DEPOSIT AND CONSULTATION OF DISSERTATION

One copy of your dissertation will be deposited in ORA (Oxford University Research Archive), where it is intended to be freely available online. In order to facilitate this, you are requested to complete and sign the form below.

Please use block capitals

Surname GE
First names (in full) YUXIN
Faculty board EDUCATION
Degree name and pathway MASTER OF SCIENCE IN APPLIED LINGUISTICS AND SECOND LANGUAGE ACQUISITION
Title of dissertation EXPLORING SPEECH SOUND ACQUISITION IN ADULT LEARNERS UNDER A DISCRIMINATIVE LEARNING MODEL N.B. The title stated here must be precisely the same as that stated on the title page of the thesis submitted. **A candidate wishing to amend the title previously approved by the faculty must apply to the faculty board for permission to do so.**
Supervisor

ELIZABETH WONNACOTT

Subject keywords *Enter your own keywords or phrases to describe your work. This information helps us describe your work in ORA*

DISCRIMINATIVE LEARNING; ERROR-DRIVEN LEARNING; SPEECH SOUND ACQUISITION; SECOND LANGUAGE ACQUISITION

Research methods used *This information helps us describe your work on SOLO for future students*

e.g. quantitative, interviews, vocabulary test, systematic review, etc.

QUANTITATIVE; ONLINE BEHAVIOURAL EXPERIMENT; ARTIFICIAL LANGUAGE LEARNING

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I understand that the Department requires that I shall deposit one copy of my dissertation in the Oxford University Research Archive (‘ORA’) where it shall be freely available online for use in accordance with ORA’s Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals, and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [www.bodleian.ox.ac.uk/ora/deposit-in-ora/deposit-licence] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.

5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]: Yuxin Ge

Date: 25/10/2021

Exploring speech sound acquisition in adult learners under a discriminative learning model

Yuxin Ge
Lady Margaret Hall
University of Oxford

A dissertation submitted in partial fulfilment of the degree of Master of Science

in *Applied Linguistics and Second Language Acquisition*

August 2021

Acknowledgements

I would like to thank my supervisors, Liz Wonnacott and Eva Viviani, for their invaluable support and guidance during the completion of my master's study, and their patience in helping me deal with any problems throughout this research project. Their passion and dedication to research work have inspired me in academic and personal life, and working with them would always be one of my best experiences at Oxford.

I want to also thank my ALSLA friends, though some of you I have never met in person due to the pandemic. It was a great pleasure talking to you, getting to know your amazing research ideas and participating in some of your experiments. Most importantly, congratulations for your well-deserved success at this master's study and best wishes for your future adventures!

Finally, I could never thank my families enough for their love, understanding and encouragement. Without them it would be impossible for me to complete my study. Words are certainly not enough to express my gratitude and love. Thank you mom and dad.

Abstract

How do we learn the sound systems of different languages from the complex speech input around us? One influential perspective is that listeners can track the statistical distribution of acoustic features in speech to promote learning. However, although strong evidence has been found for the role of input statistics in phonological development, the underlying learning mechanisms that drive statistical learning remain unclear. The current study investigates this issue under a discriminative model of language learning originated from animal learning research, which suggests that learning is driven by prediction error and cue competition. Specifically, the model proposes that a learning environment that allows for predictions to be made on the basis of the available linguistic cues leads to better learning, since learners can compare their predictions with actual outcomes and if there are discrepancies between the two, they can learn from the prediction error.

The predictions of the discriminative learning model are tested in two artificial language learning experiments with adult native speakers of Chinese. The artificial language was designed to contain a geminate consonant cue which is used contrastively (e.g. *fato* and *fatto* have different meanings) but is non-native (and hence less salient) to Chinese speakers, as well as a tonal cue that Chinese speakers are familiar with but is not used to contrast lexical meanings. Subjects learned words in the artificial language either in a discriminative (cue-outcome) order or a non-discriminative (outcome-cue) order. In the discriminative order, learners heard words immediately before seeing the referents, which enables them to make predictions about the referent (the outcome) on the basis of the acoustic features of the words (the cues). This further allows learners to adjust their expectations when a predicted referent does not occur after particular cues and learn from such prediction error. On the other hand, the non-discriminative order had referents being viewed before hearing the words, and hence there is no opportunity to make predictions about

referents and participants cannot learn from cue competition and prediction error. The predicted benefit of the discriminative order was not found in any tests of Experiment 1, however in this experiment there was limited exposure during training. Experiment 2 thus replicated the first experiment with increased exposure. There was tentative evidence in the generalization test that learners in the discriminative order condition had learned that tones were uninformative in the language, which is in line with the predictions, but no evidence that learners had learned the geminate consonant cue. Taken together, this study generally supports the discriminative approach to speech sound learning with adult learners. It suggests potential directions for future theoretical language learning research which may have further implications for the design of language teaching programs.

Contents

1. Introduction

The acquisition of a sound system represents the very first stage of language development for infants learning their native languages in the first year of life. Understanding the cognitive process underlying speech sound acquisition is an important topic of research. It has the potential to explain how children learn native phonology as rapidly as they do, and may facilitate phonology pedagogy in foreign language classes as the accurate perception and production of foreign speech sounds usually pose difficulty for adult language learners. This dissertation project examines the nature of speech sound acquisition, and in particular, non-native sound acquisition, under a discriminative learning model which suggests that language learning is a predictive process and is driven by prediction error and competition between available cues in the environment (e.g., Ramscar, Dye, & McCauley, 2013; Ramscar, Yarlett, Dye, Denny & Thorpe, 2010). Specifically, adult native Chinese speakers were trained with an artificial language which contained a non-native consonant length cue and a familiar tonal cue. The key experimental manipulation is that the non-native consonant cue is less salient to Chinese speakers but is designed to be the informative (or discriminatory) cue that contrast lexical meanings; whereas the more salient tonal cue is unreliable (or non-discriminatory) in discriminating lexical items. Under a discriminative learning account, the informative consonant cue can be learned through cue competition, which is predicted to occur only when stimuli are ordered in a particular way (with cues preceding outcomes). Two artificial language learning experiments were conducted to explore this prediction.

In this dissertation, I start by reviewing two separate but interrelated literature on speech sound acquisition (Section 2.1) and on the error-driven, discriminative learning approach (Section 2.2). In Section 2.1, I present and compare the characteristics of speech sound perception and acquisition in infancy and in adulthood (Section 2.1.1). I then provide evidence that phonological development for both infant and adult learners reflects the probabilistic distribution of acoustic features in the target language (Section 2.1.2). In Section 2.2, I review the developmental history of an error-driven learning approach, starting from its origin in the area of animal conditioning research (Section 2.2.1) and extending to its application in human learning (Section 2.2.2) and language learning (Section 2.2.3-2.2.4). In Section 2.3, the central ideas of the current study are presented. Sections 3-5 include the details of the pilot and the two language learning experiments respectively. Finally, I provide general discussions on the overall findings, as well as limitations of the current design and implications for future research (Section 6).

2. Literature review

2.1 Speech sound perception and acquisition

Speech sound acquisition is a complex process. Learners need to perceive the acoustic features in speech stream and discriminate between sounds accurately, and to find out which sets of acoustic features are relevant in the target language, given the numerous features available in speech. Additionally, they have to acquire the specific positions of speech sounds (phonotactics) and the prosodic patterns. Learners must also be able to segment continuous speech stream into meaningful units (e.g. words and phrases) and map words they hear to real-world referents in later stages of language acquisition. The focus of this dissertation is on the acquisition of speech segments, and hence I will mainly discuss literature on this aspect of phonological development in the review (though I may touch on other aspects such as prosodic development and word segmentation to some extent). Extensive research has explored the development of infants' native phonological abilities (e.g., Chládková & Paillereau, 2020; Jusczyk, 1995; Kuhl, 1979; Werker & Tees, 1984) and adults' non-native phonological skills (e.g., Best, McRoberts & Sithole, 1988; Iverson et al., 2003; Polka, 1991; Schertz, Cho, Lotto, & Warner, 2015). In this section, I start by introducing the perceptual sensitivity to native and non-native sounds by infants and adults, and illustrate the effect of linguistic experience on speech sound perception and discrimination. I then discuss a statistical (or probabilistic) learning perspective of speech perception and acquisition, which suggests that learners follow the probabilistic distribution of acoustic features in the target language to acquire speech sounds.

2.1.1 Native versus non-native sound perception and acquisition - the effect of linguistic experience

One notable finding from infant discrimination research is that in the first few months of life, infants' sensitivity to speech sound contrasts is relatively universal and not limited to their native languages, but this language-general discrimination ability declines during the first year of life. (e.g., Polka & Werker, 1994; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005; Tsushima et al., 1994; Werker & Tees, 1984). A classic study by Werker and Tees (1984) employed a conditioned head turn paradigm to explore infants' speech sound discrimination ability, where infants were trained to turn their heads in response to a change in the speech sounds. The results indicated that English-native infants of 6-8 months old were able to detect the non-native (Hindi and Salish) consonants, but by 8-10 months old fewer infants can discriminate the same non-native contrasts. By 10-12 months, performance was even poorer and was comparable to young children and adults. In addition, Polka and Werker (1994) have found similar findings for vowels, although decline in discrimination of non-native vowels occurs slightly earlier (i.e. by 6-8 months). However, the reorganization of speech sound discrimination ability during the first year of life has been found to be more complex than simply loosing sensitivities to all non-native sounds. There are non-native contrasts that remain discriminable across the developmental process (e.g., Best, McRoberts, & Sithole, 1988; Polka & Bohn, 1996). For example, Best, McRoberts and Sithole (1988) reported that 12-14-month-old infants whose native language was English were able to distinguish non-native click consonants in some African languages and this discrimination ability continued into adulthood. Moreover, discrimination of some native contrasts may improve with increased linguistic experience (e.g., Kuhl, Stevens, Hayashi, Deguchi, Kiritani & Iverson, 2006; Narayan, Werker, & Beddor, 2010; Tsao, Liu, & Kuhl, 2006). Narayan et al. (2010) found that the discrimination of native [na] and [ŋa] contrast was difficult for Filipino-learning infants of 6-8 months old, but by 10-12 months infants can discriminate this native contrast. Overall, although the developmental pathways for some specific sounds could be different, a general pattern for a number of speech sounds examined is that infants' sensitivity to non-native sounds diminishes during the first year of life as exposure to native sounds increases.

This language-specific discrimination of speech sounds by the end of early infancy continues into adulthood, making adults less sensitive to non-native acoustic features that are not used (or are used differently) to contrast meanings in their native languages. For example, it has been found that Japanese adult listeners often have difficulty discriminating the English approximants /ɹ/ and /l/ (e.g., Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Goto, 1971; Iverson et al., 2003; Miyawaki et al., 1975). Goto's (1971) early research showed that minimal pairs that differed only in /ɹ/ and /l/ consonants (e.g. light versus right) posed difficulty for Japanese listeners and they were inaccurate in discriminating such words compared to English-native listeners. Other studies have observed similar discrimination difficulty for adult listeners with non-native sounds that have different VOT patterns (e.g., Abramson & Lisker, 1970), place and manner of articulation (e.g., Trehub, 1976; Werker & Tees, 1984), and prosodic representations (e.g., So, 2005; Wayland & Guion, 2003) from their native speech sounds.

However, not all non-native contrasts pose the same degree of difficulty for all adult listeners, and the relative ease of discriminating non-native sounds varies largely with the listeners' linguistic background. For instance, Best, Traill, Carter, David Harrison and Faber (2003) provided evidence that the perception and discrimination of click consonants from an African language, !Xóõ, by non-native listeners depended on the similarity/dissimilarity between the target language and listeners' native languages. In this case, English-native listeners discriminated the click contrasts better than African listeners whose native languages contained other clicks, as African listeners were distracted by their native click categories. This linguistic experience effect on non-native sound discrimination has also been observed for vowels (e.g., Best, Hallé, Bohn, & Faber, 2003), and tonal patterns (e.g., Hallé, Chang, & Best, 2004; So, 2005). These findings suggest that listeners tend to assimilate non-native sounds to native ones if there is similarity between the two languages, which may eventually impede accurate perception of non-native sounds. It also accounts for the perceptual difficulty of /ɹ/ and /l/ by Japanese listeners, as both sounds are assimilated to a single Japanese approximant category. In contrast, if the non-native features are totally missing in listeners' native language (e.g. clicks are absent in English), they cannot be assimilated to any native sounds and the discrimination ability is preserved (Best, McRoberts & Sithole, 1988).

Such linguistic experience (i.e. native language) effect on adults' speech sound perception leaves the learning of the sound system of a second/foreign language challenging. A related question is whether and to what extent L2 learners' discrimination of non-native sounds improves after learning. Research on this topic has found that it depends both on the amount of L2 experience and the degree of initial perceptual difficulty (e.g., Best & Strange, 1992; Dobel, Lagemann & Zwitserlood, 2009; Poltrock, Chen, Kwok, Cheung, & Nazzi, 2018). For some non-native sounds that initially pose significant difficulty for learners (e.g., /ɹ/ and /l/ for Japanese learners of English), perceptual improvement can occur after intensive experience with the non-native language, but perception may still be less well compared to native speakers (Best & Strange, 1992). For those contrasts that pose less difficulty for L2 learners from the beginning, linguistic exposure tends to be a less vital factor (Best & Tyler, 2007).

To summarize, there is good evidence that humans are born to be sensitive to most speech sounds in languages, but this discrimination ability is reduced to native sounds several months after birth. As a result, adult learners of a second/foreign language usually experience difficulty in perceiving and learning the non-native acoustic features, though the degree of difficulty is largely influenced by their L1 and L2 linguistic experience. In the following section, I will introduce a theoretical approach that accounts for the above-discussed speech sound perception and acquisition patterns from a probabilistic perspective.

2.1.2 Tracking the statistical distribution in speech sound acquisition

After discussing the general patterns of phonological development, the question now turns to how the speech sound system is acquired. Theoretical accounts can be seen as sitting in the broader perspectives of nativist and constructivist/input-driven approaches. Nativist approaches basically claim that humans are born with a set of rules or grammar that operates at all levels (from phonemic to syntactic levels) and helps learners generate grammatical words, phrases and sentences (Chomsky, 1957, 1965). In the domain of speech sound learning, there are various types of innate phonological rules and knowledge (e.g. the innate distinctive features, Ambridge & Lieven, 2011, Chapter 2; the Optimality Theory, Prince & Smolensky, 1993, 2004) that have been proposed. On the other hand, constructivists argue that language learning is driven by input and no innate knowledge is assumed. Extending this input-driven view to the phonological acquisition area, speech sound perception and discrimination do not depend on an innate set of distinctive features (as suggested by nativist accounts), but emerge from exposure to acoustic input of the target language (e.g., Mielke, 2008; Pierrehumbert, 2003).

One influential distributional learning account suggests that learners (both children and adults) are able to track the probabilistic distribution of linguistic features in the massive input and figure out which features are relevant in the particular language. Following this account, speech sound acquisition results from learners capturing the distribution of acoustic cues in speech stream, acquiring the cues that discriminate meanings in the language and ignoring other uninformative cues. Such distributional learning also play a role in other aspects of phonological development, including word segmentation (e.g., Hay, Pelucchi, Estes, & Saffran, 2011; Saffran, Aslin & Newport, 1996; Saffran & Wilson, 2010) and phonotactic acquisition (e.g., Chambers, Onishi & Fisher, 2003; Jusczyk, Luce, & Charles-Luce, 1994), but the current review will focus on the learning of speech categories. Note that although this distributional learning account emphasizes on the role of input, it is not entirely incompatible with the nativist approaches. Rather, probabilistic tracking is seen as an extra constraint that is employed in addition to innate knowledge and takes a less prominent position according to nativists. Evidence from children's and adults' speech sound learning are presented to show that probabilistic distribution is indeed a key in speech sound acquisition and shapes the specialized perceptual sensitivity to native sounds.

Evidence from children's language learning

Young children and infants acquire the sound systems of their native languages rapidly and with high accuracy. Over the past few decades, various studies have examined the role of distributional information of acoustic features in children's speech sound acquisition (e.g., Maye, Weiss & Aslin, 2008; Maye, Werker & Gerken, 2002). In other words, the question of interest is whether children keep track of the structural regularities in the acoustic input to aid acquisition. Maye, Werker and Gerken (2002) examined this question by testing how the differences in the distribution of phonetic details in input influence the perception of phonemic categories by infants of 6 and 8 months old. The researchers created a continuum of sounds, moving from one endpoint of voiced unaspirated /da/ to the other endpoint of voiceless unaspirated /ta/, by manipulating the prevoicing and formant frequency dimensions. The artificial stimuli were then presented to infants with either a bimodal or an unimodal distribution. Specifically, one group of infants heard sounds towards the two endpoints of the continuum more frequently, mimicking situation in a natural language where there are two speech categories (bimodal); whereas the other group encountered more sounds at the center of the continuum, mimicking situation where there is just one category (unimodal). This difference in the frequency distribution of input led to different performances in a following da/ta discrimination test. The bimodal group did better than the unimodal group in discriminating the voicing contrasts, since the unimodal group tended to form only one category for the voicing sounds but the bimodal group had two categories. It was consistent with the prediction that infants are able to learn from input distribution and form phonological categories accordingly. The idea of manipulating the structural distribution of input was employed in later research and similar distributional effect on children's phonemic perception and discrimination has been found (Maye, Weiss & Aslin, 2008). The findings from children's language learning research suggest that tracking the statistical distribution of phonetic variations in the input plays a role in the rapid phonological acquisition in infancy, as experiments have shown that exposure of a few minutes can significantly impact children's responses to speech sounds.

Evidence from adults' language learning

Similar to infant studies mentioned above, distributional learning is also observed in adults who have already constructed a sophisticated native sound system and are learning non-native speech categories. For example, Maye and Gerken (2000) employed the same bimodal/unimodal manipulation as discussed above (Maye, Werker & Gerken, 2002) with adult learners. Participants were native speakers of English and they could all initially perceive the voiced /d/ and voiceless unaspirated /t/ as these sounds occur in English. However, the two sounds do not mark different phonemic categories in English since they occur in different environments (i.e. unaspirated /t/ is used after /s/ to form /st/ clusters but voiced /d/ never in this environment) and English speakers categorize both sounds as /d/ (Pegg & Werker, 1997). A continuum of transition from voiced /d/ to unaspirated /t/ was created and participants were presented with either a bimodal or unimodal distribution of input sounds. The bimodal group performed better in distinguishing the /d/ versus /t/ contrasts than the unimodal group after training, indicating that English-speaking adults formed a two-way category for voiced /d/ and unaspirated /t/ after exposure to the bimodal distribution, but the unimodal distribution led to a single category of the sounds. Thus, a brief distributional learning period could influence adult learners' sensitivity to speech sound categories.

The finding was replicated by Pajak and Levy (2011) with consonant length contrasts. Sonorants (/j/, /l/, /m/, /n/) and obstruents (/s/, /f/, /θ/, /ʃ/) were recorded as long consonants (e.g. [ajja], [assa]) and the consonant lengths were manipulated to form a continuum of short to long consonants. Similar to previous designs (Maye & Gerken, 2000; Maye, Werker & Gerken, 2002), adult learners were trained either with a bimodal input, hearing the long and short consonants towards the two ends of the continuum more often, or with a unimodal distribution that includes more sounds at the middle of the continuum. The results were partially as predicted: the bimodal group trained with sonorants performed better at discriminating short versus long sonorants than the unimodal group, and they could generalize the length contrast to other untrained sonorants. However, for the groups exposed to obstruents, no such distributional effect was observed, which as noted by the authors, could be attributed to the differences in the range of sonorant length and obstruent length in participants' native language.

Adults' language learning from statistical distributions is not limited to consonantal contrast acquisition. Studies that explored the acquisition of vowel contrasts show similar bimodal/unimodal distributional effects (e.g., Escudero, Benders & Wanrooij, 2011; Gulian, Escudero, & Boersma, 2007). Escudero et al. (2011) recruited Spanish-speaking adult participants who were learning Dutch and who had difficulty contrasting the Dutch /ɑ/ vs /aː/ sounds, and examined whether their categorization of the vowels improved after a bimodal training that emphasized the spectral quality differences between vowels. As predicted, a short period of training with the bimodal distribution of phonetic features helped learners categorize the sounds more accurately.

To summarize, both adult and infant language learning studies have shown the influence of probabilistic distribution of input. For adult learners, this allows them to learn features that are not used in their native languages as well as to ignore features in their native languages that are not present in the target language.

Limitations of distributional speech sound learning

Although probabilistic tracking has been widely found in child and adult speech sound learning, there are evidence that distributional learning may not lead to equally successful language acquisition in all situations (e.g., Terry, Ong, & Escudero, 2015; Wanrooij, Boersma, & Benders, 2015; Werker, Yeung, & Yoshida, 2012; Yoshida, Pons, Maye, & Werker, 2010).

As Werker et al. (2012) noted, the effect of distributional learning may vary with age and the amount of exposure to the distributional input. Infants of as young as 6-8 months could learn the distribution of phonological categories (bimodal or unimodal) after a 2.3-minute exposure (Maye et al., 2002), but 10-month-olds were found to require doubled exposure to show similar distributional learning (Yoshida et al., 2010) and adults even longer (Maye & Gerken, 2000). This indicates that the ability to learn from input statistics may decline with age, as learners become more language-specialized listeners. It also leaves out the possibility that older language learners may rely more on other methods to learn speech sounds. For example, they may make use of different aspects of linguistic knowledge (e.g. syntactic, semantic information) to aid sound learning rather than focusing on the structure of acoustic input only.

Another problem that arises from distributional learning research is that not all speech contrasts can be acquired equally well after exposure to structural input (e.g., Cristià, McGuire, Seidl & Francis, 2011). For example, Cristià et al. (2011) found that English-native infants were able to learn the Polish retroflex fricative category after exposure to structured speech stimuli, but their sensitivity to the alveolar palatal category did not improve after distributional learning. This difference in learning outcomes suggests that the relative saliency (or difficulty) of speech contrasts before training can affect probabilistic tracking in speech sound acquisition. Additionally, most existing research used artificially manipulated sound stimuli (e.g., Maye et al., 2002), creating relatively arbitrary steps along continua of acoustic dimensions (e.g. eight steps on the voicing dimension in Maye's studies). Further research needs to explore whether learners can cope with more complex probabilistic distributions in natural speech input.

Overall, existing evidence has provided support for the tracking of distributional information in phonological acquisition, but this account has limitations in accounting for speech sound learning among different age groups and in making precise predictions about which sounds can be learned and which can not. And importantly, it is not clear which mechanisms underpin such probabilistic tracking in speech sound learning. In the following section, I will introduce a theoretical approach that has the potential to account for the underlying learning mechanisms, the *discriminative learning approach*.

2.2 A discriminative learning approach to language acquisition

The core principles of the discriminative learning approach derive from learning theory in psychology and behaviourist research, and can be traced back to early studies of animal learning (e.g., Kamin, 1968; Rescorla, 1968). It has then been put forward into the area of language learning (e.g., Ramscar, Yarlett, Dye, Denny &

Thorpe, 2010). The basic idea is that learning arises from reducing learners' uncertainty about the environment, in other words, eliminating the discrepancies between learners' expectations and the actual observations (Ramscar et al., 2010). This uncertainty reduction learning process is driven by prediction error and cue competition, the mechanisms which have long been observed in animal learning studies (Kamin, 1968; Rescorla, 1968; Rescorla & Wagner, 1972). Therefore, it is necessary to first introduce the related animal studies to illustrate the fundamental principles of the discriminative learning model, before going into details about the application of this model in human language learning.

2.2.1 The origin - learning theory and the Rescorla-Wagner model

One famous animal learning observation was described by Ivan Pavlov, namely, if a dog is provided with food when a bell rings, saliva will be produced upon hearing the bell ring even without food given in the future (Pavlov, 1927). This salivary production phenomenon was once explained in terms of animals associating the bell ring (the *conditioned stimulus*, CS) with food (the *unconditioned stimulus*, US) based on the co-occurrence of the two stimuli. However, Rescorla explicitly argued in his (1988) paper that this explanation is a misunderstanding of animal's conditioning or learning process and it oversimplifies the mechanisms of learning. Instead, it was suggested that the learning process involves making predictions and using prediction error to distinguish informative cues (or CS) from uninformative cues (Kamin, 1968; Rescorla, 1968, 1988; Wagner, 1969). This proposal is supported by extensive evidence from animal studies (e.g., Annau & Kamin, 1961; Egger & Miller, 1962; Kamin, 1968; Rescorla, 1968; Wagner, 1969; Wagner, Logan & Haberlandt, 1968) and the key principles are discussed below.

Firstly, learning is error-driven and only occurs when predictions are violated (i.e. when there is prediction error). Early evidence for this came from Kamin's (1968) conditioning experiments that examined fear responses of rats. The rats were originally trained to press a bar for food, and once they had established a ‘bar-pressing for food’ pattern, they were trained with a combination of two cues (CSs), a white noise plus a light, followed by an electric shock (US). One group of rats was pre-exposed to one of the CSs (for example, noise) followed by electric shock before experiencing the combined CS (noise + light), whereas another group did not receive this pre-exposure. During the test, the rats were presented with the CS that was not presented in the pre-exposure session (in this case, light) and their responses to this cue were observed. The assumption was that the rats should elicit fear responses if they have learned the CS and hence reduce the bar-pressing frequency during the presentation of the CS as they are in fear of being shocked. The results suggested that the pre-exposed group did not show fear responses to the cue that they were not pre-exposed to (light), but the group without pre-exposure to noise did show fear responses to light alone. This finding was described as a *blocking* effect. It indicates that learning only occurs when a cue provides new information and the learners’ expectations (or predictions) about the outcomes are not fully met (Rescorla & Wagner, 1972). In this experiment, the light cue provided no additional information for the pre-exposed group as noise already fully predicted the shock, and hence there was no prediction error and the light cue was not learned. Such results would be unexplainable from a co-occurrence perspective, since for both groups (with or without pre-exposure) the light cue occurred with the shock for the same amount of trials, which should have led to similar learning outcomes.

Secondly, learning arises from various cues competing for predicting outcomes (i.e. cue competition). There are findings that the predictive value of a particular cue is influenced by other co-occurring or competing cues (e.g., Wagner, 1969). For instance, in Wagner’s (1969) study, three groups of rabbits were exposed to a tone+light CS which was followed by a shock to the eye area, leading to blinks that count as fear responses. One of the groups received additional interspersed trials where tone alone was followed by shock (the tone-shock group), and another group experienced tone alone with no shock (the tone-null group). The third group acted as a baseline comparison and received tone+light CS only. At test, the rabbits in the tone-shock group exhibited fewer fear responses to light alone compared to the baseline group, whereas the tone-null group showed greater fear responses to light alone even though they had never encountered light-shock pairings before. These findings provide evidence for cue competition. To be specific, for the tone-shock group, each time the tone alone was followed by the shock, the predictive value of the tone increased at the expense of the light, which led to weak conditioning or learning of the light cue. On the contrary, for the tone-null group, each time the expected shock was absent after the tone, the predictive power of the tone reduced (i.e. dissociating the tone) and gradually animals learned that the tone was an uninformative (or non-discriminatory) cue. At the same time, they would look for other informative (or discriminatory) cues and the predictive value of the light increased, resulting in strong conditioning of the light. Altogether, Wagner’s study demonstrated that cue competition is significant in learning, as it enables downweighting of uninformative cues as well as reinforcement of informative cues.

Additionally, learning does not simply depend on the co-occurrence of cue and outcome, but is influenced by the probability of outcome in presence as well as absence of cue. Rescorla (1968) employed the similar fear response paradigm as Kamin (1968), but provided different training trials. One group of rats were presented with a tone (CS) followed by a shock in all trials (Group G), while for another group the same amount of tone-shock trials were intermixed with other trials where the shock was presented without the tone (Group R). After training, only Group G showed conditioning of the tone. This is because only for Group G did the tone successfully predict the shock in all trials, whereas for Group R the tone failed to predict the shock in some trials, leading to reduced predictive value of the tone. It again showed that the co-occurrence account of animal learning was problematic, as the tone-shock pair occurred in the same number of trials for both groups. Instead, the distribution of outcomes with and without cues is important, since it determines the predictive value of the cue.

Bringing it together, the animal studies have shown that learning is a predictive process which is underpinned by prediction error and cue competition. A set of cues compete for predicting outcomes and learning happens when predictions diverge from observations (i.e. prediction errors occur). The Rescorla-Wagner model was developed to capture these key principles in learning (Rescorla & Wagner, 1972). It proposed that during the learning process, the predictive value of the cues that consistently predict the outcome is reinforced, whereas other cues that fail to predict the outcome and hence lead to prediction error are dissociated or downweighted. Importantly, the dissociation of an uninformative cue would lead to the upweighting of other informative cues, which is the core of the current study and will be emphasized when introducing the study design. This model is further applied in human learning research, which is covered in the next section.

2.2.2 Apply in general human learning

Rescorla & Wagner's (1972) error-driven learning model has been widely extended to account for human learning patterns, ranging from word learning to cognitive and social ability development. For example, one of the early-studied verbal learning approaches, paired-associate learning, has been accounted for under the Rescorla-Wagner model (Rudy, 1974). In paired-associate learning, multiple unrelated components (e.g. a word and a colour) are paired together to form a complex stimulus and are trained to associate with a "response" item (e.g. a number) (Underwood, Ham, & Ekstrand, 1962), which is similar to the complex CS (e.g. tone+light) and single US (e.g. shock) design in animal research aforementioned. Learners are then tested with one component of the stimulus in eliciting the response item. A typical observation is stimulus selection, meaning that not all components of the complex stimulus retrieve the response items equally well, with some components (e.g. words) overshadowing others (e.g. nonsense trigrams) (Postman & Greenbloom, 1967; Sundland & Wickens, 1962; Underwood et al., 1962). Rudy (1974) argued that such observation can be best explained following Rescorla and Wagner's error-based model. The component cues have different saliencies to learners (e.g. words being more salient than trigrams to humans). Once an association is formed between the most salient component and the response item, no further learning will be activated, and the other component will not be associated with the response because it provides no additional information and generates no prediction error.

In addition to verbal learning, the development of human contingency knowledge has also been examined under the Rescorla-Wagner model (e.g., Allan, 1993; De Houwer & Beckers, 2002; Dickinson, Shanks, & Evenden, 1984; Shanks, 1985). Forming contingency relationships is an important cognitive ability that allows humans to predict future events based on existing events (De Houwer & Beckers, 2002). For example, Dickinson et al. (1984, Experiment 2) examined participants' performance on associating two events in a video game. Prior exposure to an event A followed by an event O (A-O trials) reduced participants' contingency judgement for another potential causing event T, though compound trials (A+T followed by O) were presented after the A-O trials. This matches the prediction of the Resocrla-Wagner model in that event T provided no additional information and generated no prediction error after the A-O association was formed. However, it is worth mentioning that contingency judgement tends to be more complex than the animal conditioning process, and as noted by De Houwer and Beckers (2002), there could be multiple mechanisms involved such as a hybrid of association (as the Rescorla-Wagner model suggests) and rational reasoning.

Other cognitive-related behaviours such as category learning (Gluck & Bower, 1988) and transitive inference reasoning (Wynne, 1995) were also found to be compatible with the associative model. Additionally, in social psychology, the attractiveness judgement of strangers has been found to be affected by the presence of other strangers (and their relative attractiveness) in the same context (Cramer, Weiss, Steigleder, & Balling, 1985), analogous to the conditioning research in which multiple cues are present simultaneously in a given context and compete for predictive value.

More recent research has explored the role of prediction error in human's perception of the world (e.g., Ramscar, Matlock, & Dye, 2010 for time perception; Dye, Ramscar, & Suh, 2011 for musical pitch perception). Overall, the fundamental ideas of the Rescorla-Wagner model, prediction error and cue competition, have been largely observed in human learning. In the following sections, I will turn to the application of this approach in language learning in particular.

2.2.3 Apply in language learning

In early days, this error-driven, discriminative approach to language learning was strongly rejected by Chomsky (1959) because it is closely related to animal learning, and hence mainstream linguistics did not go further in this direction. But in more recent years, there is a revival of interest in this approach by linguists as it provides insights for the question of what mechanisms underpin human language acquisition. Researchers in this discriminative learning framework have primarily explored the order effect in language learning, showing that as predicted by the model, the order in which the linguistic stimuli (e.g. sound and visual stimuli) are presented during learning is important, as certain order allows prediction and cue competition to occur and hence leads to better learning (e.g. , Ramscar et al., 2010; Ramscar, Dye, Popick, & O'Donnell-McCarthy, 2011).

In a series of studies examining children's vocabulary acquisition, Ramscar et al. (2010; 2011) has demonstrated that the order in which labels and objects are presented influences learning outcome. To be specific, viewing objects before labels (what they called *feature-label learning*) allows for cue competition, as various features of the object compete for predicting the label, and learning from prediction error emerges. On the other hand, viewing the labels before objects (*label-feature learning*) does not allow cue competition over the features of the objects and learning thus relies on the probabilities of feature-label co-occurrence. Ramscar et al. (2010) examined this hypothesis with children of 24-30 months old who learned colour words in either a feature-label or label-feature order. The order of presentation was manipulated by contrasting English pre-nominal colour utterances (e.g. *This is a red ball*) and post-nominal colour utterances (e.g. *This ball is red*). The pre-nominal utterances provided a label-feature condition as children heard the label *red* before their attention was directed towards the object ball. In contrast, the post-nominal structure indicated a feature-label condition in which children are guided to attend to the object ball (and its related features) before hearing the word *red*, and the various features of the ball compete for predicting the label *red*. If cue competition and prediction facilitate word learning, the FL condition should lead to better knowledge of colour words after training. This is indeed what the authors observed - children from the FL group improved significantly in selecting the correct colour when being asked post-nominal questions such as *which one is red*; whereas the LF group did not show improvement in answering any types of (pre-nominal or post-nominal) colour-related questions. Similar effect of learning order has been reported by Ramscar et al. (2011) in a study of children's (aged 30-40 months) numerical word learning, where post-nominal number learning (FL, e.g. *Balls. There are two.*) benefited the development of numerical concepts but pre-nominal learning (LF, e.g. *There are two balls.*) did not. These research on children's word learning, however, lacks an explicit control over which features are competing for predictive value in the learning process. Individual children may be sensitive to different types of features of the objects such as shape, texture and size, and it is hard to know whether the same set of features enter into competition for every FL learning child. As a result, we cannot track the relative predictive value of various features after learning in order to fully test the predictions of the discriminative learning model (i.e. whether the predictive value of features other than colour decrease for colour labels).

An experiment with more controlled features and labels was carried out with adult learners in an artificial language learning experiment (Ramscar et al., 2010). A set of novel compound items were created by superimposing small novel-shaped objects onto the surface of a larger novel-shaped object (the body object), and adult participants were trained to acquire the names (or labels) of these compound items either in a FL or LF order. Since the objects were manually created, it is possible to explicitly manipulate which feature is reliable in predicting the label and which is unreliable. In this experiment, the shapes of the body objects were more salient than that of the superimposed small objects, but it was designed that body shape was a non-discriminative (or unreliable) cue. Specifically, each label corresponded to two different items, one at high frequency and the other at low frequency, and the low frequency item of one label (e.g. *wug*) shared body shape with the high frequency item of another label (e.g. *dep*). Thus, the labels cannot be discriminated based on body shapes. Instead, learners had to dissociate the body shape cue and turn to capture the actual discriminative cue - the shapes of the superimposed objects - to discriminate the novel items (especially the low frequency items), as each set of superimposed objects corresponded to a fixed label. It was predicted that only in the FL condition where visual items were presented before labels could participants learn to dissociate body shape and reinforce the discriminative cue. This is because upon viewing an object, participants can make predictions on what they would hear on the basis of the features they see and compare their predictions with the actual labels. Initially, they may rely on body shape for discrimination due to saliency. For example, when seeing the low frequency item for *wug*, they may expect to hear the word *dep* because this body shape occurred at high frequency with *dep*. However, they heard *wug* instead which is different from their prediction (i.e. prediction error), and this led them to start downweighting the body shape feature and looking for other reliable cues. In contrast, in the LF condition where labels were heard before seeing the objects, the non-discriminative cue cannot be dissociated as a result of prediction and prediction error. Participants can simply learn from co-occurrence and can learn the high frequency items only. After training, participants were tested on their ability to match labels and novel objects in forced-choice tasks. The results provided support for the predictions: though both FL and LF groups showed relatively high accuracy in classifying the high frequency items of the labels, only the FL group performed well in discriminating the low frequency items, indicating that they had learned to ignore the unreliable body shape cue and to rely on the alternative discriminative cue. Additionally, computational simulations using the Rescorla-Wagner model (Rescorla & Wagner, 1972) were carried out and confirmed the findings of the human learning experiment.

Apart from the research discussed above which focused on word learning, the discriminative, error-driven account has also been applied in the area of morpho-syntactic acquisition (e.g., Arnon & Ramscar, 2012; Ellis & Sagarra, 2010). For example, a recent study by Vujović, Ramscar and Wonnacott (2021) replicated Ramscar et al.'s (2010) order effect research in the context of learning affixes through either suffixing or prefixing. In the suffixing condition, predictions can be made based on the (semantic and phonological) features associated with the noun stem and the features compete for predicting the following suffix. Thus, this condition is predicted to lead to dissociation of the uninformative (but high-frequency) feature and better learning of the informative (but low-frequency) cue. In the prefixing condition, no such cue competition is available because the affix is heard before the features (or cues). The results supported the predictions of discriminative learning, with participants in the suffixing condition showed stronger learning of the low-frequency but discriminative cue than those in the prefixing condition. Furthermore, Arnon and Ramscar (2012) reported blocking effect of previously acquired nouns on the learning of further determiner-noun phrases for adult learners of an artificial language, as the noun in isolation was sufficient for predicting the outcome object and the addition of a determiner was considered uninformative.

Additional evidence for the discriminative learning approach came from Rost and McMurray's (2009) research on the role of input variability in infant word learning. Infants were provided with either single-talker or multiple-talker input of a novel minimal pair (that differed only in one phoneme), and it was found that infants in the single-talker group failed to discriminate the two words whereas infants in the multiple-talker group were able to map the two words to different referents. Although the authors did not explicitly discuss this finding in terms of discriminative learning, later computational simulations attempted to account for this variability effect in a way similar to the Rescorla-Wagner model (Apfelbaum & McMurray, 2011). Infants may associate uninformative talker-specific acoustic cues with words in addition to the actual informative linguistic cues if they hear the word from only a single speaker, but this distraction of talker-specific cues can be eliminated by adding multiple speakers to generate talker variability (Rost & McMurray, 2009) or increasing acoustic variation in single-talker input (Galle, Apfelbaum, & McMurray, 2015). This is generally compatible with the discriminative approach which suggests that informative and uninformative cues compete for predictive value and eventually the uninformative (or irrelevant) cues are dissociated during the learning process.

To summarize, support has been found for the discriminative learning approach in different aspects of human language learning, from phonological (e.g., Divjak, Milin, Ez-zizi, Józefowski, & Adam, 2021 for allomorphy; Olejarczuk, Kapatsinski, & Baayen, 2018), lexical (e.g., Chung, 2003; Colunga, Smith, & Gasser, 2009; Ramscar, Dye, & Klein, 2013; Ramscar, Dye, & McCauley, 2013; Ramscar et al., 2010, 2011), to morpho-syntactic level (e.g., Vujović et al., 2021). Critically, the learning order that allows for predictions to be made before encountering outcomes is more facilitative for both child and adult language learners (Ramscar et al., 2010).

2.2.4 Apply in speech sound acquisition

So far, I have discussed literature on speech sound acquisition (Section 2.1) and on the development of an error-driven, discriminative learning model (Section 2.2.1-2.2.3) separately. The present study attempts to incorporate these two areas of research and examine whether discriminative learning accounts for speech sound acquisition. Importantly, the central idea of the discriminative learning account - dissociating uninformative cues and boosting the informative ones - fits with the pattern of infant speech development which shows decreasing discriminative ability for sounds that are not used contrastively (and hence uninformative) in native language. In this section, I evaluate previous research on phonological development that provides evidence for the discriminative framework, though there is relatively scant work in this area to date (e.g. Nixon, 2020; Nixon & Tomaschek, 2020, 2021; Olejarczuk, Kapatsinski, & Baayen, 2018).

In a phonetic category learning experiment, Olejarczuk et al. (2018) trained participants on a novel lexical tone category with input that contained phonetic variations along the pitch excursion dimension. After exposure to the input sounds, participants were presented with individual sounds and were required to rate whether the heard sound belongs to the phonetic category they just learned. It was found that the learning of phonetic category did not reflect the veridical distribution of the input sounds, but was influenced to a greater extent by the infrequent exemplars in input, which gives rise to surprises and provides opportunities to adjust expectations, than by the frequent exemplars. This is in agreement with the discriminative learning model that learning is triggered by new information and violated expectations (i.e. surprises).

Furthermore, Nixon (2020) explicitly examined two predictions of discriminative learning, the blocking effect and the unlearning of uninformative cues, through artificial language learning experiments with English-native adults. The first experiment explored blocking effect in non-native acoustic cue acquisition. It was found that the acquisition of a nasality cue in predicting a visual outcome (a picture) was blocked by pre-exposure to a tonal cue before being trained with a nasality+tone compound stimulus.

A second experiment tested the role of discriminative learning in dissociating salient but non-discriminatory cues and learning discriminatory cues, which is the basic idea of the current study as well. The design closely resembled Ramscar et al.'s (2010,

2011) word learning research, with a major difference that in Ramscar et al's design visual features compete for predicting spoken words whereas in Nixon's study, acoustic cues compete for predicting the visual outcomes. English speakers were exposed to non-native speech sound stimuli from a tonal language which use pitch contours to contrast lexical meanings. Such tonal cues were expected to be less salient than the segmental information (i.e. the base syllables that carry lexical tones) for English speakers as English is a non-tonal language. Spoken words were mapped to pictures (coloured shapes, e.g. yellow triangle) and participants learned these word-picture associations either in a discriminative or a non-discriminative manner. Critically, in this experiment, in contrast to Ramscar et al. (2010), the *discriminative* condition had the order that words were heard before seeing the corresponding pictures, while the non-discriminative condition had participants viewing pictures before hearing the words. This is important since the feature (or cue) to be learned was the non-native acoustic cue rather than the features of the visual picture (as in Ramscar et al.'s study). A key manipulation was that each picture was associated with two different words, one at high frequency and the other at low frequency. The low frequency word of a picture (e.g. the yellow triangle) shared base syllable with the high frequency word of another picture (e.g. the blue square), but the two differed in terms of tones. Thus, relying on base syllables for discrimination would be unsuccessful, and learners need to acquire the less familiar but informative tonal cue to discriminate low frequency words. The prediction is that since for the discriminative learning group, words were presented first, the various acoustic cues (i.e. segmental vs tonal cues) can compete for associative strength with the shape. At first, English speakers may prefer to use segmental cues to learn words, but when an unexpected shape occurred after a base syllable (when prediction error occurs), they will dissociate the base syllable and turn to other informative cues instead. In other words, the informative cue will be learned better as a result of the downweighting of the uninformative but salient cue. For the non-discriminative learning group, cues were presented after the pictures, resulting in probabilistic learning (i.e. learning of high frequency words) only. At test, as predicted, only the FL group was above chance in selecting the appropriate low frequency word for each picture, while both FL and LF groups did equally well in discriminating the high frequency words. Note that this pattern was similar but not identical to Ramscar et al.'s (2010) findings. Ramscar et al. (2010) found a benefit of the discriminative order in the high frequency condition as well (i.e. a main effect of learning order), though the effect was strongest in the low frequency condition (i.e. an interaction effect); whereas in Nixon's report there was only an interaction effect of learning order and frequency, and the two groups were comparable in the high frequency condition. Bringing together, a common general pattern is that we expect to find a benefit of the discriminative order particularly in the low frequency condition. Following this 2020 work with adult learners, Nixon and Tomaschek (2021) constructed a computational learning model that imitated infants' speech sound learning under a discriminative learning network, and reported that the model could respond to sound pairs in a similar pattern as human infants.

Both of Nixon's (2020) experiments provided evidence for the application of the discriminative learning model in speech sound acquisition, and in particular, in non-native sound acquisition. However, to date this has only been demonstrated in this single paper in which the experiments were carried out with English native speakers on their acquisition of a particular non-native cue. It is thus important to explore if the results hold more broadly with other acoustic cues. To be specific, Nixon's (2020) Experiment 1 observed blocking of nasality cue after exposure to tonal cue, but not in the opposite direction, indicating that various non-native acoustic cues pose different degrees of learning difficulty, and only those cues that have been fully acquired (e.g. tonal cue) block the acquisition of other cues (e.g. nasality cue). Following from this, it is possible that the perception and acquisition of pitch differences are relatively easy for English speakers in Experiment 2 and this allows learners to capture the tonal cue quickly in discriminative learning, though such contrasts are not present in their native language. Thus, it is worth investigating whether the acquisition of other non-native cues by learners of different language backgrounds would exhibit similar discriminative learning effect as well. Moreover, a related question arises regarding whether the relative ease of acquiring non-native pitch contrasts implies that dissociation of (i.e. learn to ignore) tonal cues would be harder for native tonal language speakers, since such prosodic cues could be too strong to be unlearned after a short training period. Inspired by this previous research, the current study aims at extending the understanding of discriminative learning in the area of speech sound acquisition and attempts to partially address the questions (discussed above) that arise from Nixon's (2020) experiments.

2.3 The current study

In this study, I investigate the acquisition of non-native geminate consonant contrast by native Chinese speakers who rely heavily on tonal cues in lexical discrimination, and test whether this acquisition process can be accounted for by the discriminative learning model. To be specific, this study uses a reverse of Nixon's (2020) design, examining whether native speakers of a tonal language are able to dissociate a tonal cue and thus boost learning of a non-native segmental cue (in this case, a geminate consonant contrast - found in languages such as Italian - that is missing in Chinese) which is the cue that discriminates word identity. If discriminative learning mechanisms are at work, this non-native speech sound learning should be more successful in a discriminative learning order where cue competition is available.

The two language learning experiments in this study included artificial word learning tasks similar to Nixon's (2020) Experiment 2, in which participants either heard a word before seeing the associated picture (cue-outcome, discriminative learning) or saw a picture before hearing the matched word (outcome-cue, non-discriminative learning). Participants were asked to learn the word-picture mappings and thus (implicitly) to figure out which cues are reliable in discriminating the words. The design was closely based on Nixon's (2020) experiment (which in turn was based on

Ramscar et al. (2010)) with high and low frequency words. Each picture was associated with one high frequency word and one low frequency word, and the low frequency word of one picture (e.g. a ‘crown’) shared tone (e.g. a falling tone) with the high frequency word of another picture (e.g. a ‘shoe’) but the two words differed in terms of the less salient consonantal cue. This means that the tonal cue was not reliable in contrasting the words and needed to be dissociated. Instead, learners need to capture the less salient consonantal (base syllable) cue for discrimination since each base syllable corresponded to a fixed picture, making this cue discriminative in the language. In this design, it is particularly hard to learn the low frequency words as learners need to distinguish them from the high frequency items with the same tone on the basis of the non-native consonantal cue.

In the discriminative learning order, there would be competition for predictive value between acoustic cues. Chinese speakers might initially pay attention to the more salient tonal cue. For example, they might expect a ‘shoe’ after hearing a word with falling tone since this tone was frequently associated with ‘shoe’, but instead they saw a ‘crown’ in some learning trials (as this tone is also used in the low frequency word of ‘crown’). Thus, the unexpected picture (or prediction error) would reduce the predictive value of the tonal cue. At the same time, the associative strength between the less salient consonantal cue and the outcome picture should be strengthened. Gradually participants should acquire the non-native cue as discriminative in this artificial language. Thus by the end of training, at test they should show learning of the association of the discriminatory cue and discriminate low frequency words accurately. On the other hand, non-discriminative learning lacks this predictive learning process since participants see the outcome picture before hearing the acoustic cues. They will therefore show weaker learning of the discriminatory consonantal cue, leading them to confuse the two words with the same tone but different consonant gemination, and hence at test they will only recognize the high frequency words associated with pictures and fail to discriminate the low frequency words.

To summarize, it is predicted that participants in the discriminative learning group would perform better at learning the novel words than the non-discriminative learning group especially for the low frequency words, since learning is facilitated by cue competition and prediction error in the former condition, allowing learners to acquire the actual discriminative consonantal cue that is critical in discriminating low frequency items. For high frequency items, the learning order effect is expected to be small because learners could select the correct word for each picture simply based on which tone occurred frequently with the picture, without having learned the gemination contrast. Thus, I will look both for an effect of learning order specifically in low frequency items, and overall an interaction between frequency and learning order to see if there is evidence that the effect is stronger in the low frequency condition. Participant performance was measured first in a word-picture mapping test that examined whether they had learned the trained words, after which there was a generalization test which examined whether learners can extend the acquired discriminative patterns to untrained stimuli.

Before doing the main experiments, a small pilot study was carried out to test whether the non-native geminate consonant contrasts are perceivable by Chinese speakers. The purpose of this pilot is to ensure that participants are able to perceive the non-native contrasts to some extent, even though they are less salient than the tonal cues, since the short language learning experiments depend on participants being able to hear the sound differences. Italian gemination contrast (double versus single consonant) was chosen because they are absent in various Chinese dialects as well as in English, a widely-learned foreign language for Chinese speakers. Five short versus long consonant pairs were examined, three of which were selected to enter the main experiments according to the relative ease of discrimination.

3. Pilot study

3.1 Method

A discrimination task was employed to measure the relative ease of discriminating Italian double versus single consonants. The five consonantal pairs used are presented in Table 1. The consonant length contrasts always occur in the second syllables. The double consonant is pronounced as the prolonged version of the single consonant.

Table 1 Italian words with double versus single consonants

Double consonant	Single consonant
palla	pala
fatto	fato
canne	cane
cassa	casa
bevve	beve

3.1.1 Participants

Eleven native speakers of Chinese were recruited through Chinese social media to take part in the pilot study, but two participants were excluded due to technical problems and missing data. There were nine participants who successfully finished the experiment with no reported issues (aged 20-26; 6 females). All participants were adult native speakers of at least one dialect of Chinese (e.g. Mandarin, Cantonese) and had no previous knowledge of Italian. All participants reported that they know more than one language or dialect, including English. Consent was obtained from every participant before the experiment, and they took part in the experiment voluntarily knowing that there would be no payment for it. They were provided with the opportunity to discuss any questions related to the experiment or about the topic of language learning in general after the experiment.

3.1.2 Stimuli

Auditory stimuli were the ten Italian words listed in Table 1 produced by two native speakers of Italian, one female and one male, in order to generate speaker variability. Authentic speech stimuli were used with no further synthesis in order to best capture the naturally-produced Italian words and consonant lengths, avoiding any duration effects from artificial manipulation. For the words with double consonants (i.e. first column of Table 1), the duration of the second syllables ranged from around 310ms to 450ms, each of which was longer than its single consonant counterpart (~220ms to 360ms). Additionally, double consonant words have slightly shorter first syllables (~110ms to 300ms) than their single consonant pairs (~210ms to 360ms), as the lengthened second syllable tends to compress the duration of the preceding syllable.

3.1.3 Procedure

The pilot study had only one discrimination test session. In each trial, two words were played, either differing in the length of the second consonant or being exactly the same despite speaker variability. During the presentation of the words, there was a fixation cross occurring at the centre of the screen to make sure that participants were attending to the experiment. Participants were presented with two buttons on the screen after they heard the two words, one printed with “same” and one with “different”. They were asked to choose whether they think the words they heard were the same word or different words, and were instructed to disregard talker variability (i.e. female versus male voice). If there was no response in three seconds, it automatically skipped to the next trial and a 'missing' response was recorded. Each of the five consonant contrasts was tested for ten times, resulting in fifty test trials in total, with an equal number of double and single consonant words being presented. The ten trials of each consonant contrast were grouped into one block and occurred together, but the order of the five blocks were randomized (for example, some participant encountered ten *pala/palla* trials first, followed by ten *fato/fatto* trials; whereas another participant first encountered the *fato/fatto* block followed by the *beve/bevve* block). The entire pilot experiment took no more than 10 minutes to complete. All instructions were provided in both English and simplified Chinese.

3.2 Results and discussion

The average accuracy in discriminating each consonant length contrast is shown in Table 2. The above-chance accuracy in perceiving the double versus single consonants ($p<.001$ for each contrast when compared against chance level .5) indicates that Chinese listeners with no Italian experience are able to hear the consonant length difference when the contrasts are presented in pairs. A repeated measure ANOVA was conducted and it was found that there was no significant difference between the discrimination of the different consonantal contrasts ($F(4,32)=1.246$, $p=.311$). The three contrasts (fatto-fato, bevve-beve, cassa-casa) with discrimination rates of above 80% were chosen as base words in the main language learning experiments. These cues are non-native for Chinese speakers but can be successfully perceived by them, and hence may enter into cue competition in the discriminative learning design.

Table 2 Accuracy of the discrimination test

	Average accuracy
palla-pala	78.9%
fatto-fato	82.2%

canne-cane	73.3%
cassa-casa	85.6%
bevve-beve	80.0%

4. Experiment 1

Experiment 1 investigated the acquisition of three pairs of non-native consonant length contrasts by native Chinese speakers under either a discriminative (word-picture) or a non-discriminative (picture-word) order. The experiments were approved by CUREC (approval letter in Appendix B) and the principles of informed consent were strictly followed in the experiments.

4.1 Method

4.1.1 Participants

Forty participants were recruited through Chinese social media (Wechat), all of whom completed the experiment (aged 18-32; 29 females, 1 preferred not to specify). Twenty participants were randomly assigned to the discriminative learning group and the other twenty to the non-discriminative learning group. All participants were adult native speakers of at least one dialect of Chinese (including but not limited to Mandarin and Cantonese) with no known language impairments, and had never learned Italian[1] before the experiment. All participants spoke at least one dialect of Chinese and had learned English as a second/foreign language. Some participants spoke more than one Chinese dialect (4 in discriminative condition, 9 in non-discriminative condition); some participants had learned other non-Chinese languages (besides English) as second languages (7 in discriminative condition, 6 in

---

[1] 5 participants have learned Japanese before, which also contains consonant gemination. However they all learned Japanese as a foreign language in classroom, and the consonant geminate contrasts were non-native for them. Also, these participants did not do particularly well (especially in low frequency condition) compared to other learners in tests. Thus they were included since the geminate consonant contrast seems to remain less salient for them.

non-discriminative condition).

Every participant completed a consent form before the experiment with details about the experiment (including purpose, task, anonymity, participant rights, data use and contact information). They were only identified by the randomly-assigned ID number in the dataset and no personal information can be tracked. They took part in the experiment voluntarily and were informed that there would be no money payment for participation. Instead, they were provided with the opportunity to consult about any questions or issues related to language learning (especially English learning) with the researcher, who specializes in the area of language learning and had experience in language teaching.

4.1.2 Stimuli

The three pairs of Italian words that were discriminated with relatively high accuracy in the pilot (fatto-fato, bevve-beve, cassa-casa) were combined with three lexical tones (rising, falling, middle-flat) extracted from Thai to form the synthesized auditory stimuli. The Thai tonal patterns were extracted from three Thai words, *kaao* ("fishy smell", mid-flat tone), *kaǎo* ("white", rising tone) and *kaâo* ("rice", falling tone), which differ only in the tonal levels. The pitch contour of each Thai word was extracted and superimposed onto one pair of Italian base words (as shown in Table 3) using Praat, leading to six novel words. As each Italian base word was produced by two native Italian speakers to generate some talker variability and imitate variations in natural language learning, a total of twelve sound stimuli were included in the training phase. Additionally, each base word was combined with an untrained Thai tone (high or low tone), in order to form six untrained words that were not presented in the training phase but were used as sound stimuli in the generalization test after training (as shown in Table 4).

Table 3 Novel words in the training phase

Base word	bevve	beve	cassa	casa	fatto	fato
Tonal pattern	rising		falling		mid-flat	
Novel word	bevve_rising	beve_rising	cassa_falling	casa_falling	fatto_flat	fato_flat

Table 4 Untrained words for the generalization test

Base word	bevve	beve	cassa	casa	fatto	fato
Tonal pattern	low			high		
Untrained word	bevve_low	beve_low	cassa_low	casa_high	fatto_high	fato_high

Visual stimuli were three pictures of familiar objects that are frequent in daily use (crown, pencil and shoe). They were used as the referents of the novel words in the learning phase. Two additional pictures (car and apple) were used as visual choices in the generalization test. All pictures are included in Appendix A.

4.1.3 Procedure

The experiment consisted of three phases, training, test and post-test questions. The test phase contained two parts, a test on trained items and a generalization test. Participants were told that they would learn some words from an artificial language in the experiment. The experiment took around 15 minutes to complete. All instructions were provided both in English and simplified Chinese.

Training

In the training stage, three novel words occurred 15 times each (high frequency; 75% of the training trials); the other three occurred 5 times each (low frequency; 25% of the training trials), making 60 trials in total. For each consonantal contrast pair (e.g. bevve-beve), one word occurred in the high-frequency condition and the other in the low-frequency condition, and the two words corresponded to different pictures in high and low frequency conditions. The full design is provided in Table 5. Note that, for example, the rising tone was associated with “crown” in 75% of the learning trials but was mapped to “pencil” in 25% of the trials. Thus, relying on tones for discrimination would not be accurate. Instead, each base syllable corresponded with a particular picture and served as the discriminative cue in this experiment. Each pair of base syllables (e.g. bevve-beve) was heard equally from each of the male and female speakers (e.g. 8 *bevve_rising* and 2 *beve_rising* trials were from the female speaker; 7 *bevve_rising* and 3 *beve_rising* trials from the male speaker). Trial order was randomized for each participant with no ordering constraints.

Table 5 Word-picture mappings

	Crown	Pencil	Shoe
High frequency (75%, 15 exposures each)	bevve_rising	fato_flat	cassa_falling
Low frequency (25%, 5 exposures each)	casa_falling	beve_rising	fatto_flat

The key manipulation was the different ordering of stimuli in training for the two learning groups. For the discriminative learning group, participants first heard the word in each trial, and then saw the referent picture on the screen for 1 second (as illustrated in the top panel in Figure 1). In the non-discriminative condition, participants were first presented with the picture on the screen for 1 second, followed by the word (as shown in the bottom panel in Figure 1). After each trial, participants clicked on a ‘next’ button to proceed to the next trial. Each sound stimulus was around 600ms, and there was a 200ms interval between the sound and the visual stimulus. Participants could take a break at any time since the learning phase did not proceed until they clicked on the ‘next’ button.

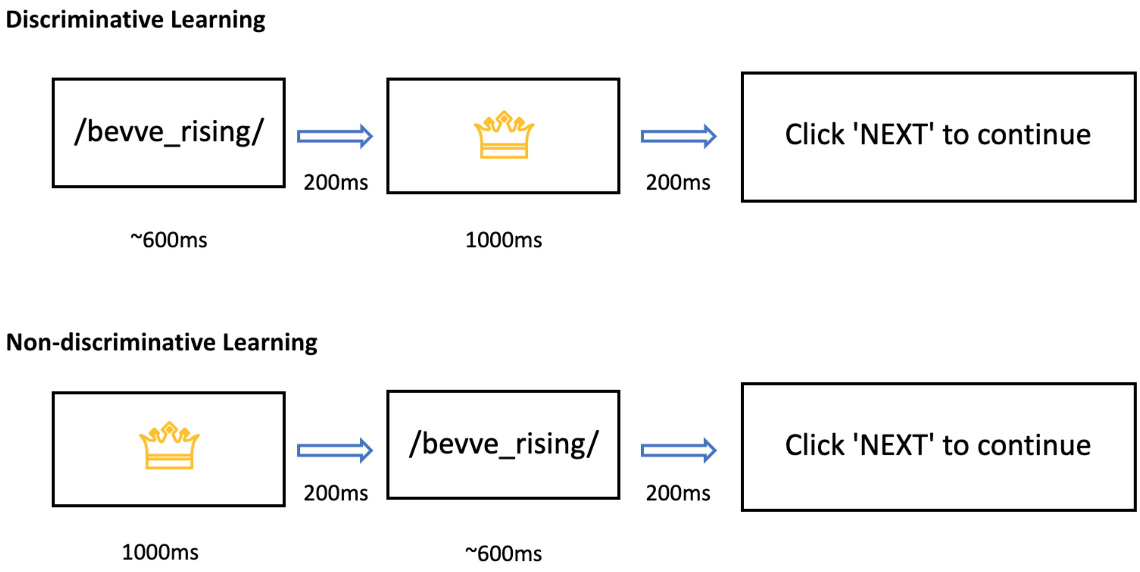

Figure 1. Training trial in the discriminative (top) and non-discriminative (botton) learning condition.

Test

The test phase included two types of tests, a familiar noun test and a generalization test.

Test 1: familiar noun test - the first test examined learning of the trained items. Participants were presented with a picture in each test trial and heard two words. One speech bubble appeared on the screen during the presentation of each word, and participants were instructed to select the appropriate word for the picture by clicking on the corresponding speech bubble (as shown in Figure 2). Eight test trials were included for each word, regardless of their frequency of exposure in the learning phase, resulting in a total of 48 trials in test 1. A subset of the recordings used in training were used and each word was paired with the other four words that mapped to different pictures from it (e.g. *bevve_rising* was paired with the four words except *casa_falling*, since *bevve_rising* and *casa_falling* refer to the same picture). There were two types of test trials in test 1 - *discriminative learning test trials* (as shown in the left panel of Figure 2) where the foil word differ from the target only in terms of consonantal length; and *baseline learning test trials* (as shown in the right panel of Figure 2) where the foil and the target words had different base syllables as well as tones. Among the eight test trials for each word, two were discriminative learning test trials and six were baseline learning test trials (with two trials for each foil word), leading to a total of 12 discriminative learning test trials and 36 baseline trials. The correct answer appeared randomly at the left or right bubble across trials. The different trial types were intermixed and the trials played in random order with no constraints. If a decision was not made within 3 seconds, it automatically skipped to the next trial and a null response was recorded. Participants were informed of this time limited and were guided to respond as quickly as possible based on their intuition.

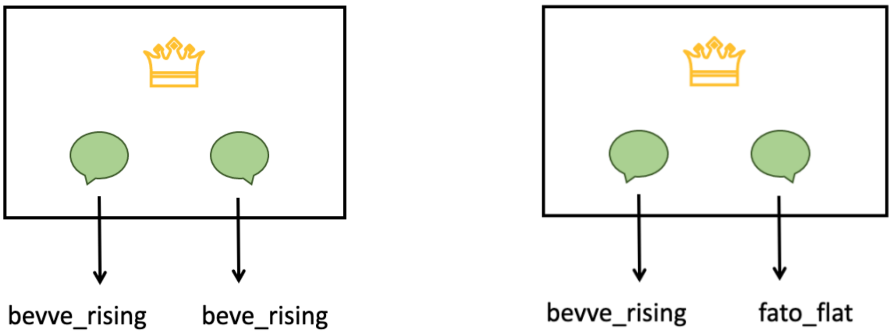

Figure 2. Example of two types of test trials in test 1 - discriminative learning test trial (left) and baseline learning test trial (right).

Test 2: generalization test - this examined whether participants had learned that the tonal cue is not discriminative in this artificial language - if so they should be able to identify the correct meaning of a word even if the tone has changed. The six Italian base words were combined with two untrained Thai tones, a low and a high tone, to create untrained items. If participants had successfully “unlearned” the tonal cue and learned that the gemination cue is discriminative, they would choose the picture for each base word which matches that in the training phase. For example, they would choose ‘crown’ for *bevve_low* because this base word was associated with ‘crown’ in the training stage.

Participants heard an untrained word in each trial, and were presented with two pictures from which they were instructed to choose the correct referent of the word (as shown in Figure 3). For half of the trials the foil picture was an unfamiliar one (as shown in the left side of Figure 3). For the other half, it was a familiar/trained picture (as shown in the right side of Figure 3). The trained foil picture was always the one that in training had been mapped to the word with different consonantal length from the heard base word (e.g. ‘pencil’ was used as the trained foil for the word *bevve_low* because it was mapped to *beve* during training). This means that in trained foil trials learners need to discriminate using the gemination cue. The inclusion of untrained pictures was to match with the use of untrained sound stimuli in this test. Each of the six base words occurred twice in test 2 (once with untrained foil and once with trained foil), leading to 12 test trials in total. The different trial types were intermixed and the trials were presented in random order with no additional constraints. As for test 1, if a decision was not made within 3 seconds, it automatically skipped to the next trial and a null response was recorded. Participants were again informed of this time limited and were guided to respond as quickly as possible.

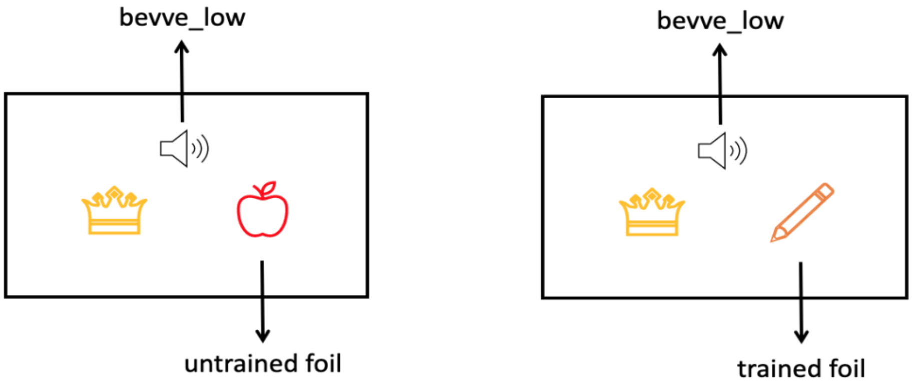

Figure 3. Example of test 2 trials with untrained foil (left) and trained foil (right).

Post-test question

After the test, participants were provided with two short answer questions about their explicit knowledge of the artificial language. The first question was about whether they noticed any difference between the artificial language and Chinese in terms of tones. The second question asked whether there were any differences between the two languages with respect to pronunciations (e.g. consonants and vowels). The two questions examined if the learners developed explicit knowledge after the brief training session and if there were any differences between answers by the two learning groups. If explicit patterns were acquired, the discriminative learning group would report that tonal cues were unreliable and that there was consonantal length contrast in the artificial language.

4.2 Analysis and results

The results from the familiar noun test (test 1) and the generalization test (test 2) were analyzed separately, as the two tests examined different lexical items using different methods, and generated different predictions.

Test 1: familiar noun test

Test 1 included two types of test trials. The baseline learning test trials consisted of foil items that differ from the targets in terms of both tones and base syllables (e.g. *bevve_rising* versus *fato_flat*); whereas the discriminative learning trials contained foils that share the same tone with the targets but had different base syllables (e.g. *bevve_rising* versus *beve_rising*). These two types of test trials tended to pose different degrees of discrimination difficulty for learners, since for baseline trials learners can rely on both tones and base syllables for discrimination but for discriminative test trials they need to detect consonantal contrasts in the base syllable. Due to this difference in discrimination difficulty, the predictions for the two types of trials were different - there is predicted to show an order effect (especially for low frequency items) in discriminative learning test trials but not in baseline learning trials.

The order effect is not predicted for baseline trials because even participants failed to learn the consonantal contrast, they could still complete the test in these trials according to tonal differences, and hence performance of the two groups should not differ significantly. The two trial types were thus analyzed separately.

Baseline learning test trials - Figure 4 shows the mean accuracy in baseline learning test trials under different frequency and learning context conditions. A 2×2 two way mixed analysis of variances (ANOVA) was carried out with the proportion of correct responses as the dependent variable, and one within-subject factor (frequency) and one between-subject factor (learning order) as the independent variables. There was a significant main effect of frequency reflecting higher performance with high frequency (average 75.4%) than low frequency words (average 38.3%) (F(1,38)=39.919, *p*<.001, η^2=.512). There was no significant main effect of learning contexts (F(1,38)=.015, *p*=.903, η^2<.001) nor any interaction between frequency and learning context (F(1,38)=.095, *p*=.760, η^2=.002). As shown in Figure 4, the high frequency baseline items in both learning contexts were correctly discriminated above chance (with 95% confidence intervals above the 0.5 chance level), while the low frequency items were below chance (the intervals are below the 0.5 level).

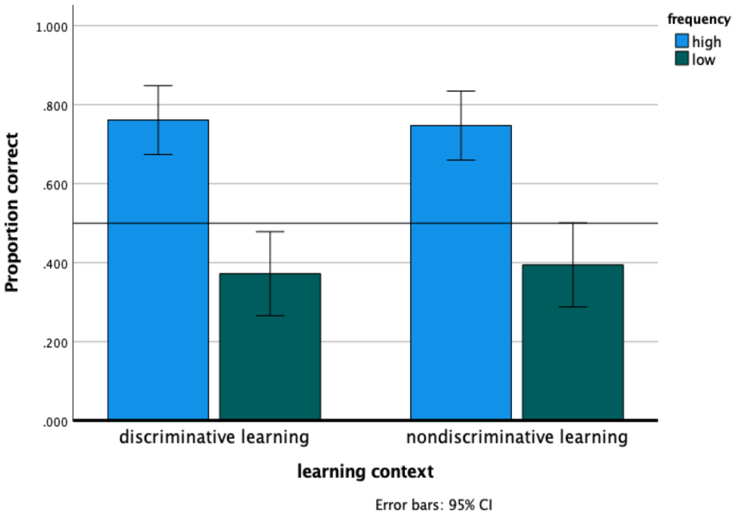

Figure 4. Accuracy of baseline learning test trials in test 1 under different conditions in Experiment 1. Error bars indicate 95% confidence intervals computed in SPSS. Line indicates chance performance.

Discriminative learning test trials - Figure 5 illustrates the proportion of correct responses for discriminative learning test items. A similar 2×2 two way mixed ANOVA revealed that performance with high frequency items (average 57.5%) was significantly better than that with low frequency items (average 34.2%) (F(1,38)=21.040, p<.001, η^2=.356). Again, there was no effect of learning order (F(1,38)=.864, p=.359, η^2=.022) nor interaction effect (F(1,38)<.001, p=1.00, η^2<.001). As can be observed from the plot, the high frequency items in non-discriminative learning context were correctly recognized above chance (the error bars showing the confidence intervals are above the chance level), whereas those in discriminative learning context were not significantly different from chance (error bars crossing the chance level). The low frequency items in both learning contexts had below-chance accuracy.

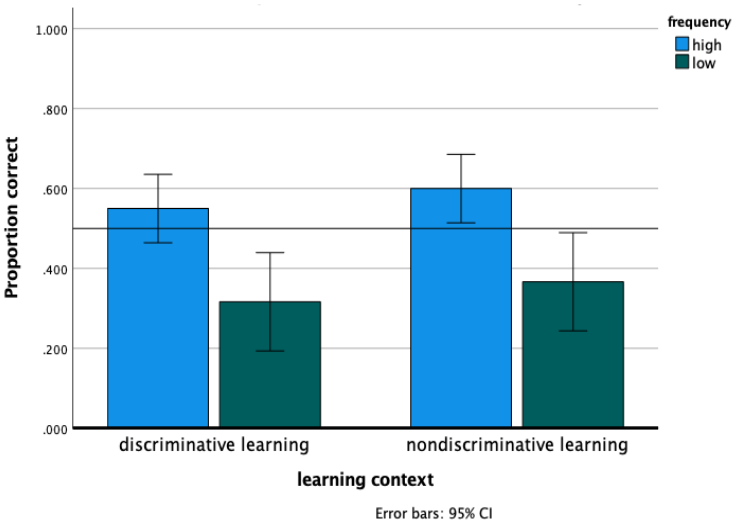

Figure 5. Accuracy of discriminative learning test trials in test 1 under different conditions in Experiment 1. Error bars indicate 95% confidence intervals computed in SPSS. Line indicates chance performance.

As the learning order effect was predicted most strongly for low frequency items, an additional independent t-test was conducted to examine if any difference would emerge between discriminative and non-discriminative learning contexts when taking into account the accuracy for low frequency items only. No significant difference was found (t_{38}=-.582, p=.564).

Test 2: generalization test

Test 2 contained two types of foil items - untrained or trained pictures, both of which required participants to ignore the tonal cue and use the consonantal cue for discrimination. Since both foil type conditions require learners to capture the non-native consonantal cue, the predictions for these two foil types were similar - discriminative order would lead to better performance than non-discriminative order especially for low frequency items, although it is possible that trials with one type of foil could be easier than the other. Thus, unlike test 1 in which the two types of test trials were analyzed independently due to the different predictions for baseline and discriminative foils, in test 2 the untrained and trained foils were analyzed together in a single ANOVA test with foil type as a factor.

Figure 6 shows the mean accuracy for test trials with untrained (left) and trained (right) foils under different learning conditions. A 2 (within-subject, frequency) × 2 (between-subject, learning order) × 2 (within-subject, foil type) three way ANOVA revealed that there were main effects of frequency and foil type. On average, performance on high frequency items (average 79.6%) was significantly better than that on low frequency items (average 37.9%) (F(1,38)=32.269, p<.001, η^2=.459). In addition, untrained foils led to higher accuracy than trained foils (F(1,38)=10.983, p=.002, η^2=.224). There was no effect of learning order (F(1,38)=1.844, p=.183, η^2=.046) nor any significant interaction among the three factors. Moreover, judgements on high frequency items were all above chance (as shown in the blue bars), and for low frequency items performances were either at chance level (with untrained foils, green bars in the left chart) or below chance (with trained foils, green bars in the right chart).

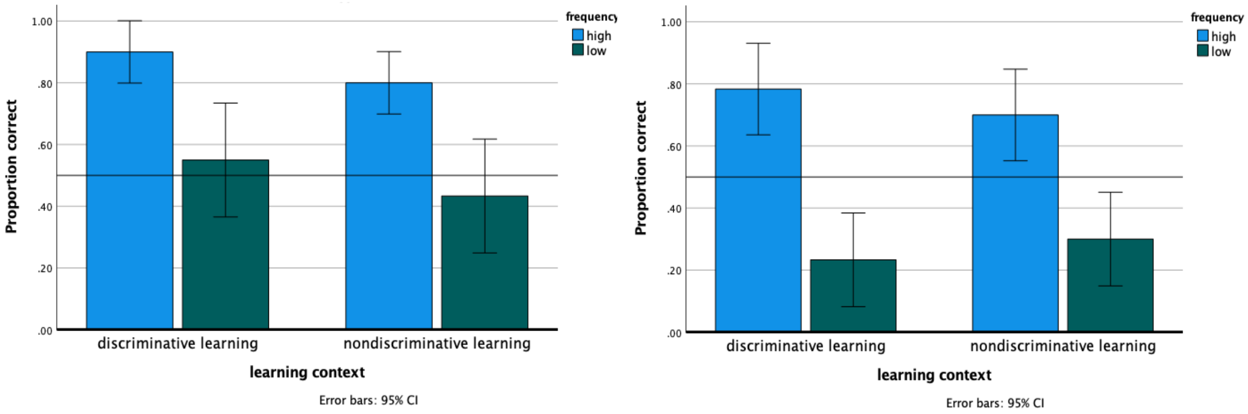

Figure 6. Accuracy with untrained (left) and trained (right) foils under different learning contexts in test 2 of Experiment 1. Error bars indicate 95% confidence intervals computed in SPSS. Line indicates chance performance.

A further analysis was conducted to specifically explore the effect of these factors on low frequency items alone, as the predicted effect of learning order is more likely to emerge in the low frequency condition. An ANOVA with learning order and foil type as factors suggest that there was only a significant effect of foil type (F(1,38)=15.138, p<.001, η^2=.285), but not learning order (F(1,38)=.059, p=.809, η^2=.002) nor interaction (F(1,38)=2.513, p=.121, η^2=.062).

Post-test question

Twenty-six participants (12 in the discriminative group) mentioned that they heard tonal differences across words in this artificial language, and there were fewer tones in this language than Chinese. Two of them (one in each learning order group) explicitly noted that they thought different tones marked different meanings in the language. As for question 2, three participants (one in the discriminative group) noticed a phoneme length contrast and two of them (one in each group) specified that it was a vowel length contrast. One additional participant (from the discriminative group) reported that there was a stop at the end of some vowels in the artificial language.

4.3 Discussion

The experiment investigated whether the order in which cues and outcomes are presented has an impact on non-native speech sound acquisition. It was predicted that for the discriminative learning test trials in test 1 and the generalization test (test 2), the discriminative learning condition would lead to greater accuracy than the non-discriminative learning condition, with the effect being strongest for the low frequency items. However, the key finding of this experiment was that the discriminative group did not show stronger learning, either overall or in the low frequency condition alone, and the predicted interaction between frequency and learning order was not found. In this section, I briefly discuss the results of each test in turn, including the patterns of where participants' performance was above chance.

Test 1: familiar noun test

In baseline learning test trials, as predicted, there was no order effect on learning but merely an impact of frequency. In fact, only high frequency words were above chance and low frequency words below chance in both order conditions. The below-chance performance on low frequency trials was not predicted. Since the baseline trials can be discriminated based on both the tonal cue and the consonantal cue, it was expected that these trials would pose less discrimination difficulty and accuracy would be at least at chance for low frequency items (in the situation where low frequency items were not learned at all and participants selected by chance). However, the results suggested that participants were more likely to choose the incorrect foils in low frequency trials. On reflection this may be accounted for on the basis of some subtypes of trials that were more tricky than they appeared to be. Specifically, take the 'crown' example, the target low frequency item was *casa_falling*. When this target was paired with *beve_rising*, it is possible that participants were more likely to choose the foil *beve_rising* because they cannot distinguish *beve_rising* from *bevve_rising* which is the high frequency word for ‘crown’. Thus, the overall accuracy in the baseline low frequency condition would be slightly below chance as influenced by such trials. When the same target was paired with other baseline foils (i.e. *fato_flat* and *fatto_flat*), performance would be around chance level since no bias would occur in these trials. Post hoc inspection was consistent with this interpretation: the mean accuracy for the harder baseline trials (with foils that share tone with the high frequency word of the picture) were 27.5% whereas for the rest of the baseline trials the accuracy was 43.7%. Therefore, although in baseline learning test trials there were two sets of acoustic cues (tone and consonant) available for discrimination, performance on low frequency items were still affected by the high frequency counterparts of foils.

In discriminative learning test trials, where participants have to discriminate the target and foil on the basis of the geminate cue (e.g. *bevve_rising* versus *beve_rising*), a possible effect of learning order was predicted, because participants in the discriminative learning condition were expected to show better learning for the discriminating consonant cue. However this expected facilitative role of the discriminative learning order did not emerge. Again, there was a frequency effect - both learning order groups performed better for high frequency items than for low frequency items. Moreover, the discriminative order group only performed at chance in the high frequency condition, and though the non-discriminative group was slightly above chance, there was no significant difference between the means of the two groups in the high frequency condition (and also no interaction). This indicates that the discriminative learning test trials pose severe difficulty for learners in both groups and most participants failed to capture the reliable consonantal cue. Thus, when they were presented with two words that differed only in terms of consonant length, they likely selected either by chance or, with more exposure to the high frequency sound stimuli, participants ignoring the consonant cue might be more likely to select the high frequency stimulus (e.g. *bevve_rising*) over the low frequency counterparts (e.g. *beve_rising*) due to familiarity. This potential bias towards high frequency items also explains the below-chance performance in low frequency conditions, as participants were less likely to select the low frequency stimulus.

Test 2: generalization test

In test 2, the two types of foils (untrained versus trained) led to different performance. Trials with untrained foil pictures were completed with higher average accuracy than those with trained foil pictures, however foil type did not interact with any of the other factors. An explanation for the greater difficulty with trained foils is that when a trained picture was used as foil, it was always the picture which in training was mapped to the consonantal counterpart of the target untrained word, and hence participants need to show strong learning of the consonantal cue for discrimination. For example, in a trained foil trial for the novel word *bevve_low*, 'pencil' was used as the foil and it occured with the base syllable beve in low frequency condition during training. In order to select the correct picture ('crown') for *bevve_low*, participants had to discriminate the base syllable *bevve* from *beve* on the basis of the consonantal cue, a type of discrimination which has been found to be unsuccessful in discriminative learning test trials of test 1. On the other hand, when the target was paired with an untrained picture, it could be easier for participants because they only need to realize that the change of tones did not alter lexical meaning and that they must thus stick with the familiar picture. There was no requirement of direct discrimination between two base syllables as in trained foil trials. Therefore, performance on untrained foil trials were higher as these trials did not require specific consonantal cue discrimination but trained foil trials did.

In addition to the foil type effect was a frequency effect, which is consistent with test

1 in that high frequency words were discriminated better than low frequency words by participants in both learning order conditions. However, the discriminative learning order did not give rise to better performance in test than the non-discriminative learning order, either overall or for low frequency items alone, and again they were above chance only in the high frequency condition, with the low frequency at chance in the untrained foil trials and below chance with the trained foils. This suggests that upon hearing a low frequency item (e.g. *beve_low*), since participants had not acquired the consonantal cue, they confused *beve_low* with *bevve_low* and were more likely to select the picture (‘crown’) that occurred more frequently with the base syllable (ignoring gemination) *bev(v)e*, leading to below-chance performance in trained foil trials. In untrained foil trials, since ‘crown’ did not occur as the foil option, participants selected by chance.

Post-test question

The post-test questions were not statistically analyzed. One general pattern regarding participants’ answers was that a number of learners (12 in the discriminative group; 14 in the non-discriminative group) noticed the tonal differences but very few reported explicit knowledge of phoneme length differences (1 in the discriminative group; 2 in the non-discriminative group). This tends to suggest that participants failed to acquire the non-native consonantal contrasts explicitly, which seems to be consistent with the previous statistical analysis. However results from self-report questions should be treated as tentative - a point which I return in the general discussion section.

In sum, the key result of this experiment was that neither the results from the discriminative learning trials in test 1, nor the generalization test, support the prediction that there will be better learning in the discriminative condition, either overall or for low frequency items, examined separately, where the learning order effect is predicted to be strongest. A potential reason for not seeing this learning order effect is that there were a limited number of training trials, especially for the low frequency items. To be specific, there were only 15 low-frequency trials in total, with 5 for each novel word. It is thus possible that participants in the discriminative learning group did not encounter enough low frequency trials to learn from prediction errors and to acquire the consonantal cue. To examine this possible influence of the amount of training, a second experiment (Experiment 2) was conducted with doubled training trials.

5. Experiment 2

Experiment 2 was a replication of Experiment 1 with an additional training block. It examined whether the lack of learning order effect in Experiment 1 was a result of the relatively small number of training trials (especially in low frequency condition).

5.1 Method

5.1.1 Participants

Seventy native speakers of Chinese were recruited from Prolific, six of them were excluded because they were not in our target age range (18-40 years old), and the remaining sixty-four participants (aged 18-40; 32 females, 2 preferred not to specify) were included for further analysis. After applying the age-related exclusion criterion, there were thirty randomly-assigned participants in the discriminative learning group and thirty-four in the non-discriminative learning group. Participants were adult native speakers of at least one dialect of Chinese with no known language impairments, and had never learned Italian before the experiment. Again, all participants spoke at least one dialect of Chinese and had learned English as a second/foreign language. Some participants spoke more than one Chinese dialect (12 in discriminative condition, 11 in non-discriminative condition); and some had learned other non-Chinese languages (besides English) as second languages (10 in discriminative condition, 6 in non-discriminative condition).

As in Experiment 1, participants completed a consent form before the experiment with details about the experiment. They received money payment (at a rate of £7.5 per hour) for participation.

5.1.2 Stimuli and procedure

The stimuli used in Experiment 2 were the same as those in Experiment 1. The procedure replicated that of Experiment 1 as well, except that the number of total training trials was doubled. After the original training session, an additional training block which was identical to the original one was added. Participants could take a break between the two training blocks and could proceed as they were ready. The tests and post-test questions in Experiment 1 were employed here without change. Experiment 2 took around 20-25 minutes to complete.

5.2 Analysis and results

Similar analyses were carried out for Experiment 2 since it was a replication of Experiment 1, as well as comparisons between the two experiments.

Test 1: familiar nouns test

As in Experiment 1, the baseline learning test trials and discriminative learning test trials of test 1 were separated and the same set of analyses were conducted for each trial type.

Baseline learning test trials - A 2 (frequency) × 2 (learning order) ANOVA showed that there was a significant effect of frequency (Figure 7). High frequency test items (average 85.9%) led to better mean accuracy than low frequency ones (average 48.5%) (F(1,62)=68.372, $p<.001$, $\eta^2=.524$). The difference between discriminative and non-discriminative learning orders was also significant (F(1,62)=4.284, $p=.043$, $\eta^2=.065$), with the discriminative learning group slightly outperforming the non-discriminative group. No interaction effect was found (F(1,62)=1.101, p=.298, η^2=.017). In both learning orders, performances on high frequency items were above chance and that on low frequency items were at chance.

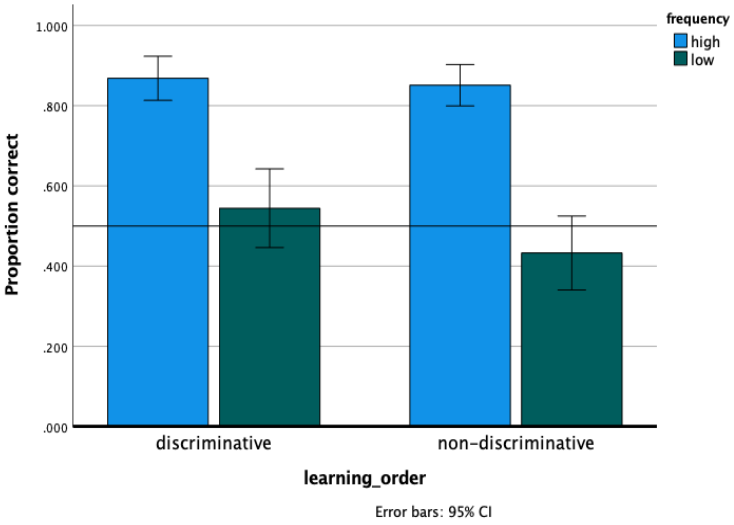

Figure 7. Accuracy of baseline learning test trials in test 1 under different conditions in Experiment 2. Error bars indicate 95% confidence intervals computed in SPSS. Line indicates chance performance.

It was unexpected that there was a learning order effect in baseline learning test trials. A further independent t-test was employed to examine whether this around-significant order effect was associated with low frequency items, as suggested by the discriminative learning model. The results turned out that performance on low frequency items alone of the discriminative and non-discriminative group did not differ significantly (t_{62}=1.665; p=.103).

Discriminative learning test trials - Figure 8 shows the accuracy in discriminative learning test trials in Experiment 2. A frequency × learning order ANOVA again found a frequency effect only (F(1,62)=14.192 p<.001, η^2=.186; high frequency above chance, averaged 61.7%; low frequency at chance, averaged 47.1%). The learning order effect (F(1,62)=.501, p=.482, η^2=.008) and interaction effect (F(1,62)=.001, p=.973, η^2<.001) were non-significant.

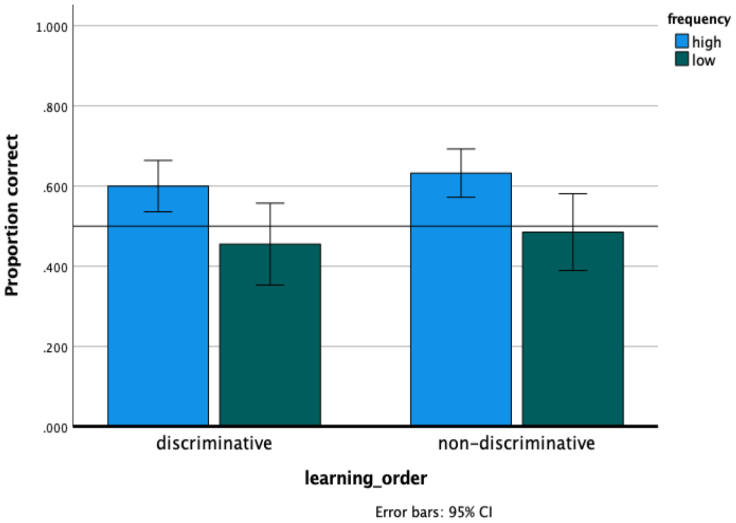

Figure 8. Accuracy of discriminative learning test trials in test 1 under different conditions in Experiment 2. Error bars indicate 95% confidence intervals computed in SPSS. Line indicates chance performance.

Similar to the analysis for discriminative learning test trials in Experiment 1, low frequency items were taken out to run an independent t-test, in order to find if learning order plays a role for these infrequent items. No significant difference between the two learning orders was found (t_{62}=-.424, p=.673).

Comparison with Experiment 1 - It is interesting to compare the results of Experiment 1 and 2, since they differed in terms of the total number of training trials and thus we can see the effect of exposure. Further analyses were therefore run with experiment treated as an additional between-subject factor. For each of the baseline and discriminative learning test trials, a 2 (frequency) × 2 (learning order) × 2 (experiment) ANOVA was carried out. The results suggested that in addition to the frequency effects discussed above, there were main effects of experiment for both baseline

(F(1,100)=19.606, $p<.001$, $\eta^2=.164$) and discriminative learning test trials (F(1,100)=5.913, $p=.017$, $\eta^2=.056$), with higher accuracy in Experiment 2 than in Experiment 1 in all learning conditions. None of the two way or three way interactions with experiment were significant.

Test 2: generalization test

The analyses for test 2 were also similar to that conducted in Experiment 1. The foil types (untrained and trained pictures) were added as a within-subject factor into the ANOVA tests.

The two charts in Figure 9 present the proportion of correct answers for test trials with untrained (left) and trained (right) foils respectively. Similar to the findings in Experiment 1, a 2 (frequency) × 2 (learning order) × 2 (foil type) ANOVA showed main effects of frequency and foil type. High frequency items were discriminated at a higher accuracy (average 78.1%) than low frequency items (average 45.1%) (F(1,62)=44.931, $p<.001$, $\eta^2=.420$). Also, accuracy in test trials with untrained foils were higher than that with trained foils (F(1,62)=26.475, $p<.001$, $\eta^2=.299$). A marginal significant effect of learning order was observed (F(1,62)=3.168, $p=.080$, $\eta^2=.049$) but the interaction with frequency was not significant (F(1,62)=2.794, $p=.100$, $\eta^2=.043$). Additionally, accuracy on high frequency items were all above chance (as shown in the blue bars); for low frequency items this differed by condition: performance was above chance with untrained foils in discriminative order, at chance level with untrained foils in non-discriminative order and with trained foils in discriminative order, and below chance with trained foils in non-discriminative order.

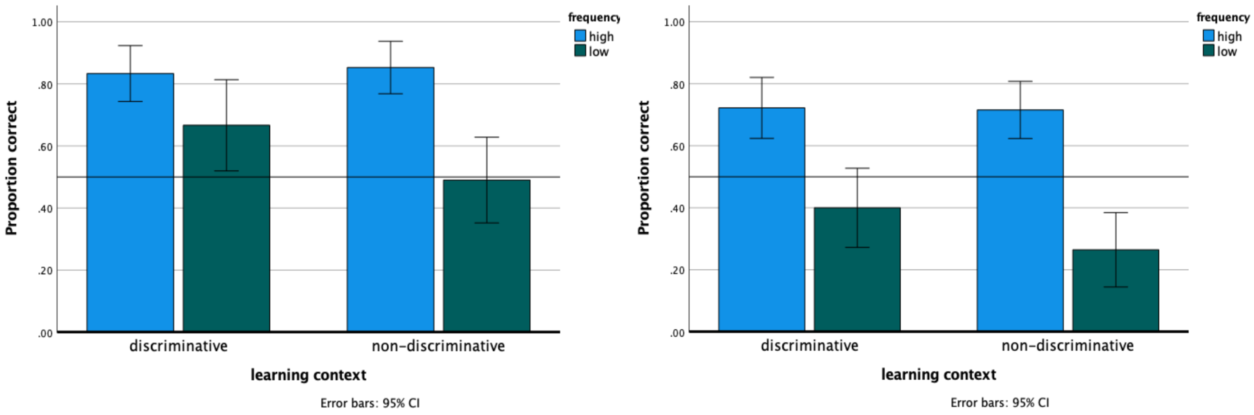

Figure 9. Accuracy with untrained (left) and trained (right) foils under different learning contexts in test 2 of Experiment 2. Error bars indicate 95% confidence intervals computed in SPSS. Line indicates chance performance.

When taking into account only the low frequency items and running a learning order × foil type ANOVA, the results were different from that in Experiment 1: there was again a significant effect of foil type (F(1,62)=16.046, $p<.001$, $\eta^2=.206$), but critically, also an effect of learning order (F(1,62)=4.713, $p=.034$, $\eta^2=.071$); however, there was no interaction (F(1,62)=.112, $p=.739$, $\eta^2=.002$).

Comparison with Experiment 1 - Experiment 1 and 2 were again compared against each other by adding a factor of experiment in the ANOVA test. A 2 (frequency) × 2 (learning order) × 2 (foil type) × 2 (experiment) ANOVA was conducted to explore this four-way interaction. There were significant main effects of foil type (F(1,100)=33.945, $p<.001$, $\eta^2=.253$), frequency (F(1,100)=77.362, $p<.001$, $\eta^2=.436$) and learning order (F(1,100)=4.454, $p=.037$, $\eta^2=.043$). There was also a significant interaction between frequency and foil type (F(1,100)=5.705, $p=.019$, $\eta^2=.054$). No other significant main effect or interaction was found.

Since an effect of learning order was found for low frequency items in Experiment 2 but not Experiment 1, another 2 (learning order) × 2 (foil type) × 2 (experiment)

ANOVA was conducted to see if there is any interaction between learning order and experiment for low frequency items. No significant interaction effect was found (F(1,100)=1.158, p=.284, η^2=.011).

Post-test question

Thirty-nine participants (18 in the discriminative group) reported that they heard tonal differences in the artificial language. Three participants (one in the discriminative group) successfully captured that the use of tones did not contrast meanings in this language. For question 2, two participants (both in the non-discriminative group) noted that there were length contrasts in pronunciations of the artificial language. One additional participant from the discriminative group explicitly mentioned that he/she used consonant rather than tone for discrimination in the tests.

5.3 Discussion

Experiment 2 explored whether the learning order effect on non-native speech sound acquisition would occur with increased exposure to the artificial language. The predictions were thus exactly the same as in Experiment 1. I again discuss findings from each test in turn.

Test 1: familiar noun test

In baseline learning test trials, there was a frequency effect as in Experiment 1, although here the low frequency items were at (rather than below) chance. This is as predicted because increased exposure to the low frequency items could lead to improved performance on these words and the accuracy on low frequency words was less severely affected by the harder trials (with foils that share tone with the high frequency word of the picture) than that in Experiment 1. Indeed a comparison between the two experiments (further discussed below) indicated that longer exposure has led to generally higher performance in this experiment (which for low frequency items, is enough to take accuracy to chance level rather than below chance). In addition, there was an unexpected main effect of learning order in the overall ANOVA test, however there was no interaction with frequency and this effect was not found looking at the low frequency items alone, which is where any effect of discriminative learning should be strongest. This suggests that this effect should be treated with caution - it could be type 1 error.

In the discriminative learning test trials, the findings were similar to that of Experiment 1 as well, where there was only an effect of frequency (with higher accuracy for high frequency items) but no difference between learning order groups, although again here the low frequency were at (rather than below) chance, in line with generally stronger performance in this experiment. Low performance in the low frequency condition indicates that even with double the number of training trials, participants were still unable to capture the consonantal cue sufficiently to identify the correct picture for the low frequency items even in the discriminative order. It is possible that the acquisition of the geminate cue to contrast lexical meanings is generally too difficult for Chinese speakers in a short training period without explicit instructions. This possibility will be further discussed in the general discussion section.

Experiment 1 & 2 comparison: There was a main effect of experiment for both types of test trials. This suggests that the greater exposure in Experiment 2 resulted in better performance in tests as compared to Experiment 1. It is worth noting that the participants in the two experiments were recruited through slightly different methods (Experiment 1 through social media and Experiment 2 through an experiment recruitment website), and though unlikely, it cannot be ruled out that this difference in recruitment methods may have an impact as well as exposure. No interaction with experiment was found to be significant, indicating that there is no evidence that the patterns of the data in test 1 are different across the experiments.

Test 2: generalization test

Similar to Experiment 1, trials with untrained foils had higher accuracy than those with trained foils, and high frequency items were discriminated at a higher rate than low frequency items. What is different from Experiment 1 is that there was a marginally significant order effect, with the discriminative learning group outperforming the non-discriminative learning group. This order effect though marginal is suggestive, since it is in line with the predictions of the discriminative learning model. Furthermore, the learning order effect was statistically significant when looking at the low frequency items alone, which is where a benefit of discriminative learning is most strongly predicted. These results suggest that there is tentative evidence for an overall benefit of discriminative learning order (i.e. across high and low frequency items) and significant evidence specifically for low frequency items.

There is no interaction between foil type and either the overall effect of learning order or the effect of learning order in the low frequency condition. This means that no evidence shows that the learning order effect is greater in one of the foil conditions than the other. However, when looking at the patterns of the mean accuracy across conditions, it is clear that the discriminative group was above chance for low frequency items in the untrained foil trials and at chance for the trained foil trials. This fits with an explanation that discriminative learning participants better captured the characteristic of the artificial language that tones do not contrast lexical meanings and hence can reject the untrained foils, but they had not learned the consonantal contrast strongly enough to be able to reject the distractor in the trained condition.

Experiment 1 & 2 comparison: There was no main effect of experiment, suggesting no evidence that the increased exposure in Experiment 2 brought further overall benefits for generalization. However, there was also no interaction between experiment and learning order even in the low frequency condition alone, indicating no evidence for the difference between the two experiments in terms of learning order effect, though the order effect was observed in Experiment 2 but not Experiment 1. These apparently contradictory findings will be addressed further in the general discussion, but one possible reason is that a greater sample is needed to generate sufficient power for a significant interaction effect to be observed.

Post-test question

The three participants who noted that tones were not used lexically in the artificial language and the other three who noticed the length contrasts (with one who explicitly mentioned using it for discrimination) performed well in discriminating low frequency items in general, with an above-chance average accuracy of 70.1% in test 1 and 61.1% in test 2. This indicates that there were participants who developed explicit knowledge and they successfully used the consonantal cue to discriminate low frequency words in the tests. But it is worth noting that there were other learners who did as well in tests and did not report the explicit knowledge, suggesting implicit learning mechanisms at play. Again, I have to acknowledge the limitations of this type of self-reporting – which is returned to in the general discussion.

6. General discussion

The present study examined the acquisition of non-native speech sounds under different learning order conditions, and investigated the hypothesis of the discriminative learning model, that is, whether presenting cues before outcomes (a discriminative order) would facilitate speech sound learning. In Experiment 1, there was strong learning of high frequency words, however no evidence for learning of low frequency words was found at all, even in the discriminative learning condition. In Experiment 2, although there was still weak performance on low frequency items and no benefit of discriminative learning in test 1, evidence was found for stronger learning in the discriminative order condition in the generalization test (test 2). In this section, I discuss potential reasons for the generally weaker learning in this study compared to previous studies, especially Nixon's (2020) research. I also address the question of why the findings of the two tests in Experiment 2 had different patterns in terms of learning to dissociate tones and learning to discriminate consonantal cues.

Although the central idea of the current study was inspired by Nixon's (2020) experiment and the design largely replicated previous discriminative learning experiments (e.g., Nixon, 2020; Ramscar et al., 2010), this study examined the acquisition of different acoustic cues from Nixon's research. Specifically, Nixon (2020) explored the acquisition of non-native tonal cues by English speakers, together with the unlearning of segmental cues. In this study, participants were native speakers of a tonal language and were trained with non-native segmental cues instead. An important difference is that though in both cases a non-native contrast was trained, the relative ease of learning the non-native contrasts may be different for the different learner groups. To be specific, although tonal cues are not used to contrast meanings lexically, pitch variations that carry linguistic functions are not totally absent in English. For example, English makes use of pitch changes to mark intonation in declaration (falling pitch at the end of sentence) and interrogation (rising pitch at the end of sentence). Pitch could also be used to convey attitudes or emotions in communication (Chuenwattanapranithi, Xu, Thipakorn, & Maneewongvatana, 2008; Xu, Kelly, & Smillie, 2013). Thus, English speakers should be sensitive to pitch changes in speech to some extent, which may facilitate their perception and acquisition non-native tonal patterns since lexical tones are also encoded via pitch changes. In contrast, for Chinese speakers in the present experiment, the consonant length contrast is entirely missing in their native language, which may make the acquisition of such non-native contrasts more difficult. Even though it was found that Chinese speakers could perceive the consonant length differences (as in the pilot study), the lack of any functional use of consonant length in their native language may distract learners’ attention from this cue in the acquisition process and they may treat this cue simply as a form of variability (similar to talker variability).

Additional support for this explanation comes from Nixon’s (2020) Experiment 1, in which pre-exposure to a tonal feature blocked the acquisition of a following nasality cue but not vice versa. The author further found that the accuracy for the tonal cue was significantly higher than that for the nasality cue during the pre-training phase, indicating that the tonal cue was acquired at a faster rate and hence was fully acquired during pre-training. This suggests that English speakers could easily capture the pitch differences in speech sounds and fully associate them with different visual outcomes. It is consistent with the assumption that the difference in the relative learning difficulty of the non-native cues in Nixon’s and the current study could be responsible for the different findings. In other words, the complete absence of the examined consonantal feature in Chinese could be one reason for participants’ failure in consonantal length acquisition, with the learning difficulty factor overcoming the order effect.

Another unintended point of difference that might have reduced learning of the discriminative cue is that the current study used pictures of daily objects (e.g. crown, pencil, shoe) to imitate more natural language learning conditions, whereas the visual pictures in Nixon’s design had two salient semantic information associated with them, colour and shape (e.g. ‘yellow triangle’). On reflection, one possibility is that Nixon’s stimuli may have encouraged participants to link two novel words to each picture, one referring to the colour and the other the shape. Since objects in the current study did not have precisely two salient visual features, participants may be biased that each object had one label only and thus less inclined to associate two labels with one object.

As a result, they may learn only the more frequent label, even in the discriminating condition. This fits with the mutual exclusivity bias, which suggests that adults as well as kids tend to assume that one object only have one associated label (e.g. Au & Glusman, 1990; Liittschwager & Markman, 1994), though this is just a bias rather than an absolute constraint.

Although the expected learning order effect is absent in familiar noun tests (test 1) of the two experiments, the generalization test of Experiment 2 did provide some support for the discriminative learning model. When acoustic cues were presented before visual outcomes (the discriminative order), participants better acquired the pattern that tones are not used discriminatively in the artificial language. This is consistent with the predictions of the discriminative learning model. In the cue-outcome learning condition, participants can make predictions on the potential visual outcomes after hearing the acoustic stimuli, and discrepancies between their predictions and actual outcomes would generate prediction error. Participants may quickly start making predictions on the basis of association between the tone (as well as those segmental features that they can discriminate) and the picture it most frequently associated with. For low frequency items, this resulted in participants predicting incorrect pictures, and then when they saw the correct pictures, they got prediction error. Participants can learn from the prediction error and figure out that discriminating on the basis of tones were inaccurate, leading them to downweight the tonal cue gradually. In contrast, participants in the outcome-cue order did not have the opportunity to make predictions on the basis of the cues and thus encounter prediction error. This explains the better performance for participants in the discriminative learning order in the generalization test. It is worth mentioning that in Experiment 1 this learning order difference was not observed, indicating that dissociating the non-discriminative tonal cue may require more exposure than participants received in Experiment 1. However, no interaction effect of experiment by learning order was obtained when treating experiment as an independent factor, making it harder to interpret the difference between the experiments. A replication of the experiments with higher power (and greater sample) would be needed to explore if the missing of the interaction effect arises from insufficient power.

Given that we saw facilitation of the discriminative learning order in the generalization test in Experiment 2, why didn’t we also see this in the discriminative learning test trials in test 1? One possibility is that the direct discrimination between two similar sound stimuli in test 1 was too difficult for participants even in the discriminative condition, as although they have learned that tone is not discriminatory, they may have not yet fully learned that the geminate consonantal contrasts *are* discriminatory. This explanation is in line with the finding that participants were at or below chance in trials with trained foils in test 2, as these trials were more challenging and required them to recognize the geminate consonant cue accurately. Though the discriminative learning model predicts that in the cue-outcome condition learners should eventually learn that the gemination cue is discriminatory, it seems likely that in this study, at least with the amount of exposure in Experiment 2, learners had enough exposure to start downweighting the salient tonal cue but perhaps not enough exposure to learn the discriminatory consonantal cue sufficiently to succeed in the discriminative learning test trials of test 1.

Although this explanation is consistent with the findings, I also have to acknowledge the fact that the forms of test 1 and test 2 were different: in test 1 participants were provided with two spoken words as options for each picture, and were mapping from pictures to words, whereas in test 2 one spoken word was presented in each trial and participants were mapping from spoken words to visual pictures. This latter test is more similar to the discriminative learning order (i.e. making predictions on visual outcomes from audio stimuli). Thus, the observation of the order effect in test 2 could arise from the design of the test. It would be interesting to explore this assumption by employing similarly designed test trials for test 1 and see if an order effect would be found.

The post-test questions intended to explore whether participants developed any explicit knowledge about the tonal and consonant length cues or learned the acoustic patterns implicitly. Note that discriminative learning is supposed to be a theory of *implicit* learning, as is typical for learning processes that are relevant in child L1 development as well as adult L2 development. The answers in Experiment 2 were interesting since a few participants seemed to explicitly capture the acoustic cue manipulation at least partially (i.e. the tonal cue was non-discriminatory and the consonantal cue was discriminatory), and these learners did well in the tests. However this data needs to be interpreted with caution because, for example, it is possible that participants learned the cues implicitly but when answering the questions they were guided to think back and figure out the patterns explicitly. This possibility is consistent with the fact that some other participants who also had high scores in tests did not show explicit knowledge when answering the questions. Thus, it suggests that at least some participants learned the acoustic cues implicitly, although a caveat here is that they might actually have learned explicitly but have difficulty articulating the patterns. The limitations in interpreting the self-report data shows that this is an imperfect way of exploring whether learning is implicit or explicit. Future work could augment the current paradigm using real time processing methods such as visual world eyetracking or ERP. For example, Batterink, Reber, Neville and Paller (2015) looked at a left parietal effect which has been found to reflect the amount of information recollected or retrieved during processing (Rugg & Curran, 2007; Vilberg, Moosavi, & Rugg, 2006), with larger brain response amplitude indicating better recollection, in response to familiar and unfamiliar test items. It was found that greater brain responses were associated with familiar words than unfamiliar words, showing that there was potential explicit recollection of information during the processing of familiar words. This ERP method can be incorporated into the current discriminative learning experiment. If participants develop explicit knowledge of the acoustic patterns, it is possible that the trained foil trials in generalization tests would elicit greater brain responses than the untrained foil trials because more recollection would be involved when the target and foil options are both trained and familiar.

Overall, the findings of the experiments provide tentative evidence for the discriminative learning model. Under this associative account, learning is not simply probabilistic, but is facilitated by learners' predictions. However, it does not mean that the discriminative learning model fails to account for the probabilistic pattern in speech sound acquisition as mentioned in Section 2.1.2. Rather, the model can explain the statistical distribution in terms of predictive value. Specifically, infants are exposed to a number of acoustic cues and gradually associate various cues with outcomes (e.g. objects and actions around them). Those acoustic features that are used contrastively in their native languages gain predictive value in predicting outcomes, and other uninformative features lose associative strength with outcomes over time due to cue competition. This results in infants losing sensitivity to non-native features and acquiring the distribution of their native phonology.

6.1 Implications for theoretical accounts of non-native speech learning and for teaching

The contribution of the current study is primarily theoretical since the purpose is to examine a theoretical approach to language learning. The results of the study (especially Experiment 2) are consistent with the discriminative learning account of language learning, and suggest a re-weighting of acoustic cues that are relevant in learners' L1 but not the L2 (in this experiment, the tonal cue) during non-native sound learning. This has the potential to address a common transfer problem in L2 acquisition, meaning that learners may transfer features from their L1 that are irrelevant or used differently in an L2 (e.g., Lombardi, 2003; Simon, 2010). The current results indicate that discriminative learning may play a role in reducing the transfer of irrelevant features from L1 (e.g. tone). However, the findings also emphasize the difficulty of learning cues that are not used discriminatively in learners' L1 at all (e.g. the geminate consonant cue for Chinese speakers). Even though the discrimination may be relatively easy (as demonstrated in the pilot study), it could nevertheless be hard for learners to fully acquire these features and use them to contrast meanings accurately.

Although this study was not designed to test specific teaching methods or practices and it does not relate to teaching directly, ultimately the discriminative learning theory may be useful in developing language training materials that, for example, present stimuli in ways that encourage formation of prediction error. The method in this study was largely artificial and it would take further work to develop into actual training programs. Furthermore, the relatively weak performance on low frequency items indicate that for some non-native features that are absent in learners' native language, this type of implicit training might not be most efficient given limited input. Future work can explore whether adding in more explicit teaching and feedback would be more effective. However, with the short period of training employed in this study, the effect of implicit learning methods cannot be completely disregarded, as it is possible that extended training (with multiple training sessions) could lead to better performance. It is also notable that most experiments that use explicit feedback contained multiple training sessions (e.g. Logan, Lively, & Pisoni, 1991), and it would be unfair to directly compare the implicit learning in this study with explicit learning given different amounts of training period.

6.2 Limitations and future directions

Some limitations and potential improvements have been identified throughout the dissertation, I summarize and extend the key points here. Firstly, I mentioned two possibilities for the absence of the learning order effect in test 1 but not test 2 in Experiment 2, one noting that participants may have dissociated the tonal cue but have not yet learned the geminate cue, and the other pointing out that the form of test 2 (but not test 1) may inadvertently benefit the discriminative condition. For the first possibility, it would be interesting to repeat the experiment with longer exposure, and see if this allows participants to learn the geminate cue better. If so, a learning order effect would emerge in test 1 as well. The second possibility can be established by matching the forms of the two tests. Secondly, I noted that the visual stimuli in this study were different from Nixon's (2020), which can unintentionally lead to word learning biases. Thus, the current experiments can be replicated using the same pictures as Nixon's research. Thirdly, the current power may be insufficient to see if there is indeed difference between test 2 of Experiment 1 and 2 in terms of the learning order effect. A replication of the experiments with higher power may provide stronger evidence for the interaction effect. Fourthly, it was discussed that the investigation of implicit versus explicit learning can be achieved in future studies through some online processing paradigms (e.g. eyetracking and ERP) rather than the short answer questions used in the current design. Additionally, in terms of the analysis of the current results, only group level performance was considered. However there could be considerable individual differences, with some participants well above chance but some below in the same learning condition. Future research with larger sample sizes can explore whether there are other factors (e.g. short-term memory, attention) that can account for the individual differences.

Future discriminative learning experiments can also consider the relative learning difficulty of the cues as a factor that influences learning processes. In Nixon's (2020) and Ramscar et al.'s (2010) experiments, the informative cues (i.e. tonal patterns in Nixon's design, visual shapes in Ramscar et al.'s) posed relatively low level of learning difficulty, since these cues can be found in the (linguistic and visual) environment around learners. In the current design, a totally absent cue tended to cause greater difficulty in learning. Future experiments can directly compare the acquisition of cues of different degrees of difficulty under different learning order conditions, and observe if there is any interaction between learning difficulty and learning order effect. Furthermore, later experiments can be carried out using the same set of stimuli, but with participants of different language backgrounds. For example, examining native English speakers' performance in such design and comparing it with Chinese speakers' responses can provide insight on whether the familiarity with the uninformative (tonal) cue influences the acquisition of other informative, non-native cues.

Last but not the least, a general limitation of the use of artificial language is that it reduces the "naturalness" of the test. Although the artificial language uses phonetic features extracted from natural languages, it is not an actually existing natural language. In particular, in this study, the artificial language was designed to have distinct tones that do not contrast lexical meanings, which is unusual in natural languages. It also follows Nixon's (2020) design in using multiple exact synonyms (e.g. *bevve_rising* and *casa_falling* both refer to 'crown') which, though may occur, are not very common in natural languages. Despite the unnatural features of the artificial language, the design allowed us to conduct an "in principle" test of how the order of stimuli affects learning and to examine critical predictions of a theoretical model within a short-term learning experiment. Overall, most empirical learning experiments are to some extent unnatural with controlled manipulations, but it is important to eventually extend these experiments to the longer term learning of natural languages.

7. Conclusion

In summary, the current study examined the acquisition of a non-native speech sound contrast by adult learners under a discriminative learning approach, which suggests that language learning (and in this particular case, speech sound learning) is driven by cue competition and prediction error. This theoretical approach precisely predicts that the order in which training stimuli (i.e. cues and outcomes) are presented influences the learning and generalization of linguistic features. The findings of this study provide tentative evidence that learning can be facilitated by a discriminative learning order, though the evidence was found in the generalization test rather than the familiar noun test. This is likely to reflect that the discriminative order led to better dissociation of tones (i.e. learning that tone is *not* a relevant cue) in the language, but not yet strong learning of the actual discriminatory cue of consonant length, possibly due to the difficulty of learning word mappings involving this non-native speech sound for Chinese speakers. Overall, the study contributes to the examination of theoretical language learning accounts and suggests direction for future research which may ultimately have implications for the development of language teaching materials.

References

Abramson, A. S. & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences,* Prague, pp. 569–573.

Allan, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin, 114*, 435-448.

Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches.* Cambridge University Press.

Annau, Z., & Kamin, L. J. (1961). The conditioned emotional response as a function of intensity of the us. *Journal of Comparative and Physiological Psychology, 54*(4), 428.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r. *Journal of Phonetics, 32*(2), 233-250.

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science, 35*(6), 1105-1138.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition, 122*(3), 292-305.

Au, T. K. F., & Glusman, M. (1990). The principle of mutual exclusivity in word learning: To honor or not to honor? *Child development, 61*(5), 1474-1490.

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

Best, C. T., Hallé, P., Bohn, O. S., & Faber, A. (2003, August). Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages. In *Proceedings of the 15th international congress of phonetic sciences* (Vol. 28892892). Barcelona, Spain: Causal Productions.

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, 14*(3), 345.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of phonetics, 20*(3), 305-330.

Best, C. T., Traill, A., Carter, A., David Harrison, K., & Faber, A. (2003). !Xóõ click perception by English, Isizulu, and Sesotho listeners.*15th International Congress Of Phonetic Sciences*.

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

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*(2), B69-B77.

Chládková, K., & Paillereau, N. (2020). The what and when of universal perception: A review of early speech sound acquisition. *Language Learning, 70*(4), 1136-1182.

Chomsky, N. (1957). *Syntactic Structures.* Berlin: Mouton de Gruyter.

Chomsky, N. (1959). A review of BF Skinner's verbal behavior. *Language, 35*(1), 26–58. JSTOR.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2008). Encoding emotions in speech with the size code — A perceptual investigation. *Phonetica, 65,* 210-230.

Chung, K. K. (2003). Effects of Pinyin and first language words in learning of Chinese characters as a second language. *Journal of Behavioral Education, 12*(3), 207-223.

Colunga, E., Smith, L. B., & Gasser, M. (2009). Correlation versus prediction in children's word learning: Cross-linguistic evidence and simulations. *Language and Cognition, 1*(2), 197-217.

Cramer, R. E., Weiss, R. E, Steigleder, M. K., & Balling, S. S. (1985). Attraction in context: Acquisition and blocking of persondirected action. *Journal of Personality & Social Psychology, 49,* 1221-1230.

Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of phonetics, 39*(3), 388-402.

De Houwer, J., & Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. *The Quarterly Journal of Experimental Psychology: Section B, 55*(4), 289-310.

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology, 36*(1), 29-50.

Divjak, D., Milin, P., Ez-zizi, A., Józefowski, J., & Adam, C. (2021). What is learned from exposure: an error-driven approach to productivity in language. Language, *Cognition and Neuroscience, 36*(1), 60-83.

Dobel, C., Lagemann, L., & Zwitserlood, P. (2009). Non-native phonemes in adult word learning: evidence from the N400m. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1536), 3697-3709.

Dye, M., Ramscar, M., & Suh, E. (2011). For the price of a song: How pitch category learning comes at a cost to absolute frequency representations. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Egger, M. D., & Miller, N. E. (1962). Secondary reinforcement in rats as a function of information value and reliability of the stimulus. *Journal of Experimental Psychology, 64*(2), 97.

Ellis, N. C., & Sagarra, N. (2010). The bounds of adult language acquisition: Blocking and learned attention. *Studies in Second Language Acquisition, 32*(4), 553-580.

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America, 130*(4), EL206-EL212.

Galle, M. E., Apfelbaum, K. S., & McMurray, B. (2015). The role of single talker acoustic variation in early word learning. *Language Learning and Development, 11*(1), 66-79.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning. *Journal of Experimental Psychology: General, 117*, 227- 247

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds" L" and" R.". *Neuropsychologia, 9*(3), 317-323.

Gulian, M., Escudero, P., & Boersma, P. (2007). Supervision hampers distributional learning of vowel contrasts. In *Proceedings of the 16th International Congress of Phonetic Sciences* (Vol. 2, pp. 1893-1896). Saarbrücken: University of Saarbrücken.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of phonetics, 32*(3), 395-421.

Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive psychology, 63*(2), 93-106.

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

Jusczyk, P. W. (1995). Language acquisition: Speech sounds and the beginning of phonology. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 263– 301). San Diego, CA: Academic Press.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants′ sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*(5), 630-645.

Kamin, L. J. (1968). Attention-like processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior:* Aversive stimulation. Miami: Miami University Press.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America, 66,* 1668– 1679.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science, 9*(2), F13-F21.

Liittschwager, J. C., & Markman, E. M. (1994). Sixteen-and 24-month-olds' use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology, 30*(6), 955.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874-886.

Lombardi, L. (2003). Second language data and constraints on manner: Explaining substitutions for the English interdentals. *Second Language Research, 19*(3), 225-250.

Maye, J., & Gerken, L. (2000, March). Learning phonemes without minimal pairs. In *Proceedings of the 24th annual Boston university conference on language development* (Vol. 2, pp. 522-533).

Maye, J., Weiss, D. J., & Aslin, R.N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science, 11*, 122–134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, 101–111.

Mielke, J. (2008). *The Emergence of Distinctive Features.* Oxford University Press.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18*(5), 331-340.

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental science, 13*(3), 407-420.

Nixon, J. S. (2020). Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition, 197,* 104081.

Nixon, J. S., & Tomaschek, F. (2020). Learning from the acoustic signal: Error-driven learning of low-level acoustics discriminates vowel and consonant pairs. In CogSci.

Nixon, J. S., & Tomaschek, F. (2020). Prediction and error in early infant speech learning: A speech acquisition model. *Cognition, 212*, 104697.

Olejarczuk, P., Kapatsinski, V., & Baayen, R. H. (2018). Distributional learning is error-driven: The role of surprise in the acquisition of phonetic categories. *Linguistics Vanguard, 4*(s2).

Pajak, B., & Levy, R. (2011). Phonological generalization from distributional evidence. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Pavlov, I. P. (1927). *Conditioned reflexes*. London: Oxford

Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America, 102*(6), 3742-3753.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning and acquisition of phonology. *Language and Speech, 46,* 115–54.

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

Polka, L., & Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America, 100*(1), 577-592.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human perception and performance, 20*(2), 421.

Poltrock, S., Chen, H., Kwok, C., Cheung, H., & Nazzi, T. (2018). Adult learning of novel words in a non-native language: Consonants, vowels, and tones. *Frontiers in psychology, 9,* 1211.

Postman, L., & Greenbloom, R (1967). Conditions of cue selection in the acquisition of paired associate lists. *Journal of Experimental Psychology, 73,* 91-100.

Prince, A. & Smolensky, P. (1993). Optimality Theory: constraint interaction in generative grammar. In J. McCarthy (ed.), *Optimality Theory in Phonology: A Reader* (pp. 3–71). Oxford: Blackwell.

Prince, A. & Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar.* Oxford: Blackwell.

Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological science, 24*(6), 1017-1023.

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of ”mouses” in adult speech. *Language,* 760–793.

Ramscar, M., Dye, M., Popick, H. M., & O'Donnell-McCarthy, F. (2011). The enigma of number: Why children find the meanings of even small number words hard to learn and how we can help them do better. *PloS one, 6*(7), e22501.

Ramscar, M., Matlock, T., & Dye, M. (2010). Running down the clock: The role of expectation in our understanding of time and motion. *Language and Cognitive Processes, 25*(5), 589-615.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of featurelabel-order and their implications for symbolic learning. *Cognitive Science, 34*(6), 909–957.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of cs in fear conditioning. *Journal of Comparative and Physiological Psychology, 66*(1), 1.

Rescorla, R. A. (1988). Pavlovian conditioning: It’s not what you think it is. *American psychologist, 43*(3), 151.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. *Classical conditioning II: Current research and theory, 2*, 64–99.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental science, 8*(2), 162-172.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental science, 12*(2), 339-349.

Rudy, J. W (1974). Stimulus selection in animal conditioning and paired-associate learning: Variations in the associative process. *Journal of Verbal Learning & Verbal Behavior, 13*, 282-296.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in cognitive sciences, 11*(6), 251-257.

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

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy, 4*(2), 273-284.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of phonetics, 52,* 183-204.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *The Quarterly Journal of Experimental Psychology Section B, 37*(1b), 1-21.

Simon, E. (2010). Phonological transfer of voicing and devoicing rules: evidence from L1 Dutch and L2 English conversational speech. *Language Sciences, 32*(1), 63-86.

So, C. K. (2005). The effect of L1 prosodic backgrounds of Cantonese and Japanese speakers on the perception of Mandarin tones after training. *The Journal of the Acoustical Society of America, 117*, 2427.

Sundland, D. M., & Wickens, D. D. (1962). Context factors in paired-associate learning and recall. *Journal of Experimental Psychology, 63*(3), 302.

Terry, J., Ong, J. H., & Escudero, P. (2015). Passive distributional learning of non-native vowel contrasts does not work for all listeners. In *ICPhS*.

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America, 120*(4), 2285-2294.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1994). Discrimination of English/rl/and/wy/by Japanese infants at 6-12 months: language-specific developmental changes in speech perception abilities. In *Third international conference on spoken language processing.*

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child development,* 466-472.

Underwood, B. J., Ham, M., & Ekstrand, B. (1962). Cue selection in paired associate learning. *Journal of Experimental Psychology, 64*, 405-409.

Vilberg, K. L., Moosavi, R. F., & Rugg, M. D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain research, 1122*(1), 161-170.

Vujović, M., Ramscar, M., & Wonnacott, E. (2021). Language learning as uncertainty reduction: The role of prediction error in linguistic generalization and item-learning. *Journal of Memory and Language, 119,* 104231.

Wagner, A. R. (1969). Stimulus selection and a “modified continuity theory”. *The Psychology of Learning and Motivation, 3*, 1–41.

Wagner, A. R., Logan, F. A., & Haberlandt, K. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology, 76*(2p1), 171.

Wanrooij, K., Boersma, P., & Benders, T. (2015). Observed effects of “distributional learning" may not relate to the number of peaks. A test of "dispersion" as a confounding factor. *Frontiers in psychology, 6,* 1341.

Wayland, R., & Guion, S. (2003). Perceptual discrimination of Thai tones by naïve and experienced learners of Thai. *Applied Psycholinguistics, 24*(1), 113-129.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7,* 49– 63.

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science, 21*(4), 221-226.

Wynne, C. D. L. (1995). Reinforcement accounts for transitive inference performance. *Animal Learning & Behavior, 23,* 207-217.

Xu Y., Kelly A. & Smillie C. (2013). Emotional expressions as communicative signals. In: S. Hancil & D. Hirst (Ed.), *Prosody and Iconicity.* Amsterdam: John Benjamins Publishing Company, 33-60.

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy, 15*(4), 420-433.

Appendix A: Visual stimuli

Pictures that used as visual referents in the experiment:

‘crown’

‘shoe’

‘pencil’

‘apple’

‘car’

Appendix B: CUREC approval letter

Note: The dissertation project is covered under my supervisor's existing research ethics approval (letter attached below). All names in the letter have been blanked out.

SOCIAL SCIENCES & HUMANITIES INTERDIVISIONAL RESEARCH ETHICS COMMITTEE

Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD
Tel: +44(0)1865 616576 Fax: +44(0)1865 280467
ethics@socsci.ox.ac.uk

6 January 2021

Department of Education

Dear

Research Ethics Approval (CUREC 2)
Ref No: R73484/RE001
Title: Language learning in adults and children

The above application has been considered on behalf of the Social Sciences and Humanities Interdivisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to the IDREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly approval has been granted.

Should there be any subsequent changes to the project that raise ethical issues not covered in the original application you should submit details to the IDREC for consideration:

https://researchsupport.admin.ox.ac.uk/governance/ethics/apply/sshidrec#collapse394916.

Please note that your study may be selected for review by the SSH IDREC during an annual audit. You may also be required to submit a brief annual progress report on each anniversary of study approval, until the study is completed.

Yours sincerely,

Research Ethics Manager

cc: