April 2021

Directing Attention in Second Language Phonological Contrast Learning

Laura Conover
University of South Florida

Directing Attention in Second Language Phonological Contrast Learning

by

Laura Conover

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Communication Sciences and Disorders
College of Behavioral and Community Sciences
University of South Florida

Major Professor: Ruth Huntley Bahr, Ph.D.
Stefan A. Frisch, Ph.D.
Amanda Huensch, Ph.D.
Catherine Rogers, Ph.D.

Date of Approval:
April 13, 2021

Keywords: Statistical Learning, Phoneme Category Development, Adult Learners, Executive Function

Acknowledgments

There are a lot of you – Dr. Bahr overall, my committee who put up with a dissertation not quite in the right field and some tight deadlines, my family who made sure I had enough coffee and cookies to stay awake, my partner who listened to me rant at the computer incomprehensibly for hours, and my cat who decided to rewrite some of my work on a whim a few times. Thank you all.

Table of Contents

List of Tables

List of Figures

Abstract

Why are some people better at learning new languages than others? There is a rich body of research examining this issue from multiple perspectives and at all levels of language. This study attempts to add to that knowledge at the most fundamental level of language by examining potential influences on the learning of novel phoneme contrasts. The purpose of this study was to explore whether individual differences in attentional capabilities would help adults learn a non-native phonological contrast, and whether providing explicit directions that would guide the learners' attention could help boost their performance.

VCV recordings of the Thai /p/ and /b/ phonemes, which are often both heard as /b/ by English speakers, were provided by four native Thai speakers. A total of 57 monolingual English-speaking adults completed the study through the Gorilla online experimental platform (www.gorilla.sc). The Attention Network Test (ANT; Fan, et al, 2002) was used to assess the three attentional networks: Alerting, Orienting, and Executive Function. To teach them the Thai /p/ and /b/ contrast, the participants underwent a statistical learning paradigm in which they had to match the sound they heard to an image on the screen. The paradigm consisted of a 40-question pretest, 480-trial learning phase, and 40-question posttest. Approximately half of the participants (n=30) were given explicit instruction on what to listen for prior to the learning phase. The other participants (n=27) were told to listen carefully but were given no specific direction.

Generalized linear models (GLM) were fit to predict the participants' posttest scores from their ANT subscores and their experimental group. A linear mixed effects model was also fit to describe the participants' performance in each block during the learning phase. The results of the GLM showed significant interactions between Executive Function and experimental group ($p=0.0398$) and significant main effects of Executive Function ($p=0.0209$). The linear mixed effects model showed a significant three-way interaction between Executive Function, block, and experimental group ($p=0.00363$) and a significant two-way interaction between Executive Function and block ($p=0.00567$). Taken together, these results imply that individuals who are better able to control their attentional focus are better able to learn novel phoneme contrasts regardless of instruction. In addition, individuals with better Executive Function abilities seem to be able to benefit more from explicit instructions as to where to direct their attention, learning the target contrast more quickly and performing better on perceptual tasks.

Introduction

There are a number of factors that impact an individual's ability to discriminate speech sounds, i.e., age, surrounding conditions, context, etc. Perhaps one of the biggest influences determining whether two phonemes will be discriminable is the listeners' native language(s) and their existing phoneme categories (Best, 1995). It has been theorized that the influence of native language phoneme categories is due to creation of the stored mental prototypes, which function as "perceptual magnets" to warp how listeners hear speech sounds (Kuhl, 1991). Early in this line of research, attempts had been made to quantify the "warping" process (Iverson & Kuhl, 1995). However, there is still much variability between individuals in their relative performance on speech discrimination tasks. Now thanks to increases in computing power and mathematical modelling, more recent work has started to align our understanding of the perceptual magnet effect and its formation of phoneme categories with a process resembling statistical learning (Feldman et al., 2009). With these methodological advances, additional variables can be accounted for, which should help to explain some of the remaining variability in performance.

This project attempts to explain differences in individuals' abilities to perceive new speech sounds by considering attention as one potential source of variability in successful phoneme category formation within a statistical learning paradigm. Although studies have suggested that statistical learning is a unique ability that is uncorrelated with other measures of cognition (Siegelman & Frost, 2015), the mere fact that cognitive processes can be independently tested does not mean that the processes have no impact on each other. For instance, it is not uncommon for models of cognition to treat attention – itself a cognitive process or ability – as a resource that fuels other cognitive processes, such as perception (Wickens, 1992). The purpose of this study was to examine if better attentional abilities would help adults learn a non-native phonological contrast in a statistical learning paradigm, and if providing explicit directions regarding how to direct attention would boost subsequent performance.

Phoneme Categories and Contrasts

Bottom-up cognitive processes take the acoustic details present in an individual's environment and lead to the creation of a perceptual system that is attuned to the phonological categories of that individual's native language (Pierrehumbert, 2003). Although this process is often associated with infants, research has suggested that monolingual speakers of a language may take a long time to finalize their phoneme categories. For instance, performance on minimal pair categorization tasks for consonants - which are normally more strongly associated with categorical perception - does not reach adult-like levels until after age 12 (Pierrehumbert, 2003). Even after this point, research involving adult second language learners implies that the phonological system is still open to some modifications later in life (Fuhrmeister et al., 2020; McMurray et al., 2018), though these changes are not as drastic as in a child's system (Flege, 1995a).

Once a phonological system has been created, the phoneme categories within that system can be considered as a collection of labels on a map of auditory-acoustic space. In this map, each label has an associated probability distribution for speech sounds the individual encounters (Pierrehumbert, 2003). Although this definition includes a key reference to probability distributions that is useful from a statistical learning standpoint, it does not provide much information on the process of categorization itself. Other researchers, however, have worked to define categorization as a process, rather than an end state. For instance, Tuller et al. (2008, p. 210), called categorization “a dynamical process that modifies perceptual space over time”. The inclusion of a time element is important to describing how humans perceive and use categories in speech perception, since the categorization of speech sounds takes years to develop, even in children learning their first language(s). This learning and development procedure likely takes place using the same bottom-up acoustic-perceptual process previously mentioned, supported by input from other speakers (Pierrehumbert, 2003).

Perceptual models. The viewpoint of categorization as a dynamic process interacts well with other models of phoneme categorization and perception because it helps provide explanations for phenomena, such as the shifting of categories when an individual is exposed to novel phonemes (Flege, 1993). It has long been understood that the mere existence of categories has the power to influence perception, such that sounds within a category are harder to discriminate than sounds across category boundaries, even if the stimuli are acoustically evenly spaced (Liberman et al., 1957). A simple awareness of phonetic categories biases a listener’s perception of a stimulus towards a category center, whereas a listener with no categories would perceive the sound with less bias and may begin to form a category with that sound at its center (Feldman et al., 2009). Even in non-speech modalities, this improved ability to distinguish two objects that lie across a category boundary, and the related decreased ability to distinguish two objects that fall within a single category, persists, such that the features that are most salient to us are the ones that are the strongest predictors of category membership (Goldstone, 1994). It is generally expected that the dissimilarity between the second language (L2) phoneme and the first language (L1) category is a predictor of how likely the two are to be distinguished from each other (Best, 1995; Flege, 1995a; Iverson & Kuhl, 1995; Myers et al., 2017). However, there are a number of different theories that predict how, and to what extent, the native language category can influence a listener's perception.

The Perceptual Assimilation Model (Best, 1995) describes how discriminable two non-native phonemes might be for an inexperienced listener based on the phoneme categories of the listener's L1. In short, discriminability is tied to the presence or absence of a mental representation for a given speech sound (Best & Tyler, 2007). Six different possible types of contrasts between the two non-native segments are described:

1) Two-Category Assimilation - each non-native segment is assimilated to a different native category;
2) Category-Goodness Difference - both non-native segments are assimilated to the same category, but they differ in how well each one fits into that native category;
3) Single-Category Assimilation - both non-native segments are assimilated to the same native category and have equal goodness of fit;
4) Both Uncategorizable - both non-native sounds are heard as speech sounds, but are not placed in any category;
5) Uncategorized vs Categorized - one non-native sound is assimilated into a native category, the other is still heard as a speech sound but does not fall into a phonetic category; and
6) Nonassimilable - neither non-native sound is heard as a speech sound at all.

In general, this model predicts that non-native sounds that fall into different categories will be more discriminable from each other than non-native sounds that fit equally well into the same category, but that is where the model technically stops. It does not explicitly provide a description of how well a non-native phoneme may be discriminated from a native phoneme in the same category. And, although it may be intuitive to conclude that non-native sounds that fall into a native category, or are more easily discriminable from each other, will ultimately be learned more easily, this model is by definition a perceptual model only, not a learning model (Best, 1995).

The Speech Learning Model (Flege, 1995a), on the other hand, does attempt to define the learning process. Individuals who have strongly established L1 phonological categories will not spontaneously create a new category for the L2 phoneme. Instead, they will instead link that phoneme to their existing L1 category, perceiving the two phonemes as one (Flege, 1995a). This process has been termed “equivalence classification”. On the other hand, individuals who do not have strongly established L1 phonological categories, such as children, will create a separate category for the L2 phoneme, even if it is similar to the L1 category, and be able to perceive the difference between the two (Flege, 1991). According to this model, without the formation of the correct mental representation, even fluent bilinguals are unlikely to be able to perceive the difference between an L2 phoneme and their L1 phoneme if both were initially assigned to the same L1 category.

The Speech Learning Model (Flege, 1995a) can be taken further, with research supporting the idea that individuals develop an interlanguage phonology as a reflection of their linguistic experience and particular phonological processing strategies (Flege, 1980; Strange & Shafer, 2008). The interlanguage system may include the formation of a new phonetic category, but it is more likely to simply shift existing categories in adult listeners (Flege, 1980). As such, even fluent bilinguals who were exposed to their L2 as adults would be expected to produce compromise values on certain dimensions between their L1 and L2, such as voice onset time (VOT) for Spanish/English bilinguals (Flege, 1991; Williams, 1979) or vowel duration for

Chinese/English bilinguals (Flege, 1993), as the influence of the L2 phonology shifts the location of their existing phoneme category along the target dimension. However, the model does predict that, should a new mental representation of the L2 phoneme be created, it could then be perceived as a unique speech sound. Other research, discussed in subsequent sections, suggests that different training paradigms can assist with the creation of these new mental representations (Lively et al., 1993; Sakai & Moorman, 2018).

A final viewpoint can be found in models of the Perceptual Magnet Effect (Iverson & Kuhl, 1995). In this model, the categories within an individual's native language perceptually attract any sounds similar to a reference sound for that category to a specific point within the category. The warping of the map of the perceptual space away from a Euclidean acoustic map can then be modeled by multidimensional scaling or signal detection techniques (Iverson & Kuhl, 1995), showing the effect on perception of the warping caused by the category magnets (Best & Tyler, 2007).

The Perceptual Magnet Effect's (Iverson & Kuhl, 1995) map of the perceptual space can also provide clues as to which L2 contrasts may be more easily learned. For instance, neither Japanese nor German phonology includes an "American" /r/ sound. However, the Japanese perceptual map includes a single category that encompasses both the American /r/ and /l/ sounds, whereas the German perceptual map does not. Therefore, one might expect it to be easier for German students to learn an American /r,l/ contrast than for Japanese students, and experiments have found exactly that trend (Iverson et al., 2003). These language-specific categories can be explicitly beneficial to a listener's perception as well. For example, Bayesian models of the Perceptual Magnet Effect suggest that the warping created by categories can help listeners perceive a signal in noisy conditions (Feldman et al., 2009).

The concept of warping the perceptual space has led to some ambiguity within models of the Perceptual Magnet Effect. Early research on phonetic category formation suggested that exemplars, rather than prototypes, are what caused the perceptual warping, due to the fact that presentation of varying contexts of a phoneme appeared to be required to learn the contrast (Lively et al., 1993). On the other hand, more recent work suggests that the shrinkage of perceptual space is due to prototypes, not exemplars, because target phonemes appear to lie at extreme values of relevant acoustic cues, rather than at typical production values (Lotto et al., 2000). While the specific class of object that causes warping may be debatable, neural research supports the idea that novel speech sounds can be learned by integration into previously existing native categories (Dobel et al., 2009).

With recent advances in computing power, many speech perception theories can now be mathematically modelled and tested more rigorously. For instance, some models take a metacognitive view of the modification of phonological categories, suggesting that a listener's awareness of phonetic categories is a product of their experience with a language. According to these models, perception of a speech sound is warped due to an awareness of the existence of phonological categories. This finding is similar to the claims of the Perceptual Magnet Effect (Feldman et al., 2009). In the mathematical representation of these metacognitive models, experience with a language allows listeners to become aware of the fact that, given a roughly Gaussian distribution of variability, they are more likely to hear speech sounds that are near the center of a phonetic distribution than those that are towards the edges, and so their perception is biased to reflect that. In order to change or shift a category boundary, an individual needs to be exposed to enough examples within suitable learning contexts to overcome that bias. The model can also explain the influence of a speaker's native language on subsequent category formation.

For instance, Polish and English have different distributions that describe the acceptable ranges in VOT for the voiced/voiceless contrast. A Polish speaker will have a poorer discrimination of range effects within the set of stimuli presented to them than an English speaker, because the Polish system allows for greater variability in VOT to make the voiced/voiceless distinction than the English system does (Keating et al., 1981). Therefore, listeners' native language is a key factor in determining how they will react to the statistical patterns of a tested set of stimuli.

Other mathematical techniques are options as well. For instance, a model of an interlanguage phonology system can be made using principles from optimality theory (Hancin-Bhatt, 2008). The listener's initial phonology is influenced by ongoing perceptual feedback, and as updating occurs, the constraints on phonology are ranked and re-ordered as the listener gains more experience with the non-native language. Pierrehumbert (2003) proposes a similar model that also includes feedback from outside sources, such as other speakers, as a part of the experience a listener might gain. In this model, both external feedback and internal monitoring of sounds are responsible for updating the rules that create learners' phonological categories. Hence, in the field of speech perception, the quantification of models using statistical techniques can help explain how information across phonetic and phonological dimensions is integrated to predict speech-sound identification, and to describe group- and individual-level patterns of differences in such processes (Silbert, 2012).

Individual differences. Not all of the imbalance in category formation ability may be due solely to the influence of the individual's native language. Even across speakers of the same language, some people appear to be better at picking up new sounds than others. Therefore, it seems reasonable to wonder what other factors may be at play. Although the most common explanation for differences in the ability to perceive non-native phonological contrasts is age of exposure to the L2 (Flege, 1995b), the current discussion focuses on an adult population. Even within this population, there are a few additional factors to consider regarding individual performance on category formation tasks.

One possible factor is each individual adult's auditory perceptual ability overall. Individual differences in perceptual abilities are reliable at the neural level, and so is the connection between perceptual abilities and a listener's ability to discriminate phonetic contrasts (Lin et al., 2017). However, individuals can vary in their perceptual ability across modalities, which may impact their discriminatory abilities (i.e., some individuals may be more auditory-biased, and others may be more visually-biased; Siegelman & Frost, 2015). For example, if a listener cannot perceive the acoustic cues within stimuli, it would be difficult or impossible for that individual to build up the appropriate probability distribution based on those cues. On the other hand, if a listener was already experienced with the statistical regularities of a set of stimuli, those expectations may help improve the listener's perception even in circumstances where perceptual acuity is challenged, such as degraded listening conditions (Conway et al., 2010).

Additionally, there is variability in how biased different individuals may be towards categorical or continuous perception of presented stimuli (Kong & Edwards, 2016). Individuals who have a bias towards perceiving the gradient features of spoken stimuli can learn new contrasts and form new phoneme boundaries more easily than those who are biased towards categorical perception, because categorical perception may impede perception of acoustic detail. Furthermore, some individuals may just not be as talented at perceiving acoustic detail in general. The difference between good and poor perceivers can be in the different types of training required to teach new categories to either group. For instance, individuals with poor perceptual abilities do better with low stimulus variability during perceptual training in order to avoid becoming overwhelmed before learning occurs, but this is not the case for individuals with good perceptual abilities (Perrachione et al., 2011).

There are also a few variables that surprisingly do not contribute to individual differences in the perception of novel phoneme contrasts. First, experience with a second language does not predict performance on contrast learning tasks in that language for adult populations (Golestani & Zatorre, 2009). For example, in a sample of adults who were not exposed to their target L2 until adulthood and who all moved to an L2-dominant country as adults, the participants' ages of exposure, length of residence in the L2-dominant country, percentage of L1 and L2 usage, and even motivation to use and improve the L2 were all uncorrelated with L2 phonological processing performance (Darcy et al., 2015). Similarly, a number of cognitive factors, such as processing speed and lexical knowledge, were not significantly correlated to L2 phonological processing in the same population (Darcy et al., 2015). Such findings can help future research narrow down the field of variables that may help explain the vast individual differences on L2 phoneme learning tasks.

Neuroimaging findings. Neural data suggests that the learning of new contrasts and the formation of new categories requires a certain amount of plasticity in the brain itself. The loss of the ability to modify one's phonological system with maturation is not due to sensory or outer neuron loss within the brain (Flege, 1995b). Thus, training and feedback could still produce both neural and behavioral changes at any age. For instance, some perceptual training can increase contrast discrimination for difficult sounds without necessarily provoking the formation of a completely new category. Patterns of neural activation for these difficult contrasts will change with training to incorporate a wider variety of brain areas, including articulatory mapping structures and auditory processing structures (Callan et al., 2003). These changes in patterns of neural activation imply a change in strategy used to discriminate the two sounds. The successful learner draws upon additional resources, such as considering “how would I produce this example,” rather than relying solely on the automatic detection processes associated with strong category boundaries.

For the creation of a new phoneme category, there appears to be two distinct steps. First, frontal lobe sensitivity to examples of the category develops, and then feedback from the frontal lobe can “tune” the automatic reactions in perceptual areas (Myers et al., 2017). Behavioral data also support a two-step process for the formation of a new phoneme category, which can start from either explicit instruction (Sakai & Moorman, 2018) or the accumulation of enough noise or variance in the speech stream signifying the possible presence of an additional category (Tuller et al., 2008). The first step is discrimination of the non-native sound from the native sound - a slow process, but not always necessary if the non-native sound is sufficiently far from any native categories - followed by a faster rate of solidification as new examples are fed into the forming category.

Influence of other cognitive abilities. As learning is typically considered a cognitive task, it is reasonable to consider what other cognitive abilities may also play a role in learning to perceive new contrasts. Some research has reported that certain aspects of memory may be relevant to phonological contrast learning. For example, higher working memory capacity is significantly correlated with more native-like L2 phonological processing (Darcy et al., 2015). There is also a correlation between short-term memory stores and phonetic discrimination ability (Silbert et al., 2015), while greater long-term memory capabilities result in a better chance for an individual to learn new contrasts (Golestani & Zatorre, 2009). A significant correlation between

L2 phonological processing and working memory capacity, a construct often strongly tied to attentional capacity, (Darcy et al., 2015) means that careful construct definition and methodological control is needed in order to tease out more specific ties than have already been demonstrated.

As a simple overview, within the field of psycholinguistics, memory is often defined as a triad: a short-term store where traces of a stimulus may be combined into stronger representations, a long-term store of phonological categories, and a store for cognitive strategies or plans (Strange, 2011). One thing that is inherent to each of these definitions is the depiction of memory as a location or holding container, not as a mechanism for directing and attuning processing units. This point contrasts with the definition of attention, which has been defined as a processing. Therefore, the interaction between attention and working memory becomes vital to consider when studying the perceptual learning process. For instance, the fact that contrasts with clear perceptual features are easier to distinguish may be connected back to attention or memory resources, as the salient acoustic and perceptual features of these contrasts reduce demands on the cognitive system during the learning process (Werker & Tees, 1984).

Statistical Learning

Many of the most promising explanatory models of phoneme categorization processes rely on the mechanics of pattern recognition and statistical learning (Lotto et al., 2000). For instance, Neary's "double-weak" model (1997) presents pattern recognition as a method of speech perception and uses logistic regression to mathematically define mapped areas. Pattern recognition is a key part of the basis for statistical learning, in both specific pattern examples and general rule acquisition (Aslin & Newport, 2012).

Statistical learning is associated with a variety of different aspects of language development: the discovery of syntactic rules, the updating of word meanings, and the categorization of speech sounds (Conway et al., 2010; Dobel et al., 2009). Although there are some caveats, in general, better statistical learning ability is correlated with better language development outcomes, particularly in regard to phonological development (Siegelman et al., 2017). However, as with all research paradigms, there are extenuating factors that must be considered before making broad generalizations about the relationships between statistical learning and speech or language development.

Definition. Siegelman and Frost present a definition of the broad concept of statistical learning as "the ability to implicitly pick up regularities of {verbal/non-verbal} information in the {visual/auditory} modality, when contingencies are {adjacent/nonadjacent}, thereby shaping behavior" (2015, p. 6). There are a few important aspects of statistical learning that are prominent in this definition.

First of all, statistical learning is an implicit ability that is stable within an individual and can exist independently of certain other cognitive abilities, such as working memory or general intelligence (Conway et al., 2010; Siegelman & Frost, 2015). This does not mean that the process of statistical learning is identical throughout the lifespan, however, just that the baseline ability stays the same. According to some research, adults and children use different statistical learning strategies due to their different levels of lexical knowledge (Peperkamp et al., 2006; Pierrehumbert, 2003, among others). Adults automatically use type statistic strategies, which are based on patterns and distributions over an entire lexicon, whereas infants or young children rely on token statistic strategies, which are based on surface-level patterns within a running speech stream (Pierrehumbert, 2003). This difference in strategy use could explain why children, who are relying more on the statistics of the physical stimuli presented to them, are better at perceiving contrasts between default phoneme productions and their allophones than adults, who are not as reliant on the physical stimuli themselves (Peperkamp et al., 2006).

Secondly, statistical learning is based on the patterns and regularities present within the stream of information individuals perceive within their environment. Although context is an important predictor of the likelihood that a learned distinction will be generalized (Idemaru & Holt, 2020), often statistical learning is considered an ability that allows for generalization to new scenarios (Aslin & Newport, 2012). This part of the definition could be considered synonymous with pattern recognition abilities, but the final clause of Siegelman and Frost's definition extends beyond simple recognition and includes the "learning" aspect that allows this process to modify behavior. The power of statistical learning comes from the fact that statistical regularities are not required to be adjacent to each other in order to be recognized, although patterns closer in proximity may be more easily recognized.

The Siegelman and Frost (2015) definition is primarily focused on the pattern extraction aspect of statistical learning, but other definitions include different key features as well. For instance, an important part of the most promising models of statistical learning is the ability both to learn conditional relations and to integrate the information gathered across examples to recognize predictive cues in new contexts (Thiessen et al., 2013). However, most models of statistical learning are greatly affected by variance in the reliability of context clues within the information participants are given. That is, if context clues are not consistently reliable, learners may fail to recognize predictive cues. Therefore, it may be more accurate to view both the "pattern recognition" and "rule learning" claims as simply microscopic and macroscopic views of the same statistical learning ability. These two facets could be linked together in human subjects by an intrinsic bias towards highlighting regularities of all types, whether they exist on a perceptual or cognitive level (Aslin & Newport, 2012). As such, statistical learning could guide an individual's thoughts and behaviors by forming multi-level cognitive architectures (in infants) or modifying existing structures (in adults). The various levels of representation within those architectures have logical dependencies weighted between them by the generalizing aspect of statistical learning, and the patterns represented within each level can reach from the most basic parametric features of phonetics, through phonological grammars, and beyond to higher levels of language processing (Pierrehumbert, 2003). Combining the key features of all of the presented arguments, a working definition of statistical learning is as follows:

> Statistical learning is the cognitive ability that allows for the automatic, implicit detection of both adjacent and nonadjacent regularities of information, and the integration of these detected patterns into rules that can be generalized to new contexts.

Some scientists have questioned whether individuals who perform well on statistical learning tasks simply have better cognitive abilities overall. However, according to current research, this is not the case: statistical learning ability is not correlated with general intelligence (IQ) or memory capacity (Conway et al., 2010; Siegelman & Frost, 2015). On the other hand, perceptual capability is correlated with statistical learning ability. Individual differences in perceptual abilities and their connection to a listener's ability to discriminate phonetic contrasts are reliable, even at the neural level (Lin et al., 2017).

Given the previous definition, it is clear that statistical learning theories may be applicable to a variety of branches of human development. For example, early models of speech perception, such as the Perceptual Magnet Effect (Iverson & Kuhl, 1995) and the double-weak model (Nearey, 1997), created maps of perceptual space in order to explain how phonological categories were formed and predicted different contrast perception results based on where a sound was located on that map. Statistical learning can be used to accomplish this same mapping technique, and the models based on it can provide additional insight into what that map may "look" like. For instance, although phonemic categories are formed and updated by the frequency of a specific value's occurrence along acoustic dimensions, phonological categories cannot be formed without some cue to indicate the correct number of contrasts these tokens should be divided into (McMurray et al., 2009). The computer model presented in this particular study was only able to successfully learn contrasts when competition was included as an aspect of the model, and the resulting phonological map included a number of unmapped spaces between categories. The shape of the frequency distribution of presented contrasts may also function as a cue to the correct number of categories. For instance, statistical learning can correctly guide the formation of two unique categories in a bimodal distribution with sufficient space between the two peaks, but it cannot form two categories in a unimodal distribution (Maye et al., 2002). These results imply that sparseness, or blank areas on the perceptual map, are key to the perception of phonological contrasts, complicating the formation of new phonological categories within areas on a map that are more densely populated.

Statistical learning paradigms. Methodologically, there are a few arguments regarding the best method of providing training to help listeners form new phonological categories in a statistical learning paradigm. Researchers debate whether participant training should focus on identification or discrimination, whether lexicality should be taken into account in stimulus design, and whether feedback should be provided during training.

Although identification training techniques are generally considered more challenging (Golestani & Zatorre, 2009), they are also preferred by participants over discrimination training techniques (Flege, 1995b). Ultimately, both identification and discrimination training of novel perceptual contrasts lead to similar performance outcomes and maintenance results among participants (Flege, 1995b), but the mechanism by which each method of training produces results is different. Identification traning is more likely to help learners develop long-term memory stores regarding the learned category (Golestani & Zatorre, 2009), but discrimination training is more likely to highlight category boundary markers in a more native-like way (Flege, 1995b).

However, the type of identification task can also play a role in listeners' abilities to learn new contrasts. Identifying abstract contrasts can be much harder than identifying meaning-based contrasts (Dobel et al., 2009). Although not all listeners need additional supports to form new categories (Tees & Werker, 1984), providing contrasts that are meaning-based during training can be helpful for many participants to connect lexicality to the contrast formed by the new category (Dobel et al., 2009). The addition of lexicality to the contrast-learning task can benefit some participants by engaging the higher-order pattern representations that are also associated with statistical learning (Peperkamp et al., 2006).

On the other hand, because speech is a multidimensional signal, any training of speech perception or development under a statistical learning methodology can also benefit from feedback after every trial and explicit instructions on how to categorize stimuli (Goudbeek et al., 2007). Providing explicit instruction at the beginning of a task, in terms of phonetic or articulatory differences that may be needed to create the new category, can increase the amount of improvement gained from perceptual training (Sakai & Moorman, 2018) as well as assist with maintenance of the category over subsequent weeks (Goudbeek et al., 2007). These instructions may be able to specifically direct attention in order for listeners to overcome the perceptual bias caused by existing categories (Darcy et al., 2016).

Feedback may also help to overcome existing perceptual biases in a statistical learning paradigm, especially within adult populations and models (McMurray et al., 2009; Peperkamp et al., 2006). Although other modalities may not require as much support (see Goldstone (1994) for category learning in the visual modality without feedback), perceptual learning of a new phoneme category requires feedback to allow self-monitoring systems to engage (Dobel et al., 2009). For example, participants cannot successfully learn to produce a non-native phoneme in a passive-exposure model of statistical learning, but they can be successful with explicit feedback from the presenter (Rugesaeter, 2014). In addition to such explicit feedback, learners can also benefit from implicit feedback (Gabay et al., 2015).

The importance of providing feedback is also supported by research including neurological data. The results of neural studies suggest that, with presentation alone, a new phoneme is simply absorbed into an existing native-language category in a process that can hamper the perception of the intended contrast (Dobel et al., 2009). Other researchers have suggested that not only is feedback from internal systems necessary for phonological category learning and subsequent contrast perception, but general external feedback from a speaker population (Pierrehumbert, 2003) or specific feedback during a training process is necessary to help make the relevant acoustic patterns salient and allow the statistical learning mechanism to ignore irrelevant patterns (Dobel, et al., 2009; Goudbeek et al., 2007).

Cognitive Factors

Differences in cognitive abilities have been shown to be related to an individual's speech perception abilities and categorization strategies (Kapnoula et al., 2017). It has long been suggested that attention capacity influences an individual's categorization ability (Goldstone, 1994), and some researchers suggest that listeners must redistribute attentional resources in order to maintain speech perception levels in difficult listening conditions (Kong & Edwards, 2016). However, the research findings are not consistent on this topic. Other researchers have suggested that perception is not correlated with similar cognitive processes, such as attentional control or attentional switching abilities (Darcy et al., 2015; Siegelman & Frost, 2015). Some of the ambiguity in the literature may be related to either the obscuring nature of group averages (Kong & Edwards, 2016), or differences in definitions. For instance, one study cited updating and shifting executive functions as indicators of an inhibitory control ability that was associated with segmentation (Darcy et al., 2016), but other researchers consider those functions part of the attentional network (Galvao-Carmona et al., 2014).

Areas of ambiguity aside, research generally suggests that some aspects of attentional inhibitory control processes may be related to both perceptual abilities and to some aspects of L2 contrast learning (Conway et al., 2010; Siegelman & Frost, 2015). It is important to keep in mind, though, that the connection between cognitive system efficiency and phonological learning processes is still not a simple one. For instance, production training of consonants gets an added boost in performance for individuals with high inhibitory control, while there is no such benefit for vowel production performance (Darcy et al., 2016). It has been suggested that the difference in performance between consonant and vowel contrast discrimination has to do with a difference in auditory short-term memory resources for each type of phoneme, with a stronger memory trace associated with vowel acoustic cues (Pisoni, 1973). Such cognitive-based findings can be easily folded into existing theories of speech perception. Perhaps the stronger memory cue from the native language serves as a magnet for vowel sounds, skewing perception and therefore making contrast learning harder (Iverson & Kuhl, 1995), whereas consonants, with their shorter memory traces, are less affected.

Additionally, aside from L2 learning, the application of attention is related to the initial learning and formation of category boundaries in L1. For instance, both English and Finnish have an /i,ɪ/ contrast, but English speakers attend to spectral cues to differentiate between them, whereas Finnish speakers attend to duration cues for their category markers. A form of attentional shifting is therefore required in order to switch between the processing strategies that highlight native language contrasts and the strategies that highlight an L2 contrast (Werker & Tees, 1984). In fact, it is a long-standing claim from the Speech Learning Model (Flege, 1995a) that gains in non-native phoneme perception are due to adjustments in selective attention.

More recent research also supports the idea that difficulty in learning non-native contrasts is due to where attention is directed. Attention-to-Dimension models suggest that the perception of these non-native contrasts can be learned by shifting attention to different acoustic dimensions, because native speakers from different languages attend to, and therefore highlight, different types of cues (Myers et al., 2017). Mathematically, this role is seen when statistical learning models suggest that attention is responsible for the weighting of the cue distributions. This weighting role of attention is what allows for the development of non-native contrast perception, because pattern extraction and integration can only be linked together when attention is turned to groups of features at the same time (Thiessen et al., 2013).

Definitions of attention. Clearly, "attention" is a word that is used in a variety of contexts, and it can be difficult to separate this particular cognitive process from other aspects of cognition. According to the Automatic Selective Perception model (Strange, 2011), selective attention affects acoustic-phonetic cue weights on subsequent behavior but is not volitional, whereas attentional focus is a voluntary, goal-driven method of attunement to particular information. Within Strange's model, attention is considered to be the deciding factor between controlled and automatic processing of information, that works alongside an internal, explicit awareness, such as a listener's metalinguistic knowledge of the aspects of a phonological structure. This model more closely associates attention with other cognitive abilities, making the construct more difficult to sum up in a single definition.

On the other hand, in developing the Attention Network Test, Fan, McCandliss, Sommer, Raz, and Posner (2002), differentiated attention from other cognitive processes by dividing the definition into three categories. *Alerting* is the aspect of attention that pulls an individual into a state of preparedness to receive information and keeps them there. *Orienting* is the part of attention that focuses on particular pieces of information about the stimulus, such as spatial location or stimulus modality. Finally, *Executive Function* allows individuals to make attentional decisions when there is conflict. These three pieces of attention are generally independent, each housed in different areas of the brain. On the other hand, the kinds of attention considered in the development of the Automatic Selective Perception model (Strange, 2011) have a much different focus than those considered for the development of the Attention Network Test. Although the theoretical basis of the Automatic Selective Perception model is useful for tying attention to other aspects of speech perception and learning, the Attention Network Test is ultimately better suited to breaking apart this broad concept for testing.

With these and similar considerations in mind, a basic explanation of attention is proposed.

> *Attention* is a cognitive function that directs and attunes processing units to particular features of presented information in order to carry out goals. The application of attention is a voluntary process, and the maintenance of attention may be voluntarily or involuntarily continued, but the cessation of attention may happen with or without intent.

This definition can also inform the discussion on what attention is not, and what aspects of human behavior are connected to other constructs. For instance, attention is a voluntary, goal-driven process in this definition, and so the impetus to apply attention to something must be internal (although, in the case of an unexpected noise or sudden movement, that impetus may be a basic desire to ensure one's own safety). It is also a cognitive function, which means that it is not an entirely unconscious process, i.e., an individual can become aware of their own attentional state. The definition provided above also allows for attention to be applied to internal sources of information, such as in the case of self-monitoring techniques. The final part of this definition points out that, although attention may provide attunement or direction to processing units, it is not responsible for the management of the information received by those units. Because the organization and interpretation of features within the information stream can be involuntary and initiated by outside sources, these activities would fall under the definition of "perception".

Connection to perception. In the above description of attention, it was noted that although the application of attention is voluntary, and the maintenance of attention can be intentional, the cessation of attention is not always voluntary. In research that relies on tasks that require attention, the cessation of attention can lead to poor results, whether caused by a gradual fade due to fatigue or a sudden switch to another stimulus. This shift in perception could lead to a dip in task performance that is unrelated to the experimenter's intended manipulations. In addition, the nature of the task itself, such as categorization, may modify a participant's attention (Goldstone, 1994) by changing the weights on perceptual processes the participant believes are needed to achieve their goals. However, it is important to note that, although categories might produce a strong influence on a listener's perception, the creation of categories does not overrule all independent aspects of perception entirely. In fact, some researchers go so far as to suggest that the influence categorization has on a listener's ability to learn novel contrasts is due to the category's direction of attention, rather than perception (Myers et al., 2017).

Finally, the different roles attention is claimed to play in practiced and unpracticed tasks may help to explain an individual's difference between perceptual abilities in L1 and L2. Because an adult's native language is so highly practiced, attention may not be needed for basic speech perception within the native language (Strange, 2011). However, speech sounds in a second language are not as over-learned as speech sounds in the listener's first language, so attention is required to detect non-native phonological contrasts (Strange, 2011). Additionally, dynamic redistribution of attention to more easily perceptible cues may help maintain speech perception performance in noisy listening conditions (Kong & Edwards, 2016). In fact, some researchers claim that attention helps to provide a better signal-to-noise ratio for desired stimuli across different modalities, as well as guiding perception with a template of features that matches expected targets and provides a boost to signals matching that template (Lu & Dosher, 1998). This template model also falls in line with speech perception, with phonological categories of a native language functioning as the expected targets for the template to match. Unless otherwise directed by attentional controls, the template would highlight the more practiced features that are likely to correlate with category boundary markers so as to improve classification ability despite interference (Lu & Dosher, 1998).

Statement of Problem

The general purpose of this study was to examine the impact of attentional abilities and explicit direction on adult listeners' ability to learn a novel phoneme contrast. To accomplish this goal, there were a number of stimulus design factors considered. These factors included the type of distinction (Beddor & Gottfried, 1995; Pisoni, 1973; Polka, 1991; Sakai & Moorman, 2018), the lexicality of the contrast (Curtin et al., 1998; Pierrehumbert, 2003; Peperkamp et al., 2006; Polka, 1991), and the variability within stimulus sets (Iverson, et al, 2003; Lim & Holt, 2011; Lively et al., 1993; Perrachione et al., 2011).

First, the types of phonemic distinction can increase or decrease the difficulty of learning a new phonological contrast (Beddor & Gottfried, 1995). For instance, gains in perceptual training are greater for obstruents than for vowels or other sonorants (Sakai & Moorman, 2018). Additionally, consonant contrasts are often less variable between participants who speak the same native language than vowels are, and therefore easier to experimentally control when using natural speech (Beddor & Gottfried, 1995; Pisoni, 1973). Although more acoustically disparate contrasts can be more salient and easier to learn (Polka, 1991), this is not likely to be the most important aspect involved in the learning process (Lively et al., 1993). In fact, it is possible to ascribe the difference provided by acoustic features back to attention or memory resources, as contrasts presented with more clear and acoustically distinct features may simply be easier to learn initially because they reduce demands on the cognitive system (Werker & Tees, 1984). There may be individuals with greater sensitivity to acoustic feature distinctiveness, but this is not necessarily the case for all learners (Kong & Edwards, 2016). For the present study, it was decided that consonants would be used as the contrast of interest, rather than vowels, in order to make the task easier for listeners.

Second, lexicality can also play a role in phonological contrast learning. It is possible that this lexicality effect is due to the role that attention plays in connecting lower-level cognitive functions, such as perception, to higher-level cognitive representations (Mirman et al., 2008). Typically, it seems that L2 contrasts that have a lexical function in a learner's L1 are easier to learn than contrasts that do not (Curtin et al., 1998; Polka, 1991). For instance, the contrast between /b/, /p/, and /p^h/ that is found in Thai, Hindi, Punjabi, and similar languages provides lexical contrast in those languages. However, the contrast between /p/ and /p^h/ has no lexical meaning in English. While English speakers are able to produce all of these sounds in specific contexts, they can rarely perceive a difference between all three productions. In addition to the fact that English has only two phonological categories encompassing these three sounds, the Thai /p/ has a voice onset time (VOT) that is closer to the English /b/ than to either the English /p/ or the Thai /p^h/. The average English /b/ has a VOT of 1 ms, with a range between 0 and 5 ms, whereas the average English /p/ has a VOT of 58 ms, with a range between 20 and 120 ms. The Thai /p/, on the other hand, sits in the gap between the two English consonants, with a range of 0 to 20 ms. However, the average Thai /p/ falls slightly closer to the average English /b/, with a VOT of 6 ms. Values for the average VOT for these consonants in each language are shown in Table 1.

Table 1. *VOT in ms for English and Thai stop consonants*

English	*Average /b/ VOT (range)*	*Average /p/ VOT (range)*	
	1 (0:5)	58 (20:120)	
Thai	*Average /b/ VOT (range)*	*Average /p/ VOT (range)*	*Average/p^h/ VOT (range)*
	-97 (-165:-40)	6 (0:20)	64 (25:100)

Note. The data for Table 1 are from Lisker & Abramson (1964)

The relative distributions of VOT between these bilabial plosives means that English speakers are more likely to assimilate the Thai /p/ into the same phoneme category as /b/ in most contexts. According to the Perceptual Assimilation Model, this should make discrimination between the two assimilated sounds more difficult (Best, 1995), which makes this contrast ideal for studying learning over time. As such, Thai /p/ and /b/ were chosen as the contrast of interest for the present study.

The final concern for phonological contrast learning involves the variability of the stimulus set overall. The amount of variability within a stimulus set can determine how heavily listeners weigh different acoustic cues, and therefore how they learn to perceive novel contrasts (Lim & Holt, 2011). High variability stimuli are more easily generalized to novel contexts (Iverson et al., 2003). This is particularly true when using stimuli from a variety of speakers (Lively et al., 1993). However, individuals with poor perceptual abilities may become overwhelmed when participating in high-variability training sets. They need lower variability training in order to decrease the processing demands of the task and give them a chance to form predictable categories before attempting to generalize (Perrachione et al., 2011). A solution suggested by Perrachione et al. (2011), that would provide the greatest benefit, is to have a training paradigm that has "blocked" variability. This includes low trial-by-trial variability within a set, but high variability between sets. In this design, multiple different speakers provide speech samples, and stimuli are presented in blocks separated by speaker. This structure was chosen for the present study.

As previously mentioned, there are likely interactions between contrast learning, perception, and other cognitive functions. Learning is easier when demands on the cognitive system are reduced (Werker & Tees, 1984), as well as when higher-level cognitive representations can be used to support perceptual learning processes (Polka, 1991). It is possible that these benefits are due to the role of attention as a bridge between lower- and higher-level cognitive functions (Mirman et al., 2008). However, "attention" itself is not a homogenous concept. The three different subsections of attention described by the Attention Network Test (Fan et al., 2002) could be expected to interact with a perceptual learning task in different ways. One of the key questions for the present study was to determine the nature of that interaction.

Based on the above considerations, this project used an identification training design with auditory stimuli matched to visual stimuli in order to promote lexical reasoning. The inclusion of a visual object tied to the nonsense words allows activation of statistical learning abilities across multiple levels of representation (phoneme, morpheme, etc.), which can improve learning outcomes (Dobel et al., 2009; Pierrehumbert, 2003; Stephens & Holt, 2010). As the participants learn the lexicality of the spoken stimuli, the hypothesis is that this higher-level change will modify what is perceived as an allophone and what is not (Peperkamp et al., 2006). Participants were provided basic feedback during the training portion of the task, which was hypothesized to help to guide the learning process (Gabay et al., 2015; Goudbeek et al., 2007). Additionally, the study experimentally addressed whether or not the provision of explicit instruction on the contrast of interest could enhance task performance by helping to direct the listener's attention to key features. A number of both planned and exploratory linear models were constructed to address the impact of interactions between explicit direction and attentional network subscores, as well as to identify any main effects for the provision of explicit direction or for scores on any of the three attentional networks (alerting, orienting, and executive function). In each model, the study's first research question was answered by examining interaction effects between the ANT scores and the experimental condition (explicit directions or general instruction). The study's second and third questions were answered by examining main effects of individual ANT scores and the experimental condition, respectively.

In sum, this study addresses the following research questions:

1. Does the ability to benefit from explicit, external direction of attention vary by ANT subtest score (i.e., alerting, orienting, and executive function) when learning a novel phoneme contrast in a statistical learning paradigm?
2. Do performances on specific ANT subtests predict an individual's ability to learn a novel phoneme contrast?
3. Does providing an explicit, external direction of attention improve a listener's ability to learn a novel phoneme contrast overall?

Methods

Participants

Talkers. Speech samples were provided by four different talkers. All talkers were adults (20-36 years old), native Central Thai speakers with no self-reported history of speech, language, or hearing impairment. There were two male and two female talkers, all of whom had spent less than 5 years in a primarily English-speaking country.

Listeners. Listeners were recruited to the experiment via an open online link. A total of 106 individuals provided informed consent to participate in the experiment and reported that they were over 18 years of age. No other demographic data were elicited. All 106 participants then took a Language Screener to certify that they were monolingual English speakers, were not heritage speakers of other languages, and did not have repeated exposure (such as conversations with close friends, coworkers, or family) to the Thai, Punjabi, or Hindi languages, which contain the contrast of interest for this study. Of the original 106 participants, 69 participants passed the Language Screener. Fifty-seven of these participants then passed a speech-in-noise and loudness discrimination task to ensure that they had normal hearing and were wearing headphones as instructed (see Screeners, below).

Materials

Screeners. Listeners' monolingual status was assessed with the Participant Screening Form in Appendix A. This screener asked participants whether they were native English speakers, if their childhood caregivers had spoken to them in a language other than English, if they had ever studied another language or were comfortable holding a conversation in another language, if they had recently lived in an area where English was not the primary language, and if they had frequent exposure to the Thai, Hindu, or Punjabi languages. These three languages (Thai, Hindu, and Punjabi) were specifically highlighted because they all have similar lexical contrasts with /b, p, p^h/.

Participant compliance in wearing headphones was checked using the loudness discrimination task described by Woods et al. (2017). In this task, the participants had to determine which of three stereo-recorded tones was the quietest. On each trial, one of the two louder tones was presented 180º out of phase across the two channels, making the task more difficult if the listener were using speakers instead of headphones. Five out of six trials had to be answered correctly for the listener to pass.

A speech-in-noise perception test was used to ensure that the listeners had their computer headphones set to an appropriate volume level. For this task, listeners had to identify English words in quiet and at -2dB SNR on a multiple-choice test. Target words were presented in multi-talker babble, and listeners had to select the correct word out of six possible options displayed on the screen. Listeners had to score no lower than 10% below the average score for normal-hearing listeners in a laboratory setting (17 out of 18 on the quiet trials, 13 out of 18 trials on the -2dB trials) in order to continue. Participants were told not to adjust their computer volume after the speech-in-noise test, as the words used in this task were normalized to the same peak amplitude as the experimental stimuli.

Attention Network Test. The Attention Network Test (ANT; Fan et al., 2002) evaluates three relevant aspects of attention: alerting, orienting, and executive function. The task was to select which direction a target arrow pointed as quickly and accurately as possible. Targets were simultaneously presented with flanking distractor arrows that participants were instructed to ignore**.** Each trial had four possible cue conditions and three possible flanker conditions. The central cue condition included a cue (an asterisk) displayed at the fixation point. The spatial cue condition displayed a cue either above or below the fixation point, at the location of the next target. The double cue condition displayed a cue both above and below the fixation point, and the no cue condition displayed only a blank screen with the fixation point. For the flanking conditions, the flankers pointed as the same direction as the target in the congruent condition and in the opposite direction as the target in the incongruent condition. The neutral flanking condition used horizontal lines as flankers rather than arrows. Overall, the ANT can be completed in approximately 20 to 30 minutes.

On every trial of the ANT, participants began looking at a fixation cross, which lasted between 400 and 1600 ms. They were then primed with one of the four cue conditions. Five hundred ms after the cue was given, the target arrow and flankers were displayed either above or below the fixation cross. Participants had to respond within 1700ms of the target's appearance by pushing the *f* key if the arrow was pointing to the left, or the *j* key if the arrow was pointing to the right. After the participant responded, or failed to respond within the designated time, the test immediately continued to the next trial. A visual summary of this procedure and an example trial are provided in Figure 1.

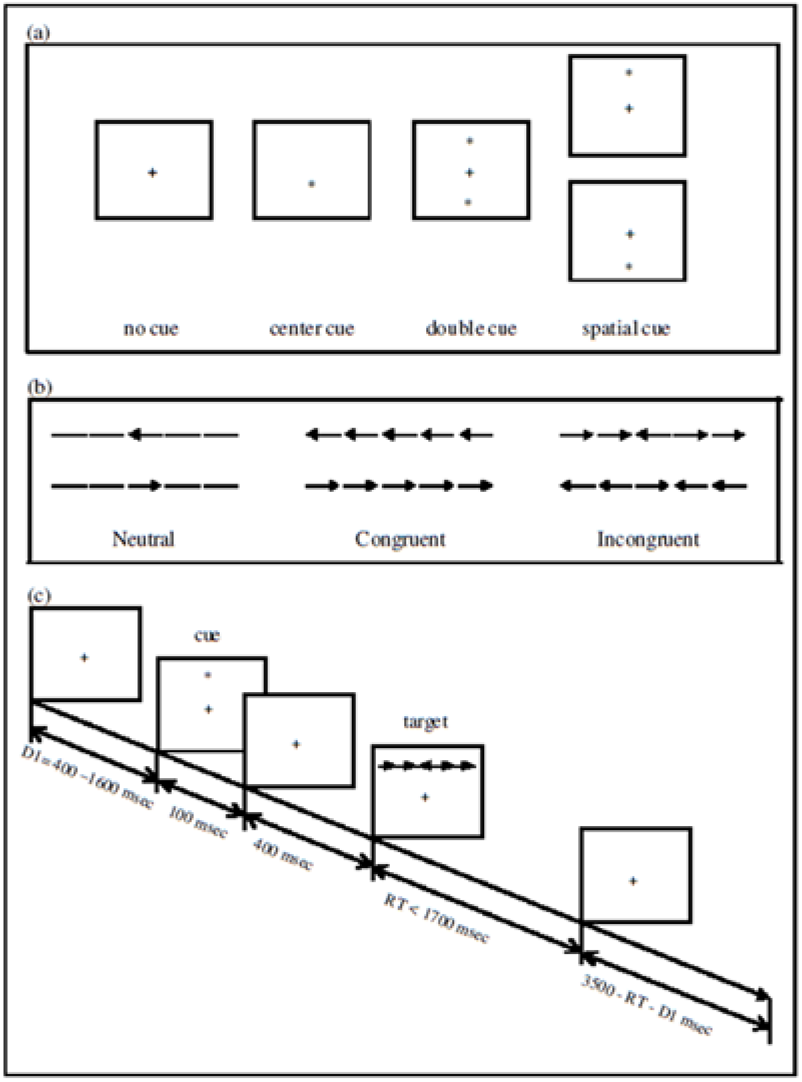

Figure 1. ANT experimental procedure. (a) All possible cue conditions; (b) All possible flanker and target conditions; (c) example of a single trial with spatial cue presentation, and an incongruent flanker condition. For this example, the participant would have to indicate the target arrow was pointing to the left within 1700ms of the target's appearance (Fan et al., 2002, copyright MIT Press)

Stimuli. The stimuli for the perceptual contrast learning experiment consisted of individual VCV nonsense words. The target contrasting phonemes were the Thai /b/ and /p/, which are often perceived by English speakers as belonging to the same category (/b/). The target phonemes were embedded between vowels from different areas of the vowel quadrilateral: /i, u, ɛ, ɔ/, plus the high, unrounded, back vowel (/ɯ/), which is native to the Thai language, but absent in English. The result was a total of 5 minimal pairs (10 nonsense words) per set. Each auditory stimulus was associated with one of two visual stimuli. Stimuli that included the /p/ consonant were associated with the Thai letter used for that sound, ป, and stimuli that included the /b/ consonant were associated with the Thai letter used for that sound, บ. See Table 2 for the list of auditory stimuli. Larger images of the two Thai letters can be seen in the example trial in Figure 2.

Table 2. *Nonsense Syllables Used in the Statistical Learning Paradigm*

Vowel context	Minimal pair
High front	/ibi/ and /ipi/
High back	/ubu/ and /upu/
Mid-low front	/ɛbɛ/ and /ɛpɛ/
Mid-low back	/ɔbɔ/ and /ɔpɔ/
High unrounded back	/ɯbɯ/ and /ɯpɯ/

Procedures

Recording. Stimuli were recorded in a quiet room (<30 dB ambient noise), using a C 420 cardioid condenser headset microphone and a Solid State Recorder. A constant microphone to mouth difference of 5 cm was used, and the sampling rate was set at 16 kHz.

Both the IPA symbols and Thai spelling of each nonsense word were displayed to the talkers on the computer screen with the instructions "Say the word on the screen". Each speaker provided four recordings of each stimulus item to provide an appropriate number of stimuli and within-talker variability (Siegelman et al., 2017). This yielded a total of 40 target recordings per talker.

Stimuli were processed using Praat software (Boersma & Weenik, 2018). The noise reduction feature was applied to all stimuli. The scale peak was set to 0.7 and 0.01 seconds of quiet were included before and after stimulus presentation. For each stimulus item, the three recordings with the least amount of remaining noise were selected, resulting in a total of 30 stimuli for each of the four talkers or 120 stimuli to be used within the experiment.

Experimental procedures. Prior to the beginning of participant recruitment, the entire experiment was pilot tested using six volunteers. They were all adult native English speakers with varying amounts of linguistic training and experience in research design. The initial pilot tester confirmed that the experiment ran smoothly in the chosen software environment**.** Pilot testers also provided feedback indicating that the screening tasks were appropriately scaled. Finally, the responses from pilot testers confirmed that the chosen phoneme contrast was initially difficult to discriminate but could be learned within the course of the study. Data from pilot testers were not included in the final experiment.

Experimental participants were recruited via flyers posted in public locations, on various social media sites including the researchers' personal social media page, and in university class message boards. The experimental part of this study had four stages: a brief screening for language status, administration of the Attention Network Test, screening for headphone use and volume settings, and the statistical learning task. The experiment was presented online through the Gorilla platform (www.gorilla.sc). Prior to accessing the study, potential participants were given a brief description of the study, participation requirements, and the access link. This link led directly to informed consent documents that described the study purpose and the basic design (see Appendix B), as well as provided the experimenter's contact information so that the participants could reach out with any questions if they chose to do so.

Following consent, volunteers completed the Language Screener (Appendix A). Those who did not meet criteria were rejected, but participants who met inclusion criteria moved on to the second stage, administration of the ANT.

Standardized directions for the ANT were presented to the participants on-screen. If participants did not respond to a trial within 1700 ms, the experiment automatically advanced to the next trial. Following completion of the ANT, participants were offered a break before the next part of the experiment.

The participants were then told to put on headphones and adjust their computer volume until a stream of white noise presented by the experiment was "loud, but comfortable to listen to". They then underwent the headphone and hearing screenings. Those who failed the headphone screening were asked to reset their headphones and volume settings and try again. Those who failed a second time were dismissed from the experiment. Following completion of this step, participants were offered a final break before the last portion of the study.

The final stage included three steps: pretest, learning, and posttest. All tasks in the next three steps were presented as two-alternative forced-choice (2A-FC) tests. Recorded stimuli were presented along with the Thai letters associated with the [b] and unaspirated [p] sounds. After hearing each stimulus, participants pressed the *f* key to select the image for [p] on the left, and the *j* key to select the image for [b] on the right. On-screen instructions for this stage are provided in Appendix C.

The pretest step in the statistical learning stage of the experiment involved presentation of a single recording of each item in the minimal pairs from each talker (40 trials). These items were presented in random order to establish baseline perceptual discrimination between the two phoneme categories. In this step, the participant did not receive feedback after selecting an image, and it was expected that they would perform at chance levels. Scores significantly different than chance (greater or less than 50% correct) were taken as evidence that the participant could already perceive the target phoneme contrast. No participants performed significantly differently from chance on the pretest. An example response screen is depicted in Figure 2.

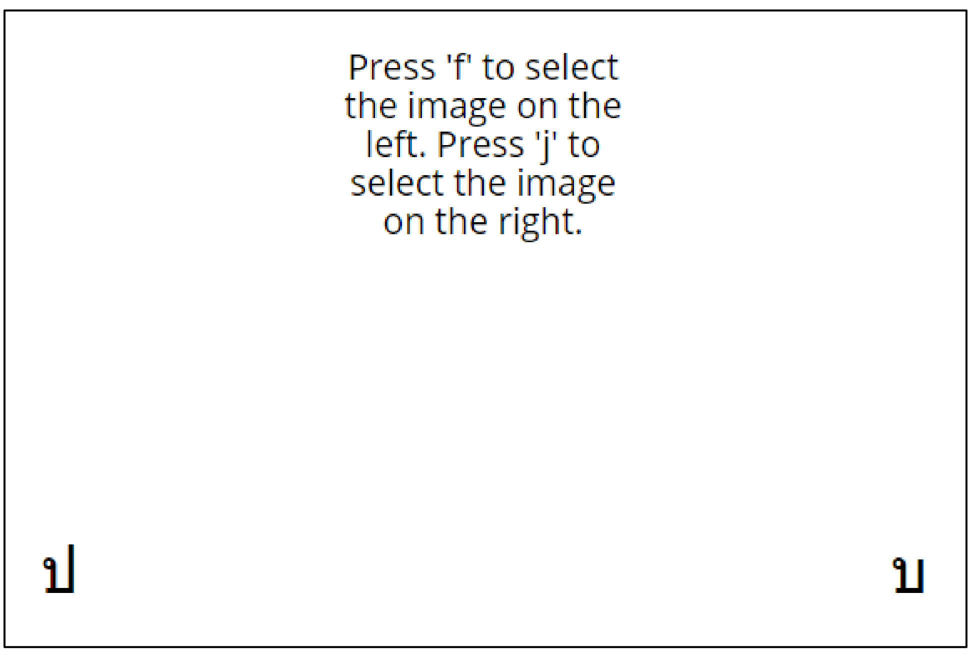

Figure 2. Statistical learning trail example. The images at the bottom of the screen are the Thai letters for the /p/ and /b/ sounds.

After the pretest, each participant was randomly assigned to one of two groups for the learning step. Group A (*n*=30) were presented with instructions that explicitly directed their attention to the target contrast with the directions “Listen carefully for the different consonant sound in each word.” The instructions given to Group B (*n*=27) did not have external cues directing them to attend to the target contrast (i.e., the directions stated “Listen carefully to each word”). The participant response task remained the same as in the pretest, but the participant received feedback after every trial in the form of a green (red) light every time they selected the correct (incorrect) answer. Within the learning step of the experiment, each of the 30 recordings from one talker were presented randomly four times within a block - that is, a total of 120 presentations per block. Blocks were separated by talker, and the 4 blocks were presented in random order to listeners. There was a total of 480 trials across the learning period. Group instructions were presented visually between blocks as a reminder.

The post-test step was identical to the pre-test step in that the participant did not receive feedback after selecting an image, although the stimuli were presented in a different random order. Once the participant completed the post-test, the software saved the experiment’s output files to a secure drive and presented a thank you message to the participants for completing the experiment.

Scoring. Pre-test, block, and post-test scores for the statistical learning task were calculated by a count of the number of correct responses made out of the total number of possible trials. ANT efficiency scores for each participant were calculated as the difference in average response time (RT) on correct trials of different types (Draheim et al., 2019). Each formula is designed to reflect the underlying cognitive process associated with each of the three attentional networks (Fan et al., 2002). The alerting effect was calculated by finding the difference between the mean RT of the no-cue and double-cue conditions. The orienting effect was calculated by finding the difference between the mean RT of the center cue and spatial cue conditions. The executive function effect was calculated without regard to cueing condition, but rather by finding the difference between the mean RT of all incongruent and congruent flanking conditions. All scoring calculations and subsequent analyses were carried out using R software (Rstudio team, 2020).

Results

Statistical analyses were carried out to ascertain factors that might affect the participants' ability to learn a novel phoneme contrast in a statistical learning paradigm. Outcome measures from a total of 57 participants included ANT scores, responses to the 40-item 2AFC perceptual pretests and posttests, and responses to the 480 learning trials. Two classes of linear models were created to analyze the data. Generalized Linear Models (GLMs) were used to model the participants' posttest scores, and Linear Mixed Effects models (LME) were used to model the participants' scores in each learning block. Since the research questions for this study focused on both individual predictors and their interactions, each linear model could assess all three questions at once. In each model, the study's first research question was answered by examining interaction effects between each ANT subscore (alerting, orienting, and executive function) and the experimental condition (explicit directions or general instruction). The study's second and third questions were answered by examining main effects of individual ANT scores and the experimental condition, respectively. Each participants' scores were rescaled into *z*-scores to reduce the impact of predictors with different magnitudes.

A few introductory tests were done to evaluate data quality prior to model formation. To confirm that the participants did not differ significantly in their ability to perceive the contrast prior to the learning phase, Welch's two-sample *t*-test was conducted to compare the pretest scores of the group who received explicit (M=19.600, SD=2.283) or general (M=18.963, SD=2.752) instructions. This test was not significant, $t(50.734)=0.945$, $p=0.349$, verifying that there was no statistically significant difference between the two groups on the pretest.

Secondly, a linear model was fit to predict pretest scores as a function of the alerting, orienting, and executive function efficiency scores (p=0.26, 0.53, 0.24, respectively). With no significant predictors in the model, it was determined that participants' efficiency scores did not significantly impact their ability to perceive the contrast prior to the learning phase. Finally, to confirm that learning did occur during the training trials, a paired t-test was conducted to compare the pretest (M=19.298, SD=2.514) and posttest (M=30.579, SD=9.071) scores of all participants. This test was significant $t(56)=9.039$, $p<0.001$, indicating that the participants in both groups were able to learn the phonetic contrast.

Posttest Scores

A Gaussian GLM with an identity link function was used to address the research questions regarding factors that could influence an individual's ability to learn a novel phoneme contrast (Model 1). The outcome measure for this equation was the participants' Posttest Scores (continuous). Fixed covariates were the Experimental Group (categorical with two levels), Alerting Score (continuous), Orienting Score (continuous), and Executive Function Score (continuous). Interaction terms were Group:Alerting, Group:Orienting, and Group:Executive Function. The theory driving the design of the ANT holds that subscores are independent and uncorrelated (Fan et al., 2002), so no interactions between the subscores were allowed within this model. Parameter estimates for Model 1 can be found in Table 3.

$$Posttest \sim Group*Alerting+Group*Orienting+Group*Executive \qquad \text{(Model 1)}$$

Table 3. *Parameter Estimates Predicting Posttest Scores in Model 1*

	Coefficient	Standard error	*t* value	*p*-value
(Intercept)	29.275	4.224	6.930	<0.001
Group	9.318	6.164	1.512	0.137
Alerting	-0.095	0.054	-1.742	0.088
Orienting	0.028	0.092	0.298	0.767
Executive Function	0.206	0.086	2.387	0.021*
Group:Alerting	-0.009	0.080	-0.112	0.911
Group:Orienting	-0.077	0.139	-0.551	0.584
Group:Executive	-0.268	0.127	-2.112	0.040*

Note: The *p*-values marked by an asterisk indicate significance below the 0.05 level. This model had 49 degrees of freedom.

Although this planned model yielded an interesting significant interaction between Experimental Group and Executive Function score, Model 1 itself is not a significant predictor of Posttest Scores, $F(7,49) = 2.179$, $p = 0.052$. This result could be due to a lack of power. There may also be methodological reasons for such a finding. For instance, despite the fact that the ANT subscores are intended to be uncorrelated and independent, some studies have found that there can be correlations between the orienting and alerting networks (Fan et al., 2002), or between the orienting and executive function networks (Galvao-Carmona et al., 2014). Such correlations could add unnecessary complexity to the model, potentially occluding true effects.

Based on the possibility that the Orienting network could be correlated with the Alerting or Executive Function networks, and the fact that neither Orienting nor its interaction term were significant predictors in Model 1, a new, exploratory model was created. Unlike Model 1, Model 2 does not include the Orienting Score as a covariate, nor the interaction term between Orienting and Group. In addition to simplifying the model, this change allows the remaining predictors to claim more of the variance in the outcome measure. Like Model 1, Model 2 is also a Gaussian GLM with an identity link function that predicts participants' Standardized Posttest Scores, with Experimental Group, Alerting Score, and Executive Function Score as fixed covariates and interaction terms between Group and both ANT subscores. Model 2 is a significant predictor of Posttest Scores, $F(5,51)=3.091$, $p=0.016$. Parameter estimates for Model 2 can be found in Table 4.

$$Posttest \sim Group*Alerting+Group*Executive \qquad \text{(Model 2)}$$

Table 4. *Parameter Estimates Predicting Posttest Scores in Model 2*

	Coefficient	Standard error	t value	*p*-value
(Intercept)	29.554	4.050	7.296	<0.001
Group	8.126	5.680	1.430	0.159
Alerting	-0.087	0.046	-1.880	0.066
Executive Function	0.200	0.082	2.424	0.019*
Group:Alerting	-0.030	0.069	-0.439	0.663
Group:Executive	-0.244	0.117	-2.080	0.043*

Note: The *p*-values marked by an asterisk indicate significance below the 0.05 level. This model had 51 degrees of freedom.

The pattern of results is similar between the two models. In Model 2, there is a significant interaction effect between Experimental Group and Executive Function score ($p=0.043$) and a significant main effect of Executive Function ($p=0.019$). In this model, there was no significant interaction between Experimental Group and Alerting score ($p=0.663$), nor was there a significant main effect of Experimental Group (*p*=0.159) or Alerting score (*p*=0.066). These results suggest the answer to the study's first research question was "yes": the benefit of providing explicit directions in a statistical learning paradigm does vary by one ANT subtest score (executive function). Similarly, the answer to the study's second research question is also "yes": performance on one ANT subtest (executive function) impacts a listener's ability to learn a novel phoneme contrast. The results of this analysis imply that the answer to the study's last question is "no": explicit directions alone did not significantly improve performance on the task.

The interaction between Experimental Group and Executive Function score can be seen in Figure 2. Those with high executive function scores benefitted more from external, explicit direction of attention when attempting to learn a novel contrast, whereas those with lower executive function scores did not. However, this additional benefit is still relatively small, as seen by the parameter estimate for the interaction term in both Model 1 and 2.

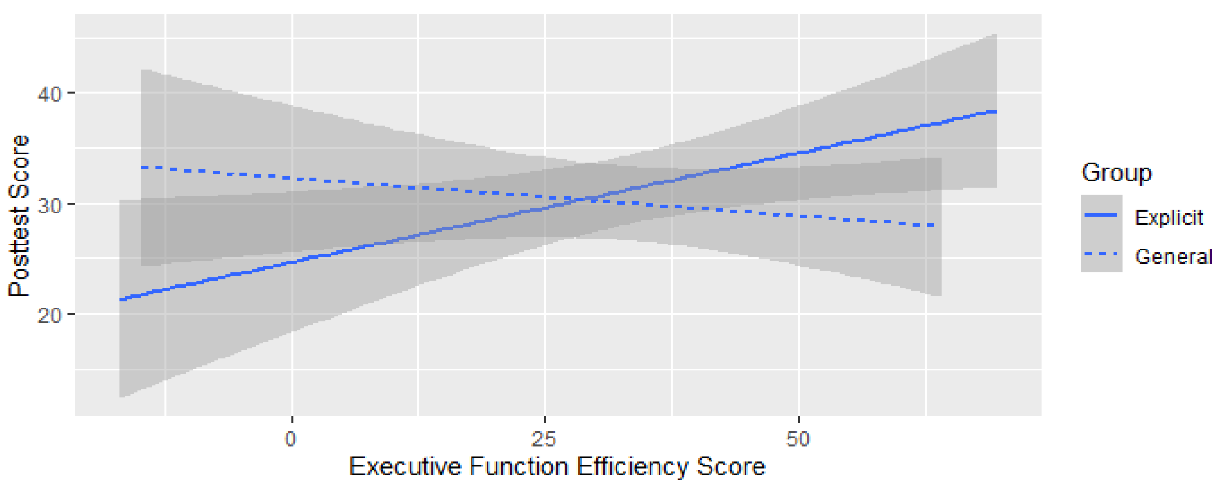

Figure 2. Interaction between experimental group and executive function as indicated by posttest scores.

Learning by Block

In addition to predictions about the participants' posttest scores, a series of exploratory LMEs were created to address factors that could impact participants' rate of learning. For all of these models, the outcome measure was Percent Correct for each participant by block. The pretest was considered Block 0 – that is, the percent correct achieved with zero blocks of training. Fixed effects included Alerting Score (continuous), Orienting Score (continuous), Executive Function Score (continuous), Experimental Group (categorical with two levels), and Trial Block (categorical with five levels, Blocks 0-4). The random effect was the Participant. As with the previous analyses, scores were standardized and rescaled into *z*-scores.

Model 3 represents the most basic model tested, including all five fixed effects, the random effect, and no interaction terms.

$$\textit{Percent Correct} \sim \textit{Group}+\textit{Block}+\textit{Executive}+\textit{Alerting}+\textit{Orienting}+(1|\textit{Participant}) \quad \text{(Model 3)}$$

Model 4 adds an interaction term between Experimental Group and Trial Block to the basic model,

$$\textit{Percent Correct} \sim \textit{Group}*\textit{Block}+\textit{Executive}+\textit{Alerting}+\textit{Orienting}+(1|\textit{Participant}) \quad \text{(Model 4)}$$

and Model 5 adds interaction terms between the ANT subscores and Trial Block to the basic model.

$$\textit{Percent Correct} \sim \textit{Group}+\textit{Executive}*\textit{Block}+\textit{Alerting}*\textit{Block}+\textit{Orienting}*\textit{Block}+(1|\textit{Participant}) \quad \text{(Model 5)}$$

Model 6 represents the full theoretical model, allowing for interaction terms between Group, Block, and each ANT subscore, but no interaction between the ANT subscores themselves.

$$\textit{Percent Correct} \sim \textit{Group*Executive*Block+Group*Alerting*Block+Group*Orienting*Block} \\ \textit{+(1|Participant)} \qquad \text{(Model 6)}$$

Finally, Model 7 is the full possible model that allows for interaction terms between Group, Block, and all ANT subscores.

$$\textit{Percent Correct} \sim \textit{Group*Executive*Block*Alerting*Orienting+(1|Participant)} \qquad \text{(Model 7)}$$

In a comparison of model fit, it was revealed that adding the interaction term between Group and Block alone did not significantly improve fit, $\chi^2(1)=0.378$, $p=0.539$. However, the addition of the two-way interaction terms between Block and each ANT subscore did significantly improve fit, $\chi^2(2)=14.815$, $p<0.001$. The inclusion of the three-way interaction terms between Block, Group, and each ANT subscore also significantly improved fit, $\chi^2(7)=23.603$, $p=0.001$. Finally, the addition of the four-way interaction between Block, Group, and all three ANT scores did not significantly improve model fit, $\chi^2(16)=23.043$, $p=0.113$. These results are summarized in Table 5.

Table 5. *Learning Curve Model Fit Comparisons*

Model	Number of Parameters	AIC	BIC	Chi-square	*df*	*p*-value
Model 3	8	-273.80	-244.58			
Model 4	9	-272.18	-239.30	0.378	1	0.539
Model 5	11	-282.99	-242.81	14.815	2	<0.001*
Model 6	18	-292.59	-226.85	23.603	7	0.001 *
Model 7	34	-283.64	-159.45	23.043	16	0.113

Note: The *p*-values marked by an asterisk indicate significance below the 0.05 level.

Because not all models were completely nested (see Models 4 and 5), model fit was compared based on a combination of the models' AIC, BIC, and chi-square significance test. Based on these results, Model 6 was selected as the model that provided the best fit for the data. Although Model 6 did not have the lowest BIC, it did have the lowest AIC, as well as showing significant improvement in fit. In this model, there was a significant three-way interaction between Block, Group, and Executive Function scores ($p=0.004$). Much like the results predicting posttest scores, this significant interaction implies that those with high executive function scores benefitted more from external, explicit direction of attention over the course of the learning phase, whereas those with lower executive function scores did not. There was also a significant two-way interaction between Block and Executive Function scores ($p=0.006$) and a significant main effect of Block ($p<0.001$) – that is, additional confirmation that participants did learn over time. There were no other significant interactions or main effects. Parameter estimates for Model 6 can be found in Table 6.

Overall, the results of these tests show that learning did occur during the course of training. Additionally, these tests also provide the same answer to this study's research questions as the first set of analyses. The results of this analysis suggest that individuals with better Executive Function scores benefited more from explicit direction of attention than others. However, attentional networks aside from Executive Function did not appear to significantly improve learning, nor did the provision of explicit directions alone.

Table 6. *Parameter Estimates Describing Learning Curve in Model 6*

	Coefficient	Standard error	t value	df	*p*-value
(Intercept)	0.548	0.027	20.448	86.547	< 0.001
Group	-0.034	0.039	-0.878	86.547	0.382
Block	0.075	0.007	10.966	220	< 0.001 *
Alerting	-0.037	0.030	-1.202	86.547	0.232
Orienting	0.039	0.031	1.262	86.547	0.210
Executive Function	0.035	0.028	1.252	86.547	0.214
Group:Block	-0.008	0.010	-0.845	220	0.399
Group:Alerting	0.0267	0.045	0.593	86.547	0.555
Group:Orienting	-0.040	0.047	-0.852	86.547	0.397
Group:Executive	-0.026	0.042	-0.620	86.547	0.537
Block:Alerting	-0.006	0.008	-0.745	220	0.457
Block:Orienting	-0.009	0.008	-1.178	220	0.240
Block:Executive	0.020	0.007	2.794	220	0.006 *
Group:Block:Alerting	-0.021	0.011	-1.821	220	0.070
Group:Block:Orienting	0.006	0.012	0.468	220	0.640
Group:Block:Executive	-0.031	0.011	-2.940	220	0.004 *

Note: The *p*-values marked by an asterisk indicate significance below the 0.05 level.

Discussion

This study investigated attentional factors and methodological features that could impact an adult's ability to learn a novel phoneme contrast. A series of GLM and LME models predicting participants' performance based on their ANT subscores and experimental condition were created in order to answer three primary research questions:

1. Does the ability to benefit from explicit, external direction of attention vary by ANT subtest score (i.e., alerting, orienting, and executive function) when learning a novel phoneme contrast in a statistical learning paradigm?
2. Do performances on specific ANT subtests predict an individual's ability to learn a novel phoneme contrast?
3. Does providing an explicit, external direction of attention improve a listener's ability to learn a novel phoneme contrast overall?

Previous research shows that some contrasts are easier or harder for speakers of a given language to perceive (Best, 1995; Flege, 1995a). For the current study, the contrast of interest involved the Thai phonemes /p/ and /b/. The Thai language has a three-way voicing/aspiration contrast to represent three distinct phonological categories for the sounds /p,p^h,b/. Although English speakers can and do produce all three of these sounds, the English language only draws a phonemic distinction between /p/ and /b/, with /p/ and /p^h/ classified as allophones that are used in different phonetic contexts. Due to a shorter voice onset time for the Thai /p/ and short lag VOT for most examples of /b/, monolingual English speakers are likely to assimilate the Thai /p/ and /b/ into the same phonological category and perceive both of them as similar to English /b/, although they will still perceive the Thai /p^h/ as an English /p/. Because Thai /p/ and /b/ fall into the same phonological category for English speakers, they should be difficult to discriminate at first (Best, 1995). This was seen in the current study, as all participants performed at chance levels upon their first exposure to the Thai phoneme contrast. However, with enough exposure to the novel contrast, English speakers should begin to develop an interlanguage phonology (Strange & Shafer, 2008) by shifting their existing phonological category boundaries (Flege, 1980), which would allow the English speakers to perceive the two Thai phonemes separately.

This experiment attempted to provide that necessary exposure to allow the listeners to start forming that interlanguage phonology. The contrast was presented in a statistical learning paradigm with 480 learning trials designed to shift listeners' perceptual boundaries, making it easier for them to recognize the two sounds as belonging to separate categories (Feldman et al., 2009). Each of the target sounds was associated with a visual object, creating a meaning-based contrast between the two sounds, and allowing participants to recruit statistical learning abilities across multiple levels of lexical representation (Pierrehumbert, 2003). Trials were presented with "blocked" variability (Perrachione et al., 2011) to promote faster mastery of the contrast, and participants were given feedback after every learning trial (Rugesaeter, 2014) in order to provide motivation and increase learning.

The experimental manipulation for this study was the provision of either explicit ("Listen carefully for the different consonant sound in each word") or general ("Listen carefully to each word") instructions during the statistical learning task. In addition to the statistical learning task, all participants completed the ANT in order to obtain efficiency scores for each of the attentional networks (Alerting, Orienting, and Executive Function). The Alerting network is responsible for preparing individuals to detect stimuli and respond to them, as well as for keeping individuals in that "ready" state. The Orienting network is responsible for allowing an individual to detect or predict information about a stimulus, such as its modality or physical location in space. Finally, the Executive Function network is a top-down control process responsible for helping individuals decide where to focus their attention if there are conflicting options.

For research question one, a specific interaction between executive function score and experimental group was predicted. In this case, individuals who had high executive function scores and were given explicit instructions would learn the contrast most quickly. It was also anticipated that there would be no interaction between the alerting score and experimental group, since individuals with high alerting scores should be more sensitive to variations in stimuli regardless of instruction. For research question two, a main effect of two of the ANT subscores was hypothesized, with better scores on either Alerting or Executive Function predicting better performance on the posttest. Because all stimuli were presented in the same modality and from the same location in space (i.e., the participant's headphones), there was no predicted effect of orienting scores on posttest performance. Finally, for research question three, the provision of explicit instruction was hypothesized to lead to better performance on the posttest overall.

The results of this experiment showed that, similar to previous research on adult second language learning, participants were able to form a new phonetic category during the course of the training. The 480 learning trials then provided enough exposure to shift most participant's perceptual category boundaries so that they were able to discriminate between Thai /p/ and /b/. Statistically, this claim was supported not only by the significant difference between pretest and posttest scores overall, but also by the significant main effect of Block on participants' performance during the learning trials. The significant main effect of Block implies that the participants' scores changed over time as participants practiced the discrimination task.

Although not one of the study's research questions, it is interesting to note that participants in this experiment did not appear to be following the two-step process of phoneme category formation, as suggested by previous research. In adults, it has previously been suggested that sensitivity to the contrast must develop first, and then perceptual tuning takes place (Myers et al., 2017). The initial development of sensitivity is a slow process, sparked by either explicit instruction (Sakai & Moorman, 2018) or extreme variance in presented speech streams (Tuller et al., 2008). The tuning and solidification of the new category occurs at a faster rate as new examples are encountered. As can be seen in the summary of participant performance in Figure 3, all trajectories are roughly linear, rather than including a slow discrimination phase followed by a faster phase for solidification. It is possible that this two-step learning process is not universal, but rather only holds for certain conditions.

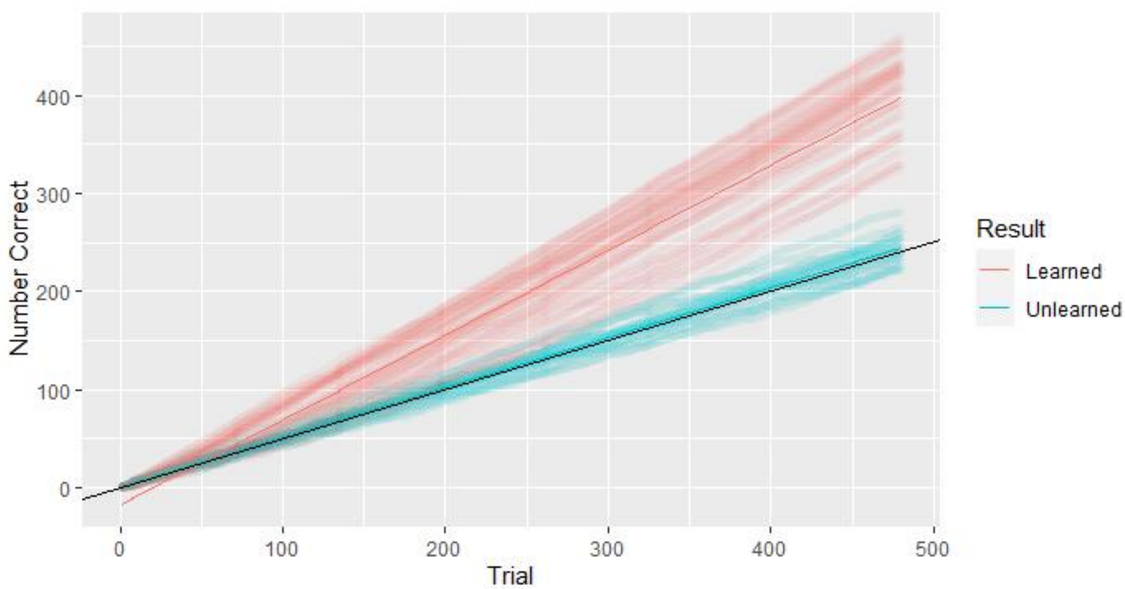

Figure 3. Number of correct responses from each participant over time.
Data comes from all participants regardless of experimental condition (i.e., general vs. explicit directions). The black line indicates chance performance.

Interactions Between ANT Subscores and Explicit Directions

The goal of this study was to offer at least a partial explanation for why some participants were successful in learning a phonological contrast when others were not. One possible reason is that innate attentional abilities may influence how an individual responds when given explicit or general instructions prior to a learning task. This idea led to the study's first research question, exploring interactions between the provision of explicit direction and ANT subtest scores that could impact a participant's ability to learn the phoneme contrast. Interaction terms in the GLM and LME models between the individual ANT scores and experimental group were used to answer question 1.

Both GLMs included a significant two-way interaction between executive function scores and experimental group, with the result that participants with high executive function scores benefitted more from explicit directions than those with lower executive function scores. This finding suggests that individuals who are better able to resolve conflict between competing attentional demands – as shown by a higher executive function score – benefit more when they are told what perceptual aspects should be prioritized. The models did not include a significant interaction between explicit direction and either of the other two ANT scores (Alerting and Orienting), which suggests that any impact these two attentional networks may have on performance is independent of the directions the participant was given.

A similar result was revealed when an LME model was fit to describe how participants performed during the learning trials. The best-fitting model included a significant three-way interaction between executive function scores, explicit direction, and block in the learning trials. This finding suggests that individuals with better executive function scores in the explicit direction group also had a faster rate of learning over time. This implied that individuals who can better direct their attention will learn faster if told where to focus that attention. In this model, there were no other three-way interactions, nor was there a significant two-way interaction between explicit direction and executive function score alone. The lack of a significant two-way interaction between directions and Executive Function in this model indicated that any benefit provided by the combination of high executive function and explicit direction took place over time as the learning process continued. This result suggested that the group of participants who performed the best at the end of the learning phase still required time to learn the novel contrast, and that they were not necessarily better at perceiving these phoneme contrasts without training.

Attention and the Learning of Phoneme Contrasts

It is also possible that some individuals are simply better at learning new phoneme contrasts than others. These individuals would not necessarily need additional instructions to direct their attention, but they would be able to learn the contrast regardless. This idea led to the creation of the current study's second research question, considering if any specific attentional networks might have a significant impact on the learning of the phoneme contrast. The GLM and LME models created to answer the first research question were used to evaluate this question as well. The answer to the study's second question is found within the main effects of individual ANT subscores.

The GLM models that were fit to predict participants' posttest scores revealed a significant main effect of executive function scores, meaning that participants with better Executive Function scored higher on the posttest regardless of experimental group. Additionally, the mixed effects model fit to describe participants' learning over time showed a significant interaction between executive function scores and learning block, meaning that participants with higher executive function scores learned the contrast more quickly than participants with lower executive function scores. These results suggested that individuals with better executive function scores were better able to learn a novel phoneme contrast. Since Executive Function can be seen as the ability to resolve conflict between competing demands on attention, the results from both sets of models suggested that the learning process was enhanced by the innate ability to prioritize key elements of the task at hand, regardless of external direction.

In addition to Executive Function, a main effect of Alerting was predicted. It was originally anticipated that individuals who had better alerting scores would be more able to attend to changes in stimuli and would learn the novel contrast better than individuals with lower alerting scores. This was not the case. There were no significant effects of alerting scores on either posttest scores or performance during learning trials. Executive function scores were the only predictor of novel contrast learning.

There are a number of potential explanations for these results. As mentioned previously, the literature is divided on the strength of the correlations between perception, statistical learning capability, and attention. For instance, while some research has supported the claim that attentional capacities are correlated with perception and learning (Darcy et al., 2016), other projects have failed to find such correlations (Darcy et al., 2015). This study's results, along with the ambiguity in the existing literature, may suggest that a more focused definition of "attention" is necessary. For instance, many studies that find significant influences of attention on phonological contrast learning appear to have definitions of attention that are more in line with the ANT's Executive Function subcategory (Darcy et al., 2016; Strange, 2011). In these definitions, attention is depicted as a control factor between how different information is processed, such as selecting what features of a contrast are important (Strange, 2011) or switching between target and distractor information (Darcy et al., 2016). Some research does suggest that processes that redistribute attention to salient or easily perceptible cues, similar to the ANT's Alerting subcategory, are important for speech perception in difficult conditions (Kong & Edwards, 2016). However, perhaps these findings only refer to conditions where the difficulty is external to the stimulus (i.e., a noisy environment), rather than difficulty caused by the stimulus itself.

However, the conflicting results from the current experiment with existing studies might also be due to imprecision in the measurement used, rather than a problem with the construct's definition. In other words, it is possible that the error variance of the ANT itself has created this issue. The current study relied on the traditional scoring method of the ANT. Although there is strong theoretical and experimental support for this method, there also some concerns that this type of scoring could produce less reliable results. Many of these concerns are based on the fact that all ANT subscores are calculated as difference scores, which have lower reliability than scores calculated as single values (Ishigami & Klein, 2010). Another issue with the traditional scoring method of the ANT is that it does not allow for interactions between subscores. According to the ANT's theoretical basis, all three of the attentional networks are considered anatomically independent, and the scores calculated for each of the networks are intended to be independent from each other (Fan et al., 2002). However, there is some evidence that orienting scores may be statistically correlated with scores from the Alerting network or the Executive Function network, implying that the Orienting network may not be completely separable from the other two attentional functions (Ishigami & Klein, 2010). Because the ANT does not allow for these interactions between networks, the reliability of models based on these scores could be decreased.

Explicit Direction and Contrast Learning

Finally, existing literature has suggested that providing participants explicit instructions on what to attend to could lead to better performance in statistical learning paradigms (Goudbeek et al., 2007; Sakai & Moorman, 2018). This idea led to the third and final research question, exploring whether giving participants specific instructions on what to listen to in the stimuli would boost their scores. The main effect of experimental group in the previously mentioned GLM and LME models was used to answer this question.

For the current study, the answer to this final question appears to be *no*. In the models that were fit to predict posttest scores and performance during learning trials, there were no significant effects of explicit instructions on either outcome variable. This result was unexpected, not only because it differs from previous findings on this topic, but also because anecdotal reports from pilot testing suggested that listeners in the general procedure condition took more trials to determine "what to listen for" prior to learning the contrast of interest.

However, the current study's findings should not necessarily be taken as contradicting prior research for a few reasons. For one, in all of the models constructed in the current study, the provision of explicit instruction did result in significant interaction terms with executive function scores. This implies that explicit instruction may be more effective in boosting the performance of individuals with higher executive function scores, who are already more likely to succeed, rather than boosting the performance of the population as a whole. These results suggest that participants may need to have a certain proficiency in resolving competing attentional demands, as demonstrated by a higher executive function score, in order to benefit from an external direction of attention in a learning task.

Another explanation for this null result might be that the experimental manipulation was not strong enough to truly capture the effect of explicitly directing attention. In the current study, the difference between the explicit instructions and general procedure was only a few words ("Listen carefully for the different consonant sound in each word" versus "Listen carefully to each word"). Since all stimuli were VCV syllables rather than real words, listeners may have had fewer options about what part of the word to focus on, even in the general procedure condition. This could have been an experimental design factor that weakened the impact of the manipulation in the current study.

Strengths, Limitations, and Future Directions

There were a few notable strengths to the current study. First, the use of online, asynchronous experimental software allowed for the recruitment of a potentially more diverse group of participants than the typical "undergraduate CSD/Psychology student" recruited for many studies on attention or speech perception. Because not all participants were necessarily college-educated, the findings of this study may be more likely to generalize to a wider population. The use of both tonal and speech-in-noise hearing screenings is also a strength. Speech-in-noise tests are better than traditional tonal hearing screenings at screening out participants with hidden hearing loss (Liberman et al., 2016), which allows greater confidence that all participants met the minimum criteria for auditory perception.

The strength of the experimental manipulation is one of a few limitations of the current study, but every limitation offers opportunities for continued research. Future studies could include a larger difference between experimental groups, including perhaps a brief introduction to articulatory or phonetic cues most important to the contrast to be learned (Sakai & Moorman, 2018) in the experimental group and no directions regarding attention to the stimuli in the control group. The distinction between groups may also be clearer if the task were more difficult, such as if the contrast of interest was embedded into more complex minimal pairs and real words rather than nonsense VCV syllables, or if the contrast of interest was a vowel sound rather than a more consistent consonant (Pisoni, 1973).

Issues regarding measurement with the ANT could be resolved in future studies by using a scoring scheme that allows for interactions between subscores. The traditional scoring method for the ANT considers the impact of cue and flanker conditions separately, but a more robust method of calculation could allow for the interaction of the attentional networks by comparing various combinations of cue and flanker conditions (see Wang et al., 2014). Alternatively, future studies could use a version of the ANT specifically designed to handle interactions between subscores, the Attention Network Test for Interactions (ANT-I; Callejas et al., 2004). In the traditional ANT, scores for the Alerting and Orienting networks are both calculated using different conditions of the same variable: the presence, absence, and location of an asterisk. The ANT-I uses an auditory warning signal as the variable used to calculate scores for the Alerting network, reserving the asterisk for the Orienting network alone. This provides more independent measures of the Alerting and Orienting networks and allows researchers to measure the degree of interaction between all three networks (Ishigami & Klein, 2010). The ANT-I was originally not used in the current study due to research suggesting that the ANT provided higher reliability on the Alerting and Executive Function networks, which were the networks of interest in this study

(ANT: MacLeod et al., 2010; ANT-I: Ishigami et al., 2016). However, the results of the current study suggest that it may be worthwhile to attempt similar research with the ANT-I instead.

Aside from addressing some of the limitations of the current study, there are a few novel directions for research in this area. Although studies have shown that perceptual capability is strongly correlated with the ability to learn new phonetic contrasts (Lin et al., 2017), the current study did not measure perceptual capability beyond simple hearing screenings. Although studies suggest that Executive Function is not correlated with perception (Siegelman & Frost, 2015), it is quite possible that underlying perceptual ability is an important covariate to the other aspects of attention measured in this experiment, such as Alerting or Orienting. The addition of a measure of auditory perception could reveal underlying patterns between perception, attention, and learning mechanisms that are not addressed in the current study.

Another interesting direction for future research may be the role of the attentional networks in the maintenance of a non-native phoneme contrast over time. There is evidence for the importance of explicit instruction in supporting category maintenance in a statistical learning paradigm (Goudbeek et al., 2007). Individuals who are given explicit instructions prior to a statistical learning task are more likely to retain the learned category boundaries after a maintenance period than individuals who are not. Although the current study only looked at the participants' performance in one session, it is possible that a longitudinal study with multiple follow-up tests could reveal additional interactions between attention and the role of instruction in phonological contrast learning.

A final interesting direction for future research might be to explore the differences in the role of attention in phoneme contrast learning across different populations. Differences in the processes underlying statistical learning in adults and children have already been discovered in previous research. In these cases, adults were more likely to rely on the patterns they have learned across an entire lexicon and children more likely to rely only on the patterns in a given speech stream (Pierrehumbert, 2003). Children and infants are known to learn the phoneme categories of their native languages without explicit instruction, whereas adults have a harder time learning certain types of new contrasts because of the influence of their native language (Best, 1995). However, it remains to be seen if these different statistical learning processes produce different end results when it comes to non-native phoneme category formation, especially when the role of attention is taken into account.

Ultimately, the results of this study add to the body of evidence regarding how adults can learn to perceive novel phoneme contrasts. Individuals who are better able to control their attentional focus can benefit from explicit instructions on where to direct that attention, learning the target contrast more quickly and performing better on perceptual tasks. On the other hand, just because an individual has a high alerting score and is aware of changes in their surroundings does not necessarily mean that they will be able to use that alerting ability to learn new patterns.

These findings challenge the idea that statistical learning of novel speech sounds is an entirely implicit process. Instead, the results of this study indicate that individuals who can focus their attention on the contrast to be learned perform better than individuals relying solely on passive exposure. This information is particularly applicable to individuals working in accent reduction or adult second language learning contexts. Outcomes for individuals who have strong executive function abilities will likely be higher than for individuals who do not. Additionally, for individuals with better attentional control, instructors should explicitly direct their pupils' attention to the target speech sound at the beginning of instruction in order to maximize their learning success.

References

Anwyl-Irvine, A.L., Massonié J., Flitton, A., Kirkham, N.Z., Evershed, J.K. (2019). Gorilla in our midst: an online behavioural experiment builder. *Behavior Research Methods, 52*(1), 388–407. doi: https://doi.org/10.3758/s13428-019-01237-x

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*(3), 170-176.

Beddor, P. S., & Gottfried, T. L. (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research* (pp. 207-232). Timonium, MD: York Press.

Boersma, P. & Weenick, D. (2018). *Praat: doing phonetics by computer [Computer program].* Version 6.0.37, retrieved from http://www.praat.org/.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research* (pp. 171-204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.S. Bohn, M. Munro (Eds). *Second-Language Speech Learning: The Role of Language Experience in Speech Perception and Production. A Festschrift in Honour of James E. Flege.* (pp. 13–34). John Benjamins; Amsterdam.

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage*, *19*(1), 113-124.

Callejas, A., Lupiánez, J., & Tudela, P. (2004). The three attentional networks: On their independence and interactions. *Brain and Cognition*, *54*(3), 225-227.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114*(3), 356-371.

Curtin, S., Goad, H., & Pater, J. V. (1998). Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, *14*(4), 389-405.

Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, *66*(4), 741-773.

Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences, 40*(1), 63-72.

Dobel, C., Lagemann, L., & Zwitserlood, P. (2009). Non-native phonemes in adult word learning: evidence from the N400m. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 3697-3709.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, *145*(5), 508-535.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*(3), 340–347.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review, 116*(4), 752-782.

Flege, J. E. (1980). Phonetic approximation in second language acquisition 1. *Language Learning*, *30*(1), 117-134.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America*, *89*(1), 395-411.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, *93*(3), 1589-1608.

Flege, J. E. (1995a). Second language speech learning: Theory, findings, and problems. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics, 16*(4), 425-442.

Fuhrmeister, P., Schlemmer, B., & Myers, E. B. (2020). Adults Show Initial Advantages Over Children in Learning Difficult Nonnative Speech Sounds. *Journal of Speech, Language, and Hearing Research, 63*(8), 2667-2679.

Gabay, Y., Dick, F. K., Zevin, J. D., & Holt, L. L. (2015). Incidental auditory category learning. *Journal of Experimental Psychology: Human Perception and Performance, 41*(4), 1124-1138.

Galvao-Carmona, A., González-Rosa, J. J., Hidalgo-Muñoz, A. R., Páramo, D., Benítez, M. L., Izquierdo, G., & Vázquez-Marrufo, M. (2014). Disentangling the attention network test: behavioral, event related potentials, and neural source analyses. *Frontiers in Human Neuroscience*, *8,* 813. doi: 10.3389/fnhum.2014.00813

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178-200.

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language*, *109*(2-3), 55-67.

Goudbeek, M., Swingley, D., & Kluender, K. R. (2007). The limits of multidimensional category learning. In *Proceedings of Interspeech 2007* (pp. 2325-2328). Causal Productions: Belgium.

Hancin-Bhatt, B. (2008). Second language phonology in optimality theory. In J. Edwards & M. Zampini (Eds.) *Phonology and Second Language Acquisition* (pp. 117-146). John Benjamins Publishing Company.

Idemaru, K. & Holt, L. L. (2020). Generalization of dimension-based statistical learning. *Attention, Perception, & Psychophysics, 82*(1), 1-19.

Ishigami, Y., Eskes, G. A., Tyndall, A. V., Longman, R. S., Drogos, L. L., & Poulin, M. J. (2016). The Attention Network Test-Interaction (ANT-I): reliability and validity in healthy older adults. *Experimental Brain Research*, *234*(3), 815-827.

Ishigami, Y., & Klein, R. M. (2010). Repeated measurement of the components of attention using two versions of the Attention Network Test (ANT): Stability, isolability, robustness, and reliability. *Journal of Neuroscience Methods*, *190*(1), 117-128.

Iverson, P & Kuhl, P. K. (1995). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling". *Journal of the Acoustical Society of America, 97*(1), 553-562.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance, 43*(9), 1594-1611.

Keating, P. A., Mikoś, M. J., & Ganong III, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *The Journal of the Acoustical Society of America*, *70*(5), 1261-1271.

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*(1), 40-57.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perceptual Psychophysics, 50*(2), 93-107.

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., & Maison, S. F. (2016). Toward a differential diagnosis of hidden hearing loss in humans. *PloS One*, *11*(9), e0162726. doi: 10.1371/journal.pone.0162726

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358-368.

Lim, S. J. & Holt, L. L. (2011). Learning foreign sounds in an Alien World: Videogame training improves non-native speech categorization. *Cognitive science*, *35*(7), 1390-1405.

Lin, Y., Fan, R., & Mo, L. (2017). Differences in phonetic discrimination stem from differences in psychoacoustic abilities in learning the sounds of a second language: Evidence from ERP research. *PloS one*, *12*(11), e0187135**.** doi.org/10.1371/journal.pone.0187135

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242-1255.

Lotto, A., K. Kluender & L. Holt (2000). Effects of language experience on organization of vowel sounds. In M. Broe & J. Pierrehumbert (eds.), *Laboratory Phonology V: Language Acquisition and the Lexicon* (pp.219-228). Cambridge: Cambridge University Press.

Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*(9), 1183-1198.

MacLeod, J. W., Lawrence, M. A., McConnell, M. M., Eskes, G. A., Klein, R. M., & Shore, D. I. (2010). Appraising the ANT: Psychometric and theoretical considerations of the Attention Network Test. *Neuropsychology, 24*(5), 637-651.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science, 12*(3), 369-378.

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology, 54*(8), 1472-1491.

Mirman, D., McClelland, J. L., Holt, L. L., & Magnuson, J. S. (2008). Effects of attention on the strength of lexical influences on speech perception: Behavioral experiments and computational mechanisms. *Cognitive Science*, *32*(2), 398-417.

Myers, E., Johns, A. R., Earle, F. S., & Xie, X. (2017). The Invariance Problem in the Acquisition of Non-Native Phonetic Contrasts: From Instances to Categories. *The Speech Processing Lexicon: Neurocognitive and Behavioural Approaches*, *22*(1), 52-84.

Nearey, T. M. (1997). Speech perception as pattern recognition. *The Journal of the Acoustical Society of America*, *101*(6), 3241-3254.

Peperkamp, S., Le Calvez, R., Nadal, J. P., & Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition*, *101*(3), B31-B41.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461-472.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2-3), 115-154.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*(2), 253-260.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America*, *89*(6), 2961-2977.

RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Rugesæter, K. P. N. (2014). Difficult contrasts: an analysis of phonemic distinctions in the English of young Norwegian learners seen against the backdrop of incidental foreign language learning. *Acta Didactica Norge-tidsskrift for fagdidaktisk forsknings-og utviklingsarbeid i Norge, 8*(1), 10. doi.org/10.5617/adno.1102.

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187-224.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*(1), 105-120.

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods, 49*(2), 418-432.

Silbert, N. H. (2012). Syllable structure and integration of voicing and manner of articulation information in labial consonant identification. *The Journal of the Acoustical Society of America*, *131*(5), 4076-4086.

Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *Journal of Phonetics*, *50*(1), 99-119.

Stephens, J. D. W. & Holt, L. (2010). Learning to use an artificial visual cue in speech identification. *Journal of the Acoustical Society of America, 128* (4). 2138-2149.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of phonetics*, *39*(4), 456-466.

Strange, W. & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. *Phonology and Second Language Acquisition*, *36*(1), 153-192.

Tees, R. C. & Werker, J. F. (1984). Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *38*(4), 579-590.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792-814.

Tuller, B., Jantzen, M. G., & Jirsa, V. K. (2008). A dynamical approach to speech categorization: Two routes to learning. *New Ideas in Psychology*, *26*(2), 208-226.

Wang, Y. F., Cui, Q., Liu, F., Huo, Y. J., Lu, F. M., Chen, H., & Chen, H. F. (2014). A new method for computing attention network scores and relationships between attention networks. *PLoS One*, *9*(3), e89733. doi.org/10.1371/journal.pone.0089733

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, *75*(6), 1866-1878.

Wickens, C.D., *Engineering Psychology and Human Performance*, Harper Collins, New York, 1992.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception & Psychophysics*, *26*(2), 95-104.

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064-2072.

Appendices

Appendix A

Language Screener Questions

1. Are you a native speaker of English (you grew up learning English from birth)?

   a. Yes - continue

   b. No - reject

2. Did your caregivers (parents, grandparents, nanny, teachers, etc.) speak a language other than English to you when you were a child?

   a. Yes - reject

   b. No - continue

3. Are you comfortable holding full, spontaneous conversations in any language aside from English?

   a. Yes - reject

   b. No - continue

4. Do you have any friends/relatives/co-workers who are frequently speak Thai, Hindu, or Punjabi in your presence?

   a. Yes - reject

   b. No - continue

5. Within the past year, did you spend more than 1 month in an area where the primary language was not English?

a. Yes - reject

b. No - continue

6. Have you studied any languages aside from English? If so, for how long, and in what contexts?

a. Yes (written text) - save data for later analysis and accept

b. No - accept

Appendix B

Informed Consent Forms

Informed Consent to Participate in Research Involving Minimal Risk

Study Title: The role of attention in adult L2 phonological contrast learning

USF IRB Study #760

Overview:

You are being asked to take part in a research study. The information in this document should help you to decide if you would like to participate. The sections in this Overview provide the basic information about the study.

Study Staff: This study is being led by Laura Conover, who is a doctoral candidate in the University of South Florida's Department of Communication Sciences and Disorders. This person is called the Principal Investigator. She is being guided in this research by Dr. Ruth Bahr. Other approved research staff may act on behalf of the Principal Investigator.

Study Details: The purpose of the study is to explore how adults learn to perceive variations between different speech sounds. The study consists of a few screening tasks, a 20-minute test of attention, and a 30-minute speech listening task.

Subjects: You are being asked to take part because you are an adult, monolingual English speaker with normal hearing and no history of speech or language difficulties.

Voluntary Participation: Your participation is voluntary. You do not have to participate and may stop your participation at any time. There will be no penalties or loss of benefits or opportunities if you do not participate or decide to stop once you start.

Benefits, Compensation, and Risk: We do not know if you will receive any benefit from your participation. There is no cost to participate. You will not be compensated for your participation.

This research is considered minimal risk. Minimal risk means that study risks are the same as the risks you face in daily life.

Confidentiality: Even if we publish the findings from this study, we will keep your study information private and confidential. Anyone with the authority to look at your records must keep them confidential.

Conflict of Interest Statement: There are no conflicts of interest to declare.

You can get the answers to your questions, concerns, or complaints. If you have any questions, concerns or complaints about this study, contact Laura Conover at lconover1@usf.edu. If you have questions about your rights, complaints, or issues as a person taking part in this study, call the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu.

You should only take part in this study if you want to volunteer. You should not feel that there is any pressure to take part in the study. You are free to participate in this research or withdraw at any time. There will be no penalty or loss of benefits you are entitled to receive if you stop taking part in this study.

I have thoroughly read and understand the consent form presented to me. I freely give my consent to take part in this study. I understand that by marking this box, I am agreeing to take part in research.

Appendix C

Statistical Learning Task On-screen Instructions

For this part of the experiment, you will hear a series of nonsense words. These nonsense words can be divided into two groups. Your task will be to learn which words are associated with each group. On every trial, you will hear one of these nonsense words. Please match the word you hear to one of the images on the screen, which are designed to represent the two groups. You will press the “f” key for the image on the left and the “j” key for the image on the right.

(Next)

At first, this will be a trial-and-error task as you learn which nonsense words are grouped together and associated with each image. There are approximately 500 trials for the whole experiment, and you are not expected to get them all right. In fact, we expect you only to get about half of the trials right at first! We'll start with a short pretest, which is designed to establish your baseline performance. After the pretest, there will be a learning period in which you will receive feedback after each trial. At the end, there will be a final test to evaluate how much you learned. Press the "Start" button when you're ready to begin.

(Pretest)

Next is the learning phase, which consists of four blocks. Your task is to select the correct image on the screen based on the nonsense word you hear. Press the ‘f’ key for the image on the left, and the ‘j’ key for the image on the right. For this section, you will receive feedback regarding how accurately you assigned each nonsense word to a group. The screen will flash a green checkmark if you guessed correctly, or a red X if you were incorrect.

(Next)

(Group A): Listen carefully for the different consonant sound in each word, and remember, you should not expect to get every trial right at first! Click the “Start” button to begin.

(Group B): Listen carefully to each word, and remember, you should not expect to get every trial right at first! Click the “Start” button to begin.

(Learning trials)

For this final test, your task is to select the correct image on the screen based on the nonsense word you hear by pressing the ‘f’ key for the image on the left, and the ‘j’ key for the image on the right, but you will no longer be given feedback on your accuracy. When you are ready to begin the post-test, press the "Start" button below.

Appendix D

IRB Determination

UNIVERSITY of SOUTH FLORIDA

EXEMPT DETERMINATION

April 21, 2020

Laura Conover
1406 Fernwood Place
Seffner, FL 33584

Dear Laura Conover:

On 4/20/2020, the IRB reviewed and approved the following protocol:

Application Type:	Initial Study
IRB ID:	STUDY000760
Review Type:	Exempt 3
Title:	The role of attention in adult L2 phonological contrast learning
Funding:	None
Protocol:	Protocol

The IRB determined that this protocol meets the criteria for exemption from IRB review.

In conducting this protocol, you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Please note, as per USF policy, once the exempt determination is made, the application is closed in BullsIRB. This does not limit your ability to conduct the research. Any proposed or anticipated change to the study design that was previously declared exempt from IRB oversight must be submitted to the IRB as a new study prior to initiation of the change. However, administrative changes, including changes in research personnel, do not warrant a modification or new application.

Ongoing IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit a new request to the IRB for a determination.

Sincerely,

Jennifer Walker
IRB Research Compliance Administrator

Institutional Review Boards / Research Integrity & Compliance
FWA No. 00001669
University of South Florida / 3702 Spectrum Blvd., Suite 165 / Tampa, FL 33612 / 813-974-5638

Page 1 of 1

Appendix E

ANT Example Figure Permissions

Andrea Louise Herbst <alherbst@mit.edu> Thu, Dec 19, 2019 at 11:11 AM
To: Laura Conover <lconover1@mail.usf.edu>
Hello Laura,
Thank you for your request. I am happy to grant non-exclusive permission to include Figure 1 from "Testing the efficiency and independence of attentional networks" in your thesis for University of South Florida (USF Scholar Commons) for academic non-commercial use only. Please include a credit line citing the material to its original MIT Press source:
Jin Fan, Bruce D. McCandliss, Tobias Sommer, Amir Raz and Michael I. Posner, 'Testing the Efficiency and Independence of Attentional Networks', Journal of Cognitive Neuroscience, 14:3 (April 1, 2002), pp. 340-347. © 2002 by the Massachusetts Institute of Technology. Published by the MIT Press.
Please let me know if you have any questions.
Best,
Andrea
Andrea Herbst
Journals Subsidiary Rights Associate
The MIT Press
One Rogers Street | Cambridge, MA 02142
617.258.0591 | alherbst@mit.edu