Perceptual flexibility during second language learning: the role of language experience in Korean-English bilinguals' pitch perception

by

Miroo Lee

Bachelor of Arts, Hanyang University, 2014

Master of Arts, Hanyang University, 2016

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Miroo Lee

It was defended on

March 4, 2024

and approved by

Alan Juffs, Professor, Department of Linguistics

Claude E. Mauk, Teaching Professor, Department of Linguistics

Seth Wiener, Associate Professor, Department of Modern Languages

Dissertation Director: Melinda Fricke, Assistant Professor, Department of Linguistics

Perceptual flexibility during second language learning: the role of language experience in Korean-English bilinguals' pitch perception

Miroo Lee, PhD

University of Pittsburgh, 2024

Listeners selectively tune in to the most relevant cues for contrasting sounds, and this process impacts their perceptual sensitivity to these cues (Nosofsky, 1986; Pisoni, 1991). For bilingual listeners, recent work suggests a shared L1/L2 phonology system with the phonetic properties of each sound established in a language-specific way (Dmitrieva, 2019; Garcia-Sierra et al., 2009). While this work has predominantly examined changes in phonemic boundaries across languages, there has been much less work examining language-specificity in cue-weightings or perceptual sensitivity. Further, as these aspects of perception have primarily been explored through the lens of cumulative effects of long-term language experience, our understanding of short-term language experience, sometimes termed 'language mode', remains unclear. The present study seeks to bridge this gap by examining the effects of both long-term and short-term linguistic context on both cue reliance and perceptual sensitivity.

The current study exploits the fact that pitch contrasts stop consonants in Korean, but not in English (Lisker & Abramson, 1964), providing a testing ground to assess the extent of perceptual changes across languages. Three research questions are posed: 1) How does the contrastive use of pitch in a listener's first language warp perceptual sensitivity? 2) To what extent are bilinguals' perceptual patterns affected by language mode? And 3) What is the relationship between second language learning history and language-specific listening strategies?

Perception of pitch of Korean-English bilinguals and English monolinguals was tested utilizing a series of listening tasks (discrimination, goodness-rating, and labeling) in two distinct sessions. In each session, listeners heard acoustically identical target stimuli, within language-specific experiment settings meant to invoke English and Korean listening modes. Results showed that Korean-English bilinguals exhibited perceptual warping of pitch through acquired similarity: pitch differences were harder to detect within Korean stop categories. This effect was limited to listeners' L1 Korean language mode, suggesting perceptual warping can be mediated by listeners' relative reliance on cue. Significant correlations were also found between L2 exposure and both L1 cue-weighting and cross-language perceptual shift, demonstrating that long-term immersion in a second language can have consequences for the organization of L1 phonetic space.

Table of Contents

List of Tables

List of Figures

Preface

I would like to thank my wonderful advisor, Melinda, for her mentorship and support throughout my PhD journey.

1.0 Introduction

Speech sounds inherently vary. The acoustic signals I generate to say “this is our cat” are not the same as those generated by my parents, who have a prominent Korean accent. Despite the variations, many of you might perfectly understand both my parents and me, getting ready to meet our feline companion. However, some might struggle to process the acoustic signal, failing to understand what my parents or I are saying until later they meet our cats and then realize maybe what they heard earlier was “this is our cat”. The point is, dealing with variation is crucial in the speech perception process, whether these variations are due to dialectal differences, accents, or even a momentary speaker fatigue leading to mumbling. So, how do we manage to recognize these diverse signals as consistent sound categories? What does learning a category mean for our perceptual development? Does knowing we are listening to English versus Korean make a difference in our perceptual strategies? To what extent can we adapt our perceptual strategies for different languages, if any adjustments are possible? These are the questions I explore in my dissertation.

The goal of the research to be reported is to examine whether speaking a language that uses pitch to distinguish sound categories affects listeners’ perceptual sensitivity to pitch, and how such effects are modulated by listeners’ linguistic exposure in both the long and short term. To achieve this objective, an empirical study examined the listening performance of Korean-English bilinguals and English monolinguals in categorization and segmental discrimination of stimuli varying in pitch, within English and Korean language contexts. The study aimed to reveal the perceptual shaping resulting from the contrastive use of pitch by comparing the segmental discrimination patterns of two groups (Korean-English bilinguals vs. English monolingual listeners) in their respective native language. Secondly, the study aimed to discover how the relation between cue-reliance (i.e. how strongly a listener typically weights pitch information during perception) and acoustic sensitivity is further modulated by language-specific listening strategies by comparing perceptual patterns of bilingual listeners in their L1 to L2 language contexts. Lastly, the study aimed to explore the relationship between any language-specific perceptual shifting and individual listeners' language learning history by examining effects of L2 proficiency level and cumulative exposure. The study employs frameworks associated with exemplar-based theories, making reference to work by Nosofsky (1986) and Gibson (1963) that posit language processing is based on the storage and retrieval of specific instances of speech sounds, words, and linguistic patterns, rather than abstract representations and phonological rules.

In the rest of this chapter, I will introduce the phonetic cue under examination, namely pitch (with a focus on level pitch), and how it is used in the context of Korean and English. Subsequently, I will examine prior literature regarding the impact of linguistic categories on perceptual sensitivity. Following that, I will review research studies that investigated pitch perception among speakers of tonal languages in particular. Finally, I will present the research questions and hypotheses.

1.1 Use of F0 in Korean vs. English

The current study focuses on the perception of Korean-English bilinguals and English monolinguals due to the significant divergence in pitch usage between the Korean and English languages. This divergence provides an interesting context for investigating the role of category membership in pitch discrimination. While English has a two-way system of stop phonemic contrasts including voiced and voiceless categories (Lisker & Abramson, 1964), Korean features a three-way system of stop phonemic contrasts known as fortis, lenis, and aspirated (Kagaya, 1974; Lisker & Abramson, 1964). These three categories are distinguished by a combination of two acoustic cues: Voice Onset Time (VOT) and Fundamental Frequency (F0) (Bang et al., 2018; M.-R. Cho, 1994; T. Cho et al., 2002; Kang & Han, 2013), the perceptual correlate of which is pitch. These features are encoded in the consonantal (VOT) and vocalic (pitch) portion of the sounds, and together convey the relevant information for stop contrasts. In Korean, stop sounds with sufficiently low F0 on the vowel immediately following the stop (for our purposes, equivalent to "low pitch") are primarily categorized as *lenis* unless they have significantly short VOT. The rest of the stop sounds with mid or high F0 on the following vowel are further categorized based on VOT, with long VOT corresponding to *aspirated* and short VOT corresponding to *fortis* (Bang et al., 2018; Schertz et al., 2015).

In contrast, in English, voiced and voiceless sounds are primarily distinguished by VOT values (Lisker & Abramson, 1964), with F0 playing a secondary role to a much smaller extent when the primary cue is ambiguous (Francis et al., 2008). This comparison highlights the role of F0, especially in the context of long VOT conditions. In Korean, stop sounds with extended VOT could potentially represent two distinct categories (*lenis* and *aspirated*), whereas in English, they are categorized under a single category (*voiceless*) (see Figure 1).

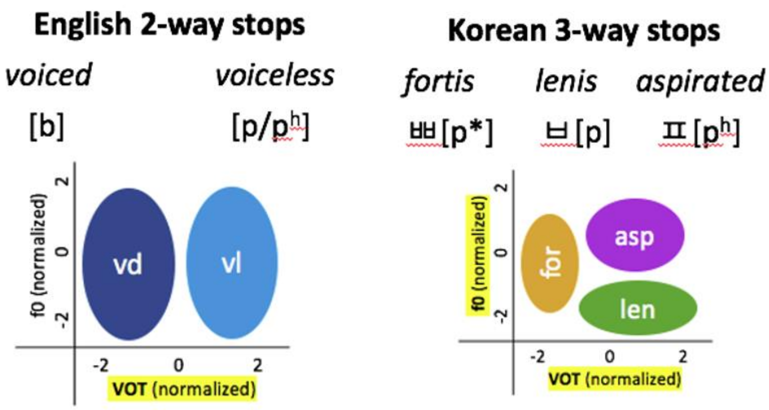

Figure 1 Stop Contrasts in English vs. Korean

Highlighted axes are those that are critical for categorization.

Previous studies on native Korean speakers' perception and production of L2 English stops revealed an intriguing pattern. These investigations highlighted how native Korean speakers utilize both Voice Onset Time (VOT) and Fundamental Frequency (F0) to differentiate L2 English stops in both production and perception (Schertz et al., 2015). In their production data, these speakers tend to produce English voiceless stops with high F0 and long VOT, somewhat resembling the characteristics of Korean aspirated stops, although the acoustic domains of the two categories do not overlap, with Korean aspirated sounds characterized by slightly higher F0 values as compared to English voiceless stops (Schertz et al., 2015). The same individuals demonstrated the similar use of cues in perception of English sounds, consistently categorizing high F0 and long VOT sounds as voiceless stop sounds.

However, what's particularly interesting is that when these individuals encountered non-canonical acoustic cues in L2 English perception, such as sounds with long VOT and low F0 or short VOT and high F0, they showed great individual variance in the relative reliance on VOT and F0 (Schertz et al., 2015, 2016). For example, Schertz et al. (2016) found that some listeners only categorized long VOT and high F0 sounds (canonical stimuli) as voiceless stops without responding voiceless to non-canonical stimuli. This indicates that these listeners relied on both

VOT and F0 to the same degree (Figure 2). On the other hand, other listeners not only categorized the canonical stimuli (long VOT and high F0) as voiceless stops but also some portions of the non-canonical stimuli as voiceless stops as well. Some listeners responded voiceless to long VOT and low F0, indicating they relatively relied more on VOT than F0 (Figure 2). On the other hand, other listeners responded voiceless to short VOT and high F0, suggesting their relative reliance was greater for F0 than VOT. Intriguingly, this individual variability in perception did not seem to be associated with L2 English proficiency levels (Schertz et al., 2015). We will return to the issue of individual variance later, but for now, the takeaway point is that Korean-English bilinguals show significant individual variance in relative reliance on F0 and VOT cues in perceiving noncanonical stimuli in an L2 English context, providing an interesting context to examine individual variance in relation to their L2 language experience.

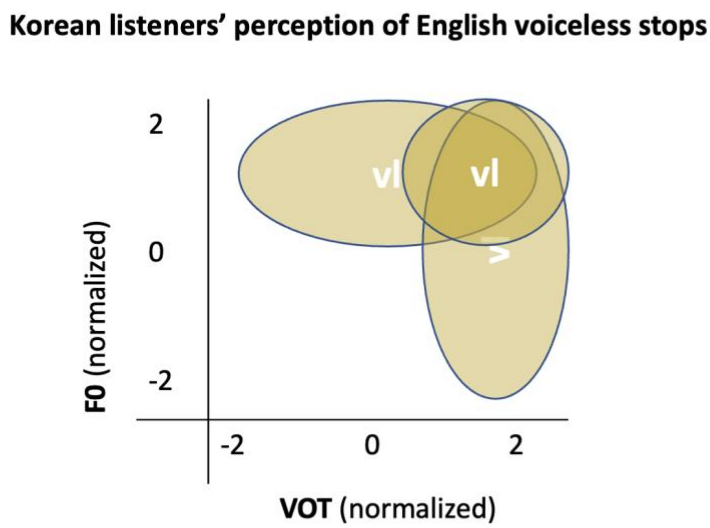

Figure 2 Individual Variance in Korean Speakers' Perception of L2 English Voiceless Stops

Each ellipse depicts one listening pattern observed in Schertz et al. (2015).

1.2 Category Learning and Perception Models

Studies show that expert language users can selectively tune into the most diagnostic acoustic parameters for contrasting sounds during speech perception (Francis & Nusbaum, 2002; Goldstone, 1994; Heeren & Schouten, 2008; Nosofsky, 1986). Such selective attention can have impacts on perceptual sensitivities, potentially restructuring our psychological space and perceived similarities between stimuli (Gibson, 1963; Goldstone, 1994; Nosofsky, 1986; Shepard, 1957). Several studies have also demonstrated that listeners' speech categories can affect their perceptual sensitivities (Francis & Nusbaum, 2002; Iverson et al., 2003; Liberman et al., 1957; Pisoni & Lazarus, 1974). In this section, I will delve into theories and experiments concerning changes in how individuals perceive a set of stimuli. In particular, I will focus on research that has explored the relationship between category learning and perception, as these studies help us understand how the speech perception process accommodates variation in acoustic signals to form coherent sound categories. While perceptual development and perceptual learning are technically not the same (Gibson, 1963), these two terms will be used interchangeably in my dissertation to refer to any changes in perception due to experience or practice.

A central concept in category learning is the notion of *perceptual distance*, which plays a crucial role in numerous models that seek to elucidate the mechanism behind categorization (Gibson, 1963; Goldstone, 1994; Nosofsky, 1986, 2011; Shepard, 1957). The idea of perceptual distance (or psychological distance) can be found in early studies on generalization (Shepard, 1957), where perception of similarities among various stimuli is interpreted as an index of perceptual distance. Shepard (1957) introduced a Multidimensional Scaling (MDS) model, which mathematically formalizes perceived similarity between items as a function of the psychological distance between them: the closer the psychological space between items, the more similar they are perceived to be, leading to increased confusion in identification. This concept of psychological distance is integrated into various perceptual models, such as the Generalized Context Model (Nosofsky, 1986), the localized warping view (Goldstone, 1994), and Natural Language Magnet (Kuhl, 1993).

Nosofsky (1986) introduced the Generalized Context Model (GCM) to conceptualize how selective attention affects our perceptual space (Nosofsky, 1986, 2011). According to the model, in situations where multiple dimensions are available, individuals might selectively focus on a specific sensory dimension, assigning greater perceptual weight to that dimension. This perceptual weight stretches the psychological space along the highly attended, relevant dimension while diminishing it along the unattended, irrelevant dimensions (Nosofsky, 1986, 2011). This mechanism finds application in phonetic learning within both first and second language acquisition contexts (Bradlow et al., 1997; Iverson et al., 2003; Jusczyk, 1993; Kuhl, 1993): learners may have to readjust their attention by allocating more focus to previously disregarded acoustic patterns and possibly directing their attention away from the information previously prioritized in their native language (Francis & Nusbaum, 2002; Pederson & Guion-Anderson, 2010). The GCM also assumes that individuals store every encountered token of a perceptual category in memory (Nosofsky, 1986). This is the foundation for exemplar-based theories, which highlight experiential representation of linguistic categories, rather than storage of category prototypes and abstract language rules (Johnson, 2007; Jusczyk, 1993; Nosofsky, 1986; Pierrehumbert, 2001).

One of the central tenets in the GCM, pertinent to the present study, posits that changes in perceptual weight result in stretching or shrinking a dimension **uniformly** across its entire range. To illustrate, consider the visualization of GCM's prediction of categorization of the stimulus *i* (Figure 3). In the top panel, which illustrates perception without selective attention, *i* is more or less equally distant from stimuli A or B. In the bottom panel, the horizontal dimension received selective attention, resulting in the horizontal distances between items being stretched while vertical distances between items are shrunk. This results in *i* being closer to A2 than any other items, leading the model to predict the item *i* to be categorized as A with higher probability. Note that this uniform application of perceptual weight to the horizontal dimension also predicts greater distance between A5 and A4, items belonging to the same category A. In other words, the GCM predicts that psychological distance will increase along the same dimension (which is the horizontal dimension in this case) regardless of whether the stimuli belong to the same category (as in the case of A5 and A4) or different categories (as in the case of A2 and B4).

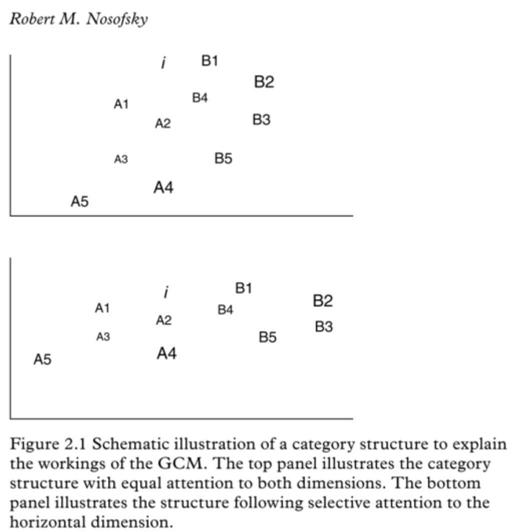

Figure 3 Visualization of the GCM

Reproduced with the permission from Nosofsky, Robert M., E. M. Photos, and A. J. Wills. "Formal approaches in categorization." (2011). Copyright 2011 Cambridge University Press.

In contrast to the prediction made by GCM, several studies reported evidence of reduced sensitivity *within* the same dimension in relation to category (Gibson, 1963; Liberman et al., 1957; Schouten & van Hessen, 1992; Werker, 2018). For example, Liberman et al. (1957) found that English listeners often exhibit better discrimination between physically different VOT stimuli when the sounds come from different categories rather than from the same category. In other words, listeners' perceptual sensitivity along the VOT dimension is better within the across-category region while it is worse within categories. Such observations contradict GCM, as the GCM model posits that psychological distances across the attended VOT dimension will equally increase, resulting in equally good discrimination of sounds along the VOT dimension regardless of category membership.

Following the perceptual changes that seem to be affected by the category membership, the localized warping model was suggested (Francis & Nusbaum, 2002; W. Heeren & Schouten, 2008; Iverson et al., 2003; Strange & Dittmann, 1984) which conceptualizes the changes in perceptual space in two different mechanisms. In the *acquired distinctiveness* view, learning categories enhances discriminability across category boundaries (Francis & Nusbaum, 2002; W. F. L. Heeren & Schouten, 2010; Iverson et al., 2003; Strange & Dittmann, 1984), therefore resulting in greater perceptual sensitivity at the across-category region. This can also be accompanied by the *acquired equivalence* view, which posits decreased discriminability for stimuli within a category (Goldstone, 1994; Pisoni, 1973, 1975; Stagray & Downs, 1993), predicting poor perceptual sensitivity at the within-category region. While these two views often accompany each other, they do not necessarily appear together. For example, previous work shows mixed evidence of acquired distinctiveness, acquired equivalence, or both (Francis & Nusbaum, 2002; Guenther et al., 1999; W. Heeren & Schouten, 2008; Livingston et al., 1998). Livingston et al. (1998) suggested that when categories are sufficiently differentiable before training, training may serve primarily to increase within-category similarity, suggesting different mechanisms of perceptual warping may arise depending on the stimuli type.

Several studies exploring the acquisition of new phoneme categories have drawn insights from the localized warping model (Francis & Nusbaum, 2002; W. Heeren & Schouten, 2008; Iverson et al., 2003; Strange & Dittmann, 1984). In a study by Francis and Nusbaum (2002), American English listeners were trained to categorize Korean stops using the F0 acoustic dimension, which (recall from Section 1.1) is contrastive in Korean but not in English. Following the training, English listeners exhibited acquired similarity within categories and acquired distinctiveness between categories in their discrimination performance, lending support for both acquired similarity and distinctiveness. Similarly, Iverson et al. (2003) observed category learning among Japanese and American English listeners along the F3 dimension in a /ra-la/ continuum. Within-category discrimination was higher for Japanese listeners, while across-category discrimination was higher for American listeners, indicating increased perceptual sensitivity around category boundaries for American listeners through acquired distinctiveness. These perceptual effects appear to be related to language experience, displaying a more native-like discrimination peak with greater language exposure (Francis et al., 2008; W. Heeren, 2004; W. Heeren & Schouten, 2008).

Evidence supporting acquired equivalence is also evident in studies examining the perceptual development of infants where babies demonstrate better discrimination of sounds relevant to their native language than language-irrelevant sounds (Pisoni, 1973, 1975; Werker & Tees, 1984). These findings are bolstered by a significant correlation between increased language experience and reduced sensitivity to within-category differences in infants: infants under 6 months show better discrimination of nonnative distinctions than nonnative adults, a phenomenon known as broad-based initial sensitivities (Werker, 2018). However, across the first year, infants' broad-based initial sensitivities undergo perceptual attunement to their native language, leading to decreased discrimination of nonnative phonetic distinctions and improved discrimination of native ones (Werker, 2018).

The concept of acquired equivalence aligns with the Native-Language Magnet (NLM) model (Kuhl, 1993, 2004), as proposed by Patricia Kuhl. The NLM suggests a "magnet effect," where category prototypes shape sensitivity, attracting similar variants toward category prototypes and thus reducing discrimination among stimuli within a category (Iverson et al., 2003; Kuhl, 1993, 2004). Differently from the localized warping model, the NLM entails only "assimilation" where perceptual space narrows but not the opposite, where space expands, which would make discrimination easier across category boundaries (Iverson et al., 2003; Kuhl, 1993, 2004).

Another notion related to the localized warping view is *categorical perception*. Categorical perception is originally proposed as a way of accounting for a certain pattern in listeners' perceptual data (Liberman et al., 1957). Categorical perception relies on observing a relationship between two tasks: a categorization task and a discrimination task. In a categorization task, participants listen to a series of sounds and classify each sound into one of several categories. Therefore, the categorization data offer insights into how listeners map the input into phonological categories as well as where the location of the boundary between categories is. On the other hand, in a discrimination task, participants listen to pairs of sounds and determine whether they are identical or different. The discrimination data provide information on where listeners demonstrate heightened or diminished perceptual sensitivity. Combining the categorization and discrimination data, the phenomenon of categorical perception is characterized by a discrimination peak aligned at the category boundary with the discrimination performance being strongly predictable from identification performance.

Several experimental studies revealed that categorical perception is not limited to speech sounds (Cutting, 1982; Cutting & Rosner, 1974; Miller et al., 1976; Mirman et al., 2004) or to humans (Kuhl & Miller, 1975; Kuhl & Padden, 1983), suggesting that the categorical perception effects may be the combined effects related to the domain-general phenomenon and language experience rather than only from the language experience. This perspective posits that phoneme boundaries in speech may simply occur at regions of heightened natural auditory sensitivities. Despite variability, studies frequently replicate the tendency for listeners to discriminate better between-category differences than within-category differences (Liberman et al., 1957; Pisoni, 1975; Strange, 1984; Van Hessen & Schouten, 1999; but see Kutlu et al., 2022 for alternative explanations).

The notion of categorical perception is also used as a framework for understanding speech perception. A strong form of categorical perception posits that listeners learn to 'ignore' any acoustic variations within category, failing to access gradient information within category (Liberman et al., 1957). In such a case, listeners' sensitivity within category is predicted to be at chance level. However, robust categorical perception patterns are not often found in empirical data; instead, the data often exhibit weak, loose categorical patterns influenced by various experimental factors, such as the type of discrimination task (Chen et al., 2015; Chen & Peng, 2016; Gerrits & Schouten, 2004; Wu & Lin, 2008) and stimulus quality (Liu & Kager, 2014; Van Hessen & Schouten, 1999). The inconsistencies in findings, coupled with the assertion of categorical perception in its strongest form, where auditory data is encoded solely in discrete categories without consideration for gradient information access within categories, have sparked criticism of the categorical perception framework (Apfelbaum et al., 2022; Kutlu et al., 2022; Massaro & Cohen, 1983; McMurray, 2022).

Massaro and Cohen (1983) investigated whether listeners' speech perception behavior is categorical vs. continuous, employing an extreme version of categorical perception that auditory information is only encoded as categories. They used the visual analogue scale (VAS) method, where listeners rate sounds by moving a pointer along a linear scale with categories A and B at each end. The researchers hypothesized that if the categorical perception model is accurate, ratings would be influenced solely by category references, resulting in a bimodal distribution of responses. For example, an average goodness rating of 5 (out of 10-point scale) for sounds would likely stem from a bimodal distribution of rating responses, with many low ratings for sound A and many high ratings for category B. On the other hand, if the continuous perception model is true, ratings would instead track the variations in the stimuli themselves. In this case, an average goodness rating of 5 for sounds would arise from a single distribution of rating responses clustered around the rating of 5, reflecting that listeners had perceived them as category ambiguous.

As alluded to earlier, one of the criticisms of the strong form of categorical perception is that the framework rejects access to gradient information within or even between categories, assuming sounds are rigidly encoded as belonging solely to category A or B. In contrast, the localized warping model views auditory input as encoded with some degree of warping. In this view, listeners can still encode category ambiguous sounds as between-category sounds, stretching the perceptual space. This enables discrimination not only among sounds from different categories but also among sounds that are outside of categories (category ambiguous). This is a crucial difference from the strong form of categorical perception which does not explicitly predict discrimination within ambiguous regions, suggesting that these sounds are only encoded as part of a category rather than outside of categories.

So far, we have compared the GCM and the localized warping model in terms of their distinctive predictions regarding perceptual space. Both the localized warping model and the GCM support the idea that higher-level cognition, such as category learning, can influence lower-level perceptual processes. However, unlike the GCM, the localized warping model predicts distinct changes in perceptual sensitivity along the relevant dimension based on category membership. The GCM can explain differences in the overall perceptual sensitivity toward certain cue(s), where a listener attending to a relevant cue demonstrates higher perceptual sensitivity compared to listener not attending to the same cue. In contrast, the localized warping model effectively accounts for differences in sensitivity within a dimension where higher sensitivity is observed between categories while lower sensitivity is observed within categories. Importantly, though, these models are not necessarily mutually exclusive, and supporting both models may be conceivable if data indicate substantial group differences (GCM) and pronounced category effects (localized warping). For example, the GCM predicts that native Korean listeners will show higher perceptual sensitivity than English listeners, indicated as a between-group difference. On the other hand, localized warping suggests that Korean listeners will show higher perceptual sensitivity at the across-category than the within-category region, indicated by significant effects within the Korean group. Such localized warping pattern is not expected for English listeners, as their perception of F0 stimuli is not affected by different category memberships.

1.3 Pitch Perception and Language Experience

Tonal languages like Mandarin Chinese, Cantonese or Thai, rely on pitch variations to distinguish words (Yip, 2002), termed "contrastive" use of pitch. In contrast, nontonal languages like English or French use pitch more for pragmatic purposes rather than altering word meanings (Hirst & Di Cristo, 1998). The distinct linguistic function of pitch has prompted numerous research endeavors seeking to unravel whether the perception of pitch varies based on listeners' experience with the linguistic function of pitch (Abramson, 1961, 1979; Gandour, 1983; Gandour & Harshman, 1978; Krishnan et al., 2005, 2009; Swaminathan et al., 2008; Xu et al., 2006). This section will review some of the studies that used the framework of categorical perception to explore how the contrastive use of pitch affects listeners' perception. As mentioned earlier, categorical perception relies on observing a relationship between two tasks: a categorization task and a discrimination task. Thus, the approach employed in these studies generally includes both identification and discrimination tasks in the methodology, aiming to locate where, if any, a discrimination peak is present.

Despite variations in tonal inventory, tones can be broadly classified into two categories: level versus contour tones, depending on perceived pitch movement. Level tones are generally perceived as static, while contour tones convey dynamic movements such as rises or falls (Abramson, 1979; Pike, 1994). Early investigations into speech perception of tones, exploring the influence of native tonal experience on pitch perception, utilized continua featuring level tones with varying pitch heights (Abramson, 1961, 1979; Francis et al., 2003). Abramson (1961, 1979) examined Thai speakers' pitch perception in identification and discrimination tasks using Thai level tones. The results showed no evidence of categorical perception; discrimination performance was consistent across the continuum without a distinct peak near the category boundary (Abramson, 1961, 1979). This result was taken as evidence that level tones are not perceived categorically.

Francis et al. (2003) conducted a series of tests on Cantonese listeners' perception, incorporating continua with a mix of level and contour tones in Cantonese. They found that the perception of pitch height versus contour exhibited different patterns. In the case of a continuum involving only level tones (and varying only in pitch height), Cantonese speakers' discrimination performance did not show a clear discrimination peak, but instead exhibited good discrimination performance across the entire stimuli overall. This lack of a clear relationship between the identification and discrimination tasks is in line with view that perception of level tone is not categorical. On the other hand, when listening to a continuum that shifted from a level to a rising pitch pattern, Cantonese speakers' discrimination performance aligned with predictions based on identification performance, featuring a smooth but significant discrimination peak near the category boundary (Francis et al., 2003). Similar outcomes were reported in other studies when stimuli continua include contrasts between contour and level tones (Chan et al., 1975; Hallé et al., 2004; Liu et al., 2017; Peng et al., 2010; Xu et al., 2006).

Studies exploring the perception of tones by non-tonal listeners have yielded varied results. The studies by Hallé et al. (2004) and Xu et al. (2006) reported that non-tonal speakers (French and English listeners) did not demonstrate a discrimination peak compared to tonal speaker groups (Mandarin Chinese listeners). However, Chan et al. (1975), Peng et al. (2010) and Liu et al. (2017) reported a significant discrimination peak for both tonal and non-tonal speakers. Interestingly, the discrimination peaks in non-tonal speakers seemed to arise from different mechanisms compared to tonal speakers (Chan et al., 1975; Liu et al., 2017; Peng et al., 2010). While tonal speakers tend to be influenced by the phonological value of tone contours, showing peak discrimination near category boundaries, nontonal speakers appear to be guided by psychophysical factors, often exhibiting peak discrimination as contour tones transition into level tones (Hallé et al., 2004; Liu et al., 2017; Peng et al., 2010; Xu et al., 2006). For instance, Chan et al. (1975) conducted a study with Chinese and English speaker groups using a Mandarin tones continuum from rising to level tones. The Chinese group showed a smooth yet significant discrimination peak near the category boundary, aligning well with their categorization patterns. In contrast, the English group exhibited a discrimination peak that did not align with their categorization pattern; instead, their peak occurred near the level tone, suggesting the peak may have been caused by some psychophysical factor. This interpretation aligns with work demonstrating that discriminating between level tones is relatively easier than discriminating stimuli involving contour tones (Francis et al., 2003).

Another interesting finding concerning category effects on tone perception is the indication of *diminished* sensitivity to pitch differences among tonal speakers (Liu et al., 2017; Stagray & Downs, 1993). The observation of diminished sensitivity takes a broader approach to investigating the effects of category, looking at not only the discrimination peak but any changes in perceptual sensitivities in relation to category. In the study by Stagray and Downs (1993), a comparison between Mandarin and English speakers was conducted to assess their ability to detect pitch differences. The results revealed that Mandarin speakers needed a higher threshold to detect pitch differences compared to English speakers. The diminished sensitivity observed in Mandarin speakers was interpreted as the result of stimuli being categorized within the same level tone category (i.e. acquired equivalence, in the localized warping framework); pitch differences within a given category were ignored to optimize perception efficiency (Stagray & Downs, 1993). Similarly, Xu et al. (2006) examined the perception of level-to-rising continua in both speech (Mandarin /i/) and nonspeech (harmonics) conditions among Chinese and English listeners. Their results indicated that Chinese speakers exhibited higher discrimination sensitivity compared to English speakers in between-category regions but showed reduced discrimination sensitivity compared to English listeners within the category. In another study by Liu et al. (2017), the discrimination performance of Mandarin and Dutch speakers on a level-to-falling continuum of

Mandarin tones was compared. The results revealed that Dutch speakers outperformed Mandarin listeners in discriminating sounds within Mandarin tone categories.

So, why do differences arise between the perception of level and contour tones? Wang (1976) offered a psychophysical explanation, suggesting that distinguishing a level tone from a rising contour might be inherently easier than differentiating between two rising contours. This implies that level tones could be inherently more discriminable, potentially overshadowing any evidence of heightened sensitivity to a particular phonologically relevant region. In an alternative interpretation, Xu et al. (2006) proposed that pitch height lacks intrinsic reference and necessitates extrinsic reference, either from the context (e.g., the larger phrase or sentence) or the best exemplar from memory (e.g., prior information about the speaker). The idea is that extrinsic references exhibit weaker categorical perception effects, and this could explain the almost nonexistent categorical perception effects with level tones compared to the observed degree of categorical perception effects with contour tones. Given these considerations, there is a basis to suspect that the perception of level pitch may engage distinct mechanisms compared to the perception of contour tones. This divergence might be influenced by varying degrees of acquired distinctiveness versus acquired equivalence in processing.

Lastly, research investigating the evidence of categorical perception of tones employed a variety of discrimination tasks (Table 1). These include the AX task, where listeners hear two stimuli on each trial and are asked to respond whether they are the same or different; the AXB task where listeners hear a triplet of stimuli and are asked to respond whether the second stimulus X matches the first (A) or the third (B) stimuli; the ABX task where listeners hear a triplet of stimuli and are asked whether the last stimulus X matches the first (A) or the second (B); and AABX task where listeners hear one pair consisting of identical members (AA) and another pair consisting of different members (BX), and are asked which pair contains different members. It is commonly posited that ABX tasks impose a higher cognitive/memory load, based on the relatively short time span of auditory memory (200-300 msec) and the assumption that the auditory memory traces may fade away by the time X is presented, resulting in listeners relying more on the abstract label. ABX tasks are therefore thought to tap into a more abstract level of representation compared to AX tasks (Liu et al., 2017). AXB tasks are considered a variant of the ABX tasks, although listeners may often ignore the third stimulus (Van Hessen & Schouten, 1999). The compiled research consistently shows use of the AX and ABX discrimination tasks. In the case of level tones, neither ABX (Abramson, 1961) nor AX tasks (Francis et al., 2003) provided evidence of a discrimination peak. Conversely, for contour tones, a discrimination peak was evident in both AX tasks (Francis et al., 2003; Liu et al., 2017; Peng et al., 2010; Xu et al., 2006) and ABX discrimination tasks (Abramson, 1961; Chan et al., 1975; Hallé et al., 2004; Liu et al., 2017). Because the current study uses an AX discrimination task to examine perception of a level pitch contour, previous results suggest that we may not see evidence of 'categorical perception' in a conventional sense, defined by a discrimination peak near the category boundary. However, it remains to be seen whether the perceptual sensitivity to level pitch differences will be impacted by language experience in other ways.

Table 1 Summary of Literature on Categorical Perception Effects in Tone Languages

	Tasks	Stimuli	Subjects	Results
Abramson 1961	Identification ABX discrimination	Thai Mid-high	Thai	no peak
Chan et al. 1975	Identification ABX discrimination	Mandarin rising-level	Chinese English	C: Peak at category boundary E: Peak near level tone region
Abramson 1979	Identification AABX discrimination	Thai High-mid-low	Thai	No peak
Stagray & Downs 1993	discrimination	Pure tone	Chinese English	C: poorer sensitivity than E
Francis et al. 2003	Identification	Cantonese	Cantonese	Level-level: no peak

	AX discrimination	Level-level Level-rising Falling-Lrising-Hrising		Level-contour: peak at boundary Contour-contour: peak btw falling and rising. No peak btw Lrising and Hrising.
Hallé et al. 2004	AXB Identification AXB discrimination	Mandarin Level-contour Contour-contour	Taiwanese French	T: peak near category boundary F: no peak
Xu et al. 2006	Identification AX discrimination	Mandarin Level-rising	Chinese English	C: peak at category boundary + poor sensitivity within-category E: no peak
Peng et al. 2010	Identification AX discrimination	Level-contour	Mandarin Cantonese German	M: peak at category boundary C: peak at category boundary G: peak near level tone region
Liu et al. 2017	Identification AX discrimination AXB discrimination	Mandarin Level-contour	Chinese Dutch	C: peak near level region + poor sensitivity within-category (AX) D: peak near level region (AX) C: peak near category boundary (AXB) D: peak near level region

1.4 Bilinguals' Language-Specific Perceptual Strategies

While the previous section provided an overview of how pitch perception may be affected by long term language experience, this section considers the broader question of how bilingual sound categories are established. Flege's Speech Learning Model (SLM; Flege, 1995) and its revised version (SLM-r; Flege & Bohn, 2021) predict development of L2 phonological categories based on their phonetic similarity to L1 phonemes. New L2 phonological categories are more likely to be produced and perceived accurately if they are dissimilar from the closest L1 phonemes. The Perceptual Assimilation Model (PAM; Best, 1995) and its extension to L2 learning (PAM-L2; Best & Tyler, 2007) predict success of discriminating non-native speech *contrasts* (i.e. pairs of contrastive sounds) based on the assimilation patterns of L2 sounds to native speech sounds. While both SLM and PAM make predictions about the difficulty in nonnative speech perception based on the similarity relations between L2 and L1 sounds, the two are different in that PAM emphasizes the discrimination between nonnative *contrast*. This means that although PAM can help understand how well listeners can tell apart different L2 categories, it doesn't directly predict how accurately they can discriminate much smaller sound differences within or between L2 categories.

While SLM and PAM primarily address non-native speech perception, the Attrition and Drift in Access, Production and Perception Theory (ADAPPT) (de Leeuw & Chang, 2023) focuses on the changes in L1 resulting from L2 acquisition. Based on the assumption that a person's L1 is always subject to change in adulthood, the ADAPPT proposes two types of changes in L1 speech: attrition and drift. The two effects share similarities as they both include L1 changes in speech perception and production, occurring within both the segmental and prosodic levels after late L2 acquisition. However, the two effects are different in that drift effects are proposed to be relatively short-term, resulting from recent L2 input where bilingual individuals are less likely to be aware of change nor have agency in the change. On the other hand, attrition effects are usually long-term, resulting from cumulative L2 input. In the case of attrition, bilingual individuals are more likely to be conscious of change and have more agency in it. It is important to note that, on the basis of the malleability of L1 over the lifespan assumed in the ADAPPT, attrition does not indicate permanent language changes but rather relatively durable changes in the L1 that can be reversed at a later point in life.

While previous literature indicates that long term exposure to a language can have important effects on the perceptual system, there is also reason to believe that the perceptual system is subject to adjustment in the short term as well. The notion of *language mode* refers to the activation state of a bilingual's languages and processing mechanisms at any given time (Grosjean, 1985, 1989, 1998). Grosjean (1985) posits that mode activation is influenced by psychosocial and linguistic factors, which are modulated by the communicative context of an interaction. For instance, a Spanish-English bilingual might be in Spanish unilingual mode when communicating in Spanish, while they might be in English unilingual mode when the situation requires that only English be used. Even in a unilingual mode, the activation of the non-dominant language is never zero but rather relatively lower. This section will review previous investigations into the potential effects of language mode on bilinguals' perception.

Previous studies examining bilinguals' perception in different language modes aimed to discover whether bilinguals adjust category boundaries in language-specific ways. Early studies did not find evidence of language-specific perception (Caramazza et al., 1973; Williams, 1977). But following studies, utilizing more naturalistic stimuli, found evidence of a shift in category boundaries between different language modes (Elman et al., 1977; Flege & Eefting, 1987). For instance, Elman et al. (1977) reported not only that bilinguals exhibited a significant shift in category boundaries between different language modes, but the extent of the shift was positively correlated with the fluency in both languages. Further evidence of boundary shifts was found in Flege and Eefting (1987) where a significant shift was observed along the VOT continuum for the /d-t/ boundary among Dutch-English bilinguals. However, unlike Elman et al. (1977), they did not find a significant correlation between the magnitude of shift and L2 proficiency level. Hazan and Boulakia (1993) also observed a comparable boundary shift over the VOT continuum for the /b-p/ contrast by French-English bilinguals. They did not find any significant relationships between boundary shift and age of acquisition. However, they did report that bilinguals' cue-weighting seemed to be governed by the dominant language of individual listeners. Lastly, research shows the degree of shift can be influenced by speakers' level of confidence in a language (Garcia-Sierra et al., 2009) as well as experimental design factors such as stimuli range (Bohn & Flege, 1993).

Prior studies on language mode effects in bilingual speech perception have primarily focused on the phonetic dimension of voice onset time or VOT (Casillas & Simonet, 2018; Gonzales et al., 2019; Gonzales & Lotto, 2013). In these studies, the tested dimension typically serves as a primary cue for contrasting categories in both languages of bilingual listeners. For example, VOT acts as a primary cue in both English and Spanish, albeit with different ranges (Elman et al., 1977; Williams, 1977). However, research has shown that language mode can also have an impact on bilinguals' cue weighting (Dmitrieva, 2019; Hazan & Boulakia, 1993; Schertz et al., 2015). For instance, Dmitrieva's (2019) study assessed how Russian-English bilinguals' reliance on glottal pulsing versus vowel duration changed between different language modes. She found that Russian-English bilinguals demonstrated a stronger reliance on glottal pulsing in Russian than in English mode, mirroring the monolingual Russian preference for the cue. Similarly, Yazawa et al. (2020) examined the relative cue reliance of Japanese learners of English in Japanese versus English sessions. They found that while Japanese learners predominantly used duration in both Japanese and English mode to categorize the /i:- ɪ/ stimuli, participants tended to rely more on spectral cues and less on durational cues in English mode. Lastly, Antoniou et al. (2012) highlights bilinguals' inclination to attend to language-relevant phonetic information. In this study, Greek-English bilinguals categorized, rated, and discriminated stop voicing contrasts in each language mode. Participants listened to natural tokens of both the English and Greek voicing contrasts within the same language-specific session. The results showed that bilinguals judged the goodness of sounds relative to the phonetic categories of that language. For example, the goodness rating of the English contrast was better in English mode than in Greek mode. These studies collectively underscore the importance of considering language mode in understanding bilinguals' cue weighting.

While the previous studies demonstrate that bilingual listeners may adjust their listening strategies in language specific ways, such as altering category boundaries or modifying reliance on pertinent cues, there is a gap in our understanding of whether these shifts in perceptual strategy also impact listeners' perceptual sensitivity; that is, their ability to detect very small differences along a single acoustic dimension. Some neuroimaging studies provide evidence for language-specific sensitivities to sound differences (Näätänen et al., 1997; Sharma & Dorman, 2000; Winkler et al., 1999; Zhang et al., 2005). To take one example, research by García-Sierra et al. (2012) testing Spanish-English bilinguals' ERPs response during a discrimination task, found that bilinguals demonstrated language-specific pre-attentive listening patterns. In terms of behavioral data, however, only a few studies addressed the effects of language mode on bilingual listeners' perceptual sensitivity (Antoniou et al., 2012; Williams, 1977). Antoniou et al. (2012) compared the discrimination performance of Greek-English bilinguals across language modes, using naturalistic tokens of stop contrasts in each language produced by native speakers. Their results showed consistent discrimination of contrasting speech categories in both language modes. While this suggests that segmental discrimination may be less influenced by language context than other tasks requiring higher-order cognitive processing, such as categorization or goodness rating, it is difficult to determine whether bilinguals' perceptual sensitivity was truly unaffected in this study. This is because the experimental design in Antoniou et al. (2012) could only test listeners' ability to discern between different categories, but not their ability to detect segmental differences smaller than the distance between distinct categories. For example, the design does not tell us how perceptual sensitivity at the within-category or across-category region changes because participants discriminated tokens of categories that were naturally produced by native speakers.

To gain insight into perceptual sensitivity at finer levels, stimuli should incorporate smaller step differences, ensuring that pairs encompass both within-category and across-category distinctions.

There is some evidence that changes in perceptual sensitivity may require longer, cumulative language experience, rather than short-term training. For example, Heeren and Schouten (2008) examined how the acquisition of new phoneme contrasts might influence listeners' sensitivity along the relevant acoustic dimensions. In their study, naive Dutch speakers were trained to learn Finnish phoneme categories /t-t:/; the authors then compared their categorization and discrimination patterns before and after training. The trained Dutch group showed a shift in categorization similarly to that of native Finnish, but their discrimination pattern only marginally changed toward the native Finnish group. Even though both laboratory-trained Dutch speakers and advanced Finnish learners demonstrated Finnish-like categorization performance, only the advanced Finnish learners exhibited a discrimination peak near the category boundary, akin to native Finnish speakers. The authors proposed that the advanced speakers of Finnish with established long term phoneme category representations can make use of the relevant acoustic dimensions better than the Dutch speakers with short-term training. Similarly, Bosch et al. (2000) examined evidence for a "prototype effect" (reduced within-category discrimination) posited by the NLM in the perception of variants of Spanish /ɛ/, Spanish /e/, and non-prototype vowels. They found that their Spanish-Catalan group showed poorer discrimination with their L1 Spanish prototype variants but not with Catalan variants, while the Catalan-Spanish group showed poor discrimination with their L1 Catalan variants but not with Spanish variants. These studies together suggest that effects of category on listeners' perception of L1 may not easily extend to L2.

Major theories in bilinguals' speech perception such as PAM-L2 and SLM-r propose that bilinguals with greater proficiency strive to maintain contrasts between L1 and L2 phonetic categories, which exist in a common phonological space. In addition to proficiency, the onset of L2 learning age is another important factor in L2 speech perception. The critical period framework predicts that later L2 onset age may lead to lower L2 performance due to maturational decay of the ability to acquire language and/or to the interference of the existing linguistic system (Archila-Suerte et al., 2012; Birdsong, 2018; Newport, 1990). However, recent studies challenge this idea, indicating that adults do not permanently lose the auditory resolution for distinguishing non-native speech sounds (Werker, 2018; Werker & Hensch, 2015; Werker & Tees, 1984).

Age of L2 acquisition is also highlighted in the SLM-r (Flege & Bohn, 2021); the development of L1 categories may negatively influence the successful development of L2 sound categories as pre-existing L1 phonetic categories interfere or block the formation of new L2 phonetic categories. While the model does not explicitly implicate "age" as a factor, within the SLM-r framework, age can be used to operationalize the relative development of L1 categories, suggesting that early age of acquisition may lead to better development in L2 categories. The model also predicts effects of learning L2 categories on L1. For example, Flegel and Eefting (1987) found that Dutch-English bilinguals showed even shorter Dutch VOT than monolingual Dutch VOT. The authors suggested that such extreme "enhancement" of the short VOT feature is evidence of new category formation in L2, mediated by the early age of acquisition.

More recently, the ADAPPT model in bilinguals' speech perception and production (de Leeuw & Chang, 2023) emphasizes how an individual's L1 is impacted by L2 by assuming that a person's L1 is always malleable and changes in adulthood once fully acquired. This was further supported by Tice and Woodley (2012) who showed increasing experience in L2 affected the segmental perception of L1, with the change in L1 sounds toward enhancing categorical differences between L1 and L2. Similarly, a study by Kellogg and Chang (2023) provides evidence that exposure to a foreign language can impact listeners' L1 speech perception. These investigations illuminate that the individual variance in L1 perceptual patterns can be partially attributed to the listeners' previous experience with L2.

1.5 Research Questions

The current study attempts to bridge existing research gaps in previous findings by investigating perceptual sensitivity to pitch for Korean-English bilinguals in different language modes. As reviewed in Section 1.1, pitch is used as a primary cue to contrast stops in Korean while the same cue does not contrast stops in English. Therefore, the two languages provide an opportunity to test how listeners' perception may change when the need for an acoustic cue is greater in one language and less in the other. The current study uses labeling, goodness-rating, and discrimination tasks, allowing us to see the connections between category-level and acoustic cue-level listening behaviors. Importantly, the current study systematically varies F0 and VOT. This design allows us to directly test how category membership affects listeners' segmental discrimination along the acoustic dimensions and how such a relationship may interact with language mode.

Below, I outline three primary research questions along with their respective predictions.

RQ1: How does contrastive use of pitch affect Korean listeners' perceptual sensitivity along the F0 acoustic dimension?

The primary question here is whether Korean listeners will show greater or qualitatively different sensitivity to pitch as compared to the control group of English listeners in their L1. If the Generalized Context Model applies to pitch perception, we can predict higher perceptual sensitivity for Korean listeners than English listeners across the entire F0 continuum. In such a case, a significant group difference is predicted without any significant category membership effect on acoustic sensitivity. On the other hand, if the Localized Warping Model applies to pitch perception, we can predict significant effects of category membership on Korean listeners' perceptual sensitivity, with poor discrimination in F0 regions corresponding to within-category regions and/or good discrimination in F0 regions corresponding to the F0 category boundary. Such results would be in line with previously reported categorical perception phenomena. Lastly, we may predict no group difference in perceptual sensitivity between the Korean and English groups, given that the manipulation of F0 in Korean corresponds phonetically to a difference in "level pitch", which in prior studies of tonal language perception showed little effects of listener group.

RQ2: *How does the experimental setting (akin to language mode) affect Korean-English bilinguals' perceptual sensitivity to pitch information?*

Under the language mode framework, we can predict Korean-English bilingual listeners will demonstrate language-specific listening strategies due to changes in language activation across experimental settings. Critically, we can predict a shift in cue-weighting between languages, with greater reliance on F0 in Korean mode and less or no reliance on F0 in English mode. This would align with the findings of redirecting attention from previous literature. However, language mode may not change listeners' perceptual sensitivity to F0, given that such low-level perception has been found to be robust and not easily changeable by conditions in the previous literature.

However, if perceptual sensitivity fundamentally depends on selective attention, the reduced F0 reliance may accompany changes in sensitivity as well.

RQ3: *How does Korean speakers' English learning experience affect their general perceptual sensitivity to pitch?*

Under the PAM-L2 framework, it may be hypothesized that greater proficiency is related to more distinct language-specific listening strategies. In such a case, a positive correlation is predicted between proficiency and the extent of perceptual shift when switching language environments. On the other hand, under the SLM-r, the development of L1 categories might be the determining factor for distinctive L2 categories (and not age, *per se*). In this case, we may predict that participants with earlier L2 age of acquisition (AoA) will more successfully develop separate and distinct L2 categories, which could be reflected in a greater degree of shift between language modes. Consequently, a negative correlation would be predicted between the age of acquisition (AoA) and the degree of language mode effect. Lastly, individual listeners' perceptual patterns may be best accounted for by the extent of their L2 experience, which is predicted by the ADAPPT model.

2.0 Methods

The objectives of the current study are three-fold. First, it aims to investigate how the contrastive use of pitch in L1 affects listeners' perceptual sensitivity along the pitch dimension. Secondly, it seeks to understand the extent of language mode effects on pitch perception. Lastly, the study seeks to identify the sources of individual variance in language-specific cue-weighting strategies by examining listeners' L2 language learning history. To address these objectives, the current study utilized a series of listening tasks (discrimination, goodness-rating, and labeling) to examine pitch perception among an English monolingual and a Korean-English bilingual group.

2.1 Overview of General Approach

In the first analysis chapter, the study explores how the use of pitch in the first language influences perceptual sensitivity along the pitch dimension, presenting a detailed analysis in two phases. As mentioned previously, in Korean, pitch serves to differentiate between aspirated and lenis stop categories. In contrast, English utilizes VOT to distinguish voiced and voiceless stops, with no contrastive use of pitch. Initially, the labeling, goodness-rating, and discrimination performance of both Korean-English bilinguals and English monolinguals in their L1 conditions are juxtaposed. This comparison aims to ascertain category membership across the F0 stimuli and the weighting strategy for the F0 and VOT cues. Subsequently, I evaluate the perceptual sensitivity levels across different F0 ranges by examining the discrimination data.

This analytical approach draws insights from the framework of perceptual development and selective attention (Goldstone, 1994; Shepard, 1957). This framework suggests that selective attention can induce systematic changes in the structure of psychological space, altering the perceived similarities among stimuli. I consider two main models, each of which corresponds to a more general conceptualization of the data: a *cue sensitivity* approach, based on the Generalized Context Model (Nosofsky, 1986), and a *category sensitivity* approach, based on the Localized Warping Model (Goldstone, 1994). The *cue sensitivity* approach predicts listeners will exhibit enhanced perception of a given acoustic dimension through selective attention. On the other hand, the *category sensitivity* approach predicts that perceptual sensitivity can change even on the same dimension based on the location of the category boundary, with poor sensitivity within a category and better sensitivity in phonetic regions that are category-ambiguous (i.e. between categories).

In the second analysis, I investigate the second research question, exploring how language mode influences the varying levels of perception among bilingual listeners. To achieve this, I compare the cue-weighting and phoneme category membership across F0 of Korean-English bilingual listeners through labeling, goodness-rating, and discrimination data, analyzed in both their native (L1) and second (L2) language modes. This examination incorporates Grosjean's Language Mode framework, which suggests that bilinguals can selectively activate one language more than the other based on the context, directing attention to language-appropriate cues. Within this framework, it's conceivable that Korean-English bilinguals may alter their cue-weighting strategies when transitioning between different language settings, reducing reliance on an irrelevant cue while intensifying reliance on a pertinent cue for the target language. If perceptual sensitivity is influenced by selective attention, then changes in cue-weighting would likely impact their perceptual acuity. However, if perceptual sensitivity is solely shaped by prolonged language experience, then changing the language setting might not lead to alterations in the perceptual sensitivity of bilinguals, irrespective of any shifts in cue-weighting strategies.

In the third chapter, I address the third research question, which examines how second language (L2) experience impacts individual listeners' language-specific listening strategies. Specifically, I concentrate on examining the potential relationships between three facets of L2 experience—namely, age of acquisition, proficiency, and input—and the manner in which listeners utilize F0 and VOT cues to categorize "non-optimal" stimuli. These noncanonical stimuli conditions have previously shown considerable individual variance in studies (Schertz et al., 2015). To accomplish this, I perform a correlation analysis between three variables obtained from the Language History Questionnaire (LHQ) data—age of acquisition, proficiency, and Full-Time Equivalent (FTE) (Flege & Bohn, 2021)—and four variables linked to listening patterns (cue-reliance in L1/L2, F0 weight in L1/L2, difference in cue-reliance between L1-L2, difference in F0 weight in L1-L2), derived from the categorization and goodness-rating performance for noncanonical stimuli. This analysis considers three frameworks; the PAM-L2 framework (Best, 1995) which predicts listeners' ability to discriminate sounds in relation to the development of second language (L2) categories; the SLM-r framework (Flege & Bohn, 2021) which posits the development of L1 categories as an important factor for successful development of L2 categories; and the ADAPPT framework (de Leeuw & Chang, 2023) which stipulates that bilinguals' L2 experience affects L1.

2.2 Participants

All participants were recruited online. English monolinguals were recruited through Prolific, an online research platform for recruiting and paying participants. They were compensated $11/hour. Korean-English bilinguals were recruited through Facebook community pages. Using the same $11/hour compensation criterion, each bilingual participant was paid $23 in Amazon gift cards.

A total of 48 monolingual English speakers and 48 Korean-English bilingual speakers participated in the study. All Korean speakers were residing in the US at the time of the study. Given the focus of language mode, it was critical to ensure that all bilingual participants were immersed in the same language environment; in this case, L2 English. All native English speakers were functionally monolingual in that none of them reported a basic level of proficiency in a second language.

Prior to starting the listening experiments, participants were administered a comprehensive language history questionnaire (LHQ). This questionnaire was designed to obtain information such as native language, age of first exposure to the L2, daily usage of language %, the length of residence in L1 and L2 speaking countries, and self-ratings of proficiency in four dimensions of both the learners' L1 and L2. In addition, the questionnaire included questions to assess any significant musical training that may impact pitch perception. Those indicating significant musical experience were excluded from the study.

Individuals who did not complete the survey were excluded from the analysis of the LHQ information. Thirty-three bilinguals and 30 monolinguals completed the language history questionnaire, which is summarized below. One bilingual participant's data (S101) were eliminated due to seemingly misunderstood (i.e. internally inconsistent) responses. Consequently, the LHQ data of 32 bilingual and 30 monolingual participants were included in the following analysis (see Table 2).

Table 2 Summary of Language History Questionnaire

	Korean-English Bilinguals N = 32		English Monolinguals N = 30	
	M	SD	M	SD
Age (years)	30.2	4.3	31.5	4.5
Eng AoA (years)	9.9	3.6	0.1	0.3
Age became fluent in Eng	20.9	6.7	2.8	3.1
Daily Eng exposure (%)	65.4	19.7	98.7	3.8
Daily Kor exposure (%)	34.0	19.9	0.7	2.9
Language use				
Eng w family (%)	12.6	24.5	100	0
Kor w family (%)	87.5	24.6	0.03	0.2
Eng w friends (%)	55.1	30.2	100	0.2
Kor w friends (%)	44.5	30.3	0	0
Eng w coworkers (%)	92.9	16.7	100	0.2
Kor w coworkers (%)	6.8	16.3	0	0
Regularly code switch	Yes 44%, No 56%		Yes 0%, No 100%	
Self-Reported Proficiency				
Eng speaking (1-5)	3.8	0.9	5.0	0
Eng listening (1-5)	4.1	0.8	5.0	0
Eng writing (1-5)	4.3	0.7	5.0	0
Eng reading (1-5)	3.3	1.2	5.0	0.2
Eng accent (1-5)	3.8	0.8	5.0	0.2
Overall Eng proficiency (sum)	19.2	3.7	25.0	0.3
Fluent in English	Yes 67%, No 10%, Partially 30%		Yes 100%	
Fluent in Korean	Yes 97%, Partially 3%		No 100%	
Language Exposure				
Daily activity in Eng (hr)	8.3	4.5	9.9	3.7
Daily activity in Kor (hr)	3.4	3.2	-	-
Daily speaking in Eng (hr)	7.1	4.6	7.0	4.2
Daily speaking in Kor (hr)	3.1	1.7	-	-
Years in Eng-Speaking Country	7.8	5.1	31.0	5.6
Years in Kor-Speaking Country	22.3	6.5	-	-
FTE	540.7	423.7	3053.4	560.8

Daily activity: watching/listening/reading contents in TV, radio, internet, books
Daily speaking: with family/friends/coworkers/others
Full-time Equivalent (FTE) Eng input: yrs of residence x daily Eng exposure %

Based on the previous literature, three measures were chosen as critical LHQ variables for examining participants' listening patterns: Age of acquisition, self-reported proficiency, and Full-

Time Equivalent (FTE) (Flege & Bohn, 2021). Korean-English bilinguals acquired their Korean as L1 and started learning English at around age 9.9 (SD = 3.6), ranging from 1 to 18. The English monolingual group acquired English at around age 0.1 (SD = 0.3) ranging from 0 to 1. Self-reported proficiency was calculated based on five scale subsections (speaking, listening, writing, reading and accentedness) which sum to 25. The average English proficiency of Korean-English bilinguals was 19.2 (SD = 3.7) ranging between 11 and 25 and the average proficiency of English monolinguals was 24.93 (SD = 0.25) ranging between 24 and 25. Full-Time Equivalent (FTE) English input was calculated by multiplying years of residence in an English-speaking country by daily English usage %. For example, an 18 year-old, monolingual English speaker who has always used English 100% of the time would therefore have an FTE of 1800 (18 years x 100%), while an 18 year-old L2 English speaker who moved to the U.S. at age 12 and estimates 70% daily English usage since then would have an FTE of 420 (6 years x 70%). The FTE English input for bilinguals ranged from 40 to 1485 with an average of 540.7 (SD = 423.7) and monolingual participants' FTE ranged from 1300 to 3800 with an average of 3053.4 (SD = 560.8).

The Korean-English participants were classified as late learners of English, having started learning the language after the age of seven, a criterion established for late learners in the phonological domain by past research (Asher & García, 1969; Birdsong, 2018; Flege & Fletcher, 1992). The majority of bilingual participants grew up in Korea and later moved to the US for college or graduate studies. They reported becoming fluent in English at an average age of 20.9 years (SD = 6.7), significantly later than the English participants, who reported fluency at an average age of 2.8 years (SD = 3.1). This means that the participants were sequential bilinguals; they had fully acquired proficiency in Korean prior to immersion in an English-dominant environment. Generally, Korean-English speakers currently used their L2 more frequently than their L1, employing their second language (L2) in a wider variety of social contexts including workplaces. However, they maintained their first language (L1) in specific situations, such as during family conversations. Approximately 44% of Korean-English bilinguals reported regularly code-switching during conversations, especially with certain friends, where they occasionally switched between English and Korean.

2.3 Stimuli

To capture natural productions of the syllable [pa], the author, a female Korean-English bilingual, recorded a list of Korean and English minimal pair words (listed in Appendix A) beginning with bilabial stop consonants. The recordings were made in a sound-attenuated booth using a Tascam recorder set at a sampling rate of 48,000 Hz. From these recordings, the word pair "bot-pot" was identified as the optimal candidate for creating CV stimuli, primarily due to their distinct coda release qualities, making them suitable for splicing CV. Subsequently, the VOT section of "pot" was spliced onto the vocalic segment of "bot" using Praat software, generating a CV syllable /pa/ consisting of 45 ms VOT and a 200 ms vowel duration.

The synthesis parameters of the syllable underwent iterative adjustments through consultation with four native Korean speakers and five native English speakers, ensuring the sound could be perceived as both a Korean lenis and an English voiceless stop. These adjustments are discussed in the following section. The syllable's overall pitch contour displayed a descending slope, with a 13 Hz difference between the maximum and minimum F0.

For F0 modifications, a Praat script was employed to linearly shift the overall pitch contour of the input audio file, thereby maintaining a consistent pitch contour across all stimuli. To create stimuli that aligned with prior research, three VOT ranges were selected, corresponding to both within-category and category-ambiguous VOT regions, resulting in a comprehensive set of stimuli spanning a two-dimensional acoustic space: 16 values of F0 ranging from 200 to 260 Hz with a 4 Hz step size and three VOT values set at 1, 25, and 45 ms conditions (as shown in Figure 4). This process resulted in the generation of 48 distinct /pa/ syllables featuring varied combinations of F0 and VOT.

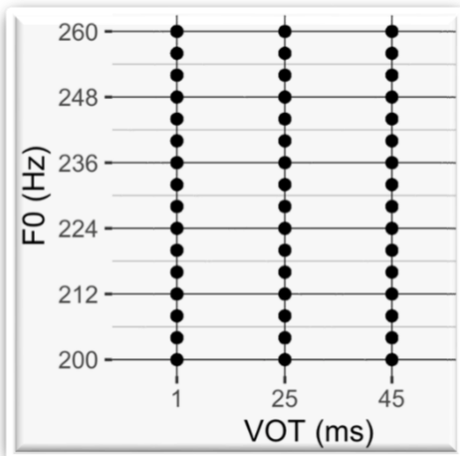

Figure 4 Stimuli Grid

2.4 Experiment Method

The experiment was conducted online and divided into two distinct testing sessions. Accepted participants engaged in the English and Korean sessions on separate days up to a week. Both oral and written instructions were provided in the respective target languages on each session day; oral instructions were pre-recorded by a native speaker of English for the English session and by a native speaker of Korean for the Korean session. The English session always preceded the

Korean session to familiarize English monolingual participants with the tasks. This ensured they could perform the same tasks without English instructions during the Korean session.

Before initiating each session, participants completed a brief headphone check test. Following the headphone test, participants performed three listening tasks. These tasks encompassed an AX discrimination task, where participants listened to pairs of syllables and discerned whether the initial consonants were the same or different; a goodness rating task, where participants listened to a syllable and rated the quality of the initial consonant sound on a 1 to 7 scale; and a labeling task, where participants listened to a syllable and categorized the initial consonant as either voiced or voiceless. Additionally, participants completed the Elevator task, a test for auditory attention (Robertson et al., 1994), in the first session only. Participants also filled out a comprehensive language history questionnaire. Participants were allowed breaks between tasks. The initial session was conducted in English, followed by the second session in Korean. Each session lasted approximately one and a half hours.

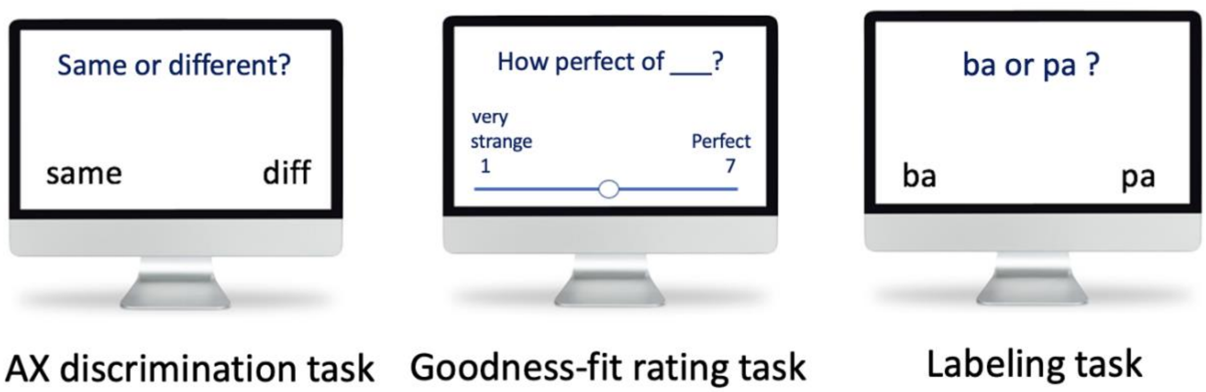

Figure 5 Illustration of Task Progression in English Session

2.4.1 Labeling Task

In the labeling task, participants were asked to categorize a single sound as either ‘pa’ or ‘ba’ (‘바’ or ‘파’ in the Korean session) by pressing keys using their keyboard. Detailed oral and written instructions were provided in the respective target languages. There were six practice trials. Across the task, 16 distinct F0 stimuli, ranging from 200 to 260 with a 4 Hz step size, were presented in three separate VOT conditions (1, 25, and 45 ms), each repeated six times. Consequently, this task encompassed a total of 288 tokens (16 x 3 x 6 = 288) per session and 576 tokens per participant over both sessions. Any labeling responses that took longer than 3000 ms were marked as “overtime” and were subsequently excluded from the analysis. Overtime responses (n=308) accounted for less than one percent of the total labeling data. Specific information regarding the data cleaning process for each analysis is presented at the beginning of the respective analysis chapters. On average, the task lasted approximately 7 minutes.

To ensure consistency in language mode effects, participants received updates on their task progress in the target language through written and audio prompts, such as "You are now 20% through this block," provided at 10% intervals. Additionally, to minimize fatigue due to repetition, participants engaged in a brief visual search task halfway through, approximately 3.5 minutes in, where they had to identify an animal image among others.

Four dependent measures were derived from the labeling data:

Voiceless Response Ratio: This measure involved the examination of raw binary labeling responses to study the effects of F0 and VOT conditions.

F0 Boundary Location and Slope: The boundary location was determined by solving the formula –(b0/b1), where b0 is the intercept and b1 is the coefficient of the predictor F0 (Xu et al., 2006). The slope of the boundary was approximated using the coefficient.

Reliance Score: The reliance scores were calculated based on the VOT 1 ms and 45 ms conditions, by subtracting the proportion of voiceless responses for short VOT-high F0 stimuli (quadrant II; see Figure 6) from long VOT-low F0 stimuli (quadrant I), adopting a similar approach from the previous literature (Schertz et al., 2016). This resulted in values ranging from -1 to 1, where -1 indicated a higher reliance on F0, 1 indicated a greater reliance on VOT, and 0 indicated an equal reliance on both cues. The rationale behind this approach is that in the absence of a canonical relationship between cues, noncanonical stimuli provide information about which of the dual-primary cues listeners weigh more heavily for categorization decisions. Therefore, the reliance score derived from the responses for the noncanonical stimuli is informative for individual listeners' relative cue reliance for each cue. See Schertz et al. (2016) for more detailed explanations.

The analysis of labeling data serves several objectives. Firstly, it aims to establish the differential utilization of F0 in categorization between Korean-English bilinguals and English monolinguals: it is predicted that Korean-English bilinguals will exhibit more pronounced F0 effects in raw categorization and a greater inclination towards F0 in their reliance score. Secondly, the analysis seeks to explore whether Korean-English bilinguals alter their categorization patters across different language modes. If bilinguals adjust their reliance on each cue in a language-congruent way, potential outcomes in their L2 English mode relative to L1 Korea mode might include diminished F0 effects and/or heightened VOT effects in categorization responses, a shift in reliance score favoring more VOT, and a less steep slope in the F0 boundary.

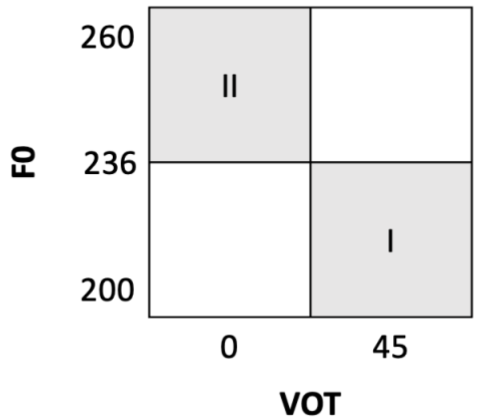

Figure 6 Schematic of Noncanonical Stimuli

Long VOT/low F0 and short VOT/high F0 were used for obtaining reliance score.

2.4.2 Goodness Rating Task

In the goodness rating task, participants listened to individual sounds and categorized them as either ‘pa’ or ‘ba’ (‘파’ or ‘바’ in Korean). Following this, they rated how well the sound exemplified the respective category by adjusting a scale from 1 to 7 using their mouse. Six practice trials were provided. To streamline the task duration, larger step-size stimuli (12 Hz) were employed, resulting in 54 trials per participant in each session. These consisted of 6 different F0 values per VOT condition, repeated 3 times (6 x 3 x 3 = 54). Lasting about 4 minutes, the test was conducted without a break. Similar to the labeling task, progress updates were communicated in the target language at every 20 % interval to ensure participants remained in the target language mode.

Two primary dependent measures were derived from the goodness rating data: the goodness rating itself and the categorization criterion. Sounds were considered ‘categorized’ if labeled as belonging to a given category over 70% (adapted from Antoniou et al., 2012). This means that for a participant's labeling response to be considered "categorized", they had to provide the same response in at least five out of six trials in the goodness rating task. This data served as supplementary information to the labeling data, contributing to the understanding of participants' category structure.

2.4.3 Discrimination Task

In the AX discrimination task, one-step (i.e. 4 Hz difference) and three-step (12 Hz difference) pairs of stimuli from the VOT 45 ms continuum were utilized to evaluate participants' perceptual processing. This task design was chosen based on prior research that explored perceptual sensitivity concerning phoneme boundaries (Francis & Ciocca, 2003; W. Heeren & Schouten, 2008). Participants were presented with pairs of stimuli from the VOT 45 ms continuum, with a 300 ms interstimulus interval (ISI), and were asked to discern "*whether the sounds were exactly the same or not*" by using the corresponding keys ("same", "different") on the keyboard.

To explore the impact of the phoneme category boundary on listeners' perception, specifically predicted by the Localized Warping Model, only stimuli from the VOT 45 condition were used with F0 being varied. The task began with six practice trials including two same pairs and four different pairs with a 20 Hz difference, designed to orient listeners' attention to differences in pitch. The actual test included a block of 44 pairs, comprising 16 same pairs, 15 different pairs with a 4 Hz difference, and 13 different pairs with a 12 Hz difference. Each pair was presented once in each order (AB, BA) to control for ordering effects. The block was repeated six times, resulting in 528 data points per participant in each session ((16 + 15 + 13) x 2 x 6 = 528) and lasting approximately 40 minutes. Additionally, the test included two breaks at regular intervals involving a visual search task. Progress updates were announced in the target language at every 20% interval to ensure the continued activation of the target language mode.

From the AX discrimination task, two key dependent measures were derived employing signal-detection-theoretic (SDT) analysis: d’ (perceptual sensitivity) and C (response bias) (see Table 3). In the context of SDT, correct responses on different pair trials are considered hits, while incorrect responses on same pair trials are termed false alarms. The hit rate thus corresponds to the probability of responding “different” on different pair trials, and the false-alarm rate corresponds to the probability of responding “different” on the same pair trials. The d’ represents perceptual sensitivity by measuring the distance between signal (hit rate) and noise (false-alarm rate) means in standard deviation units, where a value of 0 indicates an inability to discern different from same pairs, while larger values signify greater ability to distinguish these pairs. Meanwhile, the C bias measure denotes the distance between the decision criterion and the neutral point (see Stanislaw & Todorov, 1999, for further discussion). Negative values of C indicate a bias toward responding "yes" (i.e., "they are different"), whereas positive values signify a bias toward the "no" response (i.e., "they are the same"). Analysis of d’ and C values based on discrimination performance aimed to comprehend whether Korean-English bilinguals demonstrated localized warping around the presence of their category boundary. If our perceptual sensitivity is mediated by the cue sensitivity approach (i.e. the Generalized Context Model), it is anticipated that bilinguals’ d’ would generally surpass that of monolinguals. Conversely, if our perceptual sensitivity is in line with the category sensitivity approach (i.e. the Localized Warping Model), bilinguals’ d’ is predicted to be higher at the across-category region but worse at the within-category region. Given that the first prediction revolves around group differences while the second prediction is made within group, the two hypotheses are not mutually exclusive. For example, support for both cue sensitivity and category sensitivity would be evident if bilinguals display higher d' than the monolingual group overall and also demonstrate higher d' at the across-category region but lower d' at the within category region.

Table 3 Signal Detection Theory Measures

	Response: different	Response: same
Stimuli: different	HIT	MISS
Stimuli: same	FALSE ALARM	CORRECT REJECTION
$d' = z(H) - z(F)$		
$C = -(z(H) + z(F)) / 2$		

2.4.4 Catch Trials

To ensure participants' engagement with the task, "catch trials" were introduced at approximately the 30% and 80% points of the discrimination, goodness judgment, and labeling tasks. During these catch trials, an audio instruction in the appropriate language ("Press number 5 on your keyboard") was given without any corresponding written instruction displayed on the screen. If an individual provided an incorrect or delayed response (over 5000 ms) to the second catch trial, it indicated a lack of attention to the task, leading to the removal of their data for the corresponding task for that session from the analysis. Further details on data removal are provided within the respective analyses.

3.0 Analysis 1: Korean vs. English Speakers' Pitch Perception in L1

In the present chapter, I report the results of the listening experiment for Korean-English bilinguals and English monolinguals and in their respective L1 condition. One of the primary goals of this analysis is to determine whether contrastive use of F0 in native language affects segmental-level perception, and if so, how such effects arise. As F0 is a contrastive cue in Korean but not in English, a significant group difference is predicted in labeling performances. To examine listeners' segmental level perception, d' was used to measure listeners' acoustic sensitivity based on the segmental discrimination performances. Three contrasting hypotheses were made based on different perceptual learning. The first hypothesis predicts that contrastive use of F0 affects listeners' perceptual sensitivity to pitch difference following the GCM of perceptual learning. In this case, Korean listeners' d' is predicted to be higher over the entire span of the F0 continuum compared to the control group (English listeners). The second hypothesis predicts that contrastive use of F0 affects listeners' acoustic sensitivity to pitch difference under the localized warping model. In such a case, Korean listeners' d' is predicted to show localized changes near category boundaries compared to the control group. Such localized changes include poorer d' to F0 differences within category (acquired similarity) and/or better d' across categories. Note that the first and the second hypotheses are not mutually exclusive in that it is possible to observe generally greater d' of the Korean group than the English group, with Korean group exhibiting relatively higher d' at across-category range than within category. Lastly, the third hypothesis predicts that language experience does not differentially affect listeners' perception of pitch. In this case, no group difference is predicted in d' between the Korean and English groups.

For the labeling analysis, data from 32 bilinguals and 31 monolinguals were compiled, resulting in an initial dataset of 36,288 tokens. Tokens with reaction times of 3000 ms or longer were removed (n = 308). Additionally, individual participants' second catch trials were inspected for overtime or incorrect responses. One bilingual participant (subject 126) exceeded 5000 ms in responding to the second catch trial during their first session, leading to the removal of their labeling data from the first session (n = 285). After the exclusion, the labeling dataset comprised 35,695 tokens.

For the present analysis, labeling data from 31 Korean listeners in the Korean session and 31 English listeners in the English session was used, totaling 18,061 tokens. The analysis only focuses on each group's performance in their respective L1, although both groups completed both sessions in both their L1 and L2 (or foreign language for English monolinguals) contexts. Here I report categorization, F0 boundary and slope, cue-weighting strategy, goodness rating and acoustic sensitivity level (d' and response bias). In the report of these data, analyses for each of these measures are described separately. For each measure, analyses of variance were performed with participants as the random factor.

For category boundary, the data were analyzed to examine the effects of F0 and VOT. For cue-weighting strategy, listeners' reliance score was compared to identify different types of cue-weighting patterns within each group. For the goodness-rating task, the rating of voiced and voiceless were compared between groups. For acoustic sensitivity, d' and response bias were compared for each group.

3.1 Results

3.1.1 Analysis of L1 Labeling Performance

3.1.1.1 L1 Categorization Response in L1

Logistic regression using the glmer function in the *lme4* package in R was used to analyze monolingual and bilingual group's perceptual patterns in their respective L1 as a function of F0 and VOT. Fitting a separate model for each group facilitates the interpretation of complex interaction terms; below, a comparison of the two groups is also included. Voiced and 25 ms VOT were set as the reference levels, such that the statistical results show the extent to which a change in each acoustic dimension elicits more Voiceless responses. The F0 variable was standardized prior to analysis such that the resulting beta-coefficients show the change in log odds of the relevant response given a single standard-deviation increase in F0 variable.

Table 4 Categorization Analysis for Each Group

English monolingual and Korean-English bilinguals' perception of voicing categories: beta-coefficients, z-scores, and p-values from the binomial logistic mixed effects regression model of perception of the voicing stop contrast in their respective L1. The reference level is VOT 25 ms.

Effect	Voiced vs. Voiceless					
	English Monolinguals			Korean bilinguals		
	β	z	*p*<	β	z	*p*<
Intercept	1.81	6.29	0.00	-0.27	-0.96	0.34
f0	**0.43**	**6.44**	**0.00**	**1.46**	**10.15**	**0.00**
VOT1	**-4.28**	**-9.89**	**0.00**	**-2.49**	**-6.41**	**0.00**
VOT45	**1.87**	**6.17**	**0.00**	**1.74**	**9.43**	**0.00**
F0:VOT1	**-0.21**	**-2.44**	**0.01**	**-0.46**	**-4.74**	**0.00**
F0:VOT45	-0.12	-1.11	0.27	0.02	0.18	0.85

The statistical results in Table 4 show that all three variables (f0, VOT1 and VOT45) significantly influence both English and Korean listeners' response patterns in their L1. The significant effects of f0 on 25ms VOT for English groups are in line with the previous literature that showed English monolinguals use F0 as a secondary cue when the primary cue VOT is ambiguous (Schertz et al., 2015). While both English and Korean groups showed significant effects of f0, the effect size is numerically greater for the Korean group than English group (β = 1.46 vs. 0.43, respectively). On the other hand, while both English and Korean groups showed significant effects of VOT1, the effect size is numerically greater for the English group than Korean group (β = -4.28 vs. -2.49).

Table 5 shows an additional analysis explicitly comparing the two groups to one another, using a sum-coded group variable. The central focus of this analysis is to discern how each group utilizes acoustic cues (F0, VOT1, VOT45) differently for categorization. Therefore, I will focus on the group variable and any interactions associated with it. A significant main effect of group was found at the VOT 25 reference level. Additionally, significant two-way interactions were observed between the group variable and F0, as well as between the group variable and VOT 1.

Figure 7 allows for comparison of English and Korean speaker groups' responses with the response curves trend line for 1, 25, and 45 ms values of VOT across the F0 range. The relatively flat trend lines of English monolinguals in 1 and 45 ms conditions show that English monolinguals' categorization of voicing contrast was primarily dictated by VOT values. The trend lines of Korean bilinguals, on the other hand, show two noticeably different patterns. First, their categorizations of voicing contrast were affected by f0 in all three VOT conditions. Second, the effects of F0 in their responses varied as a function of VOT with the F0 effect being the smallest at 1 ms VOT compared to 25 and 45 ms.

Table 5 Summary of Categorization Analysis

English monolingual and Korean-English bilinguals' perception of voicing categories: beta-coefficients, z-scores, and p-values from the binomial logistic mixed effects regression model of perception of the voicing stop contrast in their respective L1. The reference level is VOT 25 ms.

Effect	Voiced vs. Voiceless		
	β	z	*p*<
Intercept	0.77	3.80	0.00
f0	**0.95**	**12.06**	**0.00**
VOT1	**-3.40**	**-11.63**	**0.00**
VOT45	**1.80**	**10.76**	**0.00**
group	**2.07**	**5.13**	**0.00**
F0:VOT1	**-0.35**	**-5.34**	**0.00**
F0:VOT45	-0.07	-1.02	0.31
F0:group	**-1.00**	**-6.38**	**0.00**
VOT1: group	**-1.82**	**-3.13**	**0.00**
VOT45: group	0.11	0.33	0.74
F0:VOT1:group	0.21	1.62	0.11
F0:VOT45:group	-0.16	-1.16	0.24

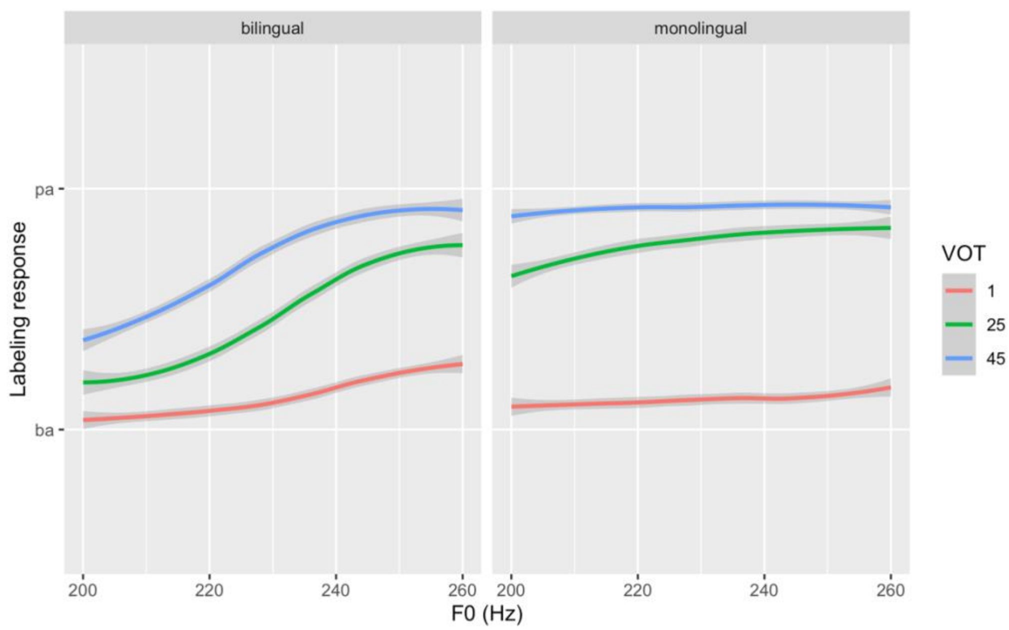

Figure 7 Labeling Response in L1 in Trend Line

3.1.1.2 Reliance Score in L1

Reliance scores were obtained based on listeners' labeling performance in the noncanonical condition (long VOT with low F0 or short VOT with high F0). Based on the previous literature (Schertz et al., 2016), I expected Korean listeners to show three types of patterns with some clustering near 1 (relying exclusively on VOT), some clustering near -1 (relying exclusively on f0), and some clustering around 0 (equal reliance on VOT and f0). On the other hand, I expected most English listeners to show clustering near 1 with most of them relying exclusively on VOT.

Reliance scores for English and Korean listeners are shown in Figure 8. In accordance with Schertz et al.'s methodology (2016), reliance score ranges were categorized into three groups based on the observed pattern within the monolingual group. Given that a substantial majority of listeners in the English monolingual group exhibited reliance scores above 0.5, the division was set at 0.5 and -0.5 to delineate distinct reliance score patterns. However, analysis of the bilingual group revealed a more continuous distribution of reliance scores, lacking discrete groupings. Consequently, the division into three groups was only used for visualization, but not for statistical analysis due to the continuous nature of reliance scores within the bilingual cohort.

The English listeners predominantly clustered together, with 27 of them relying primarily on VOT. Four listeners clustered around 0, indicating a more balanced reliance on both VOT and F0, while one listener, with a score closer to -1, exhibited a higher reliance on F0. The Korean listeners showed three categories, with one group showing a greater reliance on VOT ($n = 12$), one group showing a greater reliance on f0 ($n = 2$), and the rest of the listeners ($n = 18$) showing a more equal reliance on both. The small number of F0-relying listeners was unexpected given that the previous study showed more or less even distribution of three categories (Schertz et al., 2016). This could be due to the smaller range of stimuli used to calculate reliance scores in the current study, as the VOT in Schertz et al. (2016) study used VOT stimuli ranged between -20 to 50ms while the present study used VOT range ranged between 0 to 25ms. Another possibility is that the Korean listeners who participated in the current study were immersed in an English-speaking environment while the Korean listeners from Schertz et al. (2016) were recruited in Korea. Regardless, the current study still provides comparable patterning of reliance scores for the Korean vs. English speaker groups, with the English group showing more homogenous patterning in their cue-weighting strategy while the Korean speakers showed more variable patterns in cue-weighting strategy, with one group relying on VOT and the other group relying on both VOT and F0. While the current study showed more dominant patterns of VOT relying group than previous studies, the current findings are still in line with prominent individual variance found in Korean listeners' cue-weighting in the context of noncanonical stimuli.

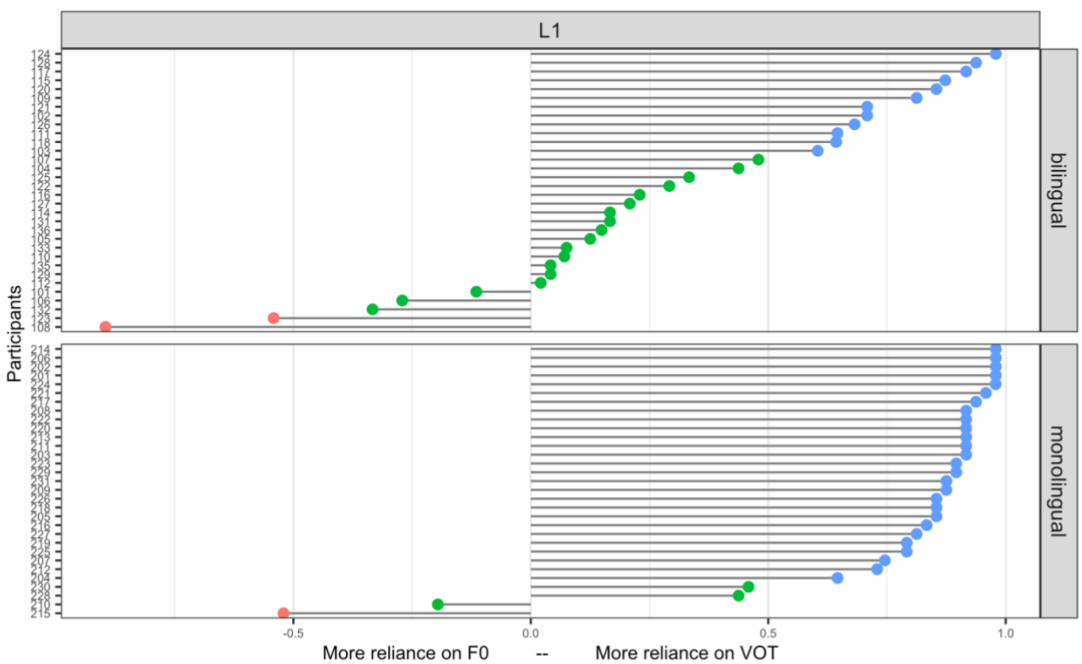

Figure 8 Individual Listeners' Reliance Score in L1

3.1.1.3 F0 Boundary Location and Slope in L1

A comprehensive dataset was compiled, including boundary locations and slopes for three distinct VOT conditions across two language sessions, initially amounting to 189 datasets. The category boundary location between the pair of phonemes for each VOT continuum (1, 25, and 45 ms) was calculated for each participant in each language mode condition. The F0 boundary location was determined exclusively for bilingual listeners since English monolinguals were not anticipated to utilize F0 for categorization. The boundary location value was calculated by solving the formula –(b0/b1) based on a logistic regression model where b0 is the intercept and b1 is the coefficient of the predictor F0 value (McDonald & Kaushanskaya, 2023; Xu et al., 2006). Such logistic regression models were constructed for each subject using the maximum available trials (n = 576) from labeling responses from 16-step F0 continuum in 3 different VOT conditions with 6 repetitions in two sessions. In addition, the slope of the categorical boundary was obtained by taking the estimated regression coefficient (Xu et al., 2006).

Post initial screening, 20 data points were excluded due to boundary location values falling outside the 0-400 Hz range, resulting in 169 datasets for analysis. The decision for the 0-400 Hz F0 boundary cutoff was informed by observed dynamic F0 adjustments in Korean listeners' categorization across varied VOT conditions. This choice ensured analysis feasibility; implementing a stricter cutoff (200 ~ 260 Hz) would have eliminated 84 data points, compromising over 40% of the dataset and impeding further analysis. The increased variability in boundary values in this study compared to prior research may be attributed to the more fluid F0 adjustments made by Korean listeners, diverging from the established VOT boundary values in previous literature.

Table 6 demonstrates a complementary relationship in bilinguals' F0 boundary location, where higher VOT corresponded to lower F0 boundary locations (and vice versa). Regarding the slope sharpness, the F0 boundary slope was numerically lowest in the short (1 ms) VOT condition and increased in ambiguous (25 ms) or long (45 ms) VOT conditions.

Table 6 Korean Listeners' F0 Boundary Location and Slope

F0 boundary location in L1 Korean

VOT	F0 boundary	slope	Number of data
1	281 (sd = 33)	0.06 (sd = 0.04)	24
25	236 (sd = 30)	0.10 (sd = 0.07)	31
45	211 (sd = 25)	0.10 (sd = 0.05)	30

The bilinguals' F0 boundary data in Table 6 and Figure 7 indicate a shifting pattern in voicing categories across the F0 continuum, contingent upon the VOT condition. For instance, at VOT 25 ms, stimuli within the 200 to 220 Hz range are perceived as "voiced". However, at VOT 45 ms, the same region exhibits category ambiguity, suggesting a dynamic boundary shift influenced by the interplay between F0 and VOT cues.

3.1.2 Analysis of Goodness-rating

3.1.2.1 Goodness-rating of Voiced Category in L1

Analyzing the goodness rating data alongside categorization data enhances our insight into how F0 and VOT influence listeners' categorization. Table 5 illustrates how F0 stimuli at 200, 224, 236, and 260 Hz are evaluated across various VOT conditions in the English and Korean groups. Clear disparities emerge between the two groups in how their goodness judgments are influenced by both VOT and F0.

For English monolinguals, VOT values predominantly influenced categorization. At VOT 1 ms, all stimuli were voiced, while at VOT 25 and 45 ms, they were classified as voiceless. Interestingly, stimuli at both VOT 25 and 45 ms were categorized as voiceless, but with lower percentages and goodness ratings for VOT 25 ms, indicating a weaker fit for voiceless sounds. Notably, there was a slight increasing trend in goodness judgment for voiceless sounds with higher F0, suggesting some weak, yet present, F0 effects at VOT 25 ms for English listeners.

For the Korean bilingual listeners, F0 seemed to influence their goodness judgment and categorization, unlike the English monolingual listeners. The categorization of stimuli in different VOT conditions varied based on F0 values. An exception to this is seen in VOT 1 ms, where the Korean listeners consistently categorized and rated the stimuli as voiced regardless of F0 values. This suggests that when competing VOT and F0 information is given, if the VOT is sufficiently low (i.e. 1 ms), Korean listeners prioritize VOT as an important cue for the voiced category. In other words, low VOT is given priority as a cue and F0 information is ignored. However, when VOT is not sufficiently low, such as mid (25 ms) or high (45 ms), F0 information becomes important in judging voicing categories. For example, at VOT 25 ms, the F0 continuum is divided into three parts: 200 ~ 224 Hz corresponding to the voiced category, 236 Hz corresponding to the across-category range, and 260 Hz corresponding to the voiceless category. This illustrates the presence of both voiced and voiceless categories with a relatively sharp boundary between them in the VOT 25ms condition. Interestingly, this patterning changed at VOT 45 as 200 ~ 224 Hz now correspond to the across-category range and 236 ~ 260 Hz corresponds to the voiceless category. Critically, at a frequency of 236 Hz, the stimulus was judged to be between-category in the VOT 25 ms condition, but as voiceless category in the VOT 45 ms condition. This suggests a change in F0 category boundary between VOT 25 and 45 ms condition.

Table 7 Goodness Rating in L1

Percentage of labeling decisions (and goodness rating 1-7) of English and Korean stops by English monolinguals and Korean bilinguals in their respective L1.

Group	VOT	Category label							
		Voiced				Voiceless			
		F0							
		200	224	236	260	200	224	236	260
English Group	1	**95%** (5.4) n=88	**95%** (5.3) 88	**91%** (5.0) 84	**94%** (4.8) 87	5% (3.0) 5	5% (4.4) 5	9% (3.3) 8	6% (3.0) 6
	25	25% (3.0) 23	17% (2.4) 16	11% (2.5) 10	15% (2.6) 14	**75%** (3.5) 69	**83%** (3.5) 77	**89%** (3.8) 83	**85%** (3.9) 79
	45	2% (4.0) 2	2% (4.0) 2	1% (5.0) 1	2% (5.0) 2	**98%** (5.4) 91	**98%** (5.2) 90	**99%** (5.3) 90	**98%** (5.5) 90
Korean Group	1	**98%** (4.7) 90	**92%** (4.2) 81	**94%** (3.8) 87	**80%** (3.8) 72	2% (2.0) 2	8% (4.0) 7	6% (2.8) 6	20% (3.2) 18
	25	**91%** (4.6) 83	**74%** (3.7) 67	49% (2.9) 46	15% (1.6) 14	9% (2.6) 8	26% (3.3) 23	51% (4.0) 47	**85%** (4.6) 78
	45	68% (5.4) 63	32% (4.9) 29	9% (3.5) 8	NA	32% (4.3) 29	68% (4.9) 63	**91%** (5.6) 83	**100%** (6.1) 91

3.1.3 Analysis of Discrimination in L1

Data points for discrimination were derived from listeners' F0 stimuli discrimination at the VOT 45 ms condition in each language session, encompassing 160 tokens from identical pairs, 150 tokens from 4 Hz-difference pairs, and 130 tokens from 12 Hz-difference pairs, totaling a maximum of 440 discrimination responses per listener per session. Responses exceeding 3000 ms were excluded, leading to the removal of 469 tokens from bilinguals and 204 tokens from monolinguals. This resulted in 54,768 tokens available for analysis, with a maximum of 440 tokens per listener per session. All participants successfully completed the second catch task. However, one bilingual participant (subject 133) and one monolingual participant exhibited prolonged reaction times in the second catch trial in their respective second language sessions (RT > 5000 ms). While these participants' data were coded separately, they were retained in the analysis.

The discrimination data were analyzed using two metrics: d' and C. The d' values quantified listeners' accuracy in distinguishing genuinely different pairs (HIT), while considering instances where listeners incorrectly identified identical pairs as different (FALSE ALARM) (see 2.4.3, Table 3). HIT values were calculated based on various stimuli conditions, including step size, F0 values, and language mode, with a maximum of 10 discrimination responses considered for each combination. Consequently, each listener generated 28 HIT values in each language session, comprising 15 HIT values for 4 Hz-difference step size and 13 HIT values for 12 Hz-difference step size. The false alarm rate (incorrect response to identical pairs) was computed based on F0 values and language mode. This involved subtracting the correct rejection (accurate response to identical pairs) from 1. The result yielded 16 FALSE ALARM values per listener in each language session. Any HIT or FALSE ALARM values of 0 or 1 were adjusted to 0.01 and 0.99, respectively. All HIT and FALSE ALARM values were z-transformed and subtracted from each other, resulting in 28 d' values per listener in each language session. C values were calculated using the equation: - (HIT + FALSE) / 2.

3.1.3.1 d' by Groups in L1

Based on the bilinguals' judgment and labeling performance of VOT 45 ms in their L1, the F0 continuum used in discrimination task was encoded under the new variable *f0Range.* The F0 range that corresponds to bilinguals' between-category region (F0 200 ~ 224 Hz) was coded as "low" and the F0 range that corresponds to bilinguals' within-category region (F0 236 ~ 260 Hz) was coded as "high". Because the main interest was to see whether the d' of the bilingual listeners changes as a function of their categorization pattern, I only focused on the F0 range whose categorization pattern was established by the judgment task, excluding the d' values outside of those categories. This resulted in the exclusion of 8 d' values from the 28 d' values per participant per language session, leaving maximum of 20 d' values available for the analysis per participant. Out of 20 data points, 10 data points correspond to between-category region (200 ~ 224 Hz) with six of them with 4 Hz step size and 4 of them with 12 Hz step size. Another 10 data points correspond to within-category region with 6 of them with 4 Hz size d' values in across-category region (236 ~ 260 Hz). With 32 bilingual and 31 monolingual participants, a total of 1,260 data points were used for the analysis.

A three-way ANOVA was performed to compare the effect of group (English vs. Korean), step (4 vs. 12), and F0 range (low vs. high) on d' in listeners' L1 condition. The analysis revealed a main effect of step ($F(1, 1252) = [486.31]$, $p = <.001$). There was no main effect of group nor F0 range. However, a significant two-way interaction was found between group and F0 range ($F(1,1252 = [7.29]$, $p <.01$). Tukey's HSD Test for multiple comparisons were performed to better understand the interaction between the group and F0 range. It was found that the mean value of d' was significantly different between bilinguals' low and high F0 region ($p <.05$). This can be observed in Figure 5 where bilinguals' d' was generally poorer in the high F0 region than in the lower F0 region. The post-hoc test also showed a significant group difference in the mean value of d' at the high F0 region ($p <.05$), which comes from poorer d' for bilinguals than monolinguals at the high F0 range.

The significant main effect of step is an expected result, as it shows that both listener groups were able to discriminate sounds more accurately when the step size was bigger. In terms of f0 range effect, the Korean and English groups showed a clear difference in their d' in that Korean group's d' at F0 stimuli corresponding to their voiceless category significantly decreased compared to the d' at F0 range corresponding to between-category. Such F0 range effect on d' is not observed in English group's data which showed consistent d' in different F0 ranges. Consequently, the English group demonstrated better d' than the Korean group at high F0 range. Given the results, the hypothesis that language experience can shape pitch perception is supported. In particular, the hypothesis that predicts perceptual development of pitch through "acquired similarity" with localized warping is supported as Korean speakers showed poor d' with stimuli that correspond to within-category as opposed to English speakers who did not show any decrease in d'.

Table 8 Summary of group and F0 range interaction

Mean d' (SD) of each group across step.

	Low F0		High F0
English Group	0.76 (1.38)	≅	**0.87(1.24)**
	≅		>
Korean Group	**0.87 (1.21)**	>	**0.64(1.20)**

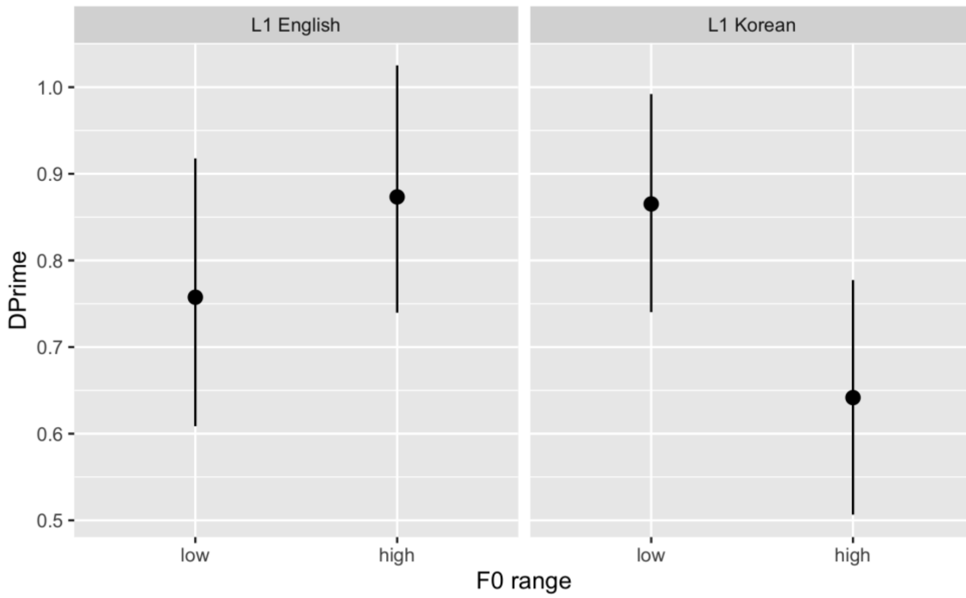

Figure 9 d' of L1 English and L1 Korean Listeners in Their Respective L1

3.1.3.2 Response Bias by Groups in L1

A two-way ANOVA analysis was performed on response bias using C (Stanislaw & Todorov, 1999) to compare the effects of group (English vs. Korean) and f0Range (low vs. high). Significant main effects were found for group ($F = 14.11$, $p<.001$) and f0Range ($F = 16.19$, $p<.001$) which further interacted with one another ($F = 28.25$, $p<.001$). Post-hoc test with Tukey multiple comparisons revealed a significant difference in Korean listeners' C between low and high F0 ranges ($p<.001$). Additionally, there was a significant group difference in response bias at high F0 range ($p<.001$). Figure 6 shows that Korean listeners showed higher response bias when listening to high F0 range, indicating that they were biased toward the "different" response.

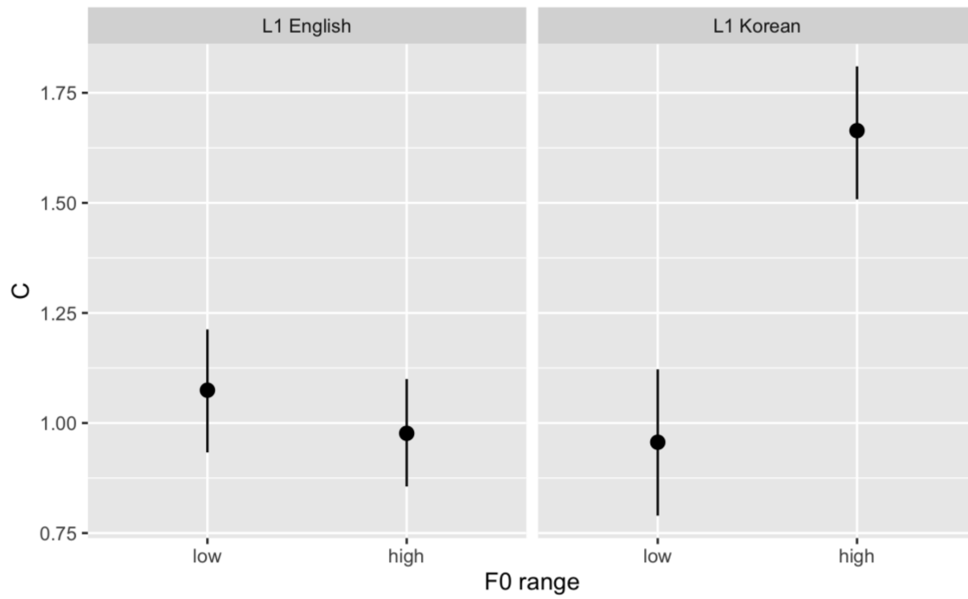

Figure 10 C of English and Korean Listeners in Their Respective L1

3.1.3.3 Post hoc analysis: correlation between L1 reliance and d'

d' and reliance score: In order to examine whether reliance score is related to listeners' d', I ran a Pearson correlation test on L1, L2 reliance score and d' at across and within-category using rcorr() function from the Hmisc package. No significant correlation was found among the variables.

d' F0 slope: In order to examine whether reduced d' is related to F0 slope at VOT 45ms, I ran pearson correlation test on L1 F0 slope at VOT 45ms and d' using rcorr() function. No significant correlation was found among the variables.

3.2 Discussion

The primary inquiry in this chapter was whether the contrastive use of pitch influences listeners' perceptual sensitivity and, if so, how these perceptual effects manifest with respect to sound categories. The above results demonstrate the distinct impact of pitch on Korean and English listeners in both categorization and discrimination tasks.

The results from the labeling and goodness-rating tasks highlight the disparity in pitch utilization between Korean and English listeners. Korean listeners exhibit a dual cue-weighting strategy involving both VOT and F0, with a dynamic shift in F0 boundary between lenis and aspirated categories depending on VOT conditions. In general, the F0 boundary was observed to decrease as VOT increases. Notably, at 1 ms VOT, Korean listeners no longer categorize stimuli as voiceless, indicating a phonological organization rule based on a minimum threshold of VOT value for the aspirated category. Interestingly, while 1 ms VOT stimuli were consistently categorized as "lenis," goodness ratings revealed the influence of F0 information, indicating a potential phonetic mapping challenge. This can be due to Korean listeners' phonetic mapping of the third phonological stop, fortis, which was not presented as part of choices in the current study. In the low VOT region, Korean lenis and fortis might share the phonetic spaces, with fortis requiring higher F0 than lenis. This nuanced interaction between cues showcases Korean listeners' adaptive use of pitch, allowing them to phonetically map their three-way stops effectively.

The labeling and goodness data provided crucial category information for F0 stimuli. In the VOT 45ms condition, the F0 continuum was divided into between-category and within-category regions, allowing an exploration of how language experience influences segmental perception. Three hypotheses were tested. First, if language experience shapes segmental perception, a significant English-Korean group difference was anticipated. If the Generalized

Context Model (GCM) applied with level pitch, higher d' across the F0 continuum for bilingual listeners was expected, without F0 range effects as perception would not be locally warped. Conversely, if the localized warping model held, Korean listeners' d' would vary between between-category and within-category, reflecting acquired distinctiveness vs. acquired similarity based on stimulus category. Lastly, if F0 perception is domain-general, independent of language experience, no difference between English and Korean groups was expected.

The results of the discrimination data show evidence for localized warping model for level pitch perception. There was a significant group difference between Korean and English listeners in the high F0 region. Given that the low F0 region corresponds to a between-category region and the high F0 region corresponds to a within-category region for Korean listeners, our data suggest that Korean listeners' may have developed acquired similarity within the "aspirated" category for relatively high F0 sounds with long VOT (45 ms). On the other hand, English listeners, even though they categorized the entire long VOT (45 ms) as a single voiceless category, they do not have F0 boundary for categories that stimulates perceptual development of acquired similarity around certain F0 area. Therefore, we can conclude that not only contrastive use of pitch can shape listeners' perceptual sensitivity, but the perceptual development also occurs in the form of acquired similarity for sounds within category. The analysis of response bias showed that Korean listeners' response bias was significantly affected by F0 range in that they tend to show higher response bias in the higher F0 range. This means Korean listeners judged the stimuli pairs to be different when they were, in fact, the same sounds. Given that d' and C are independent measures, the increased C does not contribute to the poor d'. It is possible that such increased response bias might be a strategy for Korean listeners to compensate their poor d' with higher pitch stimuli.

To conclude, this study revealed a significant influence of native language category membership on listeners' perceptual sensitivity. During the labeling task, Korean listeners exhibited selective attention to both F0 and VOT, two acoustic dimensions for distinguishing stop contrasts. English listeners, on the other hand, primarily relied on VOT. Notably, the labeling data showed a distinct division of the F0 continuum into category-ambiguous and within-category regions for Korean listeners. In contrast, English listeners exhibited consistent category membership along the F0 continuum without significant changes in category membership. The results of the perceptual sensitivity analysis highlighted that only Korean speakers were significantly affected by different F0 ranges, suggesting the impact of category membership on their perceptual sensitivity. Specifically, Korean listeners demonstrated diminished sensitivity to pitch differences when listening to sounds within the aspirated category. The d' values were markedly lower compared to those for sounds in category ambiguous regions, and significantly poorer compared to English listeners' sensitivity. These findings support the Localized Warping Model that posits that perceptual development is intertwined with the acquisition of phoneme categories. Within the framework of selective attention, listeners acquire an equivalence within categories, leading to reduced sensitivity for sounds within the same category. These observations provide valuable insights into how selective attention to acoustic dimensions shapes perceptual development and the role of the Localized Warping Model in developing phoneme categories.

4.0 Analysis 2: Perceptual Shift between L1 and L2

In this chapter, I compare the speech perception of Korean-English bilinguals in their native (L1) and second language (L2) conditions. The analysis focuses on two parts of the perceptual patterns. First, I aim to assess whether bilingual listeners alter cue-weighting in a language-specific manner by minimizing reliance on irrelevant acoustic cues. To achieve this, I compare Korean-English bilinguals' labeling data between their L1 and L2 conditions using several dependent measures including reliance scores and F0 slopes. In the Language Mode framework (e.g., Grosjean, 1999), bilingual participants are expected to activate Korean representations more in the Korean sessions and less in the English sessions. If bilingual listeners' cue-reliance is affected in a language-specific way, I predict Korean listeners to show reduced reliance on F0 (shallower F0 Slope) and/or increased reliance on VOT (numerically higher Reliance Score) in their L2 English than in their L1 Korean context, given F0 is not a primary cue in English but VOT is.

Additionally, the study investigates how altering cue-weights language-specifically affects perceptual sensitivity, which is at the segmental level of perception. The segmental level of perception is analyzed using d' in the Korean versus the English contexts. While segmental perception may be shaped by long-term phonetic learning, active attention to relevant acoustic dimensions could also play a crucial role, potentially leading to changes in sensitivity if attention shifts away from specific dimensions. Regarding discrimination, two hypotheses emerge. The robust perceptual sensitivity hypothesis posits that Korean listeners' d' will remain unaffected by language mode due to long-term perceptual development. Conversely, the attentional perceptual sensitivity hypothesis suggests that Korean listeners' d' will be influenced by language mode as they limit reliance on the F0 cue, assuming a close link between cue reliance and perceptual sensitivity.

4.1 Results

4.1.1 Analysis of Labeling Performance

For the present analysis, labeling data from 32 Korean listeners were used. Data from one participant (subject 126) in the L2 session were excluded due to poor reaction time in the second catch trial. This led to a total of 17,974 data points available for analysis.

4.1.1.1 Categorization Response in L1 vs. L2

Logistic regression using the glmer function in the *lme4* package in R was used to analyze bilingual perceptual patterns based on language mode, F0 and VOT. The dependent variable was binary response (‘voiced’ vs. ‘voiceless’), with ‘voiced’ as the reference level, such that coefficients correspond to each variable’s effect on the probability of a ‘voiceless’ response. Random factors included by-participant intercepts and slopes for F0 by participant and VOT. The model included a three-way interaction (VOT, F0 and language mode) and two-way interactions for each predictor. The language mode variable was sum coded, while VOT was treatment coded, making the reference level 25 ms VOT, and the average between the two language mode conditions. The F0 variable was standardized. The analysis focused on interactions between VOT or F0 and the language mode predictor, revealing how changes in each acoustic dimension elicited more voiceless responses.

Table 9 Categorization Analysis for Bilinguals

Korean-English bilinguals' perception of voicing categories in different language modes: beta-coefficients, z-scores, and p-values from the binomial logistic mixed effects regression model of perception of the voicing stop contrast. The reference level is VOT 25 ms with an average between two language modes.

Effect	Voiced vs. Voiceless		
	Korean bilinguals		
	β	z	*p*<
Intercept	-0.18	-0.85	0.40
F0	**1.29**	**10.99**	**0.00**
VOT1	**-2.48**	**-7.72**	**0.00**
VOT45	**1.63**	**8.08**	**0.00**
langMode	-0.11	-1.68	0.09
F0:VOT1	**-0.46**	**-5.18**	**0.00**
F0:VOT45	0.13	1.59	0.11
F0:langMode	-0.04	-0.54	0.59
VOT1:langM	0.21	1.84	0.07
VOT45:langM	**0.68**	**6.48**	**0.00**
F0:VOT1:langM	0.13	1.04	0.30
F0:VOT45:langM	-0.07	-0.59	0.55

Table 9 indicates no main effect of language mode for bilinguals at VOT 25 ms, the reference level. However, there was a significant interaction between VOT 45 ms and language mode. Post-hoc analysis, conducted using the pairs() function from the emmeans R package, revealed higher voiceless response ratios at the 45 ms VOT condition in L2 compared to L1 (Figure 11). This suggests that Korean listeners relied more on the VOT cue in L2 English, categorizing long VOT stimuli as voiceless irrespective of F0, indicating their emphasis on VOT in English mode.

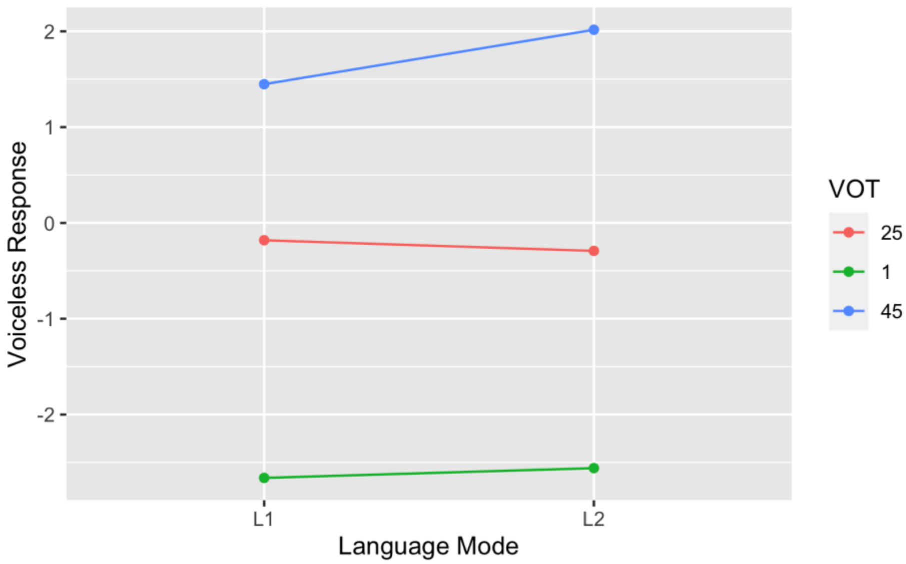

Figure 11 Visualization of Interaction between VOT 45 and Language Mode

4.1.1.2 Analysis of Reliance Score in L1 vs. L2

The reliance scores of Korean-English bilingual listeners were analyzed concerning language mode effects. The reliance score is derived from listeners' labeling responses to non-canonical stimuli (refer to Section 2.4.1). A reliance score of 1 indicates a stronger dependence on VOT, while reliance score of -1 signifies a greater dependence on F0. A reliance score closes to 0 suggests an equal reliance on both VOT and F0. Excluding one participant (subject 126) with poor reaction time in the second catch trial in session 2, the revised dataset included reliance scores from 31 bilingual participants and 31 monolingual participants. This resulted in 62 data points available for each of the Korean and English groups' analysis, respectively.

Given the simplicity of the model and there are only 61 observations per group, simple paired t-tests were conducted for each group using the t.test() function in R. The analysis revealed a significant effect of language mode in the bilingual group ($t = -0.28$, $p < .05$) but not in the monolingual group (Figure 12).

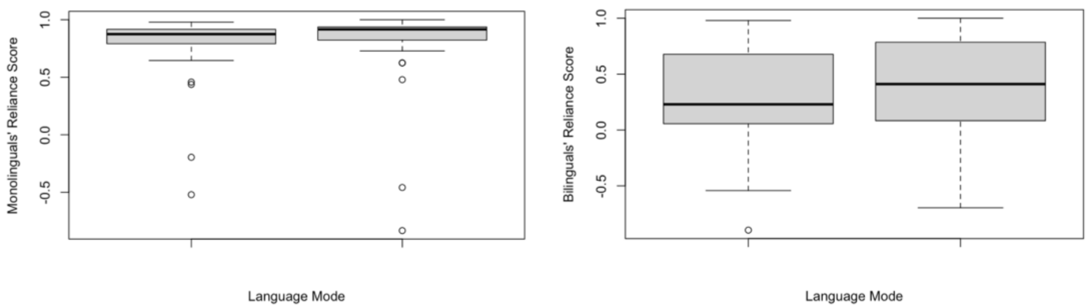

Figure 12 Monolinguals vs. Bilinguals' Reliance Score

An additional analysis was conducted to examine whether the reliance score change was affected by each listener's cue-weighting type in L1. For this analysis, two new variables were created. First, a new variable *reliance.dff* was made by subtracting reliance score in L1 from the reliance score in L2. Second, a new variable *L1type* was created with three levels (VOT, F0, or VOT+F0) based on individual bilingual listener's cue-weighting type in L1 using the cutoffs of 0.5, and -0.5 (refer to the Section 3.1.1.2). A linear model was made with reliance difference as a dependent variable and a L1cue type as a predictor. The reference level was F0+VOT cue type. The results showed a significant difference for listeners of VOT cue type ($b = -0.14$, $p < .05$) but not for F0 cue type. Figure 13 plots the average change in reliance scores for the F0+VOT versus VOT listener groups. Figure 14 shows that this difference stems from the F0+VOT group's increased reliance score in L2 English than in L1 Korean, of which reliance score 1 corresponds to exclusive reliance on VOT.

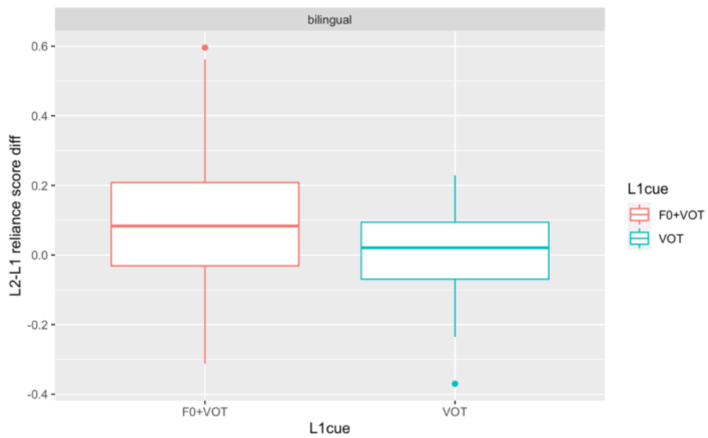

Figure 13 Bilinguals' L2-L1 Reliance Score Difference

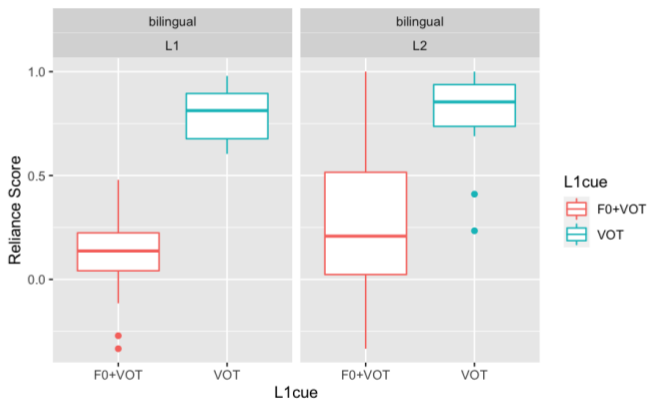

Figure 14 Bilinguals' L2-L1 Reliance Score Difference by L1 Cue-weighting Type

4.1.1.3 F0 Boundary Location and Slope in L1 vs. L2

In line with the methodology outlined in the preceding chapter, individual F0 boundary locations and slopes were determined for each VOT condition in both language sessions. Separate mixed-effects linear regression analyses were conducted to explore the impact of language mode and VOT on F0 boundary. The fixed effects comprised VOT, language mode, and their interactions. Subject was included as a random effect. Language mode was sum coded for the analysis, with a total of 169 data points.

F0 boundary location: The results indicated significant effects of VOT 1 and VOT 45 on bilinguals' F0 boundary location, with a crucial interaction observed between VOT 1 and language mode (Table 10). This was somewhat surprising, given that there was no language mode effect of the raw labeling responses related to the VOT 1 ms condition. Post-hoc analysis, utilizing the Tukey method through the emmeans() and pairs() functions from the emmeans package in R, further elucidated the interaction: bilinguals' F0 boundary location at VOT 1 ms was significantly lower in L2 compared to L1 ($p < .0001$). Conceptually, extremely high F0 boundary at VOT 1 ms in Korean indicates that listeners were in general perceived the sounds as "voiced". The fact that bilingual listeners' F0 boundary location shifted low in English mode can be interpreted as less sounds being perceived as "voiced" and more sounds being perceived as "voiceless". One way to interpret this result is that in L1 Korean mode, the listeners predominantly categorized stimuli at VOT 1 ms as "voiced," prioritizing "sufficiently low VOT" as the key cue for identifying Korean lenis sounds, irrespective of F0 differences. However, this strategy did not apply in L2 English, where they labeled VOT 1 ms stimuli as "voiced" only if both low in VOT and F0. Table 11 illustrates that in L1 Korean, bilinguals' F0 boundary was markedly influenced by VOT values, with the highest F0 boundary at VOT 1 ms and the lowest at VOT 45 ms.

F0 boundary slope: The analysis of bilinguals' F0 boundary slopes revealed a significant effect of VOT 1, indicating a steeper F0 boundary slope at VOT 25 ms compared to VOT 1 ms (Table 11). Furthermore, a significant effect of language mode was observed at VOT 25 ms, where the F0 slope was steeper in L1 than in L2 (Figure 15). Notably, there was no significant interaction between VOT 1 or 45 ms in relation to language mode.

In summary, the language mode effects on bilingual listeners' F0 boundaries were nuanced (Table 12 and Figure 15). When VOT was low (1 ms), the F0 boundary in L2 shifted slightly to lower values. Despite this shift, the raw labeling data in the VOT 1 ms condition remained largely unchanged between L1 and L2. At 25 ms VOT, there was a decrease in F0 boundary slope in L2, indicating reduced reliance on F0 in English mode, although this change didn't significantly affect raw labeling responses either. One reason for the discrepancy between boundary and raw labeling data could be due to the exclusion of some data outside of the cutoff (0 ~ 400 Hz) (refer to Section 3.1.1.3) in the boundary analysis, leading to a more robust dataset of boundary data for detecting language mode effects compared to the raw categorization data. Conversely, at high VOT (45 ms), while F0 slope and boundary remained consistent across language modes, there was a significant change in the raw labeling data. Korean listeners exhibited more voiceless responses in English mode, suggesting a greater reliance on VOT in L2 English than in L1 Korean. These findings suggest subtle shifts in cue weighting in L2, with a reduced reliance on F0 and increased reliance on VOT while maintaining the use of both cues in labeling decisions.

Table 10 F0 boundary Location Analysis

Summary of the analysis of bilinguals' F0 boundary location. The reference level is VOT 25 ms with averaged language mode.

	β	SE	*t*	*p*<
Intercept	235.28	5.75	72.23	<.001
VOT1	**23.10**	**6.85**	**3.37**	**<.001**
VOT 45	**-24.59**	**6.48**	**-3.79**	**<.001**
langMode	0.70	9.02	0.08	0.94
VOT 1: langMode	**-43.85**	**13.62**	**-3.22**	**<.001**
VOT 45: langMode	2.47	12.97	0.19	0.85

Table 11 Boundary Slope Analysis

Summary of the analysis of bilinguals' F0 boundary slope. The reference level is VOT 25 ms with averaged language mode.

	β	SE	*t*	*p*<
Intercept	0.08	0.01	10.18	<.01
VOT1	**-0.03**	**0.01**	**-3.08**	**<.01**
VOT 45	0.01	0.01	1.03	0.30
langMode	**-0.03**	**0.01**	**-2.37**	**<.05**
VOT 1: langMode	0.02	0.02	1.23	0.22
VOT 45: langMode	0.02	0.02	1.11	0.27

Table 12 Summary of Bilinguals' F0 Boundary in L1 vs. L2

F0 boundary location and slope in L1 vs. L2 estimated by mixed effects linear regression with VOT and language mode as predictors. In L1 mode, the F0 boundary between lenis and aspirated categories, while in L2 mode, voiced and voiceless categories.

LangMode	VOT 1 ms		VOT 25 ms		VOT 45 ms
			F0 Boundary (SE)		
L1	280 (8.18)	>	235 (7.30)	>	209 (7.41)
	>		=		=
L2	237 (8.04)	=	236 (7.31)	>	212 (7.64)
			F0 Slope (SE)		
L1	0.06 (0.01)	<	0.10 (0.01)	=	0.10 (0.01)
	=		>		=
L2	0.05 (0.01)	<	0.07 (0.01)	=	0.09 (0.01)

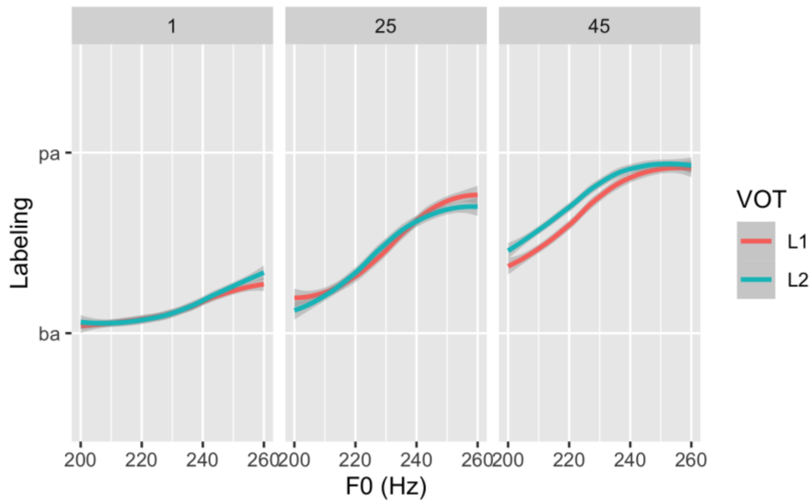

Figure 15 Labeling Responses of Bilingual Listeners in L1 vs. L2

4.1.2 Analysis of Goodness-rating in L1 vs. L2

4.1.2.1 Qualitative Assessment of Goodness-rating between L1 vs. L2

The goodness rating table (Table 13) indicates consistent judgment patterns across sessions for both groups. English monolinguals used English phoneme labels in both sessions, while Korean bilinguals used English labels in the first session and Korean labels in the second. Consequently, the analysis primarily centers on the bilingual listeners' results. Because the differences appear small, a statistical analysis was not carried out, but instead qualitative trends were considered to help with interpretation of other analyses.

Examining the table, bilinguals exhibit reduced goodness ratings for voiced sounds at 1 ms VOT as F0 increases in L1, a trend less pronounced in L2 English. The reason for this shift in goodness ratings is unclear, but it could be hypothesized that the differences in L1 and L2 relate to the activation of an additional phoneme category in Korean mode, occupying a similar acoustic space (high F0 – low VOT) as Korean lenis. Consequently, competing categories might have influenced lower goodness ratings for lenis. Moreover, the F0 effect at 25 ms VOT appears weaker in L2, in line with the reduced F0 slope observed earlier. There is no significant difference in VOT 45 across language modes. Overall, the goodness rating data align with the findings from the labeling analysis.

Table 13 Goodness Rating in L1 vs. L2

Average label proportion (and goodness rating 1-7) for English monolinguals and Korean bilinguals in their respective L1 and L2. Red highlighting corresponds to stimuli reaching the "within category" threshold value of 70%.

Group			Category label							
			voiced				voiceless			
			F0							
			200	224	236	260	200	224	236	260
English Group	1 ms	L1	**95%** (5.4) n=88	**95%** (5.3) 88	**91%** (5.0) 84	**94%** (4.8) 87	5% (3.0) 5	5% (4.4) 5	9% (3.3) 8	6% (3.0) 6
		L2	**92%** (5.2) 86	**94%** (5.3) 87	**92%** (5.1) 84	**95%** (4.9) 88	8% (5.3) 7	6% (6.5) 6	8% (5.7) 7	5% (6.8) 5
	25 ms	L1	25% (3.0) 23	17% (2.4) 16	11% (2.5) 10	15% (2.6) 14	**75%** (3.5) 69	**83%** (3.5) 77	**89%** (3.8) 83	**85%** (3.9) 79
		L2	24% (2.5) 22	16% (3.4) 15	11% (2.8) 10	10% (3.6) 9	**76%** (3.4) 71	**84%** (3.3) 77	**89%** (3.6) 82	**90%** (3.6) 84
	45 ms	L1	2% (4.0) 2	2% (4.0) 2	1% (5.0) 1	2% (5.0) 2	**98%** (5.4) 91	**98%** (5.2) 90	**99%** (5.3) 90	**98%** (5.5) 90
		L2	5% (5.0) 5	3% (4.3) 3	2% (4.0) 2	6% (4.7) 6	**95%** (5.3) 87	**97%** (5.4) 89	**98%** (5.2) 90	**94%** (5.4) 87
Korean Group	1 ms	L1	**98%** (4.7) 90	**92%** **(4.2)** 81	**94%** **(3.8)** 87	**80%** **(3.8)** 72	2% (2.0) 2	8% (4.0) 7	6% (2.8) 6	20% (3.2) 18

	L2	**98%**	**95%**	**90%**	**76%**	2%	5%	10%	24%
		(4.6)	**(4.9)**	**(4.4)**	**(4.3)**	(2.0)	(2.6)	(1.0)	(1.7)
		91	87	84	69	2	5	9	22
	L1	**91%**	**74%**	49%	15%	9%	26%	51%	**85%**
25 ms		(4.6)	(3.7)	(2.9)	(1.6)	(2.6)	(3.3)	(4.0)	(4.6)
		83	67	46	14	8	23	47	78
	L2	**82%**	66%	49%	18%	18%	34%	51%	**82%**
		(4.1)	(3.5)	(2.9)	(2.5)	(2.5)	(4.0)	(1.9)	(1.8)
		81	61	45	17	11	32	47	76
	L1	68%	32%	9%	NA	32%	68%	**91%**	**100%**
45 ms		(5.4)	(4.9)	(3.5)		(4.3)	(4.9)	(5.6)	(6.1)
		63	29	8		29	63	83	91
	L2	58%	28%	5%	2%	42%	**72%**	**95%**	**98%**
		(5.0)	(2.0)	(0.9)	(2.1)	(4.5)	(5.1)	(5.5)	(6.0)
		53	26	5	2	38	66	86	90

4.1.3 Analysis of Discrimination in L1 vs. L2

For the d' analysis, 40 data points per participant were utilized from the pool of 32 bilinguals, yielding a total of 1280 data points available for analysis.

4.1.3.1 d' in L1 vs. L2 language mode

A three-way ANOVA was conducted to examine the effects of language mode (L1 vs. L2), step (4 vs. 12), and F0 (low vs. high) on bilingual listeners' d'. All three factors were sum coded and varied within participants. The analysis revealed a significant main effect of step ($F(1, 1272) = 507.91$, $p < .001$) and a significant interaction between F0 and language mode ($F(1, 1272) = 4.93$, $p < .05$). Post-hoc tests using Tukey multiple comparisons revealed that this interaction stemmed from a significant F0 range effect in L1, which was not observed in L2. Figure 16 illustrates that bilinguals' d' was lower in the high F0 range than in the low F0 range during the L1 Korean session. The same model was applied to the monolinguals' data, revealing a significant effect of step ($F(1, 1232) = 315.08$, $p < .001$), but no significant effects for F0 range or language mode. This result suggests that Korean bilinguals' perceptual sensitivity was affected by category membership in L1 Korean mode, while it was not affected by category membership in L2 English mode.

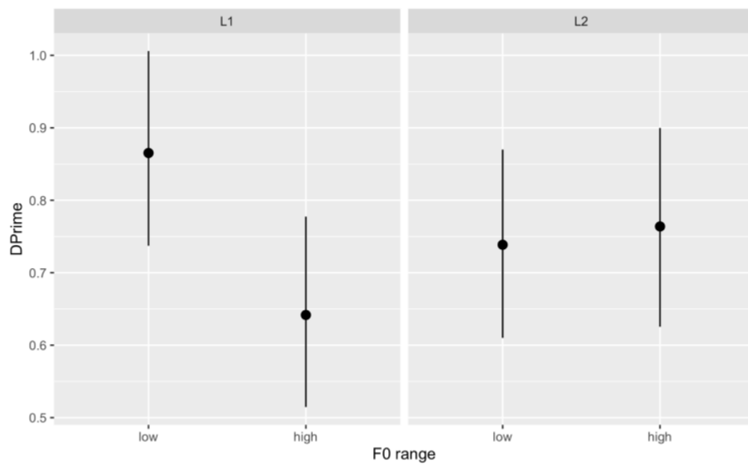

Figure 16 d' of Bilinguals in L1 Korean vs. L2 English

Table 14 Mean d' (SD) of Korean-English bilinguals in L1 vs. L2

Mean d' (SD) of Korean-English bilinguals averaged across step sizes in Korean vs. English language mode.

Language Mode	Low F0		High F0
L1 Korean	**0.87 (1.21)**	>	**0.64 (1.20)**
	≅		≅
L2 English	0.72 (1.15)	≅	0.74 (1.18)

4.1.3.2 Analysis of Response Bias by Groups

A two-way ANOVA was conducted on response bias (measured as C) to assess the influence of F0 range (low vs. high) and language mode (L1 vs. L2). The analysis revealed a significant main effect of F0 range ($F [1276] = 61.02$, $p < .001$) but no significant effect of language mode. The control group underwent the same analysis, where no significant factors were observed.

4.2 Discussion

The primary objective of this chapter was to investigate the influence of language mode on speech perception in Korean-English bilinguals, addressing two key questions: First, whether these bilinguals employ distinct cue-weighting strategies in Korean versus English language mode (i.e., language mode effects at the level of cue-weighting), and second, whether their perceptual sensitivity varies across these language modes (i.e., language mode effects at the level of perceptual sensitivity). Previous research has provided some evidence for the first question (Dmitrieva, 2019), but the answer to the second question has remained largely unclear, as perceptual sensitivity has not been directly examined in the context of the language mode literature. Different predictions were made according to various approaches to examining perceptual sensitivity. Previous studies in sound category training demonstrated that perception is relatively stable and not easily influenced by laboratory-based training (W. Heeren & Schouten, 2008). Based on this, bilingual listeners' perceptual sensitivity is predicted to remain the same across different language modes. On the other hand, theoretical models like the Generalized Context Model (Nosofsky, 1986, 2011) or the Localized Warping Model (Goldstone, 1994) propose a fundamental connection between perceptual sensitivity and attention to acoustic dimensions. This raises the possibility that shifts in attention to cues, influenced by language mode, could impact sensitivity. In this case, bilinguals' perceptual sensitivity is predicted to change along with their selective attention to different acoustic dimensions. To answer these questions, I conducted an analysis of Korean-English bilinguals' categorization and discrimination performance in both Korean and English language sessions.

The analysis of labeling and goodness rating data in Korean-English bilinguals reveals subtle but distinct cue-weighting strategies between their L1 Korean and L2 English language modes. Notably, these bilinguals exhibited a behavior more similar to English monolinguals in L2 English than in L1 Korean sessions. First, the statistical analysis of raw labeling responses indicated a significantly higher voiceless response ratio at VOT 45 ms in L2 English compared to L1 Korean. This increase in voiceless responses aligns with the categorization pattern observed in English monolinguals, where VOT 45 ms sounds are typically classified as "voiceless." Importantly, this change suggests an increase in the reliance on the VOT dimension for categorization in English compared to Korean mode. Second, the analysis of the reliance score, calculated based on listeners' categorization responses to noncanonical stimuli (VOT 1 and 45 ms), revealed an increased tendency toward VOT reliance in English mode compared to Korean mode. This tendency was most noticeable for the Korean bilingual group whose L1 reliance score corresponds to dual primary cues than the Korean bilingual group whose L1 reliance score corresponds to VOT as a primary cue. Lastly, the analysis of F0 boundary at VOT 25 ms revealed a significant decrease in F0 boundary slope in English compared to Korean, a measure that has been interpreted as degree of cue-weight (Xu et al., 2006). These results collectively highlight that Korean bilingual listeners' cue-weighting strategies shifted in L2 English mode; bilinguals showed increased reliance on VOT and reduced reliance on F0 in general, showing language-congruent perceptual strategies.

The analysis of discrimination data revealed a notable impact of language mode on bilinguals' perceptual sensitivity. In Korean mode, bilingual listeners' d' varied depending on the

F0 range, showing improved d' in the low F0 range, which corresponds to an across-category region, and diminished d' in the high F0 range, which corresponds to a within-category region. In other words, Korean bilinguals' perceptual sensitivity was affected by category membership in Korean mode with better perceptual sensitivity at the across-category region and poor perceptual sensitivity in the within-category region. However, this pattern did not replicate in the English mode. Instead, the statistical analysis showed that d' values remained consistent across different F0 ranges in English mode. Given that the high F0 range (236-260 Hz) was considered within-category for "voiceless" sounds in both Korean and English session, the concept of categorical perception alone does not fully explain why discrimination in this range was poorer in Korean but not in English. One way to think about this is to consider the *degree* of F0 reliance in each language. Bilinguals showed differences in discrimination depending on the F0 range in Korean as they were actively relying on F0 to categorize the VOT 45 ms stimuli; F0 was part of a dual primary cue system. However, bilinguals may not have shown the same F0 range effects in English as they shifted their perceptual strategy in a language-specific way, decreasing their reliance on F0 while increasing their reliance on VOT. While Korean speakers were still categorizing sounds using F0 information in L2 English, the reliance on F0 was not enough to result in an F0 range effect.

In summary, my findings indicate that language mode not only influences the cue-weighting strategy of bilingual listeners but also impacts their perceptual sensitivity. While F0 and VOT cues remained essential in both language modes, there was a subtle and language-specific shift in reliance for each cue in that the reliance on F0 decreased while the reliance on VOT increased in English mode compared to Korean mode. Importantly, this dynamic adjustment in cue-weighting across the acoustic dimension seemed to be related to listeners' discriminative sensitivity. In Korean mode, perceptual sensitivity was influenced by category membership along the F0 continuum, with poor d' values for within-category stimuli and better d' values for across category stimuli. On the other hand, in English mode where listeners' reliance on F0 was reduced, perceptual sensitivity was no longer impacted by category membership.

5.0 Understanding Sources of Individual Variance

In this chapter, I present an exploratory analysis of individual variability in listeners' cue-weighting and their L2 learning experiences. Previous studies have highlighted the considerable diversity among Korean listeners in their use of both VOT and F0 information for English voicing contrast categorization of noncanonical stimuli (Schertz et al., 2016). While understanding the origins of this variability in cue-weighting among Korean listeners is challenging, existing research suggests that L2 learning experiences can significantly influence these perceptual strategies (Antoniou et al., 2012; Flege & Eefting, 1986; Hazan & Boulakia, 1993; W. Heeren & Schouten, 2008). Therefore, the aim of the present study is to explore whether individual differences in cue-weighting observed in Korean-English bilinguals in L1 Korean and L2 English are associated with individuals' history of learning English. In addition, the analysis aimed to explore whether individual listeners' shifting of relative reliance on either F0 and/or VOT between different language modes is associated with L2 learning experiences.

Based on arguments against adjusting p-values for multiple comparisons in exploratory studies, the current study reports the correlations using unadjusted p-values (Bender & Lange, 2001). The current study reports 'significant' results as 'exploratory results' to provide preliminary information to guide further confirmatory studies. Correlational analyses (rather than regression-based analyses) were used given that the indices of VOT and F0 reliance used here were derived from performance over the entire labeling task, such that an individual listener in each language mode has a single value for each measure.

Individual listeners' cue reliance was operationalized by cue reliance scores (Schertz et al., 2016) and F0 slope functions derived from the labeling task were used to index language-specific listening strategies (refer to Section 3.1.1). Reliance scores were derived from listeners' labeling performance at VOT 1 ms and 45 ms conditions, as described in detail in the previous chapter. These scores, ranging from 0 (indicating equal reliance on both VOT and F0) to 1 (representing heavy reliance on VOT), were interpreted as indicators of VOT reliance and were encoded accordingly in the subsequent analysis. Additionally, the F0 boundary slope at VOT 25 ms was examined to assess the relationship between F0 reliance and L2 experience. The F0 slope at VOT 25 ms exhibited significant differences between L1 and L2 in the previous analysis, making it a reliable indicator of individuals' language-specific cue-weighting; F0 slopes tended to be steeper in L1 Korean compared to that of L2 English. The F0 slope variable, in conjunction with reliance scores, provides a comprehensive way to analyze the participants' language-specific listening strategies as each variable was derived from labeling responses to different subsets of the stimuli.

To index L2 learning experience, age of acquisition, self-reported proficiency, and FTE scores (Flege & Bohn, 2021) were used. Proficiency levels were determined through self-rated categories (speaking, listening, writing, reading, and accentedness) in the language questionnaire. FTE scores were calculated by multiplying the number of years spent in an English-speaking country by the daily exposure percentage to English.

One potential association can be made about individuals' L2 experience such as greater English exposure and/or higher English proficiency levels with increased reliance on VOT in English. Such pattern might be possible based on the premise that L2 cue-weighting could be influenced by L2 language experiences, potentially aligning with cue-weighting patterns observed in native English listeners. Consequently, positive correlations between the three variables from the language history questionnaire (LHQ) and the L1 and L2 reliance scores were anticipated, if anything.

In addition, another potential pattern could emerge for the change in F0 slope and L2 learning experiences. Such association would be possible based on the PAM-L2 where L2 learners refine their understanding of language-specific phonetic categories through learning experience. Given that F0 is a significant cue in Korean but less so in English, it was hypothesized that individuals with more refined language-specific phoneme categories might decrease their reliance on F0 in L2 English compared to L1 Korean. Prior research has successfully quantified cue weight through the steepness of the boundary slope (Dmitrieva, 2019), a method adopted in this analysis. The F0 slope difference was obtained by subtracting each participant's L1 slope value from their L2 slope value. The negative values in F0 slope differences mean slope was steeper in L1 than in L2 mode, indicating greater F0 reliance in L1 than in L2. On the other hand, the positive values in F0 slope differences would mean steeper slope in L2 than in L1, suggesting greater F0 reliance in L2 English than in L1 Korean; this would be unanticipated because F0 is an important primary cue in stop contrasts in Korean (L1) but not in English (L2). Therefore, the directionality of correlation was examined for the three L2 experience variables to test the hypothesis whether a heightened awareness of the distinct phonetic properties of Korean and English might lead to a more substantial perceptual shift in a language-congruent way.

To summarize, exploratory analysis was conducted to investigate potential associations among L2 experience (English proficiency, L2 AoA, and FTE) and language-specific listening strategies.

5.1 Results

Two distinct datasets were created from the 32 bilingual participants for separate correlation analyses. The first dataset encompassed listening performance data in L2, while the second dataset contained listening performance data in L1. Both datasets additionally included three variables representing L2 language learning experience: AoA, FTE, and proficiency. FTE was log transformed to achieve log-normality. Participant 101 was excluded from both datasets due to extreme values in FTE and AoA (more than 3 SD away from the other participants' mean). Additionally, the data from participants 120 and 126, whose portions of F0 boundary values were removed earlier due to their F0 boundary being outliers (outside of 0 ~ 400 Hz cutoff) and left with boundary data points for either one of the language mode, were removed from the correlation test to ensure that all individuals in the correlation analysis contribute both L1 and L2 data. Additionally, participant 105 whose slope value showed extreme values (more than 3 SD away from the other participants' mean) was removed from the analysis. This led to the exclusion of four participants (subjects 101, 105, 120, and 126). Consequently, the final datasets comprised 28 participants each, ensuring a matching number of participants overall.

5.1.1 Analysis of L2 cue-weighting and L2 language experience

Pearson's correlation coefficient tests were conducted using the rcorr function from the Hmisc package to analyze the relationships between these variables. As for the main interests of relations between cue-weighting in L2 English mode and L2 language experience, no significant correlations were found (Table 15). Several significant correlations were found among other parameters in the analysis. First, a statistically significant negative correlation emerged between

F0 slope and VOT reliance score (r = -0.45, p <.05) (Figure 17a), suggesting that listeners who showed a higher reliance score on VOT (as indexed by higher reliance score) tended to show a less reliance on F0 (as indexed by less steep F0 slope). This finding is as expected and indicates that even though the reliance score and F0 slopes were calculated from responses to different subsets of the stimuli, both variables are providing an index of listeners' cue-weighting strategies. There was also a significant positive correlation among LHQ factors, in particular, between FTE and proficiency (r = 0.40, p<.05) (Figure 17b), indicating that bilingual listeners with greater exposure to English tended to rate their proficiency higher. In summary, the results of the correlation analysis revealed that none of the L2 experience variables (FTE, AoA, and proficiency) were significantly related to bilingual listeners' cue-weighting (F0 slope and VOT reliance score) in L2 English.

Table 15 Summary of Correlation Table

Pearson correlations (p values) of LHQ and cue-weighting in each language mode. Correlations significant at or below the *p* = .06 are marked in bold.

Parameter	EN F0 slope	EN Rel score	KR F0 slope	KR Rel score	F0 slope difference	Rel score difference
FTE	-0.04 (0.84)	0.23 (0.23)	**-0.37 (<0.06)**	0.29 (0.13)	**0.43 (<0.05)**	-0.11 (0.57)
Proficiency	0.08 (0.70)	0.13 (0.49)	0.16 (0.43)	0.16 (0.43)	-0.11 (0.57)	-0.04 (0.84)
AoA	0.12 (0.55)	0.00 (1.00)	0.08 (0.67)	-0.02 (0.90)	0.03 (0.89)	0.05 (0.79)
F0 slope diff	0.29 (0.14)	0.06 (0.75)	**-0.52 (<0.05)**	0.23 (0.24)	-	-0.33 (0.08)
Rel score diff	-0.16 (0.43)	0.27 (0.16)	0.12 (0.55)	-0.22 (0.26)	-0.33 (0.08)	-

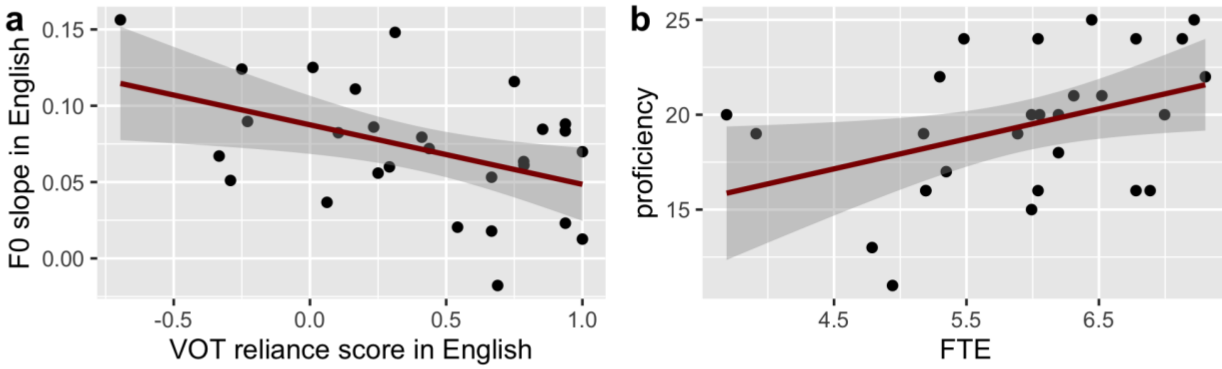

Figure 17a-b. a. Correlation between VOT Reliance Score and F0 Slope; b. Correlation between FTE and Proficiency

5.1.2 Analysis of L1 cue-weighting and L2 language experience

The same Pearson's correlation coefficient tests were conducted on the Korean data using the rcorr function from the Hmisc package to analyze the relationships between L2 learning variables and L1 cue-weighting. The results are presented in Table. As in the English data, there was a significant negative correlation between slope and reliance ($r = -0.51$, $p < .01$). Interestingly, two additional correlations were found significant which were not significant in the English data. First, a significant correlation emerged between F0 slope and slope difference ($r = -0.52$, $p < .01$) (Figure 18a). Figure 18a shows that higher F0 slopes in L1 Korean relate to more negative values in L1-L2 slope difference. This suggests that listeners with greater reliance on F0 in L1 showed a bigger change in F0 reliance between L1 and L2, with greater F0 reliance in L1 than in L2. On the other hand, less reliance on F0 in L1 tends to be related to little change in F0 reliance between L1 and L2. In other words, listeners who rely more heavily on F0 in Korean are more likely to shift their listening strategy when listening in English, whereas listeners who do not rely heavily on F0 in Korean tend not to shift their listening strategy when listening to English. Secondly, there was a marginally significant negative correlation between FTE and F0 slope in L1 Korean (r = -0.37, p <.06) (Figure 18b), suggesting that the greater the FTE (i.e., greater cumulative English experience), the less steep the F0 slope when listening in Korean. Neither AoA nor proficiency level showed significant correlations with either of the L1 cue-weighting variables nor F0 slope shift between language modes.

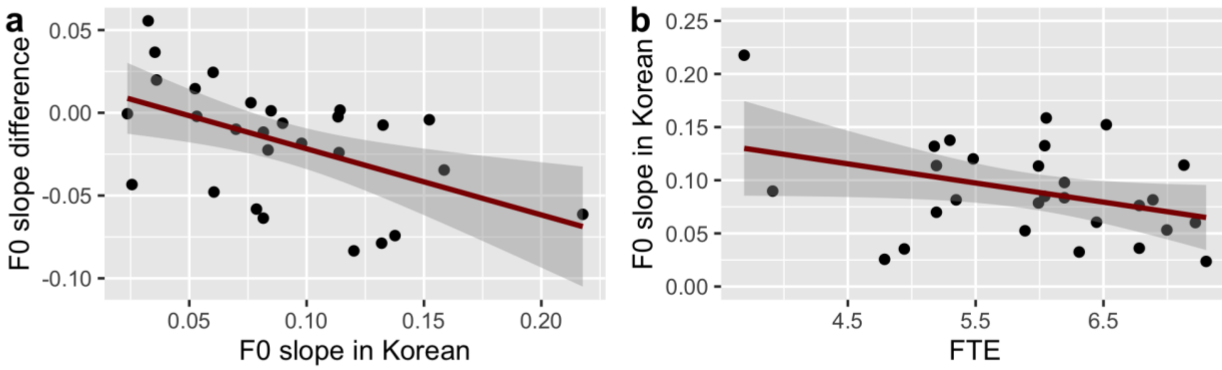

Figure 18a-b. a. F0 Slope in Korean and L1-L2 Slope Difference; b. FTE and F0 Slope in Korean

5.1.3 Analysis of L1-L2 difference in cue-weighting and L2 experience

There was a significant positive correlation between FTE and the shift in F0 slope between L1 and L2 and (Figure 18). Figure 18 shows that bilinguals with lower FTE values (i.e. less English experience) tend to show negative values in slope difference, meaning that their F0 reliance was greater in L1 Korean than in L2 English. However, bilinguals with greater FTE tend to show slope difference values around 0, suggesting that their reliance on F0 was more or less equal in both L1 Korean and L2 English. It seems that bilingual participants with less lifetime English exposure demonstrated greater shifting in F0 reliance between languages while bilingual participants with more cumulative exposure to English demonstrated less shifting in F0 reliance between languages. This finding is somewhat counterintuitive and will be discussed in the following section.

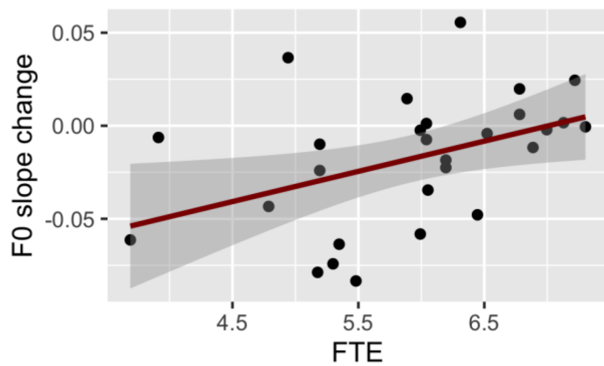

Figure 19 Correlation between FTE and F0 Slope Difference

5.2 Discussion

One of the main objectives of this chapter was to conduct an exploratory analysis to test potential associations among Korean-English bilinguals' L2 English experience and their cue-weighting strategies. This analysis was motivated by previously reported individual variance in Korean-English bilinguals' cue-weighting in an L2 English context (Schertz et al., 2016). Another objective of this chapter is to investigate whether the observed shifts between L1 and L2 in cue-weighting are related to L2 learning experiences. Given that voicing categories are contrasted by VOT by native English speakers, I expected that Korean-English bilinguals who developed more target-like categories would exhibit greater reliance on VOT in an L2 English context. I also predicted that Korean-English bilinguals with more distinct L1 vs L2 categories would show greater perceptual shift between different language modes.

Correlation analyses were conducted considering three factors from individual learners' L2 language: age of acquisition, proficiency, and FTE. Individual listeners' language-specific cue-weightings were approximated by two variables. First, *F0 slope* was used to measure individual listeners' reliance on F0 information in an ambiguous VOT condition (25 ms). Second, a *reliance score* was used to measure individual listeners' reliance on VOT information in non-canonical conditions (VOT 1 and 45 ms). In addition, individual listeners' degree of perceptual shift was quantified by measuring differences in *F0 slope* between Korean and English modes as well as differences in *reliance scores* between Korean and English modes.

If age of acquisition plays a primary role in developing separate L2 categories for second language learners, we would predict that speakers with lower age of acquisition will have more successfully developed separate L2 categories, and therefore show greater reliance scores on VOT in an English context. On the other hand, if proficiency was expected to drive the distinct development between L1 and L2 categories, proficiency was predicted to have a positive correlation with VOT reliance score in an English context. Lastly, if the amount of L2 experience was the determining factor for L2 category development, we expect a significant correlation between FTE and VOT reliance score. In addition, if L2 experience affects L1 cue-weighting as predicted by the ADAPPT, we may predict significant correlations between FTE and cue-weightings in L1.

The results revealed no significant correlations between age of acquisition and either L2 cue-weighting or perceptual shift. The results also showed no significant correlations between proficiency and either L2 cue-weighting or perceptual shift. However, there were significant correlations between L1 cue-weighting patterns and FTE. The greater English exposure approximated by higher FTE was related to less reliance on F0 and greater reliance on VOT in L1 Korean mode. This supports the hypothesis based on ADAPPT that significant L2 exposure can lead to perceptual changes in L1. This result also partially explains why the majority of Korean-English bilinguals in our current study, immersed in an English-speaking country at the time of testing, showed cue-weighting of either VOT or VOT+F0 with almost no one showing an F0-relying pattern in non-canonical stimuli conditions, differently from previous studies (Schertz et al., 2015, 2016) whose Korean participants were recruited in Korean at the time of testing.

Another result found in the current analysis is a significant positive correlation between F0 slope and L1-L2 F0 slope difference (Figure 18a). It is possible that higher FTE indicates that the listeners are relatively more L2 dominant, hence behaving more similarly to English monolinguals with less reliance on F0 and more reliance on VOT even in their L1 Korean. Such a listening pattern leads to little need for cue-weighting shift in the L2 English condition. On the other hand, listeners with lower FTE seem to maintain more distinctive cue-weighting between their Korean vs. English mode, demonstrating perceptual adaptability of increasing and reducing their reliance on F0 between languages.

6.0 General Discussion

This work examines the impact of the contrastive use of F0 in L1 on Korean-English listeners' perceptual sensitivity, focusing on the relative reliance on two cues (F0 and VOT) within the context of contrasts in L1 Korean and L2 English. This is the first study to test the perceptual warping effects for pitch information for speakers of a non-tonal language, Korean. To the best of the author's knowledge, this is also the first study to systematically examine bilingual listeners' perceptual sensitivity using the same stimuli in different language modes. In addition, the current study explores the relationship between bilinguals' L2 learning history and their cue-weighting patterns. A comparison of the Korean-English and monolingual English groups in their respective L1 reveals clear group differences in sensitivity to F0: not only did the Korean-English group require both VOT and F0 for contrast, but they also showed poor perceptual sensitivity when listening to pitch differences within the same category. In terms of language-specific listening patterns, when compared to their L1 Korean mode, bilingual listeners in English mode demonstrated a subtle change in their relative reliance on F0 and VOT, as well as in their perceptual sensitivity. When examining the second language learning history of individual listeners, it was found that bilinguals with less exposure to English (operationalized in terms of FTE, or "Full Time Equivalent"; Flege & Bohn, 2021) tended to make a greater shift in their relative reliance on F0 between languages. In the sections that follow, these findings will be discussed in turn.

6.1 Contrastive Use of Pitch for Stop Consonants and Perceptual Sensitivity

The discrimination performance of the bilingual listeners in their L1 Korean mode in the present study suggests that the contrastive use of level pitch in the native language can have an impact on their perceptual sensitivity. In line with the hypothesis of the Localized Warping Model, the discrimination performance of the bilingual group varied between across-category and within-category regions of the F0 space; the same pitch differences were perceived as more similar when they occurred within-category (higher F0 region) as opposed to when they occurred across category (lower F0 region). The observed decrease in perceptual sensitivity within the aspirated stop region is comparable to the previous work demonstrating poorer discrimination performance for pitch differences in within-category regions among tonal language listeners compared to non-tonal language listeners (Liu et al., 2017; Xu et al., 2006). In a similar line, Stagray and Downs (1993) also found evidence of poorer perceptual sensitivity of tonal speakers than non-tonal speakers. Using synthetic standard tones at 125 Hz and 1000 Hz with variable tones differing by increments of 0.3 Hz (for 125 Hz tone) or 0.5 Hz (for 1000 Hz tone) from each standard tone, they tested how well each group of listeners can detect the differences between the standard and variable tones. Based on the results showing that Mandarin listeners required a higher threshold to detect frequency differences than English listeners, the authors suggested that Mandarin listeners' language experience of grouping sounds of similar frequency together may have led to poor perceptual sensitivity to detect frequency differences.

Perhaps as expected, the absence of a discrimination peak at the across-category region in the present study stands in contrast to findings from previous research demonstrating discrimination peaks with contour tones (Chan et al., 1975; Francis et al., 2003; Hallé et al., 2004; Liu et al., 2017; Peng et al., 2010; Xu et al., 2006). Instead, the current findings align with previous research on level pitch stimuli which did not present discrimination peaks (Abramson, 1961, 1979; Francis et al., 2003). In the current study, the direct group comparison did not show evidence of superior perceptual sensitivity among the Korean listeners compared to the English control group. This result was somewhat unexpected and contradicts the predictions of the Generalized Context Model, which suggests that selective attention to a particular acoustic dimension would generally enhance the sensitivity across the dimension. This finding also contrasts with work demonstrating greater perceptual sensitivity among tonal speakers compared to nontonal speakers (Hallé et al., 2004; Xu et al., 2006). In fact, Korean listeners' perceptual sensitivity within the aspirated category region was poorer compared to English listeners. This observation suggests that in the context of level-pitch perception, the notion of *acquired equivalence* (Goldstone, 1994; Goldstone & Hendrickson, 2010; Pisoni & Lazarus, 1974) may be most relevant. Stagray and Downs (1993) suggested that Mandarin listeners exhibit diminished sensitivity to level tone differences, potentially due to their categorization strategy that emphasizes grouping similar frequencies together while disregarding smaller distinctions. Similarly, for our current finding, it is possible to posit that Korean listeners have adopted a similar categorization strategy for stop sounds, accentuating the likeness among various frequencies within categories, while retaining strong discrimination across other frequency ranges that are not within categories. Although Korean listeners exhibit poorer perceptual sensitivity within category than English listeners, both English and Korean listeners exhibit comparable perceptual sensitivity across F0 ranges outside the Korean aspirated stop category. Such comparable sensitivity observed in both groups is consistent with a general robustness in perceiving level pitch differences among most listeners (Wang, 1976). However, this sensitivity may be compromised in instances if listeners develop a categorization strategy of disregarding pitch differences, resulting in shrinking of the psychological space within category along the F0 dimension.

In summary, the findings regarding native Korean listeners' perceptual sensitivity to pitch when listening in L1 Korean indicate that the contrastive use of level pitch does impact their perception of this cue. However, in contrast to the predictions of the GCM, the mechanism of acquired equivalence, rather than acquired distinctiveness, appears most relevant for understanding how the L1 category space relates to perceptual sensitivity to pitch in an L1 listening context.

6.2 Language Mode Effects on Bilinguals' Speech Perception

Previous studies have shown that bilingual listeners' perceptual strategies can differ based on language mode (Antoniou et al., 2012; Elman et al., 1977; Garcia-Sierra et al., 2009; Gonzales et al., 2019), and the changes in bilinguals' cue reliance between L1 and L2 in the present study are largely in line with this work. Garcia-Sierra et al. (2009) examined how Spanish-English bilinguals and English monolinguals label a continuum varying in VOT (/ga-ka/) in Spanish and English contexts. They found that only the bilingual group showed a significant correlation between the magnitude of the shift and the level of confidence in using English. Similarly, Gonzales and Lotto (2013) investigated the labeling performance of early Spanish-English bilinguals and English monolinguals using stimuli varying in VOT embedded in pseudo words (bafɾi-pafɾi) in Spanish and English contexts. Their results show only bilinguals significantly shifted their boundary at short-lag VOT, a range identified as voiced in English language but as voiceless in Spanish. Gonzales et al. (2019) presented similar findings for French-English and

Spanish-English bilinguals using a /ba-pa/ continuum where both bilingual groups showed significant boundary shifts to be lower in the French or Spanish context compared to the English context. The present study found evidence for a language-specific F0 boundary demonstrated by a significantly lower F0 boundary in English than in Korean mode among Korean-English bilinguals in the VOT 1 ms condition, in line with work exhibiting evidence of a "double boundary" among bilinguals (Gonzales et al., 2019; Gonzales & Lotto, 2013; Schertz et al., 2020). Furthermore, in the present study, there was a notable increase in voiceless responses at VOT 45 ms in English mode compared to Korean mode. This finding contributes to the evidence suggesting language-specificity in phonetic representations in bilinguals' speech perception.

Importantly, the current results highlight subtle but significant shifts at the level of cue-weighting, demonstrated by changes in relative reliance on VOT and F0 information in a language-congruent way. These results stand in contrast with previous work suggesting language specificity in bilinguals' perception may not extend to cue-weighting (Schertz et al., 2015, 2020). Schertz et al. (2015) tested L1 Korean listeners' weighting of F0 and VOT cues between L1 Korean and L2 English conditions and concluded the Korean listeners maintained stable cue-weighting across languages. This difference in results may have come from the different experimental design and methodology adopted for analysis. One difference between studies involves the way in which language mode was encouraged. In Schertz et al. (2015), the authors provided both oral and written instructions in the target language during each session, with stimuli embedded in carrier sentences in either English or Korean. The current study not only delivered oral and written instuctions in the target language, but participants also received regular progress reports in each task (see Section 2.4) in the target language, both in written and spoken form. This additional reinforcement of the language mode may have contributed to participants in the current study exhibiting more distinct language patterns in their listening performances. Another difference between studies include the type of analysis adopted in the study. Using stimuli varied in both VOT and F0 in nine steps, Schertz et al. (2015) obtained the perceptual weight for each cue by taking the coefficients of each variable (VOT and F0) from a fitted logistic regression model. The perceptual weights between different language sessions were examined for significant correlations. On the other hand, the current study examined individual listeners' relative reliance on cue (approximated by reliance score) based on their labeling performance in non-canonical stimuli. Additionally, the current study investigated relative reliance on F0 in ambiguous VOT condition between different languages. These dependent measures were analyzed using a linear regression model with *language mode* as a predicting variable. It can be argued that the analysis adopted in the current study allowed for a more direct test of the effects of language mode on listeners' relative cue reliance. Schertz et al. (2020) tested secondary cue-weighting of Spanish-English bilinguals in different language modes and found that bilingual listeners maintained stable reliance on their secondary cue. A major difference between that study and the current one is that Schertz et al. used Spanish and English contexts where VOT is arguably equally relevant in both languages, whereas in the current design, the language contexts differed from another in the relevance of F0, with F0 serving as a part of a dual-primary cue system in Korean while F0 is only a secondary cue in English. The current findings mirror those of Dmitrieva (2019), who explored the relative dependence of Russian-English listeners on perceptual cues varying in importance across each language. Our results complement their evidence in that language mode had an impact on bilingual listeners' reliance on cues in a manner consistent with each language. Thus, it is plausible that language mode can influence cue-weighting, particularly for primary cues differing in relevance between languages.

Furthermore, the current findings reveal a complementary relationship in the shifts of relative reliance between the F0 and VOT cues, with Korean-English bilingual listeners enhancing their reliance on VOT while reducing their reliance on F0 in English mode. This aligns with findings by Dmitrieva (2019), where Russian-English bilinguals exhibited heightened reliance on vowel duration and diminished reliance on glottal pulsing in Russian mode. Such observations resonate with the concept of finite attentional resources proposed by the GCM, suggesting that an increase in reliance of one cue would accompany a decrease in reliance on the other cue (Nosofsky, 1986, 2011). Therefore, it's plausible that bilingual listeners' alterations in relative reliance on multiple cues may be influenced by the optimal distribution of their selective attention resources, and that as they direct more attention to cues that are language relevant, attention to other less relevant cues decreases.

Previous studies have differed in terms of their predictions for how L1 and L2 category structure should be expected to relate to perceptual sensitivity. Antoniou et al. (2012) proposed that within a discrimination task, bilinguals integrate information from both languages, without employing language-specific strategies. As mentioned earlier, Antoniou et al. (2012) investigated the discrimination of stop voicing contrast among Greek-English bilinguals in various language contexts. They discovered that the bilinguals' discrimination performance remained unaffected by language mode, closely resembling that of English monolinguals. The authors proposed that language mode effects might manifest in tasks that explicitly designate a language for judgments, such as categorization or goodness judgment. However, in tasks where response options lack overt language specificity, such as discrimination, bilinguals may default to their dominant language. The current study is different from Antoniou et al. (2012) in the type of discrimination focused on. While their study focused on listeners' discrimination performance of contrastive categories, such as distinguishing between voiced and voiceless stops, our study examines listeners' discrimination performance both within and across categories. Specifically, we investigate how listeners differentiate sounds within the same category and those from category-ambiguous (across) regions. This design allows for detailed examinations for listeners' processing of gradient acoustic information varying in category memberships.

The present study demonstrates notable effects of language mode on the perceptual sensitivity of Korean-English bilingual listeners. Specifically, listeners displayed diminished perceptual sensitivity within categories only in the Korean language mode, contrasting with English mode. This discovery challenges the notion that language-specific perceptual strategies only emerge in tasks involving explicit language judgment. Rather, it suggests that language mode can also affect bilinguals' perceptual sensitivity as their sensitivity is variably influenced by the structure of their native and second language categories. These aspects of the current results align more closely with Bosch et al. (2000), where bilinguals exhibited evidence of reduced sensitivity within their L1 category but not within L2 category. As in Bosch et al. (2000), the goodness ratings of the voiceless category in this study were comparable across L1 and L2, suggesting that the stimuli were an equally good representation of both L1 and L2 categories. So why was there reduced sensitivity in L1 Korean but not in L2 English mode?

To explain the asymmetry, I first consider the possibility that reduced relative reliance on F0 in an L2 English listening context may impact how category membership affects perceptual sensitivity. Category-derived warping effects, such as acquired similarity, may only be observed when listeners pay sufficient attention to the relevant domain, which is F0 in our case, positing a fundamental dependence of perceptual sensitivity on selective attention. In this scenario, we might anticipate several outcomes; First, given that the psychological space within category may shrink due to the localized warping effects, we may expect a *negative* correlation between the degree of reliance on F0 (F0 slope) and sensitivity (d') *within* category in L1, such that listeners with greater F0 reliance may show worse d' within category in L1. In a similar vein, we may expect a *positive* correlation between the degree of reliance on F0 in L1 and the change in sensitivity within category between L1 and L2, such that listeners with greater F0 reliance may show larger changes in their perceptual sensitivity between languages. The correlation test results were not significant for either of these cases. However, somewhat surprisingly, there was a significant *negative* correlation between F0 slope in L1 and change in d' in *across* category regions ($r = -0.52$, $p < .05$). At first, this result seems to contradict the earlier hypothesis in two ways: First, the correlation of F0 reliance was found not between *within* category, but between *across* category. Second, there was a *negative* correlation between the degree of F0 reliance and the change in sensitivity, instead of a positive correlation posited in our assumption that greater F0 reliance in L1 relates to greater change in perception between languages, not less change in perception between languages.

However, upon visualizing this correlation (Figure 19), it appears that it might be more fitting to conceptualize the change in perception between languages in terms of 'directionality' rather than 'size'. Since the change in perception was determined by subtracting d' in L1 from L2 (L2-L1), negative values suggest greater discriminability in L1 (Korean) than in L2 (English), whereas positive values indicate the opposite. Therefore, the negative correlation indicates that listeners with greater F0 reliance tend to show greater discriminability in across category in L1 Korean than in L2 English while listeners with less F0 reliance tend to show greater discriminability in across category in L2 English than in L1 Korean. As the category effects involve worse sensitivity within the same category but better sensitivity within the across category region, the increase in the sensitivity in the across category region indicates more pronounced category effects. The current results show that listeners with high F0 reliance tend to show more pronounced category effects in Korean than in English. This aligns with a previous finding (Section 4.1.3.1) where category effects on perception were only significant in Korean, not in English. The correlation indicates that this perceptual pattern tends to be more pronounced in listeners with a stronger reliance on F0, characterized by greater sensitivity across category and poorer sensitivity within category. Hence, although the results are not straightforward, they still provide support for the relationship between selective attention and the warping effects.

The view that listeners' perceptual abilities are closely linked to their selective attention has been previously suggested (Pisoni et al., 1994). For example, Francis et al. (2008) showed that adult English speakers learning Korean can demonstrate perceptual learning patterns along the F0 dimension. Such findings suggest perceptual sensitivities can be modified with selective attention. The view also aligns with exemplar-based phonology models where the activation of exemplars in response to input and the spread of that activation in a network of exemplars are affected by both phonetic and nonphonetic properties (Johnson, 2007). For example, it is possible to posit that bilinguals' different language mode acts as contextual, nonphonetic properties to impact the relative activation of the network of exemplars, ultimately resulting in greater activations of exemplars conveying contrastive pitch information in Korean mode than in English mode. In such a case, the degree of activation of exemplars can be related to the degree of perceptual warping effects. The current findings suggest that bilingual listeners' perceptual distance can be modified in a language-specific way and such process is mediated by altering attentional processes in that the perceptual warping is primarily due to selective attention rather than permanent changes in sensory mechanisms.

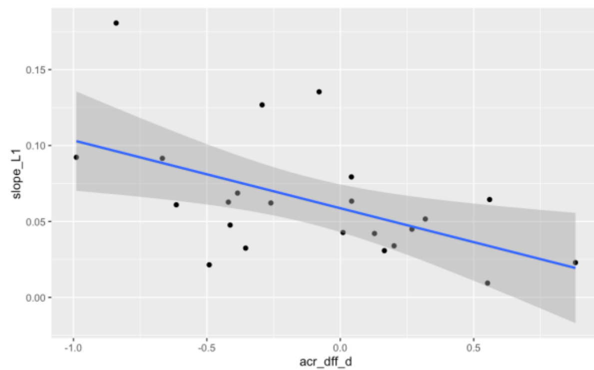

Figure 20 Correlation between Change in d’ between Languages (x-axis) and F0 Slope (y-axis) Negative values in x-axis indicate greater discriminability in the across category region in L1 Korean than in L2. Listeners who relied more on F0 tend to show greater discriminability in L1, suggesting more pronounced category effects in L1 Korean than in L2 English.

Alternatively, we may consider the possibility that the asymmetry in category effects on perception between L1 Korean and L2 English could stem from differences in memory usage across language contexts. Pisoni (1973) posited that short-term memory recruited in speech perception comprises two components: a continuous *auditory short-term memory,* which holds relatively accurate representations of sounds but tends to decay rapidly, and a *phonetic short-term memory*, characterized by longer but less veridical retention of items due to their connection with representations residing in long-term memory. Pisoni and Tash (1974) found that reaction time for discriminating sounds takes longer when the sounds are phonetically similar but acoustically different (A-a) than the sounds that are phonetically and acoustically the same (A-A). They suggested that such difference in reaction time reflects multi-levels of processing of sounds involving acoustic and phonetic evaluations. Based on this model, sounds between categories are encoded with both categorical phonetic information and continuous acoustic information, while sounds within categories are only encoded with continuous acoustic information. McMurray (2022) suggested that such model can account for better discriminability of sounds between categories than within category without assuming any perceptual warping effects. Additionally, cognitively more demanding task designs, such as those incorporating a longer Inter-Stimulus Interval (ISI) and the inclusion of brief interrupting noise between stimuli, are typically associated with the activation of a phonetic memory mode (Guenther et al., 1999; Massaro & Cohen, 1983). In task designs where the ISI is neither particularly long nor short, it is possible that the language mode could play a crucial role in determining the utilization of different memory modes. Since other studies that found categorical perception in AX discrimination tasks adopted a slightly longer ISI of 500 ms, I suspect that the relatively short ISI of 300 ms in the present study may have led listeners to employ different memory modes depending on the language context. It is possible that our Korean-English bilingual participants, when in Korean mode, relied more on phonetic memory due to having more robust long-term representations in L1 Korean, while in English mode, where long-term representations may be less robust, they may have relied more on auditory memory, with long-term representations that are less robust in their memory. This memory-based account also provides a coherent explanation for the findings from Bosch et al. (2000) where highly proficient bilinguals displayed an acquired similarity effect in their L1 but not in L2. Bosch et al.'s AX discrimination task used an ISI of 250 ms, which could similarly allow bilingual listeners to employ either of the memory modes depending on the language mode. At present, however, this explanation is only speculative, and more systematic perceptual experiments need to be done in order to gain a better understanding of how language mode may interact with ISI to modify the involvement of different memory systems.

6.3 L2 Experience and Speech Perception

The final question in this study concerns the relationship between language-specific cue-weighting and L2 learning experience including age of acquisition, self-rated proficiency, and the degree of lifetime L2 exposure (FTE). While certain predictions for variables were made prior to the analysis, it's important to note that the analysis was exploratory in nature with a primary aim to uncover any potential patterns within the data to guide the direction for the future confirmatory studies. The present study found exploratory correlations between the amount of L2 English input (calculated by the length of residence and the daily use of English) and listeners' cue-weightings in L1. Specifically, individuals with a history of greater L2 English input tended to show less reliance on F0 (indexed by a shallower F0 slope) when listening in Korean mode and when VOT was equal to 25ms. Given that the F0 dimension contrasts stops in Korean but not in English, this result is in line with work showing that bilinguals with greater L2 dominance tend to present perceptual strategies impacted by their long-term dominant languages (Antoniou et al., 2012; Hazan & Boulakia, 1993). Such findings can be accommodated by the recently proposed ADAPPT model (de Leeuw & Chang, 2023), which generally predicts that L1 perception strategies will come to resemble those of L2 over time. The ADAPPT posits that L1 perception and production always remain malleable and subject to change as a result of L2 exposure and acquisition. The model suggests two levels of changes: *drift*, which refers to short-term changes caused by recent/ongoing L2 experience and *attrition*, which refers to long-term changes caused by the totality of L2 experience that is not necessarily recent or ongoing (de Leeuw & Chang, 2023). It is possible to posit that Korean listeners may exhibit a case of *attrition* in their reliance on F0 in L1, as their cumulative L2 English experience increases, adopting a cue-weighting pattern close to that of monolingual English speakers. For Korean listeners who reside in Korea, their dominant language will be Korean. In such a case, we may expect to see more listeners relying on F0 assuming an active tonogenesis in Korean language. In fact, previous studies on dual-cue reliance among Korean listeners residing in Korea (Schertz et al., 2015, 2016) showed more people rely on F0 cue in noncanonical English context than the participants in the current study. Given that the Korean listeners in the previous studies demonstrated high proficiency level with Korean language as a dominant language, changes in listeners' cue-reliance patterns might be most strongly driven by the language dominance than the proficiency level.

In addition to finding that Korean listeners with greater L2 English experience tend to rely less on F0, the current study also found an exploratory correlation between the degree of English exposure and the degree of perceptual shift across language modes; there was a tendency for individuals with *less* long-term English exposure to demonstrate *greater* perceptual shift between language modes. However, our findings also revealed a significant correlation between L2 English exposure and self-rated proficiency. Taken together, this was somewhat surprising, as it contradicts the prediction of the PAM, which suggests that increased L2 proficiency leads to a more distinct, separate development of L2 categories from L1 categories. In order to understand these results, the assumption of L2 effects on L1 from the ADAPPT needs to be considered. In the current study, bilinguals with greater English exposure seemed to adopt perceptual strategies akin to English speakers, even in their L1 Korean perception, thereby maintaining similar strategies in both L1 and L2 language contexts. In contrast, the L1 Korean perception of bilinguals with lower English exposure was less affected by their long-term L2 English experience, such that they appeared to maintain more distinctive perceptual strategies between their L1 and L2, resulting in greater adjustments between language modes. It appears that neither age of acquisition nor proficiency was related to individual variance in the cue-weighting data, despite being two factors commonly considered when examining second language learners' speech perception. Instead, our data indicate that the ADAPPT model's assumption of L2 effects on L1, especially in the context of potential attrition where cumulative L2 experience impacts L1, provided the best account for the individual variance in language-specific cue-weighting as well as the degree of shift across language modes. The current results indicate that examining relative reliance on cues for listeners with dual primary cues could provide a good testing ground to examine the viability of the ADAPPT model for its predictions of L2 experience on L1.

6.4 Conclusion

This dissertation sought to answer the questions of 1) whether Korean listeners' contrastive use of pitch in their L1 affects their perceptual sensitivity to pitch differences, 2) how language mode affects Korean-English bilinguals' listening strategy at the cue-weighting and perceptual sensitivity levels, and 3) whether individual bilingual listeners' language-specific cue-weighting can be explained in relation to their L2 learning experience. The current study is the first study to test the localized warping effect toward pitch information for speakers of a non-tonal language, and one of the first studies to systematically examine language mode effects on perceptual sensitivity. In examining the perception of level pitch by Korean-English bilinguals in Korean and English modes, the current work replicated previous findings of acquired similarity within L1 categories by demonstrating reduced sensitivity within L1 Korean stop categories. These bilingual individuals demonstrated flexible perceptual strategies at the levels of their relative reliance on two distinct phonetic cues while making categorization decisions, as well as their perceptual sensitivity when detecting very small differences between stimuli. Our results offered limited support for the fundamental relationship between selective attention and the effects of category. It was suggested that category effects may require sufficient attention to the given dimension. If the attention is redirected, indicated by reduced reliance on the dimension, listeners may not exhibit category effects. This interpretation can be understood in terms of exemplar-based theories of speech perception, in which previously experienced tokens can be differently weighted depending on their perceived relevance to the task. An alternative interpretation was suggested in that bilingual listeners may employ different memory systems (i.e., phonetic vs. auditory) when listening in their two languages, such that their perceptual sensitivity is affected by category membership in their L1 but not in L2. The present study also suggests that prolonged L2 exposure may result in bilingual listeners' utilization of more L2-like cue-weightings across both L1 and L2 conditions, while bilinguals with less L2 exposure may preserve more language-specific cue-weightings, displaying greater perceptual shift between language modes. A limitation of this study includes a lack of direct evidence for the memory-based account of asymmetry in the acquired equivalence effect. An avenue for future study could include testing the interaction between working memory load in varying ISI conditions and language specificity in listeners' perceptual warping. Testing the effects of language dominance by directly comparing bilinguals immersed in different countries would also help elucidate the applicability of the ADAPPT model.

Appendix A Korean and English minimal pairs

Lenis			Aspirated		
[bat]	밭	'field'	[pʰat]	팥	'red beans'
[bal]	발	'feet'	[pʰal]	팔	'to sell'
[ba. d͡ʒi]	바지	'pants'	[pʰa. d͡ʒi]	파지	'scrape papers'
[ba.da]	바다	'sea'	[pʰa.da]	파다	'to dig'
Voiced			**Voiceless**		
box			pox		
bot			pot		

Bibliography

Abramson, A. S. (1961). Identification and Discrimination of Phonemic Tones. *The Journal of the Acoustical Society of America*, *33*(6_Supplement), 842–842. https://doi.org/10.1121/1.1936844

Abramson, A. S. (1979). The noncategorical perception of tone categories in Thai. In *Frontiers of Speech Communication Research* (Vol. 1, pp. 127–134). Academic Press.

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics*, *40*(4), 582–594. https://doi.org/10.1016/j.wocn.2012.05.005

Apfelbaum, K. S., Kutlu, E., McMurray, B., & Kapnoula, E. C. (2022). Don't force it! Gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical Society of America*, *152*(6), 3728–3745. https://doi.org/10.1121/10.0015201

Archila-Suerte, P., Zevin, J., Bunta, F., & Hernandez, A. E. (2012). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism*, *15*(1), 190–201. https://doi.org/10.1017/S1366728911000125

Asher, J. J., & García, R. (1969). The Optimal Age to Learn a Foreign Language. In *Source: The Modern Language Journal* (Vol. 53, Issue 5). https://about.jstor.org/terms

Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, *66*, 120–144. https://doi.org/10.1016/j.wocn.2017.09.005

Bender, R., & Lange, S. (2001). Adjusting for multiple testing-when and how? In *Journal of Clinical Epidemiology* (Vol. 54).

Best, C. T. (1995). A direct realist view of crosslanugage speech perception. In *Speech perception and linguistic experience* (pp. 171–204). https://www.researchgate.net/publication/229078334

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In *Language experience in second language speech learning* (pp. 13–34). https://doi.org/10.1075/lllt.17.07bes

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. In *Frontiers in Psychology* (Vol. 9, Issue MAR). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2018.00081

Bohn, O.-S., & Flege, J. E. (1993). Perceptual switching in Spanish/English bilinguals. In *Journal of Phonetics* (Vol. 21).

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*(2), 189–221. https://doi.org/10.1080/09541446.2000.10590222

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Human, A. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, *54*(2), 421–428. https://doi.org/10.1121/1.1913594

Casillas, J. V., & Simonet, M. (2018). Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics*, *71*, 51–64. https://doi.org/10.1016/j.wocn.2018.07.002

Chan, S. W., Chuang, C.-K., & Wang, W. S.-Y. (1975). Cross-linguistic study of categorical perception for lexical tone. *The Journal of the Acoustical Society of America*, *58*(S1), S119–S119. https://doi.org/10.1121/1.2001896

Chen, F., & Peng, G. (2016). Context Effect in the Categorical Perception of Mandarin Tones. *Journal of Signal Processing Systems*, *82*(2), 253–261. https://doi.org/10.1007/s11265-015-1008-2

Chen, F., Yan, N., Wang, L., Yang, T., Wu, J., Zhao, H., & Peng, G. (2015). The development of categorical perception of lexical tones in Mandarin-speaking preschoolers. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2015-January*, 3130–3134. https://doi.org/10.21437/interspeech.2015-630

Cho, M.-R. (1994). Acoustic characteristics of Korean stops and perception of English stop consonants. In *Order Number*.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. In *Journal of Phonetics* (Vol. 30). http://www.idealibrary.comon

Cutting, J. E. (1982). Plucks and bows are categorically perceived, sometimes. In *Perception & Psychophysics* (Vol. 31, Issue 5).

Cuttingt, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music*. In *Perception & Psychophysics* (Vol. 16, Issue 3).

de Leeuw, E., & Chang, C. B. (2023). Phonetic and Phonological L1 Attrition and Drift in Bilingual Speech. In *The Cambridge Handbook of Bilingual Phonetics and Phonology*.

Dmitrieva, O. (2019). Transferring perceptual cue-weighting from second language into first language: Cues to voicing in Russian speakers of English. *Journal of Phonetics*, *73*, 128–143. https://doi.org/10.1016/j.wocn.2018.12.008

Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America*, *62*(4), 971–974. https://doi.org/10.1121/1.381591

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, *January 1995*, 233–277. https://doi.org/10.1111/j.1600-0404.1995.tb01710.x

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning*.

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, *43*(4), 155–171. https://doi.org/10.1159/000261768

Flege, J. E., & Eefting, W. (1987). CROSS-LANGUAGE SWITCHING IN STOP CONSONANT PERCEPTION AND PRODUCTION BY DUTCH SPEAKERS OF ENGLISH. In *Speech Communication* (Vol. 6).

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*, *91*(1), 370–389. https://doi.org/10.1121/1.402780

Francis, A. L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *The Journal of the Acoustical Society of America*, *114*(3), 1611–1621. https://doi.org/10.1121/1.1603231

Francis, A. L., Ciocca, V., & Chit Ng, B. K. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics*, 1029–1033.

Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America*, *124*(2), 1234–1251. https://doi.org/10.1121/1.2945161

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349–366. https://doi.org/10.1037/0096-1523.28.2.349

Gandour, J. (1983). Tone perception in Far Eastern languages. In *Journal of Phonetics*.

Gandour, J., & Harshman, R. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*(1), 1–33.

Garcia-Sierra, A., Diehl, R. L., & Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech Communication*, *51*(4), 369–378. https://doi.org/10.1016/j.specom.2008.11.005

García-Sierra, A., Ramírez-Esparza, N., Silva-Pereyra, J., Siard, J., & Champlin, C. A. (2012). Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study. *Brain and Language*, *121*(3), 194–205. https://doi.org/10.1016/j.bandl.2012.03.008

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics*, *66*(3), 363–376. https://doi.org/10.3758/BF03194885

Gibson, E. J. (1963). PERCEPTUAL LEARNING. *Annual Review of Psychology*, *14*(1), 29–56. www.annualreviews.org

Goldstone, R. (1994). Influences of Categorization on Perceptual Discrimination. *Journal of Experimental Psychology*, *123*(2), 178–200.

Goldstone, R., & Hendrickson, A. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69–78. https://doi.org/10.1002/wcs.026

Gonzales, K., Byers-Heinlein, K., & Lotto, A. J. (2019). How bilinguals perceive speech depends on which language they think they're hearing. *Cognition*, *182*, 318–330. https://doi.org/10.1016/j.cognition.2018.08.021

Gonzales, K., & Lotto, A. J. (2013). A Bafri, un Pafri: Bilinguals' Pseudoword Identifications Support Language-Specific Phonetic Systems. *Psychological Science*, *24*(11), 2135–2142. https://doi.org/10.1177/0956797613486485

Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual and Multicultural Development*, *6*, 467–477. http://www.informaworld.com

Grosjean, F. (1989). Neurolinguists, Beware! The Bilingual Is Not Two Monolinguals in One Person. *BRAIN AND LANGUAGE*, *36*, 3–15.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149. https://doi.org/10.1017/s136672899800025x

Grosjean, F. (1999). The bilingual's language modes. In *One Mind, two Languages: Bilingual language processing* (pp. 1–22).

Guenther, F. H., Husain, F. T., Cohen, M. A., Shinn-Cunningham, B. G., Lane, H., Pisoni, D., Beck, J., Beale, J., & Chiu, P. (1999). Effects of Categorization and Discrimination Training on Auditory Perceptual Space Stevens for helpful discussions regarding this work. *Journal of the Acoustical Society of America*, *106*, 2900–2912.

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. https://doi.org/10.1016/S0095-4470(03)00016-0

Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. In *LANGUAGE AND SPEECH* (Vol. 36, Issue 1).

Heeren, W. (2004). PERCEPTUAL DEVELOPMENT OF A NEW PHONEME CONTRAST BY ADULT AND 12-YEAR-OLD LISTENERS. *Proceeding from Sound to Sense*.

Heeren, W. F. L., & Schouten, M. E. H. (2010). Perceptual development of the Finnish /t-t:/ distinction in Dutch 12-year-old children: A training study. *Journal of Phonetics*, *38*(4), 594–603. https://doi.org/10.1016/j.wocn.2010.08.005

Heeren, W., & Schouten, M. (2008). Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories. *The Journal of the Acoustical Society of America*, *124*(4), 2291–2302. https://doi.org/10.1121/1.2967472

Hirst, Daniel., & Di Cristo, Albert. (1998). *Intonation systems : a survey of twenty languages*. Cambridge University Press.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, ich, Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*. https://doi.org/10.1016/S0

Johnson, K. (2007). Decisions and Mechanisms in Exemplar-based Phonology. In *Experimental Approaches to Phonology* (pp. 25–40).

Jusczyk, P. W. (1993). From general to language-specific capacities: the WRAPSA Model of how speech perception develops. In *Journal of Phonetics* (Vol. 21).

Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. In *Journal of Phonetics* (Vol. 2).

Kang, Y., & Han, S. (2013). Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study. *Lingua*, *134*, 62–74. https://doi.org/10.1016/j.lingua.2013.06.002

Kellogg, J., & Chang, C. B. (2023). Exploring the Onset of Phonetic Drift in Voice Onset Time Perception. *Languages*, *8*(1). https://doi.org/10.3390/languages8010078

Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*, *21*(6), 1092–1105. https://doi.org/10.1162/jocn.2009.21077

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161–168. https://doi.org/10.1016/j.cogbrainres.2005.05.004

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *Journal of Phonetics*, *21*, 125–139.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. K., & Miller, J. D. (1975). Speech Perception by the Chinchilla: Voiced-Voiceless Distinction in Alveolar Plosive Consonants. *Science (American Association for the Advancement of Science)*, *190*(4209), 69–72.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries, for the place feature in macaques. *The Journal of the Acoustical Society of America*, *73*(3), 1003–1010.

Kutlu, E., Chiu, S., & McMurray, B. (2022). Moving away from deficiency models: Gradiency in bilingual speech categorization. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1033825

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). THE DISCRIMINATION OF SPEECH SOUNDS WITHIN AND ACROSS PHONEME BOUNDARIES 1. *Journal of Experimental Psychology*, *54*(5).

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Liu, L., Chen, A., & Kager, R. (2017). Perception of tones in Mandarin and Dutch adult listeners. *Language and Linguistics*, *18*(4), 622–646. https://doi.org/10.1075/lali.18.4.03liu

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385–394. https://doi.org/10.1016/j.cognition.2014.06.004

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical Perception Effects Induced by Category Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 732–753.

Massaro, D. W., & Cohen, M. M. (1983). CATEGORICAL OR CONTINUOUS SPEECH PERCEPTION: A NEW TEST. In *Speech Communication* (Vol. 2).

McDonald, M., & Kaushanskaya, M. (2023). Bilingual Children Shift and Relax Second-Language Phoneme Categorization in Response to Accented L2 and Native L1 Speech Exposure. *Language and Speech*. https://doi.org/10.1177/00238309231176760

McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America*, *152*(6), 3819–3842. https://doi.org/10.1121/10.0016614

Miller, J. D., Dooling, R., Wier, C. C., Pastore, R. E., Kelly, W. J., & Doolingõ, R. J. (1976). Discrimination and labeling of noise buzz sequences with varying noise lead times: An example of categorical perception. *Article in The Journal of the Acoustical Society of America*. https://www.researchgate.net/publication/22160438

Mirman, D., Holt, L. L., & McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *The Journal of the Acoustical Society of America*, *116*(2), 1198–1207. https://doi.org/10.1121/1.1766020

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Livonen, A., Vainio, M., Alku, P., Iimoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.

Newport, E. L. (1990). Maturational Constraints on Language Learning. *Cognitive Science*, *14*(1), 11–28. https://doi.org/10.1207/s15516709cog1401_2

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Nosofsky, R. M. (2011). The generalized context model: an exemplar model of classification. In Andy J. Wills & Emmanuel M. Pothos (Eds.), *Formal Approaches in Categorization* (pp. 18–39). https://www.ebsco.com/terms-of-use

Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America*, *127*(2), EL54–EL59. https://doi.org/10.1121/1.3292286

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In *Typological studies in language* (Vol. 45, pp. 137–158).

Pike, E. V. (1994). Tonal Contrast on the Word Level. *WORD*, *45*(2), 173–176. https://doi.org/10.1080/00437956.1994.11435923

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*(2), 253–260. https://doi.org/10.3758/BF03214136

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition*, *3*, 7–18.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustic Society of America*, *55*(2), 328–333.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual Learning of Nonnative Speech Contrasts: Implications for Theories of Speech Perception. In *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121–166). https://www.ebsco.com/terms-of-use

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories*. In *Perception & Psychophysics* (Vol. 15, Issue 2).

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The Test of Everyday Attention (TEA)*. https://www.researchgate.net/publication/267552527

Schertz, J., Carbonell, K., & Lotto, A. J. (2020). Language specificity in phonetic cue weighting: Monolingual and bilingual perception of the stop voicing contrast in english and Spanish. *Phonetica*, *77*(3), 186–208. https://doi.org/10.1159/000497278

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. https://doi.org/10.1016/j.wocn.2015.07.003

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, and Psychophysics*, *78*(1), 355–367. https://doi.org/10.3758/s13414-015-0987-1

Schouten, M. E. H., & van Hessen, A. J. (1992). Modeling phoneme perception. I: Categorical perception. *The Journal of the Acoustical Society of America*, *92*(4), 1841–1855. https://doi.org/10.1121/1.403841

Sharma, A., & Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *The Journal of the Acoustical Society of America*, *107*(5), 2697–2703. https://doi.org/10.1121/1.428655

Shepard, R. N. (1957). STIMULUS AND RESPONSE GENERALIZATION: A STOCHASTIC MODEL RELATING GENERALIZATION TO DISTANCE IN PSYCHOLOGICAL SPACE. *Psychometrika*, *22*(4), 325–345.

Stagray, J. R., & Downs, D. (1993). DIFFERENTIAL SENSITIVITY FOR FREQUENCY AMONG SPEAKERS OF A TONE AND A NONTONE LANGUAGE / 使用声调语言和非声调语言为母语的人对声音频率的分辨能力. In *Journal of Chinese Linguistics* (Vol. 21, Issue 1). https://www.jstor.org/stable/23756129

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *3*, 37–149.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131–145.

Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *NeuroReport*, *19*(11), 1163–1167. https://doi.org/10.1097/WNR.0b013e3283088d31

Tice, M., & Woodley, M. (2012). Paguettes & bastries: Novice French learners show shifts in native phoneme boundaries. *LSA Annual Meeting Extended Abstracts*.

Van Hessen, A. J., & Schouten, M. E. H. (1999). Categorical perception as a function of stimulus quality. *Phonetica*, *56*(1–2), 56–72. https://doi.org/10.1159/000028441

Wang, W. S. (1976). Language Change. *Annals of the New York Academy of Sciences*, *280*(1), 61–72. https://doi.org/10.1111/j.1749-6632.1976.tb25472.x

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics*, *39*(4), 703–728. https://doi.org/10.1017/S0142716418000152

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, *66*, 173–196. https://doi.org/10.1146/annurev-psych-010814-015104

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*(6), 1866–1878. https://doi.org/10.1121/1.390988

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, *21*(4), 289–297. https://doi.org/10.3758/BF03199477

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R. J., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*, 638–642.

Wu, X., & Lin, H. (2008). Perception of Mandarin Tones by Mandarin and English Listeners. *Journal of Chinese Language and Computing*, *18*(4), 175–187.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, *120*(2), 1063–1074. https://doi.org/10.1121/1.2213572

Yazawa, K., Whang, J., Kondo, M., & Escudero, P. (2020). Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research*, *36*(4), 557–581. https://doi.org/10.1177/0267658319832645

Yip, M. (2002). Introduction. In *Tone* (pp. 1–16).

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *NeuroImage*, *26*(3), 703–720. https://doi.org/10.1016/j.neuroimage.2005.02.040