to facilitate statistical learning (SL) (H1); whether greater evidence of statistical learning, a process underlying the perception of novel auditory speech, can be observed in the structured visual speech stream (H2); and finally whether the language background of the participants will affect the neural processing of the visual silent speech (H3). Our results showed neural, and partially also behavioural, evidence supporting each of the three working hypotheses.

The present results suggest that effects of statistical learning can be recorded during the exposure to silent visual speech using neuroimaging methods, which proved to be more sensitive to the effects of SL compared to behavioural methods which failed to detect the said effects. We were able to demonstrate the brain's ability to detect the underlying statistical patterns in structured artificial speech devoid of all auditory or contextual cues, as well as the synchronisation of the neural activity to the articulatory salient points in the speech stream. Our findings show that this synchronisation is gradual but almost instant and can be effectively captured by online neuroimaging methods. Although we were able to capture the real-time learning process, we were not able to detect evidence of statistical learning via a post-exposure behavioural task, leading us to believe that the effects of SL may not be reliably detected by post-exposure assessment alone.

We were able to detect a significant difference in processing of the structured and random speech stream in hearing adults (Group A) using the inter-trial phase coherence (ITPC), suggesting that the participants were able to detect the underlying statistical words through the process of inattentive learning. Furthermore, the present study was able to demonstrate substantial differences in neural processing of the silent visual speech caused by the primary mode of communication (spoken vs signed) and the native language (Czech vs English). The present results suggest that the neural processing of the VS in the hearing-impaired participants with a visual primary mode of communication differed from that of the hearing participants possibly due to the brain's neural adaptation and plasticity, or due to the different strategies used in real-time processing of a spoken and a sign language. Moreover, a non-native articulation of the stimulus and different phonotactics of one's native language can result in insufficient neural evidence of statistical learning as shown in the results of the English speaking participants (Dal Ben, Souza, & Hay, 2021). The effects of primary form of communication, hearing-impairment, and one's native language on statistical learning suggested that the familiarity with the modality and the articulation potentially aid the detection of the underlying patterns, however, a bigger sample would be needed to unequivocally show whether these effects are real.

The results of the present study can contribute to the ongoing research on neural tracking of speech and the mechanisms and cognitive functions underlying this process. The present study can help refine the methods used for assessing statistical learning, especially in research using challenging stimuli, such as visual speech input. Finally, it also contributes to the pool of research on silent speech processing, which may help understand how hearing-impaired individuals process speech and speech-related tasks and how this processing differs from that of individuals with normal hearing, representing a scaffolding for future EEG and fMRI studies.

6. REFERENCES

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line Assessment of Statistical Learning by Event-related Potentials. *Journal of cognitive neuroscience*, *20*(6), 952–964. https://doi.org/10.1162/jocn.2008.20058

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-Month-Old infants. *Psychological Science*, *9*(4), 321–324.

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78.

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex; a journal devoted to the study of the nervous system and behavior*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, *67*(1), 1–48.

Bear, H.L., Harvey, R.W., Theobald, BJ., Lan, Y. (2014). Which Phoneme-to-Viseme Maps Best Improve Visual-Only Computer Lip-Reading?. In: , *et al.* Advances in Visual Computing. ISVC 2014. Lecture Notes in Computer Science, vol 8888. Springer, Cham. https://doi.org/10.1007/978-3-319-14364-4_22

Bernstein, L. E., Jiang, J., Pantazis, D., Lu, Z. L., & Joshi, A. (2011). Visual phonetic processing localized using speech and nonspeech face gestures in video and point‐light displays. *Human brain mapping*, *32*(10), 1660-1676.

Binnie, C. A., Jackson, P. L., & Montgomery, A. A. (1976). Visual intelligibility of consonants: A lipreading screening test with implications for aural rehabilitation. *Journal of Speech & Hearing Disorders, 41*(4), 530–539. https://doi.org/10.1044/jshd.4104.530

Bourguignon, M., Baart, M., Kapnoula, E. C., & Molinaro, N. (2020). Lip-Reading Enables the Brain to Synthesize Auditory Features of Unknown Silent Speech. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *40*(5), 1053–1065. https://doi.org/10.1523/JNEUROSCI.1101-19.2019

Buiatti, M., Peña, M., & Dehaene‐Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage*, *44*(2), 509–519.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127-132.

Caldwell-Harris, C. et al. (2015). Factors influencing sensitivity to lexical tone in an artificial language: Implications for second language learning. Studies in Second Language Acquisition. 37. 335-357. 10.1017/S0272263114000849.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science (New York, N.Y.)*, *276*(5312), 593–596. https://doi.org/10.1126/science.276.5312.593

Capek, C. M., Macsweeney, M., Woll, B., Waters, D., McGuire, P. K., David, A. S., Brammer, M. J., & Campbell, R. (2008). Cortical circuits for silent speechreading in deaf and hearing people. *Neuropsychologia*, *46*(5), 1233–1241. https://doi.org/10.1016/j.neuropsychologia.2007.11.026

Christiansen, M. H. (2019). Implicit Statistical Learning: a tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468–481.

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: evidence from neural entrainment. *Psychological Science*, *31*(9), 1161–1173.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1980). Segmenting speech into words. *The Journal of the Acoustical Society of America*, *67*(4), 1323–1332. https://doi.org/10.1121/1.384185

Cooper, H., Holt, B. and Bowden, R. (2011) Sign Language Recognition. In: Moeslund, T.B., et al., Eds., Visual Analysis of Humans, Springer, London, 539-562. http://dx.doi.org/10.1007/978-0-85729-997-0_27

Crosse, M. J. et al. (2015) "Investigating the temporal dynamics of auditory cortical activation to silent lipreading," *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, Montpellier, France, 2015, pp. 308-311, doi: 10.1109/NER.2015.7146621.

Cunillera, T., Toro, J. M., Sebastián-Gallés, N., & Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain research*, *1123*(1), 168-178.

Dal Ben, R., De Hollanda Souza, D., & Hay, J. F. (2021). When statistics collide: The use of transitional and phonotactic probability cues to word boundaries. *Memory & Cognition*, *49*(7), 1300–1310.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods, 134*(1), 9-21.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158–164.

ELAN (Version 6.5) [Computer software]. (2024). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language?. *Psychological science*, *24*(7), 1243–1252. https://doi.org/10.1177/0956797612472207

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological bulletin*, *145*(12), 1128–1153. https://doi.org/10.1037/bul0000210

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20-25.

Hall, D. A., Fussell, C., & Summerfield, A. Q. (2005). Reading fluent speech from talking faces: typical brain networks and individual differences. *Journal of cognitive neuroscience*, *17*(6), 939–953. https://doi.org/10.1162/0898929054021175

Hayes, JR., Clark, HH. (1970). Experiments in the segmentation of an artificial speech analog. In: Hayes, JR., editor. Cognition and the Development of Language. New York: Wiley.

Kabdebon, C., Peña, M., Buiatti, M., & Dehaene‐Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language*, *148*, 25–36.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). “lmerTest Package: Tests in Linear Mixed Effects Models.” *Journal of Statistical Software*, *82*(13), 1–26.

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, *104*(51), 20600-20605.

Lüdecke, D. (2018). “ggeffects: Tidy Data Frames of Marginal Effects from Regression Models.” *Journal of Open Source Software*, *3*(26), 772.

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., Williams, S. C., Woll, B., & Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: an event-related fMRI study. *Neuroreport*, *11*(8), 1729–1733. https://doi.org/10.1097/00001756-200006050-00026

MacSweeney, M., Calvert, G. A., Campbell, R., McGuire, P. K., David, A. S., Williams, S. C., Woll, B., & Brammer, M. J. (2002). Speechreading circuits in people born deaf. *Neuropsychologia*, *40*(7), 801–807. https://doi.org/10.1016/s0028-3932(01)00180-4

MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A. S., McGuire, P. K., Calvert, G. A., & Brammer, M. J. (2004). Dissociating linguistic and nonlinguistic gestural communication in the brain. *NeuroImage*, *22*(4), 1605–1618. https://doi.org/10.1016/j.neuroimage.2004.03.015

McGurk, H., & MacDonald, J. B. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

Mohammed, T. E., MacSweeney, M., & Campbell, R. (2003). Developing the TAS: Individual differences in silent speechreading, reading and phonological awareness in deaf and hearing speechreaders. In *AVSP 2003-International Conference on Audio-Visual Speech Processing*.

Mohammed, T., Campbell, R., Macsweeney, M., Barry, F., & Coleman, M. (2006). Speechreading and its association with reading among deaf, hearing and dyslexic individuals. *Clinical linguistics & phonetics*, *20*(7-8), 621–630. https://doi.org/10.1080/02699200500266745

Morgan, J. L., & Saffran, J. R. (1995). Emerging Integration of Sequential and Suprasegmental Information in Preverbal Speech Segmentation. *Child Development*, *66*(4), 911–936. https://doi.org/10.2307/1131789

Muthukumaraswamy, S. D., Johnson, B. W., Gaetz, W. C., & Cheyne, D. O. (2006). Neural processing of observed oro-facial movements reflects multiple action encoding strategies in the human brain. *Brain research*, *1071*(1), 105–112. https://doi.org/10.1016/j.brainres.2005.11.053

Myers, B. R., Lense, M. D., & Gordon, R. L. (2019). Pushing the envelope: Developments in neural entrainment to speech and the biological underpinnings of prosody perception. *Brain sciences*, *9*(3), 70.

Obleser & Kayser, 2019

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife*, *5*, e14521. https://doi.org/10.7554/eLife.14521

Payne, B. R., Ng, S., Shantz, K., & Federmeier, K. D. (2020). Event-related brain potentials in multilingual language processing: The N's and P's. In *Psychology of learning and motivation* (Vol. 72, pp. 75-118). Academic Press.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in cognitive sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Podlipský, V.J., Chládková, K., Paillereau, N., Šimáčková, Š. Native variety influence on speech segmentation in a novel language. In Carlet, A. et al (Eds.) *Book of Abstracts. New Sounds 2022.* Barcelona, p. 133.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rice, G. E., Lambon Ralph, M. A., & Hoffman, P. (2015). The roles of left versus right anterior temporal lobes in conceptual knowledge: an ALE meta-analysis of 97 functional neuroimaging studies. *Cerebral Cortex*, *25*(11), 4374-4391.

Saffran, J. et al. (1996a). Word Segmentation: The Role of Distributional Cues. Journal of Memory and Language. 35 606-621.10.1006/jmla.1996.0032.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996b). Statistical learning by 8-Month-Old infants. *Science*, *274*(5294), 1926–1928.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature neuroscience*, *5*(7), 700–703. https://doi.org/10.1038/nn873

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

The MathWorks Inc. (2022). MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc.

Tune & Obleser (2022): Chapter on Neural Oscillations in Speech Perception. Retrieved from osf.io/6b7eg

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *Journal of cognitive neuroscience*, *21*(10), 1934–1945. https://doi.org/10.1162/jocn.2009.21131

Wong, C., & Gallate, J. (2012). The function of the anterior temporal lobe: a review of the empirical evidence. *Brain research*, *1449*, 94–116. https://doi.org/10.1016/j.brainres.2012.02.017

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4.

7. RESUMÉ

Předmětem této studie bylo zkoumání mechanismů, které zprostředkovávají zpracovávání a percepci neznámé tiché řeči, tedy řeči, která se skládá pouze z artikulačních pohybů a neobsahuje žádná zvuková vodítka. Tato práce se zaměřila především na procesy segmentace a statistického učení, které jsou neodmyslitelně spojeny s úspěšnou percepcí nepřerušované řeči, a to díky posluchačově schopnosti podvědomě pracovat s pravidelností v řečovém inputu, a tak odhalovat hranice slov. Předchozí literatura se však zabývala pouze přítomností statistického učení při zpracovávání auditivní řeči, a tak vystala otázka, zda statistické učení napomáhá percepci řeči obecně, bez ohledu na její modalitu. Na základě této výzkumné otázky byly formulovány tři hlavní **hypotézy**, a to sice: (H1) tiché artikulační pohyby tiché řeči budou dostatečným vodítkem pro statistické učení, (H2) výsledky statistického učení budou dominantnější ve strukturovaném proudu tiché řeči, který obsahuje skrytou statistickou strukturu, (H3) výsledky statistického učení budou ovlivněny jazykovým pozadím účastníka, a to konkrétně primárním způsobem komunikace, tzn. mluveným či znakovým, a mateřským jazykem. Nástin problematiky, kterou se práce zabývá, výzkumný cíl a hypotézy jsou uvedeny v úvodní části práce.

Teoretická kapitola této práce shrnuje předchozí poznatky v oblasti statistického učení a implicitního učení, neurálního trackování řeči a zpracování tiché řeči. První část teoretické kapitoly shrnuje studie zabývající se statistickým učením. Významnými studiemi pro tuto práci jsou zejména studie autorů Saffran a kol. (1996a, 1996b) a Batterink a Paller (2017). Studie Saffran a kol. prokázaly, že dospělí i kojenci používají prosodická a statistická vodítka k tomu, aby identifikovali hranice slov v neznámé nepřerušované řeči. Již osmiměsíční kojenci jsou schopni identifikovat hranice slov v nepřerušovaném toku čtyř opakujících se tříslabičných slov, a to pomocí podvědomého výpočtu pravděpodobnosti souvýskytu slabik, která je menší vně slova (.33) než uvnitř slova (1) (Saffran a kol., 1996a, 1996b). Autoři Batterink a Paller poukazují ve své studii na to, že behaviorální metody, které se běžně používají při zkoumání statistického učení během percepce řeči, nemusí být ideální metodou testování, jelikož se zaměřují pouze na jeden komponent statistického učení, a to na úspěšné uložení statistických slov do paměti a jejich zpětné vybavování (Batterink a Paller, 2017). K monitorování prvního komponentu statistického učení, tedy postupnému získávání znalostí na základě statistických vodítek, autoři navrhují použití neurosnímacích metod. Používání neurosnímacích metod k monitorování statistického učení během percepce řeči dokazuje, že tuto metodu lze efektivně použít k zachycení výsledků učení a to pomocí výpočtu tzn. *inter-trial phase coherence* či *ERPs (event-related potentials)*. Tyto metody jsou často používané při zkoumání tzv. *neurálního trackování řeči (neural speech tracking)*, při kterém je měřena synchronizace neurální aktivity s vnějším řečovým stimulem. První část teoretické části také v neposlední řadě doplňuje zjištění, že statistické učení je obecným principem, který se nachází i ve vizuální modalitě, jak dokazují studie, které se zabývají podvědomým učením statistických pravidelností v řadě tvarů a obrazců (Fiser et al., 2001). Úspěšnost statistického učení je však ovlivněna několika faktory, a to například familiaritou se stimulem a jeho formou (Frost a kol., 2019).

Druhá část teoretické kapitoly shrnuje předchozí poznatky o zpracovávání tiché řeči, která byla zkoumána zejména v kontextu lokalizace zvýšené neurální aktivity při percepci tiché řeči. Statistické učení v rámci této percepce nebylo dosud prozkoumáno. Jelikož konfigurace artikulátorů je jedním z prvků, který formuje identitu fonému (či v tomto případě vizému), tichá řeč obsahuje nezbytné vizuální informace, díky kterým může být rozklíčován její lingvistický obsah. Předchozí studie ukazují, že tichá řeč se liší od jiných tichých pohybů artikulátorů tím, jak se neurálně zpracovávaná, a to právě díky tomu, že obsahuje lingivistické informace (Muthukumaraswamy a kol., 2006). Schopnost rozklíčovat tyto lingvistické informace je však ovlivněna schopností odezírat a možná také obecnou schopností číst (Mohammed a kol., 2006).

Několik předchozích studií se zabývalo tím, jakou změnu v neurální aktivitě evokuje tichá řeč. Bylo zjištěno, že tichá řeč aktivuje sluchový kortex a evokuje podobnou neurální aktivitu jako řeč auditivní (Bourguignon a kol., 2020). Zároveň bylo zjištěno, že lokalizace zvýšené neurální aktivity evokované percepcí tiché řeči souvisí se schopností slyšet, jelikož absence sluchu mění funkce některých částí mozku (MacSweeney a kol., 2002). Tichá řeč tedy aktivuje jiné části mozku u neslyšících dospělých než u slyšících dospělých. Zároveň také u neslyšících dospělých dochází k silnější aktivaci, což může být způsobeno plasticitou mozku, který se adaptuje na absenci sluchu a mění funkce svých částí, které slouží ke zpracovávání auditivních informací (Capek a kol., 2008).

V **metodologické části** práce jsou připomenuty výzkumné cíle a hypotézy a zároveň je podrobně popsána metoda, která byla použita k dosažení těchto cílů. První část metodologické kapitoly popisuje, jací účastníci byli zahrnuti do této studie. Do této studie bylo zahrnuto 25 dospělých účastníků, kteří byly rozděleni do tří skupin na základě jejich primárního způsobu komunikace (auditivní vs vizuální) a jejich mateřského jazyka (čeština vs angličtina). Skupina A se tedy skládala z českých rodilých mluvčích s normálním sluchem, skupina B se skládala z neslyšících mluvčích českého znakového jazyka a nedoslýchavých dospělých, kteří se významně spoléhali na vizuální vodítka v řeči, a skupina C se skládala z anglických rodilých mluvčích s normálním sluchem. Všechny skupiny byly vystaveny všem částem experimentu.

V další části metodologické kapitoly jsou popsány stimuly, které byly použity v této studii. Tato práce adaptovala set stimulů použitých ve studii autorů Batterink a Paller (2017). Jedná se o typické stimuly, které jsou nejčastěji používané při výzkumu statistického učení při percepci řeči a které vycházejí ze stimulů použitých ve studii Saffran a kol. (1996a, 1996b). Tyto stimuly byly dále upraveny tak, aby všechny vizémy byly vizuálně rozlišitelné a splňovaly tak záměr této práce. Výslednými stimuly byly dva nepřerušované proudy tiché řeči. První proud tiché řeči byl strukturovaný a obsahoval čtyři opakující se tříslabičná slova, jejichž hranice mohly být detekovány jen na základě statistických informací, konkrétně na základě pravděpodobnosti souvýskytu slabik vně a uvnitř slova. Druhý proud tiché řeči byl náhodným tokem slabik bez jakékoliv statistické struktury. Proudy řeči byly nahrány rodilou mluvčí češtiny. Následně byl z nahrávek odebrán zvuk, čímž vznikla videa s tichou řečí.

Následující podkapitola ilustruje průběh experimentu. Účastníci experimentu byly vystaveni dvou tokům tiché řeči, přičemž byla pomocí EEG měřena jejich neurální aktivita. Po sledování proudu tiché řeči byl účastníkům administrován hodnotící test, v němž měli ohodnotit familiaritu jednotlivých slov, která se v toku řeči objevila či neobjevila. Tento test byl administrován se cílem testovat efekty statistického učení behaviorálně. Následně byl administrován test odezírání, který byl vytvořen pro účely této práce a který se skládal ze dvou částí. V první části účastníci odezírali izolovaná slova ve svém mateřském jazyce, v druhé části pak odezírali věty ve svém L1. Správné odpovědi vybírali účastníci z obrázkových možností.

V další části metodologické kapitoly je popsáno, jak byla získaná data analyzovaná. Synchronizace nahrané neurální aktivity a řečového stimulu byla vypočítána na frekvenci slabik (3,3 Hz) a slov (1,1 Hz) pomocí tzv. *inter-trial phase coherence*. Neurální aktivita synchronizovaná s frekvencí výskytu skrytých tříslabičných slov by naznačovala, že účastníci byli schopni detekovat hranice těchto slov pomocí statistického učení. Obě podmínky (strukturovaná a náhodná) byly rozděleny na dva bloky, aby bylo možné analyzovat změnu synchronizace v průběhu experimentu. Behaviorální hodnotící test byl analyzován na základě správnosti jednotlivých odpovědí. Vypočítáno bylo tzn. *rating score* a *rating accuracy* podle metody použité ve studii Batterink a Paller (2017). Test odezírání byl také obdobně hodnocen na základě správnosti odpovědí. Následně byla provedena statistická analýza, která ověřovala signifikanci jednotlivých výsledků. V posledním úseku metodologické části této práce byly popsány EEG výsledky, výsledky behaviorálních testů a jejich korelace.

V následující kapitole diplomové práce byly podrobně diskutovány získané výsledky. **Diskuze** je rozdělena do několika podkapitol na základě diskutovaných hypotéz. První podkapitola se zabývá EEG výsledky skupiny A, tedy rodilými mluvčími češtiny s normálním sluchem. U účastníků s normálním sluchem se synchronizace neurální aktivity na frekvenci slov zvyšovala během experimentu a dosáhla nejvyšších jednotek ve druhém bloku strukturované podmínky, což naznačuje, že účastníci byli schopni odhalit statistické pravidelnosti v proudu tiché řeči a to jen na základě artikulačních vodítek. Tyto výsledky jsou v souladu s naší první hypotézou, která tvrdí, že (H1) tiché artikulační pohyby tiché řeči budou dostatečným vodítkem pro statistické učení, a také s druhou hypotézou, která tvrdí, že (H2) výsledky statistického učení budou dominantnější ve strukturovaném proudu tiché řeči, který obsahuje skrytou statistickou strukturu.

Druhá podkapitola diskuze se zabývá behaviorálními výsledky skupiny A. Přestože EEG výsledky vykazují výsledky statistického učení, hodnotící test statistické učení neodhalil, což může naznačovat nedostatky behaviorálních metod testování použitých v některých studiích. Test odezírání odhalil, že i účastníci, jejichž mateřská řečová modalita nebyla vizuální, dokázali úspěšně odezírat slova a věty ve svém mateřském jazyce.

Třetí podkapitola se zaměřuje na srovnání výsledků jednotlivých skupin. Tato podkapitola se nejprve věnuje vlivu sluchu na zpracovávání tiché řeči a poté vlivu mateřského jazyka. Zpracování tiché řeči bylo skutečně rozdílné u slyšících a neslyšících účastníků a také u rodilých mluvčích češtiny a angličtiny. Lokalizace zvýšené neurální aktivity ukazuje, že účastníci ve skupině A zpracovávali druhý blok strukturované podmínky primárně v pravé hemisféře, zatímco účastníci ve skupině B jej zpracovávali primárně v levé hemisféře. Zároveň u účastníků ve skupině B nevidíme signifikantní rozdíl mezi zpracováváním jednotlivých proudů řeči, vidíme však zvětšování synchronizace neurální aktivity v průběhu experimentu. U účastníků ve skupině C vidíme silnější synchronizaci v náhodném toku řeči, což může být způsobeno mateřským jazykem těchto účastníků, jehož struktura je možná lépe reflektovaná v náhodném toku řeči než ve tříslabičném strukturovaném toku řeči. Zajímavé je, že zvýšená neurální aktivita u skupiny C byla lokalizována v levé hemisféře, podobně jako tomu bylo u neslyšících a nedoslýchavých účastníků. Tyto výsledky ukazují, že statistické učení může být ovlivněno familiaritou s formou inputu, fonotaktikou rodného jazyka a plasticitou mozku, což je v souladu s naší třetí pracovní hypotézou, která tvrdí, že (H3) výsledky statistického učení budou ovlivněny jazykovým pozadím účastníka, a to konkrétně primárním způsobem komunikace, tzn. mluveným či znakovým, a mateřským jazykem.

Závěrečná kapitola shrnuje celou práci, včetně výzkumného záměru, dosažených výsledků a závěrů, které lze z těchto výsledků vyvodit. Kromě toho se kapitola zaměřuje na vědecký přínos této práce. Tato studie svými poznatky přispívá k výzkumu neurálního trackování řeči, mechanismů a kognitivních funkcí, které tomuto procesu napomáhají. Přispívá také k porozumění, jaké metody testování jsou vhodné či nevhodné pro výzkum statistického učení při percepci řeči, zejména pokud jde o složité jazykové stimuly, jako je tichá řeč. Práce rovněž přispívá k pochopení zpracovávání tiché řeči a poskytuje cenné informace o tom, jak jedinci s narušeným sluchem zpracovávají řeč ve srovnání s jedinci s normálním sluchem. Tyto poznatky mohou sloužit jako základ pro další studie zkoumající percepci řeči u neslyšících osob pomocí EEG a fMRI, a tím přispět k dalšímu rozvoji této oblasti výzkumu.

8. APPENDIX

8.1 Excerpts of the Transcripts for Both Conditions

Structured Condition:

ti bu fe pa be ku go la tu ti bu fe pa be ku ti bu fe go la tu pa
be ku go la tu da fo pi go la tu pa be ku da fo pi go la tu da fo
pi pa be ku go la tu ti bu fe go la tu ti bu fe go la tu ti bu fe
go la tu pa be ku ti bu fe pa be ku go la tu pa be ku ti bu fe go
la tu da fo pi ti bu fe go la tu pa be ku go la tu ti bu fe pa be
ku ti bu fe go la tu ti bu fe pa be ku da fo pi ti bu fe go la tu
ti bu fe pa be ku da fo pi pa be ku go la tu pa be ku ti bu fe go
la tu pa be ku ti bu fe go la tu da fo pi ti bu fe da fo pi go la
tu pa be ku da fo pi ti bu fe da fo pi pa be ku ti bu fe go la tu
pa be ku go la tu pa be ku da fo pi pa be ku ti bu fe go la tu pa
be ku go la tu pa be ku da fo pi ti bu fe da fo pi pa be ku da fo
pi ti bu fe pa be ku ti bu fe go la tu pa be ku ti bu fe da fo pi
ti bu fe pa be ku da fo pi ti bu fe go la tu da fo pi ti bu fe go
la tu da fo pi ti bu fe go la tu ti bu fe pa be ku go la tu da fo
pi ti bu fe da fo pi go la tu da fo pi pa be ku da fo pi pa be ku
ti bu fe go la tu da fo pi pa be ku ti bu fe da fo pi go la tu da
fo pi go la tu ti bu fe pa be ku da fo pi pa be ku da fo pi go la
tu da fo pi pa be ku da fo pi go la tu da fo pi ti bu fe pa be ku
da fo pi ti bu fe da fo pi ti bu fe go la tu pa be ku da fo pi go
la tu pa be ku da fo pi ti bu fe pa be ku ti bu fe pa be ku ti bu
fe da fo pi ti bu fe go la tu ti bu fe pa be ku go la tu pa be ku
da fo pi pa be ku go la tu pa be ku ti bu fe da fo pi ti bu fe da
fo pi ti bu fe da fo pi ti bu fe pa be ku go la tu da fo pi go la
tu da fo pi go la tu da fo pi ti bu fe go la tu ti bu fe go la tu
ti bu fe pa be ku go la tu pa be ku go la tu da fo pi go la tu da
fo pi ti bu fe da fo pi ti bu fe go la tu ti bu fe da fo pi pa be
ku da fo pi pa be ku go la tu da fo pi go la tu pa be ku da fo pi
pa be ku da fo pi go la tu pa be ku da fo pi pa be ku da fo pi go
la tu pa be ku da fo pi go la tu ti bu fe da fo pi ti bu fe da fo
pi pa be ku go la ...

Random Condition:

go tu be la pi tu go bu pi la pa da la bu pa fo la tu ku be tu ku
pi la pa la go pi fe bu ku go ku tu la fo be tu pi la ku da tu fo
go da be ti be da pi fe go da ku be ti bu pi fo pi da fe fo la pi
la tu bu pa fe tu ti pi fe tu fo la da fe tu pa pi go pa tu la be
pi pa tu da go be go fe tu da tu fo tu ku da ti tu pi fo go pa be
fo fe ku tu pi fo la da fo bu be pa fe fo la da pi fo la pi fo pa
fe pi ti pa tu pi be go ku fo la ku be la fe pi da pa fe la ku tu
la ku bu pa pi ku la da ti bu pa ti tu go bu pi ku go ti pa fo ti
pa bu da pa la ku ti tu go pi go fo be pa ku pi ti pi bu be bu fe

```
fo fe fo bu fe bu be pa be pa bu tu bu tu fo ku pi bu fe la da go
tu bu fe pa da pi ti tu fe go bu ku pi be go la fo da fe ku pi da
fo la pa la fo go pi go fe fo be fo bu ti da pa la pi ku da pi tu
ti fe go bu fo be fe ku pa ku da ti go ti bu pa be fo la ku la ku
fe pa la go fo da pi fo ku ti la pi da pi ti fo be pa pi da fo ku
fe tu bu da fe be la fo la ti bu fo ti da be fe ti be fo be ku tu
be pi da tu pi tu go be go da tu ti pa la pi ti da tu be fo be ku
da fe tu ti fe go bu da be fe be ti la ku bu fo ku fo tu la ti da
fe fo go ti da be bu la go ti tu be fe la ti pa tu la fe be pa ku
be la fe tu fo fe ku tu da pi da go bu go fo ti fe ti be go da fe
ti be fe go fe tu pa la tu fo pa la ku fe la da fo pa da pa la ti
bu fe be bu pa go la pi tu ku be ku da pi fe da be fo be pi fe bu
fo ku go bu ti fo fe pi bu ku ti pa da ku la ku go fo ti fo bu ku
da bu be fe pa ti pi la bu be da fe da ti be bu tu fe pi fo pa ku
fe go ti fe da pi ti ku be fo ti pi ku pi be la go fe pi ku ti pi
ti be da be go tu fe tu go ti fo ti tu pi pa la pi fe pa pi fe pi
tu ku fe ku da fe go tu ti fo pi pa ti be da fo ti pi pa fe go be
la pa fo ...
```

8.2 Forced-Choice Task

Presented order: tibufe, pitugo, tuda, pabeku, bufe, lafobe, gotifo, pabe, golatu, fego, dafopi, bepafe

Pseudo-words: pabeku, tibufe, golatu, dafopi
Part-words: pabe, bufe
Non-words: pitugo, adobe, gotifo, bepafe, tuda, fego

8.3 Lip-Reading Test

<u>Czech version</u>:

Word-targets: ryba, slunce, prase, komín, osel, hrnec, hruška, koláč, balón, okno

Sentence-targets:

cs_s1: Žena má modré kalhoty.
cs_s2: Mladá žena pere červené triko.
cs_s3: Žena má v ruce činku.
cs_s4: Na stole je ryba.
cs_s5: U domu je šedá ovce.
cs_s6: Dívka venčí psa v lese.
cs_s7: Muž a dvě ženy sedí u vlaku.

English version:

Word-targets: fish, airplane, dolphin, chimney, donkey, saucepan, cherry, pancake, balloon, window

Sentence-targets:

en_s1: The woman is wearing blue trousers.
en_s2: The young woman is washing a red T-shirt.
en_s3: The woman is holding a dumbbell.
en_s4: There is a fish on the table.
en_s5: A grey sheep is standing next to the house.
en_s6: The girl is walking her dog in the forest.
en_s7: A man and two women are sitting near the train.