The Role of Individual Differences in Attention and Working-Memory in Learning a Nonnative Phoneme Category

by

Josh Topping

A thesis submitted to the Faculty of Graduate Studies and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Master of Arts

in

Psychology

Carleton University

Ottawa, Canada

Abstract

Speech perception is characterized by categorical perception of phonemes. Certain speech sounds varying along the continuous acoustic dimension known as voice onset time (VOT) are perceived as either voiced /b/ or unvoiced /p/ phonemes by English listeners. A third VOT prevoiced /p^h/ phoneme category is used in Thai and is indistinct from the /b/ category in English. Some listeners can learn to perceive speech sounds belonging this third VOT category with a small amount of training. The cognitive mechanisms underlying the variation in individuals' ability to perceive the prevoiced /p^h/ phoneme are not well understood. The current experiment investigated the role of attention and working-memory in facilitating listeners' ability to learn to perceive the prevoiced /p^h/ phoneme. A consistent relationship between attentional and working-memory measures and prevoiced perceptual learning was not found. Musical ability was a good predictor of performance on a prevoiced phoneme identification task.

Acknowledgements

I would like to thank John Logan, my supervisor, for his guidance and assistance throughout the completion of this thesis. Thank you to all my friends and family who were willing to be guinea pigs for my experiment. Thank you to Etelle Bourassa, Graduate Studies Administrator, for all your help, guidance, support, and encouragement throughout this challenging process. I would like to thank my committee members: Guy Lacroix and Kevin Cheung. Finally, I would like to thank my parents for their unconditional love and support.

Table of Contents

List of Tables

List of Figures

List of Appendices

The Role of Individual Differences in Attention and Working-Memory in Learning a Nonnative Phoneme Category

The ability to perceive and produce language is one of the most markedly human qualities that distinguishes us from other species. Our first language forms the auditory lens through which we selectively attend to the acoustic cues underlying speech. This selective bias toward the acoustic cues acquired through childhood language experience (i.e., the first language or L1) can affect the ability to perceive nonnative (i.e., second language or L2) phonetic distinctions (Kuhl & Iverson, 1995). The mechanism underlying L2 phoneme perception has been widely investigated from a variety of perspectives. The present study was designed to investigate individual differences in cognitive functioning that may affect L2 phoneme perception. The goal was to determine the extent to which individual differences in attention and working memory affect the ability of monolingual English listeners to learn to perceive an L2 phoneme category. The L2 phoneme category chosen for the present study was the prevoiced bilabial stop category, which is phonemically distinct in several southeast Asian languages. This category is related to the English phoneme categories /b/ and /p/; these phonemes vary along the acoustic dimension of voice-onset time (VOT). In the sections that follow, I will outline research on speech perception and its underlying cognitive mechanisms. I begin by providing an overview of some of the basic findings associated with speech perception, focusing on the perception of nonnative phonemes. Next, I will review research on individual differences in attention and working memory, and how they are related to the performance of cognitive tasks, including language. In the final section of the introduction I will describe the study and expected pattern of results.

Speech Perception

A defining characteristic of speech perception is categorical perception. Adult listeners perceive L1 speech stimuli that are evenly spaced apart on a physical continuum as grouped together into distinct phoneme categories. Moreover, their ability to discriminate between speech sounds from different categories exceeds their ability to discriminate sounds from within the same category, despite equivalent physical differences. Collectively, these characteristics define the phenomenon known as categorical perception (Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Pisoni & Tash, 1974).

Phoneme categories arise in part from characteristics of the human auditory system (see Stevens & Blumstein, 1975). However, differences in phonemic categories and category boundaries between languages suggest that the environment also plays a role in shaping the categorical perception of speech. The Quantal Theory developed by Stevens and Blumstein posits that speech sounds that are easily and reliably produced tend to be more common among the various languages of the world. This theory accounts for an early perceptual advantage for particular kinds of acoustic contrasts. Nevertheless, invariant properties of both acoustic stimuli and the human auditory system cannot account for all cross-language differences in perceptual sensitivity for speech sounds (Swan & Myers, 2013). For example, phonetic boundaries along the same acoustic continuum vary between languages (Lisker & Abramson, 1964). Adults are highly skilled at perceiving phonetic differences corresponding to phonemic contrasts in their native language but will often have difficulty perceiving phonetic differences that are not phonemic contrasts in their native language (e.g., Werker , Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984)).

Despite having a perceptual bias toward perceiving phonetic differences based on the phonemic categories of a listener's native language, there is considerable evidence that listeners remain perceptually sensitive to sub-phonemic variation under certain task conditions (Pisoni & Tash, 1974; Werker & Logan, 1985; Polka, 1992; Werker, 1994; Hayes, 2002; McMurray, Tanehaus, & Aslin, 2002). For example, although adults have a reduced sensitivity to nonnative phonemic contrasts, they maintain a limited ability to discriminate among them (Hayes-Harb, 2007). According to Pisoni and Tash (1974), phonetic information is available to listeners within phonemic categories, but retrieval of this information depends on the level of processing used to attend to the speech sounds. Such findings suggest that perception of category boundaries between speech sounds and the ability to perceive within-category differences is based on a combination of invariant acoustic properties and universal properties of the human auditory system, as well as on learned features of the language system.

The way that discontinuous perception of phonetic categories arises as new phonetic categories are acquired is not well understood (Swan & Myers, 2013). Adults can improve their perception of novel phoneme contrasts by mere exposure to a second language, yet little is known about how this change in perceptual sensitivity is accomplished (Hayes-Harb, 2007). Researchers have suggested that discontinuous perception of phonetic categories is the result of both bottom-up information such as the statistical distribution of speech sounds providing cues to novel phoneme categorization (e.g., Hayes-Harb, 2007; Kuhl et al., 1992; Maye, Werker, & Gerken, 2002), as well as explicit top-down information such as category labels, phoneme-grapheme correspondences, or minimal pairs (e.g., Francis & Nusbaum, 2002; Yeung & Werker, 2009). Many researchers have suggested that knowledge of the meaning of words can mediate a learner's ability to acquire phonemic contrasts (e.g., Jusczyk, 1985; Werker & Pegg, 1992; Lalonde & Werker, 1995). Minimal pairs also contribute to learning phonemes. Minimal pairs are sound strings that differ in one novel contrast only and have different lexical meanings, for example [rak] and [lak]. Hayes-Harb (2007) found that although adults can demonstrate evidence of perceptual learning based on statistical information alone, the additional availability of minimal pairs facilitates greater accuracy in identification of a novel speech contrast.

L2 learners are often able to discriminate between L2 phonemes in a laboratory or classroom setting, yet are unable to generalize this ability in a linguistically useful way. For example, Hayes-Harb (2007) found that participants who were able to discriminate between novel phoneme contrasts in a perception task were unable to encode the same contrasts in a word learning task. This parallels findings in language acquisition literature (see Aslin et al., 1998) suggesting that infants demonstrate the ability to discriminate sounds without having the ability to contrastively match the sounds to word meanings in lexical tasks. Further, L2 teachers frequently observe that adult learners can learn to perceive and produce novel L2 phonemic contrasts in classroom tasks while being unable to implement this ability in real world and lexical settings (Hayes-Harb, 2007). These findings suggest that there may be an intermediate processing stage between having a basic perceptual sensitivity to novel speech contrasts and being able to represent the contrasts in a lexically useful way.

Even in a highly controlled classroom or laboratory environment there are certain combinations of L1 language background and L2 phoneme categories that are particularly difficult for learners to discriminate. For example, findings from several studies suggest that Japanese-English bilingual listeners have difficulty discriminating between the English phonemes /l/ and /r/, even when they have considerable experience conversing in English (e.g. Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fijimura, 1975; Logan, Lively, & Pisoni, 1991; Yamada & Tohkura, 1992). Several factors that contribute to the discriminability of L2 phonemic contrasts include the degree of conflict between the phonemic inventories of L1 and L2, age of the learner, and amount of second language input.

According to Hayes-Harb (2007), the degree of conflict between L1 and L2 phonemic inventories is generally accepted as the principal determinant of a learner's ability to perceive L2 phonemes. This observation has been formalized in three models of nonnative speech perception: Best's Perceptual Assimilation Model (Best, 1994), Flege's Speech Learning Model (Flege, 1995), and Kuhl's Native Language Magnet Model (Kuhl, 1991). These models posit that native speech experience provides the organizational framework for a listener's perception of speech that shapes discrimination and identification of unfamiliar speech contrasts (Best, McRoberts, & Goodell, 2001). All three models presume that an adults' ability to discriminate L2 speech contrasts is systematically informed by their native speech system. These models hold that nonnative speech sounds belonging to categories that overlap or conflict with native phoneme categories are particularly difficult to perceive because listeners have a natural tendency to perceptually assimilate these sounds to L1 speech sound categories. Discrimination of these speech sounds require a listener to inhibit the influence of the native language perceptual system in order to distinguish these novel speech sounds from nearby or overlapping L1 phoneme categories.

Discrimination of L2 speech sounds belonging to phoneme categories that overlap L1 categories is inhibited by the listener's pre-existing phonemic system because a listener must inhibit the natural tendency to perceive the speech sounds as L1 phonemes. An L2 speech sound perceived as a poor example of an L1 phoneme is more easily discriminated from the L1 phoneme than an L2 sound that is more perceptually similar to the nearest L1 phoneme (Best et al., 2001). In other words, the degree of assimilation to an L1 phoneme can act as the basis for discrimination of an L2 speech sound. Japanese speakers, for example, are likely to poorly discriminate between the English phonemes /r/ and /l/ because they perceive both of them as poorly assimilating to either the Japanese phoneme /r/ or /w/ to the same extent (e.g. Best & Strange, 1992; Takagi & Mann, 1995; Yamada & Tokura, 1992). Moreover, although Japanese listeners can learn to discriminate between /r/ and /l/ (Logan et al., 1991), to do so requires approximately 15 hours of training. In contrast, Pisoni, Aslin, Percy, and Hennessy (1982) found that monolingual English speakers can learn to distinguish the previously phonemically indistinct Thai prevoiced bilabial stop consonant (/p^h/) from the English phoneme /b/ after a brief 10 minute training session. Even though /p^h/ is initially perceived as belonging to the category /b/ in English, listeners can learn to discriminate between /p^h/ and /b/ relatively efficiently because /p^h/ is a poor example of /b/. Attention to category goodness of fit may underlie this learning process. Further, in order for a listener to interpret goodness of fit, attention must be directed toward the acoustic property of the speech sound that differentiates /p^h/ from /b/. The current study involves a training procedure modelled after Pisoni et al. (1982) (discussed in detail below) which is designed to direct the listener's attention to an unattended region of voice onset time (VOT)—the acoustic dimension that differentiates /p^h/ and /b/. Both early and late L2 learners are able to gain perceptual access to features of L2 speech sounds that are not used to contrast L1 phonemes (e.g., Schirru, & MacKay, 2003; Flege & MacKay, 2004). Adults can learn to detect these differences and store this information in durable long-term memory representations (Flege, 1987; Flege & Hammond, 1982). Individuals can learn to identify L2 phonemes with categorical-like accuracy with relatively little training when the training is designed in a way that facilitates attending to the relevant stimulus property (e.g., Pisoni et al., 1982).

The training procedure used by Pisoni et al. (1982) is based on the relationship between native language speech sounds and prototype theory. The underlying perceptual organization of phonetic categories is consistent with the prototype theory (see Rosch, 1975; 1978; Posner & Keele, 1968), which holds that categorical prototypes have a unique perceptual status (Kuhl & Iverson, 1995). Good instances of categories (i.e., those exemplars that are most similar to the category prototypes) are more readily classifiable and more quickly recalled than less exemplary instances (e.g., Mervis & Rosch, 1981). For example, a robin is more easily recognized as belonging to the category of birds than an ostrich, and a hammer is more exemplary of the tool category than an awl. Similarly, adult listeners of a particular language are highly skilled at identifying exemplars of phonetic categories in their native language (Kuhl & Iverson, 1995). The training procedure by Pisoni et al. (1982) uses category exemplars and response feedback to engage listeners and focus their attention toward the differences between the speech sounds used in their experiment. Listeners are presented with best instances of /p^h/ (VOT = -70 ms), /b/ (VOT = 0 ms), and /p/ (VOT = 70 ms) and required to identify each stimulus. Response feedback indicates whether the listener has correctly or incorrectly identified the phoneme category for each training trial. The difference between the prevoiced and voiced categories is emphasized by only using category exemplars, rather than the entire continuum of 15 stimuli ranging from -70 to 70 ms VOT. Response and feedback facilitate learning by requiring listeners to actively engage with the stimuli. The purpose of this training procedure is to focus the listeners' attention toward the acoustic differences between /p^h/ and /b/ in order to facilitate development of a durable representation of the prevoiced bilabial stop consonant that is perceptually distinct from the voiced phoneme category. The current experiment uses the same training procedure because it has been shown to be successful in teaching naive monolingual English listeners to identify the novel prevoiced bilabial stop consonant with a level of accuracy suggestive of categorical perception.

Kuhl and Iverson (1995) state that the category prototypes defined by a listener's native language structure do not alter the ability to discriminate L2 speech contrasts at a sensory level (e.g., Logan, Lively, & Pisoni, 1991; MacKain, Best, & Strange, 1981), but they do alter the perceptual system underlying speech perception in a way that reduces the prominence of certain phonetic distinctions when compared to the language-general initial state that humans are born with. This change is thought to occur at a higher level of processing involving memory and/or attention.

Working Memory

The relationship between unfamiliar L2 speech sounds and established L1 phonemic categories is a critical factor in determining an individual's ability to learn speech sounds from a nonnative language (Hayes-Harb, 2007). The framework of categorical perception has become an indispensable component facilitating investigations of L2 speech perception both from a developmental perspective (e.g., Aslin, Pisoni, Hennessey, & Perey, 1981) and from an adult perspective (e.g., Pisoni, Aslin, Perey, &

Hennessey, 1982). Accounts of how acoustic auditory information is transformed into more durable phonetic code (e.g., Pisoni, 1973) have enabled the integration of the categorical speech perception framework into mainstream psychology. Despite these advances, little research has addressed the role of attention and working memory limitations in this process (Schoenherr & Logan, 2014). The present study included measures of attention and working memory in order to explore the potential relationship between these cognitive processes and the learning process involved in developing a durable representation of a novel phoneme.

Working memory is a system of domain-specific stores or formats for temporarily holding information along with an executive attention or domain-general supervisory mechanism (Engle, 2010). Engle (2010) suggests that there may be a few dozen domain-specific stores whose contents can be thought of as temporarily activated representations of long-term memory information, acting as “pointers” linking information to existing representations.

The phonological loop is a domain-specific store that is relevant to spoken language, and is suspected of playing an important role in language development (Enlgle, 2010). It is specialized for the retention of verbal information for short periods of time (Baddeley, Gathercole, & Papagno, 1998). It is comprised of two components, a phonological store responsible for holding information in its phonological form, as well as a rehearsal process responsible for preserving phonological representations in the phonological store. The phonological loop is manifested in the excellent ability of humans to repeat what they hear, especially when they hear strings of speech sounds from their native language. Butterworth, Campbell, and Howard (1986) have argued that the phonological loop is an aspect of short-term memory that is not actively involved in the process of working memory. It has been shown that many individuals with specific deficits in short-term phonological memory do not have any serious difficulties performing everyday cognitive tasks (Baddeley et al., 1998). Individuals with serious deficits in phonological loop capacity have typically demonstrated normal speech production and language comprehension abilities (Shallice & Butterworth, 1977; Vallar & Shallice, 1990). In contrast, Baddeley et al. (1998) suggest that the function of the phonological loop is not to remember familiar words, but to aid in learning new words. From this point of view, the ability to repeat a string of familiar digits or words is an evolutionary byproduct of the more fundamental human capacity to generate more durable representations of brief and novel phonological material—our ability to learn new words. In the context of the present investigation, it is not clear what role the phonological loop might play. This will be considered in a later section.

Engle (2010) found evidence suggesting that individuals differ in practice-developed skill for various coding formats of the domain-specific working-memory stores. The vast majority of research on individual differences in working-memory has utilized complex span tasks to measure working-memory capacity (WMC). Historically, the prototypical measures of short-term memory have been simple span tasks such as digit span. A simple span task presents the participant with a series of letters, words, or numbers, one at a time, and then asks the subject to recall the items in the correct order. In one of the most widely cited findings in cognitive psychology (Gorenflo & McConnell, 1991), Miller (1956) reported that the typical human recall capacity for a simple span task is approximately seven items, plus or minus two items. These simple tasks were eventually discarded as measures of short-term memory because they lacked reliability and were inconsistently valid (Engle, 2010). In contrast, complex span measures modelled after the reading span (RSPAN) task originally developed by Daneman and Carpenter (1980) have demonstrated moderate reliability and consistent validity (Engle & Kane, 2004) in predicting performance in a wide array of higher-level and real world cognitive tasks including listening comprehension, language comprehension, ability to follow oral and spatial directions, vocabulary learning, note taking ability, writing, reasoning, hypothesis generation, bridge playing, and complex task learning such as the ability to learn to write computer programs in a computer language (respectively: Daneman & Carpenter, 1983; Daneman & Merikle, 1996; King & Just; 1991; Engle, Carullo, & Collins, 1991; Daneman & Green, 1986; Kiewra & Benton, 1988; Benton, Kraft, Glover, & Plake; 1984; Barrouillet, 1996; Kyllonen & Christal, 1990; Dougherty & Hunter, 2003; Clarkson-Smith & Hartley, 1990; Kyllonen & Stephens, 1990). It is conceivable that the RSPAN task measures a cognitive construct that is involved in an individual's ability to transform a relatively unstable acoustic representations of a novel sound into a more durable representation, which would facilitate performance on a novel phoneme identification task. This is the rationale behind exploring the relationship between individual differences in the RSPAN task and individual differences in identification performance for the prevoiced bilabial stop consonant categorization in the current experiment.

A complex span task involves a series of easy yet nontrivial cognitive tasks presented alongside a series of items to be recalled. The cognitive tasks provide a distraction that interferes with an individual's ability to recall the series of items. This forces the reliance on higher-level cognitive resources for recall, compared to a simple span task, because the items must be preserved in memory while attending to the concurrent task. In the RSPAN task used by Unsworth, Heitz, Shrock, and Engle (2005) the participant is given a brief period to read a sentence followed by a letter. The participant is then required to make a judgment regarding whether the sentence makes sense. After three to seven of these trials have been presented, the participant is required to recall the letters that followed each of the sentences. Another complex span task is the operation span task (Unsworth et al., 2005), which consists of presenting subjects with a series of simple mathematical operations to perform, each followed with a single letter for recall after three to seven trials. Both of these complex span tasks, as well as several different spatial span tasks used by Engle et al. (2005) have accounted for similar variance in a wide array of higher-order cognitive tasks and have been strongly correlated with a construct that has been associated with fluid intelligence. The high degree of agreement between the various complex span tasks and their ability to predict higher-order cognitive abilities suggests that these tasks reflect a unitary psychological construct

(Engle, 2010). Studies associating differences in WMC with performance of higher-order cognitive tasks involving a degree of interference (e.g.,, Unsworth et al., 2005) but not with those with minimal interference (e.g., Kane & Engle, 2002), suggest that WMC is associated with attentional control. This idea was further supported by a study conducted by Kane et al. (2007) that investigated the relationship of WMC to mind wandering. Kane et al. (2007) found that individuals with low WMC were more likely to have their mind wander than high WMC individuals as cognitive tasks became more challenging, required more effort, or a higher amount of concentration. There is a degree of interference involved in attending to a novel speech sound at an acoustic, rather than phonemic, level of processing. The novel speech sound used in the present study is automatically perceived as a voiced bilabial stop consonant by monolingual English listeners. Therefore, the listeners must inhibit interference from the phonemic level of attending in order to perceive the acoustic differences between the prevoiced and voiced phonemes.

Engle (2010) argues that speech-based, along with visually- and spatially-based coding formats for working memory, require limited-capacity attentional control that functions under complex situations that involve distraction and interference. He suggests that the link between attention control and complex span tasks lies in the psychological mechanism involved in transferring relevant information between active and inactive memory. According to Engle (2010), the source of individual differences in working memory capacity tasks is the ability to control attention in a manner that enables functional access to representations of task-relevant information in either active memory or easily accessible inactive memory. This attentional ability is most relevant when an individual is engaged in a task involving interference from competing representations. For example, in a dichotic listening task, Conway, Cowan, and Bunting (2001) found that a significantly lower percentage (20%) of high WMC individuals reported hearing their name after it was presented through the unattended ear, compared to low WMC individuals of which 65% reported hearing their name. Similarly, Unsworth, Schrock, and Engle (2004) found that low WMC individuals were more likely to make unintentional errors during a task where individuals attempted to avoid the innate urge to direct their gaze toward a moving stimulus on a computer screen. In addition, they found that low WMC and high WMC individuals did not differ in ability to direct their gaze toward a moving stimulus when that was the objective. These studies, therefore, both suggest that individual differences in working memory capacity are highly associated with differences in attentional control―specifically with the ability to avoid directing attention toward distracting stimuli unrelated to the task at hand, even for very low-level attention tasks.

Attention

Posner and Rothbart (2007) describe attention as a fundamental set of mechanisms underlying our awareness of the world as well as our ability to intentionally regulate our thoughts and emotions. Functional neuroimaging data have indicated that attention is comprised of three networks, each responsible for different attentional processes (Fan, McCandliss, Fossella, Flombaum, & Posner, 2005). These networks consist of distinct anatomical areas in the brain and associated cognitive processes of alerting, orienting, and executive control of attention. Alerting is the process of achieving and maintaining a high degree of sensitivity to incoming stimuli; orienting refers to the process of selecting information from sensory input; and executive attention consists of the mechanisms involved in monitoring and resolving conflict among thoughts, feelings, and actions. Duncan et al. (2000) argue that the executive attention network is involved in self-regulating emotions as well as a variety of higher-order cognitive tasks underlying intelligence.

The attention network task (ANT) was developed by Posner and his colleagues in order to investigate individual differences in efficiency of the alerting, orienting, and executive attention networks (Fan, Flombaum, McCandliss, Thomas, & Posner, 2002). Figure 3 provides an illustration of the ANT experimental procedure. The ANT uses differences in reaction time (RT) between several conditions to measure the performance of each network in an individual. The task consists of simply identifying the direction of an arrow. The orienting and alerting responses consist of a cue presented shortly before the target indicating that the target will occur or the location where the target will occur, or both. A no-cue condition presents the target with no warning. The target occurs either above or below the point of fixation, and consists the central target arrow and flanking arrows surrounding the target that are either congruent or incongruent with the direction of the target. The efficiency of the executive attention network is measured by subtracting the RTs from the congruent condition from those of the incongruent condition. Subtracting RTs from the cue condition that does not inform the location of the target from the RTs from the no-cue conditions provides an indication of the efficiency of the alerting network. Finally, subtracting RTs obtained from the condition where target location is cued from the RTs from the trials where the cue is present but not spatially informative provides a measure of the efficiency of the orienting network. Fan et al. (2002) used the ANT to examine individual differences in efficiency of each attentional network and found no correlation among orienting, alerting, and executive scores among a sample representative of the normal population. The current experiment required the listener to 1) maintain a high degree of sensitivity to incoming speech sounds for an extended period of time, 2) attend toward the acoustic nature of the speech sound, and 3) retain attentional control in order to inhibit perceiving incoming auditory stimuli as speech while a new phoneme category is being learned. These attentional processes are expected to be associated with the alerting, orienting and executive control networks, respectively. Individual differences in any of the three dimensions of attention measured with the ANT may therefore be associated with individual differences in identification performance for prevoiced stimuli.

Working Memory, Attention, and L2 Speech Perception

Working memory capacity, specifically the efficiency of the phonological loop, has been implicated in the acquisition of native (e.g., Gathercole, Willis, Emslie, & Baddeley, 1992) and nonnative (e.g., Papagno, Valentine, & Baddeley, 1991) durable lexical representations. Pisoni and Geers (2000) suggest that spoken language processing and working memory are highly interrelated processes that share common reciprocal links and processing resources involved in speech perception and production, language comprehension, and reading. They found that children with cochlear implants who performed better on a digit span task also performed better on a series of language tasks. Pisoni and Greers were attempting to provide some explanation for the large discrepancy in development of language related abilities among children with cochlear implants. Analogously, if speech perception is closely associated with WMC, then it is probable that individual differences in WMC within the normal population are associated with variations in ability to acquire L2 phonemes. Further, Pisoni, Lively, and Logan (1994) suggest that selective attention is the psychological mechanism principally responsible for perceptual reorganization during perceptual learning of speech. Neuroanatomically based functional divisions in attentional capacities as measured by the ANT (Posner & Rothbar, 2007) may shed light on the degree to which the attentional networks are involved in the acquisition of L2 phonemic categories.

Redick and Engle (2006) found that individual differences in WMC are related to individual differences in performance on the executive control component of the ANT, but not to alerting or orienting. Although both the ANT and the RSPAN provide a measure of executive control of attention, which is arguably a major cognitive component involved in learning a novel phoneme, the advantage to administering both tasks rather than either the ANT or the RSPAN alone in the present study warrants discussion. As discussed earlier, the ANT provides a measure of individual efficiencies of three distinct attentional networks. The alerting network is associated with the ability to maintain a sensitivity to incoming stimuli in the environment, which may be associated with maintaining focus on the speech sounds used in the present experiment. The orienting network is associated with the ability to align attention with a source of sensory signals, and may therefore be associated with a listener's ability to focus on the specific auditory cues presented during the current experiment in a way that maximizes the level of stimulus information available to the listener for further processing after input. Finally, executive control of attention is associated with the ability to inhibit attention towards stimulus properties that are irrelevant to the task at hand. In the context of the present study, executive attention is conceivably involved in inhibiting attending to the prevoiced stimuli at the phonemic level so that the listener can focus attention on the acoustic nature of the stimuli during the process of new category formation. Executive control of attention is also associated with WMC, but WMC also involves the ability to maintain, access, and develop durable representations of novel information during the learning process. The RSPAN task will be administered in the current study in addition to the ANT because complex span tasks provide a measurement of a construct related to attending, updating, storing, and retrieving information from active memory stores. WMC is highly related to executive control of attention in a way that facilitates inhibiting attention towards irrelevant aspects of stimuli, but is also related to storage and retrieval. This active memory updating process may be related to an individual's ability to develop new phoneme categories while attending to the acoustic dimension of incoming speech sounds.

The present study focuses on identification of the phonemic status of a series of bilabial stop consonant syllables varied along a single acoustic dimension—the VOT continuum. The procedure is an adaptation of the experimental design used by Pisoni et al. (1982) and Schoenherr and Logan (2014).

Pisoni et al. (1982) provided strong evidence that the perceptual abilities underlying stop consonant identification are not permanently fixed by early linguistic experience, and that the perceptual abilities of monolingual speakers of English can be selectively modified. They found that, contrary to widely accepted views at the time, individuals could learn to identify phonemes that were not phonemically distinctive in their native language, specifically, the prevoiced bilabial stop consonant /p^h/. They presented participants with synthetic speech stimuli evenly spaced along an acoustic dimension (VOT) and representing three possible phonemes. Two of the three phoneme categories encompassed by the synthesized VOT series (see Lisker & Abramson, 1967) represented a contrast in English (/b/-/p/), whereas the third category /p^h/ encompassed speech sounds that are not phonemically distinct in English but are often produced in English, as free variants of /b/ in word-initial contexts and as allophonic variants of /b/ in intervocalic phonetic contexts (Strange & Dittmann, 1984). Therefore, according to Strange and Dittmann, participants had considerable experience with the phonetic categories being trained, although not as phonemically distinctive (functional) categories in their native language. The key to the success of Pisoni et al. involved modifying experimental conditions in order to reduce participants' uncertainties and strategically directing participants' attention toward acoustic, rather than phonetic properties of the stimuli.

Pisoni et al. (1982) conducted three experiments. The purpose of the first experiment was to determine whether a large group of naive participants could identify the three perceptual categories along the VOT continuum with no formal training or systematic feedback during identification. The second experiment was conducted in order to determine whether successful three-category labelling would modify subsequent discrimination performance. The stimuli used in their experiment consisted of 15 synthetic labial stop consonant-vowel (CV) syllables varied along the VOT continuum from -70 ms to +70 ms VOT. The experiment involved two one-hour sessions conducted on separate days. Prior to the experimental tasks, participants were presented with one exemplar representing each one of the three phoneme categories (-70, 0, and 70 ms VOT) ten times each in order to familiarize them with stimulus contrasts and correct responses. It is important to note that these 30 trials did not require any response from participants so no feedback was available to listeners. One group of participants completed a two and three category identification (ID) task on separate days, and another group completed the same identification tasks followed by an ABX discrimination test. For each trial of the identification task, participants were presented with a stimulus. They were then required to indicate which phoneme category the stimulus best represented (i.e., either /b/ or /p/

for the two category condition, and either /pʰ/, /b/, or /p/ for the three category condition). The ABX discrimination test included all 13 two-step pairs from the 15 stimuli (ie: -70/-50, -60/-40, -50/-30, etc.), with stimuli in the triads presented 500 ms apart. For each trial, participants were presented with a stimulus pair followed by one of the stimuli from the pair. They were then required to indicate whether the third stimulus was identical to the first or second stimulus. The ID task consisted of 150 trials each day, and the discrimination task consisted of 208 trials.

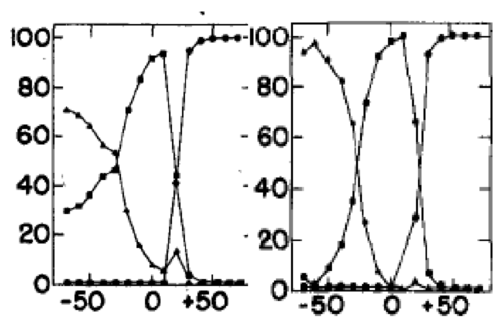


Figure 1. Results from Pisoni et al. (1982). Experiment 1 (left) includes full sample and Experiment 3 (right) includes only the half of participants who scored above 85% correct identification on the training procedure. ID functions for the percentage of VOT stimuli identified as /pʰ/, /b/, or /p/. VOT is shown on the x-axis and proportion of responses for each category on the y-axis.

All participants showed reliable and consistent two-category ID functions and in the three-category condition only two of the 45 participants failed to reliably identify stimuli belonging to the third phoneme category which was not phonemically distinctive in English (see Figure 1, left panel). Many participants showed two peaks in their ABX discrimination functions at boundaries separating the three voicing categories, suggesting that participants could perceive a novel prevoiced category. Peaks in discrimination functions closely corresponded to category boundaries in the ID functions providing further evidence that participants learned categorical perception of a nonnative phoneme category.

The correspondence between ID and discrimination functions arguably could have been a result of the ID task influencing performance on the ABX test. Pisoni et al. (1982) tested this possibility in Experiment 2 which involved an ABX task with no preceding ID task. Additionally, they tested the effect of response feedback on ABX performance. Pisoni et al. found that ABX functions from the no-feedback condition were similar to those obtained in Experiment I, demonstrating that the peaks in discrimination performance at category boundaries were not entirely the result of a prior ID task. The results from Experiment 2 suggested that some underlying sensory or psychophysical factor, rather than perceptual categorization in and of itself, is responsible for discontinuities in VOT discrimination performance typically found at specific regions of the acoustic continuum. The feedback condition did not significantly enhance discrimination performance compared to the no-feedback condition.

Experiment 3 in Pisoni et al. (1982) investigated whether an enhanced training regimen would improve discrimination performance and sharpen category boundaries indicated by ID functions. Pisoni et al. were also interested in reducing intersubject variability and response consistency which had been issues of minor concern in

Experiments 1 and 2. This was accomplished with a training procedure involving response feedback.

Experiment 3 involved four one-hour sessions occurring on consecutive days. Consistent with Experiments 1 and 2, participants were initially presented with the three category exemplars (-70, 0, and 70 ms VOT) to familiarize them with the stimuli and required responses. Next, a training procedure presented participants with the three salient exemplars of each perceptual category presented in random order. Participants were presented with each exemplar 80 times, for a total of 240 training trials. After each stimulus was presented, participants were required to identify which of the three phoneme categories it belonged to. After each response, feedback was provided indicating whether each response was correct or incorrect.

Day two consisted of a 75-trial training task identical to that of the previous day followed by an three-category ID task and an ABX discrimination task identical to those of Experiment 1. Participants who correctly identified at least 85% of the stimuli during the feedback training task on day 1 participated in the experiment on days 3 and 4. For these days, testing consisted of one block of 150 ID trials, and one block of 250 ABX discrimination trials. No response feedback was provided during the ID task, however, response feedback was provided after each correct response during the ABX test.

Pisoni et al.'s (1982) Experiment 3 results indicated, as in Experiment 1, that there was variability among individual participants. Six out of twelve participants performed the initial feedback training task with at least 85% accuracy, and these participants showed a high degree of consistency in labelling stimuli in the prevoiced category. As shown in the right panel of Figure 1, the average ID function was characterized by sharp slopes at both category boundaries, similar to the voiced-unvoiced category boundary obtained in the ID function of Experiment 1. This indicates the presence of three discrete and well-defined perceptual categories. Pisoni et al. also found that the slope for the group-identification function of the prevoiced category was steeper for Experiment 3 than for Experiment 1, suggesting that a training procedure using exemplars, participant responses, and response feedback, increases response consistency in perceptual categorization. I will be administering a training procedure identical to that used in Pisoni et al.'s Experiment 3 in the current experiment, due to the success of this method.

The overall results of Pisoni et al. (1982) indicated that when given the appropriate experimental procedures, naive listeners can learn to perceive differences in the prevoiced region of the VOT continuum. Therefore, these results demonstrate that the perceptual system of adults is capable of responding to a phoneme category that is not phonemically distinctive in the listener’s native language. These results indicated that perceptual reorganization of the system involved in interpreting auditory input at the phonemic level is possible, although the specific mechanisms underlying this kind of perceptual reorganization remain largely unexplored.

As discussed earlier, nonnative speech perception requires a degree of inhibition of the native language system in order to restrict inappropriate categorization of nonnative speech sounds. This is especially true during the perceptual learning process involved in developing novel phoneme categories. This type of inhibition requires several lower levels of attention in order to focus on the acoustic nature of the novel speech sounds, as well as executive control of attention in order to inhibit native language system from automatically categorizing nonnative speech sounds. This learning process also requires a degree of working-memory in order to facilitate the formation of durable representations of nonnative phoneme categories. In order to investigate the roles of attentional control and working-memory capacity in facilitating the process of learning to perceive a nonnative linguistic phoneme, Schoenherr and Logan (2014) conducted a study modelled after Pisoni et al. (1982). Additionally, Schoenherr and Logan assessed the roles of explicit and implicit sources of information during perceptual learning at the phonemic level by asking participants to provide confidence ratings after each identification response. Ashby, Alfonso-Reese, Turken, and Waldron (1998) suggest that both explicit and implicit cognitive systems are involved in categorical learning. Explicit systems focus attention on a small number of stimulus dimensions used when initially learning to categorize stimuli. As learning progresses, an implicit procedural-learning system that uses more multidimensional stimulus information becomes a more dominant system involved in categorizing stimuli. Few models of speech perception have incorporated categorization models involving explicit and implicit systems (but see Chandrasekaran, Koslov, & Maddox, 2014 for a recent attempt to develop such a model).

Schoenherr and Logan (2013) have argued that when confidence ratings exceed performance during an ID task, individuals' subjective awareness is focused on the phonemic level of incoming auditory information. By contrast, when individuals are required to focus on acoustic stimulus properties during a phoneme identification task such as learning to perceive the prevoiced bilabial stop consonant, they will exhibit lower confidence ratings. Schoenherr and Logan suggest that acoustic properties of speech sounds are implicit when performing an ID or discrimination task involving nonnative phonemes. They argue that divergent evidence for explicit and implicit representations of speech sounds can be obtained from measures of immediate memory and attention. Schoenherr and Logan (2014) assessed the relationship between individual measures of immediate memory and attention, explicit and implicit representations of the prevoiced VOT category, and ID performance using the laboratory training procedure employed by Pisoni et al. (1982), with the addition of confidence, attentional network, and immediate memory measures. They posited that implicit representations of speech sounds should be evidenced in ID performance, measures of orientation obtained from the attentional network task, and measures of immediate memory components requiring less attentional control.

Prior to the main experimental task, Schoenherr and Logan (2014) measured individual differences in immediate memory using a reading span (RSPAN) task adapted from Engle et al. (1999). They presented each participant with 42 sentences in sets ranging from 2 to 5 sentences (there were three cases of each set length, i.e., $2 \times 3 + 3 \times 3 + 4 \times 3 + 5 \times 3 = 42$ sentences). The order of each set length was randomized. Each sentence was presented visually and auditorily, and participants were instructed to read sentences aloud along with the auditory presentation. A random letter followed each sentence. Participants were required to make a judgment after each sentence-letter pair was presented regarding whether the sentence was semantically meaningful. For example, when presented with the sentence “The young pencil kept his eyes closed until he was told to look”, a participant should press the keyboard button associated with a “No” response to indicate that the sentence was not semantically correct. After each set of sentence-letter pairs participants were prompted to recall the set of letters that followed each sentence. The ability to correctly recall letter sets was assumed to reflect the executive function of immediate memory. Other measures of memory assumed to be associated with implicit representations requiring less attentional control consisted of simple letter span tasks. Ten sets of eight random letters were presented to participants who were required to recall each set either forwards or backwards immediately following the presentation of each letter set. Letter sets were presented either auditorily or visually.

The principal task in Schoenherr and Logan (2014) involved training with feedback, and four experimental blocks involving an ID task (two-category ID with confidence response, two-category ID without confidence response, three-category ID with confidence response, three-category ID without confidence response). The training component was identical to that used in Experiment 3 of Pisoni et al. (1982), except that each category exemplar was presented 60 times instead of 80 times. Each experimental block consisted of 10 repetitions of each of the 15 synthetic bilabial stop consonant stimuli ranging from -70 to +70 ms VOT, presented in random order. As in Pisoni et al., participants responded to each stimulus by indicating its categorical status as /p^h/, /b/, or /p/ in the three-category condition, and as /b/ or /p/ in the two-category condition. For the ID tasks with confidence reporting participants additionally indicated how confident they were with each category response by pressing keys representing a six point scale ranging from 50% to 100% certainty, in 10% increments.

Schoenherr and Logan's (2014) results provided further evidence that listeners are able to use acoustic properties of stimuli along a speech continuum in order to identify speech sounds belonging to a nonnative linguistic category that was previously not phonemically distinct. They argued that listeners became sensitized to specific regions of the VOT continuum characteristic of categorical perception. Confidence reporting, however, was not significantly related to ID performance in the 3-category condition. Schoenherr and Logan suggest that this could be taken as evidence that the process of perceptual reorganization involved in nonnative categorical perception occurs at a cognitive level below subjective awareness. Additionally, they suggest that this implies that the acquisition of nonnative speech sounds is not affected by the additional attentional requirements involved in performance monitoring. They also suggest that this implies that using acoustic properties to develop a nonnative contrast is primarily a function of implicit processes.

Schoenherr and Logan (2014) concluded by suggesting that immediate response feedback during training allocates attention to specific acoustic stimulus properties, as evidenced by correct identification of exemplars during the training task. This provides further rationale for my use of this training procedure in the present study. Further, they make the claim that attentional capacity appears to affect ID performance by determining the extent that either acoustic or phonemic stimulus features are used for processing. Greater attentional capacity is associated with more sharply defined ID functions. However, immediate memory did not play a significant role in the context of their investigation. Nevertheless, I included an RSPAN task as a measure of working memory to investigate the potential relationship between phoneme category learning and working memory over the time course of my experimental task.

Objectives and Hypotheses

The general aim of the present study was to investigate the cognitive mechanisms involved in learning to perceive L2 speech sounds. The method used in the present study was adapted from Schoenherr and Logan (2014) which, in turn, was modelled after Pisoni et al. (1982). I modified the methods used in these earlier studies by separating the phoneme identification task into five blocks in order to obtain a more fine-grained understanding of the learning process. Although evidence of L2 phoneme category acquisition as a result of a short experimental training session has been provided in several studies (e.g., Pisoni et al., 1982; Schoenherr & Logan, 2013; Schoenherr & Logan, 2014), the time course of the acquisition process has not yet been investigated. It is possible that new phoneme categories are learned through a gradual perceptual shaping process that enables better perception of previously unperceived phoneme categories in small increments over the course of an experimental session. Another possibility is that there is a learning threshold, or tipping point, where an individual's mode of attending switches from acoustic to phonemic, and absolute categorical perception of an L2 phoneme occurs suddenly at a specific point in time over the course of an experimental session. Blocking of the experiment will enable the individual ID responses to be assessed in a block to block comparison of ID performance. Aside from blocking the experimental procedure, this experimental paradigm was essentially the same as the feedback training task from Experiment 3 and three-category ID response task from Experiment 1 by Pisoni et al. (1982).

The training task for the current study consisted of presenting three exemplar stimuli from the VOT continuum ranging from -70 ms to 70 ms. Each stimulus represents an ideal case of either the prevoiced, voiced, or unvoiced bilabial stop consonant phoneme category and was presented 80 times, in random order, for a total of 240 trials. After each stimulus was presented, the participant was required to identify the phoneme category associated with the stimulus. The participant was given feedback whether the response was correct or incorrect. I used this particular training procedure because it was demonstrated to be effective in facilitating prevoiced phoneme category learning (e.g., Pisoni et al., 1982; Schoenherr and Logan, 2013; 2014). The use of exemplar stimuli during training facilitates the formation of a durable representation of a prevoiced category phoneme in memory, and the use of feedback focuses the listener's attention toward the relevant acoustic differences that differentiate the prevoiced category from the voiced category.

For the current experiment, I employed the three category identification task used by Pisoni et al. (1982). I did not employ a two-category identification task because Pisoni et al. and Schoenherr and Logan (2013, 2014) have clearly demonstrated that monolingual native language English speakers can consistently and accurately differentiate the voiced and unvoiced phoneme categories. I also did not employ the ABX discrimination task used by Pisoni et al. because discrimination tasks do not necessarily require the listener to attend to the stimuli at the phonemic level. I was concerned with the way that learning is related to the interface between acoustic and phonemic stimulus properties, and the requirement of phoneme category labelling for the identification task requires attention toward the phonemic properties. By only employing a three-category identification task, I was able to increase the number of trials and obtain more identification data from each participant. The experimental task consisted of a 15-item stimulus set from along the VOT continuum that spanned the range from -70 ms to 70 ms, with each stimulus presented six times per block, resulting in 30 presentations of each stimulus, for a total of 450 experimental trials. Considering that this study involves the ANT, the RSPAN task, and a relatively substantial training procedure, I concluded that 450 trials would best maximize the amount of data that could be collected without risking serious degradation in identification performance due to particiapant fatigue.

I expected that the ID functions obtained in the present investigation would resemble those obtained by Pisoni et al. (1982) (see the left-hand panel of Figure 1, above). These ID functions are characterized by a sharply defined category boundary between voiced /b/ and unvoiced aspirated /p/ ID functions, but a more gradual transition from the voiced to prevoiced ID functions.

Based on the models of nonnative speech perception (the Perceptual Assimilation Model, the Speech Learning Model, and the Native Language Magnet model), I expected that participants would base their interpretation of the novel speech sounds on their relationship to L1 phoneme categories. This would be sufficient for identifying the stimuli belonging to the /p/ category (+30 to +70 VOT). However, attending to phonemic information alone will not be sufficient as a means of distinguishing between stimuli in the -70 to +20 VOT range. According to the models of L2 speech perception, both the /p^h/ and the /b/ (the stimuli ranging from -70 to +20 ms VOT) will, by default, be perceived as belonging to the /b/ category by native speakers of English. However, with a short period of training English listeners can reliably perceive differences in the prevoiced (negative VOT) region of the VOT continuum (Pisoni et al., 1982). Participants received training intended to facilitate their ability to selectively attend to the relevant acoustic property (VOT) of the prevoiced category /p^h/. Because participants lacked well-established phonemic knowledge of this category, they had to rely on their ability to selectively attend to acoustic properties as a means of categorizing the prevoiced stimuli. As noted above, the range of stimuli in the prevoiced category include speech sounds often produced in English as free variants in word-initial contexts, and as allophonic variants in intervocalic phonetic contexts (Strange & Dittmann, 1984). Therefore, English listeners will have had experience with the prevoiced category at an acoustic/phonetic level, but not as a phonemically distinct category.

I expected that some participants would identify stimuli as either /b/ or /p^h/ based on the degree of perceived goodness of fit to either the exemplar of the /b/ category or of the newly formed /p^h/ category, in accordance with the Perceptual Assimilation Model (PAM) (Best et al., 2001). The difference between prevoiced stimuli and the voiced unaspirated phonemic exemplar (i.e., /b/) is termed a Category Goodness (CG) difference by the PAM. The prototypical /b/ has a VOT of 0 ms, and stimuli with VOT ranging from -70 to +20 ms VOT will be by default perceived phonemically as /b/ by English listeners. A CG difference is more readily perceived than the difference between stimuli when Single Category assimilation is used to perceptually classify phonemes, but not as well as when the stimuli belong to separate categories. Consistent with the PAM, Schoenherr and Logan (2013) found that for adjacent /p^h/ and /b/ categories, participants showed partial evidence of phonemic categorization, whereas a sharp category boundary was present between the L1 categories /p/ versus /b/ (see Figure 2 below). The presence of one clear-cut category distinction (between /b/ and /p/) and one somewhat gradual category distinction (between /p^h/ and /b/) suggests that both high-level phonemic and lower-level acoustic modes of attending are being employed, depending on where the stimulus is situated on the VOT continuum. A sharply defined category boundary is highly suggestive of attending toward stimuli purely at the more abstract phoneme level; acoustic-level stimulus differences can be automatically processed below the level of explicit awareness. Less clear-cut category boundaries may be more suggestive of attending toward both acoustic and phonemic levels in order to categorize stimuli. This kind of attending is required when a newly developed phoneme category is not yet rigidly established and when this new category overlaps or interferes with an L1 phoneme category.

I expected that an acoustic mode of attending would be employed in order to identify stimuli with a VOT below +20 and that this would be a way to assess the goodness of fit to either the /b/ or to the newly developed /p^h/ phoneme category. A high goodness of fit with /b/ would be identified as /b/, with stimuli having smaller VOT values gradually being perceived as poorer representations of /b/ and better representations of /p^h/ as the VOT decreases from 0 ms (the VOT of an exemplary /b/ stimulus) to -70 ms. Stimuli with the poorest assimilation to the /b/ category and best assimilation to the /p^h/ were expected to be identified as /p^h/, and this is what Schoenherr and Logan (2013) found (see Figure 2). I proposed that a gradual slope for the category boundary between /p^h/ and /b/ in ID performance is indicative of an acoustic mode of attending, in addition to the phonemic mode of attending used to label the stimulus category. Absolute identification was not expected to initially occur for these categories because attention would initially need to be focused at the acoustic level rather than at the higher-order phoneme level that more or less automatically categorizes speech sounds according to phoneme categories. Schoenherr and Logan (2014) found evidence supporting an acoustic mode of attending for prevoiced stimuli based on confidence ratings and individual measures of attention and immediate memory.

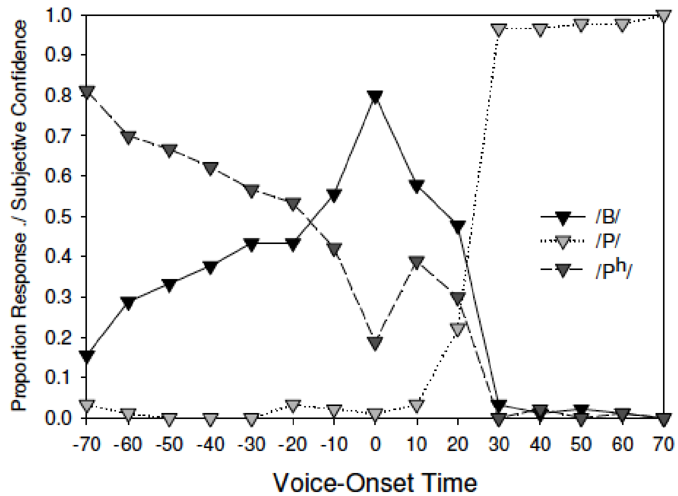


Figure 2. From Schoenherr and Logan (2013). ID functions for prevoiced, voiced, and unvoiced categorization of bilabial stop consonants ranging from -70 to +70 ms VOT. There is a relatively clear-cut category boundary between /b/ and /p/, and a more gradual boundary between /b/ and /p^h/.

The present study included measures of working-memory (the RSPAN task) and attentional network efficiency (the ANT) in order to investigate whether a relationship exists between working memory, executive control of attention, lower levels of attention, and the process of perceptual learning. Learning was measured by ID performance for stimuli situated within the prevoiced region of the VOT, over the course of five experimental blocks. A link between the more implicit, lower-level, measures of attention—the orienting and alerting components of the ANT—was expected to be associated with the ability to maintain and focus attention toward stimulus properties along the acoustic dimension, thus contributing to the ability to accurately categorize stimuli. The more explicit, higher-level measure of attention—the executive control component of the ANT—was expected to be associated with the ability to inhibit perception of the prevoiced stimulus using L1 phoneme categorization, thus facilitating attending to prevoiced stimuli at the acoustic level. Working memory was expected to be associated with executive control of attention, as well as the ability to negotiate the interface between attention and memory in order to access either the newly formed prevoiced category, or the previously established voiced category, while maintaining focus on acoustic stimulus properties in order to label prevoiced stimuli.

Previous studies of L2 phoneme learning (e.g., Pisoni et al., 1982; Logan et al., 1991; Golestani & Zatorre, 2009) have found a high degree of individual variability in ID performance. These differences may be related to individual differences in some form of executive attention or working memory capacity, but relatively few studies have investigated the relationship between these cognitive processes and L2 speech perception. The ability to identify a stimulus as either /b/ or /p^h/ in this type of experiment involves the cognitive ability responsible for inhibiting the impulse to use phonemic encoding mechanisms to categorize the stimulus into a pre-existing L1 phoneme category. Otherwise, all stimuli with a VOT of +20 or less would be perceived as /b/. I expected that individual differences in executive control of attention and working memory capacity are associated with the ability to inhibit interference from automatic categorization of prevoiced stimuli based on the L1 phonemic system. The more an individual is able to focus attention on acoustic stimulus properties and retain a representation of the stimulus in a working memory store, the better that individual will be at making a judgment regarding how well (or poorly) the acoustic properties of the stimulus correspond with a phoneme category. Further, it has been suggested that attending to acoustic-level stimulus properties require more short-term memory resources than attending at the phonemic level (Werker & Logan, 1985). Therefore, measures of memory and attention at the level of executive control as well as lower-level orienting and alerting mechanisms may be involved in processing the prevoiced speech stimuli at the acoustic level.

Learning to perceive nonnative phonemes, especially those situated within pre-established L1 phoneme categories, is extremely difficult for individuals with no previous linguistic experience with the given L2 phoneme. Eimas (1978) even argued that the limited ability to make distinctions between speech sounds within the same phoneme category suggested that the neural mechanisms mediating VOT perception degenerate after a certain critical period of development. Pisoni et al. (1982) demonstrated that with proper training techniques, the perceptual mechanisms used by adults to categorize stop consonants can in fact be modified. Pisoni et al. were able to show that American English (AE) listeners can distinguish a prevoiced unaspirated category not present in AE. Further, when individual differences are minimized by only using data from participants who perform at a high level on a training procedure, very sharp identification functions can result for all three phoneme categories (see Figure 1, right panel) compared to the group ID functions (see Figure 1, left panel). The underlying source of this individual variability is not well understood, but may be related to individual differences in working memory capacity and attentional networks. This potential relationship was investigated in the present study by comparing individual differences in working memory and attention measures to perceptual learning as measured by identification performance over the course the phoneme categorization experiment.

Based on the proposed association between executive control of attention and L2 phoneme identification performance, as well as individual variability in L2 phoneme category acquisition found by Pisoni et al. (1982), I hypothesized that individual executive control of attention would be positively related to metalinguistic awareness of phonemic stimulus properties. Each participant's executive control of attention was measured with the executive attention component of the ANT as well as with a portion of the reading span (RSPAN) task (e.g., Daneman & Carpenter, 1980). The metalinguistic awareness of phonemic stimulus properties was measured by performance on the phoneme identification task.

Based on the gradual transition between /b/ and /p^h/ in the identification functions provided by Shoenherr and Logan (2013) (Figure 2) and in Experiment 1 of Pisoni et al. (1982) (Figure 1, left), I posited that the early stages of metalinguistic phoneme identification are based on the ability to recognize the degree of acoustic relationship between the prevoiced stimuli and category exemplars. I expected that efficiency of the lower-level, more implicit, attention networks (orienting and alerting) would be positively related to early ID performance in the prevoiced-voiced region. I expected that the perceptual system responsible for making comparisons based on small acoustic differences would become gradually more attuned over the course of the short experimental session and as a result, ID performance would improve over the course of the five experimental blocks. As a caveat, I predicted that some participants would have markedly better ID performance and produce ID functions consistent with those found in Experiment 3 of Pisoni et al. (1982) (Figure 1, right). I hypothesized that these individuals would have developed durable phonemic representations of /p^h/ at some point of the course of the experiment and the accompanying shift in mode of attending from acoustic and phonemic to purely phonemic would be characterized by a sudden change in ID functions between blocks. This sudden shift in mode of attending would be evidenced by ID functions with two clear-cut phoneme category boundaries similar to those obtained in Experiment 3 of Pisoni et al. (Figure 1, right panel), rather than one clear-cut boundary and one gradual transition (Pisoni et al. (Figure 1, left panel). It is at this point in the course of L2 learning that I expected a new durable representation of the prevoiced phoneme category would be developed. Once a new phoneme category has been acquired, acoustic properties become more or less automatically processed and phoneme identification can rely on absolute categorization based on attending to stimulus information at the phonemic level. This change in perception may be also characterized as a transition from an effortful process to a more automatized process that is typically associated with L1 phoneme processing.

I expected that acquisition of L2 phoneme categories relies on cognitive processes involved in language learning including executive control of attention. Therefore, I hypothesized that WMC and efficiency of the executive control attention network would be positively related to learning efficiency. More efficient executive function was expected to lead to faster L2 category formation, and steeper prevoiced-voiced category boundaries would occur earlier in the experiment. Executive control measures were expected to be related to ID performance at the beginning of the task as a result of their role in inhibiting interference from the L1 phonemic system and preserving acoustic representations of the stimuli. I predicted that this ability would facilitate learning, and that individuals with higher executive function would be able to transition toward purely phonemic attending more quickly as a result of faster category development. Once an L2 category has been acquired, I hypothesized that measures of memory and attention at both higher and lower levels of processing is not implicated in stimulus processing. This is because it has been suggested by Werker and Logan (1985) that phonemic level stimulus encoding requires less short-term memory resources than encoding at the level of acoustic stimulus properties. Thus, once a stable representation of the prevoiced phoneme category has been formed, stimulus identification can rely on the largely automatic process of categorization at the phonemic level.

In addition to working memory capacity and attentional measures, it is possible that other individual differences in experience would affect categorization performance. In the present study, I examined individual differences in musical ability. Specifically, I expected that relatively highly skilled musicians are skilled at attending to fine acoustic stimulus properties, but they may have difficulty developing a new phonemic category due to their well developed ability to attend to stimulus properties at the acoustic level. I hypothesized that skilled musicians may artificially create a phonemic boundary based on acoustic features; they would produce clear-cut ID functions highly suggestive of categorical perception of the prevoiced phoneme for the duration of the experiment; however, attending will not be an automatized phonemic process, rather, musicians will continue to attend to acoustic stimulus features for the duration of the experimental procedure. To investigate this hypothesis, a short musical experience survey was administered to participants prior to the experiment. ID functions for experienced musicians were compared those of nonmusicians to investigate potential differences in ID performance over the course of the experiment.

Method

Participants

35 participants were recruited for this experiment from the university's online study participation system and from a community sample (20 females, 15 males). All participants spoke English as their first language and had no significant knowledge of any Southeast Asian language (this was to ensure that prior exposure to languages that use the prevoiced category was kept to a minimum). Participants had normal or corrected vision and had no significant hearing or neurological problems. They signed an informed consent form prior to the experiment. Each Carleton University student recruited through the SONA system received a 1.5% course credit.

Materials

Identification Task Stimuli. Fifteen synthetic speech stimuli were obtained from the Haskins Laboratories website (Haskins Laboratories, 2014). The stimuli originally were generated by Lisker and Abramson (1967). The stimuli varied along the VOT continuum from -70 to 70 ms VOT, in 10 ms increments. These stimuli correspond to labial stop consonant-vowel (CV) syllables representing prevoiced unaspirated, voiced unaspirated, and unvoiced aspirated stops. The latter two categories are present in the inventory of English phonemes while the former is not. The sounds were originally recorded on reel-to-reel tape and later converted into AIFF format at Haskins Laboratories. Once downloaded, the stimuli were pre-processed using a DC offset correction to eliminate clicks present in the AIFF versions and then converted into .WAV files using Audacity software. It should be noted that these stimuli were also used by Schoenherr and Logan (2013, 2014) but are not the same stimuli used by Pisoni et al. (1982).

Individual Differences Measures. Participants were required to complete three individual difference measures: 1) a music and language experience questionnaire, 2) a reading span task (RSPAN), and 3) the Attentional Network Test (ANT). The music and language experience questionnaire was designed to assess individual differences in language and musical experience that could have affected performance of the speech perception task (see Appendix A for the music and language experience questionnaire).

Working memory capacity was measured using a version of the the reading span

(RSPAN) task adapted from Turner and Engle (1989). This measure assesses an individual's ability to maintain and retrieve task relevant information in active working memory stores while attention is focused on performing an unrelated simple, yet nontrivial, cognitive task. In this version of the RSPAN task participants were presented with a series of sentences in both auditory and visual form. Participants were required to read the sentences displayed on the screen aloud as they were presented auditorily through headphones. Each sentence was followed immediately with a letter to be recalled following the presentation of a span of two to five sentences. Participants were asked to make judgments after reading each sentence and letter pair regarding whether the sentence was semantically correct. Judgments were made by pressing the letter "Y" on the computer keyboard to indicate a semantically correct judgment, or "N" to indicate that the sentence was not semantically correct. For example, when presented with the sentence-letter pair "*Andy was stopped by the policeman because he crossed the yellow heaven. R*" the participant should have pressed "N" to indicate that the sentence was not semantically correct. Once a span of two to five sentence-letter pairs was presented participants were asked to recall the letters from the set in the order they were presented by writing them down on a response sheet. Three sets of each length of spans were presented. There were three two-sentence spans, three three-sentence spans, three four-sentence spans, and three five-sentence spans. Thus, a total of 42 sentences were presented over twelve sets. The RSPAN task took approximately 12 minutes to complete.

The RSPAN scores were obtained based on two scoring methods: 1) the traditional absolute span scoring method (e.g., Daneman & Carpenter, 1980) and 2) the Partial scoring method (e.g., Conway et al., 2005). The absolute scoring method obtains scores for each individual participant by adding up the lengths of sets for which all letters were correctly recalled. For example, if a participant recalled all letters correctly from three of twelve sentence spans, one two-sentence span, one three-sentence span, and one five-sentence span, that participants absolute RSPAN score would be 10 (2 + 3 + 5). The partial scoring method awards points for each letter correctly recalled in the correct position. The score obtained is the number of letters (out of 42 in this case) correctly recalled in the correct position. The partial scoring method would give a participant who correctly recalled four out of five letters in a five-sentence span four points for that span, compared to zero points that would be given using the absolute method.

The Attention Network Task (ANT), developed by Fan et al. (2002) and retrieved from https://sacklerinstitute.org/cornell/assays_and_tools/ant/jin.fan/, measures each participant's ability on three distinct attentional functions: alertness, orienting, and executive control of attention (also referred to as conflict resolution). For each trial participants were required to indicate the direction (either left or right) of a target arrow presented to them either above or below a central fixation cross on a computer monitor by pressing the "left arrow" or "right arrow" keys. Target arrows were presented in either in the middle of a row of flankers facing either the same way as the target arrow (congruent) or the opposite way (incongruent), or with no flankers (neutral condition). Prior to the target presentation, participants were given either no cue, or a cue indicating that the target would be presented shortly. The cue was an asterisk presented either centrally, both above and below the fixation, or in the location of the target (either above or below the fixation). The ANT consisted of four blocks which included a practice block with response feedback, and three test blocks. The ANT took approximately 25 minutes to complete. See Figure 3 for an illustration of the ANT.

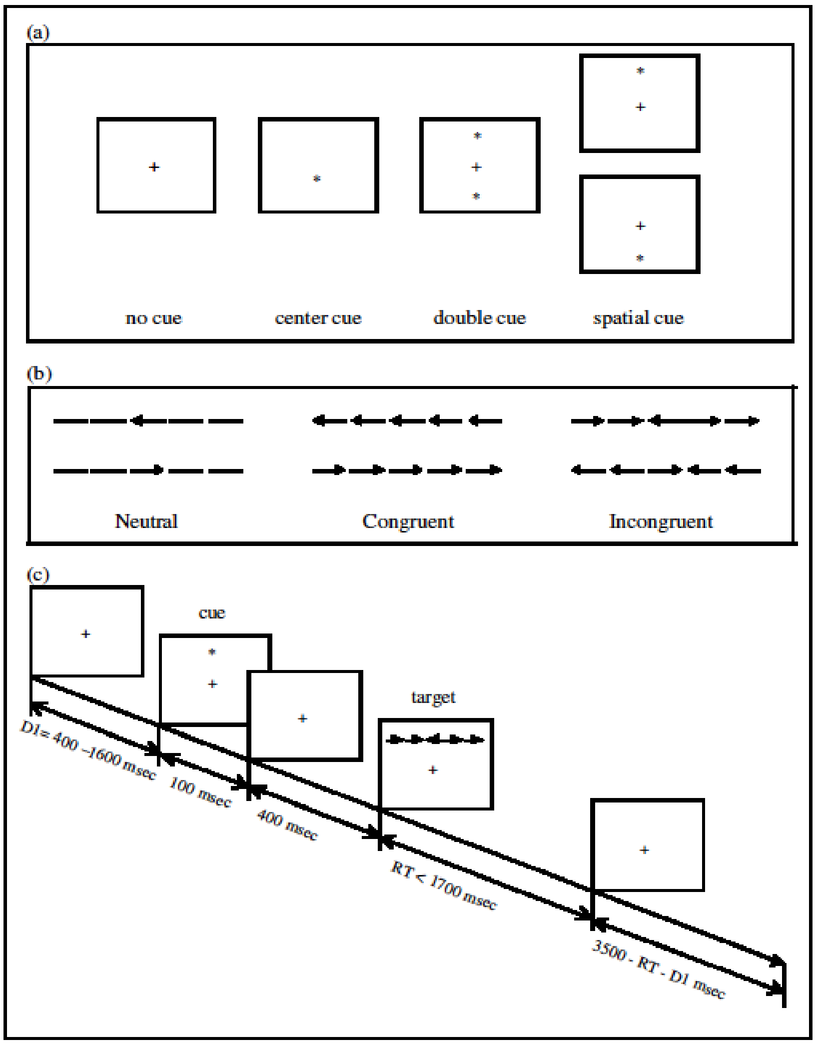


Figure 3. An illustration of the ANT experimental procedure. (a) The four cue conditions; (b) the six stimuli presented; (c) an example of the procedure for one trial. (From Fan et al., 2002, p. 341).

Procedure

The individual difference measures were administered prior to the speech perception training and experimental procedures. The auditory experience questionnaire was administered first, followed by the RSPAN and the ANT. The order of administration of the RSPAN and ANT was counterbalanced. PsychoPy software (Pierce, 2007) was used to control presentation of stimuli and collection of responses for the RSPAN, training, and experimental tasks. The ANT is a Java program that was administered using JAR (Java Application Runtime environment) Launcher software.

Training. The training procedure and phoneme identification task were adapted from those originally used by Pisoni et al. (1982). Participants were seated in front of a 14" laptop screen in a sound attenuated booth. All auditory information was presented through headphones adjusted to a comfortable listening level. A training task was administered prior to the principal experimental task to familiarize participants with the experimental task and the association between the stimuli and the three phonemic categories. Participants were presented with exemplars from each of the three phoneme categories: /p^h/, /b/ and /p/. These exemplars were represented by stimuli with VOTs of -70, 0, and 70 ms, respectively. After hearing each stimulus, participants indicated whether the speech sound belonged to the /p^h/, /b/, or /p/ category by pressing the “V”, “B”, or “N” keys, labelled as “_B”, “B”, and “P”, respectively. Participants were given feedback after each response indicating whether the response was correct or incorrect. Feedback was given by presenting the word “Correct” or “Incorrect” visually on the monitor. Each of the three stimuli were presented 80 times during the training phase, for a total of 240 training trials. The stimuli were presented in random order. The training task took approximately 12 minutes to complete.

Experimental Task. Once training was complete participants were given the main experimental task. This task involved presenting participants with all 15 stimuli from the VOT continuum ranging from -70 to 70 ms VOT. As in the training procedure, participants were required to identify each stimulus as /p^h/, /b/, or /p/ by pressing either the 'V', 'B', or 'N' key, labelled /_B/, /B/, or /P/, respectively. However, in this phase of the experiment no response feedback was provided. The experimental task consisted of five blocks with a short break between blocks. Participants were presented with each stimulus six times in each block for a total of 90 trials per block, yielding 450 experimental trials. Stimuli were presented in random order. The test phase took approximately 15 minutes to complete. Overall, the study took approximately an hour and ten minutes for each participant to complete.

Results

Individual Difference Measures. The means and standard deviations of the attention network efficiencies, working-memory measures, and number of years of musical training are presented in Tables 1, 2, and 3, respectively. A median split was used to separate participants into high and low groups for the ANT (see Table 1) and RSPAN (see Table 2) measures. For the auditory experience measure, participants were categorized into high and low experience groups based on the number of years of musical training (see Table 3). A series of analyses indicates that there was a significant amount of variation in individual difference measures. Analyses of variance indicated significant differences between low and high levels of the ANT alerting efficiency scores, $F(1,33) = 38.51$, $p < .01$, ANT orienting efficiency scores, $F(1,33) = 38.67$, $p < .01$, ANT executive control of attention efficiency scores, $F(1,33) = 36.04$, $p < .01$, ANT average response times, $F(1,33) = 52.40$, $p < .01$, RSPAN scores using the absolute scoring method, $F(1,33) = 68.85$, $p < .01$, RPAN scores using the partial scoring method, $F(1,33) = 64.60$, $p < .01$ and number of years of musical training, $F(1,33) = 57.24$, $p < .01$.

Most participants (N=24) had at least one year of musical training experience and this measure was therefore used as a measure of musical experience for the sample as a whole. This provided a strong rationale to include musical training as an auditory experience measure for the general population of English speaking individuals. Additionally, seven of the 35 participants currently played an instrument two or more hours per week. These participants were classified as musicians, and all other participants were classified as nonmusicians. The musician/nonmusician classification was used in addition to the musical training measure because it was clear from the individual ID functions that participants who currently played an instrument more than two hours per week were better than the rest of participants (see Figure 8). Although the musician/nonmusician distinction clearly played a role in ID performance, only twelve participants currently played an instrument (seven of those twelve played more than two hours per week), therefore this measure was not as well suited as musical training as a predictor variable. See the middle and right columns of Tables 1, 2, and 3 for the means and standard deviations of median-split participant groups for each individual difference measure.

Table 1
Mean efficiency (RT average in seconds) for attentional networks.

		High Efficiency	Low Efficiency
Network	M (SE)	M (SE)	M (SE)
Alerting	32.86 (24.08)	15.12 (11.75)	50.71 (20.52)
Orienting	37.46 (23.28)	20.12 (19.96)	54.52 (11.06)
Executive Control	133.97 (37.57)	106.88 (14.56)	159.54 (34.59)
Overall RT	592.37 (87.25)	526.46 (15.56)	661.46 (72.35)

Table 2
Mean RSPAN score .

			High		*Low*	
Scoring Type	*M (SE)*		*M (SE)*		*M (SE)*	
Absolute	15.06	(9.90)	23.29	(3.40)	6.94	(7.38)
Partial	25.29	(8.92)	32.59	(5.03)	18.00	(5.54)

Table 3
Mean number of years of musical training.

			High		*Low*	
	M (SE)		*M (SE)*		*M (SE)*	
Years	5.34	(6.09)	10.29	(5.22)	0.59	(0.87)

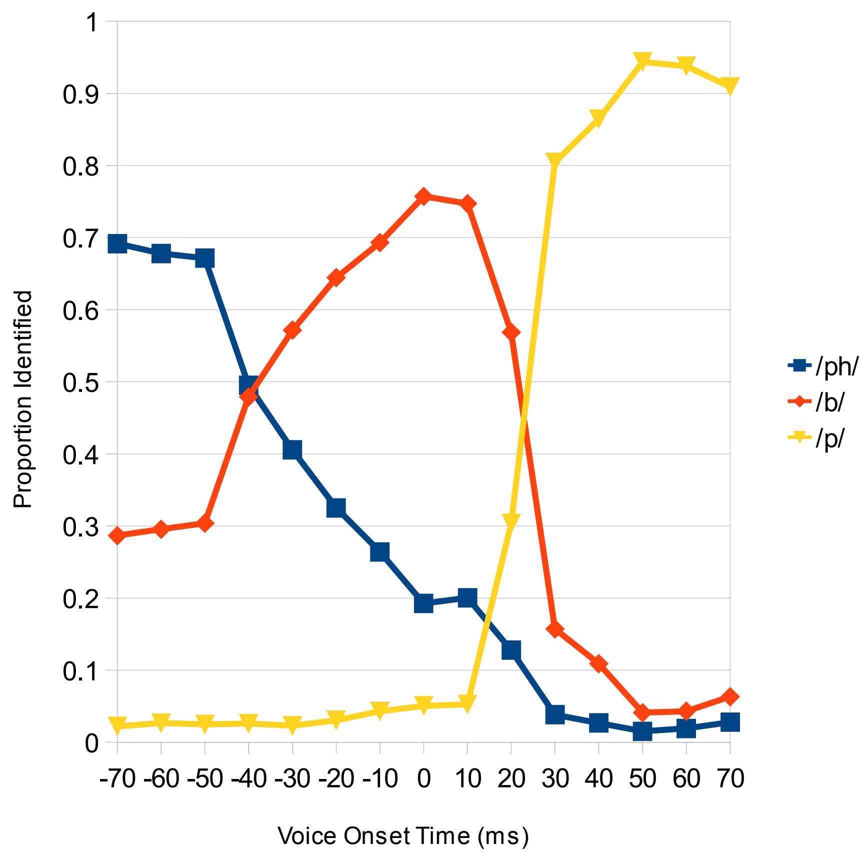


Figure 4. Overall ID functions for /pʰ/, /b/, and /p/ categories.

Identification responses. The results of the overall ID response functions averaged across all blocks and participants (see Figure 4) were very similar to those of Pisoni et al. (1982) (see left-hand panel of Figure 1) and Schoenherr and Logan (2013, 2014). About half of the participants were quite proficient at distinguishing between the prevoiced and voiced stop consonant phoneme categories, and almost all participants demonstrated the ability to distinguish between the voiced and unvoiced aspirated categories. The musicians' ID functions indicated that musicians were better at distinguishing between prevoiced and voiced stimuli than nonmusicians (see Figure 8). The block-by-block ID functions indicated that there was not a substantial or consistent change in individuals' ID performance across blocks.

ID functions for were created using each individual difference measure median-split participant group. The ID functions created from median-split groups based on ANT response time, RSPAN scores using Partial scoring strategy, and number of years of musical training revealed distinct groups (See figures 5, 6 and 7). Participants in the faster ANT response time group were better at distinguishing between prevoiced and voiced stimuli (see the left-hand panel of Figure 8) compared to participants with slower ANT response time (see the right-panel of Figure 8). Participants in the high RSPAN group, according to the Partial scoring method, were better at distinguishing between prevoiced and voiced stimuli (see the left-hand panel of Figure 9) compared to participants in the low RSPAN group (see the right-hand panel of Figure 9). Participants with more years of musical training were better at distinguishing between prevoiced and voiced stimuli (see the left-hand panel of Figure 10) compared to participants with less musical training (see the right-hand panel of Figure 10).

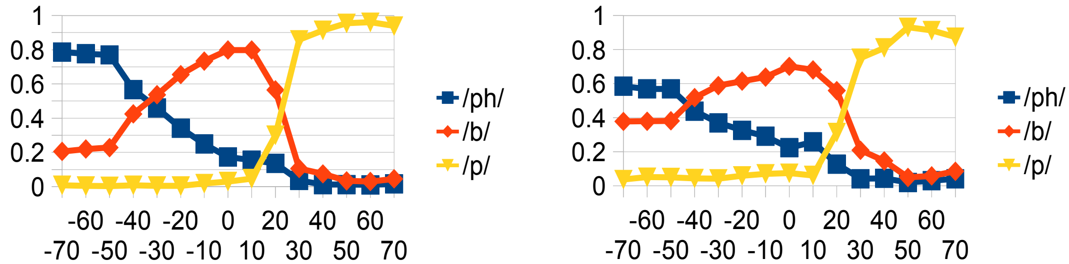


Figure 5. ID functions for overall performance, for fast and slow ANT response time groups. Left: ID function for fast ANT response time group. Right: ID function for slow ANT response time group.

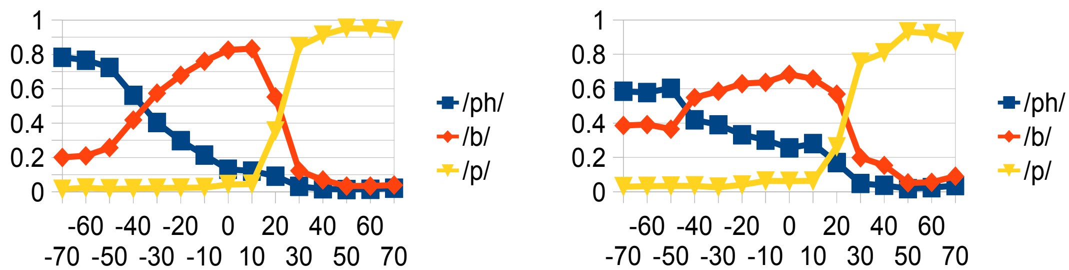


Figure 6. ID functions for overall performance for high and low RSPAN (Partial scoring method) groups. Left: ID function for high RSPAN group. Right: ID function for low RSPAN group.

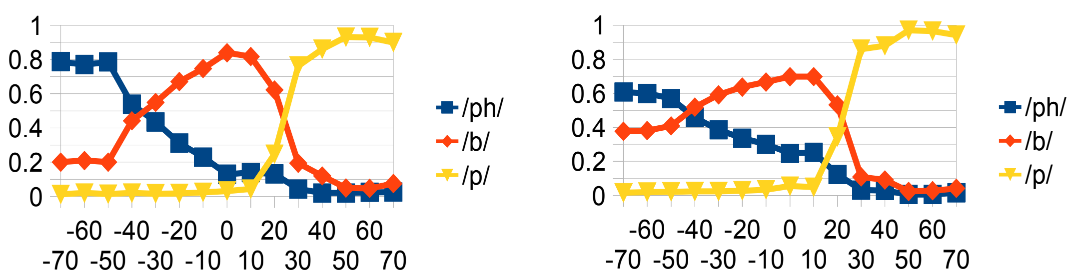
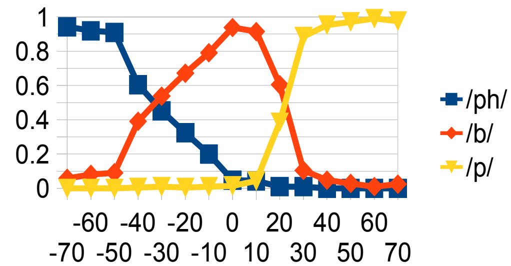


Figure 7. ID functions for overall performance for high and low number of years of musical training. Left: ID function for high musical training group. Right: ID function for low musical training group.



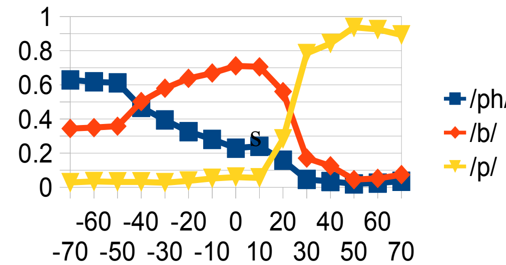


Figure 8. ID functions for overall performance for musicians (N=7) and nonmusicians (N=28). Musicians were visibly better at identifying prevoiced stimuli compared to nonmusicians.

Logistic Regression. In order to quantify these observations, the identification performance for prevoiced responses for all 15 VOT stimuli was tabulated for each block, as well as overall, for each participant. Identification frequency of responses for prevoiced categorization for all 15 VOT stimuli was tabulated for each block, as well as overall, for each participant. Logistic functions were fitted to individual participants' ID functions for prevoiced category responses for each block, as well as for overall performance, using the R software (R Core Team, 2014) library DRC (Dose Response Curve; Ritz & Streibig, 2005). The DRC software yielded slope, lower limit, upper limit, and 50% crossover parameters for each plot. (The 50% crossover is defined as the point on the VOT continuum that corresponds to where 50% of the prevoiced responses are located). See Figure 9 for an example of two logistic functions fitted to prevoiced identification plots. Plots were created in order to facilitate efficient categorization of prevoiced identification performance as either "successful" or "unsuccessful" for each block. RCommander (Fox, 2005) was used to apply a hierarchical means clustering solution in order to sort the individual prevoiced identification plots for each block according to the lower limit, upper limit, and 50% crossover parameters from each logistic model. See Figure 10 for an example of a hierarchical means clustering dendrogram.

Roughly half of participants successfully distinguished between the prevoiced and voiced stimuli in each block. Although the hierarchical means clustering strategy was generally effective in grouping all prevoiced ID performances as either "successful" or "unsuccessful", some inconsistencies led to some plots being categorized incorrectly. For example, some successful participants tended to draw the distinction between the prevoiced and voiced stimuli at longer VOT values than most other successful participants, and as a result these participants were improperly classified with the unsuccessful ID performance group by the hierarchical means clustering analysis (see Figure 11, right). Another inconsistency resulted from logistic functions that coincidentally had the correct 50% crossover category boundary but with numerous inappropriate responses for stimuli in the positive VOT range (see Figure 11, middle). Finally, some participants were biased to make prevoiced responses which also had an effect on the ability of the hierarchical means clustering analysis to correctly group prevoiced ID performances into successful and unsuccessful groups (see figure 11, left). These inconsistencies were not due to any inherent failure of the clustering algorithm but occurred as a result of simply applying criteria for separating groups that did not take into account the characteristics or features of an “ideal” categorical identification function based on linguistic and psychoacoustic consideration. As a consequence of these inconsistencies in the clustering solutions, all logistic fitted plots were visually inspected and recategorized as either successful or unsuccessful prevoiced identification as necessary. This manual check of the clustering solutions resulted in anywhere from four to eight functions being reclassified as belonging to the opposite group (i.e., successful to unsuccessful or unsuccessful to successful), for each block. Reclassification required a consensus of two observers.

Figure 9. Examples of logistic functions fitted to plots for a "successful" prevoiced ID individual participant's performance block (left) and an "unsuccessful" ID performance (right). Y-axis represents VOT stimuli from -70 ms (represented by 0) to 70 ms (represented by 15). Prevoiced response frequency is represented along X-axis.

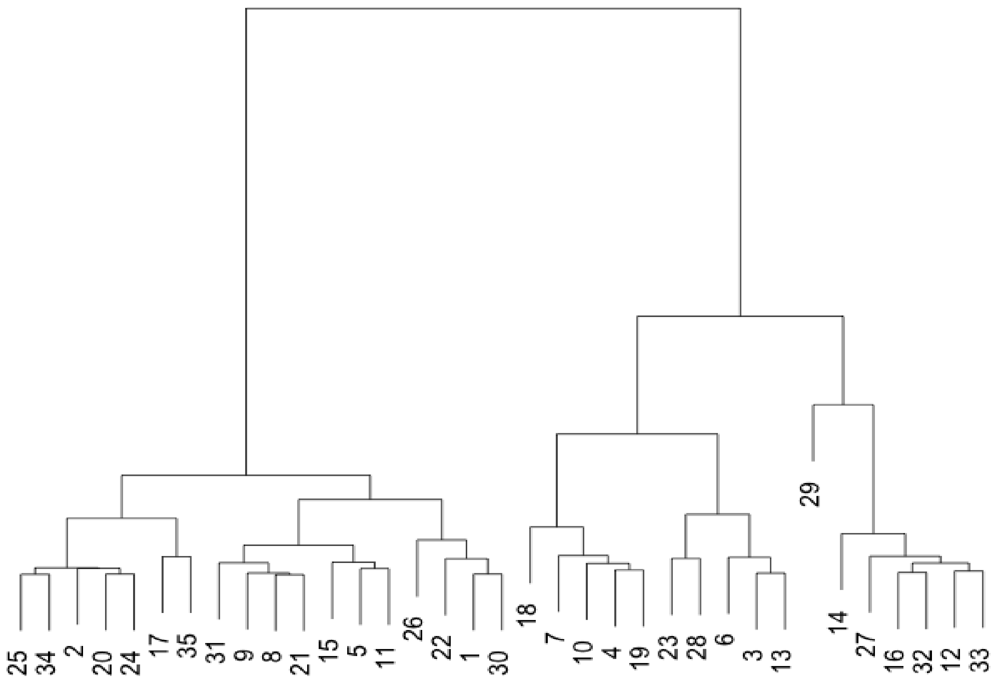


Figure 10. Hierarchical means clustering of individual prevoiced logistic functions for block 3. Brach 25 corresponds to the plot on the left side of Figure 5 ("successful") and branch 18 corresponds to the plot on the right side ("unsuccessful").

Figure 11. Logistic fitted plots classified incorrectly by hierarchical means clustering analysis. Left: high prevoiced response bias. Middle: coincidentally "successful" logistic model fit. Right: Prevoiced/voiced boundary inappropriately shifted toward voiced category.

Once each participants' prevoiced ID responses for each block were categorized as either successful or unsuccessful, data from participants who had the 10 most ambiguous logistic functions were removed from each block, leaving data from 25 participants for further analysis. Next, predictor variables were selected for use in a logistic regression to predict successful or unsuccessful categorization of the prevoiced stimuli. The predictor variables were chosen from the ANT measures (alerting, orienting, executive), plus the overall ANT response times, the three RSPAN measures (absolute RSPAN scores, and partial RSPAN scores) and finally, two measures from the auditory experience questionnaire (number of years of music experience and number of hours per week spent playing an instrument). The predictors chosen were RSPAN scores using Partial scoring, ANT average overall response time, and number of years of musical training. The choice of these variables was based on the ID functions for each median-split individual difference measure. These three measures revealed the largest differences between the high and low median-split groups. Results from from this regression analysis are presented in Tables 4 through 9.

The logistic regression output tables merit some explanation. The slopes of the constant and predictor variables and their standard errors are presented in the first data column (*B (SE)*). These are the coefficients used in the logistic regression model. The odds ratio of each predictor is presented in the third data column (*Odds ratio)* and is an indicator of the change in the odds of successful prevoiced categorization resulting from a change in one unit of the predictor. The lower and upper confidence intervals for the odds ratio are presented in data columns 2 and 4 (*Lower* and *Upper*). Confidence values greater than 1 signify a positive relationship between the predictor and the outcome variable (prevoiced categorization success in this case). Confidence values less than 1 signify a negative relationship between the predictor and the outcome variable. Confidence intervals that span across 1 indicate that there is a chance that in the population the direction of the relationship between the predictor and outcome variables is the opposite of what was observed in this experiment. A measure of the logistic regression analogue to R^2 is provided for three different methodologies: Hosmer and Lemeshow (1989), Cox and Snell (1989) and Nagelkerke (1991). These measures indicate how well the data fit the logistic regression line. The model chi-square statistic measures the difference between the model with the included predictors and the model with only the constant. A significant chi-square statistic for a logistic regression indicates that the model predicts the probability of a certain outcome significantly better with the predictor(s) included in the model. The R^2 and chi-square values are reported as a "Note" below each regression output table.

ANT overall average response time was a significant predictor of prevoiced identification performance success for block 1. Number of years of musical training was a significant predictor of prevoiced identification performance success from block 4 and block 5. These results indicate that when these three predictors are included the logistic regression model predicted prevoiced categorization performance success significantly better than when these predictors were not used in the model for all five blocks, as well as overall performance collapsed across blocks.

The role of ANT and RSPAN measures changes, however, when the analyses separate the categorization performance of musicians from nonmusicians. A binomial logistic regression analysis that included only nonmusicians indicated that when ANT average response time and RSPAN score are used as predictors, the logistic regression model did not predict prevoiced categorization performance success significantly better than when these predictors were not used in the model for four of five blocks, as well as overall (see Tables 10 through 15). The exception was for block 3, which indicated that the model was significantly better at predicting prevoiced categorization performance when the ANT average response time and RSPAN score were included in the model than when the predictors were not included in the model.

Table 4

Block 1 binomial logistic regression results with ambiguous cases removed. N=25.

	B (SE)	*95% CI for odds ratio*		
		Lower	*Odds ratio*	*Upper*
Included				
Constant	15.94 (9.35)			
ANT Response Time	-0.031 * (0.02)	0.93	0.97	0.99
RSPAN	0.05 (0.10)	0.86	1.06	1.36
Years of Musical Training	0.21 (0.14)	0.98	1.23	1.78

Note. R^2= .57 (Hosmer-Lemeshow), .55 (Cox-Snell), .73 (Nagelkerke). Model $\chi 2$ (3) = 19.80, $p < .0005$.
*$p < .05$.

Table 5

Block 2 binomial logistic regression results with ambiguous cases removed. N=25.

	B (SE)	*95% CI for odds ratio*		
		Lower	*Odds ratio*	*Upper*
Included				
Constant	19.2 (10.11)			
ANT Response Time	-0.03 . (0.02)	0.93	0.97	0.99
RSPAN	-0.09 (0.07)	0.86	1.06	1.36
Years of Musical Training	0.23 (0.16)	0.98	1.23	1.78

Note. R^2= .47 (Hosmer-Lemeshow), .55 (Cox-Snell), .73 (Nagelkerke). Model $\chi2$ (3) = 16.12, $p < .005$. . $p < 0.1$

Table 6

Block 3 binomial logistic regression results with ambiguous cases removed. N=25.

	B (SE)	*95% CI for odds ratio*		
		Lower	*Odds ratio*	*Upper*
Included				
Constant	11.15 (11.74)			
ANT Response Time	-0.02 (0.02)	0.93	0.98	1.00
RSPAN	0.06 (0.10)	0.88	1.06	1.32
Years of Musical Training	0.24 . (0.13)	1.03	1.27	1.78

Note. R^2= .47 (Hosmer-Lemeshow), .48 (Cox-Snell), .64 (Nagelkerke). Model $\chi2$ (3) = 16.40, $p < .001$. . $p < .1$.

Table 7
Block 4 binomial logistic regression results with ambiguous cases removed. N=25.

	B (SE)	*95% CI for odds ratio*		
		Lower	*Odds ratio*	*Upper*
Included				
Constant	12.98 (12.42)			
ANT Response Time	-0.02 (0.02)	0.93	0.97	0.99
RSPAN	(-0.02) (0.09)	0.86	1.23	1.78
Years of Musical Training	0.22 * (0.11)	0.98	1.06	1.36

Note. R^2= .39 (Hosmer-Lemeshow), .55 (Cox-Snell), .73 (Nagelkerke). Model $\chi 2$ (3) = 16.40, $p < .001$. * $p < .05$.

Table 8
Block 5 binomial logistic regression results with ambiguous cases removed. N=25.

	B (SE)	*95% CI for odds ratio*		
		Lower	*Odds ratio*	*Upper*
Included				
Constant	-2.35 (3.95)			
ANT Response Time	0.00 (0.01)	0.99	1.00	1.01
RSPAN	-0.03 (0.06)	0.85	0.97	1.08
Years of Musical Training	0.31 * (0.16)	1.07	1.36	2.06

Note. R^2= .23 (Hosmer-Lemeshow), .27 (Cox-Snell), .36 (Nagelkerke). Model $\chi 2$ (3) = 7.81, $p < .05$. * $p < .05$.

Table 9

Overall binomial logistic regression results with ambiguous cases removed. N=25.

	B (SE)	95% CI for odds ratio		
		Lower	*Odds ratio*	*Upper*
Included				
Constant	19.45 (12.05)			
ANT Response Time	-0.03 (0.02)	0.92	0.97	1.00
RSPAN	-0.08 (0.08)	0.77	0.92	1.07
Years of Musical Training	0.24 . (0.14)	1.02	1.27	1.85

Note. R^2= .44 (Hosmer-Lemeshow), .45 (Cox-Snell), .60 (Nagelkerke). Model χ2 (3) = 15.07, $p < .005$. . $p < .1$.

Table 10

Block 1 binomial logistic regression results with musicians removed. N=28.

	B (SE)	95% CI for odds ratio		
		Lower	*Odds ratio*	*Upper*
Included				
Constant	6.08 (3.95)			
ANT Response Time	-0.01 . (0.01)	0.97	0.99	1.00
RSPAN	-0.01 (0.05	0.9	0.99	1.09

Note. R^2= .11 (Hosmer-Lemeshow), .15 (Cox-Snell), .20 (Nagelkerke). Model χ2 (2) = 4.08, $p > .1$. . $p < .1$

Table 11

Block 2 binomial logistic regression results with musicians removed. N=28.

	B (SE)	*95% CI for odds ratio*		
		Lower	*Odds ratio*	*Upper*
Included				
Constant	7.85 . (4.13)			
ANT Response Time	-0.01 * (0.01)	0.97	0.99	1.00
RSPAN	0.02 (0.05)	0.92	1.02	1.12

Note. R^2= .19 (Hosmer-Lemeshow), .25 (Cox-Snell), .32 (Nagelkerke). Model $\chi 2$ (2) = 7.17, $p < .05$. **p* *< .05.*, . *p* < .1.

Table 12

Block 3 binomial logistic regression results with musicians removed. N=28.

	B (SE)	*95% CI for odds ratio*		
		Lower	*Odds ratio*	*Upper*
Included				
Constant	-7.83 . (4.07)			
ANT Response Time	0.01 (0.01)	1	1.01	1.02
RSPAN	0.16 * (0.07)	1.05	1.17	1.38

Note. R^2= .24 (Hosmer-Lemeshow), .31 (Cox-Snell), .39 (Nagelkerke). Model $\chi 2$ (2) = 9.22, $p < .01$. **p* *< .05.*, . *p* < .1.

Table 13

Block 4 binomial logistic regression results with musicians removed. N=28.

	B (SE)	95% CI for odds ratio		
		Lower	Odds ratio	Upper
Included				
Constant	0.56 (3.22)			
ANT Response Time	0.00 (0.00)	0.99	1	1.01
RSPAN	0.04 (0.05)	0.95	1.04	1.14

Note. R^2= .04 (Hosmer-Lemeshow), .06 (Cox-Snell), .07 (Nagelkerke). Model $\chi2$ (2) = 1.49, $p > .1$.

Table 14

Block 5 binomial logistic regression results with musicians removed. N=28.

	B (SE)	95% CI for odds ratio		
		Lower	Odds ratio	Upper
Included				
Constant	0.14 (3.01)			
ANT Response Time	0.00 (0.00)	0.99	1	1.01
RSPAN	0.00 (0.04)	0.91	1	1.08

Note. R^2= .00 (Hosmer-Lemeshow), .00 (Cox-Snell), .00 (Nagelkerke). Model $\chi2$ (2) = 0.02, $p > .1$.

Table 15

Overall binomial logistic regression results with musicians removed. N=28.

	B (SE)	*95% CI for odds ratio*		
		Lower	*Odds ratio*	*Upper*
Included				
Constant	0.56 (3.07)			
ANT Response Time	0.00 (-0.49)	0.99	1	1.01
RSPAN	0.02 (0.04)	0.94	1.02	1.12

Note. R^2= .01 (Hosmer-Lemeshow), .02 (Cox-Snell), .03 (Nagelkerke). Model $\chi 2$ (2) = 0.55, $p > .1$.

Discussion

ID functions for all three phoneme categories collapsed over the five experimental blocks were characterized by a sharply defined category boundary between voiced /b/ and unvoiced aspirated /p/ ID functions that contrasted with a more gradual transition from the /b/ to /p^h/ ID functions. These ID functions were similar to the ID functions obtained in Experiment 1 of Pisoni et al. (1982) and Schoenherr and Logan (2013, 2014). They demonstrated that participants could accurately categorize stimuli in the unvoiced /p/ region of the VOT continuum. The smaller distinction between the prevoiced /p^h/ and voiced /b/ ID functions suggests that participants could generally identify stimuli in the prevoiced VOT region from stimuli in the voiced VOT region but that this distinction was far less reliable. Furthermore, the category boundary between the prevoiced and voiced regions was characterized by a much more gradual transition than the voiced-unvoiced category boundary. See Figure 12 for a side-by-side comparison of the current results to those from Experiment 1 of Pisoni et al. (1982).

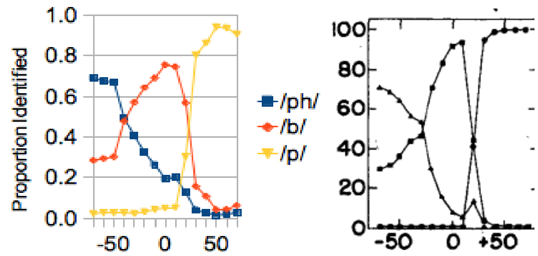


Figure 12. Overall ID functions for prevoiced (blue), voiced (red), and unvoiced (yellow) categories obtained from the current experiment (left-hand panel). Three-category ID functions from Experiment 1 of Pisoni et al. (1982) (right-hand panel).

As expected, the stimuli closest to the category exemplars presented during training were most accurately identified by participants, even for stimuli in the prevoiced region from -70 ms to -20 ms VOT. The slope of the prevoiced and voiced ID functions was less steep than the slope of the unvoiced ID function, suggesting that participants attended to acoustic-level stimulus properties (i.e., VOT) as well as more abstract phonemic properties in order to categorize the prevoiced stimuli. A plateau in the prevoiced ID response curve between -70 and -50 ms suggests that some participants were able to attend to phonemic-level stimulus properties in order to categorize the prevoiced stimuli.

Also consistent with Pisoni et al. (1982), about half of the participants could reliably and accurately identify stimuli in the prevoiced VOT region. Moreover, contrary to what I expected, participants' ID performance did not vary substantially across blocks. The cognitive processes underlying an individual's ability to learn to perceive a nonnative phoneme may be somewhat resistant to change over the course of such a short experimental session.

The results did not provide a clear picture of the cognitive mechanisms underlying the individual differences observed in ID performance. Results from the ANT and RSPAN did not consistently predict ID performance. There appeared to be a relationship between ID performance and both ANT average response time, as well as RSPAN scores. ANT response time significantly predicted ID performance in block 1 in the logistic regression analysis that included musicians, and in block 2 in the analysis that did not include musicians, and moderately predicted ID performance in block 2 in the first analysis and block 1 in the second analysis. Participants who respond quicker may categorize stimuli based on more implicit cognitive processes that utilize acoustic-level cues more effectively than explicit, rule-based processes, and this in turn may result in a better ability to identify prevoiced stimuli, especially for early blocks.

The complete lack of relationship between prevoiced ID functions and individual differences in the alerting, orienting, and executive control of attention as measured by the ANT suggests that either these constructs do not underly perceptual learning of nonnative phonemes or that the ANT is an inappropriate conceptualization of attentional dimensions for predicting the cognitive processes underlying phoneme learning. An alternative to attentional networks is attentional control theory which traditionally focuses on the relationship between anxiety and performance on cognitive tasks. However, it may be a useful conceptualization of attention for future studies of perceptual learning related to speech perception (Eysenck, Derakshan, Santos, & Calvo, 2007). Attentional control theory holds that attentional control underlying processing efficiency for a given cognitive task is associated with three attentional functions: inhibition, shifting, and updating. Inhibition comprises one's ability to intentionally inhibit automatic or dominant responses that interfere with task completion (Eysenck et al., 2007). The inhibition function of Attentional Control Theory could conceivably facilitate the inhibition of native phonological system required in order to accurately perceive nonnative phonemes. Shifting defines the cognitive process used to allocate attention toward task relevant stimuli. This could be related to an individual's ability to shift from acoustic-level to phoneme-level processing in order to categorize novel speech sounds. Finally, the updating function is associated with the processes of monitoring and updating information in working memory. This could underly the accuracy of a comparison between novel phonemes and previously heard category exemplars. Although the ANT measures attentional efficiency of three distinct networks—alerting, orienting, and executive control of attention—an individual difference task that measures attentional functions as posited by attentional control theory may be worth exploring as a potential way to account for the cognitive processes that underlie perceptual learning of novel phonemes.

Visual inspection of the overall ID functions from the high and low RSPAN median split groups suggests that participants with a higher RSPAN were somewhat better at accurately identifying stimuli in the prevoiced region (see Figure 6), however, RSPAN was for the most part not a significant predictor of ID performance for stimuli in the prevoiced VOT region. Although working-memory capacity may underly nonnative phoneme learning in some way, either the sample size for this experiment may have been too small or the RSPAN task may not have been an ideal measure of working memory.

Two potential issues with way that the RSPAN implemented in the present study may have affected the results. Requiring participants to read sentences out loud may have affected recall ability differently for different participants. For individuals who are extremely proficient at reading out loud this part of the RSPAN may not inhibit letter recall as much as it would for participants who are not as proficient at reading out loud. In this situation the RSPAN is providing a measure of ability to read aloud that is confounded with the RSPAN as a measure of working-memory capacity. Another issue is that recall strategies may vary between participants. For example, some participants may utilize a repetition strategy, some may use mnemonics, and some may use no particular strategy. RSPAN score would therefore provide an indication which strategy works best rather than a measure of working memory capacity.

Another measure of working memory that has become popular, especially in neuroimaging studies, is the *n*-back task. An *n*-back task presents a sequence of a certain kind of stimulus such as letters or pictures, and requires a participant to continuously recognize whether the current item matches the item presented *n* items ago (Kane, Conway, Miura, & Colflesh, 2007). The *n*-back task has received little empirical validation as a measure of working memory compared to complex span tasks, however, an *n*-back *recall* task (as opposed to recognition) has been shown to be strongly correlated with results from a complex span task (Shelton, Metzger, & Elliott, 2007). Future studies on individual differences in working memory and nonnative speech perception could investigate whether the *n*-back can be used to better predict identification performance compared to complex span tasks such as the RSPAN.

Other processes that may be associated with learning to perceive a nonnative phoneme may be delayed gratification or impulse control. Delayed gratification is the ability to resist the temptation to indulge in an immediate reward based on the knowledge that this will lead to a larger reward at some point in the future. An individual's ability to delay gratification is associated with impulse control, which requires cognitive control. Thus, a measure of gratification delay, such as the Delaying Gratification Inventory (DGI) (Hoerger, Quirk, & Weed, 2011) could be administered in future phoneme learning studies to investigate the possible relationship between impulse control and a person's ability to inhibit the L1 phonological system during nonnative speech perception. Another measure that could be administered in order to measure the degree that individual differences in impulse control may be related to novel phoneme identification performance is a Go/No-go test. These tests are used to provide a measure of a participant's ability to sustain attention toward relevant stimuli and control responses. Usually a Go/No-go test involves pressing a button in response to a certain kind of stimulus, and refraining from pushing the same button in response to a different kind of stimulus. This process may be similar to the inhibition of the L1 phonological system when learning a novel L2 phoneme category, and thus a Go/No-go test may be a useful measure of individual differences in future studies.

In conclusion, the present study found limited support for the involvement of attention and working memory in the learning of a novel phoneme category as a function of block. Future work that examines individual differences in nonnative phoneme learning should consider alternative measures of attention and working memory to determine if the pattern of results obtained in the present study were due to limitations in the individual difference measures used. Additional participants may also yield a wider range of performance that may make it easier to see the role of individual differences in attention and working memory. Finally, the differences between musicians and nonmusicians deserves further examination. The possibility that becoming proficient at playing an instrument develops a skill that is transferable to learning a novel phoneme category implicates general auditory processing being recruited to deal with speech perception. Such a finding suggests that the modularity of mind championed by Fodor (1983) and others (e.g., Pinker, 1996) may be limited, especially as Fodor used speech perception as the model system to illustrate modularity.

References

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development, 52,* 1135-1145.

Barrouillet, P. (1996). Transitive inferences from set-inclusion relations an dworking memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1408-1422.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. (1998). *Psychological Review, 105*(1), 158-173.

Benton, S. L., Kraft, R. G., Glover, J. A., Plake, B. S. (1984). Cognitive capacity differences among writers. *Journal of Educational Psychology, 76,* 820-834.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: the transition from speech sounds to spoken words* (167-224). Cambridge, MA: MIT Press.

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* (167-200). York, UK: Timonium, MD.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception on approximants. *Journal of Phonetics, 20,* 305-330.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrastsvarying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109,* 775-794.

Best. C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance., 4,* 45-60.

Best, C. T., & Strange, W. (1992) Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305-330.

Bongaerts, T., Planken, T., & Schils, B. (1995). Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton & Z. Lengyel (Eds.) *The age factor in second language acquisition* (30-50). Clevedon, UK: Multilingual Matters.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3,* 219-252.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49,* 155-180.

Butterworth, B., Campbell, R., & Howard, D. (1986). The uses of short-term memory: A case study. *Quarterly Journal of Experimental Psychology, 38;* 705-737.

Calderon, J., & Best, C. T. (1996). Effects of bilingualism on the perception of nonnative consonant voicing contrasts. *Journal of the Acoustical Society of America., 99,* 2602.

Chandrasekaran, B., Koslov, S. R., & Maddox, W. T. (2014) Toward a dual-learning model of speech category learning. *Frontiers in Psychology,* Hypothesis and Theory Article. doi:10.3389/fpsyg.2014.00825

Clarkson-Smith, L., & Hartley, A. A. (1990). The game of bridge as an exercise in working memory and reasoning. *Journal of Gerontology, 45,* 233-238.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O. W., & Engle, R. W. (2005). *Psychonomic Bulletin & Review, 12*(5), 769-786.

Cox, D. R., & Snell, D. J. (1989). *The analysis of binary data* (2nd ed.). London: Chapman & Hall.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour, 19,* 450-466.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25,* 1-18.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: a meta-analysis. *Psychonomic Bulletin and Review, 3,* 422-433.

Dougherty, M. R. P., & Hunter, J. (2003). Probability judgment and subadditivity: the role of working memory capacity and constraining retrieval. *Memory and Cognition, 31,* 968-982.

Diehl, R. & Kluender, K. R. (1989). On th objects of speech perception. *Ecological Psychology, 1,* 1-45.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., et al. (2000). A neural basis for general intelligence. *Science, 289,* 457-460.

Eimas, P. D. (1978). Developmental aspects of speech perception. In R. Held, H. Leibowitz, & H. L. Teuber (Eds.) *Handbook of sensory physiology: VIII. Perception.* New York, NY: Springer-Verlag, 1978.

Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology, 51,* S1-S10.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.) *The psychology of learning and motivation, vol. 44* (145-199). New York, NY: Elsevier.

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research, 84,* 253-262.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics, 16,* 513-521.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the /r-l/ distinction by young infants. *Perception & Psychophysics, 18,* 341-347.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303-306.

Eysenck M. W., Derakshan, N., Santos, R., Calvo, M. G. (2007) Anxiety and cognitive performance: Attentional Control Theory. *Emotion, 7*(2), 336-353.

Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., & Posner, M. I. (2002). Cognitive and brain mechanisms of conflict. *Neuroimage, 18,* 42-57.

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *Neuroimage, 26,* 471-479.

Flege, J. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics, 15,* 47-65.

Flege, J. (1995). Second language speech learning: Theory, findings and problems. In Winifred Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233-277. Baltimore: York Press.

Flege, J. (2005). Origins and development of the Speech Learning Model. *1st ASA Workshop on L2 Speech Learning.* Lecture conducted from Simon Fraser Univ., Vancouver, BC. Slides retrieved from http://jimflege.com/files/Vancouver_April_2005.pdf.

Flege, J., & Eefting, W. (1987). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America, 83*(2), 729-740.

Flege, J., & Hammond, R. (1982). Non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition., 5,* 1-17.

Flege, J., & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition , 26,* 1-34.

Flege, J., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language & Speech, 24, 25-146.*

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics, 34,* 153-175.

Flege, J., Schirru, C., & MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40,* 467-491.

Fodor, J. (1983). The modularity of mind: An essay on faculty psychology. Cambridge, MA: MIT Press.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3-28.

Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics, 48*(6), 559-570.

Fox, J. (2005). The R commander: A basic-statistics graphical user interface to R. *Journal of Statistical Software, 14,* http://www.jstatsoft.org/

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 349-366.

Frieda, E. M., Walley, A. C., Flege, J. E., Sloane, M. E. (1999) Adults' perception of native and nonnative vowels: implications for the perceptual magnet effect.

Perception and Psychophysics, 61(3), 561-577.

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

Gorenflo, D., & McConnell, J. (1991). The most frequently cited journal articles and authors in introductory psychology textbooks. *Teaching of Psychology, 18,* 8-12. doi: 10.10370333-295X.101.2.195

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*(5), 2711-2724.

Haskins Laboratories (2014). Abramson/Lisker VOT Stimuli. Retrieved 05/12/2014. From http://www.haskins.yale.edu/featured/demo-liskabram/index.html.

Hayes, R. L. (2002). The perception of Japanese consonant length by non-native listeners. *Laboratory Phonology, 8,* New Haven, 2002. Poster.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research, 23*(1), 65-94.

Hoerger, M., Quirk, S. W., & Weed, N. C. (2011). Development and validation of the delaying gratification inventory. *Psychological Assessment, 23,* 725-738.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied Logistic Regression.* New York: Wiley.

Jusczyk, P. W. (1985). On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), *Neonate cognition: beyond the blooming, buzzing* *confusion* (199-229). Hillsdale, NJ: Lawrence Erlbaum.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual differences perspective. *Psychonomic Bulletin and Review, 9,* 637-671.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, J. H. (2007). Working memory, attention control, and the N-Back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 615-622.

Kiewra, K. A., & Benton, S. L. (1988). The relationship between information processing ability and notetaking. *Contemporary Educational Psychology, 13,* 33-44.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language, 30,* 580-602.

Kluender, K. (1994). Speech perception as a tractable problem in cognitive science. In M. A. Gernsbacher (Ed.) *Handbook of Psycholinguistics*. San Diego, CA: Academic.

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

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.) *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (121-154). Baltimore, MD: York Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age.

Science, 255, 606-608.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389-433.

Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences, 2,* 129-160.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development, 18,* 459-475.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243,* 489-494.

Liberman, A. M., Cooper, F. S., Shankweiler, D. S., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 421-461.

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

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

Lisker, L., & Abramson, A. S. (1967). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences.* Prague: Academia.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89,* 874-886.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2,* 369-390.

MacMillan, N. A., & Creelman, D. (1991). *Detection Theory: A User's Guide.* New York, NY: Cambridge.

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

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86,* B33-B42.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. doi: 10.1037/h0042158

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fijimura, O. (1975). An effect of linguistic experience: the discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception and Psychophysics, 18,* 331-340.

Negelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78,* 691-692.

Pallier, C., Bosch, L., & Sebastian-Galles, N. (1997). A limit on behavioural plasticity in speech perception. *Cognition, 64,* 9-17.

Pinker, S. (1994). The language instinct. New York: Harper Perennial Modern Classics.

Pisoni, D. B., Aslin, R. N., Percy, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 297-314.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics, 15:* 285-290.

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

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics, 52,* 37-52.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353-363.

Posner, M. I., & Rothbar, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology, 58,* 1-23.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Redick, T. S., & Engle, R. W. (2006). Working memory capacity and attention network test performance. *Applied Cognitive Psychology, 20,* 713-721.

Ritz, C., & Streibig. (2005). Bioassay Analysis using R. *Journal of Statistical Software, 12,* http://www.jstatsoft.org/

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7,* 532-547.

Schoenherr, J. R., & Logan, J. (2014). Attentional and immediate memory capacity limitations in the acquisition of non-native linguistic contrasts. (2014). *In Press.*

Shallice, T., & Butterworth, B. (1977). Short-term memory impairment and spontaneous speech. *Neuropsychologia, 15,* 729-735.

Shelton, J. A., Metzger, R. L., & Elliot, E. M. (2007). A group- administered lag task as a measure of working memory. *Behavior Research Methods, 39,* 482-493.

Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics, 3:* 215:233.

Swan, K., Myers, E., (2013). Category labels induce boundary-dependent perceptual warping in learned speech categories. *Second Language Research, 29*(4), 391-411. doi:10.1177/0267658313491763

Swoboda, P. J., Kass, J., Morse, P. A., & Leavitt, L. A. (1978). Memory factors in vowel discrimination of normal and at-risk infants. *Child Development, 49,* 332-339.

Takagi, N., & Mann, V. A. (1995). Signal detection modelling of Japanese listeners' /r/-/l/ labelling behaviour in a one-interval identification task. *Journal of the Acoustical Society of America, 97,* 563-574.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127-154.

Unsworth, N., Heitz, R. P, Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498-505.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1202-1321.

Vallar, G., & Shallice, T. (1990). *Neuropsychological impairments of short-term memory.* Cambridge, England: Cambridge University Press.

Werker, J. F. (1994). Cross-language speech perception: developmental change does not involve loss. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: the transition from speech sounds to spoken words* (167-224). Cambridge, MA: MIT Press.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37,* 35-44.

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: models, research, implications* (285-311). York, UK: York Press.

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

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52,* 349-355.

Yamada, R. A. & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception and*

Psychophysics, 52, 376-392.

Yeni-Komshian, G., Robbins, M., & Flege, J. (2001). Effects of word class differences on L2 pronunciation accuracy. *Applied Psycholinguistics, 22,* 283-299.

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition, 113,* 234-243.

Appendix A

Auditory Experience Questionnaire

Language experience

Please indicate your date of birth: ___/___/______

Your place of birth:_____________________

1. Please list, to the best of your knowledge, where you have lived during your life. Please list Country, Province, and City and at what age you were when you lived there:

Age	Country	Province	City
_______	______________________	_____________	______________________
_______	______________________	_____________	______________________
_______	______________________	_____________	______________________
_______	______________________	_____________	______________________

2. While in elementary or primary school, did you receive any instruction in a language other than English?

_____Yes _____No

3. If Yes, which language? _________________________
If yes, approximately how many hours a day were you exposed to this language?

____

4. Did any parent/guardian/daycare provider have a first language other than Canadian English?

_____Yes _____No

5. If Yes, which language? _________________________

6. Did any parent/guardian/daycare provider speak a language other than Canadian English to you on a regular basis?

_____Yes _____No

7. If Yes, which language? ________________________

Musical Experience

1. Do you have any musical training? _____Yes _____No

2. State your musical training, including age and duration, and type of schooling

Age	Duration of training	Instrument	Schooling
_____	__________________	_____________	_________________________________
_____	__________________	_____________	_________________________________
_____	__________________	_____________	_________________________________
_____	__________________	_____________	_________________________________
_____	__________________	_____________	_________________________________
_____	__________________	_____________	_________________________________

3. Rate your musical skill level (circle answer)

beginner intermediate good excellent professional

4. What instrument(s) can you play? (List in order from best to worst).

____

5. How many hours per week on average do you play a musical instrument currently?

______

6. Do you come from a musical family? (Check best option)

______Yes, very much.

______Somewhat.

______Not at all.

7. How many hours per week on average do you listen to music?

____

8. How many hours per week on average do you listen to music without being involved in any other activities (ie: listening to music and not doing anything else)?

____

9. What type of music do you listen to most often? (Check best options)

_____Classical

_____Jazz

_____Rock

_____Pop

_____Folk

_____Other (please specify)

10. How often do you enjoy listening to music? (Check best option)

_____Always.

_____Most of the time.

_____Sometimes.

_____Almost never.