Learning of Linguistic Tonal Cues Follow the same Learning Principle as for other Linguistic Cues

Elin Varissara Arnholm & Julia Lorgen Hartvigsen
Candidate numbers: 126 & 120

Master Thesis
Master's Programme in Health Sciences
Logopedics
Department of Biological and Medical Psychology
Faculty of Psychology

Spring 2018

Preface

As two previous linguistic students, we have always been very interested in language, language learning, and second language learning. We both had a common interest in language impairments, and a goal of studying speech language therapy. We believe that greater insight in language learning processes is important for understanding treatment of language disabilities. Our passion for language learning theories lead us to the field of statistical learning. After consulting with our supervisor, his passion for this field inspired us to conduct this study. Since we were interested in an experimental research design, this topic fit us well. By using the advantage of one of us being a Thai-Norwegian bilingual, we could construct an experiment using Thai as stimuli.

We want to thank our supervisor, Arve Asbjørnsen, for his passion for statistical learning, which in turn lead us to be inspired and engaged in this topic as well. Without his guidance, none of this would have been possible.

We would also like to thank Theeraporn Ratitamkul, for being a great help in constructing the stimulus material, as well as recording the audio clips. We appreciate her valuable work and collaboration. Without her competence, our execution of the language learning experiment would not have gone as well as it did.

We thank Alexander Grøndalen Kristiansen and Erling Wogn-Henriksen, for their collaboration. Collaborating with them was a great advantage, and they were very helpful throughout the process.

At last, we want to thank all participants participating in our study.

Table of contents

Sammendrag

Denne studien undersøker statistisk læring i et naturlig tonespråk, samt om eksekutive funksjoner har en innvirkning på språklæring i en slik læringssituasjon. I tillegg til dette, undersøker studien i hvilken grad språklæringen skjer ved bevissthet. De 40 voksne som deltok i studien ble eksponert for setninger på thai i et språklæringseksperiment. Et selvrapporteringsskjema ble brukt for å undersøke deltakernes bevissthet (eller fraværet av bevissthet) om læringen. Videre ble de testet med fire kognitive tester som måler oppmerksomhet, inhibisjon, verbal flyt og arbeidsminne. Resultatene viste at voksne var i stand til å skille familiariserte ord fra nye ord, i tillegg til å ha en preferanse for tonale mønstre. Resultatene fra selvrapporteringsskjemaet viste ingen generelle sammenhenger mellom svarene de avga på spørreskjemaet, og hvordan deltakerne gjorde det under språklæringseksperimentet. Det var heller ingen systematiske korrelasjoner mellom eksekutive funksjoner og resultatene av læringseksperimentet. Resultatene støtter tidligere funn i studier på statistisk læring, som viser at deltakere er i stand til å skille familiariserte ord fra nye ord.

Abstract

This study investigated statistical learning in a natural tonal language, and whether executive functions has an impact on language learning in such learning conditions. In addition, the study investigates whether learning happens with awareness. The 40 adults that participated were exposed for Thai sentences in a language learning experiment. A self-report questionnaire was administered to examine the participants' awareness. They were then tested with four cognitive tests which measures attention, inhibition, verbal fluency and working memory. The results show that adults are able to discriminate familiarized words from novel words, in addition to having a preference for tonal cues. The results from the self-report questionnaire revealed that those who had detected certain patterns related to tone did not overall perform better in the language learning experiment than those who did not. There were no systematic correlations between executive functions and the results from the language learning experiment. The results support previous statistical learning studies which show that participants are able to discriminate familiarized word from novel words.

Statistical Learning of Tonal Cues

Theoretical Approaches to Language

How is language acquired? The debate about language acquisition has traditionally had two main views (Lidz & Gagliardi, 2015). The contrast between these views, is the distinction between those who view the process of language acquisition as similar to the acquisition of other complex skills, and those who assume that humans have an innate language module, or universal grammar (Berggreen & Tenfjord, 1999). The latter we refer to as nativism. Nativists consider language acquisition as domain specific. As a nativist, Chomsky view language as a faculty that is a particular component of the human mind (Chomsky, 1986). He and others argue that language is too complex to be acquired by general learning mechanisms and suggest that humans are born with what they call universal grammar (Arciuli & Torkildsen, 2012). Very briefly explained, universal grammar is a theory involving that some structural rules are innate in humans from birth, and it is not dependent of any sensory experience (Lidz & Gagliardi, 2015). Chomsky is also known for describing a domain specific language acquisition device (LAD), by arguing that all children share the same internal constraints in grammar learning, and that children only need to hear correct utterances to learn, rather than be corrected by adults (Behme & Deacon, 2008). Chomsky emphasizes that children acquire language fast and with precision, and that LAD plays a role in this.

The former view is focused around the language-input, and cognitive capacities, and claims humans can generalize patterns beyond the input, and also across domains (Lidz & Gagliardi, 2015). According to this view, humans are not born with a universal grammar, but rather domain general cognitive capacities.

Most of the research on implicit and statistical learning is empiricist, rather than nativist, oriented (Rebuschat, 2015). Rebuschat (2015) presents implicit learning, statistical learning and second language acquisition as three approaches to the same phenomenon. Implicit learning is a fundamental feature of human cognition, and it is the process of acquiring unconscious (implicit) knowledge (Rebuschat, 2015). Implicit learning is in other words incidental, and non-intentional. In contrast, explicit learning is awareness of what one is learning. An example of explicit knowledge is knowledge of grammatical rules in a foreign language. For instance, you can learn grammatical rules of a language, with or without having the intention to learn how to actually speak the language. An example of implicit knowledge is grammatical rules you unconsciously know in your first language but cannot explain to a foreigner. You know when a sentence is grammatically incorrect, but you may often experience that it is difficult to explain why.

Statistical learning is an approach to language acquisition, that suggests that humans can use statistical properties of linguistic input to detect patterns in language (Saffran, 2003). These include sound patterns associated with words and grammar. It is a theory harbouring the idea that understanding and prediction is the main goal of learning (Armstrong, Frost, & Christiansen, 2017). The thought is that there are statistical probabilities in language and our brain unconsciously process these probabilities when we hear language (Erickson & Thiessen, 2015). Research on statistical learning and implicit learning target how we acquire information from the environment. As mentioned, it is argued that implicit learning, statistical learning and second language acquisition are three approaches to one phenomenon. It has been suggested to combine the terms of implicit learning and statistical learning to implicit statistical learning (Rebuschat, 2015).

Krashen's monitor model implies a similar distinction, regarding second language acquisition. He distinguishes acquisition and learning as two separate processes, and consider them qualitatively different from one another (Krashen, 1981). According to Krashen, acquisition is an unconscious process. Acquisition is a result of meaningful communication, and he considers this process as the same process that takes place when children acquire their first language. This is an opposition to learning, which he considers as a conscious process (Krashen, 1981). Learning can take place in a classroom situation, where the learners learn grammatical rules of a language. The results of learning is that the learner gain knowledge about the language (Berggreen & Tenfjord, 1999).

Krashen's learning theory comprise explicit processing of data, intent to learn purposefully, explicit instruction, rule practice and many other behaviours. As such, the perspective is wider than in the generative theory, where explicit learning is only about what takes place during processing (VanPatten & Rothman, 2015). Rebuschat argues that implicit knowledge might be retained longer and more easily than explicit knowledge (Rebuschat, 2015).

In second language acquisition we also find the implicit-explicit learning debate. Cognitive second language acquisition theories focus on mental processes regarding language acquisition. In this theory, second language acquisition can be understood based on how the human brain process and acquire new information and skills. The brain processes all types of information, and linguistic data is one type of data, among many others (Berggreen & Tenfjord, 1999). We can compare language acquisition to the acquisition of other skills. Cognitive second language acquisition theories emphasize the formation of hypothesis, hypothesis testing and feedback. For instance, when a second language learner gets exposed to the language he is learning, he will maybe notice different endings (Berggreen & Tenfjord, 1999) in verbs, and make a hypothesis about verb inflection (among other things). The learner will acquire some information about the regular form, and maybe generalize that information to verb inflection in general (hypothesis formation) (Berggreen & Tenfjord, 1999). Then he might use the inflection on some verbs (hypothesis testing). If he uses the regular form of inflection on irregular verbs, he might get positive or negative feedback by the listener explicitly telling him it's ungrammatical, by being understood, or by being misunderstood. The learner takes advantages of the feedback he is getting and decides whether or not he should keep the hypothesis or form new hypotheses about verb inflection. By forming and testing hypotheses, and receiving feedback, the learner acquires information about the language without explicitly learning the grammatical rules of the language (Berggreen & Tenfjord, 1999). Cognitivists view negative feedback as necessary for learning. Nativists on the other hand, view negative feedback as irrelevant and not pedagogical. They view positive feedback and the innate universal grammar as the essentials to language learning (Berggreen & Tenfjord, 1999). This can relate to the logical problem of language acquisition: when children learn language, they are exposed to inconsistent and incomplete input, but acquire adequate language nevertheless, even without taking advantage of negative feedback (MacWhinney, 2004). There is, in other words, a gap between the input the children get and the competence they achieve. Nativists believe the explanation of this gap lies in the innate universal grammar (Berggreen & Tenfjord, 1999).

Explicit knowledge is knowledge of the rules one is using, for instance, if you learn the rules of a foreign language while learning the language. Explicit learning happens if you are instructed to look for patterns in a situation where learning is intentional. This result in conscious knowledge (Rebuschat & Williams, 2012). Implicit learning happens without knowledge of rules. It is the ability to acquire unconscious knowledge (Rebuschat & Williams, 2012).

Executive Functions' Role in Language Learning

Research show that as children develop, so does their ability to control their thoughts and actions, that is, their executive functions. Executive function, or cognitive control, is a term for the many cognitive processes that ensures controlled cognitive processing. These functions make us capable of controlling our attention and actions by overruling dominant responses. They are crucial in demanding situations where you need to adapt to a changing environment quickly (Huizinga, Dolan, & van der Molen, 2006).

Studies have shown that executive functions develop over a long time, from early childhood into adolescence. These studies indicate that different executive function tasks reach the level of an adult in different ages ranging from childhood to adolescence. For instance, some studies found that working memory develops gradually from childhood to youth, and inhibition increases through childhood up until approximately 12 years of age (Huizinga et al., 2006). A study on 48,537 online participants show that there is a diverse range in age at when different cognitive abilities peak. For instance, the results showed that you do not reach peak performance in vocabulary until you are about 50 years old, but you peak at cognitive tasks like “Digit span” when you are between 20-25 years old (Hartshorne & Germine, 2015). There might exist similar peaks to language learning. If language learning follows a biological schedule, it makes more sense that there will be biological constraints on language learning as well (Long, 1990). Some researchers on language acquisition believe that after a certain age, our ability to learn language starts decreasing. For instance, it is generally accepted that after a certain age, you cannot train enough to become a professional ballerina, or a sensational piano player. The same principle can be applied to medicine, where we know that some medicine must be given before a certain time for it to have full effect, and, for instance, the prognosis of aphasia differs depending on the age the damage happened (Long, 1990).

To what extent do executive functions explain variance in language learning? Specific language impairment (SLI) has been linked to a disadvantage in executive functions such as working memory and inhibition. As mentioned, in statistical learning tasks, learners are supposed to track patterns of regularities in input like syllables and tones. The learners are typically exposed to a stream of syllables or sounds that is constructed with a set of simple statistical regularities (Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999). Even though there is no instruction, adults, young children and infants are able to discover the regularities connecting the stream elements.

However, research suggest that children with SLI are less able to detect these patterns. Evans, Saffran & Robe-Torres (2009) asked if children with SLI are impaired in their ability to keep track of the sequences of syllables they hear in a stream of speech. To be able to track these sequences, children have to realize what sounds fit together and make words by adjusting native language speech perception, discover the language phonological structure and segment words from fluent speech input (Saffran & Estes, 2006). This is a fundamental implicit learning process to the early stages of word learning. If children with SLI are unable to discover such word boundaries their challenges might be about more than implicit artificial grammar learning (Evans, Saffran, & Robe-Torres, 2009).

In the first experiment the participants were 113 children, of which 35 had SLI and the rest were typically developing. The stimuli were six words with three syllables and the participants needed to use transitional probabilities to establish word boundaries. In the two parts of the second experiment 30 children from the first experiment participated 6 months later. Half the group had SLI. The results from the experiments conducted by Evans et al. (2009) support the hypothesis that typically developing children can use statistical information to notice word boundaries, and it seems to be a domain-general ability being similar across speech and tone conditions. The results for the children with SLI were somewhat unclear, but it seems that these children are not as able to use statistical information to discover word boundaries. One of the most apparent difficulties the children with SLI had, was to discriminate newly learned target words from words sounding very similar that were test words. The researchers argue that might be because children with SLI do not keep in memory a detailed phonological form of the target words. This might indicate that memory capacity may be of importance for these cognitive tasks.

The difference in ability to establish words boundaries by statistical information might be because of differences in executive functions such as working memory and attention capacities. This difference could not be explained by factors like intelligence, age or language (Evans et al., 2009). If intelligence was crucial for language development, we should have been able to measure IQ differences in children learning their first language (Long, 1990). There would be correlation with IQ and the rate and achievement in first language acquisition. Such a difference does not seem to exist. Ludden and Gupta (2000) learned that performance by adults is decreased in statistical learning tasks when memory and attention are reduced. Other studies have also shown that executive functions have an impact on working memory performance in children with SLI, regardless of domain and that tasks requiring a great amount of executive functions are more challenging for children with SLI than typically developing children. These children also showed weaker abilities in inhibition and attention control tasks than their typically developing peers (Marton, 2008).

A study investigating the role of attention in statistical learning had half of their participants listen to a speech stream passively, while the other half were distracted in three different ways in three respective experiments by pushing a button at different cues (Toro, Sinnett, & Soto-Faraco, 2005). They found that the participants passively listening to the stimuli were able to extract the nonsense words of the stimuli. The participant that were distracted, on the other hand, the performance on word segmentation decreased to chance level (Toro et al., 2005).

Brooks (2006) did an experiment with 60 adults to examine if they were able to inflect masculine and feminine Russian case nouns. Every noun was paired with a matching drawing. The participants went through six training sessions and a test phase. The experiment involved 4 blocks, first a listen and repeat task, then a noun comprehension task, a case comprehension task and a case production task. In addition to a language learning task, Brooks (2006) wanted to measure individual differences in the participants. Verbal working memory was tested with the “reading span task”. Phonological memory was tested using a Non-word span task. Nonverbal intelligence was tested using the Cattell culture-fair test of nonverbal intelligence. The results from this experiment showed that there was great variation in the participants ability to learn inflection. They also observed that only learners with sufficient attentional resources could benefit from larger training vocabulary to generalize the rich input (Brooks, Kempe, & Sionov, 2006).

Thus, differences in attention, inhibition and working memory might play a role in statistical learning mechanisms.

However, if statistical learning happens implicitly, we might see that executive functions will not have an effect on the ability to learn language. Reber believed that implicit learning is independent of intelligence, but that explicit learning, on the other hand, is dependent on intelligence Reber (1993). Studies that support this hypothesis are the studies of Gebauer and Mackintosh (2007) and Kaufmann (2010). Gebauer and Mackintosh (2007) found no correlation between tasks of implicit artificial grammar learning, and different measures of intelligence. However, they did find a significant correlation between scores on the learning tasks and the measures of intelligence when the participants were given explicit rules (Gebauer & Mackintosh, 2007). Kaufman et al. (2010) concluded that implicit learning was related to some personality traits, but not to psychometric intelligence or working memory.

Previous Research of Statistical Learning

Jenny Saffran has conducted several studies with focus on the topic of language acquisition in accordance with learning processes, mainly in the field of statistical learning (Pelucchi, Hay, & Saffran, 2009; Saffran, 2002; Saffran et al., 1996; Saffran & Wilson, 2003). Saffran argues that the similarities across languages are not accidental, and that languages are shaped by learning mechanisms in human beings (Saffran, 2003). Aspects of language that are more likely to persist are those aspects which are more learnable. Research show that learners seek regularities (Erickson & Thiessen, 2015). Statistical learning has been assessed in different ways, and there have been studies on infants, children, adolescents as well as adults (Arciuli & Torkildsen, 2012).

Most of the evidence accumulated thus far comes from studies experimenting with artificial language (Pelucchi et al., 2009). Gomez (2002) tested forty-eight adults in one of two artificial languages made up by an aXb, cXd, eXf structure. Structures like these are nonadjacent dependencies. Nonadjacent dependencies happen over several units in language and learners must track discontinuous sequential relationships (Eidsvåg, Austad, Plante, & Asbjørnsen, 2015). An example of a nonadjacent dependency can be found in the sentence "I was sleeping". The grammatical morphemes "was" and "-ing" indicates past tense. Both of these morphemes are necessary to comprise a grammatical sentence. Since the lexical morpheme "sleep" is between the two grammatical morphemes, the relationship between the grammatical morphemes are nonadjacent (Santelmann & Jusczyk, 1998). Research has shown that these dependencies are learned by using statistical learning mechanisms (Sandoval & Gomez, 2013). Some examples of the input Gomez used is "pel wadim rud" and "vot kicey tood". The first and last elements was the nonadjacent dependencies. They were also exposed to different set sizes, namely 2, 6, 12 and 24 (Gomez, 2002). Gómez found that performance increased significantly in the largest set size and that this can suggest that high variability makes learning of nonadjacent dependencies more effective.

Gómez then tested forty-eight 17-19-month-old infants on nonadjacent dependencies (Gomez, 2002). This experiment had the same structure as the one with the adults, but the infants were tested on two nonadjacent dependencies. Participants were exposed to set size 3, 12 and 24. They were tested using the head turn preference procedure. This involves measuring how long the infant's head is turned towards the sound stimuli. First, the infant's attention is turned towards the middle of a small room, using a doll or something similar. Afterwards, one out of two red lights on each side of the room is switched on. When the infant turns to the light, the sound stimuli start playing. The measuring ends when the infant looks away for 2 seconds, or when the entire stimuli is played to the end (Nelson et al., 1995).

Gómez used one of two artificial languages as stimuli in the experiment, and they were exposed to two nonadjacent dependencies in three-element strings. Element one and three was the nonadjacent dependencies. Element two varied depending on the set size. Each participant listened to auditory strings for about 3 minutes. To test whether variability played a role in learning of nonadjacent dependencies, she varied the size of the pool of the middle element in the strings. Each set size consisted of 3, 12 or 24 middle elements. Participants in the low-variable group (set size 3), heard each of the 6 training strings eight times. Participants in the high-variable group (set size 24) heard 48 unique strings. A test trial was presented after the exposure to the strings. The results show that almost all of the infants, except one, exposed to the set size 24 listened significantly longer to untrained strings. These results suggest that they were able to discriminate the two types of stimulus. Infants exposed to set size 3 and 12, did not show any discrimination ability.

The results show that a greater proportion of participants in the set size 24 showed perfect discrimination. These results suggest that learning of nonadjacent dependencies is more likely with higher variability.

Eidsvåg et al. (2015) conducted a study that showed results in accordance with the findings of Gómez (2002). The experiment examined if variability in stimulus input affects learning of gender affixes in a natural language. 40 adults were exposed to Russian words containing masculine and feminine affixes. Half of the group were exposed to 32 unique root words in a high-variability condition. The remaining participants heard 16 unique root words that were repeated twice in a high-repetition condition. The results from this study showed that the participants in the high-variability condition were able to rapidly learn the input, while the high-repetition condition learned after additional input (Eidsvåg et al., 2015). This study therefore also show that rich input variability is an advantage in language learning.

Evidence suggests that learners, including infants, can use statistical properties of linguistic input to discover structure, including sound patterns, words, and the beginnings of grammar (Saffran, 2003). Saffran, Aslin and Newport (1996) investigated whether eight-month-old infants were able to segment words of fluent speech, only based on the statistical relationships between neighbouring speech sounds. In the familiarization-phase, the participants were exposed to a speech stream of four three-syllable nonsense words. The nonsense-words were repeated in random order, so the only cues to the word-boundaries were the transitional probabilities between syllable pairs. Transitional probability is the probability of one event given the occurrence of another event, for instance between syllable sequences. This statistic refers to the chance of which one element follows another (Pelucchi et al., 2009). In English, the probability that “ba” is to be followed by “by”, to make “baby”, is higher than “lo”, to make “balo”, which is a nonsense word. Thus, syllable pairs that had higher transitional probabilities where syllables that constituted words (Saffran et al., 1996). Afterwards the participants went through a test phase, where the non-words from the familiarization phase was presented with new non-words which was not presented in the familiarization phase but consisted of the same syllables in different orders. The results show that the infants showed a significant discrimination between the non-words they had been familiarized with, and the new non-words. They listened longer to the new non-words. These findings suggest that infants can recognise the difference between familiarized words and new words (Saffran et al., 1996).

The vast majority of studies that assess statistical learning use artificial languages (Arciuli & Torkildsen, 2012). One advantage with using artificial languages is that participants have never heard the language before, and one can easier control for pre-linguistic experience. Also, using artificial languages allow the experimenter to generate the input in a way that learning can be ascribed solely to the use of the cues they are exposed to under the experiment, this is known as experimental control (Arciuli & Torkildsen, 2012). The downside with using artificial languages is the lack of ecological validity (Arciuli & Torkildsen, 2012). Natural languages are more complex on many levels; the phonetic level, morphological level as well as the syntactic level, to name a few. On the phonetic level, the acoustic variability for instance, is much more complex (Arciuli & Torkildsen, 2012). Do the findings from studies on artificial languages apply for natural languages? To answer this question, we need more studies on statistical language learning in natural languages. There are, however, some studies that use natural language as stimuli.

Pelucchi, Hay and Saffran studied statistical learning by English 8-month-old infants. They investigated the infants' ability to track transitional probabilities in Italian speech, by doing a similar study to the study by Saffran, Aslin and Newport (1996) mentioned above.

These infants were also tested with the head turn preference. The difference from the study by Saffran, Aslin and Newport (1996) is that Pelucchi et al. (2009) used Italian, a natural language. The participants were familiarized with some Italian sentences, and afterwards tested on familiar words and Italian words they had not heard before. Results suggested that infants could discriminate the items presented during familiarization from novel words, despite the use of foreign language materials, because of the significant preference for familiar words. The results do not only support the assertion that infants are born with statistical learning mechanisms, but also that the mechanisms work on natural languages (Pelucchi et al., 2009).

Natural language has more complex morphology, syntax, phonology, as well as intonation, than artificial languages have been able to recreate. E.g. in natural language, words will vary acoustically depending on which words they precede and follow, because words cannot be produced by speakers without any assimilation (Erickson & Thiessen, 2015). Though artificial language has provided a lot of evidence for statistical learning (Gomez, 2002; Wang & Saffran, 2014), this type of stimuli clearly lack the complexity of a natural language and that problem has been recognized in studies using this type of stimuli (Arciuli & Torkildsen, 2012). In the study by Gomez (2002) we can see the typical artificial language design. There are few words, and the words are repeated very frequently during the familiarization phase. In addition, Pelucchi et al. (2009) explains that the materials often lack any rhythmic pattern and lack variation in pitch, which is something that exist in all natural languages.

The study by Friederici, Mueller, and Oberecker (2011) investigated natural grammar learning in 4-month-old infants. They examined this by testing the participants on nonadjacent dependencies in a novel natural language. The participants were German monolinguals, and the stimulus material was in Italian. The stimulus material consisted of four-word grammatically correct Italian sentences. The sentences contained rule-based nonadjacent dependencies. There were 64 sentences in the familiarization phase, which they listened to four times. The familiarization phase lasted approximately 13 minutes. After the familiarization phase, a test phase followed. In the test phase, the participants were exposed to both correct and incorrect sentences. The results show an indication of discrimination between the incorrect and the correct sentences. These findings suggest that there is evidence for infants learning nonadjacent dependencies in a novel language (Friederici et al., 2011).

Awareness

What is the role of awareness in language acquisition? And is it possible to learn language without awareness? The questions surrounding awareness, and how to measure awareness in language experiments, are widely discussed. There are multiple ways to assess awareness in language experiments. To measure awareness, one can for instance use verbal reports, direct and indirect tests and subjective measures (Hamrick & Rebuschat, 2012). Verbal reports include anything the subjects might have noticed while participating in the experiment, that they can verbalize. If their performance is above chance, despite them being unable to verbalize what they have learned, it can indicate unconscious knowledge. Confidence ratings (CRs) are the most typical subjective measure (Wierzchoń, Asanowicz, Paulewicz, & Cleeremans, 2012). Subjects get asked how confident they are on their responses. If they performed above chance, while they believed to be guessing, it can indicate unconscious knowledge: the subjects' accuracy on the test results are unrelated to their confidence of their answers. A direct test instructs subjects to use their conscious knowledge to perform, while an indirect test do not instruct them to use acquired knowledge (Hamrick &

Rebuschat, 2012). If the latter test indicates a learning effect, it can be an indication of unconscious knowledge.

Statistical learning studies often don't have any measures of awareness as a part of the experiment. The aim of the study conducted by Hamrick and Rebuschat (2012), was to fill in this gap. They wanted to measure whether the subjects acquired conscious or unconscious knowledge in a typically statistical learning experiment, by testing thirty English-speaking participants. They applied subjective measures and verbal reports for measuring awareness. The participants were assigned to one of two groups. Half of the participants were in the group with incidental learning conditions, whereas the other half were in the intentional learning condition-group. The difference between these groups was that the intentional learning-group were told that they were participating in a word-learning experiment. They were explicitly instructed to learn word meanings, and that they soon were going to be tested. The incidental learning-group was not given the same instructions. The instruction the researchers presented to them, was to count animate objects. The stimuli material contained 27 pseudo-words, matched with one or more drawings. In the exposure phase, participants were presented with two pictures, while simultaneously hearing a set of two auditory strings.

In the test phase, four pictures from the exposure phase appeared on the screen at once, while they heard a pseudo-word. They were instructed to choose which picture they thought matched the pseudo-word. Since the study wanted to measure awareness, the subjects were asked how confident they were on their answers. They were also instructed to evaluate if their response was a guess, intuition or memory. At the end of the test phase, participants were instructed to answer a questionnaire concerning if they had learned any of the referents to the pseudo-words, if they had used any strategies, and if so, what kind of strategies.

The results indicate that both groups had a learning effect, but the intentional learning-group had a greater effect. Regarding confidence ratings, participants in the intentional-group were partially aware of learning during the experiment, whereas participants in the incidental-group were not aware.

The role of awareness is widely debated, and which method one is using, can affect the results. Williams (2005) wanted to investigate whether learning without awareness is possible. By using offline measures (retrospective verbal reports), he concluded that learning without awareness is possible (Rebuschat, Hamrick, Riestenberg, Sachs, & Ziegler, 2015). Leow (2000) conducted a study to investigate the effects of awareness in foreign language behaviour. To measure awareness, he used think-aloud protocols (online measures). The results suggest that there is no evidence for learning without awareness. Rather, it supports the claim that awareness is crucial in processing of a second language.

The methodological differences between Williams' study and the study of Leow, is the use of measurements of awareness. While offline measures measure whether exposure results in unconscious knowledge, online measures measure the role of awareness at the time of encoding. These differences result in measures of awareness in different stages. While Williams measured the *product*, Leow measured awareness of the *process* of learning (Rebuschat et al., 2015).

Hama and Leow (2010) combined online and offline measures, and conducted a study where they adapted Williams methodology, and extended it by adding think-aloud protocols. This was done to gather data in both stages: in the process of learning, and the product of learning. The results show no evidence for the participants being aware during exposure of the stimulus, however there was evidence for awareness in the think-aloud protocols in the test phase. The results also suggest that the participants who showed a learning effect, were those who became aware of the hidden regularity.

Rebuschat, Hamrick, Riestenberg, Sachs & Ziegler (2015) also conducted an extension study to the study of Williams (2005). The intention of the study was to compare three measures of awareness: retrospective verbal reports, concurrent verbal reports (think aloud protocols) and subjective measures. This was the first study to compare the three measures of awareness. By triangulating these measures of awareness, they could investigate what participants became aware of, at what time they became aware (if they became aware), and the differences of the awareness measures.

The study consisted of three experimental groups. The participants in the first group thought aloud during training, the second group thought aloud during training and testing, and the third group remained silent. For the think-aloud groups, the participants were recorded while thinking aloud, and the recordings were transcribed later. A fourth control group also participated. In addition, every participant in every group went through a forced choice task, subjective measures (confidence ratings and source attributions), and got interviewed after (retrospective verbal reports).

The subjective measure responses (confidence ratings and source attribution) were collected in the test phase. Regarding the confidence ratings, participants had to rate how confident they were on their answers after each test item, by selecting one of four response options (100% confident, very confident, somewhat confident or not confident at all). Participants were also instructed to evaluate their responses for the source attribution. They were asked to select one of the following options: guess, intuition, memory, or rule knowledge.

The retrospective verbal reports were collected as short interviews after the exposure and test-phase. The researcher asked them about the reasons for their responses during the test-phase. They were also asked if they had made any reference for living-nonliving or similar distinctions, and if so at what point they had made these references. The participants were also asked for their thoughts regarding source attribution. If they had picked the “rule knowledge” response, they were asked to explain why.

Results from the test phase shows that the silent-group performed best. Both the silent group and the think aloud during exposure-group performed above chance, indicating a learning effect. Regarding the confidence ratings and source attributions, all groups scored significantly above chance that they were very confident, or 100% confident, and chose the memory or rule knowledge category. The silent and the think aloud during exposure-group also scored significantly above chance in the intuition category, which indicates some implicit knowledge. The verbal reports indicate that a considerable amount of participants began to actively search for rules during the test phase, and reported awareness. In other words, the results indicate that the participants had acquired both explicit and implicit knowledge (Rebuschat et al., 2015).

By testing and comparing different types of measure for awareness, one can more clearly detect what type of awareness each method is measuring, and that results of awareness can differ depending on method. These differences in methods should be taken account for when studying awareness. By using more than one type of measure of awareness, one can collect more detailed information. In the present study, we apply a self-report questionnaire which contains both retrospective verbal reports, and confidence ratings.

Tone in Statistical Learning

Tone play a major part in many of the world languages, and it’s therefore interesting to study how linguistic tones plays a part in statistical learning. There have been few studies on statistical learning with linguistic tone as the focus thus far, which is part of the reason it is such an interesting topic to investigate further. Saffran, Johnson, Aslin, & Newport (1999) did an experiment using non-linguistic tone. They were curious whether the same statistical learning ability, as with spoken input, could be applied to musical tones. They created a sound stream identical in its statistical properties to the syllable stream employed by Saffran, Newport, & Aslin (1996). Each of the syllables was substituted with a distinct tone. For example, “bu” became the musical tone D. There were six trisyllabic artificial words that were translated into a sequence of three notes following each other. Three experiments were executed with two groups of adults and one group of infants (Saffran et al., 1999).

The results of the adult subjects in the first two experiments performed equivalent to the subjects from the previous study, they managed to segment the tone stream (Saffran et al., 1999). The same results were also found with eight-month-old participants, whose results on the tone segmentation task paralleled the results from the previous speech segmentation task by Saffran, Newport, & Aslin (1996). According to Saffran et al. (1999), these findings suggest that linguistic stimuli are not privileged to statistical learning processes. The same ability to detect patterns can be applied to tones. However, the significance of tone sequence segmentation is less evident than word segmentation (Saffran et al., 1999).

Most of the studies mentioned are either conducted with artificial language, or Italian/Russian as the natural language. Most of the participants in the studies (both on artificial language and natural language) were either English, American or German. English and German are both west Germanic languages (Konig & Van der Auwera, 2013). Since English, German as well as Italian are in the same language family, The Indo-European language family, we need more studies from other language families to increase the validity. One reason for this, is that we want to know if statistical learning applies to all types of languages. Germanic languages are only a small part of the world's languages, and therefore one cannot generalize these findings to all languages, without investigating additional and unrelated languages.

German is a synthetic language (Kaminska, 2007), which means that a word can consist of many morphemes, for instance by inflecting a word. Opposite we have analytic languages, which are languages where words consist of few morphemes. Some refer to English as an analytic language, but English do contain synthetic elements (Rissanen, Ihalainen, Nevalainen, & Taavitsainen, 1992). Synthetic and analytic languages ought to be understood as a continuum with them being the pole on each opposite end. Languages that are completely analytic with no inflections and few morphemes per word, are referred to as isolating languages. Mandarin Chinese is an example of an isolating language (Li & Thompson, 1989).

Since Mandarin morphology structure is very different from English and German, according to the synthetic-analytic continuum, it is of interest to study statistical learning in languages like Mandarin, or test participants of Chinese origins to see if the results differ from participants with a Germanic language.

Mandarin is a tonal language. Whilst the morphology of Mandarin is simple, compared to English, the tonal system is making the phonology much more complex.

One study which focused on linguistic tone is the study of Wang and Saffran (2014). They studied statistical learning of a tonal language and used an artificial language to test the influence of bilingualism and pre-linguistic experience. They performed three experiments, with participants of different origins. In the first experiment, the participants were English monolingual adults. In the familiarization phase, they were exposed to an artificial tonal language, with words consisting of three syllables. Each syllable had a tone. Wang & Saffran (2014) also used two conditions (A and B) to control for arbitrary listening preferences during testing. The speech stream consisted of three trisyllabic tonal words. For condition A, one word would be "tadugu", while for condition B the word would be "guduta". In the test phase afterwards, they were exposed to both words from the familiarization phase and new words. The participants did not discriminate words from new words better than chance (Wang & Saffran, 2014).

The second experiment, the stimulus material and the procedure were the same. The participants however, were Mandarin monolinguals and Mandarin-English bilinguals. They compared these groups to see whether the results differ. Mandarin-speakers are familiar with the concept of tones, due to the use of tone in Mandarin. The results show that the Mandarin monolinguals performed better than chance. The Mandarin-English bilinguals outperformed the monolinguals.

In the third and final experiment, they wanted to test if bilingual participants performed better in this test, only due to them being bilingual. Therefore, they tested non-tonal bilingual participants, with the same material and procedure. The results show that the non-tonal bilinguals performed significantly better than chance. Only the Mandarin-English bilinguals performed better than this group. Results from the last group, the non-tonal bilinguals suggested that bilingualism facilitate learning in the task they were given.

As discussed in the paper, one reason why the Mandarin-monolinguals might have found the test hard, is that Mandarin is a disyllabic language, and the artificial language they were exposed to was three-syllabic. English words can consist of one to five syllables, and English monolinguals are therefore used to that word can consist of more than two syllables (Wang & Saffran, 2014). This should be taken account for in future studies.

To our knowledge, there are not many studies on statistical learning in relation to linguistic tones. Wang and Saffran's (2014) used an artificial language as stimulus material. It would be interesting to investigate if findings like those in the studies mentioned above, applies to natural tonal languages. By conducting studies on natural languages the ecological validity will also increase. It would be interesting to investigate whether similar results will come from an experiment with a tonal language as the stimulus. The results of this study and the fact that statistical learning works on tone, visual and tactile modalities (Frost, Armstrong, Siegelman, & Christiansen, 2015), might suggest that the same statistical learning principles may be applied to the suprasegmental level that is linguistic tone.

Typological Description of Thai

As mentioned, Mandarin is an isolating language with simple morphology, but a complex tone system. We find similar language characteristics in other languages as well. Thai is a language spoken in Thailand. As Mandarin, Thai is an isolating language (Nakatani & Minegishi, 2011). This means that in Thai they use verbs, adverbs or nouns where other languages might use a tense marker. The words mostly consist of few syllables, and linguistic tone play a big role in languages like Thai (Nakatani & Minegishi, 2011) (and Mandarin, mentioned in previous studies). In these perspectives Thai is different from the artificial languages and the European languages that have been used as test materials in many other studies (Gomez, 2002; Pelucchi et al., 2009; Saffran et al., 1996). The language has a simple morphology, but like Mandarin, a complex tone system. By investigating statistical learning in a language like Thai, we can increase our knowledge on how statistical learning operates in tonal languages. Standard Thai has a complex tone system, with five different lexical tones. Linguists have described these tones as mid, low, high, falling and rising (Morén & Zsiga, 2006). In contrast to Norwegian, which has a simple tone system, with only a two-way basic contrast (Vanvik, 1973), Thai monosyllabic words are tone bearing units. In the study by Wang and Saffran (2014), an artificial tonal language was used. As mentioned, they tested if Mandarin-speaking participants had an advantage due to Mandarin being a tonal language. The words in the artificial language consisted of three syllables, with each syllable having its own tone. Words like this don’t usually exist in isolating languages, due to the low morpheme per word ratio.

Wayland & Guion (2004) examined language learners’ ability to separate Thai tones. They wanted to compare and contrast the ability to discriminate a tone contrast in standard Thai. The mid and low tones were compared. They used a group of native English listeners and a group of native Mandarin and Taiwanese listeners as participants. The study aimed to observe the difference in the ability to perceive tones among these listeners before and after auditory training. The results of this study suggest that the Chinese listeners were superior to the English listeners at discriminating phonetic variations of the low and mid tone in Thai.

They also argue that the Chinese listeners are better at tracking voice movement, direction, or change at word level, because of their pre-linguistic knowledge, as Chinese also is a tonal language (Wayland & Guion, 2004). The results in this experiment may suggest that language experience may be a factor in language learning.

By testing participants in a natural language like Thai, we can increase the ecological validity, by using syllables and words that actually exist in isolating languages.

Intention and Research Questions

The intention of this study is to supplement previous research on statistical learning. The aim is to get more knowledge about how statistical learning mechanisms operates in a language that differ from most languages used in previous experiments. There are not many experiments done on tonal languages. A study on a natural tonal language will increase our knowledge on statistical language learning, considering that many of the world's languages are tonal languages. Tone is a suprasegmental phoneme, which means it is a phoneme that may extend over series of segmental phonemes (such as consonants or vowels). Thus, it is of interest to investigate how statistical learning applies at the suprasegmental level.

In our study, we want to test whether Norwegian adults can track phonological tone patterns in Thai. As mentioned previously, Norwegian has a simple tone system. The participants are therefore familiar with the concept of linguistic tone, and that words that consist of two or more syllables can differ in meaning only by changing tone. However, they are not familiar with a complex tone system, and that monosyllabic words can contrast only by tone.

Testing with Thai as stimuli is interesting because of the difference in morphology, syntax and especially tone, compared to the type of European stimuli many other studies have used. Evidence indicates that linguistic experience as well as bilingualism gives the learner an advantage in a language learning situation (Wang & Saffran, 2014; Wayland & Guion, 2004). As mentioned above, Norwegian and Thai both have linguistic tone, and this makes it interesting to see whether Norwegians will be successful in learning the tone patterns in Thai.

As mentioned, previous experiments on the syntactic level show that higher variability leads to quicker learning. We want to test if the same pattern is true for the tonal level. Our hypothesis is that higher variability leads to quicker learning in a natural tonal language.

We also want to investigate whether the same mechanisms that apply to the syllable level, apply to the tonal level as well. Additionally, we wish to detect potential differences in the levels. Are there any differences between learning of syllables and learning of tones? Does learning happen faster in one of the levels?

We also want to investigate whether executive functions have an impact on the ability to implicitly acquire language. In the present study we examined some cognitive abilities. We tested the participants' executive functions by using four recognized tests that challenges the participants' attention, inhibition, verbal fluency and working memory.

In addition, we will also investigate whether learning happens with awareness. Are the participants aware of the rules, or does learning happen unconsciously? To gather data of participants' awareness, a questionnaire will be administered.

Method

Research Design

This study is a quantitative study and will have an experimental research design for collecting and analysing data. Since we wanted to investigate statistical learning in a natural tonal language, and in addition investigate whether or not awareness occurred, and if executive functions are related to statistical learning, we chose an experimental design. The reason for this is that we wanted to measure if learning happened, and to what extent awareness and executive functions plays a role. Since there is limited research on statistical learning of linguistic tone, we wanted to supply previous research with an experiment. We tested 40 participants, and to recruit them we used consecutive sampling. That is, every person who signed up for the experiment and fulfilled the inclusion criteria was accepted until we reached 40 participants. The study has a consecutive sampling, due to time and cost considerations.

We also have a limited access to potential participants, due to our exclusion- and inclusion criteria. To ensure a balanced distribution of participants in each group, we applied a permuted block randomization (Polit & Beck, 2017). The participants were therefore assigned to different treatments in two groups: high variability group vs. low variability group. To get as similar groups as possible in the high variability group and the low variability group, the males and the females were split in half. Thus, 10 males and 10 females constituted the respective groups. The respective groups were then again split in half after the experiment. During the cognitive test dichotic listening, half of the participants were asked to focus on their left ear first, then the right. The other group were asked to do the opposite, that is, first focus on the right ear, and lastly their left ear. This was done to make the groups as similar as possible, and control for other factors (sex, listening preferences). Each participant got a number, which make them anonymous, and makes it possible to connect the different parts of the examination.

We will analyse the participants' scores in the language learning experiment and analyse the potential correlation between the experiment and the experience of awareness and executive functions.

Participants

Participants included in this study are 40 Norwegian students. 20 males and 20 females, with an age range from 19 to 35 (mean age: 24). Our inclusion criteria for age was 18-35, as changes in cognition are age-related. It has been found that between the ages 20-30, we experience cognitive decline (Salthouse, 2009). It is shown that cognitive abilities have a wide range of ages for peak performance (Hartshorne & Germine, 2015). Thus, we want as homogeneous groups as possible. To make it easier to compare the participants' results, and control for factors of different mother tongues, each participant must have Norwegian as their native language. The majority of Norwegian students are Norwegian-English bilinguals, as they learn English in primary-, secondary- and high school. To Norwegians, Swedish and Danish are mutually understandable languages. The participant must not have any self-reported learning difficulties that they know of (for instance language impairment, dyslexia). They must have normal hearing, and don't suffer from hearing loss on one or both ears. The participants must not have any previous experience with Thai, as this can affect the results. We want to measure whether they are able to track certain patterns in Thai, hence they must not have any experience with patterns they are going to be exposed for.

The participation is voluntary, and the participants have the right to withdraw from the study at any time, without risking sanctions. Informed consent was obtained.

Two participants were excluded due to an error in the setup in the low variability stimuli. Only half of the stimulus material was presented to the participants. This error was discovered after the two first participants had already been tested. Thus, we had to recruit two additional participants.

Procedure

To prepare for the experiment, we tested the procedure on each other, as well as on two volunteers in a pilot study. The pilot study resulted in feedback from the volunteers, who thought the stimulus material was presented too slowly. We took this feedback into consideration and shortened the time in between the strings to 0,5 seconds.

To prevent an expectation bias, we only recruited people with no previous knowledge of the experiment. Recruitment was done by hanging up posters in different areas/campuses of University of Bergen. Posters were also published on social media.

The participants were examined one by one. As soon as they were accompanied to the experiment room, they read and signed the consent form in a laboratory room. Since all our participants are adults, they are mentally able to make their own decisions. Thus, they were able to give informed consent before the experiment began. After they had confirmed to voluntary participation, they were escorted in to the soundproof room within the same lab, where the language learning experiment was going to take place. To prevent possible experience of discomfort, one of the experimenter always went in the soundproof room with them. E-prime 2.0 Proffesional (Schneider, Eschman, & Zuccolotto, 2002) was used to present the stimuli and to collect data. The language learning experiment lasted approximately 30 minutes. After the 30 minutes, they went back to the lab room (out of the soundproof room), where one of the other researchers handed them a questionnaire. After the questionnaire was answered, the participants went through four cognitive tests, which lasted about twenty minutes. As the whole examination lasted approximately an hour, the participants got a monetary compensation of 150 NOK.

The results were treated at group level. Results were based on how the participants scored as a high variability group versus a low variability group. Thus, individual scores will not be published or revealed. Since the participants are anonymized, their results from the experiment, questionnaire and cognitive tests were stored separately from their signature of consent and signature of receiving money.

Stimulus Material for the Language Learning Experiment

In the familiarization phase, the participants listened to a set of auditory strings, presented by two speakers; a man and a woman with Thai as their native language. By using both a man and a woman's voice to read the stimulus, the ecological validity increased. The strings were constructed as adjacent dependencies. These dependencies were constructed for this study, and do not occur as a grammatical rule in the Thai language. However, the strings contained grammatically correct Thai sentences. The stimulus material in the familiarization phase in the high variability group consisted of 64 unique strings. The full stimulus material for the familiarization phases is available in the Appendix. Each sentence was semantically and lexically different. These 64 strings were divided into four semantic subgroups, with 16 strings in each subgroup. Every subgroup had a "target", consisting of a three-worded sentence with a rising-falling-high tone pattern (R-F-H) as illustrated in Table 1. In addition to the target, every sentence had a context clause or additional words with grammatical meaning. To increase the variability, the target appeared before, in the middle and after these context phrases.

Thus, in subgroup one we had a target sentence (lǎan mây lóm) with 16 different context phrases. Together they made 16 unique grammatically correct sentences. In five of the sentences, the target appeared before the context phrase. The target appeared in the middle of the context phrase in five sentences. In six of the sentences, the target appeared after the context. Subgroup two, three and four had the same structure, but each with its own target and context phrases. The only thing these 64 strings had in common was the rising-falling-high tone pattern (R-F-H) in the target.

The stimulus material in the familiarization phase for the low variability group consisted of half of the 64 unique strings. Strings from all four semantic subgroups were used, but the subgroups were halved. Instead of having 16 strings from each subgroup, each subgroup in the low variability group consisted of eight strings. Thus, the stimulus material in the low variability group consisted of 32 unique strings. To make sure all the participants from each group was presented for the same amount of strings, the high variability group was presented for the 64 strings two times, whereas the low variability group was presented for the 32 strings four times. Thus, every participant listened to 128 strings in each familiarization phase. With three familiarization phases, the participants listened to 384 familiarization strings in total.

In the test phases, a third native Thai speaker presented the test strings. These strings were divided into four categories. Category one consisted of four strings, namely the four unique target strings, with the rising-falling-high tone pattern, from the familiarization phase, (see Table 2.1).

Category two consisted of four three-word strings, which had the same tonal structure as the target strings, but with different syllables. In other words, the syllable structure (words) in the strings were different from category one, but the number of syllables, as well as the tone pattern (R-F-H), were the same (see Table 2.2).

Category three also consisted of four three-word strings, but unlike category two, the strings had the same syllable structure as category one. Here, it was the tone pattern that was different. Every syllable in this category was the same as the target, but the tone was different (which changes the meaning of the words.) Every string had its own tonal structure (M-L-F, H-R-F, L-H-L, F-L-L) (see Table 2.3).

Category four consisted of four three-word strings. Both the syllable structure and the tonal pattern in these strings were different from category one/the targets in the familiarization phase (see Table 2.4).

Thus, the test items consisted of sixteen different test strings: four strings which was the same as the target in the familiarization phase, four strings with the same tonal structure as the target, but with different syllable structure, four strings with the same syllable structure as the target, but with different tonal structure, and four strings with both different syllable and tonal structure. With four test phases, the participants listened to 64 test strings in total. In total, the participants listened to 448 test- and familiarization strings.

Tests and Measurements

The examination took place in a quiet room at University of Bergen. The experiment is divided into three parts. Part A was the language learning experiment. Part B consisted of a self-report questionnaire, related to part A. In part C, the participants went through cognitive tests.

Part A. In the first part, the participants were seated in a soundproof room with headset, to prevent disturbances. The language learning experiment (part A) was divided into a pre-test phase, three test phases and three familiarization phases. Each familiarization phase lasted approximately eight minutes. When the experiment started, a message on the computer screen appeared. The message informed them that they would hear different sentences, and they were instructed to judge which of the sentences were grammatically correct. The participants were presented for the sixteen different test items. Every item was grammatically correct Thai sentences, but “grammatically correct” in our experiment, means strings which have the R-F-H tone pattern.

A pre-familiarization test was applied as a control for whether the results from the test phases that follow, significantly differ from chance. In this pre-familiarization test, we expected theirs answers not to be better than chance.

After the pre-familiarization test phase, the first familiarization phase started. The high variability group heard the 64 unique strings two times. The low variability group was presented for half of the strings (32 unique strings) four times. Both groups were presented to the same amount of strings, but the lower variability group was presented to fewer unique strings. Thus, all the participants were presented for 128 strings in random order. By dividing participants in two different variability groups, we can test whether variability is a factor for learning.

After the first familiarization phase, a test phase followed. They were once again presented for the 16 different test items, in random order. The task was to judge which of the strings were grammatically correct, and which strings were not grammatically correct. Right after they were presented for each string from the test phase, a smiley face and a frowny face appeared on the screen. If they thought the sentence was grammatically correct, they were instructed to click on the smiley face. If they thought the sentence was grammatically incorrect, they were instructed to click on the frowny face.

After the test phase, they once again went through the familiarization phase and test phase two more times. The experiment (three familiarization phases and four test phases) lasted approximately 30 minutes.

Part B. When the language learning experiment (part A) ended, they were accompanied into another room, to answer a self-report questionnaire (part B). The questionnaire contained questions about the participants' language background. The full self-report questionnaire is available in the Attachments. We wanted to know what dialect they have, and if they spoke other languages. As mentioned, Norwegian is a language with lexical tones, which means that words can contrast only by changing the tone. However, not all Norwegian dialects have this feature. Some dialects do not use tone as a phoneme. Since tone is essential to this study, it is important to know whether participants have this feature in their dialect, as it can affect the results. It is also important to how many languages they speak, and how well they speak the languages, as this can affect their metalinguistic awareness. If they speak a language that is typologically similar to Thai, it could also affect the results. An inclusion criterion for participating was that they could not have any previous experience with Thai. To ensure this, they were asked to rate their knowledge about Thai.

The questionnaire's main intention was to investigate participants' own awareness. By using offline measures, we gathered awareness-data of the product of learning (what has been learned, opposed to the process of learning). The offline measures we used, was retrospective verbal reports and subjective measures. The retrospective verbal reports were collected through questions in the self-report questionnaire. Some researchers classify self-report questionnaires and group discussions as verbal report (Kormos, 1998). Thus, we classify our questionnaire as a retrospective verbal report. They were asked which test phase they thought they managed best and why, to what extent they thought their answers were arbitrary, to what extent they were looking for certain patterns, and if they found any patterns. If they had found certain patterns, they were asked to describe the pattern they had found. They were then asked to rate how correct they thought their pattern was (not at all, a little, to a large extent or to a very large extent).

Regarding the subjective measures, confidence ratings was collected in the last part of the questionnaire. In the last questions, participants listened to four audio clips from the familiarization phase. Two of the clips were sentences from the RFH-core: [lǎan mây lóm] and [sʉ̌a lɔ̂ɔ mót]. Each participant has listened to each of these sentences 100 times; 32 times with context sentences in each familiarization phase, and once (without context) in each test phase. With three familiarization phases and four test phases in total, they have listened to each sentence 100 times. The other two clips contained parts of two context sentences: [khâw pay nay pàa] and [mây dâay rɔ̀k]. Each participant in the high variability group heard each of these two sentences 6 times during the experiment. The low variability group heard each of these sentences 12 times during the experiment. These parts of the context sentences do not have the same syllable nor tonal structure as the RFH-core. Thus, we wanted to detect if there were any difference in the answers of the sentences with the RFH-core, and the sentences that was a part of the context sentences. After each clip was played, participants had to determine whether it was one of the sentences they had listened to in the experiment. They also had to rate how confident they were on their answer (a scale from 0 to 100). It took approximately five minutes to answer the questionnaire.

The measures of awareness were organized in a way that we could detect different levels of awareness. The first questions surrounding awareness, reveal whether the participants *noticed* any patterns, and to what extent their answers were arbitrary. The second level was about if they could *recall* any patterns and describe them. The third level was the lowest level of awareness: if they could *recognise* a pattern in the four audio clips (get the answers right), without describing it.

By using different types of measurement, we gather knowledge about several levels of awareness. Thus, it provides us with more insight in our participants' awareness. To investigate this even further, one could apply both offline and online measures. Most of our questionnaire gathers information about the participants' awareness as a product of learning. However, the last part of the questionnaire, participants had to rate how confident they were on their answers right after they listened to each audio clip. This kind of measure can be compared to online measures (measures participants awareness during the experiment), but as a less extensive variant. By supplying with online measures during the experiment, one can gather data about the awareness of the *process* of learning.

Retrospective verbal reports/self-report questionnaires have limitations. Since these measures solely rely on participants' verbalization, they can fail to report conscious knowledge, even if they were partially aware. Some participants may not realize that their thoughts are relevant, and thus lack confidence on reporting it. Some participants did not answer the questions surrounding which pattern they had detected. If they have an option of not responding, awareness does not get detected, even if they were aware. We do not know if they were aware or not, by them not responding. It can indicate unconscious knowledge, but it can also indicate that the participants lack confidence in responding. Lack of answer does not always correspond to a lack of awareness. Since these offline measures gathers data of awareness *after* the experiment, the participants' awareness may have decayed in memory by the time they answer the questionnaire. Thus, these measures of awareness may not be sensitive enough to gather all relevant data (Rebuschat et al., 2015).

Part C. When the questionnaire was answered, part C started. In part C, the participants went through four cognitive tests.

Handedness. Before testing, the participants' hand preference was mapped. This was meant to supplement the cognitive tests, especially dichotic listening. If the participant is right handed, he is more likely to have left hemisphere language dominance and therefore a right ear advantage (REA) (Van der Haegen, Westerhausen, Hugdahl, & Brysbaert, 2013). Thus, a test of this kind can provide signs concerning cerebral representation of speech (Rasmussen & Milner, 1977). The participants were asked to answer 15 questions about hand preference by circling the letters H (right), B (both) or V (left) for each of the 15 tasks the questionnaire asked about. An example of the questions asked is "with which hand do you draw?". The hand preference questionnaire was a Norwegian translation of a questionnaire developed by Raczkowski, Kalat, and Nebes (1974).

Dichotic listening. Dichotic listening is often used to assess language lateralisation and auditory attention (Asbjørnsen & Helland, 2006). Kimura (1961) linked right ear advantage (REA) to brain laterality. REA refers to where the language faculties are located in the brain, which is usually in the left hemisphere in right handed individuals (Kimura, 1961). This technique is used to study a range of cognitive processes related to brain laterality, hemispheric asymmetry, attention, learning and memory (Hugdahl, 1995). However, in the current study dichotic listening is used as a measure of attention.

The dichotic stimuli consisted of paired presentations of sounds containing one of six stop-consonants /b, p, d, t, k, g/ and the vowel /a/ (Asbjørnsen & Hugdahl, 1995; Hugdahl, 2004). The six stop consonants, balanced on voiced and unvoiced, combined with the vowel /a/ give a robust REA. Syllable pairs of the sort /ba/ and /ga/ was presented to the participant simultaneously to the right and left ear. In sum there were 36 pairs. The six homonymic pairs function only as control and is not a part of the final score. This was combined with attentional instructions. Attentional instructions to either non-forced or attention forced; to the right ear input or the left ear input. The change in performance during the attention tasks yield a measure of executive functions (Asbjørnsen & Hugdahl, 1995).

The participants were seated at a desk with headphones on and a discman was used to administer the test. The experimenter follows the participants' response on the scoring sheets. The test is divided in three rounds, were the sounds are randomized differently (Asbjørnsen & Hugdahl, 1995).

In the first round, the instructions were to repeat the sound the participant heard most distinctively. This was the non-forced round. The test leader marked the answer on a scoring sheet. In the forced left round, the attentional instructions were to repeat what they heard on their left ear, thus isolating one specific sound. In the forced right round, the attentional instruction were to say the sound they heard in their right ear out loud.

The order in which they were given the forced left or the forced right round after the non-forced round was randomized so that half of the high variable group and half of the low variability were tested first with non-forced, then with forced left and lastly with forced right. The other half of both groups had attentional instructions in the order non-forced, then forced right and lastly forced left.

Dichotic listening is challenging not only because you are presented with different sounds in each ear, but you must be able to repeat what you heard out loud. In the second round no matter which ear is forced, you must be able to isolate attention to the sound you hear in one ear over the input you receive in the other ear. In the last round you must do a sudden switch in attention from the ear you were focusing on, to the sound you hear on the other ear.

Stroop colour-word test. The Stroop colour-word test (Stroop, 1935) requires the participant to be able to inhibit linguistic input and suppress habitual responses (Spreen & Strauss, 1991). In the statistical learning task, if the participants are able to inhibit the patterns they know from their language and be open to recognizing new patterns, a correlation in scores might occur. The Stroop colour-word test is a test that measures response inhibition (Golden & Freshwater, 1978). The Stroop-test used in this experiment is inspired by Lund-Johansen, Hugdahl, and Wester (1996).

The test consisted of three pages. On the first page, the words are printed in regular, black ink. The words "RED", "BLACK", "BLUE", "GREEN", "YELLOW" and "WHITE" (in Norwegian) are written first in one separate line as a trial. Then the words are randomly repeated in a section below, functioning as the test. The instructions given is to first read the one line out loud, and when that is read correctly, the participant is told to the same thing through the entire section, as fast as possible. A phone timer was used to measure time spent reading. On the second page, there are coloured circles. The colours are the same as those that were written on the first page. The participant is told to say the colours out loud on the first line, and afterwards he is timed while saying the colours in the section out loud. On the last page the same words as on the first page are printed, but in colours. However, the words that are written and the colours they are printed in do not necessarily match. The word “RED” can for instance be printed in yellow ink. Instructions to this page are to say the colour of the word, and not to read the written word, thus ignoring the verbal content. After reading the first line as a test, and getting it right, the participant reads the rest of the section while being timed. The scoring sheets contain the same words and colours as the test pages, allowing the test leader to keep track of errors made by the participants. Each of the three test pages gives a time score and a number of errors.

Digit span. We tested the participants with the tasks digit span forwards and digit span backwards from the Wechsler Adult Intelligence Scale (WAIS) (Reitan & Wolfson, 1990). By testing the participants with digit span we will be able to investigate their working memory and attention capacities. If there are deviations on the test scores that correlate with the acceptance rates from the language learning experiment, it might be that these abilities are connected. The aim of the digit span tasks was to see how many numbers forwards and backwards the participants were able to remember.

The first part of the task was to repeat numbers in the same order they were presented. The test leader is supposed to start with a span of three numbers read with approximately one second intervals. When the numbers are presented, the participant repeats the numbers he remembers. The amount of numbers per span increases with one, until the maximum of 9 numbers, for each sequence the participants successfully repeats. If the participant does not remember the digits, or answers wrong, he will get a second attempt with another digit span of the same length. If both attempts fail, part two of the test begins.

The second part of the test is similar to the first part. The only difference is that the participant is supposed to repeat the digit spans backwards. This means that if the test leader reads 7-2, the correct response would be 2-7. The digit span starts with two numbers and increases for each one correctly repeated and the test is finished at eight numbers or when the participant fails both attempts. This might be more challenging: while digit span forwards measure general attention, backwards digit span is a task of working memory (Conklin, Curtis, Katsanis, & Iacono, 2000).

FAS-test. Initial letter fluency tests are often used to test executive functions. The participant is supposed to produce words spontaneously under restricted search conditions (Spreen & Strauss, 1991). This test requires the participant to generate words from initial letters. The challenge of this task is being able to pay attention, initiate search and subsequently be able to retrieve data from the lexicon, and to say the words out loud (Hurks et al., 2006). This test measures verbal fluency. We are curious whether this will impact the participants acceptance rates in the language learning experiment.

The instructions were similar to what is described in (Spreen & Strauss, 1991), however 30 seconds, instead of 60 were used. The participants were asked to find as many words as possible starting with the letters F, A and S in three respective rounds of 30 seconds. The participants were told that they were going to be presented with a letter, and that they would have 30 seconds to produce as many words as possible starting with said letter. They were also told to avoid names, so the answers were constrained to words. They were given each letter just before the timer started. All the words were written down by the test leader, so that eventual repeats were accounted for.

Data Analysis

To analyse the data collected from this experiment we used the computer program IBM SPSS statistics (SPSS IBM, 2011). To see if the results indicate a learning effect, we conducted an analysis of variance (mixed ANOVA). Repeated-measures ANOVA for mixed designs was conducted, because we collected data from two groups, measured at multiple points. We also had both a within-subject and a between subject factor. With a mixed ANOVA we would also analyse potential interaction effects (Polit & Beck, 2017).

Learning was defined as an increase in acceptance rate for the target items. Combined with reduction in acceptance rate of the so called ungrammatical items, consisting of the None-category. In addition, acceptance rate of the two categories of T and S will indicate a preference for either tonal or syllable information. Thus, we analysed the test results with two mixed ANOVAs, using the experiment groups (high vs. low variability) as the between factor. And the acceptance rate for the categories as dependent measures

The first ANOVA was conducted with the two experimental groups (high vs. low variability), and the cycles (1-4) as the independent variables. The dependent variables were the two categories (Tone and syllable (T&S), and None (N)). Since the experiment is conducted comparing two groups, the high/low variability is a between-subject variable.

The second ANOVA had the Tone (T) and the Syllable (S) category as dependent variables, to analyse whether there as a preference for either syllable or tonal cues.

To analyse the self-report questionnaire, a correlation analysis (Pearson's *r*) of the measures of awareness and the acceptance rates from the language learning experiment was conducted. In addition, a mixed ANOVA was conducted to analyse if there were any significant findings regarding which pattern the participants reported, and acceptance rates in the language learning experiment. To analyse this further, we conducted a correlation analysis, and a post hoc with an independent t-test. The t-test tests differences in group means (Polit & Beck, 2017). We conducted an independent t-test as a post hoc analysis, to uncover whether there was a significant difference between the selected means in the two groups (high/low variability).

To investigate if there were significant correlations between the scores of the cognitive tests and the acceptance rates of the language learning experiment, we executed a correlation analysis in SPSS. Correlation is used to examine relationships between variables (Polit & Beck, 2017). We wanted to measure if the variation in one variable was related to another variable, namely the scores of the cognitive tests and the acceptance rates in the language learning experiment. Since the acceptance rates collected in language learning experiment are scale variables, we did a Pearson's *r* correlation analysis.

Outcomes of the Present Study

The results of the ANOVA analysis with the T&S and N category revealed that there was a main effect for cycles, which indicated that the participants had higher acceptance rates in different cycles. In addition, the interaction between the cycle by category was significant. The effect is due to no change in the None category, but an increase in the T&S category. There were also significant findings, a main effect, for categories, which means that the participants had significantly higher acceptance rate for the T&S category than the N category. We found no main effect for groups (high variability and low variability). Thus, we cannot conclude whether variability is a factor for learning in our study. This is illustrated in Figure 1.

The ANOVA analysis with the T and S category, revealed that there were significant findings regarding the categories. Thus, participants had significantly higher acceptance rate for tonal cues than for syllable cues. There was no main effect for cycles or groups. This is illustrated in Figure 2.

The correlation analysis of the self-report questionnaire and the last cycle in the language learning experiment revealed no significant results. The results of the ANOVA of the self-report questionnaire revealed no significant results. There were significant findings in the correlation analysis regarding reported pattern and acceptance rates of the language learning experiment. The independent t-tests revealed that the participants who answered that they had detected a cadence pattern, had significantly higher acceptance rate regarding tone in cycles four, and significantly low acceptance rate regarding syllable in cycles two.

The results from the correlations analysis of the cognitive tests and the language learning experiment showed no systematic patterns of correlations between any of the tests that assumingly should measure different aspects of cognitive control. Except for a spurious correlation of r=.35 between the right ear score during dichotic listening, Forced Left-condition and the acceptance rate of NONE of the third cycle. And also, errors on the Stroop colour-word test and the T&S category during the third cycle r=.31.

Results that are not discussed in the article. We conducted a mixed ANOVA to analyse the response time for the categories (N, S, T, and T&S). We found the following main effect for cycles: $F(3, 114) = 34.058$ $\eta_p^2 = 0.473$ $p<0.05$. A post hoc test revealed that all four categories has significantly faster response time in the last cycle compared to the first cycle (paired sample t-test $p<0.05$). The participants used longest time in the first cycle. One reason for this, might be due to the first cycle being a pre-familiarization test, and the participants may not feel confidence in what they are doing. They might feel more confident in the following test phases, which came after the familiarization phases.

Validity

High external validity implies that the results from a study can be generalized to a greater population. The external validity is weakened due to the homogenic participant group. The group homogeneity is caused by the study's inclusion- and exclusion criteria. Most of the participants are studying higher education. This is not representative for the Norwegian population, as only 32,9% of Norwegian citizens had finished higher education by 2016 (Statistisk sentralbyrå, 2017). Some participants also have experience with language courses during their studies, as people studying language often are more interested in participating in language learning experiments. People with experience with language courses may be more observant towards language patterns. Every participant has Norwegian as their native language. Thus, we must be careful to generalize the findings to people with other native languages. The study only investigates people in a particular age range. Thus, we cannot for instance generalize the results to people older than the age range. The results of this study will, in other words, only give us information about our participants. The results cannot be generalized to Norway's population or other people in general. Nor can we test Norwegian monolingual people, as almost every Norwegian citizen in this age group are bilingual.

However, the homogenic participant group helps to control external factors which potentially can influence the results. These are factors like the influence of native language, hearing disorders, language disorders, and age. The external validity is strengthened by having the same amount of men and women as participants. Thus, the results can be generalized across sex. This dividing reflects the sex representation in the real world.

If a study has high internal validity, it implies that the researchers have control over possible biases. The internal validity is strengthened by making different people read the stimulus material and the testing. This is to limit potential biases that are related to the voice or pitch of the voice belonging to the people reading the auditory strings. If only one person is reading the strings, an insignificant characteristic of his/her voice may be prominent. By making two people read in the familiarization phase, and a third person read in the test phase, we can easier control for whether the participants are noticing the tone patterns, instead of some other patterns due to voice characteristics. By using both a man and a woman's voice, they will listen to the stimulus material in different pitch, as women have higher pitch than men. This will also increase the ecological validity, as you hear both men's and women's voices daily in the real world. The ecological validity is also strengthened by using a natural language as stimulus material. Natural languages, in comparison to artificial languages, have more variations in both sound and grammar. Since most of us speak a natural language in the real world, a study using natural language strengthens the ecological validity. As mentioned, the stimulus material consisted of strings with different context sentences. There are 64 unique context sentences, which increases the variability. This strengthens the ecological validity, due to the variability in everyday speech. The RFH-cores appear before, in the middle as well as after the context sentences. This also increases the variability, and thus also the ecological validity. Artificial languages don't have this variability as natural languages.

In part B, the participants answered a questionnaire. Biases to be aware of related to questionnaires are response biases. In one of the questions, they were asked if they tried to look for certain patterns. This is perhaps a leading question. Some can consciously, or subconsciously, be induced to respond according to what they expect the study tries to achieve. To limit this bias as much as possible, we restricted the amount of information given to the participants. This prevented them from understanding the full extent of the research.

Reliability

The reliability of this experiment increases because it is very much possible to replicate. The same stimuli can be used by anyone who wishes to conduct the same experiment. The questionnaire can easily be translated, and the cognitive tests are commonly used in research. This makes it easy for other researchers to implement an identical experiment.

The procedure was as identical as possible for every participant. One of the researchers accompanied the participant into the computer room where they were exposed to the stimuli, while another researcher conducted the questionnaire and the cognitive tests. 40 participants volunteered for the experiment, so both the stimuli and the cognitive tests were presented 40 times. The experiment was carried out as a collaboration with four experimenters. All four experimenters tested 10 participants each. Experimenter bias was reduced due to the used of E-prime (Schneider et al., 2002), because the participants were exposed to the exact same stimuli. However, because we were four experimenters, we did not fulfil the exact same conditions for every participant while conducting the cognitive tests manually. To ensure that the reliability would still be good during the cognitive testing, the four researchers went through the instructions and tested each other so they could make sure that they all did the procedure as similar as possible. We also used a notebook to keep record of every possible deviation, so that any abnormal results could be traced back to the circumstances which it happened.

However, researcher error is still likely to occur, especially with four different people conducting tests. The way the participant was prepped before each test will vary, as well as measurement in the Stroop colour-word test and the FAS-test. Those two tests required the use of a timer, the judgement of the researcher is involved, like when to stop and start the timer. Human error will also be a possible weakness during these tests because of the researchers' reaction time. A factor to strengthen the reliability is that the experiment was conducted in the same room, in the same environment. The only noticeable environmental difference was the time of day the participants were tested, which could vary from 08.00 to 21.00. Participant changes might also be a relevant factor in this experiment because it lasted approximately an hour. The participant could experience drowsiness while sitting in the soundproof room listening to sentences and this could affect their performance. Some participants also showed up for the experiment directly after work, so tiredness could easily affect them.

Inter-rater reliability refers to measurements by several researchers using the same instrument (Polit & Beck, 2017). In total, four researchers conducted this experiment. This means that the reliability of the experiment is weakened because human observers will not necessarily interpret answers the same way. For example, the researchers might disagree as to what makes a valid answer or not. In the FAS-test test, it was discovered that the researchers had different opinions on whether names were a valid answer or not. We cleared it up so that names did not count as a valid answer and those rules were used for the remainder of the experiment. As far as the questionnaire goes, it was well designed to investigate awareness with the participants. However, there were some limitations. We should have designed the questionnaire so that the relevant questions for the awareness was the first thing the participants were asked to answer so the patterns they may have noticed was fresh in their short-term memory (Kormos, 1998). Instead, background questions like where they were from and what languages they spoke came first. We also discovered that the participants had troubles understanding the scale of how certain they were of their answer in the recognition task, but after we noticed we would point it out very carefully how they should respond. It might also be discussed whether the participants should have been informed in advance that they would be given a questionnaire, as this foreknowledge in itself might affect the performance of the participants (Kormos, 1998). However, the participants have the right to know what they are expected to be doing in the experiment. The reliability of the scores is increased because all four of the experimenters went over the scores several times, double checking that the results were correctly registered.

We compared the results from learning of a tonal structure to learning of syllable structures. The results show that participants learned the tonal structure slightly better than the syllable structures. It must be taken into account that the participants were only exposed for *one* tonal pattern (RFH), as opposed to *four* syllable structures (the four sentences in test category 1). Hence, to score as high on syllable and tone, one must learn four patterns in the syllable category, as opposed to one pattern in the tone category. To control for this, one could try a similar experiment with four tonal patterns in addition to four syllable patterns. Using the same amount of patterns is more comparable. Maybe the results would differ in a way that the score of syllable would be higher?

Ethical Considerations

There are a few ethical considerations that must be taken into account. The *Belmont Report* has three principles for protecting study participants (Polit & Beck, 2017). The first is to minimize harm and maximize benefits. The study must not inflict stress on the participants in any way, thus we must make sure the procedure isn't stressful, by making each participant sit alone in a quiet room with one researcher.

The second principle is respect for human dignity (Polit & Beck, 2017). The right to self-determination involves that the participant can volunteer to take part in a study but can at any point ask questions and answer as much or little as they like, they can refuse to give information or withdraw from the study (Polit & Beck, 2017). They can quit at any time, without being afraid of prejudicial treatment. No participant is going to be pressured to participate, they have their own free will to decide. We also fully describe the participants rights, our responsibility and potential risks and benefits. The participants must sign an agreement before performing the experiment.

Justice is the third principle, and that includes fair treatment of participants and their right to privacy (Polit & Beck, 2017). This is considered by anonymising every participant. The data they provide is also going to be kept confidential. The right to privacy ensures that the participants are not having to give more information about themselves than they need to (Polit & Beck, 2017). The data that are collected will be stored confidentially. The participants also received 150 NOK as monetary compensation for the hour they spent contributing to our study.

The participants had to sign a consent form to make sure they understood if they wanted to take part in the study or not. Informed consent was obtained by participants reading and signing the consent form in Attachments. By taking these ethical considerations into account, the participants' safety and rights will be preserved.

Limitations

The variability groups in the present study might be too similar. The results from the study of Gomez (2002) showed that infants and adults were able to discriminate familiarized words from novel words better with higher set sizes. Eidsvåg et. al. (2015) used set sizes 16 and 32, and the results shows that the participants exposed to the set size 32 showed evidence of learning before the participants exposed to the set size 16. The set sizes in our study is, as mentioned, 32 and 64. Since our set sizes are this large, the low variability might not be low enough, since it is a much larger set size than 12. Our low variability group can be compared to the high variability group of Eidsvåg et. al. (2015). We can therefore not conclude whether variability is a factor for learning. We suggest using a lower set size for a low variability group to see clear differences in input variability. The results of the present study also show that participants had higher acceptance rate for tonal cues. One reason for this could be due to them being exposed for targets with only one tonal pattern (RFH), as opposed to four syllable patterns ([lǎan mây lóm], [mǎa khwâaŋ kíaw], [mǔu phôn phít], and [sɯ̌a lɔ̂ɔ mót]). Thus, participants were exposed to the "correct" tonal structure more than each of the "correct" syllable structures. There are many studies assessing statistical learning of syllable cues, but not as many studies focusing on tonal cues. Thus, we were interested in investigating whether participants could show a learning effect when using tonal cues. The present study is therefore a pilot study with focus on tonal cues.

Our study had inclusion- and exclusion criteria regarding what kind of participants we wanted to test. One of the criteria was that they must not have any self-reported learning difficulties. To be more precise, we should also have added that they must not have any self-reported attention difficulties (such as ADHD), as this may affect the results of both the language learning experiment and cognitive tests. Another criteria for participating, was that participants must not have any previous experience with Thai. To control for pre-linguistic knowledge of linguistic tone, we may should have changed the criteria, and excluded participants who had previous experience with languages with a complex tone system (such as Mandarin).

During the language learning experiment, participants wore headphones. To make the input clearer, and limit potential noise in the room, sound isolating headphones might be a better alternative than ordinary headphones.

Regarding the four last questions in the questionnaire, where participants listened to audio clips, participants may choose the strings with RFH-pattern due to the prosody, rather than the tonal pattern. This is due to two of the clips that were parts of two context sentences, have a slightly different prosody than the two RFH-sentences. Since two of the clips are cut from context sentences, the words are less stressed, and are said with a higher speed. One of these sentences also had four syllables, in contrast to the other sentences, which had three syllables. To control for this, we could have had two extra sentences, with similar prosody, which the use was restricted to the questionnaire.

Regarding the question where participants were instructed to explain the pattern they potentially discovered during the experiment, it is hard to interpret exactly what the participants meant. Since the participants are Norwegians they answered the questionnaire in their first language. Almost half of the participants answered “tonefall”, which in this case can be translated to “cadences”. Thus, we cannot know for sure if they mean intonation or linguistic tone. The meaning of the word itself is a synonym to “intonation”. Since the participants probably not have studied linguistics, they might not be aware of this difference. Hence, they might mean linguistic tone, even though the word means intonation, or they might believe that they mean linguistic tone, when in fact they mean intonation. Maybe they don’t differentiate this distinction. This makes it complicated to interpret exactly what they mean.

Concluding Remarks

In the present study, we conducted a language learning experiment using new stimulus materials from a natural tonal language. We investigated whether adults were able to learn linguistic tonal cues. In addition, we examined whether executive functions influenced the performance in the language learning experiment. To investigate whether statistical learning is an implicit learning process, we administered a self-report questionnaire.

The results revealed that adults are able to discriminate familiarized words from novel words, with a preference for tonal cues (as opposed to syllable cues). However, it is somewhat unclear to what extent learning happened, as our findings reveal that even though there was an increase in acceptance rate for familiarized words (T&S), we found no change in acceptance rate for the unheard stimuli (None). The cognitive tests and the self-report questionnaire showed no systematic correlations with the participants' performance in the language learning experiment.

Because the cognitive tests showed only spurious correlations with the language learning experiment, executive functions could not explain variance in language learning in the present study. Since this stimulus material in our language learning experiment never have been used before, this study serves as a pilot study. To investigate participants' preference for either syllable or tonal cues, we suggest future studies to apply the same amount of tonal cues as for syllable cues.

References

Arciuli, J., & Torkildsen, J. (2012). Advancing Our Understanding of the Link between Statistical Learning and Language Acquisition: The Need for Longitudinal Data. *Front Psychol, 3*, 324. doi:10.3389/fpsyg.2012.00324

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research: past, present and future. *Philos Trans R Soc Lond B Biol Sci, 372*(1711). doi:10.1098/rstb.2016.0047

Asbjørnsen, A. E., & Helland, T. (2006). Dichotic listening performance predicts language comprehension. *Laterality, 11*(03), 251-262.

Asbjørnsen, A. E., & Hugdahl, K. (1995). Attentional effects in dichotic listening. *Brain and Language, 49*(3), 189-201.

Behme, C., & Deacon, S. H. (2008). Language learning in infancy: Does the empirical evidence support a domain specific language acquisition device? *Philosophical Psychology, 21*(5), 641-671.

Berggreen, H., & Tenfjord, K. (1999). *Andrespråkslæring* (2. ed.). Oslo: Gyldendal Norsk Forlag AS.

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics, 27*(2), 185-209.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*: Greenwood Publishing Group.

Conklin, H. M., Curtis, C. E., Katsanis, J., & Iacono, W. G. (2000). Verbal working memory impairment in schizophrenia patients and their first-degree relatives: evidence from the digit span task. *American Journal of Psychiatry, 157*(2), 275-277.

Eidsvåg, S. S., Austad, M., Plante, E., & Asbjørnsen, A. E. (2015). Input variability facilitates unguided subcategory learning in adults. *Journal of Speech, Language, and Hearing Research, 58*(3), 826-839.

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*, 66-108.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical Learning in Children With Specific Language Impairment. *Journal of Speech Language and Hearing Research, 52*(2), 321-335. doi:10.1044/1092-4388(2009/07-0189)

Friederici, A. D., Mueller, J. L., & Oberecker, R. (2011). Precursors to natural grammar learning: preliminary evidence from 4-month-old infants. *PLoS One, 6*(3), e17920. doi:10.1371/journal.pone.0017920

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in cognitive sciences, 19*(3), 117-125.

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 34.

Golden, C. J., & Freshwater, S. M. (1978). *Stroop color and word test*: Stoelting Chicago.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychol Sci, 13*(5), 431-436. doi:10.1111/1467-9280.00476

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition, 32*(3), 465-491.

Hamrick, P., & Rebuschat, P. (2012). How implicit is statistical learning. *Statistical learning and language acquisition*, 365-382.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science, 26*(4), 433-443.

Hugdahl, K. (1995). Dichotic listening: Probing temporal lobe functional integrity. *Brain asymmetry, 1*, 123-156.

Hugdahl, K. (2004). Dichotic Listening in the Study of Auditory Laterality. *The asymmetrical brain*, 441.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017-2036.

Hurks, P., Vles, J., Hendriksen, J., Kalff, A., Feron, F., Kroes, M., . . . Jolles, J. (2006). Semantic category fluency versus initial letter fluency over 60 seconds as a measure of automatic and controlled processing in healthy school-aged children. *Journal of Clinical and Experimental Neuropsychology, 28*(5), 684-695.

Kaminska, A. (2007). Distribution of word classes in old English and old high German: A preliminary contrastive study based on The battle of Maldon, Hildrebrandslied and Ludwigslied. *Research in Language*(5), 135-146.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116*(3), 321-340.

Kimura, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology/Revue canadienne de psychologie, 15*(3), 166.

Konig, E., & Van der Auwera, J. (2013). *The germanic languages*: Routledge.

Kormos, J. (1998). Verbal reports in L2 speech production research. *TESOL Quarterly, 32*(2), 353-358.

Krashen, S. D. (1981). The "fundamental pedagogical principle" in second language teaching. *Studia Linguistica, 35*(1-2), 50-70.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition, 22*(4), 557-584.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*: Univ of California Press.

Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annu. Rev. Linguist., 1*(1), 333-353.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(3), 251-285.

Ludden, D., & Gupta, P. (2000). *Zen in the art of language acquisition: Statistical learning and the less is more hypothesis.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

Lund-Johansen, M., Hugdahl, K., & Wester, K. (1996). Cognitive function in patients with Parkinson's disease undergoing stereotaxic thalamotomy. *Journal of Neurology, Neurosurgery & Psychiatry, 60*(5), 564-571.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of child language, 31*(4), 883-914.

Marton, K. (2008). Visuo-spatial processing and executive functions in children with specific language impairment. *International Journal of Language & Communication Disorders, 43*(2), 181-200.

Morén, B., & Zsiga, E. (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory, 24*(1), 113-178.

Nakatani, H., & Minegishi, M. (2011). Description of Thai as an isolating language. *Social Science Information, 50*(1), 62-80.

Nelson, D. G. K., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant behavior and development, 18*(1), 111-116.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Dev, 80*(3), 674-685. doi:10.1111/j.1467-8624.2009.01290.x

Polit, D. F., & Beck, C. T. (2017). *Nursing research: Generating and assessing evidence for nursing practice* (Tenth ed.): Wolters Kluwer.

Raczkowski, D., Kalat, J. W., & Nebes, R. (1974). Reliability and validity of some handedness questionnaire items. *Neuropsychologia, 12*(1), 43-47.

Rasmussen, T., & Milner, B. (1977). The role of early left-brain injury in determining lateralization of cerebral speech functions. *Annals of the New York Academy of Sciences, 299*(1), 355-369.

Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. New York: Oxford University Press.

Rebuschat, P. (2015). *Implicit and explicit learning of languages* (Vol. 48): John Benjamins Publishing Company.

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015). Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition, 37*(2), 299-334.

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33*(04), 829-856.

Reitan, R. M., & Wolfson, D. (1990). A consideration of the comparability of the WAIS and WAIS-R. *Clinical Neuropsychologist, 4*(1), 80-85.

Rissanen, M., Ihalainen, O., Nevalainen, T., & Taavitsainen, I. (1992). *History of Englishes: New methods and interpretations in historical linguistics* (Vol. 10): Walter de Gruyter.

Saffran, J. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*(1), 172-196. doi:10.1006/jmla.2001.2839

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

Saffran, J., & Estes, K. G. (2006). Mapping sound to meaning: Connections between learning about sounds and learning about words. *Advances in Child Development and Behavior, Vol 34, 34*, 1-38. doi:Doi 10.1016/S0065-2407(06)80003-0

Saffran, J., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52.

Saffran, J., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*(2), 273-284. doi:Doi 10.1207/S15327078in0402_07

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*(4), 110-114. doi:Doi 10.1111/1467-8721.01243

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of aging, 30*(4), 507-514.

Sandoval, M., & Gomez, R. L. (2013). The development of nonadjacent dependency learning in natural and artificial languages. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(5), 511-522.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69*(2), 105-134.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide.* Pittsburgh, PA: Psychology Software Tools.

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

SPSS IBM. (2011). IBM SPSS statistics for Windows, version 20.0. *New York: IBM Corp*.

Statistisk sentralbyrå. (2017, 15.06.2017). Befolkningens utdannigsnivå. Retrieved from https://www.ssb.no/utniv

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6), 643.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*(2), B25-B34.

Van der Haegen, L., Westerhausen, R., Hugdahl, K., & Brysbaert, M. (2013). Speech dominance is a better predictor of functional brain asymmetry than handedness: A combined fMRI word generation and behavioral dichotic listening study. *Neuropsychologia, 51*(1), 91-97.

VanPatten, B., & Rothman, J. (2015). What does current generative theory have to say about the explicit-implicit debate? In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 91-116).

Vanvik, A. (1973). A phonetic-phonemic analysis of Standard Eastern Norwegian. Part II. *Norwegian Journal of Linguistics, 27*, 101-139.

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Front Psychol, 5*, 953. doi:10.3389/fpsyg.2014.00953

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning, 54*(4), 681-712.

Wierzchoń, M., Asanowicz, D., Paulewicz, B., & Cleeremans, A. (2012). Subjective measures of consciousness in artificial grammar learning task. *Consciousness and cognition, 21*(3), 1141-1153.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*(2), 269-304.

Learning of Linguistic Tonal Cues Follow the Same Learning Principle as for other Linguistic Cues

Elin Varissara Arnholm & Julia Lorgen Hartvigsen

Department of Biological and Medical Psychology
Faculty of Psychology

Authors' note

Elin Varissara Arnholm, Master's Programme in Health Sciences - Logopedics, University of Bergen.

Julia Lorgen Hartvigsen, Master's Programme in Health Sciences – Logopedics, University of Bergen.

Special thanks to Theeraporn Ratitamkul at the Department of Linguistics, Chulalongkorn University, for constructing and recording the Thai stimulus material. We also thank our fellow students, Alexander Grøndalen Kristiansen and Erling Wogn-Henriksen, for the collaboration during the construction and conducting of the experiment. Thanks to Arve Asbjørnsen, Professor at the Department of Biological and Medical Psychology, University of Bergen, for supervision on the present study.

Abstract

This study investigated statistical learning in a natural tonal language, and whether executive functions had an impact on language learning in such learning conditions. The 40 adults that participated were exposed to Thai sentences. They were then tested with four cognitive tests. The results showed that adults were able to discriminate familiarized words from novel words, in addition to having a preference for tonal cues. There were no correlations between executive functions and language learning ability. The results support previous studies that investigate statistical learning of tonal cues.

Keywords: Statistical learning, tonal language, Thai, executive functions, awareness

Learning of Linguistic Tonal Cues Follow the Same Learning Principle as for other Linguistic Cues

Statistical Learning of Tonal Cues

Results from studies on statistical learning of syllable structures suggest that both infants and adults are able to use statistical properties of linguistic input to detect patterns in language (Friederici, Mueller, & Oberecker, 2011; Gomez, 2002; Pelucchi, Hay, & Saffran, 2009; Saffran, Aslin, & Newport, 1996). In the present study we investigated whether these learning mechanisms apply to learning of linguistic tone, and to what extent differences in executive functions can explain variance in language learning.

It was of interest to investigate whether the findings on the syllable level applied to the suprasegmental level. Linguistic tone is a suprasegmental phoneme, which means it is a phoneme that may extend over series of segmental phonemes. Even though Saffran, Johnson, Aslin and Newport (1999) studied statistical learning by using non-linguistic tone, their findings are relevant for learning more about how statistical learning applies to tonal cues. They were curious whether the same statistical learning ability could be applied to musical tones. Infants and adults were exposed to six sequences consisting of three different tones following each other. The results indicated that the learning mechanisms also can be applied to non-linguistic-stimuli, such as tones. According to Saffran et al. (1999) these findings suggest that linguistic stimuli are not privileged to statistical learning processes. The same ability to detect patterns can be applied to tones. However, the significance of tone sequence segmentation is less evident than word segmentation (Saffran, Johnson, Aslin, & Newport, 1999).

Wang and Saffran (2014) conducted three experiments with certain tone patterns applied to an artificial tonal language. The language was trisyllabic, where each syllable had a tone. In the first experiment, English monolingual adults participated. The results revealed that the participants did not discriminate familiarized words from new words better than chance. The second experiment consisted of Mandarin monolingual and Mandarin-English bilingual participants. The Mandarin monolinguals did perform better than chance, but the Mandarin-English bilinguals outperformed the monolinguals. In the last experiment, the participants were non-tonal bilinguals of different origins. The non-tonal bilinguals performed significantly above chance, only outperformed by the Mandarin-English bilinguals.

Results from this study suggest that being bilingual is a major advantage when tracking regularities in a tonal language, at least with the task they were given in this study (Wang & Saffran, 2014). As such, Wang and Saffran (2014) laid the foundation for the research on how statistical learning operates at the suprasegmental level by asking whether adults can track regularities in a tonal language.

Wayland & Guion's (2004) study had similar findings as the study by Wang and Saffran (2014). Thai, which is a tonal language, was used as stimulus material. The mid and low tones were contrasted and compared. The participants were English, Chinese and Taiwanese. The Chinese participants were superior to the English participants at discriminating phonetic variations in Thai (Wayland & Guion, 2004). Many Chinese languages are also tonal languages, which indicates that pre-linguistic knowledge is a factor for language learning.

Many of the world's languages are tonal languages, and it is therefore of interest to study learning of linguistic tones. The study by Wang and Saffran (2014) mentioned above, was conducted using an artificial language as stimuli. Despite the fact that Mandarin-monolinguals have pre-linguistic experience with tone, they were outperformed by other bilinguals (Wang & Saffran, 2014). One of the reasons for this, could be that the artificial language they were exposed to was three-syllabic. Since Mandarin is a disyllabic language, the other participants with experience from languages with polysyllabic words, may have had a pre-linguistic advantage, in regard to morphology.

In the present study, we investigated the learning of tones, using stimuli from a natural language. To our knowledge, there are not many studies on statistical learning in relation to linguistic tones. It is therefore of interest to investigate whether the results from this study support the findings from Wang and Saffran's (2014) study. The monolingual English-speaking participants did not perform above chance. In contrast to English, Norwegian have two linguistic tones. The present study investigated whether Norwegian speaking adults were able to track tonal cues. Do Norwegian adults have an advantage, considering Norwegian is a tonal language?

Since this study applied a natural language as stimulus material, the ecological validity also increased. In addition to investigating whether participants were able to learn a tonal pattern, we applied tests of executive functions to uncover whether executive functions correlate with language learning abilities. A questionnaire was also administered to detect if participants were aware of learning. We also asked whether the findings from Gomez (2002) and Eidsvåg et al. (2015), concerning variability, applied to learning of tonal cues.

The stimulus material used in the present experiment, consisted of Thai phrases containing a tonal pattern. Thai is a language spoken in Thailand, and is characterized as an isolating language (Nakatani & Minegishi, 2011). What characterizes an isolating language, is among other things, the lack of morphological inflection. The language has a simple morphology, but a complex tone system. Thai monosyllabic words are tone bearing units. Standard Thai has five different lexical tones. Linguists have described these tones as mid, low, high, falling and rising (Morén & Zsiga, 2006). The word เสือ [sǔa] has a rising tone and means “tiger”. If we change the rising tone to a falling tone, as in เสื้อ [sûa], the word changes lexical meaning to “shirt”. This is in contrast to Norwegian, which has a simple tone system with only a two-way basic contrast (Vanvik, 1973). English on the other hand, is not a tonal language, and do not contrast word meanings only by pitch. By investigating statistical learning in a language like Thai, we can increase our knowledge on how statistical learning operates regarding tonal cues.

The words from the stimulus material used in Wang and Saffran’s study, consisted of three syllables, with each syllable having a certain tone. Due to the low morpheme per word ratio in isolating languages, words like this do not usually exist. There are a few words in Thai consisting of more than one syllable. However, a trisyllabic word does not have three tones, but rather a stressed tone bearing syllable. In contrast, the present study is used syllables and words that actually exist in isolating languages, which increased the ecological validity.

The Role of Executive Functions in Statistical Learning

In the present study we examined some cognitive abilities. We measured the participants’ executive functions by using tests that challenged the participants’ attention, inhibition, verbal fluency and working memory. These traits have been thought to be of importance for language learning (Brooks, Kempe, & Sionov, 2006; Evans, Saffran, & Robe-Torres, 2009; Winke, 2005). Executive function, or cognitive control, is a term for the many cognitive processes that ensures controlled cognitive processing (Rodriguez-Fornells, De Diego Balaguer, & Münte, 2006). These functions make us capable of controlling our attention and actions by overruling dominant responses. They are crucial in demanding situations where you need to adapt to change, for instance in language learning (Huizinga, Dolan, & van der Molen, 2006). Toro, Sinnett, and Soto-Faraco (2005) found that when participants are distracted, it compromised their ability to extract words. Executive functions play a part in the regulation and control of the language you currently speak. Bilinguals, for instance, can separate languages without much effort, using inhibition to partially suppress the language not in use at the moment (Rodriguez-Fornells et al., 2006). When an adult is learning a second language there are a lot of linguistic input that needs to be acquired simultaneously. They need to consider aspects of the language such as word boundaries, syntax and morphology (Brooks et al., 2006; Evans et al., 2009; Gomez & Gerken, 1999; Rodriguez-Fornells et al., 2006; Thiessen & Saffran, 2003). Adult second language learners also have a set of rules from their native language. They need to supress these existing rules to an extent, in order to learn the new rules. In addition, they must pay enough attention to notice these new rules (Toro et al., 2005). Thus, second language learning could be a cognitively challenging task.

A study by Hartshorne and Germine (2015) show that you peak at different executive functions at different ages. For instance, you might not reach peak performance in vocabulary until you are about 50 years old, but you peak at cognitive tasks like “Digit span” when you are between 20-25 years old. Can we explain variance in language learning by referring to differences in executive functions?

Studies suggest that people with specific language impairment (SLI) or language-based learning disabilities (LLD) need more variability or longer exposure before they are able to generalize grammar from an artificial language (Evans et al., 2009; Torkildsen, Dailey, Aguilar, Gómez, & Plante, 2013). This variety in language learning could possibly be explained by differences in executive functions such as working memory and attention capacities. Ludden and Gupta (2000) learned that performance by adults is decreased in statistical learning tasks when memory and attention are reduced.

However, Plante, Vance, Moody, and Gerken (2013) did an experiment where they wanted to investigate whether there was a difference in the way children with and without SLI conceptualize artificial language input, but found no differences between these groups.

Various studies examining executive functions in people with SLI indicate that there is a difference in executive functions between people with and without SLI (Marton, 2008; Montgomery, 2003; St Clair-Thompson & Gathercole, 2006). Marton (2008) found that working memory have an impact on performance in children with SLI, and that tasks requiring a great amount of executive functions are more challenging for these children than for typically developing children. These children also showed weaker abilities in inhibition and attention control tasks than their typically developing peers (Marton, 2008).

Cognitive abilities, such as verbal working memory and the ability to focus attention, have been suggested to explain individual differences in how successful second language learners will be at learning (Brooks et al., 2006; Winke, 2005).

Brooks, Kempe and Sionov (2006) also hypothesized that successful learning depends on the learners’ cognitive abilities in addition to the variables in the input. In this experiment, 60 adults were tested with Russian masculine and feminine nouns. The nouns were associated with a drawing to match the meaning. There were six training sessions and a test phase. After their language learning task, they conducted three tests to examine individual differences in verbal working memory, phonological memory and nonverbal intelligence (Brooks et al., 2006). They found some correlations in that the nonverbal intelligence task predicted learners' success. That might indicate that there is a correlation between cognitive abilities and language learning abilities. Since there is a limited amount of research on cognitive abilities' impact on language learning in general (Brooks et al., 2006), we wanted to investigate further whether we could find similar results regarding executive functions and statistical language learning.

Reber (1993) believed that implicit learning is independent of intelligence. As opposed to explicit learning, which he believed is dependent on intelligence. Studies that support this hypothesis are the studies by Gebauer and Mackintosh (2007) and Kaufmann (2010). Gebauer and Mackintosh (2007) found no correlation between implicit artificial learning tasks and different measures of intelligence. However, they did find a significant correlation between scores on the learning tasks and the measures of intelligence when the participants were given explicit rules (Gebauer & Mackintosh, 2007). Kaufman et al. (2010) concluded that implicit learning was related to some personality traits, but not to psychometric intelligence or working memory.

Rebuschat (2015) presents implicit learning, statistical learning and second language acquisition as three approaches to the same phenomenon. Implicit learning is a fundamental feature of human cognition, and it is the process of acquiring unconscious (implicit) knowledge (Rebuschat, 2015). Implicit learning is in other words incidental, and non-intentional. In contrast, explicit learning is awareness of what one is learning. The role of awareness is crucial in the implicit/explicit debate and has been assessed in various statistical learning studies (Hama & Leow, 2010; Leow, 2000; Rebuschat, Hamrick, Riestenberg, Sachs, & Ziegler, 2015; Williams, 2005).

To what extent does awareness play a role in statistical learning? Do participants acquire implicit or explicit knowledge? While offline measures of awareness measure whether exposure results in unconscious knowledge, online measures measure the role of awareness at the time of encoding (Rebuschat et al., 2015). Williams (2005) collected retrospective verbal reports (offline measures) by interviewing the participants *after* the learning experiment. He concluded that learning without awareness is possible (Rebuschat et al., 2015). Leow (2000) used think-aloud protocols (online measures) to measure awareness in foreign language behaviour, and participants had to verbalize their thoughts *during* the experiment. The results suggest that there is no evidence for learning without awareness. Rather, it supports the claim that awareness is crucial in processing of a second language. These methodological differences result in measures of awareness in different stages. While Williams measured the *product* of learning, Leow measured awareness of the *process* of learning (Rebuschat et al., 2015).

Rebuschat, Hamrick, Riestenberg, Sachs & Ziegler (2015) compared three measures of awareness: retrospective verbal reports, concurrent verbal reports (think aloud protocols) and subjective measures (confidence ratings and source attributions). The findings revealed that the groups who did not think aloud performed best. The subjective measures revealed that the participants were confidence in answering and had acquired knowledge of rules. The verbal reports indicated that a considerable number of participants began to actively search for rules during the test phase and reported awareness. In other words, the results indicate that the participants had acquired both explicit and implicit knowledge (Rebuschat et al., 2015).

These differences in methods should be taken account for when studying awareness. By using more than one type of measure of awareness, one can collect more detailed information. Thus, in the present study we apply a self-report questionnaire which contains both retrospective verbal reports, and confidence ratings.

The Present Study

In the present study, we expose young adults to a tone pattern, consisting of three different tones in a specific order (rising-falling-high). The relationship between these tones can be considered an adjacent dependency. The tones we apply in this study are tones that exist in Thai. However, there are no syntactic rules for distribution of tones in Thai sentences. Hence, the tone pattern/adjacent dependency is constructed for this experiment. Each tone is paired with a syllable. We designed the experiment in a way that we can test if participants will be able to learn the regularities with either syllable cues, tonal cues or both. We ask if adults are able to detect tonal cues in a natural language using statistical learning mechanisms. On the basis of previous research, regarding input-variability (Eidsvåg, Austad, Plante, & Asbjørnsen, 2015; Gomez, 2002), we expect to find that higher variability leads to higher acceptance rate of familiarized words.

We also apply cognitive tests to measure whether executive functions (attention, inhibition, verbal fluency and working memory) influence the participants' performance in the language learning experiment. In addition, we investigate whether learning happens with awareness, by applying retrospective verbal reports and confidence ratings in a self-report questionnaire.

Method

Participants

40 Norwegian adults participated in this study, 20 females (M = 22;9 [years;months]) and 20 males (M = 25;5 [years;months]). Every participant had Norwegian as their native language, normal hearing, and no previous experience with Thai. None of the participants had any self-reported learning difficulties. Every participant spoke English in addition to their native language.

The participants were assigned to two groups (high vs. low variability). To make sure we had an even distribution of sex in each group, the participants were pseudorandomized. Two participants were excluded due to an error in the setup of the low variability stimuli. This error was discovered after the two first participants had already been tested. Thus, two additional participants were recruited.

Stimulus Material for the Language Learning Experiment

In the familiarization phases, the participants listened to a set of auditory strings. Targets consisted of four correct three-syllable sentences with a RFH-tonal pattern (rising, falling, high). See Table 1 for target sentences. The full stimulus material is available in the Appendix. These syllables constituted adjacent dependencies. Each of these targets were embedded in 16 grammatically correct Thai strings, where the target sentence was either in the beginning, the middle or the end of the strings. This created a total of 64 unique strings as context for the targets. One of the targets was [lǎan mây lóm]. An example of the target embedded in a grammatically correct string is [lǎan mây lóm tàe kɔ̂ɔ rɔ́ɔŋ hâay] (tonal pattern R-F-H-L-F-H-F). The only thing these 64 strings had in common, was the rising-falling-high tone pattern in the target. In addition, the strings were recorded with both a male and a female voice. Thus, giving 128 unique auditory stimuli.

The high variability group was familiarized to the 64 unique strings two times. The low variability group was presented for half of the strings (32 unique strings) four times. Thus, all the participants were presented for 128 strings in random order in each familiarization phase.

In the test phases, a third native Thai speaker presented the test strings. A message on the computer screen appeared. The message informed the participants that they would hear different sentences, and they were instructed to judge which of the sentences were grammatically correct. Every item was grammatically correct Thai sentences, but "grammatically correct" in our experiment meant strings which had the R-F-H tone pattern. Immediately after they were presented for each string from the test phase, a smiley face and a frowny face appeared on the screen. If they thought the sentence was grammatically correct, they were instructed to click on the smiley face. If they thought the sentence was grammatically incorrect, they were instructed to click on the frowny face.

These test strings were divided into four categories. Category one consisted of four strings, namely the four unique target strings, with the rising-falling-high tone pattern, from the familiarization phase (see Table 2.1).

Category two consisted of four three-word strings, which had the same tonal pattern as the target strings, but with different syllables. In other words, the syllable structure (words) in the strings were different from category one, but the number of syllables, as well as the tone pattern (R-F-H), were the same see Table 2.2).

Category three also consisted of four three-word strings, but unlike category two, the strings had the same syllable structure as category one. Here, it was the tone pattern that was different. Every syllable in this category was the same as the target. However, the tone was different, which changes the meaning of the words. Every string had its own tonal structure (M-L-F, H-R-F, L-H-L, F-L-L) (see Table 2.3).

Category four consisted of four three-word strings. Both the syllable structure and the tonal pattern in these strings were different from category one/the targets in the familiarization phase (see table 2.4).

In order to perform well in the test phases, the participant had to detect the relationship between the adjacent dependency in the tonal pattern and/or syllable pattern.

A pre-familiarization test was also applied as a control for whether the results from the test phases that follow, significantly differ from chance. In this pre-familiarization test, we expected their performance not to be better than chance.

Measures of Executive Functions

To measure executive functions, we conducted four recognized tests. The participants were asked to answer 15 questions about hand preference by circling the letters H (right), B (both) or V (left) for each of the 15 tasks the questionnaire asked about. One example of the questions asked was “with which hand do you draw?”. A test of this kind can provide signs concerning cerebral representation of speech (Rasmussen & Milner, 1977). The hand preference questionnaire was a Norwegian translation of a questionnaire developed by Raczkowski, Kalat, and Nebes (1974).

Dichotic listening is a technique used to study a range of cognitive processes related to brain laterality, hemispheric asymmetry, attention, learning and memory (Hugdahl, 1995). The dichotic stimuli consisted of paired presentations of sounds containing one of six stop-consonants /b, p, d, t, k, g/ and the vowel /a/ (Asbjørnsen & Hugdahl, 1995; Hugdahl, 2004). Syllable pairs such as /ba/ and /ga/ was presented to the participant simultaneously to the right and left ear. In total there were 36 pairs including six homonymic pairs. This was combined with attentional instructions. Attentional instructions are given to either non-forced or attention forced; to the right ear input or the left ear input. The change in performance during the attention tasks yield a measure of executive functions (Asbjørnsen & Hugdahl, 1995). In the non-forced round, the participants were only instructed to repeat the most distinct sound they heard. In the forced left round, the attentional instructions were to repeat what they heard on their left ear, thus isolating one specific sound. In the forced right round, the participants were asked to repeat the sound they heard in their right ear.

The order in which they were given the forced left or the forced right round after the non-forced round was randomized. Half of the high variability group and half of the low variability group were tested first with the non-forced condition, then with the forced left condition and lastly with the forced right condition. The other half of both groups had attentional instructions in the order non-forced, then forced right and lastly forced left.

Stroop colour-word test measured the participant's response inhibition ability (Spreen & Strauss, 1991). The Stroop-test used in this experiment was inspired by Lund-Johansen, Hugdahl, and Wester (1996). The test consists of three pages.

On the first page, the words were printed in regular, black ink. The words "RED", "BLACK", "BLUE", "GREEN", "YELLOW", "WHITE" (in Norwegian) were written first in a separate line as a trial. Then, the words were randomly repeated in a section below, functioning as the test. The participant was instructed to read the words in the section as fast as he managed.

On the second page, there were coloured circles. The colours were the same as those written on the first page. The participant was told to say the colours out loud as fast as he managed.

On the last page, the same words as on the first page were printed, but in colours. However, the words that were written and the colours the words were printed in do not necessarily match. The word "RED" could for instance be printed in yellow ink. The instruction is to say the colour of the word, not reading the written word, thus suppressing the verbal content.

The Digit Span test was taken from the Wechsler Adult Intelligence Scale (WAIS) (Reitan & Wolfson, 1990). This test contains two parts, Digit Span forwards and Digit Span backwards.

In part one, Digit Span forwards, the test leader started with a span of three numbers read with approximately one second intervals. When the numbers were presented, the participant repeated back the numbers he remembered. The amount of numbers per span increased with one for each sequence the participant successfully repeated. If the participant did not remember the digits, or repeated the span wrongly, he would get a second attempt at another digit span of the same length. If both attempts failed, or he managed to repeat the longest span of 9 numbers, part two of the test began.

The second part, Digit Span backwards, utilized the same rules where the spans increased with one number and the participant had two attempts per span. The instructions differed in that the participant were supposed to repeat digits in reverse order. The Digit Span backwards task started with two numbers and increased for each one correctly repeated. The test was finished at eight numbers or when the participant failed both attempts.

The last test was the FAS-test. The participants were asked to find as many words as possible starting with the letters F, A and S in three respective rounds (Spreen & Strauss, 1991) of 30 seconds. The participants were told that they were going to be presented with a letter, and that they would have 30 seconds to produce as many words as possible starting with said letter. They were also told to avoid names. They were presented with each letter just before the timer started. All the words were written down by the test leader, so that eventual repeats were accounted for.

Measures of Awareness

The questionnaire is based on the method of Rebuschat et al. (2015) and modified after Eidsvåg et al. (2015). The questionnaire contained questions about language background (dialect, language spoken, knowledge about Thai), as well as questions surrounding their awareness of patterns. The main intention of the questionnaire was to gather information about the participants' awareness of the tonal pattern (RFH). The questionnaire served as an offline measure containing retrospective verbal reports and subjective measures. Participants were asked which test phase they thought they managed best and why, to what extent they thought their answers were arbitrary (not at all, a little, to a large extent or to a very large extent), how aware they were of looking for regularities (not aware, a little aware, aware, aware to a large extent, aware to a very large extent), and if they had found any patterns (not at all, a little, to a large extent, to a very large extent). If they found certain patterns, they were asked to describe the pattern they had found. They were then asked to rate how correct they thought their pattern was (not at all, a little, to a large extent or to a very large extent).

Confidence ratings was collected in the four last questions of the questionnaire. Participants listened to four audio clips. After each clip, they were instructed to determine whether it was one of the sentences they had listened to in the language learning experiment. In addition, they had to rate how confident they were in their answers on a scale from 0 to 100 (0=not confident, 100=confident). Two of the clips contained an RFH-sentence from the target. The two remaining clips contained parts of two context sentences used in the familiarization phase.

Procedures

The language learning experiment. To recruit participants, we used consecutive sampling. That is, every person who signed up for the experiment and fulfilled the inclusion criteria was accepted until we reached our desired amount. Recruitment was done by hanging up posters in areas/campuses of University of Bergen. Posters were also published on social media. The experiment took place in a research lab at University of Bergen. Upon arrival in the laboratory, the participants were screened with the inclusion and exclusion criteria and signed the consent form. The approved participants were seated in front of the computer, in a soundproof room, with headphones. To prevent possible experience of discomfort, one of the experimenters always went in the soundproof room with them. To conduct the experiment, E-prime 2.0 Professional (Schneider, Eschman, & Zuccolotto, 2002) was used to present the stimuli and to collect data.

During the pilot experiment, we discovered that the stimulus material was presented too slowly. Thus, we shortened the time between the strings. To make the familiarization sentences resemble natural speech, the audio clips were played with 0,5 seconds silence in between each string.

The familiarization and test phases were repeated a total of three times. The total duration of the language learning experiment was approximately 30 minutes.

The participants went through four cognitive tests. They were first asked to answer a handedness questionnaire containing 15 tasks of hand preference. The first test of executive function administered was dichotic listening, where we measured the ability to control attention and focus on sounds presented to different ears. After the dichotic listening test, we conducted the Stroop colour-word test which challenges the participants' ability to inhibit responses. The third cognitive task was the Digit Span task from WAIS to investigate the participants' working memory capacity. The last test the participants went trough was the FAS-test that required them to list words starting with a certain letter, thus examining their ability for verbal fluency. The cognitive tests took about 20 minutes to conduct.

A written self-report questionnaire was also administered. The researcher was in the same room, in case the participants had any questions. In the four last questions, they listened to four auditory strings through headphones. It took approximately five minutes to answer the self-report questionnaire.

Statistical Analysis

Learning was defined as an increase in acceptance rate for the target items, combined with reduction in acceptance rate of the so called ungrammatical items, consisting of the None-category. In addition, acceptance rates of the two categories of T and S will indicate a preference for either tonal or syllable information. Thus, we analysed the test results with a mixed ANOVA, using the experiment groups (high vs. low variability) as the between factor, and the acceptance rate for the four categories as dependent measures. SPSS Statistics (SPSS IBM, 2011) was used to analyse the results.

Results

To measure whether the results from the language learning experiment showed a learning effect, two analyses of variance (ANOVA) were conducted with two experimental groups (high vs. low variability), and the cycles (1-4) entered as between groups, independent factors. The dependent variables were the acceptance rates of the categories. Since the experiment was conducted comparing two groups, the high/low variability was treated as a between-subject factor.

An ANOVA including experimental group as independent variable and acceptance rate of the T&S and N as dependent measures, yielded a main effect of stimulus category ($F(1, 38) = 92.4$, $\eta_p^2 = 0.709$ $p<0.05$), with higher acceptance rate for the T&S stimuli. This is illustrated in Figure 1. The results from both the high variability group and the low variability group revealed that the acceptance rates were significantly above chance in category T&S (one sample t-test $p<0.05$). There was also a significant effect of cycles ($F(3, 114) = 10.34$, $\eta_p^2 = 0.246$ $p<0.05$) with an increase in acceptance over cycles.

In addition, these two categories of test items also gave an interaction effect with cycles ($F(3, 114) = 4.44$, $\eta_p^2 = 0.105$, $p<0.05$), the effect is due to no change in the None category, but an increase in the T&S category. A post hoc test with a paired sample t-test showed that participants had significantly higher acceptance rate in T&S in the last cycle than in the first cycle (pre-familiarization test) ($p<0.05$). There were no significant findings comparing acceptance rates for cycles one and four of the N category (paired sample t-test $p>0.05$). Thus, the results from the post hoc analysis (paired sample t-test) of cycle one and four of T&S indicate that there is a learning effect in category T&S (syllable and tone), since there is an increase in the acceptance rate of the correct items. The acceptance rate for the N-category was significantly below chance (one sample t-test $p<0.05$). The results from the post hoc analysis (paired sample t-test) did not show a decrease in acceptance. Thus, it does not indicate that there is a learning effect in the N category.

We found no main effect of the two experimental groups in the analysis (ANOVA) of N and T&S ($F(1, 38) = 2.227$, $\eta_p^2 = 0.055$ $p>0.05$).

To uncover whether there was a preference for either tonal or syllable cues, we conducted a similar ANOVA analysis with the remaining categories Tone (T), and Syllable (S). We found a main effect of category: $F(1, 38) = 5.402$, $\eta_p^2=0.124$ $p<0.05$. This is illustrated in Figure 2. Thus, participants had higher acceptance rates for the Tone category than the Syllable category. The results from both the high variability group and the low variability group reveal that the acceptance rates were significantly above chance in categories T and S (one sample t-test $p<0.05$). The results from the pre-test are also significantly above chance, which indicate that there is no learning effect for the S and T categories. There were no main effect in the cycles ($F(3, 114 = 2.096$, $\eta_p^2= 0.052$ $p>0.05$), or of the two experimental groups in the analysis of the T and S ($F(1, 38) = 1.026$, $\eta_p^2= 0.026$ $p>0.05$).

To investigate if there were significant correlations between the scores of the cognitive tests and the acceptance rates of the language learning experiment, we executed a Pearson's *r* correlation analysis. Interestingly, we found no systematic patterns of correlations between any of the tests that assumingly should measure different aspects of cognitive control, except for a spurious correlation of $r=.35$ between the right ear score during dichotic listening (Forced Left-condition) and the acceptance rate of the None category of the third cycle. In addition to errors on the Stroop colour-word test and the T&S category during the third cycle $r=.31$.

The self-report questionnaire's main purpose was to measure the participants' awareness. To measure whether there were significant correlations between the acceptance rates for the four categories (None, Tone, Syllable, and Syllable and Tone) in cycle four, and questions from the questionnaire, we executed a Pearson's *r* correlation. The results revealed no significant findings.

Regarding the question about whether they had found any patterns during the language learning experiment, most of the participants described their thoughts. Since the participants were Norwegian, they answered the questionnaire in their native language. 15 of the participants answered that they had detected a cadence-pattern. In Norwegian they answered "tonefall", which can be interpreted ambiguously. In this case, we translate this to "cadence", which the English Oxford Dictionary defines as "A modulation or inflection of the voice. A fall in pitch of the voice at the end of a phrase or sentence." (Oxford University Press, 2018). The remaining 25 participants answered patterns unrelated to cadence/tone or did not report any pattern. Thus, we divided the results from this questions in two groups: cadence vs. no cadence. To see if the results between the groups differed, we did an analysis of variance (mixed ANOVA) to analyse the acceptance rate of the categories None, Syllable, Tone, and Tone & Syllable. The independent variables are the two pattern groups (cadence vs. no cadence), and the cycles (1-4). The dependent variables are the four categories (Tone and syllable, tone, syllable, and none). The cadence vs no cadence is a between-subject variable.

The results from the ANOVA revealed no main effect of groups (cadence/no cadence): $F(1, 38) = 0.654$, $\eta_p^2=0.017$ $p>0.05$. Results from a correlation (Pearson's *r*) analysis with the four categories and cadence revealed that the groups (cadence vs. no-cadence) significantly differ in cycle four in the tone category ($r=.405$ $p<0.05$), and in cycle two in the syllable category ($r=-.346$ $p<0.05$). To analyse this further, we conducted a post hoc analysis with an independent t-test. With cadence/no cadence as the independent variable, and mean acceptance rate for category T, cycle 4 as the dependent variable, the results showed significant findings ($p <0.05$). The results from the independent t-test with mean acceptance rate for category S, cycle 2 as the dependent variable, revealed significant findings ($p<0.05$).

In summary, the ANOVA revealed that there was not a general tendency that the participants who reported a cadence pattern performed better than those who did not. However, the t-test show that the participants who answered that they had detected a cadence pattern, had significantly higher acceptance rate of the tone category in cycles four, and significantly low acceptance rate of the syllable category in cycles two. This means that the participants who thought they had detected a cadence pattern, accepted the test items with different syllable structure, but with the same tonal pattern as in the familiarization phase, in the last test phase, more than those who did not report a cadence pattern. Those who did not report a cadence pattern accepted the test items with different tonal pattern, but with the same syllable pattern as in the familiarization phase, in the second test phase, more than those who did report a cadence pattern.

Discussion

Variability

One explanation for no significant differences in the results between the groups (high variability and low variability) could be that the groups do not differ enough. As mentioned, the participants listened to 128 strings in each familiarization phase. The high variability group listened to each string two times, whereas the low variability group listened to each string four times. Is four times per string enough to make a difference, when the participants listened to 128 strings in total in each phase? Studies on statistical learning with focus on variability may have a high variability group and a high repetition group (such as Eidsvåg et al., 2015). It can be argued that our low variability group do not have high enough repetition, and that the low variability stimuli is not low enough, only lower than the high variability stimuli.

Higher variability in input may promote learning. The results from the artificial language experiments by Gómez (2002) show that the participants who were exposed to sets with higher variability, performed better than those exposed to the sets with lower variability. The results of the study by Eidsvåg et al. (2015) support these findings. They do not only show that variability can facilitate language learning, but that this also applies to natural languages.

The findings from our study concerning variability do not support the findings from previous studies (Eidsvåg et al., 2015; Gomez, 2002; Onnis, Monaghan, Christiansen, & Chater, 2004). As discussed, the stimulus material for our high- and low variability groups may not differ enough.

Another reason for no significant results regarding variability, could be explained by our homogenic participant groups. Torkildsen, Dailey, Aguilar, Gomez and Plante (2013) investigated the role of variability in statistical learning. They compared learners with normal language and learners with language-based learning disabilities. The results show that learners with normal language managed to generalize the underlying grammar in both the high variability and the low variability condition. Of the learners with language-based learning disabilities, only those in the high variability condition managed to show generalization of the grammar (Torkildsen et al., 2013). Since none of our participants have language learning disabilities, and the majority is studying at advanced level, the participant groups might be too homogenous to generalize our findings to a greater population.

Cycles

The only category with significant findings in regard to cycles was the T&S (Tone and Syllable) category compared to the N (None) category. The participants had higher acceptance rate for the T&S category in each cycle. One explanation for this could be that participants got more exposure for this stimulus for each cycle. Thus, they might accept the sentences from the familiarization phase more (category T&S).

Preference for Tonal Cues

The category with the highest acceptance rate was the T&S category (Tone and Syllable). This might be due to salience with both a tonal cue and syllable cues. The T (Tone) category also had higher acceptance rates than the S (Syllable) category. One reason for this could be that participants got exposed to only one tonal pattern (RFH) as opposed to four syllable patterns ([lǎan mây lóm], [mǎa khwâaŋ kíaw], [mǔu phôn phít], and [sɯ̌a lɔ̂ɔ mót]). Thus, participants only had to learn one tonal pattern to perform well, in contrast to learning four syllable patterns. Concerning differences in categories, our findings are consistent with the findings by Pelucchi et al. (2009). In other words, participants are able to discriminate familiarized words from novel words.

Executive Functions' Effect on Language Learning

We wanted to investigate whether executive functions have an impact on the ability to acquire language. To measure executive functions, we used four recognized cognitive tests, which measured attention, inhibition, working memory and verbal fluency. If there had been significant correlation between, for instance, one of the cognitive tests and the participants that had higher acceptance rates for target items, it would give reason to believe that this executive function is important for using statistical language learning mechanisms.

The fact that there were no systematic patterns of correlation might suggest that executive functions have limited impact on statistical learning. If language learning is dependent on cognitive skills, we would expect to find a pattern of significant correlations. Since we did not find any systematic correlation, we cannot determine that cognitive skills affect language learning.

Reber's (1993) hypothesis that implicit learning is independent of cognitive abilities was supported by the results of the studies by Gebauer and Mackintosh (2007) and Kaufman et al. (2010). The findings show that individual differences in cognitive abilities are more prominent in explicit learning. Implicit learning is a process that happens automatically and is not dependent on cognitive skills (Cleeremans, Destrebecqz, & Boyer, 1998; Kaufman et al., 2010), at least in this learning situation with this participant group. Because our study only showed spurious correlations between two executive functions and the acceptance rates in the language learning experiment, it could indicate that learning happened implicitly in our experiment.

Another explanation for our results may be that we used such a homogenous group of participants. The majority of the participants were students at the University of Bergen, so they were highly educated. One might imagine that you need to have certain cognitive abilities to study at an advanced level. We also had an inclusion criterion that the participants should not have any known language impairment. Thus, the results not correlating may be as expected. This is in coherence with studies investigating correlations between education levels and executive functions, because people that are highly educated seem to score better at executive functioning tasks, like Stroop colour-word test, digit span and verbal fluency-tests, than their less educated peers (Tombaugh, Kozak, & Rees, 1999; Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006; Zimmermann, Cardoso, Trentini, Grassi-Oliveira, & Fonseca, 2015).

The lack of correlation between the cognitive tests and the language learning experiment does seem to agree with previous findings from other studies (Gebauer & Mackintosh, 2007; Kaufman et al., 2010; Reber, 1989). However, these studies did not investigate executive function in conjunction with statistical learning tasks. Therefore, it is still an interesting topic to research further, to understand if statistical learning is a learning process that happens implicitly, independent of executive functions.

Measures of Awareness

Results from the self-report questionnaire revealed that there was not a general tendency that the participants who reported a cadence pattern performed better than those who did not. There were no significant results from the correlation analysis.

As mentioned, the verbal reports in Rebuschat et al. (2005), indicated that participants actively search for rules during the test phase, and reported awareness. Since almost every participant reported a pattern in our study, it can indicate that they too actively searched for rules. With 15 reporting a cadence pattern, the results may indicate that they had acquired both explicit and implicit knowledge.

Retrospective verbal reports/self-report questionnaires have constraints. Since these measures solely rely on participants verbalization, they can fail to report conscious knowledge, even if they were partially aware. Some participants may not realize that their thoughts are relevant, and thus lack confidence in reporting it (Rebuschat et al., 2015). Some participants did not answer the questions surrounding which pattern they had detected. If they have an option of not responding, awareness does not get detected, even if they were aware. We do not know if they were aware or not, by them not responding. It can indicate unconscious knowledge, but it can also indicate that the participants lack confidence in responding. Lack of answer does not always correspond to a lack of awareness (Rebuschat et al., 2015).

Another constraint regarding the answers of patterns detected, could be that the participants do not have enough linguistic knowledge to explicitly express their thoughts. Unless you have studied linguistics, it may be difficult to know that there is a difference between linguistic tone and intonation, and what that difference is. Since most of the answers regarding tone is cadence, we cannot know for sure if they really meant linguistic tone or intonation, or if there are individual differences of what meaning they denoted to “cadence”. In other words, we cannot know for sure if they meant the tone of the words or the tone of the whole sentences.

Since these offline measures gathers data of awareness *after* the experiment, the participants’ awareness may have decayed in memory by the time they answer the questionnaire. Thus, these measures of awareness may not be sensitive enough to gather all relevant data. However, these measures gather data about knowledge of the *product* of learning (Rebuschat et al., 2015).

Natural Language as Stimuli

In the present study, we conducted a language learning experiment, with a natural language as stimuli. These kinds of stimuli are more complex than small artificial languages typically used in statistical learning studies. The stimulus material used in this study was more varied than stimuli from an artificial language. This is due to the large syllable set, a varied usage of linguistic tones, and unique context sentences with very little word repetition. We limit factors as pre-linguistic experience by not including participants with previous experience with Thai. Using a natural language increased the ecological validity. However, the adjacent dependencies in the stimulus material were constructed solely for this experiment, and do not exist as a grammatical rule in Thai. Thus, we have elements from both artificial and natural language stimuli.

When using artificial language as stimuli one can easier control potential factors affecting the results. As artificial languages are created solely for this purpose, one can ensure that the participants have never heard the language before. As such, the experimental control minimizes the confounding effect of earlier linguistic and pre-linguistic experience.

A main difference between artificial and natural languages, is that natural languages are more complex. The complexity of natural languages can be found in many levels: the phonetic level, morphological level as well as the syntactic level, to name a few (Arciuli & Torkildsen, 2012).

Studies using natural languages as stimuli, can further assess if findings from studies on artificial languages applies to natural languages. Findings from statistical learning studies with natural languages as stimulus show that infants can discriminate familiarized words from new words in Italian (Pelucchi et al., 2009), that participants learned nonadjacent dependencies in a foreign language (Friederici et al., 2011), and that adults showed evidence for learning gender marking in Russian (Eidsvåg et al., 2015).

Limitations of the Present Study

As discussed, our variability groups might be too similar. Gomez (2002) exposed infants and adults to different set sizes. The results showed that they were able to discriminate familiarized words from novel words better in higher set sizes. The infants did not manage to discriminate in set size 3 and 12, but managed this in set size 24. Eidsvåg et. al. (2015) used set sizes 16 and 32, and the results shows that the participants exposed to the set size 32 showed evidence of learning before the participants exposed to the set size 16. The set sizes in our study is, as mentioned, 32 and 64. Since our set sizes are this large, the low variability might not be low enough, since it is a much larger set size than 12. Our low variability group can be compared to the high variability group of Eidsvåg et. al. (2015). We can therefore not conclude whether variability is a factor for learning. We suggest using a lower set size than half of our high variability set size to discover differences in input variability.

The results show that participants had higher acceptance rate for tonal cues. As mentioned, this could be due to them being exposed for only one tonal pattern, as opposed to four syllable patterns. Because this was a pilot study, we did not know whether the participants would show a learning effect of the tonal cues, and we were interested in investigating the possibility of this. If the stimulus were to be improved for future experiments, we would suggest adapting it to have the same amount of patterns, of syllable and tones, for the learners to generalize.

The questionnaire’s structure may have had an impact on the participants answers. Since the questions surrounding awareness came after the questions about participants’ language background, their thoughts about awareness may have decayed in memory. To ensure that the participants’ memory had not been compromised, the questions about dialect and spoken languages should have been answered either before the language learning experiment begun or at the end of the questionnaire.

Regarding the four last questions in the questionnaire, where participants listened to audio clips, participants may have chosen the strings with RFH-pattern due to the prosody, rather than the tonal pattern. This is due to two of the clips that were parts of two context sentences that have a slightly different prosody than the two RFH-sentences. Since two of the clips was cut from context sentences, the words were less stressed, and was presented with a higher speed. One of these sentences also had four syllables, in contrast to the other sentences, which had three syllables. To control for this, we could have used two extra sentences restricted to the questionnaire, with similar prosody to the other test items.

Conclusion

In summary, our results support previous studies on artificial languages, and indicate that adults are able to discriminate familiarized words from novel words. This is also the case when the stimulus input involves tonal cues from a natural language. However, our participants did not manage to generalize the tonal pattern to novel words with different syllable structures. Our findings reveal that even though there was an increase in acceptance rate for familiarized words (T&S), we found no change in acceptance rate for the unheard stimuli (None).

We found no evidence to support the previous findings regarding high variability leading to rapid learning. Regarding awareness, participants reporting a cadence pattern performed better in the Tone category in the last cycles, than those who did not report such a pattern. We did not find any general tendency for participants reporting a cadence pattern performing better. We found no correlation in terms of the cognitive tests that were conducted and the performance in the language learning experiment. This finding suggests that high scores on the cognitive tests did not result in a better performance in the language learning experiment. In the present study, executive functions could therefore not explain variance in language learning.

References

Arciuli, J., & Torkildsen, J. (2012). Advancing Our Understanding of the Link between Statistical Learning and Language Acquisition: The Need for Longitudinal Data. *Frontiers in Psychology, 3*, 324. doi:10.3389/fpsyg.2012.00324

Asbjørnsen, A. E., & Hugdahl, K. (1995). Attentional effects in dichotic listening. *Brain and Language, 49*(3), 189-201.

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics, 27*(2), 185-209.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in cognitive sciences, 2*(10), 406-416.

Eidsvåg, S. S., Austad, M., Plante, E., & Asbjørnsen, A. E. (2015). Input variability facilitates unguided subcategory learning in adults. *Journal of Speech, Language, and Hearing Research, 58*(3), 826-839.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical Learning in Children With Specific Language Impairment. *Journal of Speech Language and Hearing Research, 52*(2), 321-335. doi:10.1044/1092-4388(2009/07-0189)

Friederici, A. D., Mueller, J. L., & Oberecker, R. (2011). Precursors to natural grammar learning: preliminary evidence from 4-month-old infants. *PLoS One, 6*(3), e17920. doi:10.1371/journal.pone.0017920

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 34.

Gomez, R. (2002). Variability and detection of invariant structure. *Psychol Sci, 13*(5), 431-436. doi:10.1111/1467-9280.00476

Gomez, R., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-135.

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition, 32*(3), 465-491.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science, 26*(4), 433-443.

Hugdahl, K. (1995). Dichotic listening: Probing temporal lobe functional integrity. *Brain asymmetry, 1*, 123-156.

Hugdahl, K. (2004). Dichotic Listening in the Study of Auditory Laterality. *The asymmetrical brain*, 441.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017-2036.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116*(3), 321-340.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition, 22*(4), 557-584.

Ludden, D., & Gupta, P. (2000). *Zen in the art of language acquisition: Statistical learning and the less is more hypothesis.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

Lund-Johansen, M., Hugdahl, K., & Wester, K. (1996). Cognitive function in patients with Parkinson's disease undergoing stereotaxic thalamotomy. *Journal of Neurology, Neurosurgery & Psychiatry, 60*(5), 564-571.

Marton, K. (2008). Visuo-spatial processing and executive functions in children with specific language impairment. *International Journal of Language & Communication Disorders, 43*(2), 181-200.

Montgomery, J. W. (2003). Working memory and comprehension in children with specific language impairment: What we know so far. *Journal of communication disorders, 36*(3), 221-231.

Morén, B., & Zsiga, E. (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory, 24*(1), 113-178.

Nakatani, H., & Minegishi, M. (2011). Description of Thai as an isolating language. *Social Science Information, 50*(1), 62-80.

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). *Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

Oxford University Press. (2018). Definition of cadence in English. Retrieved from https://en.oxforddictionaries.com/definition/cadence

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development, 80*(3), 674-685. doi:10.1111/j.1467-8624.2009.01290.x

Plante, E., Vance, R., Moody, A., & Gerken, L. (2013). What influences children's conceptualizations of language input? *Journal of Speech, Language, and Hearing Research, 56*(5), 1613-1624.

Raczkowski, D., Kalat, J. W., & Nebes, R. (1974). Reliability and validity of some handedness questionnaire items. *Neuropsychologia, 12*(1), 43-47.

Rasmussen, T., & Milner, B. (1977). The role of early left-brain injury in determining lateralization of cerebral speech functions. *Annals of the New York Academy of Sciences, 299*(1), 355-369.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of experimental psychology: General, 118*(3), 219.

Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. New York: Oxford University Press.

Rebuschat, P. (2015). *Implicit and explicit learning of languages* (Vol. 48): John Benjamins Publishing Company.

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015). Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition, 37*(2), 299-334.

Reitan, R. M., & Wolfson, D. (1990). A consideration of the comparability of the WAIS and WAIS-R. *Clinical Neuropsychologist, 4*(1), 80-85.

Rodriguez-Fornells, A., De Diego Balaguer, R., & Münte, T. F. (2006). Executive control in bilingual language processing. *Language Learning, 56*(s1), 133-190.

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

Saffran, J., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52. doi:Doi 10.1016/S0010-0277(98)00075-4

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide.* Pittsburgh, PA: Psychology Software Tools.

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

SPSS IBM. (2011). IBM SPSS statistics for Windows, version 20.0. *New York: IBM Corp*.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly journal of experimental psychology, 59*(4), 745-759.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology, 39*(4), 706.

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of clinical Neuropsychology, 14*(2), 167-177.

Torkildsen, J., Dailey, N. S., Aguilar, J. M., Gómez, R., & Plante, E. (2013). Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability. *Journal of Speech, Language, and Hearing Research, 56*(2), 618-629.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*(2), B25-B34.

Van der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J., & Jolles, J. (2006). The Stroop color-word test: influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment, 13*(1), 62-79.

Vanvik, A. (1973). A phonetic-phonemic analysis of Standard Eastern Norwegian. Part II. *Norwegian Journal of Linguistics, 27*, 101-139.

Wang, T. L., & Saffran, J. R. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00953

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning, 54*(4), 681-712.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*(2), 269-304.

Winke, P. (2005). *Individual differences in adult Chinese second language acquisition: The relationships among aptitude, memory and strategies for learning.* Georgetown University,

Zimmermann, N., Cardoso, C. d. O., Trentini, C. M., Grassi-Oliveira, R., & Fonseca, R. P. (2015). Brazilian preliminary norms and investigation of age and education effects on the Modified Wisconsin Card Sorting Test, Stroop Color and Word test and Digit Span test in adults. *Dementia & Neuropsychologia, 9*(2), 120-127.

Table 1

Target strings with RFH tone pattern.

Thai	Transcribed	Translation	Meaning
หลาน ไม่ ล้ม (RFH)	lǎan mây lóm	grandchild Negation fall	"Grandchild did not fall."
หมา ขว้าง เกี๊ยว (RFH)	mǎa khwâaŋ kíaw	dog throw wonton	"The dog threw a wonton."
หมู พ่น พิษ (RFH)	mǔu phôn phít	pig spit venom	"The pig spat venom."
เสือ ล่อ มด (RFH)	sɯ̌a lɔ̂ɔ mót	tiger trick ant	"The tiger tricked the ant."

Table 2.1

Test items Category 1: The targets as used in the familiarization strings.

Thai	Transcribed	Translation	Meaning
หลาน ไม่ ล้ม (RFH)	lǎan mây lóm	grandchild negation fall	"Grandchild did not fall."
หมา ขว้าง เกี๊ยว (RFH)	mǎa khwâaŋ kíaw	dog throw wonton	"The dog threw a wonton."
หมู พ่น พิษ (RFH)	mǔu phôn phít	pig spit venom	"The pig spat venom."
เสือ ล่อ มด (RFH)	sɯ̌a lɔ̂ɔ mót	tiger trick ant	"The tiger tricked the ant."

Table 2.2

Test items Category 2: Stimuli with the same tonal structure as the targets, but with different syllables.

Thai	Transcribed	Translation	Meaning
ขวัญ อุ้ม น้อง (RFH)	khwǎn ʔûm nɔ́ɔŋ	Khwan hold younger sibling	"Khwan (female name) held her younger sibling."
หมี เลื่อย ไม้ (RFH)	mǐI lɨ̂aj máaj	bear saw wood	"The bear sawed wood."
ผม นั่ง ยิ้ม (RFH)	phǒm nâŋ yím	I (male) sit smile	"I (male) sat smiling."
หมอ ว่าย น้ำ (RFH)	mɔ̌ɔ wâaj náam	doctor swim water	"The doctor swam."

Table 2.3

Test items Category 3: Stimuli with the same syllable structure as the targets, but different tonal structures.

Thai	Transcribed	Translation	Meaning
ลาน ใหม่ ล่ม (MLF)	laan mày lôm	court new collapse	"The new court collapsed/was not successful."
ม้า ขวาง เกี้ยว (HRF)	máa khwǎaŋ kîaw	horse block palanquin	"The horse blocked a palanquin."
หมู่ พัน ผิด (LHL)	mùu phón phìt	sergeant be acquitted	"The sergeant was acquitted."
เสื้อ หล่อ หมด (FLL)	sʉ̂a lɔ̀ɔ mòt	shirt handsome all gone	"Handsome shirts are all gone."

Table 2.4

Test items Category 4: Stimuli where both tonal and syllable structure is different from the targets.

Thai	Transcribed	Translation	Meaning
ลุง ลืม ของ (MMR)	luŋ ~~lɯɯm~~ khɔ̌ɔŋ	uncle forget thing	"Uncle forgot something."
ไกล จาก รถ (MLH	klay càak rót	far from car	"Far from the car."
ฟ้า สี คราม (HRM)	fáa sǐi khraam	sky colour indigo	"The sky is indigo."
โต๊ะ ตัว ใหญ่ (HML)	tó tua yày	table classifier big	"The big table."

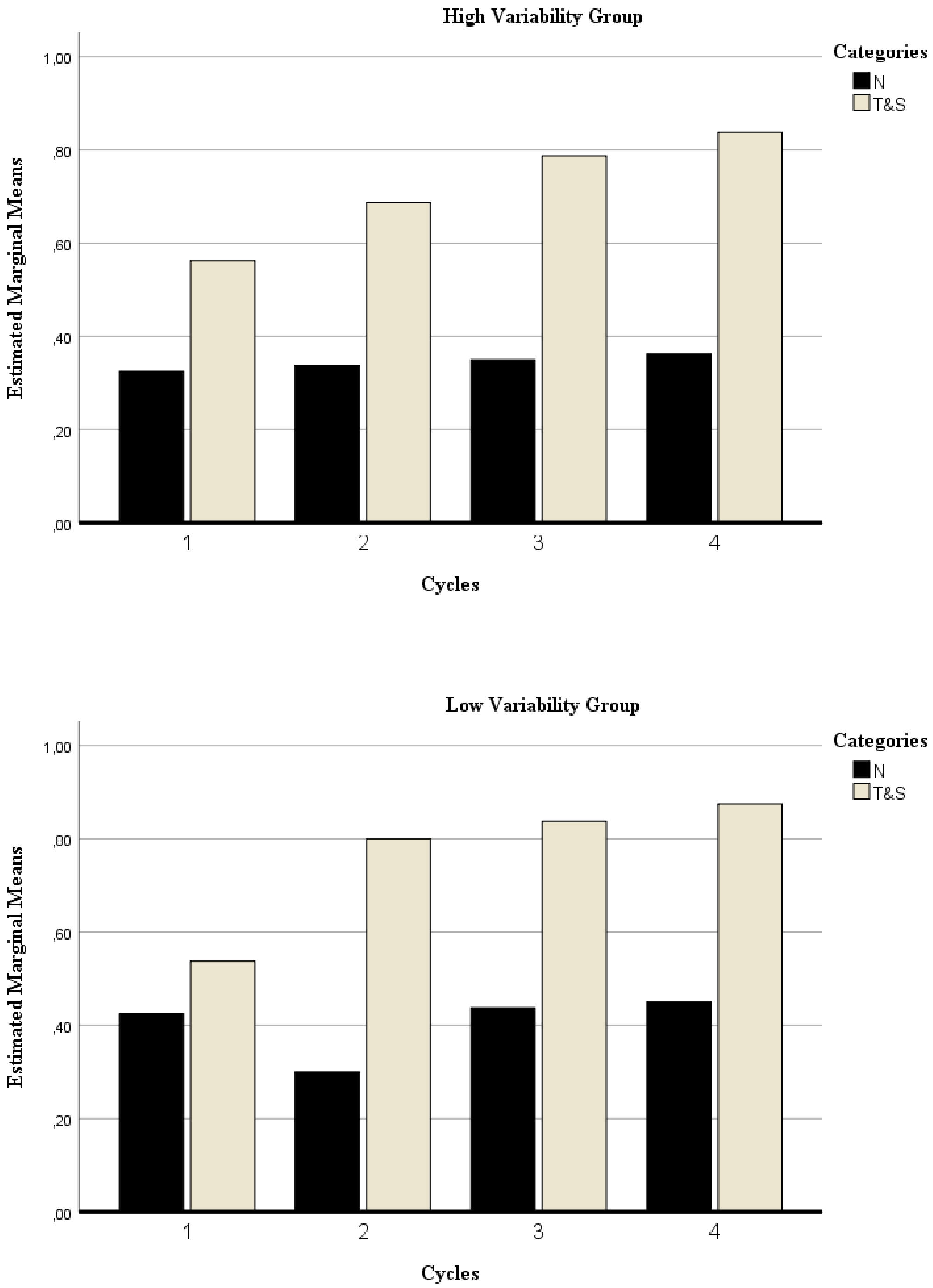

Figure 1. ANOVA analysis of N and T&S. This figure illustrates the mean acceptance rates in four cycles of the categories None, and Tone and Syllable in the high variability group and the low variability group, respectively.

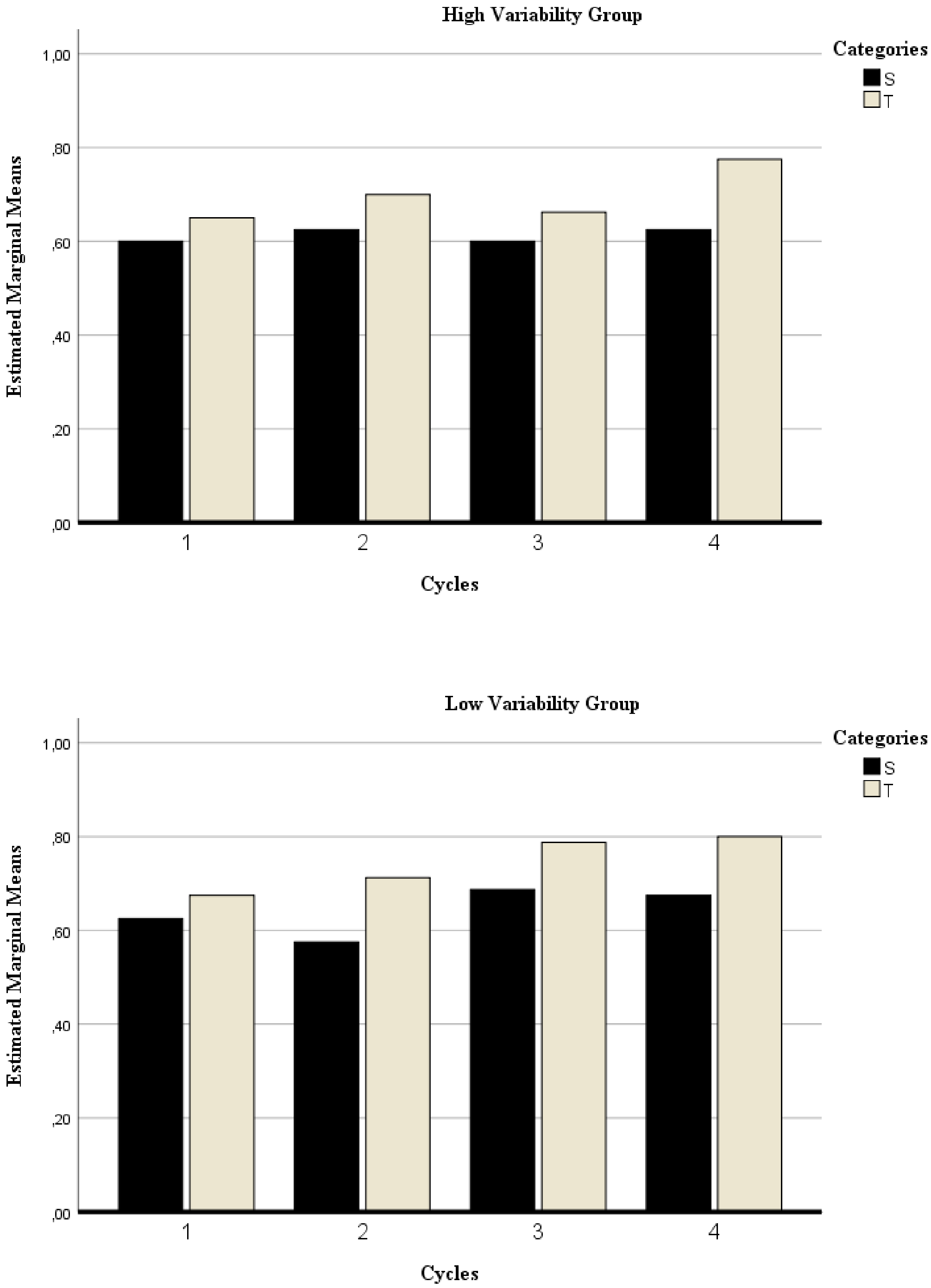

Figure 2. ANOVA analysis of S and T. This figure illustrates the mean acceptance rates in four cycles of the categories Syllable and Tone in the high variability group and the low variability group, respectively.

Appendix

Table 3

Familiarization strings (16 context sentences for each of the RFH core strings).

Thai	Transcribed	Translation	Meaning
หลาน ไม่ ล้ม แต่ ก็ ร้อง ไห้	lǎan mây lóm tàe kɔ̂ɔ rɔ́ɔŋ hâay	grandchild negation fall but connector cry	"Grandchild did not fall, but (he/she) still cried."
หลาน ไม่ ล้ม แล้ว	lǎan mây lóm láew	grandchild Negation fall Aspect	"Grandchild will not fall anymore."
หลาน ไม่ ล้ม จริง จริง	lǎan mây lóm ciŋ ciŋ	grandchild Negation fall really	"Grandchild really did not fall."
หลาน ไม่ ล้ม เลย ตอน วิ่ง เร็ว เร็ว	lǎan mây lóm ləəy tɔɔn wîŋ rew rew	grandchild Negation fall Particle when run fast	"Grandchild did not fall at all when running fast."
หลาน ไม่ ล้ม หรอก ค่ะ	lǎan mây lóm rɔ̀k khà	grandchild Negation fall Particle Polite particle	"Grandchild will not fall."
เขา เห็น ว่า หลาน ไม่ ล้ม แน่	khǎw hěn wâa lǎan mây lóm nâe	he/she see that grandchild Negation fall certain	"He/she saw that Grandchild certainly would not fall."
ถ้า หลาน ไม่ ล้ม น้า ก็ จะ สบาย ใจ	thâa lǎan mây lóm náa kɔ̂ɔ cà sabaay cay	if grandchild Negation fall aunt/uncle Connector Modal happy	"If Grandchild does not fall, Aunt/Uncle will be happy."
พี่ บอก ว่า หลาน ไม่ ล้ม ง่าย ง่าย หรอก	Phîi bɔ̀ɔk wâa lǎan mây lóm ŋâay ŋâay rɔ̀k	Older sibling tell that grandchild Negation fall easy Particle	"Older sibling said that Grandchild would not fall easily."
พื้น ลื่น แบบ นี้ หลาน ไม่ ล้ม เหรอ	phɨ́ɨn lɨ̂ɨn bàep níi lǎan mây lóm rə̌ə	floor slippery like this grandchild Negation fall Question particle	"A slippery floor like this, won't Grandchild fall?"

เมื่อ วาน หลาน ไม่ ล้ม นะ	mʉ̂a waan lǎan mây lóm ná	yesterday grandchild Negation fall Particle	"Yesterday, Grandchild did not fall."
ดี ใจ ที่ หลาน ไม่ ล้ม ยาย ล้ม แต่ หลาน ไม่ ล้ม	dii cay thîi lǎan mây lóm yaay lóm tàe lǎan mây lóm	glad that grandchild Negation fall grandmother fall but grandchild Negation fall	"(I) am glad that Grandchild did not fall." "Grandmother fell, but Grandchild did not fall."
บอก แล้ว ไง ว่า หลาน ไม่ ล้ม	bɔ̀ɔk láew ŋay wâa lǎan mây lóm	tell Aspect Particle that grandchild Negation fall	"(I) already told (you) that Grandchild did not fall."
ไม่ อยาก จะ เชื่อ ว่า หลาน ไม่ ล้ม	mây yàak cà chʉ̂a wâa lǎan mây lóm	Negation want Modal believe that grandchild Negation fall	"(I) could not believe that Grandchild did not fall."
วิทย์ แปลก ใจ ที่ หลาน ไม่ ล้ม	Wít plàek cay thîi lǎan mây lóm	Wit surprised that grandchild Negation fall	"Wit (male name) was surprised that Grandchild did not fall."
บัว โก หก แม่ ว่า หลาน ไม่ ล้ม	Bua koohòk mâe wâa lǎan mây lóm	Bua lie mother that grandchild Negation fall	"Bua (female name) lied to Mother that Grandchild did not fall."
หมา ขว้าง เกี๊ยว ไป ไกล	mǎa khwâaŋ kíaw pay klay	dog throw wonton go far	"The dog threw a wonton far away."
หมา ขว้าง เกี๊ยว ครับ	mǎa khwâaŋ kíaw khráp	dog throw wonton Polite particle	"The dog threw a wonton."
หมา ขว้าง เกี๊ยว ให้ แมว	mǎa khwâaŋ kíaw hây maew	dog throw wonton give cat	"The dog threw a wonton to the cat."

หมา ขว้าง เกี๊ยว รึ เปล่า	mǎa khwâaŋ kíaw rɨ́ plàaw	dog throw wonton Question particle	"Did the dog throw a wonton?"
หมา ขว้าง เกี๊ยว เพราะ ไม่ อยาก กิน	mǎa khwâaŋ kíaw phrɔ́ mây yàak kin	dog throw wonton because Negation want eat	"The dog threw a wonton because (he) did not want to eat (it)."
หมา ขว้าง เกี๊ยว เกือบ ถึง กำ แพง	mǎa khwâaŋ kíaw kɨ̀ap thɨ̌ŋ kam phaeŋ	dog throw wonton almost reach wall	"The dog threw a wonton almost to the wall."
ไม่ รู้ ว่า หมา ขว้าง เกี๊ยว ได้ ด้วย	mây rúu wâa mǎa khwâaŋ kíaw dâay dûay	Negation know that dog throw wonton able also	"(I) did not know that the dog could also throw a wonton."
คิด ว่า หมา ขว้าง เกี๊ยว ได้ ไกล กว่า ช้าง	khít wâa mǎa khwâaŋ kíaw dâay klay kwàa cháaŋ	think that dog throw wonton able far than elephant	"(I) think that the dog can throw a wonton farther than the elephant."
เขา ฝัน ว่า หมา ขว้าง เกี๊ยว อยู่ ใน ครัว	khǎw fǎn wâa mǎa khwâaŋ kíaw yùu nay khrua	he/she dream that dog throw wonton stay in kitchen	"He/she dreamt that a dog was throwing a wonton in the kitchen."
พ่อ ตก ใจ ที่ หมา ขว้าง เกี๊ยว ออก ไป ข้าง นอก	phɔ̂ɔ tòk cay thîi mǎa khwâaŋ kíaw ʔɔ̀ɔk pay khâaŋ nɔ̂ɔk	father shocked that dog throw wonton exit go outside	"Father was shocked that the dog threw a wonton out."
เธอ เชื่อ ว่า หมา ขว้าง เกี๊ยว ได้ จริง จริง เหรอ	thəə chɨ̂a wâa mǎa khwâaŋ kíaw dâay ciŋ ciŋ rə̌ə	you believe that dog throw wonton able really Question particle	"Do you really believe that a dog can throw a wonton?"

ไม่ เคย เห็น หมา ขว้าง เกี๊ยว	mây khəəy hěn mǎa khwâaŋ kíaw	Negation ever see dog throw wonton	"(I) have never seen a dog throwing a wonton."
กระ ต่าย บอก ว่า หมา ขว้าง เกี๊ยว	krà tàay bɔ̀ɔk wâa mǎa khwâaŋ kíaw	rabbit tell that dog throw wonton	"The rabbit said that the dog threw a wonton."
เด็ก เด็ก อยาก ดู หมา ขว้าง เกี๊ยว	dèk dèk yàak duu mǎa khwâaŋ kíaw	child want watch dog throw wonton	"The children wanted to watch the dog throw a wonton."
ฉัน นั่ง มอง หมา ขว้าง เกี๊ยว	chǎn nàŋ mɔɔŋ mǎa khwâaŋ kíaw	I sit watch dog throw wonton	"I sat and watched a dog throw a wonton."
น้อย ฝึก ให้ หมา ขว้าง เกี๊ยว	nɔ́ɔy fɨ̀k hây mǎa khwâaŋ kíaw	Noy train give dog throw wonton	"Noy (female name) trained the dog to throw a wonton."
หมู พ่น พิษ ร้าย แรง มาก	mǔu phôn phít ráay raeŋ mâak	pig spit venom malignant very	"The pig spat very malignant venom."
หมู พ่น พิษ เยอะ	mǔu phôn phít yə́	pig spit venom a lot	"The pig spat a lot of venom."
หมู พ่น พิษ ใส่ กระ รอก จริง เหรอ	mǔu phôn phít sày kra rɔ̂ɔk ciŋ rə̌	pig spit venom put squirrel true Question particle	"Is it true that the pig spat venom at the squirrel?"
หมู พ่น พิษ โดน ใคร	mǔu phôn phít doon khray	pig spit venom touch who	"Who did the pig spit venom at?"
หมู พ่น พิษ ไม่ ได้ หรอก	mǔu phôn phít mây dâay rɔ̀k	pig spit venom Negation able Particle	"The pig cannot spit venom."

เจอ หมู พ่น พิษ ที่ โรง เรียน	cəə mǔu phôn phít thîi rooŋ rian	see pig spit venom at school	"(I) saw a pig spitting venom at school."
ไม่ รู้ เลย ว่า หมู พ่น พิษ สาม ครั้ง แล้ว	mây rúu ləəy wâa mǔu phôn phít sǎam khráŋ láew	Negation know Particle that pig spit venom three time Aspect	"(I) did not know that the pig has spat venom three times already."
คิด ว่า หมู พ่น พิษ ได้ นะ	khít wâa mǔu phôn phít dâay ná	think that pig spit venom able Particle	"(I) think that the pig can spit venom."
อย่า ไป เดิน ตรง ที่ หมู พ่น พิษ ไว้	yàa pay dəən troŋ thîi mǔu phôn phít wáy	do not go walk at place pig spit venom keep	"Do not walk near the spot where the pig spat venom."
ถ้า โดน หมู พ่น พิษ ใส่ ต้อง รีบ ล้าง ออก	thâa doon mǔu phôn phít sày tɔ̂ŋ rîip láŋ ʔɔ̀ɔk	if Passive pig spit venom put must hurry wash out	"If the pig spits venom at you, you must quickly wash it out."
สง สัย หมู พ่น พิษ ออก มา แล้ว	sǒŋ sǎj mǔu phôn phít ʔɔ̀ɔk maa láew	wonder pig spit venom out come Aspect	"(I) suspect that the pig has already spat venom."
ทุก คน กลัว หมู พ่น พิษ	thúk khon klua mǔu phôn phít	everyone be afraid pig spit venom	"Everyone was afraid that the pig would spit venom."
เมื่อ คืน หมู พ่น พิษ	mʉ̂a khʉʉn mǔu phôn phít	last night pig spit venom	"Last night, the pig spat venom."
เขา สะ ดุ้ง ตอน เห็น หมู พ่น พิษ	khǎw sa dûŋ tɔɔn hěn mǔu phôn phít	he/she be startled when see pig spit venom	"He/she was startled when seeing the pig spitting venom."
ย่า ไม่ เดือด ร้อน ที่ หมู พ่น พิษ	yâa mây dʉ̀at rɔ́ɔn thîi mǔu phôn phít	Grandmother Negation be troubled that pig spit venom	"Grandmother was not troubled that the pig spat venom."

เกลียด หมู พ่น พิษ	klìat mǔu phôn phít	hate pig spit venom	"(I) hate the pig that spits venom."
เสือ ล่อ มด ไม่ เป็น	sʉ̌a lɔ̂ɔ mót mây pen	tiger trick ant Negation be	"The tiger did not know how to trick the ant."
เสือ ล่อ มด เก่ง	sʉ̌a lɔ̂ɔ mót kèŋ	tiger trick ant well	"The tiger was good at tricking the ant."
เสือ ล่อ มด เข้า ไป ใน ป่า	sʉ̌a lɔ̂ɔ mót khâw pay nay pàa	tiger trick ant enter go in wood	"The tiger tricked the ant into the wood."
เสือ ล่อ มด ได้ ด้วย เหรอ	sʉ̌a lɔ̂ɔ mót dâay dûay rə̌ə	tiger trick ant able also Question particle	"Can the tiger trick the ant?"
เสือ ล่อ มด เป็น ประ จำ	sʉ̌a lɔ̂ɔ mót pen pràcam	tiger trick ant regularly	"The tiger regularly tricked the ant."
เสือ ล่อ มด เพื่อ ความ สนุก	sʉ̌a lɔ̂ɔ mót phʉ̂a khwaam sanùk	tiger trick ant for fun	"The tiger tricked the ant for fun."
เชื่อ เถอะ เสือ ล่อ มด เอง	chʉ̂a thə̀ sʉ̌a lɔ̂ɔ mót ʔeeŋ	believe Particle tiger trick ant self	"Believe (me), the tiger tricked the ant himself."
อยาก เห็น เสือ ล่อ มด ใกล้ ใกล้	yàak hěn sʉ̌a lɔ̂ɔ mót klây klây	want see tiger trick ant close	"(I) want to see a tiger tricking the ant closely."
รู้ ไหม ว่า เสือ ล่อ มด ได้	rúu mǎy wâa sʉ̌a lɔ̂ɔ mót dâay	know Question that tiger trick ant get	"Do (you) know that the tiger can trick the ant?"
เขา สอน ให้ เสือ ล่อ มด อยู่	khǎw sɔ̌ɔn hây sʉ̌a lɔ̂ɔ mót yùu	He/she teach give tiger trick ant Progressive	"He/she is teaching the tiger to trick the ant."

สิง โต ชวน ให้ เสือ ล่อ มด ด้วย กัน	sĭŋ too chuan hây sŭa lɔ̂ɔ mót dûay kan	lion persuade give tiger trick ant together	"The lion persuaded the tiger to trick the ant together."
นัก วิ จัย กำ ลัง สัง เกต เสือ ล่อ มด	nák wí cay kam laŋ sáŋ kèt sŭa lɔ̂ɔ mót	researcher Progressive observe tiger trick ant	"The researcher is observing a tiger tricking an ant."
มา ดู เสือ ล่อ มด	maa duu sŭa lɔ̂ɔ mót	come watch tiger trick ant	"(I) came to watch the tiger tricking the ant."
ปู่ กัง วล ที่ เสือ ล่อ มด	pùu kaŋ won thîi sŭa lɔ̂ɔ mót	Grandfather worried that tiger trick ant	"Grandfather was worried about the tiger tricking the ant."
ต้น บอก เมย์ ว่า เสือ ล่อ มด	tôn bɔ̀ɔk mee wâa sŭa lɔ̂ɔ mót	Ton tell May that tiger trick ant	"Ton (male name) told May (female name) that the tiger tricked the ant."
ใคร ก็ ไม่ เชื่อ ว่า เสือ ล่อ มด	khray kɔ̂ɔ mây chûa wâa sŭa lɔ̂ɔ mót	who Connector Negation believe that tiger trick ant	"Nobody would believe that the tiger tricked the ant."