Title
Cue Integration and Contrast Shifts: Experimental and Typological Studies

Author
Yang, Meng

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Cue Integration and Contrast Shifts:

Experimental and Typological Studies

A dissertation submitted in the partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Linguistics

by

Meng Yang

2019

ABSTRACT OF THE DISSERTATION

Cue Integration and Contrast Shifts:

Experimental and Typological Studies

by

Meng Yang

Doctorate of Philosophy in Linguistics

University of California, Los Angeles, 2019

Professor Patricia Keating, Co-Chair

Professor Megha Sundara, Co-Chair

Auditory Enhancement has been put forth as an explanation for why certain acoustic phonetic cues co-vary to signal phonological contrasts more often than others. Under this account, listeners more readily associate two cues if they produce the same auditory effect, making the cues perceptually inseparable. Traditionally, evidence for enhancement has come from studies showing perceptual integration between enhancing cues, but even cues that do *not* share the same auditory effect have been shown to perceptually integrate. Further, language experience with co-variation between cues is often a confound in these studies.

In this dissertation, I present new evidence in favour of auditory enhancement from four experiments and one typological study.

In the first set of experiments, I use a modified cue weighting paradigm that mimics diachronic contrast shifts. Listeners categorizing synthesized speech stimuli were forced to shift

their attention between a pair of acoustic cues based on how informative each cue was to the contrast. This was done for a pair of enhancing cues, pitch and breathiness, and a pair of non-enhancing cues, pitch and vowel duration, both of which have been shown to perceptually integrate. For each pair of cues, I tested two groups of listeners – English listeners, who had no phonemic experience with either cue pair, and Hani (Tibeto-Burman) listeners who had experience with both pairs of cues co-varying in the same contrast. The extent to which listeners were able to shift attention between non-enhancing cues was predicted to reflect their language experience. For enhancing cues, attentional shift was predicted to also be conditioned by whether the cues were in an enhancing relationship. These predictions were borne out, but there was an unpredicted finding that shifting between the enhancing cues was asymmetric.

This asymmetry was further explored in two experiments. The first of these investigated whether the asymmetry could be caused by both listener groups having more linguistic experience with pitch than with breathiness. Two additional groups of listeners were thus tested using the same paradigm: Tone listeners, who used pitch phonemically, and Phonation listeners, who used breathiness phonemically. Both of these groups also exhibited the same asymmetry, showing that the phenomenon is language-general.

In the final experiment, I tested the hypothesis that the asymmetry in attentional shift was caused by an asymmetric perceptual dependency between pitch and breathiness. Listeners categorized stimuli for which one cue was informative but the other was completely neutralized. The amount of attention listeners paid to the uninformative cue was predicted to differ if the percept of one cue was dependent on the other but not vice versa. Results from this experiment provided weak evidence in favour of the hypothesis.

Finally, I conducted a cross-linguistic typological survey of the synchronic co-variation and diachronic contrast transfer between the cue pairs I tested experimentally. While the cues in both pairs co-vary synchronically, only the enhancing cues participate in contrast transfer. Furthermore, the transfer of phonological contrast between the enhancing cues occurs overwhelmingly in the direction that matches the asymmetry in attentional shift observed in the lab.

The experimental and typological studies in this dissertation provide support for Auditory Enhancement, demonstrating that cues that converge on the same auditory effect are treated differently by listeners compared to cues that do not. Based on the results, I argue that i) auditory enhancement and perceptual integration should remain separate notions, and ii) perceptual associations that are not learned through experience may be asymmetric, but learned associations are necessarily symmetric.

The dissertation of Meng Yang is approved.

Bruce Hayes

Jody Kreiman

Patricia Keating, Committee Co-Chair

Megha Sundara, Committee Co-Chair

University of California, Los Angeles

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I owe my gratitude to many people for their contributions to this dissertation. First and foremost, I am deeply grateful to Megha Sundara and Pat Keating, who were as different as advisors could be, but in their own ways, gave me the support I needed to pursue my research. Megha, thank you for your unforgiving honesty about my work. It made me a better thinker, presenter, and researcher. Pat, thank you for sharing your amazing wealth of knowledge about phonetics. I would not be the phonetician I am today without your mentorship.

I would also like to thank my committee members, Jody Kreiman and Bruce Hayes, who didn't hesitate to come along for the ride. Jody, your expertise with Voice Synthesis was invaluable, as were your words of encouragement, which you always offered with tea. Bruce, thank you for your curiosity about speech perception and your excitement about hats.

Thanks to Marc Brunelle, my M.A. advisor at the University of Ottawa, whose work has informed my interests from the very beginning. Thanks also to everyone who was curious enough to ask about my research and patient enough to listen. Your willingness to engage with me on my work made it all the more meaningful. I am also grateful to the Hani community in Nanuoxiang, Yunnan, China for sharing their language with me.

There were many more people who helped me to complete my dissertation in less direct but equally important ways. Thank you, Joshua Lai, for being my best friend and biggest fan. You were on this journey as much as I was, and your patience allowed me to endure. Thank you, Deborah Wong, for sharing your food, your car, and your love of ballet. Because of you, I was much less productive, but much happier. I am also eternally grateful to my family at CBCWLA for holding me up in prayer and for being the manifestation of God's love in my life. God is,

indeed, good! Finally, I am thankful to my parents, who supported me through every decision I made that they advised me against, including studying linguistics and getting a Ph.D.

This dissertation was supported by the UCLA Dissertation Year Fellowship and the NSF Doctoral Dissertation Improvement Grant [BCS-1823851].

VITA

2012	B.A., Linguistics and Language Studies York University, Glendon College
2013	M.A., Linguistics University of Ottawa
2016	Graduate Summer Research Fellowship University of California, Los Angeles
2017-2018	Mellon Foundation Pre-Dissertation Year Fellowship University of California, Los Angeles
2015-2018	Teaching Assistant, Associate, Fellow/Instructor University of California, Los Angeles
2018-2019	Dissertation Year Fellowship University of California, Los Angeles
2018-2019	Doctoral Dissertation Improvement Grant National Science Foundation

CHAPTER 1: INTRODUCTION

1.1. The multiplicity of acoustic phonetic cues to phonological contrasts

Phonological representations of speech sounds are minimal lists of feature values, where neighbouring sounds are distinguished by a single featural change. While these features often correspond to articulatory (e.g. [±ATR]), acoustic (e.g. [±strident]), or perceptual (e.g. [±sonorant]) characteristics of the sounds they describe, they are an abstraction from the phonetic information available to the listener as they encounter the physical speech signal.

Voiced and voiceless stops differ in their values for the feature [voice], typically to reflect the presence or absence of voicing during the stop closure respectively. But closure voicing is not a consistent cue to this contrast in some languages, and it is never the only cue to the contrast. In English for example, word-initial stops are typically phonetically voiceless, with aspiration of the voiceless series being the main cue that distinguishes it from the voiced series. Further, as many as 16 acoustic phonetic properties have been identified as co-varying in the production of the English stop voicing contrast (Lisker, 1986). These include properties of the closure (e.g. closure duration, intensity of voicing), properties of the preceding vowel (e.g. vowel duration, fundamental frequency (f0) and F1 frequency before the closure), and properties of the following vowel (e.g. f0 and F1 frequency after the closure). Another example is consonant place, which is identified not only by the transition of the formants into and out of the consonant (Delattre, Liberman, & Cooper, 1955) but also by the energy distribution in the burst spectrum (Blumstein & Stevens, 1979; Alwan, Jiang, & Chen, 2011). The tense-lax vowel contrast in English and other Germanic languages is cued by both vowel quality, itself comprising several different formant cues, and vowel duration. Tone contrasts which are characterized by differences in the level or contour of pitch, are typically signaled by additional cues such as

voice quality (e.g. Brunelle, 2009 on Northern Vietnamese; Kuang, 2013b on Miao; Yu & Lam, 2014 on Cantonese; Kuang, 2017 on Mandarin;), particularly in systems where there are many tonal categories crowding the space. In addition, the duration of the tone-bearing unit almost always varies systematically in tone contrasts between level and contour tones (Yu, 2010 and references therein) and between different level tones (see Faytak & Yu, 2011). In the languages of Southeast Asia, register contrasts (Henderson, 1952) are commonly attested. These contrasts typically employ a cluster of cues which may include pitch, voice quality, vowel duration, vowel quality, and even consonantal features such as voice onset time (VOT) (Brunelle & Kirby, 2016).

It should not be surprising, given the abundance of phonetic information, that listeners might attend to more than one cue for each contrast. It has been shown, for example, that Cantonese speakers who use pitch as the primary cue for lexical tone contrasts nevertheless listen for creak to distinguish the low-falling Tone 4 from other tones. Tone 4 is identified with more accuracy when accompanied by creak, and creaky stimuli were more likely to be identified as bearing Tone 4, especially when the duration of creak was longer (Yu & Lam, 2014).

Numerous studies have also shown that categorical boundaries along a primary cue dimension can be shifted by changing a secondary cue dimension. For example, English listeners' categorization of stop-consonant voicing along the VOT continuum can be shifted by manipulating the f0 at the onset of the following vowel. That is, raising the f0 onset biases listeners to identify stops as voiceless at shorter VOTs (e.g. Abramson & Lisker, 1985). It has also been shown that English listeners' categorization of /l/ and /ɹ/, primarily cued by the relative frequency of F2 and F3, can be shifted by changing the abruptness of the F1 transition from the liquid to the vowel, where a shorter, more abrupt transition biases listeners towards an /l/ response (Polka & Strange, 1985).

Attending to multiple, redundant cues to a single contrast has obvious advantages. When one or more cues are neutralized or obscured (e.g. in noise), listeners can still recover the contrast by shifting their attention to other cues that are available. This is most convincingly demonstrated by studies showing that even when the most important (i.e. primary) cue to a contrast is unavailable, listeners are still able to categorize speech sounds with higher-than-chance accuracy. Numerous studies on various languages have shown that neutralizing pitch information for tone contrasts does not render listeners incapable of correctly identifying underlying tone categories (e.g. Abramson, 1962 using Thai; Jenson, 1958 using Swedish and Norwegian; Whalen & Xu, 1992 using Mandarin; Liu & Samuel, 2004 using Mandarin; Gao & Hallé, 2013 using Shanghainese). The studies on Mandarin (e.g. Blicher, Diehl, & Cohen, 1990; Whalen & Xu, 1992) and Shanghainese (Gao & Hallé, 2013) suggest that these listeners are using duration information when pitch is not available.

In sum, there is ample research showing that cues co-vary in speech production and that listeners attend to multiple acoustic dimensions cuing the same categorical difference. Attending to cue co-variation allows listeners to categorize speech sounds even when some cues are unavailable.

1.2. Why do listeners perceptually associate cues?

Listeners may come to associate co-varying cues such that their perception of one cue dimension becomes dependent on variations along the other. Three types of explanations have been put forth to explain why this occurs: *associative learning* accounts (e.g. Holt & Lotto, 2006), *articulatory* accounts (e.g. Kohler, 1984), and *auditory enhancement* accounts (e.g. Kingston & Diehl, 1994), These are discussed in turn.

1.2.1. Associative Learning Account

According to the *associative learning* account, listeners learn to couple cues if they co-vary reliably in the input signal. In support of this view are studies showing that listeners learn to weight cues in a multi-dimensional acoustic space (e.g. Holt & Lotto, 2006). The relative weighting of cues to categories can be acquired through unsupervised learning processes (Toscano & McMurray, 2010) and successful learning of categories is dependent on the distributional properties of cues (e.g. Holt & Lotto, 2006). That is, listeners assign perceptual weight (i.e. attend) to a cue to a contrast if the categories of that contrast are statistically differentiated along that acoustic dimension. In a multi-dimensional space, the most informative cue will get the highest weight, and other cues will also receive weight if they convey the same categorical distinctions. Under this account, listeners should learn to weight and therefore associate *any* cues that co-vary to signal a contrast, and they should be able to learn both positive and negative co-variations of cues. To demonstrate this, Holt et al. (2001) trained Japanese quail on stimuli in which VOT and vowel-initial f0 had a positive correlation, as in most languages with a voicing contrast, or on stimuli in which the two cues had a negative correlation, or on stimuli in which the two cues were uncorrelated. Birds trained in the positive correlation condition showed a bias toward the voiced category when f0 was low and a bias toward the voiceless category when f0 was high. Conversely, birds trained in the negative correlation condition showed a bias in the opposite direction. Those trained on the uncorrelated condition showed no bias. Thus, both correlations can be learned through exposure to co-variation.

Though it is virtually uncontested that listeners learn speech patterns through experience, the cue co-variation patterns that emerge cross-linguistically nevertheless suggest that pure

learning accounts may be inadequate. In particular, certain cues (e.g. VOT and f0) co-vary more frequently and consistently across languages than would be predicted if all cue co-variations were equally learnable. Diachronically, some cue pairs also are more likely to undergo contrast transfer, a process whereby a phonological contrast primarily signaled by one cue is shifted onto another co-varying cue (Kingston, 2011). Together, the cross-linguistic patterns suggest that some cues have a special relationship, and are preferentially combined to signal category differences in speech. This preference is likely phonetically grounded, rooted in either articulation or in perception. These are discussed below.

1.2.2. Articulatory Account

The most straightforward explanation for why listeners associate some cue pairs is because these cues, and not others, are controlled by the same articulatory mechanisms. Listeners are aware that producing changes in one cue dimension necessarily changes the other, thus they couple them in perception as well.

The co-variation between vowel-initial f0 and voicing on the preceding stop consonant is often attributed to articulation. Cross-linguistically, f0 tends to be lower following a voiced stop and higher following a voiceless stop. From an aerodynamic perspective, transglottal airflow is slowed during the closure period of a voiced stop as oral pressure increases and subglottal pressure decreases. Thus, f0 is low at the time of release, and rises as transglottal airflow returns to normal after the release. After a voiceless (aspirated stop), transglottal airflow is faster, thus increasing the rate of vocal fold vibrations and raising pitch initially before returning to normal (Hombert et al., 1979 and references therein). Another articulatory account supposes that differences in vocal fold tension in the production of voiced and voiceless consonants is

responsible for the differences in f0 on following vowels. In order to produce voicing during the closure of an oral stop, the larynx can be lowered to increase the size of the oral cavity and maintain transglottal airflow. This action of lowering the larynx also has the effect of relaxing vocal fold tension and slowing the rate of vocal fold vibrations (Hombert et al., 1979; Moisik et al., 2019 and references therein).

While articulatory accounts could explain why certain cues co-vary across languages, they are not without their limitations. First, the articulatory account predicts that cues should co-vary continuously, rather than categorically. In the case of f0 as a cue to voicing, the onset f0 of a vowel should depend on the degree of voicing on the previous consonant. However, evidence suggests that the phonological category of the consonant, rather than its phonetic properties, determines f0 onset of the following vowel (Keating, 1984; Kingston & Diehl, 1994; Kingston, 2007). For example, in a study comparing Spanish and English, Dmitrieva et al. (2015) found that vowel-initial f0 was significantly different after Spanish voiced stops and voiceless unaspirated stops, which are separate phonological categories. However, after English voiced and voiceless unaspirated stops, which pattern together as [+voice], there was no significant difference in vowel-initial f0. Rather, for English speakers there was a significant difference after voiceless unaspirated and voiceless aspirated stops, which are separate phonemic categories. In this case, articulation of voicing clearly does not motivate the differences in vowel f0 following stop consonants.

Other evidence against a purely articulatory account comes from studies on the perceptual integration of co-varying cues. Before discussing the findings and implications of these studies, I will first explain the basis for this line of research. The idea is that listeners can perceptually associate, or *integrate*, co-varying cues to varying degrees such that their pairing

increases the perceptual distance between the categories they signal. Of course, pairing the cues such that their correlation reverses their pattern of co-variation can have the opposite effect. Perceptual integration has been traditionally tested using the Garner paradigm (Garner, 1974) which measures the perceptual dependence between two cues. Four stimuli are created in a two-dimensional acoustic space in which the dimensions are equated psychoacoustically (e.g. using just-noticeable differences). This is schematized in Figure 1.1.

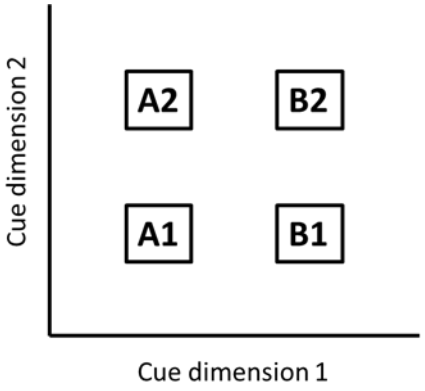

Figure 1.1. Stimuli distribution from Garner paradigm (Garner, 1974)

Listeners classify pairs of stimuli from the distribution, either A1 vs. B1, A2 vs. B2, A1 vs. A2, B1 vs. B2, A1 vs. B2, or A2 vs. B1. The pairs A1 vs. B1 and A2 vs. B2 test listeners' discrimination of categories varying along just Dimension 1. Similarly, the pairs A1 vs. A2 and B1 vs. B2 test listeners' discrimination of categories varying along just Dimension 2. The pairs A1 vs. B2 and A2 vs. B1 test listeners' discrimination of categories when the cues are positively or negatively co-varied respectively. Reaction times or d-prime measures are compared, where faster reaction time and higher d-prime scores for a correlated pair compared to the single-dimension pairs indicates ease of categorization and greater perceptual distance when the cue measures are co-varied in a certain way. Conversely, slower reaction times and lower d-prime scores for the opposite correlation of cues indicates that cues paired this way cause perceptual

interference. If either an advantage or interference is observed, the cues are considered to be integral.

Studies on perceptual integration using this paradigm have shown that there are mismatches between the articulatory contingencies of a set of cues and the way they are perceived by listeners. In a study on listener's perception of English voicing cues, Kingston et al. (2008) found that only a subset of the cues that co-vary with voicing were integrated. Both f0 and F1 on the following vowel perceptually integrated with voicing during the stop closure, but not with each other, and none of these cues integrated with stop closure duration. The non-integration of voicing cues to closure duration may be surprising from an articulatory standpoint given that the difference in closure duration between voiced and voiceless segments is also motivated by constraints on the production of voicing. Namely, since voicing is more difficult to maintain during the closure of an oral stop, voiced stops typically have shorter closures than voiceless stops. If the pairing of cues is solely a matter of having articulatory dependencies, then closure duration should be equally likely to integrate with voicing.

Furthermore, cues that are articulatorily independent have also been shown to co-vary and perceptually integrate. Notably nasalization, produced by lowering the velum, F1, controlled by tongue height, and breathy phonation, controlled at the larynx, co-vary phonetically and phonemically across languages. Vowel height has been shown to co-vary with nasalization in English (Carignan et al., 2011), Portuguese (Shosted et al., 2015), and Hindi (Shosted et al., 2012). These two cues have also been shown to perceptually integrate (Kingston, 1992; Kingston & Macmillan, 1995). Breathiness, a voice quality produced with a wide glottal aperture, has also been shown to co-vary with nasalization in three Loloish languages (Garellek, Ritchart, & Kuang, 2016). Finally, all three cues – nasalization, f1, and breathiness – have been shown to co-

vary in Southern French (Carignan, 2017). The language-general nature of the co-variation between these cues begs for a phonetic explanation, but one that is not based in articulation, since the cues in question do not have any obvious articulatory contingencies.

1.2.3. Auditory Enhancement Account

In response, it has been proposed that cues co-vary and integrate if they reinforce a single auditory effect (Diehl & Kluender, 1989; Kingston & Diehl, 1994; Diehl et al., 1995; Diehl & Molis, 1995). For example, Kingston and Diehl (1994) propose that consonant voicing co-varies with f0 because both voicing during the closure and low f0 on adjacent vowels contribute to the auditory effect of there being low-frequency energy in or near the closure. If the motivation for f0 to co-vary with voicing is perceptual, then the fact that it patterns with phonological categories is expected. Low f0 can be expected near phonologically voiced obstruents, whether the obstruent is phonetically voiced or not. We might then also expect that amongst cues that co-vary, those that contribute to the same auditory effect are more easily integrated by the listeners, as was found for English cues to voicing (Kingston et al., 2008). And of course, if the privileged status of cue pairs is rooted in perception and not in articulation, then we should also observe that cues that are articulatorily independent co-vary if they converge on the same auditory effect. Such is the case for nasalization, breathiness, and F1. Nasalization and breathy phonation both strengthen the percept of a stronger first harmonic (H1) compared to a weaker first formant (Maeda, 1993; Bickley, 1982), and degree of nasalization and tongue height both serve to raise or lower F1 (e.g. Maeda, 1993).

The perceptual account is also consistent with results from several studies showing that even acoustic dimensions that are not detectable or not present in a contrast can be used as cues

to categorical distinctions. In Cantonese, where pitch information is important for tone identification, perturbations of vowel f0 from neighbouring consonants are shorter than the duration necessary for listeners to actually detect the pitch change. Nevertheless, when the f0 on the following vowel was extended to a detectable duration, Cantonese listeners were able to use it as a cue to consonant voicing (Francis et al., 2006). Francis et al. took this to mean that listeners were associating f0 to the voicing contrast, even though they had no perceptible experience with the co-variation between f0 and voicing. In a similar vein, Lee and Katz (2016) show that some cues can be perceptually integrated without co-varying in the signal. In their study on the Korean plain-fortis contrast in fricative consonants, they show that listeners integrate voicing during frication with f0 on the following vowel, even though f0 is not actually a correlate in this contrast. Listeners' association of a cue to a contrast in the absence of experience with its co-variation can be explained if the basis for this association is the auditory percept to which both cues contribute.

Kingston and colleagues use *enhancing* as a specialized term to refer to cues such as those described above that reinforce a single auditory effect. They argue that enhancing cues are represented by a higher-level auditory unit, an integrated perceptual property (IPP), mediating between individual acoustic correlates and the features they distinguish (e.g. [voice]), and making them perceptually inseparable. Under this view of enhancement, the enhancing property of cue pairs is language-general, and influences what cues are likely to be used to signal a given contrast across all languages.

The notion of enhancement as defined by Kingston and colleagues is similar to featural enhancement (e.g. Stevens & Keyser, 1989) or gestural enhancement (e.g. Stevens & Keyser, 2010) where either a secondary phonological feature or a secondary articulatory gesture,

respectively, are coupled with a primary distinctive feature or gesture to increasing the acoustic distance between two sounds. However, Stevens and Keyser are more concerned with the outcome of the co-variation, but not necessarily with why two features or gestures might co-vary. In their view, enhancement is language-specific. Thus, between languages, different secondary features or gestures could be recruited to achieve the outcome of increasing the robustness of the same contrast.

In this dissertation, I will use the term *enhancing* in the Kingstonian sense – for cue pairs that converge on a auditory effect. The evidence I have cited in favour of *auditory enhancement* as the explanation for why cues co-vary is not without issues. The first is that even the most conservative results are confounded with language experience. That is, evidence for perceptual advantages with cue pairings are usually tested on listeners for whom these cues already co-vary to some degree. It thus becomes difficult to distinguish the effect of experience with co-variation from enhancement effects, which should be language-general and independent of experience. Second, the strongest evidence for enhancement between two cues comes from experiments showing that they are perceptually integrated, as we saw above. This is largely to demonstrate that enhancing cues are not perceptually separable, as predicted by a model in which these cues form a single IPP. However, there is also strong evidence that cues that do not contribute to the same auditory effect also co-vary cross-linguistically and are perceptually integrated by listeners. Thus, the uniqueness of so-called *enhancing cues* is called into question.

1.3. This dissertation

This dissertation aims to address both of the issues above. Teasing apart enhancement from language experience will be the primary purpose of Chapter 2, in which I present

perception data obtained from a novel paradigm that provides a better control for listeners' experience with co-varying cues. In Chapter 3 I draw a distinction between enhancement and perceptual integration by presenting evidence differentiating enhancing cues – those that converge on a single auditory effect – from those that do not but nevertheless integrate perceptually. In Chapter 4, I explore the apparent asymmetric relationship between enhancing cues found in Chapter 2. In Chapter 5, I show that the experimental patterns observed in Chapters 2-4 are observable in the ways cues co-varying cross-linguistically and in the way that they participant in sound change. Finally, in Chapter 6, I discuss the implications of these findings and future directions for this research.

CHAPTER 2: ENHANCING CUES

2.1. Introduction

Auditory Enhancement (Kingston & Diehl, 1994) has been proposed to account for the prevalence of the co-variation between some cue pairs in signalling contrasts across the world's languages. This is the idea that some cues have a special relationship because they contribute to the same auditory effect and are jointly represented at a higher-level, abstract perceptual node referred to as an integrated perceptual property (e.g. Diehl & Kluender, 1989; Kingston & Diehl, 1994; Diehl et al., 1995; Diehl & Molis, 1995). When one of these cues is used for a contrast, other cues that produce the same auditory effect are preferentially recruited to enhance the percept of differences between the categories being distinguished.

In Chapter 1, I reviewed two pieces of evidence that enhancement is independent of language experience. The first is that English listeners who have experience with the co-variation of many acoustic cues in the stop voicing contrast nevertheless only integrate a subset of those cue pairs to the exclusion of others (Kingston et al., 2008). The second is that Korean listeners who do not have experience with the co-variation between voicing during frication and f0 on the following vowel in the contrast between plain and fortis fricatives nevertheless integrate these two cues in distinguishing between these two sound classes (Lee & Katz, 2016). These studies suggest that experience with co-variation is neither sufficient nor necessary for listeners to perceptually associate two enhancing cues to a contrast. Unfortunately, both of these studies fail to completely remove language experience as a confound. What Kingston et al.'s results do not rule out is that this asymmetry could be due to the differing extent to which English listeners have experience with different cue pairs in the stop voicing contrast. Lee and Katz (2016) themselves point out that the lenis-fortis contrast in Korean initial *stops* is cued by an f0

difference on the following vowel (Cho et al., 2002). Thus, the effect observed in fricatives described above may well be due to generalization from learned co-variation in another series of consonants.

The primary goal of this chapter is to provide further evidence that enhancement is independent of language experience obtained using a cue weighting paradigm that allows me to experimentally equalize listeners' experience with the cues selected. The basis for this paradigm will be explained in Section 2.1.1.

2.1.1. Cue weighting of multidimensional stimuli

Given multidimensional stimuli, listeners rely on some cues more than others, a phenomenon referred to as *cue weighting* (e.g. Holt and Lotto, 2006, Mayo et al., 2011). To relate this to more familiar concepts, the cue that receives the highest weight is the primary cue, whereas cues that are weighted less are secondary. For example, for English listeners, the VOT cue for consonant voicing receives the most weight, making it the primary cue, while initial f0 on the following vowel receives less weight, making it a secondary cue (Abramson and Lisker, 1985, Gordon et al., 1993, Lisker, 1978, Whalen et al., 1993).

Listeners might come to rely more on one cue than another based on their language experience. For instance, listeners are more likely to attend to cues that have a wider range of values compared to those that have a narrower range in the input (Lutfi, 1993). Further, listeners assign higher cue weights to more distinctive cues, that is, cues with less distributional overlap between tokens belonging to distinct categories, compared to less distinctive cues (Holt & Lotto, 2006). Relatedly, in categorization tasks, listeners respond more confidently and show a sharper response curve when there is less within-category variance (Clayards, et al., 2008). If attention to

cues has a direct relationship with the distributional informativeness of that cue to categories, then we expect the most distinctive acoustic correlate to be the primary cue, and the less distinctive correlates to be secondary cues.

Secondary cues can play a more crucial role in categorization when the primary cue is obscured. As discussed briefly in Chapter 1, Mandarin listeners are able to categorize tones using duration and phonation cues when pitch information is removed from portions of the stimuli (Liu & Samuel, 2004). Similarly, Alwan, Jiang, and Chen (2011) demonstrate that secondary cues to the perception of labial/alveolar distinctions in English (e.g. F1 and F2 onset frequencies, F2 and F3 frequency changes) become increasingly important as the signal to noise ratio reduces. Such studies indicate that cues are re-weighted as a result of changes in informativeness in the speech signal.

Listeners can also re-weight cues as a result of experience with a second language. For example, Japanese listeners are known to have difficulty distinguishing between English /l/ and /ɹ/ (e.g. Goto, 1971; Miyawaki et al., 1975) because they attend to F2 frequency cues, which are unreliable for this contrast, rather than F3 frequency cues, which are reliable and well-attended to by native English listeners (Iverson et al., 2003). However, their ability to distinguish English /l/ and /ɹ/ can be improved with exposure to synthesized (Iverson, Hazan, & Bannister, 2005) and natural stimuli (e.g. Hazan et al., 2005; Logan et al., 1991; Lively et al., 1993; Bradlow et al., 1999) with reduced F3 variability, but high F2 variability. The change in variability causes listeners to up-weight F3 and down-weight F2 as they learn that one cue is more informative than the other.

Further, changes in cue weights are proportional to changes in the signal. Consistent with this idea, not only do listeners down-weight reliance on a secondary cue when they hear

"accented" speech with atypical cue relations (Idemaru & Holt, 2011, Idemaru & Holt, 2014, Liu & Holt, 2015), but the extent of cue down-weighting bears a linear relationship to the proportion of accented speech they hear (Lehet & Holt, 2016).

In sum, there is ample evidence that listeners learn to assign and/or alter cue weights for category learning when they are exposed to co-variation between cues in the input. I thus used cue weighting as a tool to probe whether listeners draw inferences about enhancing cue pairs even when they are not supported by input distributions. To this end, I chose a pair of enhancing cues, pitch and breathiness, and, using a modified cue weighting paradigm, tested experienced (Hani – Loloish, Tibeto-Burman) and inexperienced (English) listeners' ability to shift weights between the cues. The cues will be described in Section 2.1.2., and the two language groups in Section 2.1.3., followed by the predictions for the results in Section 2.1.4.

2.1.2. Pitch and breathiness as enhancing cues

Pitch and breathiness were selected as the cue pair in these experiments. Pitch here refers to the percept of change in fundamental frequency (f0). Breathiness is a voice quality with many acoustic correlates including H1-H2 (Gordon & Ladefoged, 2001; Keating et al., 2011, 2012), H1-A1, H1-A2, H1-A3 (Klatt & Klatt, 1990; Blankenship, 2002; Esposito, 2012; DiCanio, 2009)[1], Cepstral Peak Prominence (CPP) (Hillenbrand et al., 1994; Blankenship, 2002; Garellek & Keating, 2011), and Harmonics-to-Noise Ratio (HNR) (Garellek et al., 2012). In these experiments, the only measure manipulated for breathiness was source H1-H2, the difference between the amplitude of the first and second harmonics in the voice source, where a larger difference in amplitudes corresponds to more breathiness, and a smaller difference between

[1] Amplitudes of harmonics (H1, H2) and formants (A1, A2, A3) are adjusted post-VoiceSauce (Shue et al., 2011).

amplitudes corresponds to less breathiness (modal voice). H1-H2 is one of the most common cues used for signalling phonation differences cross-linguistically (Chen, 2011 and citations therein), and it has been demonstrated that manipulation of this measure alone is sufficient for listeners to hear changes in breathiness (Kreiman & Gerratt, 2010).

These cues were chosen because Kingston (2011) claims this cue pair is enhancing. Since breathiness is characterized by a strong first harmonic and weaker higher harmonics, breathy voice strengthens the percept of low frequency energy. Low pitch, being characterized by a low f0, also strengthens the percept of low frequency energy. Thus, breathy voice quality and low pitch have the same auditory effect for the listener. In other words, a negative relation between these cues is enhancing: lower pitch (low f0) is coupled with more breathiness (high H1-H2), and higher pitch (high f0) is coupled with less breathiness (low H1-H2). The positive relation between these cues is non-enhancing.

Evidence of the perceptual integration between breathiness and pitch comes from studies that have shown that a) listeners' perception of spectral slope is affected by changes in pitch (Li & Pastore, 1995), b) listeners' perception of pitch is affected by changes in spectral shape (Silverman, 2003; Kuang & Liberman, 2015), and c) pitch and voice quality show Garner interference (Brunelle, 2012).

Of course, since breathiness and pitch are both laryngeal cues, they also have articulatory contingencies, though the prediction about the way in which the cues should be related is less clear. On one hand, larynx lowering is common to the production of breathy voice (Henderson, 1952; Gregerson, 1976; Hirano, 1981; Thongkum, 1991) and lower pitch (Ohala, 1972; Ohala & Ewan, 1973; Ohala, 1978). Also, less activation of the lateral cricoarytenoid muscles (LCA), which leaves a longitudinal gap between the vocal folds to generate breathiness, simultaneously

reduces the tenseness of the vocal folds, causing the vocal folds to vibrate at a slower rate (Hombert, 1978). This suggests that listeners should associate breathy voice with lower pitch and vice versa. On the other hand, increasing subglottal pressure to allow more air to pass through the glottis produces more breathiness. But if subglottal pressure is increased sufficiently, say as a compensatory response to the rapid drop in pressure due to the wider glottal aperture, it could also have a pitch raising effect (Silverman, 1997). If this occurs, then breathiness would be associated instead with higher pitch. There is some evidence that both associations are possible. Diachronically, breathy phonation has become realized in tone systems as the low tone in a great number of languages (see Chapter 5 for fuller discussion) and as the high tone in other languages, possibly Jeh (Gradin, 1966) and Quiotepec Chinantec (Robbins, 1968). Synchronically, though breathiness and pitch are not phonemically contrastive in English (see Section 2.1.3), at least two studies have found a positive correlation between f0 and H1*-H2* in English vowels (Kreiman, Gerratt, & Antonanzas-Barroso, 2007; Iseli, Shue, & Alwan, 2007).

Thus, while the perceptual account makes clear predictions that listeners should associate lower pitch with breathy voice and vice versa, the predictions made by the articulatory account are less straightforward.

2.1.3. English and Hani listeners' experience with pitch and breathiness

English listeners were chosen as the group without experience with the cue pair. There is no segmental contrast in English to which both pitch and breathiness are cues. There are also no contrasts to which pitch or breathiness are cues individually. That is, neither tone nor phonation are contrastive in English.

Hani listeners were chosen for having experience with pitch and breathiness as a cue pair. Maddieson and Ladefoged (1985) recorded Hani speakers in Mojiang, Yunnan, China, and found that in this dialect, the tense-lax contrast is realized by differences in pitch, voice quality, and vowel duration[2] (among other dimensions). Specifically, lax syllables have lower pitch, a larger difference between the first two harmonics (H1-H2), and longer vowel duration. This is summarized in Table 2.1. below.

	Pitch	Vowel Duration	Breathiness
Lax	Lower	Longer	Breathier
Tense	Higher	Shorter	Less breathy

Table 2.1. Hani tense-lax contrast and its correlates (adapted from Maddieson & Ladefoged 1985)

Given the way in which these measures map onto the two register categories, lower pitch (low f0) co-varies with greater breathiness (high H1-H2) and higher pitch (high f0) co-varies with less breathiness (low H1-H2). This is congruent with the enhancing relation described above. While Mojiang Hani shows this pattern, the same was not found for Lüchun Hani (Kuang & Keating, 2012) spoken approximately 150km Southeast of Mojiang.

For practical reasons in the field, my Hani perception data was collected in Nanuoxiang, Yunnan, China, a region different from the two above. The three locations are given in the map in Figure 2.1. The lightly shaded region on the map is Yunnan province. The three locations for data collection are marked with red stars.

[2] Vowel duration will become relevant in Chapter 3, but will be described here as well as it also co-varies with pitch and phonation in the tense-lax distinction.

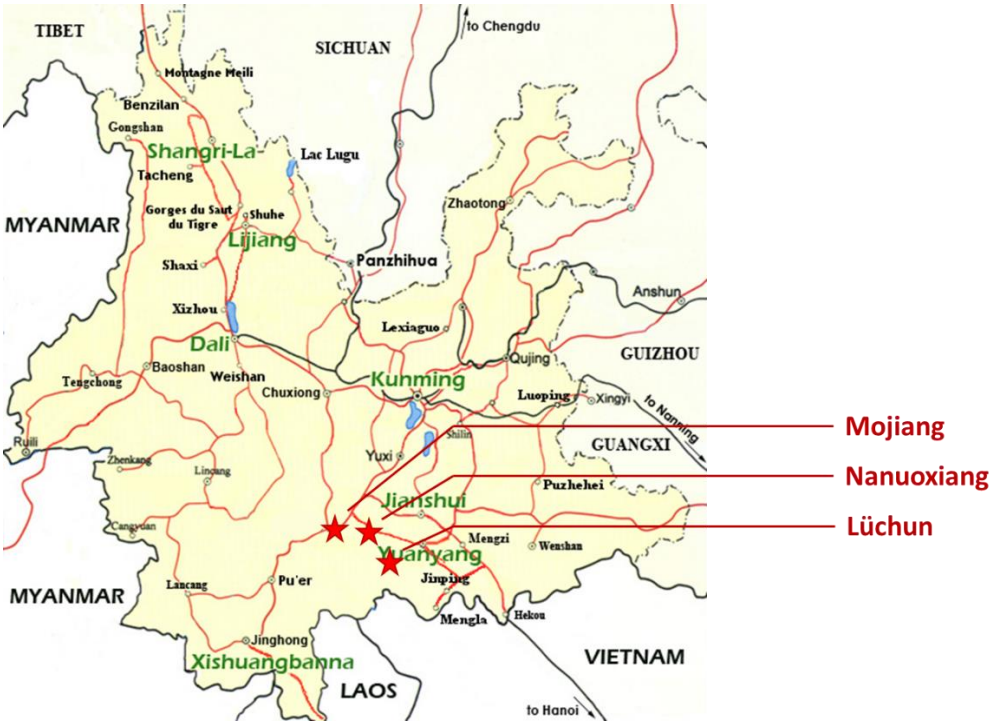

Figure 2.1. Locations for data collection: Mojiang (Maddieson & Ladefoged, 1985), Nanuoxiang (present study), Lüchun (Kuang & Keating, 2012).

Given that there are discrepancies in what cues co-vary to signal the tense-lax contrast between different dialect regions, it was necessary to determine whether the cues relevant to my study – pitch, breathiness, and vowel duration – were being used by my participants to distinguish the tense-lax contrast. I therefore collected production data from 11 of my participants who also did the perception experiment. All were male speakers between 25-60 years old who were employees at the Nanuoxiang government office. They reported using Hani for most of their communication both at home and at work. 10 of the speakers were from Nanuoxiang, though not from the same village. One speaker, we later found, was from the neighbouring area, Yangjiexiang, and so will not be included in the discussion below.

The elicitation list consisted of 33 monosyllables that formed minimal sets contrasting the tense and lax categories. However, I was only able to compare data from a small number of minimal pairs for the following reasons: Four of the speakers, who were all under 30 years old,

produced no difference between the tense and lax categories, often commenting that the words being elicited “sound the same” (i.e. were homophonous). This suggests that younger speakers in this region are merging/have merged the tense-lax contrast. Three of the remaining speakers had difficulty producing the monosyllables out of context, and instead produced them in multisyllabic phrases, which were unusable for my purposes. The discussion that follows thus centers on the production data collected from three participants (M1, M6, and M11) who clearly had a tense-lax contrast and were comfortable producing the elicitation items in isolation. However, due to disagreements on the citation tones of some of these words, only six pairs of words were considered in the comparison between the tense and lax registers. These are given in Table 2.2.

Segments and Tone	Lax	Tense
/ba31/	白 ‘white’	坏 ‘broken’
/ba55/	薄 ‘thin’	扛 ‘to carry’
/ka31/	种 ‘to plant’	坝田 ‘to flatten land’
/na55/	停 ‘to stop’	深 ‘deep’
/tse55/	冷 ‘cold’	犁地 ‘to plow’
/de55/	按 ‘to press’	活 ‘to live’

Table 2.2. Minimal pairs in which Speakers M1, M6, and M11 agreed on underlying tones. Glosses in the Lax and Tense columns are given first in Chinese characters, then in English.

The vowel portion of all target tokens from the recordings were segmented in Praat (Boesma & Weenink, 2016) and measurements of pitch (f0 using the Straight algorithm) and breathiness (H1*-H2*) were obtained using VoiceSauce (Shue et al., 2011). F0 and H1*-H2* values averaged across the syllable were compared. Vowel duration was obtained by subtracting the start time of each segment from its end time. Given the small number of data points, the following discussion will primarily be descriptive.

Overall, all three cues were used to distinguish between the tense and lax categories, though to differing degrees by different speakers. Speaker M1 produced a pitch difference of greater than 14 Hz for two minimal pairs (f0 averaged across the whole syllable), and less than 8 Hz for the remaining pairs. Speaker M6 produced a pitch difference of greater than 8 Hz for all pairs except one. Speaker M11 did not seem to use pitch at all, with most pairs having less than a 3 Hz difference. H1*-H2* was higher for the lax token in all but four pairs across the three speakers. The magnitude of the difference ranged between 0.39 dB and 3.75 dB, with most pairs differing by more than 1.5 dB. Vowel duration was the least consistent cue across speakers. In 2/3 of the elicited pairs across speakers, duration was longer in the lax syllable, typically by more than 10% of the duration of the tense syllable. In the remaining one third of the elicited pairs, duration was longer in the tense syllable than in the lax syllable.

From this coarse description of the small amount of production data, it's clear that there is a non-negligible amount of variation in the use of the three cues. More extensive work would need to be done to see whether there is less variability within speakers from the same village and the same speaker generation. However, it was also very evident that speakers from different villages in Nanuoxiang were in frequent contact with each other. Thus, as listeners, they were exposed to linguistic input in which all three cues play a role in distinguishing this contrast, even if they themselves may not make use of all three cues equally.

2.1.4. Experimental design and predictions

Listeners' experience with these cues was controlled in two ways. First, as discussed in the previous section, listeners were selected for these studies based on their language background. Pitch and breathiness are not linguistically relevant for English listeners, while they

signal a phonemic contrast for Hani listeners. Experience was also experimentally controlled with the distribution of stimuli across these two cues in a cue weighting paradigm. Listeners were first presented with a distribution of stimuli that biased them to weight one cue higher, then they were given a different distribution to induce a shift in attention to the other cue. The relation between category labels in the initial learning phase and the shift phase was manipulated such that the relationship between pitch and breathiness was either enhancing or non-enhancing.

Recall that Hani listeners are experienced with the enhancing relation between pitch and breathiness in the tense-lax contrast. Thus, we expect Hani listeners to have difficulty shifting their attention from one cue to another if the experimental mapping runs counter to their experience (non-enhancing mapping). If English listeners also exhibit the same difficulties when the mapping reverses the enhancing relation, then this could not be attributed to their experience, but rather to listeners perceptual association enhancing cues.

2.2. Methods

2.2.1. Participants

For English listeners, 150 undergraduate participants (age 18-31) were recruited from the Subject Pool at UCLA. Four subjects were excluded for having experience with languages that have a phonation or tone contrast. The remaining subjects were native speakers of English and had no experience with such languages, as self-reported on a Language Background form. Nine additional subjects did not complete the study and were also excluded for having incomplete data.

Hani listeners for these experiments were recruited in the same way as for the pitch and vowel duration experiments. 103 participants were recruited, 79 from the local middle school in

Nanuoxiang (age 9-16) and 24 others from the area (age 18-64). All Hani listeners recruited also speak at least one variety of Mandarin as self-reported on a Language Background form. Nine participants were excluded for being less than 12 years old, 15 were excluded for using Hani less than 50% of the time, and two were excluded because their data was incomplete.

2.2.2. Stimuli

All stimuli were the syllable [tɑ] with a specific breathiness and pitch value on the vowel. In this section, I first describe the method for scaling these two cues so that they were matched to be equally discriminable to English listeners. Then I describe the distribution of stimuli within the acoustic space, as well as how they were synthesized.

2.2.2.1. Perceptual scaling

The acoustic measure used to manipulate Pitch was fundamental frequency (f0) in Hertz (Hz). The Pitch scale ranged from 96 Hz to 126 Hz. This 30 Hz range was also set at 10 times the JND for English listeners, that is, approximately 3 Hz. This pitch range is within the normal range for the human male voice. Pitch was scaled using Hertz despite JND for pitch being typically measured using psychoacoustic scales (i.e. 3 mel for modal voice, Kollmeier et al., 2008) for practical reasons relating to speech synthesis. The synthesis program used to generate the stimuli only produces whole-number Hertz values, making it impossible to generate equally spaced f0 values converted from mels to Hertz. The decision to use the acoustic scale also seemed appropriate given that the relationship between Hertz and mels is essentially linear below 500 Hz (Stevens et al., 1937).

The acoustic parameter used to control Breathiness was (source spectrum) H1-H2, the amplitude of the first harmonic minus the amplitude of the second harmonic (e.g. Fischer-Jorgensen, 1967; Gordon & Ladefoged, 2001; Garellek et al., 2016). To manipulate the difference in amplitude between H1 and H2, H2 was held constant while the H1 value was adjusted. The H1-H2 values ranged from -3.67 to 33.03 dB. This range of 36.7 dB was set at 10 times the just-noticeable difference (JND) of this measure for English listeners, that is, 3.67 dB (Kreiman & Gerratt, 2010), to match the range for pitch. The minimum H1-H2 used in the experiment corresponds to the lower bound for modal voice and the maximum H1-H2 corresponds to the upper bound for breathy voice. While the overall range is larger than what is typically employed by speakers (see Garellek et al., 2016), two trained phoneticians verified that it was within a reasonable range for this cue given auditory impressions of the stimuli.

2.2.2.2. Stimulus distribution

The experimental paradigm, adapted from Holt & Lotto (2006), involved two sets of training stimuli and a set of test stimuli. Each set of training stimuli was synthesized to contain 86 unique tokens varying in the two-dimensional space delineated by Pitch (f0) and Breathiness (H1-H2), as shown in Figure 2.2.

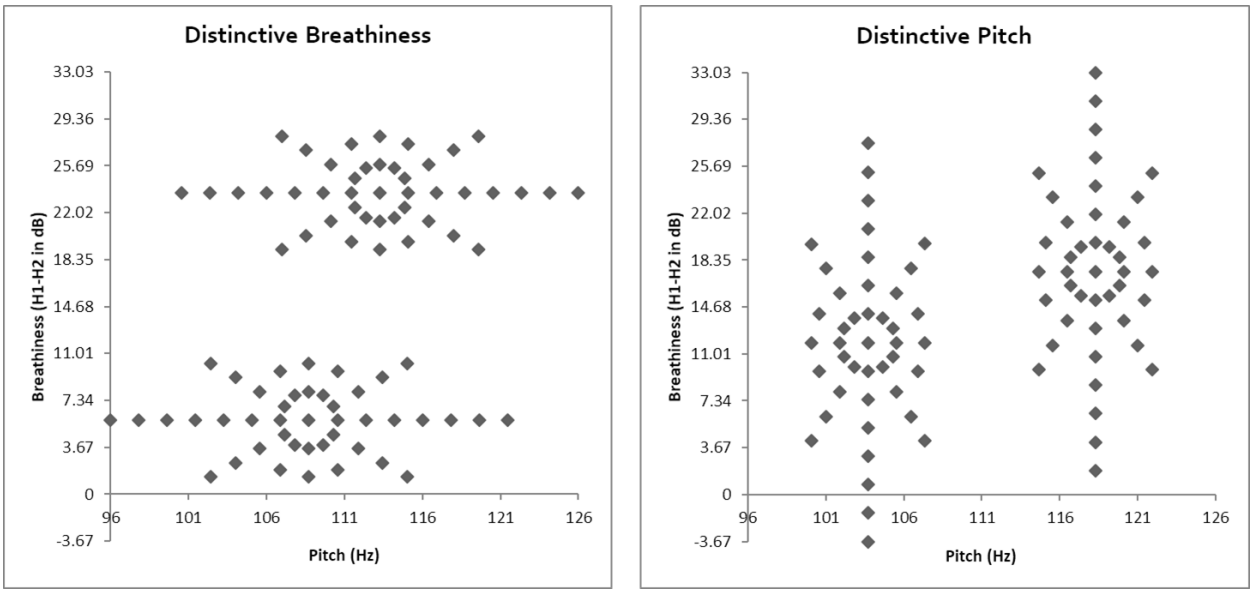

Figure 2.2. Training stimuli: Distinctive Breathiness (left) and Distinctive Pitch (right) distributions. Each training stimulus has a breathiness (H1-H2 in dB) value and a pitch (Hz) value, represented by a black point in the two-dimensional space.

Each stimulus token is represented by a point on the graph, and belongs to one of two categories, arbitrarily Category A and Category B, which are visually distinguishable as the two clusters of points. In both distributions, *Distinctive Breathiness* (left) and *Distinctive Pitch* (right), Category A had relatively lower f0 and lower H1-H2, while Category B had higher f0 and higher H1-H2.

The stimuli in each training set were designed to cause participants to favour one cue over the other (i.e. give a higher weight to one cue than the other). For the Distinctive Breathiness stimuli (Figure 2.2., left), optimal categorization would be obtained by attending more to the breathiness cue, and for the Distinctive Pitch stimuli (Figure 2.2., right), optimal categorization would be obtained by attending only to the pitch cue. Cue distinctiveness was manipulated by controlling the difference in mean values between categories and range of values within categories. In the *Distinctive Breathiness* training set, no tokens in either category had overlapping breathiness values with tokens in the other category (within-category range = 2.4

JNDs or 8.8 dB, distance between category means = 4.8 JNDs or 17.6 dB), whereas along the Pitch range, 93 percent of the tokens in one category had overlapping pitch values with tokens in the other category (within-category range = 8.3 JNDs or 25 Hz, distance between category means = 1.3 JNDs or 4 Hz). Thus, in this set, participants should find Breathiness to be more informative of the contrast than Pitch and should therefore give it a higher weight. Similarly, in the *Distinctive Pitch* training set, no tokens in either category had overlapping pitch values with tokens in the other category (within-category range = 2.3 JNDs or 7 Hz, distance between category means = 4.7 JNDs or 14 Hz), whereas along the Breathiness range, 93 percent of the tokens in one category had overlapping breathiness values with tokens in the other category (within-category range = 8.5 JNDs or 31.2 dB, distance between category means = 1.5 JNDs or 5.5 dB). Thus, in this set, Pitch was more informative of the contrast than Breathiness, and was therefore expected to get a higher weight.

Note that the correlation between Breathiness and Pitch in each distribution is not the co-variation that is enhancing: as described above, increased breathiness (larger H1-H2) together with lower pitch enhances low frequency energy. Instead, we chose to give listeners the non-enhancing, positive, correlation to avoid giving listeners any experience with the enhancing, negative, correlation. Thus, if this distributional correlation biased them toward one of the cue relations at all, it would be for the non-enhancing relation.

A set of 50 test stimuli was also created in which Breathiness and Pitch varied orthogonally within the same two-dimensional space. These were withheld during training. They are shown in Figure 2.3.

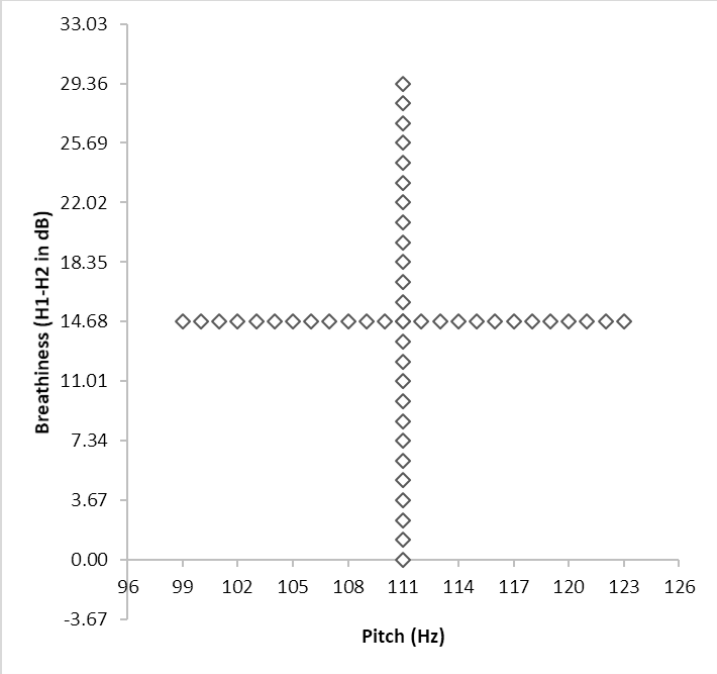

Figure 2.3. Test Stimuli for all conditions. Each test stimulus is represented by a point in the two-dimensional space. Vertically arranged points have the same pitch (111 Hz) but vary in breathiness. Horizontally arranged points have the same breathiness (14.68 dB) but vary in pitch.

For the vertically arranged points (25 tokens), Pitch was held constant at 111 Hz while Breathiness changed in 1/3 JND (1.22 dB) increments from 0 to 29.36 dB. For horizontally arranged points (25 tokens), Breathiness was held constant at 14.68 dB, while Pitch was changed in 1/3 JND (1 Hz) increments from 99 to 123 Hz. Since one dimension is always held at a constant value in the middle of the scale where categorization is ambiguous, the category choice made by participants on these tokens should be primarily conditioned by changes along the other dimension. The same set of test stimuli was used to measure cue weights for both the *Distinctive Breathiness* and *Distinctive Pitch* training sets. Pitch and Breathiness values for all training and test tokens can be found in Appendix A.

2.2.2.3. Stimuli synthesis

The 222 unique stimuli tokens – 86 training tokens for the Distinctive Breathiness training set, 86 training tokens for the Distinctive Pitch training set, and 50 test tokens – were synthesized using the free program Voice Synthesis (Antoñanzas-Barroso, Kreiman, and Gerratt, 2006). First, a natural voice sample, the same one used for the Pitch and Vowel Duration stimuli, was inverse-filtered to obtain the harmonic part of the glottal source. A male voice sample ([ɑ], f0 = 111 Hz, H1-H2 = 3.6 dB) that had been processed in this way was used as the base for all the stimuli in this study. In Voice Synthesis, Pitch was first manipulated by changing the f0 parameter, then Breathiness was manipulated by increasing or decreasing the amplitude of the first harmonic, thereby changing the amplitude difference between the first and second harmonic (H1-H2) without affecting the rest of the harmonic spectrum. Inharmonic information (e.g. noise, vocal tremors, jitter and shimmer, and formant frequencies and bandwidths) was then reintroduced to approximate the original voice. After the vowel was manipulated, an unaspirated [t], composed of a period of silence and two pulses, was spliced onto each token to form the syllable [tɑ]. Figure 2.4. shows the resulting full-audio spectrum for two test stimuli that have the same pitch (111 Hz) but are at either ends of the Breathiness continuum.

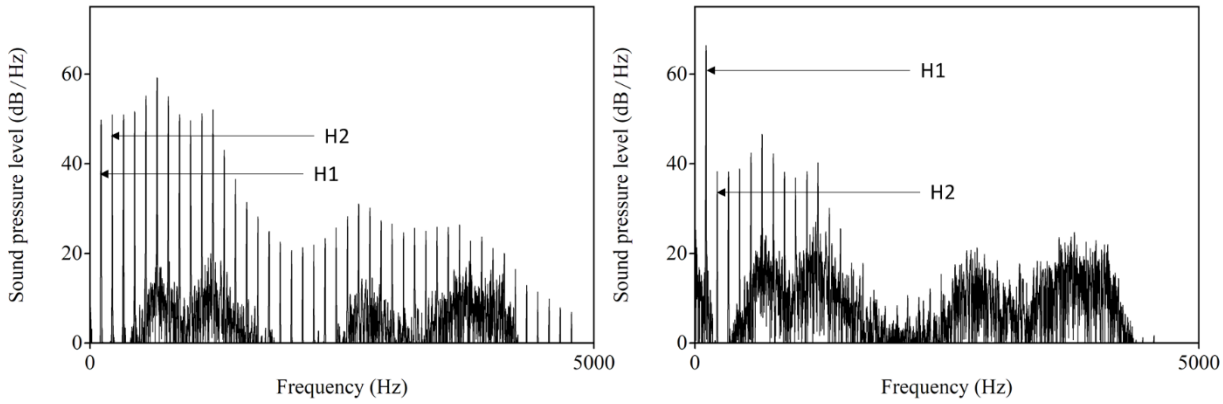

Figure 2.4. Examples of synthesized stimuli. Modal (left): H1-H2 = 0 dB. Breathy (right), H1-H2 = 29.36 dB.

Note that integration between cues has been reported to be specific not only to the cues but also to the range of values being tested (e.g. Kingston et al., 1997). We manipulated H1-H2 in the same range of values; future research is needed to evaluate the extent of generalization to other acoustic correlates of breathiness (e.g. H1-A1, HNR, CPP) or to different ranges of these cues.

2.2.3. Procedure

English and Hani listeners were tested in slightly different settings and Hani listeners required slight modifications to the presentation of the stimuli. Thus, in this section, I first lay out the general procedure that was common for both listener groups, then I describe the specific procedures used for each group where they differ. All participants were trained on a Language 1 (L1) and then a Language 2 (L2). This is shown schematically in Figure 2.5.

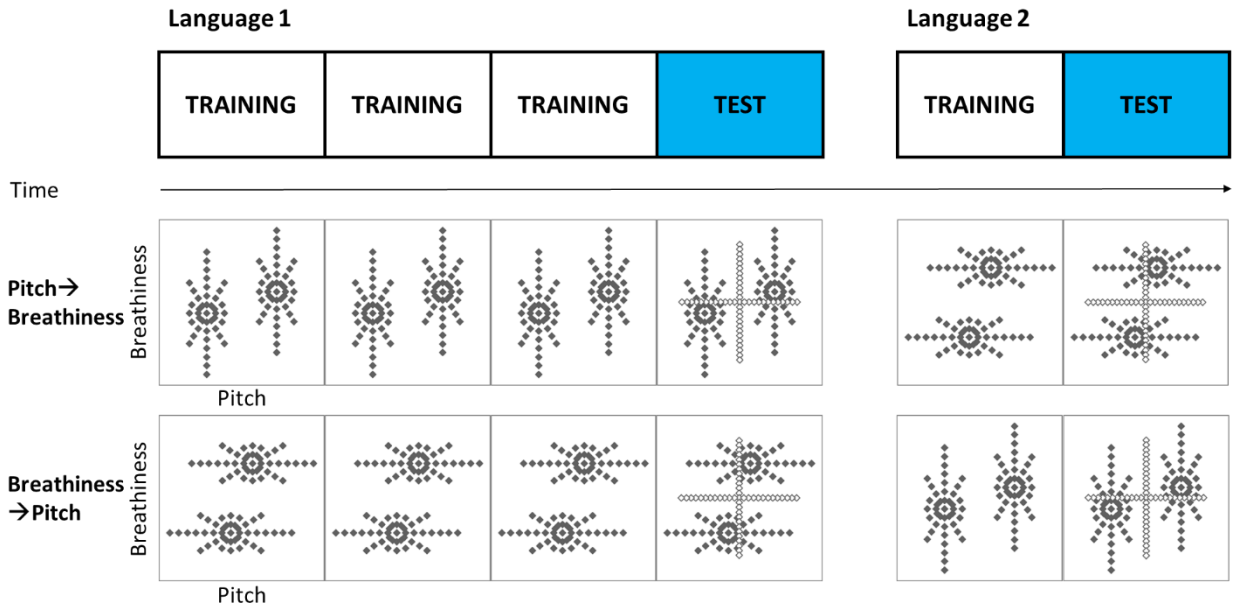

Figure 2.5. Design of the experiment: Participants completed training blocks and test blocks in order from left to right (L1: 3 training, 1 test; L2: 1 training, 1 test). Pitch→Breathiness participants heard the Distinctive Pitch stimuli in L1 and Distinctive Breathiness stimuli in L2. Breathiness→Pitch participants heard the Distinctive Breathiness stimuli in L1 and Distinctive Pitch stimuli in L2. Stimuli presented in each block are displayed for each condition under each block type (black points = training stimuli, white points = test stimuli).

The direction of the shift was counterbalanced such that half of the participants in each language group, English and Hani, were trained and tested on the *Distinctive Pitch* stimulus set as their L1 and the *Distinctive Breathiness* set as their L2, while the other half of the listeners were trained and tested on the *Distinctive Breathiness* set as their L1 and the *Distinctive Pitch* set as their L2. In Language 1, all participants heard three blocks of training stimuli each consisting of 86 randomized trials (labeled "Training" in Figure 2.5.), then one test block (labeled "Test" in Figure 2.5.) which included 136 randomized trials consisting of both training and test stimuli. In Language 2, participants heard one block of new training stimuli, then one block with the same training stimuli plus the test stimuli.

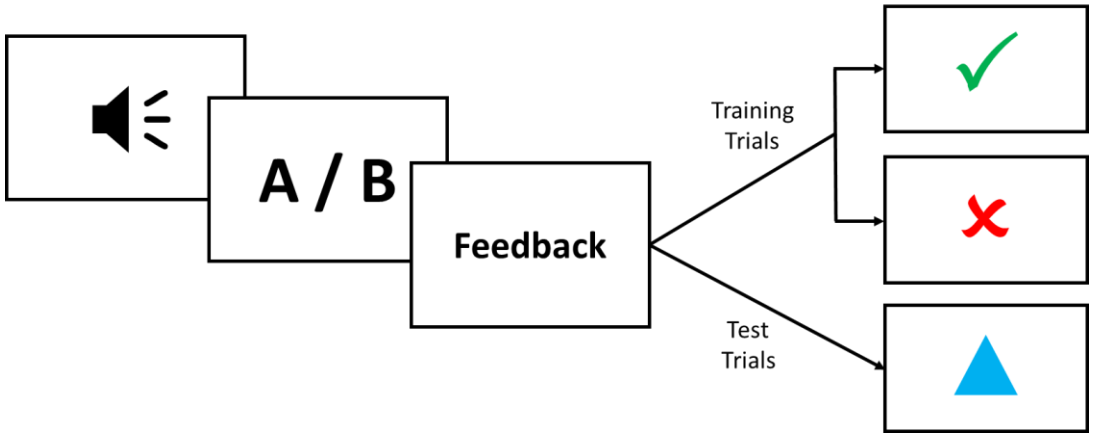

Figure 2.6. Procedure during each trial: Participants heard a sound, made a choice between Category A and Category B, then received feedback: Correct (green check mark), Incorrect (red ex), Unknown (blue triangle).

The sequence of events per trial is given schematically in Figure 2.6. On each trial, participants listened to a single stimulus token and decided which category, A or B, the sound belonged to. After pressing one of the two keys, participants received visual feedback. For training trials, the feedback informed them whether their response was correct or incorrect. Participants were not told what to listen for. They were instructed to guess at first, then use the feedback to get as many trials correct as possible. During the test blocks at the ends of L1 and L2, participants continued to receive informative feedback on the training trials, but feedback was an uninformative blue triangle for the novel test trials.

The procedural differences were as follows: English listeners completed the entire process in a quiet lab, unsupervised by an experimenter. Stimuli were presented using the online Appsobabble platform (Tehrani, 2015) and participants listened to the stimuli on 3M Peltor HTB79A-02 headphones. The instructions were given in written form as part of the experimental interface. English listeners gave their responses on a QWERTY keyboard, pressing either the S key if they thought the word they heard was ‘sea’ (Category A) or the L key if they thought the word was ‘land’ (Category B). They completed a Background Questionnaire form after

completing the experiment. This form asked them for their age and asked them to list out the languages they speak, when they began to learn each one, and how fluent they are (beginner, intermediate, functional, or fluent).

Hani listeners were given oral instructions by the experimenter in Mandarin. The experimenter also obtained their background information orally prior to the experiment. The questionnaire for Hani listeners asked for information about their age, the village they are from, and how often they use Hani to communicate (in percentages). Before beginning the experiment, they were also given 8 trials for practice, which were different from those used in the experiment itself. Hani listeners did the experiment on touch screen devices and listened to the stimuli on 3M Peltor HTB79A-02 headphones. Rather than selecting an arbitrary key, they touched a picture on the screen, either a rabbit (Category A) or a turtle (Category B), to indicate their choice.

2.2.4. Conditions

As in the pitch and vowel duration experiments, the crucial manipulation was the mapping between categories in L1 and L2 such that the cues had a positive or negative relation. The two resulting conditions are thus the positive *Non-Enhancing* mapping condition, in which the change in category labels from L1 to L2 reverses the enhancing co-variation between pitch and breathiness, and the negative *Enhancing* mapping condition, in which the labeling respects the enhancing co-variation. Combined with the counterbalancing of Direction (Pitch to Breathiness vs. Breathiness to Pitch), this creates four between-subjects conditions, schematized below in Figure 2.7.

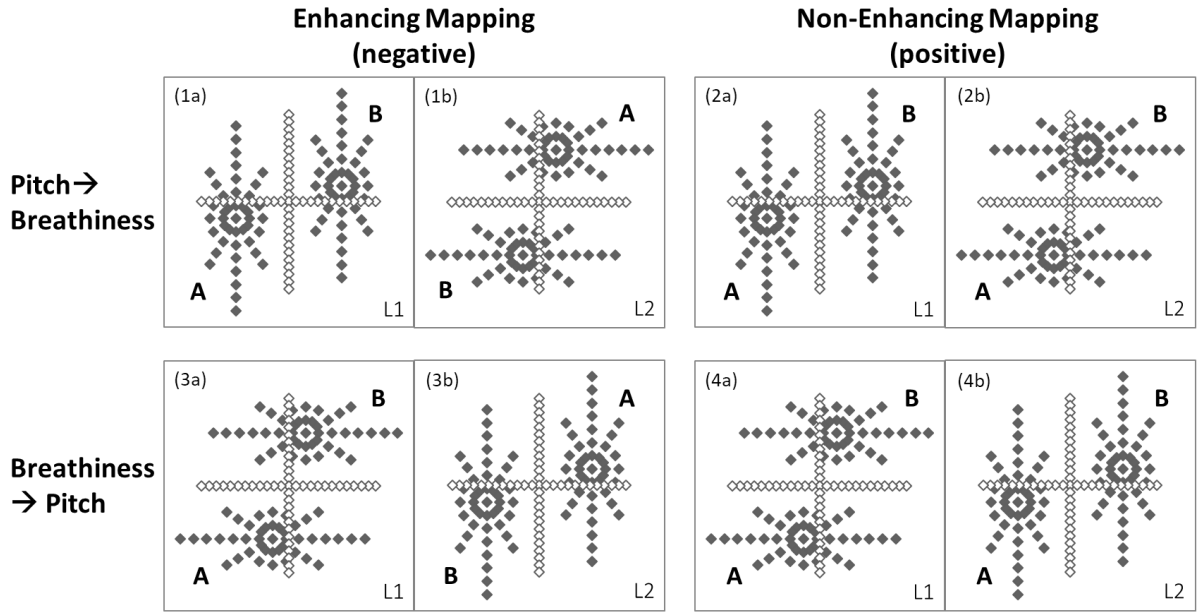

Figure 2.7. Experiment Conditions: Direction (Pitch → Breathiness, upper panels vs. Breathiness → Pitch, lower panels) × Mapping Relation (Enhancing, left panels vs. Non-Enhancing, right panels). Category labels (A or B) are labeled for each set of training stimuli in each panel.

In all four conditions, the category labels were the same for L1, the left-most stimulus set in each pair. That is, for L1 in every condition, the category with relatively low f0 and H1-H2 (bottom left quadrant) was arbitrarily labeled *A* and the category with relatively high f0 and H1-H2 (top right quadrant) was labeled *B*. Thus, L1 in the Enhancing mapping condition is identical to L1 in the Non-Enhancing condition for each Direction, providing a built-in replication of the results.

The Enhancing and Non-Enhancing conditions differ only in the labels assigned to the categories in L2. In L2 of the *Enhancing* mapping conditions, the category with relatively low f0 and H1-H2 (bottom left quadrant) was labeled *B* and the category with relatively high f0 and H1-H2 (top right quadrant) was labeled *A*. In L2 of the *Non-Enhancing* conditions, the category with relatively low f0 and H1-H2 (bottom left quadrant) was labeled *A* and the category with the relatively high f0 and H1-H2 (top right quadrant) was labeled *B*.

The rationale behind the Mapping Relation manipulation is as follows: Participants learn to attribute more weight to the more distinctive cue in L1, and then are forced to transfer weight onto a different cue in L2. Suppose a participant is trained first on the set of stimuli in which Breathiness is more distinctive, and, by the end of training, learns to rely more on the Breathiness cue than on the Pitch cue to categorize stimuli. That is, they have learned that less breathy tokens belong to Category A, and more breathy ones belong to Category B. When they are given the new stimulus set in which Pitch is more distinctive, they must shift cue weight onto Pitch in order to be accurate in the categorization task since changes in breathiness are now less informative. If listeners are aware of the enhancing relation between pitch and breathiness, the participant will expect the category with lower pitch in L2 to have the same label, B, as the breathier category in L1. Similarly, they will expect the category with higher pitch in L2 to have the same label, A, as the category with less breathiness in L1. The category labels in the *Enhancing* condition match these expectations, while the category labels in the *Non-Enhancing* condition reverse these expectations.

Given the two sets of conditions, the experiment has a two-by-two design with four conditions in total: Pitch→Breathiness – Enhancing (Pitch-Enhancing), Pitch→Breathiness – Non-Enhancing (Pitch-NonEnhancing), Breathiness→Pitch – Enhancing (Breath-Enhancing), and Breathiness→Pitch – Non-Enhancing (Breath-NonEnhancing). Participants were randomly assigned to one of these four conditions.

2.2.5. Analysis

The purpose of training in L1 was to control listeners' experience with cues in the experiment such that they were all weighting the L1 distinctive cue higher than the L1 non-

distinctive cue. Therefore, participants who were clearly not using the distinctive cue to categorize in L1 were excluded. Specifically, a participant was excluded if their performance on the training trials in the test block was below chance, that is, if the probability of obtaining the observed number of correct responses was greater than 0.05 given the hypothetical number of correct responses in a binary choice task. For a participant who responded to all the training trials (86 trials), 53 or more correct responses (> 62%) was considered above-chance performance.

From the English group, 14 participants were excluded for being below this performance threshold. Including participants who were excluded for their language background and those who did not complete the study, a total of 27 participants were excluded from the English group. In the final analysis, there were 30 participants in the Pitch-Enhancing condition, 31 participants in the Pitch-NonEnhancing condition, 30 participants in the Breath-Enhancing condition, and 32 participants in the Breath-NonEnhancing condition, totalling 123 participants.

From the Hani group, 14 participants were excluded for performing below criteria for inclusion on the training trials of the test block in L1. Including those excluded for age, using Hani less than 50% of the time, and those who did not complete the study, a total of 40 participants were excluded from the Hani group. In the final analysis, there were 15 participants in the Pitch-Enhancing group, 16 participants in the Pitch-NonEnhancing group, 16 participants in the Breath-Enhancing group, and 16 participants in the Breath-NonEnhancing group, totalling 63 participants.

Two pairs of cue weights were obtained for each participant: one weight for each cue, Pitch and Breathiness, from L1, and one weight for each cue from L2. The pair of cue weights from each Language was calculated from the test trials in the test block of that Language only. Following Holt and Lotto (2006), a logit binomial regression was run using the listeners'

Category Choice on the test trials as the dependent variable and the Pitch and Breathiness values for each test trial as independent predictors. Cue weights were taken as the coefficients of Breathiness and Pitch from this logit binomial regression. These coefficients are a measure of how well changes in each dimension, Breathiness or Pitch, was able to predict the responses of a participant. For example, if Breathiness has a higher coefficient than Pitch, then Breathiness is a better predictor of the participant's category choice. The logit binomial regression was implemented in R (R Development Core Team, 2015) using the built-in glm function. The absolute values were normalized to sum to one. Note that the normalization of weights does not take into account the accuracy of listeners' categorization, but rather gives a better idea of the *relative* contribution of each cue for each listener. These normalized cue weights were the dependent variable in all subsequent analyses.

The normalized cue weights were then analyzed using a mixed effects linear regression model, implemented in R, using the *lme4* package (Bates et al., 2008). P-values were obtained from the t-statistic. Pairwise Tukey's HSD post-hoc tests were run using the *lsmeans* package (Lenth, 2016) to identify which pairs were significantly different in significant interactions. P-values from these tests are adjusted for multiple comparisons.

2.3. Results

The results for Hani listeners will be presented first in Section 2.3.3, then the results for English listeners will be presented in Section 2.3.4. For each language group, L1 results will be presented first, showing that the initial learning of cue weights was not different across conditions. This is so that differences in cue weights in L2 between conditions can be attributed to experimental manipulations, rather than to differences in initial learning.

2.3.1. Hani

2.3.1.1. Language 1

The normalized cue weights for the Distinctive and Non-Distinctive cues from the test block of L1 are given in Figure 2.8., grouped by condition.

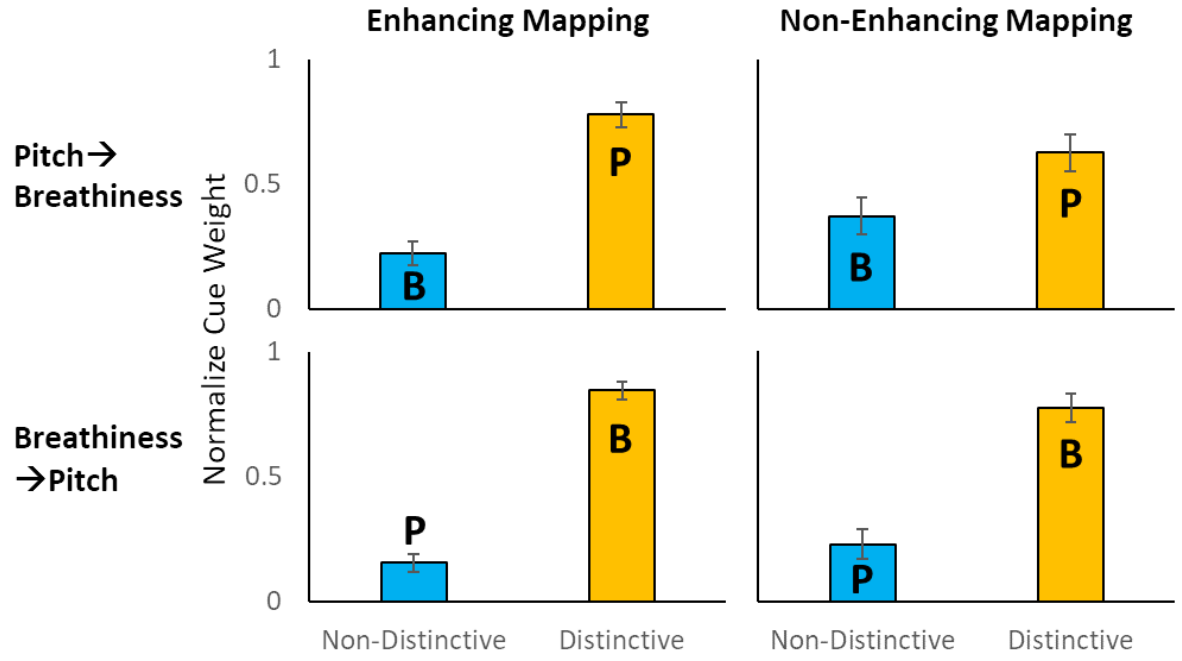

Figure 2.8. Enhancing cues: Language 1 Distinctive (yellow) and Non-Distinctive (blue) normalized cue weights for Hani listeners, by Direction (Pitch→Breathiness vs Breathiness→Pitch) and Mapping Relation (Enhancing vs. Non-Enhancing). The specific cue is labeled as P for Pitch and B for Breathiness. Error bars = Standard Error.

Figure 2.8. shows that the Distinctive cue (yellow) was weighted higher than the Non-Distinctive cue (blue) in L1 for Hani listeners in every condition.

These L1 data were analyzed using a mixed effects model with a random intercept of Subject, and fixed effects were Direction (Pitch→Breathiness vs. Breathiness→Pitch), Mapping Relation (Enhancing vs. Non-Enhancing), and Distinctiveness (Distinctive vs. Non-Distinctive).

All 2- and 3-way interactions were also included. This was the highest level of random effects structure that converged. The model results are given in Table 2.3.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.16	.06	2.76	.006	**
Distinctiveness = *Distinctive*	0.69	.08	8.70	<.001	***
Mapping Relation = *Non-Enhancing*	0.07	.08	0.93	.351	
Direction = *P→B*	0.07	.08	0.83	.407	
Direction × Mapping Relation = *P→B & Non-Enhancing*	0.08	.11	0.68	.496	
Direction × Distinctiveness = *P→B & Distinct.*	-0.13	.11	-1.17	.241	
Mapping Relation. × Distinctiveness = *Non-Enhancing. & Distinctive*	-0.15	.11	-1.32	.188	
Direction × Distinct. × Mapping Rel. = *P→B & Distinct. & Non-Enh.*	-0.15	.16	-0.96	.335	

Table 2.3. Lmer results from Hani listeners' performance on Language 1 for enhancing cues.

There was a significant random effect of Subject, indicating that there was individual variation in the cue weights given to the Non-Distinctive cue in the Breathiness-Enhancing condition. Of the fixed effects, only Distinctiveness was significant. A pairwise Tukey's HSD test on the model confirmed that the Distinctive cue was weighted significantly higher than the Non-Distinctive cue in all four conditions (Pitch-Enhancing, $\beta = 0.56$, $p < .001$; Breath-Enhancing, $\beta = 0.69$, $p < .001$; Pitch-NonEnhancing, $\beta = 0.25$, $p < .029$; Breath-NonEnhancing, $\beta = 0.54$, $p < .001$). There was also no significant difference between the Distinctive cue weights in different conditions and between the Non-Distinctive cues in different conditions (Pitch→Breathiness, Enhancing vs. Non-Enhancing, $\beta = 0.15$, $p = .570$; Breathiness→Pitch, Enhancing vs. Non-Enhancing, $\beta = 0.07$, $p = .983$; Enhancing, Pitch→Breathiness vs. Breathiness→Pitch, $\beta = 0.07$, $p = .9915$; Non-Enhancing, Pitch→Breathiness vs. Breathiness→Pitch, $\beta = 0.14$, $p = .6105$.

Thus, subjects in all 4 conditions learned to weight the Distinctive cue higher than the Non-Distinctive cue in L1.

2.3.1.2. Language 2

The normalized cue weights for the Distinctive and Non-Distinctive cues from the test block of L2 are given in Figure 2.9., grouped by condition.

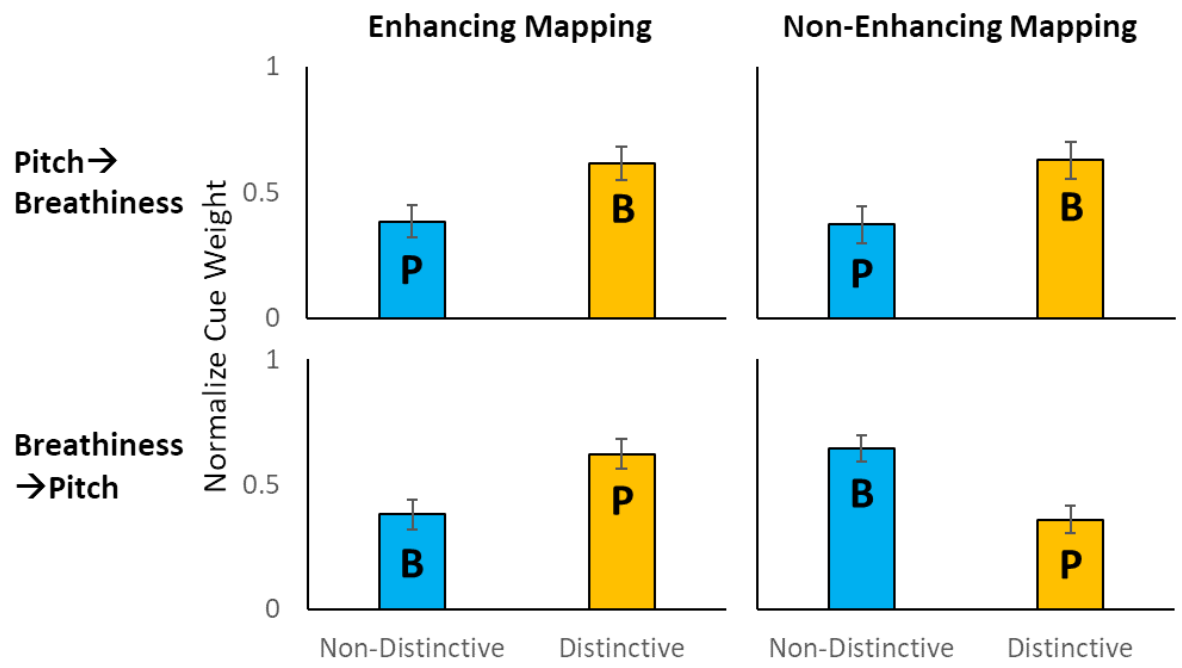

Figure 2.9. Enhancing cues: Language 2 Distinctive (yellow) and Non-Distinctive (blue) normalized cue weights for Hani listeners, by Direction (Pitch→Breathiness vs Breathiness→Pitch) and Mapping Relation (Enhancing vs. Non-Enhancing). The specific cue is labeled as P for Pitch and B for Breathiness. Error bars = Standard Error.

In L2, the distinctiveness of the Pitch and Breathiness cues was switched. Breathiness was now the Distinctive cue in the Pitch→Breathiness conditions and Pitch the new Distinctive cue in the Breathiness→Pitch conditions. If participants successfully shifted cue weight onto the new Distinctive cue (blue), then cue weights from the test trials should show a higher weight for Breathiness and a lower weight for Pitch in the Pitch→Breathiness conditions, and the opposite

weighting in the Breathiness→Pitch conditions. This was the case for every condition except the Breathiness→Pitch – Non-Enhancing condition, where the Distinctive cue from L1 (blue) retained a higher weight than the new Distinctive cue from L2 (yellow).

These data were analyzed using a mixed effects logistic regression which included the random intercept of Subject and the fixed effects Direction (Pitch→Breathiness vs. Breathiness→Pitch), Mapping Relation (Enhancing vs. Non-Enhancing), Distinctiveness (Distinctive vs. Non-Distinctive), as well as all 2- and 3-way interactions. The model results are in Table 2.4.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.38	.06	6.07	<.001	***
Distinctiveness = *Distinctive*	0.24	.09	2.72	.006	**
Mapping Relation = *Non-Enhancing*	0.26	.09	2.96	.003	**
Direction = *P→B*	0.01	.09	0.06	.956	
Direction × Mapping Relation = *P→B & Non-Enhancing*	-0.27	.13	-2.17	.030	*
Direction × Distinctiveness = *P→B & Distinct.*	-0.01	.13	-0.08	.937	
Mapping Relation. × Distinctiveness = *Non-Enhancing. & Distinctive*	-0.52	.13	-4.18	<.001	***
Direction × Distinct. × Mapping Rel. = *P→B & Distinct. & Non-Enh.*	0.55	.18	3.07	.002	**

Table 2.4. Lmer results from Hani listeners' performance on Language 2 for enhancing cues.

Here also, the random effect of Subject is significant, as in previous model. There was a significant interaction between Direction × Cue Relation × Distinctiveness. This was driven by an effect that is unique to the Breath-NonEnhancing condition. As shown by a pairwise Tukey's HSD test on the three-way interaction, the Distinctive cue was weighted marginally higher than the Non-Distinctive cue for the Pitch-Enhancing condition (β = 0.25, p = .184), the Breath-

Enhancing condition (β = 0.24, p = .116), and the Pitch-NonEnhancing condition (β = 0.25, p = .076). However, in the Breath-NonEnhancing condition, the Distinctive cue was weighted significantly *lower* (β = -0.28, p = .030). The effect did not reach significance for the first three conditions likely due to the lack of power given a small *n*. With a larger number of participants matching the *n* of the English group, I expect these effects to be significant. However, the trend suggests that these listeners were able to shift cue weights to some degree such that the Distinctive cue from L1 was no longer weighted higher. This was clearly not true for the last condition, Breath-NonEnhancing, in which listeners failed to give a higher weight to the new Distinctive cue.

2.3.2. English

2.3.2.1. Language 1

The normalized cue weights for the Distinctive and Non-Distinctive cues from the test block of L1 are given in Figure 2.10., grouped by condition.

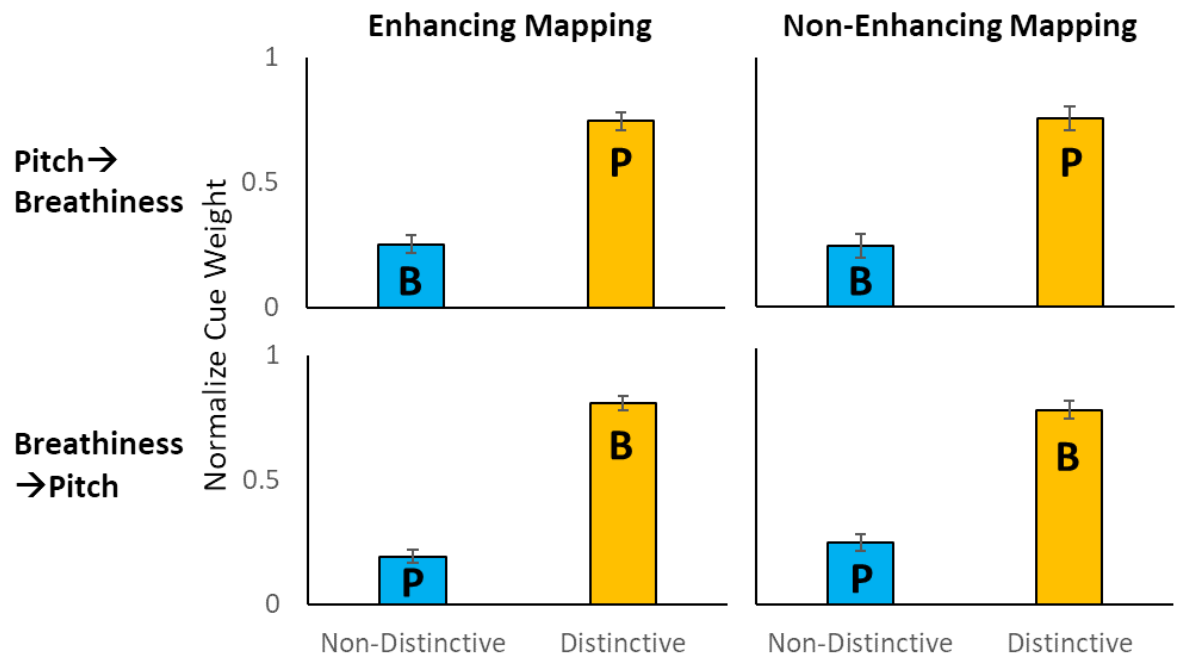

Figure 2.10. Enhancing cues: Language 1 Distinctive (yellow) and Non-Distinctive (blue) normalized cue weights for English listeners, by Direction (Pitch→Breathiness vs Breathiness→Pitch) and Mapping Relation (Enhancing vs. Non-Enhancing). The specific cue is labeled as P for Pitch and B for Breathiness. Error bars = Standard Error.

Figure 2.10. shows that the Distinctive cue (yellow) was weighted higher than the Non-Distinctive cue (blue) in L1 for English listeners in every condition. Results from the mixed effects model on Language 1 data confirmed this. In addition to the random intercept of Subject, the fixed effects included the between-subjects variables Direction (Pitch→Breathiness vs. Breathiness→Pitch) and Mapping Relation (Enhancing vs. Non-Enhancing), and the within-subjects variable Distinctiveness (Distinctive vs. Non-Distinctive). All 2- and 3-way interactions were also included. Again, this is the most complex random effects structure that converged. The model results are given in Table 2.5.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.19	.04	5.16	<.001	***
Distinctiveness = *Distinctive*	0.61	.05	11.66	<.001	***
Mapping Relation = *Non-Enhancing*	0.05	.05	0.91	.361	
Direction = *P→B*	0.06	.05	1.14	.251	
Direction × Mapping Relation = *P→B & Non-Enhancing*	-0.06	.07	-0.77	.444	
Direction × Distinctiveness = *P→B & Distinct.*	-0.12	.07	-1.62	.105	
Mapping Relation. × Distinctiveness = *Non-Enhancing. & Distinctive*	-0.09	.07	-1.29	.196	
Direction × Distinct. × Mapping Rel. = *P→B & Distinct. & Non-Enh.*	0.11	.10	1.08	.279	

Table 2.5. Lmer results from English listeners' performance on Language 1 for enhancing cues.

These results mirror those from the Hani group. Distinctiveness was the only significant fixed effect. A pairwise Tukey's HSD test on the model confirmed that the Distinctive cue was weighted significantly higher than the Non-Distinctive cue in all four conditions (Pitch-Enhancing, $\beta = 0.49$, $p < .001$; Breath-Enhancing, $\beta = 0.61$, $p < .001$; Pitch-NonEnhancing, $\beta = 0.51$, $p < .001$; Breath-NonEnhancing, $\beta = 0.52$, $p < .001$). There was also no significant difference between the Distinctive cue weights in different conditions and between the Non-Distinctive cues in different conditions (all *p*-values close to 1.0). Thus, subjects in all 4 conditions learned to weight the Distinctive cue higher than the Non-Distinctive cue in L1.

2.3.2.2. Language 2

The normalized cue weights for the Distinctive and Non-Distinctive cues from the test block of L2 are given in Figure 2.11., grouped by condition.

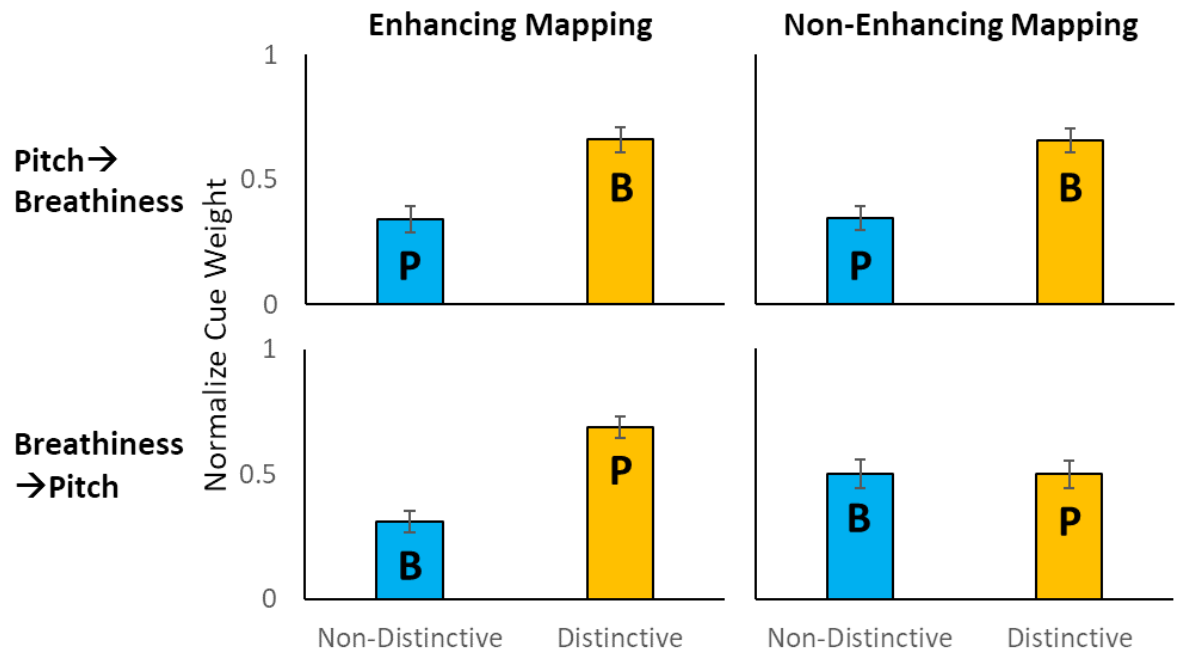

Figure 2.11. Enhancing cues: Language 2 Distinctive (yellow) and Non-Distinctive (blue) normalized cue weights for English listeners, by Direction (Pitch→Breathiness vs. Breathiness→Pitch) and Mapping Relation (Enhancing vs. Non-Enhancing). The specific cue is labeled as P for Pitch and B for Breathiness. Error bars = Standard Error.

Again, in L2, the distinctiveness of cues was switched. In Figure 2.11., successful cue weight shifting is indicated by a greater weight for the new Distinctive cue (yellow) compared to the new Non-Distinctive cue (blue). This was the case for every condition except the Breath-NonEnhancing condition, where the cue weights of the Distinctive and Non-Distinctive cues were not different.

This is confirmed by results from the lmer model. Again, the model included the random intercept of Subject and the fixed effects included between-subjects variables Direction (Pitch→Breathiness vs. Breathiness→Pitch) and Mapping Relation (Enhancing vs. Non-Enhancing), and the within-subjects variable Distinctiveness (Distinctive vs. Non-Distinctive). All 2- and 3-way interactions were also included. The model results are in Table 2.6.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.31	.05	6.18	<.001	***
Distinctiveness = *Distinctive*	0.37	.07	5.30	<.001	***
Mapping Relation = *Non-Enhancing*	0.19	.07	2.69	.007	**
Direction = *P→B*	0.03	.07	.42	.677	
Direction × Mapping Relation = *P→B & Non-Enhancing*	-0.18	.10	-1.85	.064	.
Direction × Distinctiveness = *P→B & Distinct.*	-0.06	.10	-0.59	.555	
Mapping Relation. × Distinctiveness = *Non-Enhancing. & Distinctive*	-0.38	.10	-3.81	<.001	***
Direction × Distinct. × Mapping Rel. = *P→B & Distinct. & Non-Enh.*	0.37	.14	2.62	.009	**

Table 2.6. Lmer results from English listeners' performance on Language 2 for enhancing cues.

There was a significant interaction of Direction × Cue Relation × Distinctiveness. This was driven by an effect that is unique to the Breath-NonEnhancing condition. As shown by a pairwise Tukey's HSD test on the three-way interaction, the Distinctive cue was weighted significantly higher than the Non-Distinctive cue for the Pitch-Enhancing condition ($\beta = 0.32$, $p < .001$), the Breath-Enhancing condition ($\beta = 0.38$, $p < .001$), and the Pitch-NonEnhancing condition ($\beta = 0.31$, $p < .001$), but not for the Breath-Non-Enhancing condition ($\beta = 0.00$, $p = 1$). That is, listeners in all conditions shifted cue weights successfully except for those shifting from Breathiness onto Pitch when the mapping between categories was non-enhancing.

2.4. Discussion

In this chapter, I used cue weighting experiments to test for effects of enhancement that are independent of language experience. Two groups of listeners were chosen for their specific experience with the co-variation between the enhancing cues, pitch and breathiness: Hani

listeners, for whom pitch and breathiness both signal a single phonemic contrast, and English listeners, for whom pitch and breathiness do not signal any phonemic contrasts. These listeners were asked to weight the cues in one distribution that favoured the use of one of the cues, then re-weight the cues in a distribution that favoured the use of the other cues. The mapping of the categories from the first distribution to the second was either enhancing or non-enhancing. Listeners were expected to have difficulty shifting cue weights only if the mapping relations reversed either their expectations about how the cues should co-vary. The expectations about cue relations could either come from language experience, or from language-general perceptual enhancement between the cues. Hani listeners were predicted to have difficulty shifting cue weights when the mapping relation was positive (non-enhancing) because they have language experience with the negative (enhancing) co-variation between these cues. English listeners were also predicted to have difficulty shifting cue weights when the mapping relation was positive, but only if they expect the co-variation between pitch and breathiness to produce an enhancing effect.

Overall, the results from the two listener groups mirrored each other. All listeners in each group weighted the more distinctive cue higher than the less distinctive cue after training with the first distribution, regardless of whether pitch or breathiness was more informative. Hani listeners, who had experience with the enhancing relationship between breathiness and pitch, were unable to shift cue weights when the mapping relation was positive (non-enhancing). However, this was unexpectedly only true when listeners were shifting cue weight from breathiness onto pitch, and not from pitch onto breathiness. Though directionally asymmetric, these results from the Hani listeners set a baseline for how we can expect English listeners to behave if they come with expectations of how two cues should co-vary, despite their lack of

experience. Results showed that English listeners had difficulty shifting weight when the mapping relation was non-enhancing, and, like Hani listeners, this was only true for listeners shifting from breathiness onto pitch.

Setting aside the directional asymmetry momentarily, the results from English listeners complement earlier studies (e.g. Lee & Katz, 2016) in showing that experience with co-variation is not necessary for listeners to perceptually couple two cues. This is not predicted by the associative learning account, where all and only those acoustic patterns present in the speech signal should be learned by the listener. Since pitch and breathiness do not signal the same phonemic contrast in English, and they do not individually signal any phonemic contrasts, English listeners have no linguistically meaningful experience with co-variation between them. Thus, a pure learning theory of speech perception cannot account for the fact that these listeners experienced difficulty in shifting attention from breathiness to pitch when the relation between the cues was specifically non-enhancing.

There are two additional sources from which listeners could have learned that pitch and breathiness co-vary. The first of these is in their language experience with the speech signal. Though pitch and breathiness do not co-vary phonemically in English, these cues can both be influenced by the same articulatory gestures, thus it is possible that they may co-vary phonetically. If they do, then listeners could theoretically learn to associate them. However, recall that studies on large corpora of native English speech show that these two acoustic properties are actually *positively* correlated (e.g. Kreiman et al., 2007; Iseli et al., 2007). Thus, if listeners were influenced by their knowledge of this correlation in the current study, they should have had difficulty with shifting cue weights when the mapping was *negative* and enhancing, rather than the reverse. However, the results of this study were the opposite – that is, English

listeners had difficulty when the mapping *matched* their phonetic experience. It is also possible for listeners to have gained some experience with the co-variation between pitch and breathiness from the distributions in the experiment. Recall that in all the distributions presented to listeners, regardless of which cue was more informative, the category with higher pitch was always breathier, and the category with lower pitch was always less breathy. So in the study itself, listeners were exposed to a positive, non-enhancing correlation between the two cues. This should effectively bias them to prefer cue shifting when the mapping relation was positive over cue shifting when the mapping relation was negative. Again, this was not the case, as listeners actually had more difficulty with the task in the positive mapping condition. With both phonetic experience and short-term exposure in the experimental setting biasing listeners in the opposite direction, I take the results of this study to be even stronger evidence for auditory enhancement.

I now return briefly to the question of why the enhancement effect was *asymmetric* in this study. It was predicted that both Hani and English listeners would have difficulty shifting cue weights when the mapping relation was non-enhancing, and no difference was predicted between listeners shifting from breathiness to pitch and those shifting in the reverse direction. However, results from both groups of listeners clearly showed that listeners shifting weight from pitch to breathiness were able to do so just as well as when the mapping relation was enhancing. The only listeners who had difficulty with the cue shift were those shifting from breathiness to pitch when the mapping was non-enhancing.

First, it is possible that this asymmetry was caused by an unequal perceptibility between the two cues. But, this is unlikely since i) the cues were scaled using JNDs for English listeners, and ii) there were no differences in cue weighting when listeners were learning weights from the initial distribution. Another likely cause is that both Hani and English listeners have more

experience with pitch as a cue than breathiness – Hani listeners because they also have a tone contrast, and English listeners because they use pitch as an intonational cue. This possibility will be ruled out in additional experiments presented in Chapter 4.

Finally, enhancing cues are defined by Kingston et al. as cues which i) contribute to the same auditory effect, and ii) are perceptually integrated at a higher auditory node, the IPP. Perhaps because of the second part of this definition, much of the supporting evidence for auditory enhancement comes from studies showing that two cues are perceptually integrated. However, it is also the case that there are cues that do not contribute to the same auditory effect, but are perceptually integrated. This begs the question of whether the behaviour of listeners in this study was a result of the two cues being truly enhancing and having the same auditory effect, or whether the same effect could be obtained with two cues that were just perceptually integral. This is the purpose of the study in Chapter 3, where I use the same cue weighting paradigm to test two integral but non-enhancing cues, pitch and vowel duration.

CHAPTER 3: NON-ENHANCING CUES

3.1. Introduction

In Chapter 2, we saw evidence that listeners who did not have any language experience with the co-variation between enhancing cues, pitch and breathiness, nevertheless expected them to have an enhancing correlation. However, the question of whether this effect could be caused by perceptual integration alone was raised.

Recall that based on Kingston's definition of enhancing cues, they contribute to the same auditory effect and converge on a single intermediate perceptual property (IPP), making them perceptually inseparable. Much of the evidence in support of two cues being enhancing has focused on the perceptual inseparability of the cues. These are studies showing that two cues are perceptually integrated, including those that give the strongest evidence that enhancement is independent of experience (i.e. Kingston et al. 2008; Lee & Katz, 2016). That is, they show, using the Garner paradigm, that categories (e.g. voiced vs. voiceless stops in English, plain vs. fortis fricatives in Korean) are more discriminable when the relevant cues were correlated in a certain way (either negatively or positively, depending on the cues). This has led to a conflation of the terms *enhancement* and *integration* in the literature. However, while enhancing cues must, by definition, be perceptually integrated, integrated cues need not be enhancing. Cues that clearly do not share auditory similarities have been shown to integrate perceptually. Pitch and duration, for example, have been shown to be integral (Sandor, 2004), even though higher or lower pitch do not produce the same auditory effect as longer or shorter duration. In a study using the Garner paradigm, Sandor (2004) showed that listeners were faster at categorizing stimuli when pitch and vowel duration were correlated than when they were not correlated. Interestingly, the effect was significant for both the positive and negative correlation, though listeners were fastest to respond

when pitch and duration were positively correlated. This indicates that listeners do integrate the two cues, though without any constraint on the kind of relationship they bear.

Besides this experiment, a number of studies have found that pitch affects perceived duration (Lehiste, 1976; Pisoni, 1976; Yu, 2010; Gussenhoven & Zhou, 2013) and at least one study has found that duration affects the percept of tone (Blicher, Diehl & Cohen, 1990), suggesting that the two cues may be perceptually integral. Specifically, stimuli with the same duration are perceived to be longer when the f0 is higher and shorter when the f0 is lower.

Though similar results for non-linguistic stimuli (Brigner, 1988) has led some to propose that this pattern is rooted in domain-general psychoacoustics or psychophysics (e.g. Yu, 2010), no explanation is given as to why pitch and duration should be integral. These effects have also been attributed to perceptual compensation (Gussenhoven & Zhou, 2013). Cross-linguistically, segments with lower f0 tend to be produced with longer duration while those with higher f0 tend to be produced with shorter duration, as is the case in many tone languages (e.g. Faytak & Yu, 2011). Thus, upon hearing stimuli with higher pitch versus stimuli with lower pitch, listeners will expect the former to be shorter than the latter. Given stimuli with the same measured duration, listeners will compensate and perceive a longer duration for the high-pitched stimuli compared to the low-pitched stimuli, giving rise to the effect observed in perceptual studies. Importantly, neither of these accounts base their explanation on whether or not two cues contribute to the same auditory effect.

Besides the definitional difference between integrated cues and cues that Kingston calls enhancing, there may also be a difference in the level of processing at which integration and enhancement can/should be observed. In the studies where an integration effect is found, listeners are typically asked to judge which of two segments is longer (Lehiste, 1976; Pisoni,

1976) or give a duration rating for each stimulus (Yu, 2010; Gussenhoven & Zhou, 2013). These tasks typically force listeners to focus on low level differences in cue measures. In comparison, Kingston emphasizes that the role of enhancing cues is to convey contrastive information (e.g. Kingston & Diehl, 1994). Thus, we should expect to observe enhancement effects only when listeners are engaged in linguistically meaningful categorization. Indeed, if the perceptual object when listeners hear two enhancing cues is an Intermediate Perceptual Property (IPP), e.g. low frequency energy, then they have already abstracted away from the independent cue dimensions. Thus, we must ask whether all integrated cues are the same and would produce the same effects observed in Chapter 2, or whether cues that are integrated but also contribute to the same auditory effect are treated differently by listeners. These questions are addressed in this chapter, which uses the same paradigm to test the non-enhancing but perceptually integrated cue pair, pitch and vowel duration. In Section 3.1.1, I discuss the perceptual relation of the two cues and why they were selected for this study. I then introduce the language groups as well as the way in which the target cue pair is present in each language (Section 3.1.2). In Section 3.1.3., I present the design of the experiments and lay out the predictions given our understanding of way in which pitch and vowel duration relate. The experimental methods and results will be presented in Sections 3.2. and 3.3. respectively, followed by a discussion of these results in Section 3.4.

3.1.1. Pitch and vowel duration as non-enhancing cues

Pitch and vowel duration were selected as the non-enhancing cues. Pitch refers to the percept of change in fundamental frequency (f0), and vowel duration refers to the percept change in duration. These cues were chosen because they do not, in the Kingstonian sense, enhance each others' percept. That is, they do not contribute to the same auditory effect in the same way that,

for example, breathy voice and low pitch do: In breathy voice, the amplitude of the first harmonic is relatively high compared to that of the harmonics above it, thus re-enforcing the percept of low frequency energy. Low pitch naturally also reinforces the same percept, thus the two cues are enhancing. Since higher or lower pitch does not produce any auditory effect that longer or shorter vowel duration also produces, these do not constitute an enhancing cue pair.

3.1.2. English and Hani listeners' experience with pitch and vowel duration

English listeners were chosen as the group without experience with the cue pair. There is no segmental contrast in English to which both pitch and vowel duration are cues. There are also no contrasts to which pitch and vowel duration are cues individually. That is, neither tone nor vowel length are contrastive in English[3].

Note that pitch and vowel duration are known cues to English stress. However, given the nature of their relation in this prosodic contrast and the nature of particular stimuli in these experiments, I maintain that the use of these two cues in this study for this particular group of listeners is still valid. Stressed syllables in English have more extreme pitch and have longer vowel duration, whereas unstressed syllables have less extreme pitch and shorter vowel durations. Since pitch in stressed syllables can be either higher or lower than pitch in unstressed syllables, pitch and vowel duration do not have a strictly positive or negative relationship. Also, since the stimuli used in these experiments are all monosyllabic, stress does not bear on the way listeners are categorizing sounds, and so any knowledge about the pitch-duration relation listeners may have from a stress contrast in their language should also not come into play for this particular task.

[3] Of course, English has distinctions, such as the tense-lax contrast (e.g. Wells & Wells, 1982), in which duration is an important cue, but there are none in which duration is the primary cue.

Hani listeners were chosen for having experience with pitch and vowel duration as a cue pair, which have a negative co-variation pattern in the language. This is supported by production data described in Section 2.1.3.

	Pitch	Vowel Duration	Breathiness
Lax	Lower	Longer	Breathier
Tense	Higher	Shorter	Less breathy

Table 3.1. Hani tense-lax contrast and its correlates

3.1.3. Experimental design and predictions

Listeners' experience with these cues was controlled in the same way as in Chapter 2. First, as discussed in the previous section, listeners were selected for these studies based on their language background. Pitch and vowel duration are not linguistically relevant for English listeners, while they signal a phonemic contrast for Hani listeners. Experience was also experimentally controlled with the distribution of stimuli across these two cues in the same cue weighting paradigm. Listeners were first presented with a distribution of stimuli that biased them to weight one cue higher, then they were given a different distribution to induce a shift in attention to the other cue. The relation between category labels in the initial learning phase and the shift phase was manipulated such that the relationship between pitch and vowel duration was either positive or negative.

Recall that Hani listeners are experienced with the negative relation between pitch and vowel duration in the tense-lax contrast. Thus, we expect Hani listeners to have difficulty shifting their attention from one cue to another if the experimental mapping runs counter to their experience (positive mapping). Based on the premise that listeners do not associate non-enhancing cues to contrast, English listeners are not expected to have any difficulty with either

experimental mapping since they are not constrained by their experience with co-variation between these cues.

3.2. Methods

3.2.1. Participants

For English listeners, 172 undergraduate participants (age 18-29), different from those who participated in the experiment in Chapter 2, were recruited from the Psychology Subject Pool at UCLA. 29 subjects were excluded for having experience with languages that have a tone or length contrast. The remaining subjects were native speakers of English and had no experience with such languages, as self-reported on a Language Background form. Data was missing for 5 participants and there were technical difficulties for 2 participants, thus they were also excluded.

100 additional Hani listeners were recruited for this experiment in the same way as the listeners for the pitch and breathiness experiment. 76 were from the local middle school in Nanuoxiang (age 9-16) and 24 others from the area (age 23-64). All Hani listeners recruited also speak at least one variety of Mandarin as self-reported on a Language Background form. Data from the younger participants was judged to be unreliable by the experimenter, possibly due to the nature of the task. I thus set a cut-off age of 12 and excluded the 10 participants 11 years old and under. Additionally, 11 listeners were excluded for using Hani less than 50% of the time and one participant was excluded because of experimenter error.

3.2.2. Stimuli

All stimuli were the syllable [tɑ] with a specific pitch and vowel duration value on the vowel. In this section, I first describe the method for scaling these two cues so that they were matched to be equally discriminable to English listeners. Then I describe the distribution of stimuli within the acoustic space, as well as how they were synthesized.

3.2.2.1. Perceptual scaling

The acoustic measure used to manipulate Pitch was fundamental frequency (f0) in Hertz (Hz). The f0 scale ranged from 96 Hz to 126 Hz. This 30 Hz range equals 10 times the JND for English listeners, which is approximately 3 Hz. This pitch range is well within the normal range for the human male voice. Pitch was scaled using Hertz despite JND for pitch being typically measured using psychoacoustic scales (i.e. 3 mel for modal voice, Kollmeier et al., 2008) for practical reasons relating to speech synthesis and because the relationship between Hertz and mels is linear below 500 Hz (Stevens et al., 1937).

To ensure that Vowel Duration was equally salient compared to Pitch, the same range, 10 times the JND, was used. In a pilot, I used the minimum JND of 10% increment on the base duration from Bochner et al. (1988) for English listeners. However, this JND was too small, causing English listeners to attend to Pitch more easily in the initial learning phase. This could be attributed to a difference in the experimental design used in the Bochner et al. study and that used in the current study. To offset the difference in learning between Vowel Duration and Pitch, I increased the JND for Vowel Duration to an 11% increment on the base duration. With this JND, listeners attended equally to the two cues (see results). I chose 150 ms as the lower bound of Vowel Durations used. The full range was calculated as 150 ms × 1.11^{10}, giving a range of

276 ms with an upper bound of 426 ms. This is a normal range for vowel duration in English (Klatt & Cooper, 1975; Lefkowitz, 2017).

3.2.2.2. Stimulus distribution

The stimulus distributions for the non-enhancing cues exactly mirrors that those of the enhancing cues. Each set of training stimuli was synthesized to contain 86 unique tokens varying in the two-dimensional space delineated by Pitch (f0) and Vowel Duration (ms), as shown in Figure 3.1.

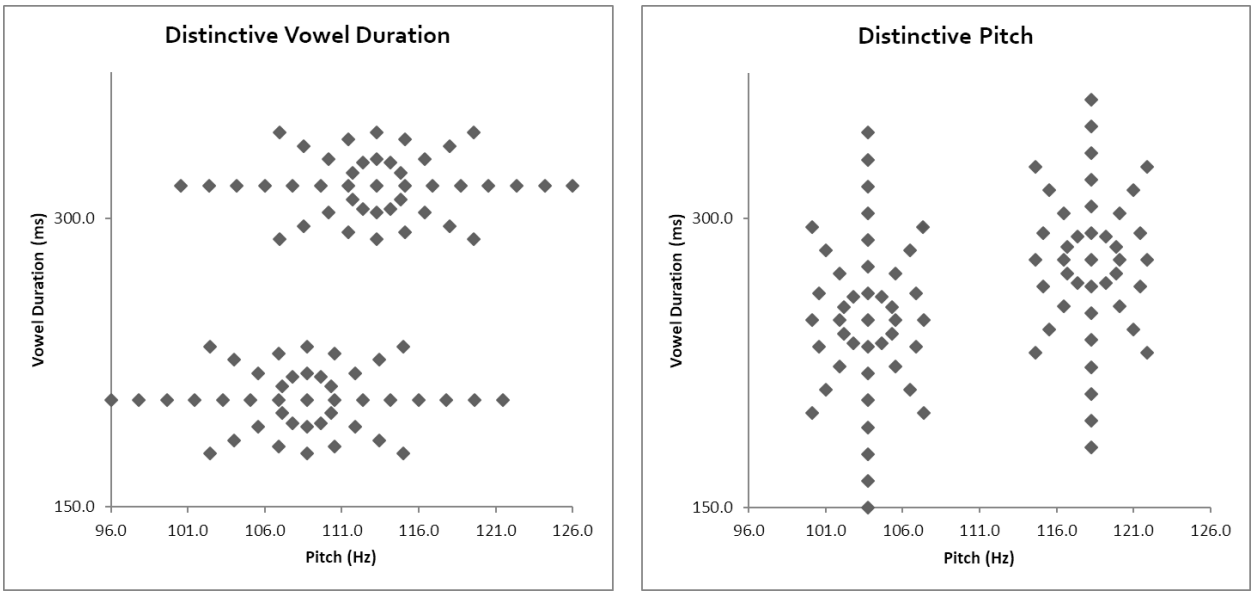

Figure 3.1. Training stimuli: Distinctive Vowel Duration (left) and Distinctive Pitch (right) distributions. Each training stimulus has a duration (ms) value and a pitch (Hz) value, represented by a black point in the two-dimensional space.

Each stimulus token is represented by a point on the graph. In both distributions, Distinctive Pitch (left) and Distinctive Vowel Duration (right), the [tɑ] syllables in the cluster with lower pitch and shorter duration are arbitrarily assigned to Category A and those in the cluster with higher pitch and longer duration are arbitrarily assigned to Category B.

To encourage listeners to attend to one cue over the other, cue informativeness was manipulated such that the optimal categorization would be achieved by attending more to one

cue than another. This was done by controlling the difference in mean values between categories and the range of values within categories. In the Distinctive Pitch training set, there was no overlap between the two categories along the Pitch dimension (within-category range = 2.3 JNDs or 7 Hz, distance between category means = 4.7 JNDs or 14 Hz), whereas along the Distinctive Vowel Duration 93 percent of the tokens in one category had overlapping duration values with tokens in the other category (within-category range = 8.6 JNDs or 252.6 ms for the longer category and 218.6 ms for the shorter category, distance between category means = 1.4 JND or 36.6 ms). In this set, participants should find Pitch to be more informative of the contrast than Vowel Duration, and should therefore give it a higher weight. Similarly, in the Distinctive Vowel Duration training set, there was no overlap between the two categories along the Vowel Duration dimension (within-category range = 2.5 JNDs or 83.5 ms for the longer category and 49.9 ms for the shorter category, distance between category means = 4.9 JNDs or 130.3 ms), whereas along the Pitch range, 93 percent of the tokens in one category had overlapping values with tokens in the other category (within-category range = 8.3 JNDs or 25 Hz, distance between category means = 1.3 JNDs or 4 Hz). Thus in this set, Vowel Duration was more informative of the contrast than Pitch and was therefore expected to get a higher weight.

A set of 50 test stimuli as also created in which Pitch and Vowel Duration varied orthogonally within the same two-dimensional space. They are shown in Figure 3.2.

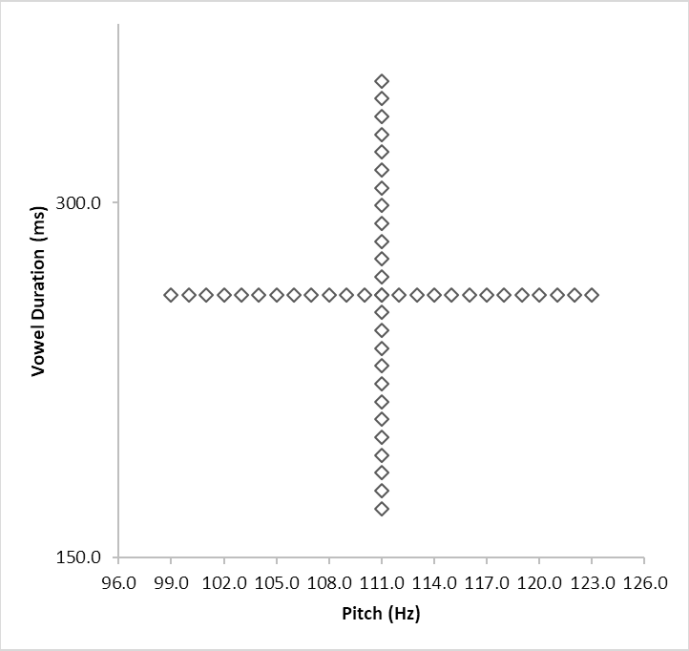

Figure 3.2. Test stimuli for all conditions. Each test stimulus is represented by a point in the two-dimensional space. Vertically arranged points have the same pitch (111 Hz) but vary in vowel duration. Horizontally arranged points have the same vowel duration (250.7 ms) but vary in pitch.

For the vertically arranged points (25 tokens), Pitch was held constant at 111 Hz while Vowel Duration changed in 1/3 JND increments from 165.2 to 380.6 ms. For horizontally arranged points (25 tokens), Vowel Duration was held constant at 250.7 ms, while Pitch was changed in 1/3 JND (1 Hz) increments from 99 to 123 Hz. Since one dimension is always held at a constant value in the middle of the scale where categorization is ambiguous, the category choice made by participants on these tokens should be primarily conditioned by changes along the other dimension. The same set of test stimuli was used to measure cue weights for both the Distinctive Pitch and Distinctive Vowel Duration training sets. Pitch and Vowel Duration values for all training and test tokens can be found in Appendix B.

3.2.2.3. Stimuli synthesis

The 222 unique stimulus tokens – 86 training tokens for the Distinctive Pitch training set, 86 training tokens for the Distinctive Vowel Duration training set, and 50 test tokens – were synthesized using Praat (Boersma & Weenink, 2016). A natural male voice sample of the vowel [ɑ] was used as the base token. From the base, a manipulation object was created, and pitch and duration were changed to a specific value for each token. These [ɑ] vowels were then saved as .wav files and a [t] was spliced onto each token to form the syllable [tɑ].

3.2.3. Procedure

The procedure for these experiments is identical to the procedure for the pitch and vowel duration experiments in Chapter 2. All participants were trained on a Language 1 (L1) and then a Language 2 (L2). This is shown schematically in Figure 3.3.

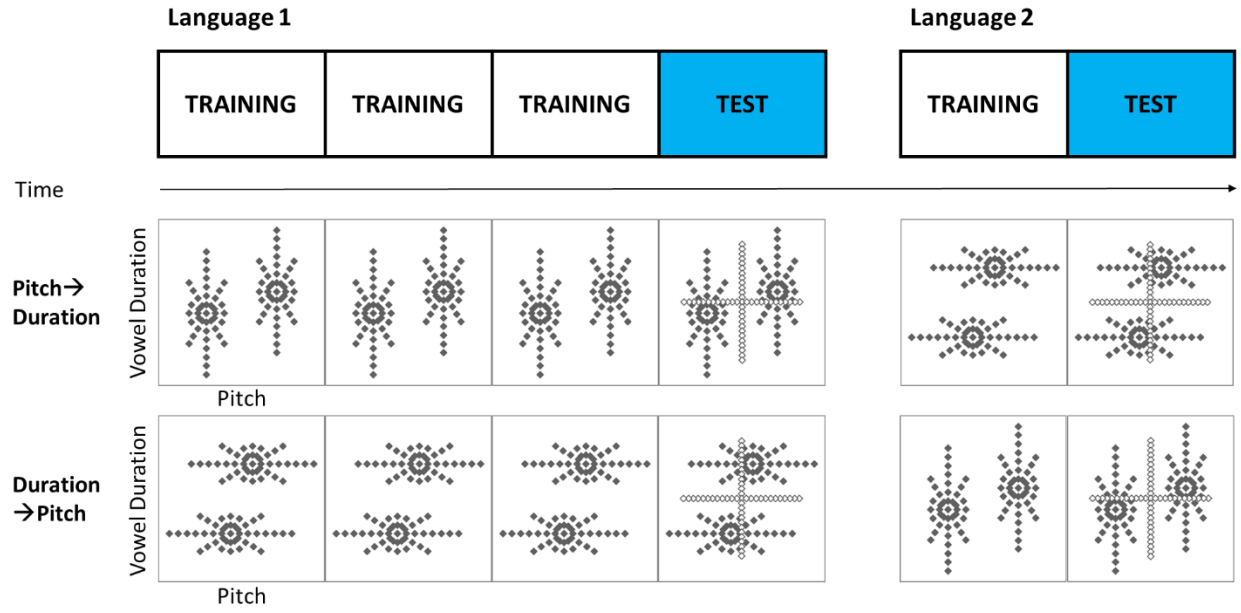

Figure 3.3. Design of the experiment: Design of the experiment: Participants completed training blocks and test blocks in order from left to right (L1: 3 training, 1 test; L2: 1 training, 1 test). Pitch→Duration participants heard the Distinctive Pitch stimuli in L1 and Distinctive Vowel Duration stimuli in L2. Duration→Pitch participants heard the Distinctive Vowel Duration stimuli in L1 and Distinctive Pitch stimuli in L2. Stimuli

presented in each block are displayed for each condition under each block type (black points = training stimuli, white points = test stimuli).

The direction of the shift was counterbalanced such that half of the participants in each language group (English and Hani) were trained and tested on the *Distinctive Pitch* stimulus set as their L1 and the *Distinctive Vowel Duration* set as their L2, while the other half of the listeners were trained and tested on the *Distinctive Vowel Duration* set as their L1 and the *Distinctive Pitch* set as their L2.

In Language 1, all participants heard three blocks of training stimuli each consisting of 86 randomized trials (labeled "Training" in Figure 3.3), then one test block (labeled "Test" in Figure 3.3) which included 136 randomized trials consisting of both training and test stimuli. In Language 2, participants heard one block of new training stimuli, then one block with the same training stimuli plus the test stimuli.

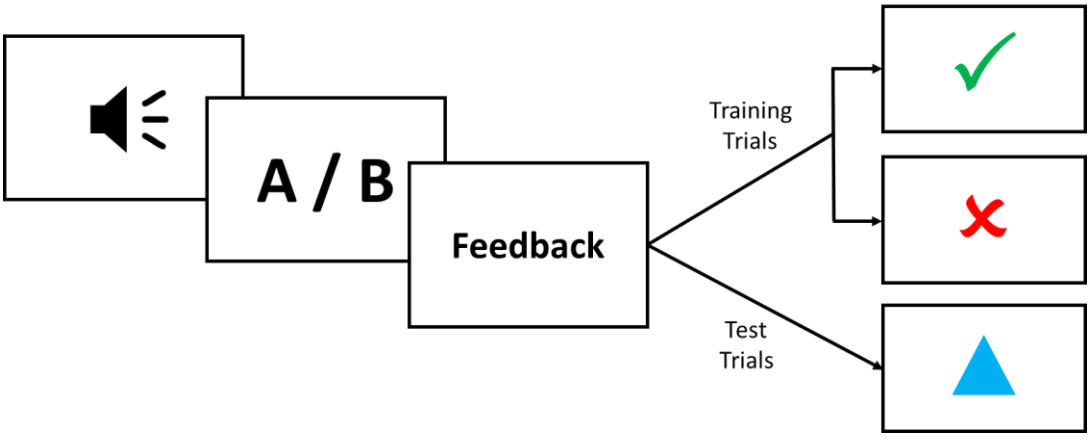

Figure 3.4. Procedure during each trial: Participants heard a sound, made a choice between Category A and Category B, then received feedback: Correct (green check mark), Incorrect (red ex), Unknown (blue triangle).

The sequence of events per trial is given schematically in Figure 3.4. On each trial, participants listened to a single stimulus token and decided which category, A or B, the sound belonged to. After pressing one of the two keys, participants received visual feedback. For

training trials, the feedback informed them whether their response was correct or incorrect. Participants were not told what to listen for. They were instructed to guess at first, then use the feedback to get as many trials correct as possible. During the test blocks at the ends of L1 and L2, participants continued to receive informative feedback on the training trials, but feedback was an uninformative blue triangle for the novel test trials.

English listeners completed the entire process in a quiet lab, unsupervised by an experimenter. Stimuli were presented using the online Appsobabble platform (Tehrani, 2015) and participants listened to the stimuli on 3M Peltor HTB79A-02 headphones. The instructions were given in written form as part of the experimental interface. English listeners gave their responses on a QWERTY keyboard, pressing either the S key if they thought the word they heard was 'sea' (Category A) or the L key if they thought the word was 'land' (Category B). They completed a Background Questionnaire form after completing the experiment. This form asked them for their age and asked them to list out the languages they speak, when they began to learn it, and how fluent they are (beginner, intermediate, functional, or fluent).

Hani listeners were given oral instructions by the experimenter in Mandarin. The experimenter also obtained their background information orally prior to the experiment. The questionnaire for Hani listeners asked for information about their age, the village they are from, and how often they use Hani to communicate in percentages. Before beginning the experiment, they were also given 8 trials for practice, which were different from those used in the experiment itself. Hani listeners did the experiment on touch screen devices and listened to the stimuli on 3M Peltor HTB79A-02 headphones. Rather than selecting an arbitrary key, they touched a picture on the screen, either a rabbit (Category A) or a turtle (Category B), to indicate their choice.

3.2.4. Conditions

The crucial manipulation in this experiment is how the category labels in L1 map onto the category labels in L2, resulting in either a negative mapping between categories or a positive mapping between categories. This is shown in Figure 3.5.

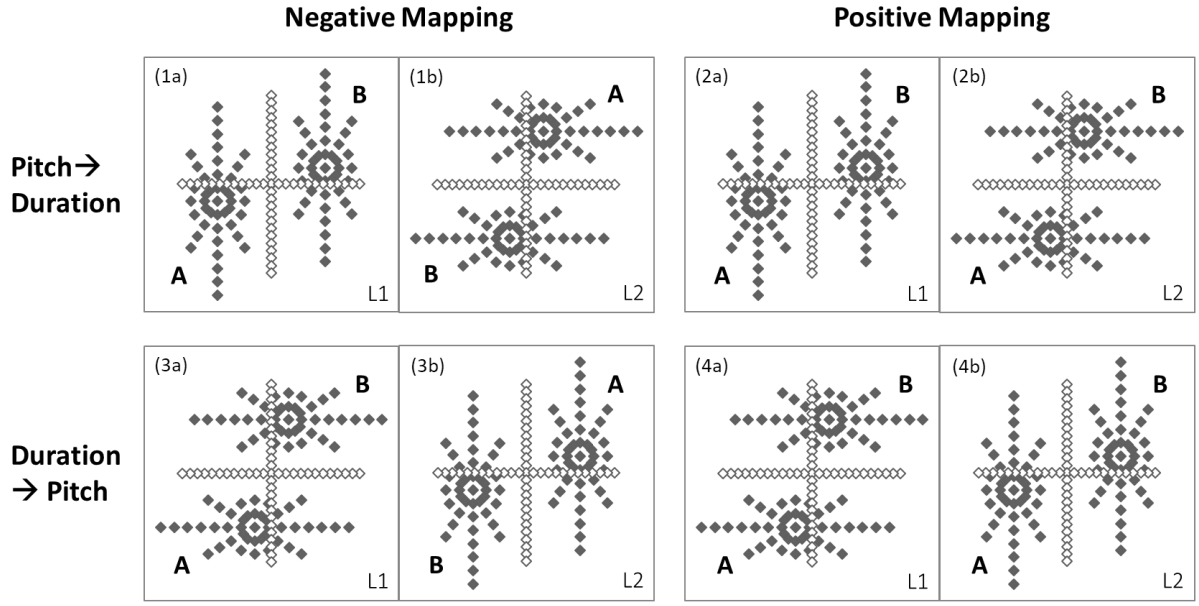

Figure 3.5. Experiment Conditions: Direction (Pitch→Vowel Duration, upper panels vs. Vowel Duration→Pitch, lower panels) x Mapping Relation (Negative, left panels vs. Positive, right panels). Category labels (A or B) are labeled for each language (L1 and L2) in each panel.

In all four conditions, the category labels were the same for L1, the left-most stimulus set in each pair. That is, for L1 in every condition, the category with the relatively low f0 and short duration (bottom left quadrant in each panel) was arbitrarily labeled *A* and the category with relatively high f0 and long duration (top right quadrant in each panel) was labeled *B*. Note that this means the L1s for each condition with the same Direction are identical, providing a built-in replication of the results.

The Negative and Positive Mapping conditions differ only in the labels assigned to the categories in L2. For participants shifting from Pitch to Vowel Duration, a negative mapping

between L1 and L2 labels meant that the category with the lower pitch in L1 had the same label as the category with longer vowel duration in L2, and the category with higher pitch in L1 had the same label as the category with shorter vowel duration in L2. Similarly, for participants shifting from Vowel Duration to Pitch, a negative mapping meant that the category with the shorter vowel duration in L1 had the same label as the category with higher pitch in L2, and the category with the longer vowel duration in L1 had the same label as the category with lower pitch in L2. For the two conditions in which there was a positive mapping, the category with lower pitch in L1 mapped onto the category with shorter vowel duration in L2 and vice versa, and the category with higher pitch in L1 mapped onto the category with longer vowel duration in L2 and vice versa.

The rationale behind the Mapping Relation manipulation is the same as for the experiments in Chapter 2: Participants learn to attribute more weight to the more distinctive cue in L1, and then are forced to transfer the weight onto a different cue in L2. Suppose a participant is trained first on the set of stimuli in which Vowel Duration is more distinctive, and, by the end of training, learns to rely more on the Vowel Duration cue than on the Pitch cue to categorize stimuli. That is, they have learned that shorter tokens belong to Category A, and longer ones belong to Category B. When they are given the new stimulus set in which Pitch is more distinctive, they must shift cue weight onto Pitch in order to be accurate in the categorization task since changes in vowel duration are now less informative. If listeners have experience with a particular co-variation between Vowel Duration and Pitch, the participants will have expectations about how the categories from L1 map onto categories from L2. If these two cues have a *negative* co-variation in their language for example, the participant will expect the category with lower pitch in L2 to have the same label, B, as the longer category in L1.

Similarly, they will expect the category with higher pitch in L2 to have the same label, A, as the shorter category in L1. In this particular case, the category labels in the Negative mapping condition match these expectations, while the category labels in the Positive mapping condition reverse these expectations.

Since pitch and vowel duration are not enhancing, and English listeners do not have experience with their co-variation in a phonemic contrast, the English participants in this experiment have no a priori knowledge about how the cues relate to each other. Thus, their ability to shift attention should not be affected by the particular mapping between category labels. On the other hand, Hani listeners do have experience with the *negative* co-variation between these cues in their language. Thus, they should have difficulty shifting attention between cues when the experimental Mapping Relation is Positive.

Given the Direction and Mapping factors, there are four conditions: Pitch→Vowel Duration, Negative Mapping (Pitch-Negative), Pitch→Vowel Duration, Positive Mapping (Pitch-Positive), Vowel Duration→Pitch, Negative Mapping (Duration-Negative), and Vowel Duration →Pitch, Positive Mapping (Duration-Positive).

3.2.5. Analysis

As with the experiments on pitch and breathiness, performance threshold on the training stimuli in the test block of L1 was used to exclude participants who did not learn to use the distinctive cue in L1 to do the categorization task.

From the English group, 18 participants were excluded for being below this threshold. Including the participants who were excluded for their language background, 54 participants were excluded in total. In the final analysis, there were 30 participants in the Pitch-Negative

condition, 28 participants in the Pitch-Positive condition, 30 participants in the Duration-Negative condition, and 30 participants in the Duration-Positive condition, totalling 118 participants.

From the Hani group, 16 participants were excluded for being below performance threshold on the training trials of the test block in L1. Including those excluded for age, using Hani less than 50% of the time, and experimenter error, a total of 38 participants were excluded from the Hani group. In the final analysis, there were 16 participants in the Pitch-Negative condition, 15 participants in the Pitch-Positive condition, 16 participants in the Duration-Negative condition, and 15 participants in the Duration-Positive condition, totalling 62 participants.

Two pairs of cue weights were obtained for each participant: one weight for each cue, Pitch and Vowel Duration, from L1, and one weight for each cue from L2. The pair of cue weights from each Language was calculated from the test trials in the test block of that Language only. Following Holt and Lotto (2006), a logit binomial regression was run using the listeners' Category Choice on the test trials as the dependent variable and the Pitch and Breathiness values for each test trial as independent predictors. Cue weights were taken as the coefficients of Pitch and Vowel Duration from this logit binomial regression. These coefficients are a measure of how well changes in each dimension, Pitch or Vowel Duration, was able to predict the responses of a participant. For example, if Vowel Duration has a higher coefficient than Pitch, then Vowel Duration is a better predictor of the participant's category choice. The logit binomial regression was implemented in R (R Development Core Team, 2015) using the built-in glm function. The absolute values normalized to sum to one. Note that the normalization of weights does not take into account the accuracy of listeners' categorization, but rather gives a better idea of the relative

contribution of each cue for each listener. These normalized cue weights were the dependent variable in all subsequent analyses.

The normalized cue weights were then analyzed using a mixed effects linear regression model, implemented in R, using the lme4 package (Bates et al., 2008). P-values were obtained from the t-statistic. Pairwise Tukey's HSD post-hoc tests were run using the *lsmeans* package (Lenth, 2016) to identify which pairs were significantly different in significant interactions. P-values from these tests are adjusted for multiple comparisons.

3.3. Results

The results for Hani listeners will be presented first in Section 3.3.1., then the results for English listeners will be presented in Section 3.3.2. For each language group, L1 results will be presented first, showing that the initial learning of cue weights was not different across conditions. This allows differences in cue weights in L2 between conditions to be attributed to experimental manipulations, rather than to differences in initial learning.

3.3.1. Hani

3.3.1.1. Language 1

The normalized cue weights for the Distinctive and Non-Distinctive cues from the test block of L1 are given in Figure 3.6., grouped by condition.

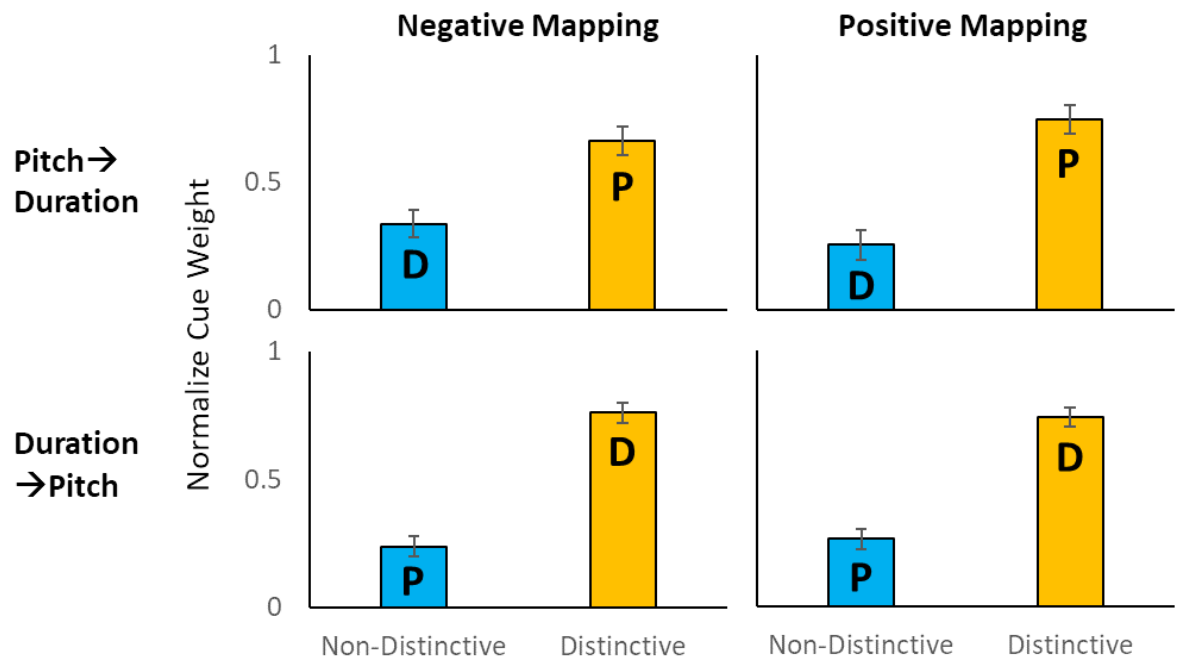

Figure 3.6. Non-enhancing cues: Language 1 Distinctive (yellow) and Non-Distinctive (blue) normalized cue weights for Hani listeners, by Direction (Pitch→Duration vs Duration→Pitch) and Mapping Relation (Negative vs. Positive). The specific cue is labeled as P for Pitch and D for Duration. Error bars = Standard Error.

Figure 3.6. shows that the Distinctive cue (yellow) was weighted higher than the Non-Distinctive cue (blue) in L1 for Hani listeners in every condition. These L1 data were analyzed using a mixed effects model with the same structure as for the English listeners. There was a random intercept of Subject, and fixed effects were Direction (Pitch→Duration vs. Duration→Pitch), Mapping Relation (Negative vs. Positive), and Distinctiveness (Distinctive vs. Non-Distinctive). All 2- and 3-way interactions were also included. The model results are given in Table 3.2.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.24	.05	4.90	<.001	***
Distinctiveness = *Distinctive*	0.52	.07	7.64	<.001	***
Mapping Relation = *Positive*	0.02	.07	0.37	.714	
Direction = *P→D*	0.10	.07	1.47	.140	
Direction × Mapping Relation = *P→D & Positive*	-0.11	.09	-1.14	.254	
Direction × Distinctiveness = *P→D & Distinct.*	-0.20	.10	-2.08	.037	*
Mapping Relation. × Distinctiveness = *Positive & Distinctive*	-0.05	.10	-0.52	.604	
Direction × Distinct. × Mapping Rel. = *P→D & Distinct. & Positive*	0.22	.13	1.61	.106	

Table 3.2. Lmer results from Hani listeners' performance on Language 1 for non-enhancing cues.

There was a significant interaction of Direction × Distinctiveness. A pairwise Tukey's HSD test on this interaction shows that the difference between the Distinctive and Non-Distinctive cues may be greater in the Duration→Pitch conditions (β = 0.50) compared to the Pitch→Duration conditions (β = 0.41). However, both of these differences were highly significant ($p < .001$), thus regardless of the direction in which listeners were shifting cue weights, the new Distinctive cue was weighted higher than the Non-Distinctive cue. Additionally, I ran a pairwise Tukey's HSD test on the full model, and found that the Distinctive cue was weighted significantly higher in each of the four conditions (Pitch-Negative, β = 0.33, $p < .001$; Duration-Negative, β = 0.52, $p < .001$; Pitch-Positive, β = 0.49, $p < .001$; Duration-Positive, β = 0.48, $p < .001$). Again, there was no significant difference between the Distinctive cue weights in different conditions and between the Non-Distinctive cues in different conditions (p-values between .8 and 1.0). Thus, Hani listeners across all conditions weighted the Distinctive cue higher than the Non-Distinctive cue in L1.

3.3.1.2. Language 2

The normalized cue weights for the Distinctive and Non-Distinctive cues from the test block of L2 are given in Figure 3.7., grouped by condition.

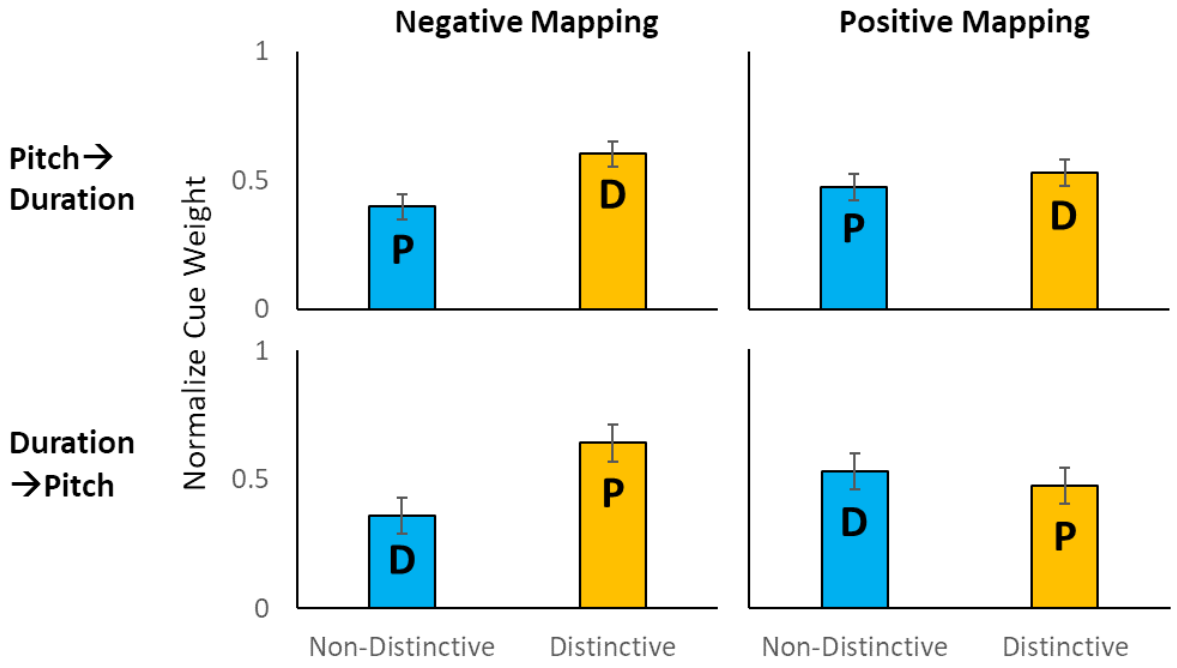

Figure 3.7. Non-enhancing cues: Language 2 Distinctive (yellow) and Non-Distinctive (blue) normalized cue weights for Hani listeners, by Direction (Pitch→Duration vs Duration→Pitch) and Mapping Relation (Negative vs. Positive). The specific cue is labeled as P for Pitch and D for Duration. Error bars = Standard Error.

Again, the distinctiveness of cues was switched. In Figure 3.7., successful cue weight shifting is indicated by a greater weight for the new Distinctive cue (yellow) compared to the new Non-Distinctive cue (blue). For Hani listeners, this was true just in the two Negative mapping conditions. In the Positive mapping conditions, the new Distinctive cue was not weighted higher than the Non-Distinctive cue.

These data were, again, analyzed using a mixed effects logistic regression which included the random intercept of Subject and the fixed effects Direction (Pitch→Duration vs.

Duration→Pitch), Mapping Relation (Negative vs. Positive), Distinctiveness (Distinctive vs. Non-Distinctive), as well as all 2- and 3-way interactions. Since there was a significant interaction of Direction × Distinctiveness in L1, I also wanted to verify that any differences observed in listeners' ability to shift cue weights in L2 was not caused directly by a baseline difference in learning in L1. Thus, this model also included cue weight differences for each pair of cues for each participant from L1 as a covariate. The model results are in Table 3.3.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.36	.06	5.82	<.001	***
L1 Cue Weight Difference	-1.35	0.06	-2.20	1.00	
Distinctiveness = *Distinctive*	0.28	.09	3.20	.001	**
Mapping Relation = *Positive*	0.17	.09	1.95	.052	.
Direction = *P→D*	0.03	.09	0.38	.706	
Direction × Mapping Relation = *P→D & Positive*	-0.09	.12	-0.72	.469	
Direction × Distinctiveness = *P→D & Distinct.*	-0.07	.12	-0.54	.592	
Mapping Relation. × Distinctiveness = *Positive & Distinctive*	-0.34	.12	-2.75	.006	**
Direction × Distinct. × Mapping Rel. = *P→D & Distinct. & Positive*	0.18	.17	1.03	.304	

Table 3.3. Lmer results from Hani listeners' performance on Language 2 for non-enhancing cues.

There was a significant interaction of Mapping Relation × Distinctiveness. A pairwise Tukey's HSD test on this interaction shows that the Distinctive is weighted higher than the Non-Distinctive cue when the mapping is Negative (β = 0.25, p = .001), but not when the mapping is Positive (β = 0.00, p = 1.00). Thus, listeners were able to shift their attention to the newly Distinctive cue only when the category mapping from L1 to L2 was negative.

3.3.2. English

3.3.2.1. Language 1

The normalized cue weights for the Distinctive and Non-Distinctive cues from the test block of L1 are given in Figure 3.8., grouped by condition.

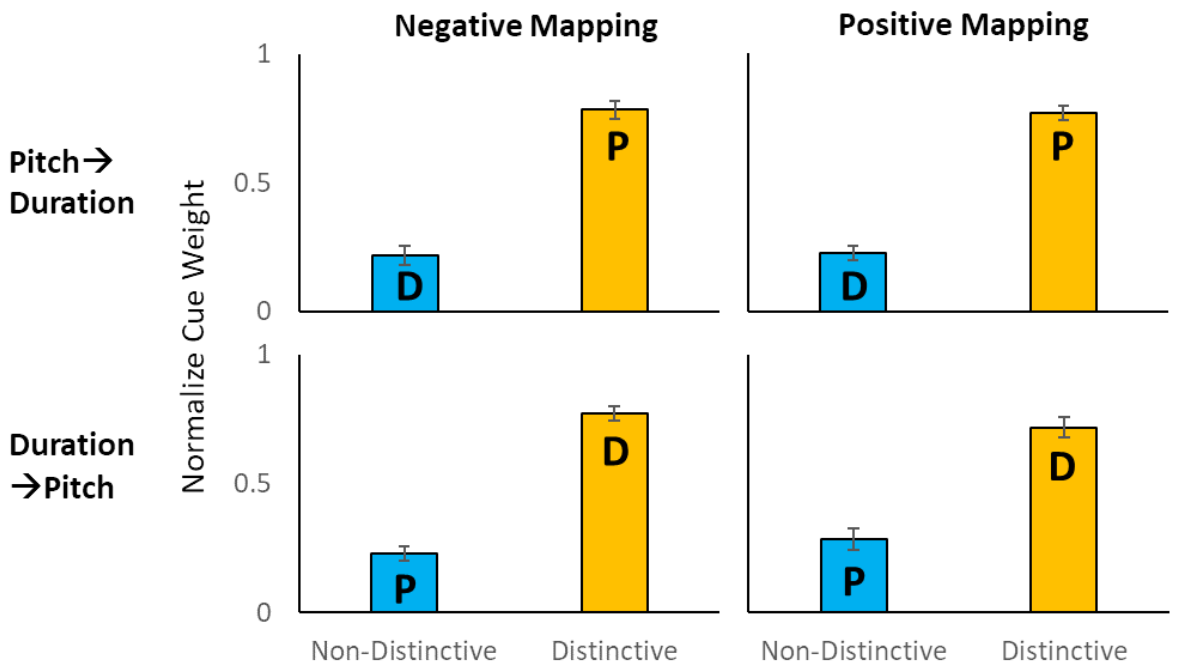

Figure 3.8. Non-enhancing cues: Language 1 Distinctive (yellow) and Non-Distinctive (blue) normalized cue weights for English listeners, by Direction (Pitch→Duration vs Duration→Pitch) and Mapping Relation (Negative vs. Positive). The specific cue is labeled as P for Pitch and D for Duration. Error bars = Standard Error.

Figure 3.8. shows that the Distinctive cue (yellow) was weighted higher than the Non-Distinctive cue (blue) in L1 for English listeners in every condition. Results from the mixed effects model on Language 1 data confirmed this. In addition to the random intercept of Subject, the fixed effects included the between-subjects variables Direction (Pitch→Duration vs. Duration→Pitch) and Mapping Relation (Negative vs. Positive), and the within-subjects variable Distinctiveness (Distinctive vs. Non-Distinctive). This was the highest level of random effects

structure that converged. All 2- and 3-way interactions were also included. The model results are given in Table 3.4.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.23	.04	5.83	<.001	***
Distinctiveness = *Distinctive*	0.54	.06	9.70	<.001	***
Mapping Relation = *Positive*	0.05	.06	0.95	.341	
Direction = *P→D*	-0.01	.06	-0.23	.820	
Direction × Mapping Relation = *P→D & Positive*	0.04	.08	0.49	.621	
Direction × Distinctiveness = *P→D & Distinct.*	0.03	.08	0.32	.748	
Mapping Relation. × Distinctiveness = *Positive & Distinctive*	-0.11	.08	-1.34	.179	
Direction × Distinct. × Mapping Rel. = *P→D & Distinct. & Positive*	-0.08	.11	-0.70	.484	

Table 3.4. Lmer results from English listeners' performance on Language 1 for non-enhancing cues.

Distinctiveness was the only significant fixed effect. A pairwise Tukey's HSD test on the model confirmed that the Distinctive cue was weighted significantly higher than the Non-Distinctive cue in all four conditions (Pitch-Negative, $\beta = 0.57$, $p < .001$; Duration-Negative, $\beta = 0.54$, $p < .001$; Pitch-Positive, $\beta = 0.38$, $p < .001$; Duration-Positive, $\beta = 0.43$, $p < .001$). There was also no significant difference between the Distinctive cue weights in different conditions and between the Non-Distinctive cues in different conditions (p-values between .7 and 1.0). Thus, subjects in all 4 conditions learned to weight the Distinctive cue higher than the Non-Distinctive cue in L1.

3.3.2.2. Language 2

The normalized cue weights for the Distinctive and Non-Distinctive cues from the test block of L2 are given in Figure 3.9., grouped by condition.

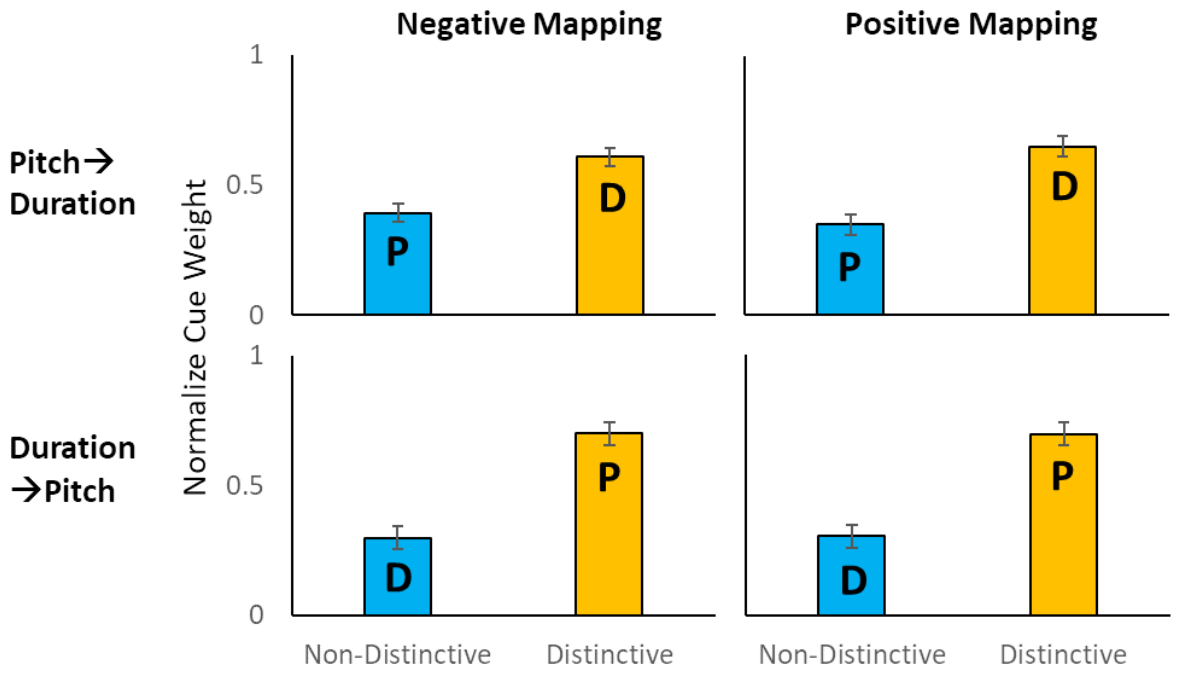

Figure 3.9. Non-enhancing cues: Language 2 Distinctive (yellow) and Non-Distinctive (blue) normalized cue weights for English listeners, by Direction (Pitch→Duration vs. Duration→Pitch) and Mapping Relation (Negative vs. Positive). The specific cue is labeled as P for Pitch and D for Duration. Error bars = Standard Error.

In L2, the distinctiveness of the Pitch and Vowel Duration cues was switched. Vowel Duration was now the Distinctive cue in the Pitch→Duration condition and Pitch the new Distinctive cue in the Duration→Pitch condition. If participants successfully shifted cue weight onto the new Distinctive cue (yellow), then cue weights from the test trials should show a higher weight for Duration and a lower weight for Pitch in the Pitch→Duration conditions, and the opposite weighting in the Duration→Pitch conditions. This was the case across conditions, as confirmed by results from the lmer model.

Again, the model included the random intercept of Subject and the fixed effects included between-subjects variables Direction (Pitch→Duration vs. Duration→Pitch) and Mapping Relation (Negative vs. Positive), and the within-subjects variable Distinctiveness (Distinctive vs.

Non-Distinctive). All 2- and 3-way interactions were also included. The model results are in Table 3.5.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.30	.04	6.85	<.001	***
Distinctiveness = *Distinctive*	0.40	.06	6.55	<.001	***
Mapping Relation = *Positive*	0.01	.06	0.09	.931	
Direction = *P→D*	0.09	.06	1.50	.132	
Direction × Mapping Relation = *P→D & Positive*	-0.05	.09	-0.55	.580	
Direction × Distinctiveness = *P→D & Distinct.*	-0.19	.09	-2.13	.034	*
Mapping Relation. × Distinctiveness = *Positive & Distinctive*	-0.01	.09	-0.12	.903	
Direction × Distinct. × Mapping Rel. = *P→D & Distinct. & Positive*	0.09	.12	0.78	.434	

Table 3.5. Lmer results from English listeners' performance on Language 2 for non-enhancing cues.

There was a significant interaction of Direction × Distinctiveness. A pairwise Tukey's HSD test on this interaction shows that the difference between the Distinctive and Non-Distinctive cue may be greater in the Duration→Pitch conditions (β = 0.40) compared to the Pitch→Duration conditions (β = 0.26). However, both of these differences were highly significant (p <.001), thus regardless of the direction in which listeners were shifting cue weights, the new Distinctive cue was weighted higher than the Non-Distinctive cue. Additionally, I ran a pairwise Tukey's HSD test on the full model, and found that the Distinctive cue was weighted significantly higher in each of the four conditions (Pitch-Negative, β = 0.21, p = .022; Duration-Negative, β = 0.40, p < .001; Pitch-Positive, β = 0.30, p < .001; Duration-Positive, β = 0.39, p < .001). Again, there was no significant difference between the Distinctive cue weights in different conditions and between the Non-Distinctive cues in different conditions (p-values between .8 and 1.0). Thus,

English listeners across all conditions weighted the Distinctive cue higher than the Non-Distinctive cue in L2 as well, showing that they were able to successfully shift their attention. Crucially, there was no difference between listeners' ability to shift cue weights onto the newly Distinctive cue whether the mapping was Negative or Positive.

3.4. Discussion

In this chapter, I used a cue weighting experiments to test whether the enhancement effects found in the previous chapter could be due to perceptual integration only. Listeners shifted attention between two acoustic phonetic cues, pitch and vowel duration, that do not have the same auditory effect but are shown to be perceptually integral. Listeners from the same two language groups as the previous experiments were recruited. Hani listeners had native language experience with the co-variation between pitch and vowel duration as cues to a single contrast, while English listeners did not. The crucial manipulation was the mapping between category labels from the first to the second experimental distribution. The mapping was either positive, which neither group of listeners have experience with, or negative, which the Hani listeners have experience with but not the English listeners. The direction of shifting, from pitch to vowel duration or vice versa, was counterbalanced.

Under the premise that non-enhancing cues would not be perceptually dependent in signalling a contrast, the hypothesis was that listeners' behaviour would be entirely determined by their experience with these cues in their native language. English listeners were predicted to be able to shift attention from one cue to the other regardless of the mapping relation between category labels given their inexperience with either co-variation pattern. Hani listeners were predicted to be able to shift attention only when the mapping matched the negative co-variation

they are experienced with, but not when the mapping was the reverse. The results confirm these predictions: English listeners' cue weighting in the second experimental language, after the attentional shift was induced, showed a higher weight for the newly distinctive cue in all conditions. Hani listeners' cue weights at the end of the experiment differed depending on the mapping condition. When the mapping was negative, the cue weights for the newly distinctive cue was higher than the non-distinctive cue, matching the relative distributional informativeness of the two cues in the second experimental language. In contrast, when the mapping was positive, there was no difference between the weights of the distinctive and non-distinctive cues, showing that listeners were unable to attend more to the newly distinctive cue, even though it was more informative of the contrast in the second distribution.

Since inexperienced English listeners were able to shift attention in both mapping conditions, the language-general integration between pitch and vowel duration seems not to have affected their behaviour. If integration played a role, English listeners should have had more difficulty shifting attention either when the mapping was positive or when it was negative. This alternative prediction was not borne out. The results from the previous chapter and this chapter provide evidence that enhancement and perceptual integration are distinct notions. What has yet to be addressed is why the enhancement effects observed in Chapter 2 were asymmetric. That is, why, in the non-enhancing condition, listeners only had difficulty shifting cue weight from breathiness to pitch but not vice versa. This will be further explored in Chapter 4.

CHAPTER 4: PERCEPTUAL ASYMMETRY

4.1. Introduction

In Chapters 2 and 3, I demonstrated that listeners shifting attention between enhancing cues, pitch and breathiness, behaved differently from listeners shifting between non-enhancing cues, pitch and vowel duration. As predicted by the Auditory Enhancement theory, even English listeners who do not have experience with the co-variation between enhancing cues were unable to shift cue weights when the mapping between the first and second experimental language reversed the enhancing relationship between the cues. However, the experiment produced the unexpected result that for both language groups, English and Hani, the difficulty in shifting attention was unidirectional, only occurring when shifting from the breathiness cue to the pitch cue, but not the reverse. This directional asymmetry will be the focus of this chapter.

Under the Auditory Enhancement account, two cues are enhancing if they contribute to the same auditory effect. Such cues form an IPP (intermediate perceptual property) (Kingston & Diehl, 1994) and thus either cue can be recruited to enhance the percept of the other in signalling a contrast. If two such cues are equated for perceptibility and salience, then they should have symmetric perceptual dependencies. For the experiments presented in Chapter 2, this would predict that listeners would have difficulty shifting attention from breathiness to pitch as well as from pitch to breathiness, when the mapping relation is non-enhancing.

Why, then, was directional asymmetry observed? The obvious culprit is that the distinctive cue was not learned equally well by listeners initially trained on breathiness and those initially trained on pitch. If breathiness was learned better as a distinctive cue, then shifting weight away from breathiness could have been more difficult than shifting weight away from pitch. Recall that I controlled for this by equating the perceptibility of the cues in the design of

the experiment. Specifically, the scales of each cue were equated by using the same number of JNDs measured for English listeners. So, for listeners in the English group at least, changes along the pitch dimension should not be easier to perceive compared to changes along the breathiness dimension and vice versa. Therefore, the two cues should be equally easy or difficult to learn. To ensure that all listeners were weighting pitch and breathiness the same before I induced the attentional shift to the other cue, I first compared the relative weighting of the distinctive and non-distinctive cues in the first experimental language. Indeed, the results showed that for both language groups, English and Hani, the weightings were not different, whether they were training on the distribution that favours pitch or breathiness.

Second, the asymmetry could have arisen because listeners' have unequal experience with the cues in their native language. I controlled for this by selecting groups of listeners for having or not having *phonemic* experience with both or neither cues. However, these cues are still present in the signal and are used to differing degrees for different purposes. One could, for example, argue that English listeners have more linguistic experience with pitch, which is used to signal phrase-level meaning differences (i.e. intonation), than breathiness, which is used to signal paralinguistic information such as gender (Klatt & Klatt, 1990; Mullenix et al., 1995), attractiveness (Babel et al., 2014), valence of new information (Freese & Maynard, 1998), etc. This could lead English speakers to give a higher weight to pitch in a linguistic task despite equal perceptibility of both cues. The same argument could be made for Hani, which, in addition to having a tense-lax contrast in which pitch co-varies with breathiness, also has a tone contrast, in which pitch is the primary cue. To rule out the possibility that the directional asymmetry was caused by unequal experience, I conducted and report in this chapter a study on two more groups of listeners who have distinct language experience with either pitch or breathiness, using the

same cue shifting paradigm. As we will see, these new language groups differ in their initial weighting of the two cues, but both groups exhibit the same directional asymmetry.

The universality of this asymmetry will be addressed in Section 4.3, where I propose that all of these results can be accounted for by a perceptual asymmetry in the dependency between the cues. That is, I propose that pitch is perceived independently of breathiness, but breathiness is not perceived independently of pitch, so listeners shifting their attention from breathiness to pitch experience interference, while listeners shifting from pitch to breathiness can treat the latter as a novel cue. This proposal makes specific hypotheses about cue weighting, which I then experimentally test in this same chapter.

4.2. Experiment I: Directional asymmetry, a cross-linguistic phenomena

The purpose of Experiment I was test whether the directional asymmetry observed in Chapter 2 was due to language-specific factors. To do this, I tested two additional groups of listeners with distinct differences in their exposure to pitch and breathiness as cues to phonemic contrast. These included i) a tone language group, for whom pitch is used as the primary cue to contrast word meanings but breathiness is not, and ii) a phonation language group, for whom breathiness is used as the primary cue to word contrasts, but pitch is not. Since for both English and Hani listeners the directional asymmetry was observed only in the Non-Enhancing mapping condition, I only test the listeners in these new language groups in this condition. If the asymmetric enhancement effect was caused by listeners' language experience, the tone group is expected to perform like English listeners, given their extensive experience with pitch, whereas the phonation group is expected to have the opposite directional asymmetry. The methods are described in detail below.

4.2.1. Methods

4.2.1.1. Participants

44 participants (age 18-39) were recruited at UCLA for the Tone group and were either given course credit through the Subject Pool or paid for their participation. These participants were native speakers of Vietnamese or one or more dialects of Chinese and had no experience with languages that use phonation as a primary cue to a phonemic contrast, as self-reported in the Language Background Questionnaire. Note that the Vietnamese participants most likely speak the southern variety, in which phonation cues play a very minor role at best in the perception of tones (Brunelle, 2009). Five subjects were excluded for being self-reported non-fluent speakers, and seven subjects were excluded for not completing the study.

32 participants (age 18-27) were recruited at UCLA and the University of Sourthern California for the Phonation group and paid for their participation. These participants were all native speakers of Gujarati, who have been shown to be sensitive to H1-H2 as a cue for breathiness (Bickley, 1982; Esposito, 2006). These participants had no experience with languages that use pitch as a primary cue to a phonemic contrast as self-reported in the Language Background Questionnaire. One subject was excluded because technical difficulties occurred during the experiment.

All participants in the Tone and Phonation groups also speak English at varying proficiencies.

4.2.1.2. Stimuli and procedure

The stimuli from the experiments in Chapter 2 were used. There were two distributions of stimuli, Distinctive Breathiness and Distinctive Pitch. These are shown schematically in Figure 4.1. (adapted from Figures 2.2. and 2.3. in Chapter 2).

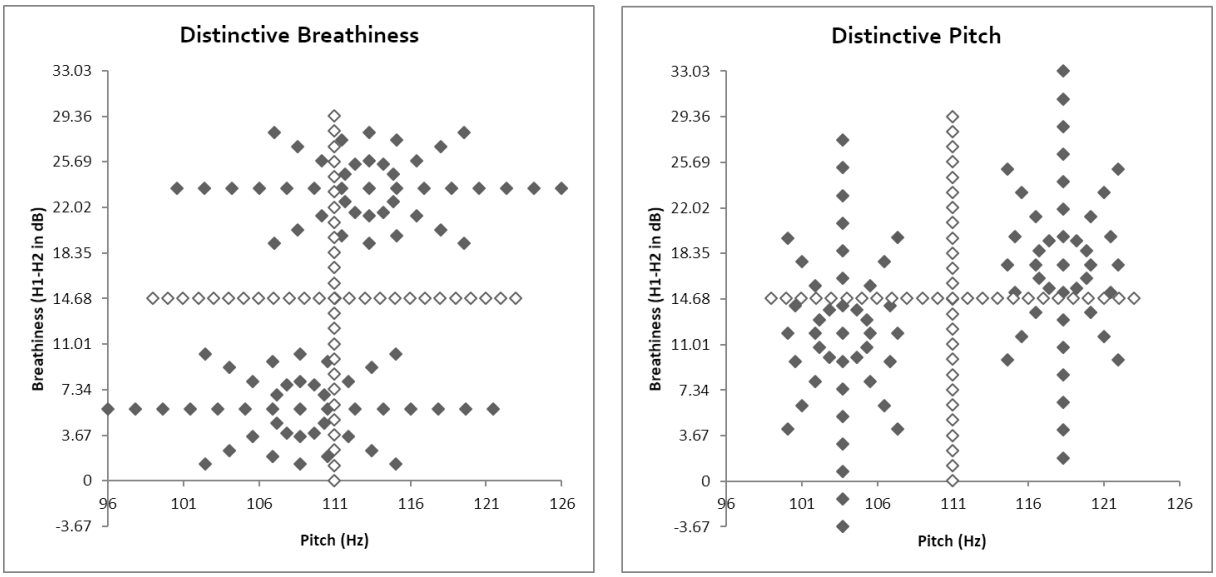

Figure 4.1. Training stimuli (black) and test stimuli (white) in Distinctive Breathiness (left) and Distinctive Pitch (right) distributions. Each stimuli has a breathiness (H1-H2 in dB) value and a pitch (Hz) value in the two-dimensional space.

The 86 training stimuli (black) in each distribution form two categories, one with lower pitch and lower H1-H2 values, and the other with higher pitch and higher H1-H2 values. In the Distinctive Breathiness distribution, the two categories are well-separated along the breathiness dimension but overlapped along the pitch dimension. In the Distinctive Pitch distribution, the two categories are well-separated along the pitch dimension but overlapped along the breathiness dimension. The test stimuli (white) in both distributions, composed of 50 stimuli, are identical. 25 of these stimuli are held at a constant average pitch of 111 Hz but they vary along the breathiness dimension from 0 to 29.36 dB by increments of 1/3 JND (~1.22 dB). The other 25 test stimuli are held at constant average H1-H2 of 14.68 dB but they vary along the pitch dimension from 99

to 123 Hz by increments of 1/3 JND (1 Hz). See Chapter 2 for scaling and synthesis of the stimuli.

The procedures were also identical to those used in the previous experiments. All participants were trained on a Language 1, which had either the Distinctive Breathiness or the Distinctive Pitch distribution, then on a Language 2, which had the other distribution. Figure 4.2. given below is replicated from Chapter 2.

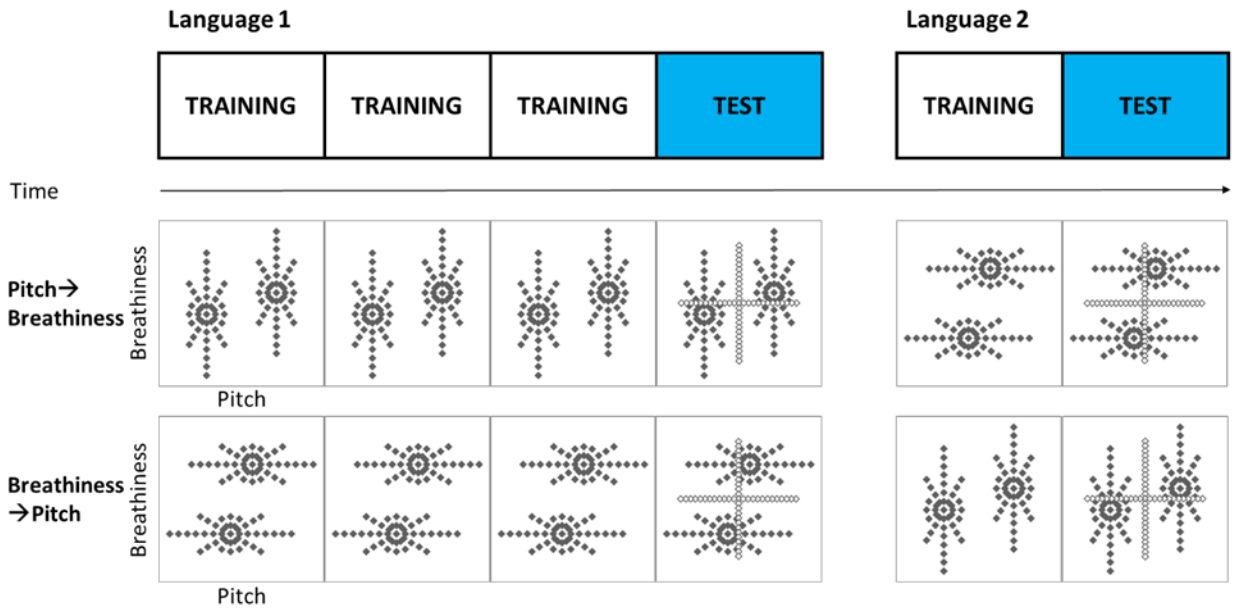

Figure 4.2. Overall procedure: Participants completed training blocks and test blocks in order from left to right (L1: 3 training, 1 test; L2: 1 training, 1 test). Direction of shifting was counterbalanced between Pitch→Breathiness and Breathiness→Pitch.

As in the previous experiments, all participants completed three training blocks (training stimuli only) and one test block (training stimuli and test stimuli) in L1, then one training block and one test block in L2.

The procedure for each trial is schematized in Figure 4.3., replicated from Chapter 2.

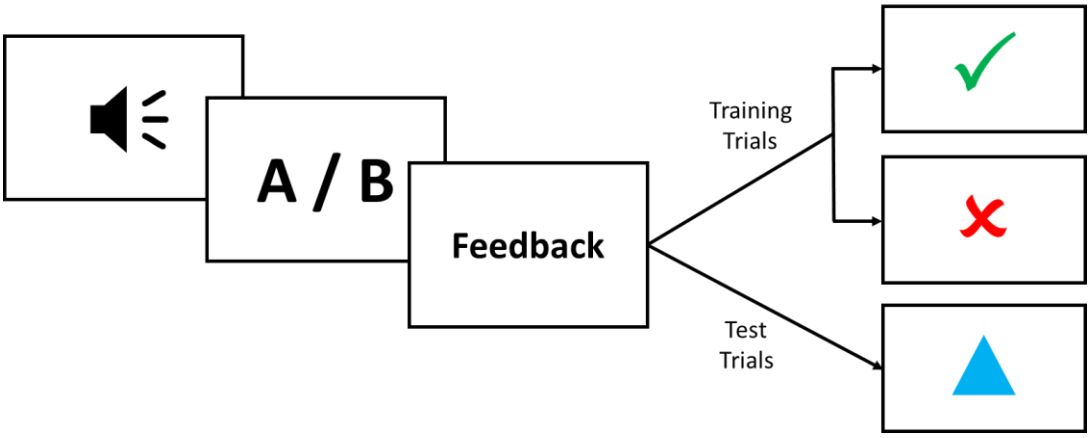

Figure 4.3. Procedure during each trial: Participants heard a sound, made a choice between Category A and Category B, then received feedback: Correct (green check mark), Incorrect (red ex), Unknown (blue triangle).

On each trial, participants listened to a single stimulus token and decided which category, A or B, the sound belonged to. After pressing one of the two keys, participants received visual feedback, which either informed them of whether their response was correct (training trials) or did not provide this information (test trials). They were instructed to guess at first, then use the feedback to get as many trials correct as possible.

Listeners from both the Tone and Phonation groups were tested in quiet spaces on university campuses. The equipment and presentation methods used for them were identical to those used for English listeners.

4.2.1.3. Conditions

All participants were assigned to the Non-Enhancing condition in which the mapping relation of the category labels from L1 to L2 is *positive*, the reverse of the enhancing co-variation between pitch and breathiness. This is schematized below in Figure 4.4.

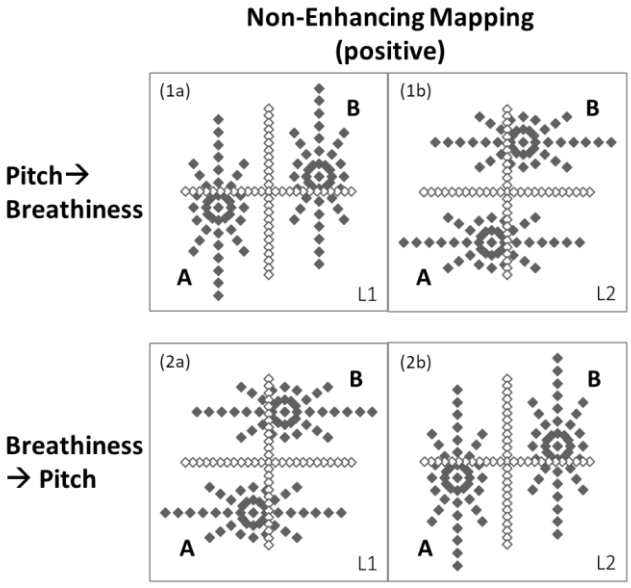

Figure 4.4. Non-Enhancing Experimental Conditions: Pitch→Breathiness (upper panel) and Breathiness→Pitch (lower panels)

Half of the participants in each language group were in the Pitch→Breathiness condition and the other half were in the Breathiness→Pitch condition. In the Pitch→Breathiness condition, the L1 category with lower pitch had the same label as the L2 category with lower H1-H2, and the L1 category with higher pitch had the same label as the L2 category with higher H1-H2. In the Breathiness→Pitch condition, the L1 category with lower H1-H2 had the same category label as the L2 category with lower pitch, and the L1 category with higher H1-H2 had the same label as the L2 category with higher pitch.

4.2.1.4. Analysis

Participants were excluded if the number of correctly categorized training stimuli in the L1 test block was not above chance (see Analysis in Chapter 2). We took this to mean that they had not learned to use the distinctive cue in the categorization task. One subject from the Tone group was excluded for being below this performance threshold. 31 participants in this group

were included in the final analysis, 16 in the Pitch→Breathiness condition and 15 in the Breathiness→Pitch condition. Seven participants from the Phonation group were excluded for being below the performance threshold. 24 participants from this group were included in the final analysis, 12 in the Pitch→Breathiness condition and 12 in the Breathiness→Pitch condition.

Two pairs of normalized cue weights, one from L1 and one from L2, were obtained from each participant (see Analysis in Chapter 2). The cue weights of all participants for each experimental language was then analyzed using a mixed effects linear regression model, implemented in R, using the *lme4* package (Bates et al., 2008). P-values were obtained from the t-statistic. Pairwise Tukey's HSD post-hoc tests were run using the *lsmeans* package (Lenth, 2016) to identify which pairs were significantly different in significant interactions. P-values from these tests are adjusted for multiple comparisons.

4.2.2. Results

The L1 results for both the Tone and Phonation groups will be presented first, followed by the L2 results.

4.2.2.1. Language 1

Figure 4.5. shows the normalized cue weights for L1 in all conditions, separated by Language Group and Direction.

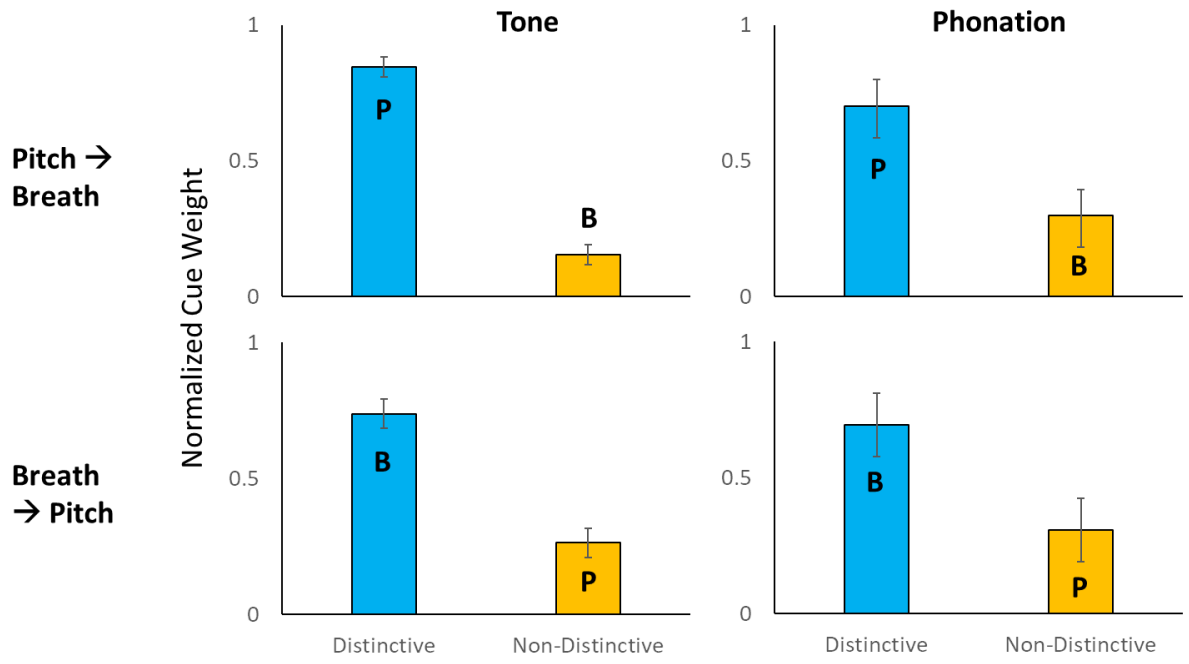

Figure 4.5. L1 Distinctive (blue) and Non-Distinctive (yellow) normalized cue weights by Direction (Pitch→Breathiness, upper panels vs. Breathiness→Pitch, lower panels) and Language Group (Tone vs. Phonation). All participants were tested on the Non-Enhancing Condition.

Though in all conditions, the Distinctive cue is weighted higher than the Non-Distinctive cue, the difference is smaller for the Phonation group when Pitch is Distinctive and Breathiness is Non-Distinctive. This is confirmed by the results from the lmer model. The model included the random intercept of Subject and the fixed effects included between-subjects variables Direction (Pitch→Breathiness vs. Breathiness→Pitch), Language Group (Tone vs. Phonation), and Distinctiveness (Distinctive vs. Non-Distinctive). All 2- and 3-way interactions were included. The results are given in Table 4.1.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.22	.05	4.12	<.001	***
Distinctiveness = *Distinctive*	0.56	.07	7.54	<.001	***
Language Group = *Tone*	0.04	.07	-0.63	.530	
Direction = *P→B*	0.17	.07	2.22	.027	*
Direction × Group = *P→B & Tone*	-0.27	.10	-2.76	.006	**
Direction × Distinct. = *P→B & Distinct.*	-0.33	.11	-3.14	.002	**
Group × Distinct. = *Tone & Distinct.*	-0.09	.10	-0.89	.375	
Direction × Distinct. × Group = *P→B & Distinct. & Tone*	0.55	.14	3.90	<.001	***

Table 4.1. Lmer results from Tone and Phonation group listeners' performance on Language 1.

Though there was a significant main effect of Distinctiveness, showing that the Distinctive cue was weighted higher than the Non-Distinctive cue overall, there were also significant two-way interactions between Distinctiveness and Direction, Direction and Language Group, as well as a significant three-way interaction between Distinctiveness, Direction and Group. A pairwise Tukey's HSD test on the three-way interaction showed that the Distinctive cue was weighted significantly higher than the Non-Distinctive cue in all conditions, (P-primary – Tone, $\beta = 0.69$, $p < .001$; B-primary – Tone, $\beta = 0.47$, $p < .001$; P-primary – Phonation condition, $\beta = 0.23$, $p = .040$; B-primary – Phonation, $\beta = 0.56$, $p < .001$). The 3-way interaction likely stems from the smaller effect in the P-primary – Phonation condition. Given a significant 3-way interaction, we took the cue weight difference (Distinctive weight – Non-Distinctive weight) for each participant in L1 and included this as a covariate in the mixed effects model for Language 2 (see below), to control for initial differences in the learning of L1.

4.2.2.2. Language 2

Figure 4.6. shows the normalized cue weights for L2 in all conditions, separated by Language Group and Direction.

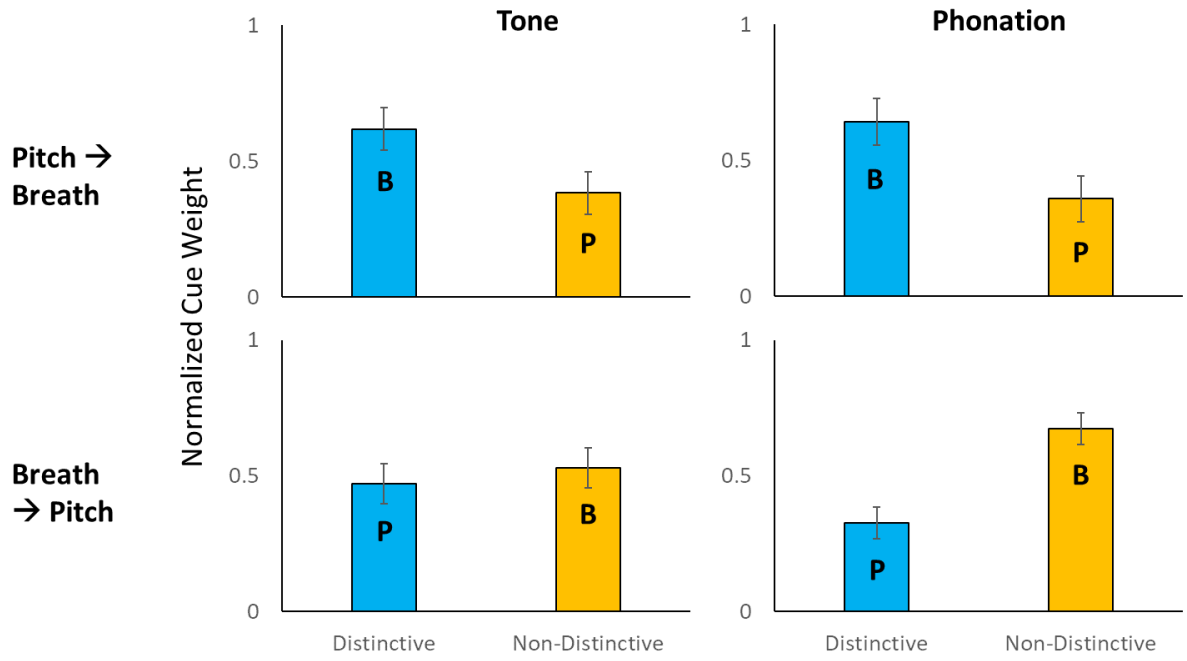

Figure 4.6. L2 Distinctive (blue) and Non-Distinctive (yellow) normalized cue weights by Direction (Pitch→Breathiness, upper panels vs. Breathiness→Pitch, lower panels) and Language Group (Tone vs. Phonation). All participants were tested on the Non-Enhancing Condition.

In Language 2 as well, successful learning of the new distribution is indicated by a higher weight for the Distinctive cue and a lower weight for the Non-Distinctive cue. This was observed in the two language groups when they were shifting from Pitch to Breathiness, but not when they were shifting from Breathiness to Pitch. The linear mixed effects model for Language 2 included the random intercept of Subject, cue weight differences from Language 1 as a covariate, the fixed effects of Direction (Pitch→Breathiness vs. Breathiness→Pitch), Language Group (Tone vs. Phonation), and Distinctiveness (Distinctive vs. Non-Distinctive), as well as all 2-way and 3-way interactions between the fixed effects. The results are in Table 4.2.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.67	.08	8.30	<.001	***
L1 Cue Weight Difference	1.22	.08	0.00	1.00	
Distinctiveness = *Distinctive*	-0.35	.12	-3.03	.002	**
Language Group = *Tone*	-0.14	.11	-1.32	.188	
Direction = *P→B*	-0.32	.12	-2.67	.008	**
Direction × Group = *P→B & Tone*	0.17	.16	1.10	.292	
Direction × Distinct. = *P→B & Distinct.*	0.63	.16	3.87	<.001	***
Group × Distinct. = *Tone & Distinct.*	0.29	.15	1.86	.062	
Direction × Distinct. × Group = *P→B & Distinct. & Tone*	-0.34	.22	-1.54	.123	

Table 4.2. Lmer results from Tone and Phonation group listeners' performance on Language 2.

Results show that there was a significant interaction between Direction and Distinctiveness, β = 0.63, $p < .001$. A pairwise Tukey's HSD test showed that in the Pitch→Breathiness conditions, the Distinctive cue was weighted higher than the Non-Distinctive cue (β = 0.26, p = .004), but that in the Breathiness→Pitch conditions, the Distinctive cue was weighted *lower* than the Non-Distinctive cue (β = 0.20, p = .041), indicating that participants, regardless of language background, were unable to shift cue weights from Breathiness onto Pitch. This effect was largely driven by the cue weight difference in the Phonation group, though the difference between the Tone and Phonation group was not significant in the three-way interaction (β = -0.34, p = 0.123).

The lack of distinction between Language Groups was somewhat surprising given the visual discrepancy between them in Figure 4.6. Thus, the model was run once more after excluding one participant from the Phonation group in the Breathiness→Pitch condition who had

the opposite cue weighting (Pitch > Breathiness) in L2. Crucially, in this model, the interaction between Direction and Distinctiveness was still significant ($\beta = 0.71$, $p < .001$), confirming the cue shifting asymmetry. New in this model was the significant interaction between Group and Distinctiveness ($\beta = 0.37$, $p = .020$). A Tukey's HSD test showed that this effect was driven by the fact that in the Tone group, the Distinctive cue was numerically higher than the Non-Distinctive cue, but in the Phonation group, the Non-Distinctive cue was numerically higher than the Distinctive cue. However, neither of these differences were significant. Thus, overall, listeners in either the Tone or the Phonation group, like the English listeners, were unable to shift cue weights in the Non-Enhancing, Breathiness→Pitch condition.

4.2.3. Discussion

The current experiment was designed to rule out the possibility that more linguistic experience with either pitch or breathiness could explain the asymmetry. For this, two additional groups of listeners were tested, speakers of lexical tone languages, which use pitch as a primary cue, and speakers of Gujarati, a language that uses breathiness contrastively. This experiment replicated the findings from Chapter 2. Like English and Hani participants, listeners of either a tone language or a language where breathiness is phonemic successfully learned to use either Pitch or Breathiness as a primary cue when trained on Language 1, though the difference between cues was smaller for Gujarati listeners in the Pitch→Breathiness condition. Then, like English and Hani participants, these listeners also failed to shift cue weight from Breathiness to Pitch, when the enhancing relationship was reversed. Though not significantly different, Gujarati listeners nevertheless seemed to behave differently from the English, Hani, and Tone groups when shifting cue weights from breathiness to pitch with the enhancing relationship reversed.

English, Hani, and Tone language listeners were able to re-weight cues to some extent, but were unable to shift enough weight to Pitch such that it became the primary cue. In comparison, Gujarati listeners seem not to have shifted weights at all, maintaining a higher cue weight for Breathiness and a lower cue weight for Pitch despite the distributional evidence that Pitch is more informative.

Nonetheless, the pattern for all four groups of listeners – one that uses neither pitch nor breathiness phonemically (English), one in which pitch and breathiness co-vary in signalling a contrast (Hani), one in which pitch is a primary cue to a contrast but breathiness is not (Tone), and one in which breathiness is a primary cue to a contrast but pitch is not (Phonation) – is consistent. All of these groups exhibit the same directional asymmetry when shifting cue weights between pitch and breathiness: when the enhancing relationship between the cues is reversed, all listeners have difficulty shifting from breathiness to pitch but not from pitch to breathiness.

The uniformity of the behaviour across these languages makes it difficult to explain the directional asymmetry as an effect of language experience. If English and Hani listeners both use pitch more heavily than breathiness, then we can expect the Tone group to exhibit the same pattern of asymmetry. Our results confirm this. However, Gujarati listeners who rely on breathiness to distinguish a native contrast, also showed the same asymmetry. The identical pattern of asymmetry of listeners is difficult to reconcile with the idea that either the enhancement effect or the directional asymmetry can be attributed to language experience alone.

One could argue that Gujarati listeners' inability to shift weight from breathiness onto pitch could have been due to the difficulty of the task and/or insufficient training in conjunction with their language experience. That is, rather than using an unfamiliar cue, pitch, to learn a new mapping between stimuli and category labels, these listeners may have simply found it easier to

keep using a cue they are familiar with in their native language. And we see some evidence for this – the smallest difference between the Distinctive and Non-Distinctive cue weights in Language 1 was for Gujarati listeners when they were trained on the Pitch-first condition. With more training, they may well have promoted Pitch to the same extent as the Tone language listeners. A similar argument can also be made for the critical condition in Language 2. With more training on the second artificial language in which pitch is more distinctive, Gujarati listeners may well have learned to shift cue weights onto pitch. In other words, Gujarati listeners' language experience alone could explain their difficulty in shifting weights onto pitch.

However, results from the Tone group do not support this interpretation. Given the same task difficulty, the same training, and their native advantage with pitch, we would then expect these listeners to have difficulty shifting from pitch to breathiness but not vice versa. Instead, we found the opposite result: this group of listeners also could promote breathiness but could not promote pitch to a primary cue with a reversal of the enhancing cue relationship, much like the Gujarati listeners.

Rather than language experience, I propose that the reoccurring asymmetric enhancement effect we have observed in these studies is language-general and rooted in perception. That is, these effects emerge because listeners perceive pitch relatively independently of breathiness, but fail to perceive breathiness independently of pitch. Thus, if pitch is learned as the primary cue in Language 1, listeners are able to treat salient breathiness in Language 2 as novel, and learning the distribution of the pitch cue does not interfere with learning the distribution of the breathiness cue because the percept of pitch is not strongly tied to the percept of breathiness. Importantly, the independence of the perception of pitch from the perception of breathiness predicts that any mapping relation, positive or negative, should be learnable. This is indeed what was observed for

all listeners shifting from pitch to breathiness. Conversely, when breathiness is learned as the primary cue in Language 1, the listener's familiarity with this cue is tightly coupled with pitch (breathier voice being coupled with lower pitch, and less breathy voice being coupled with higher pitch). When this enhancing relationship is respected in the cue-shift, the transfer of cue weights is facilitated, and when this enhancing correlation is disrupted, the transfer of cue weights is hindered. In the next section, I present results from a further perception experiment that lends support to this proposal.

4.3. Experiment II: Directional asymmetry rooted in perception

Experiment II tests the proposal that listeners associate differences in breathiness with changes in pitch, but they do not associate differences in pitch with changes in breathiness. To test the asymmetric dependency hypothesis more directly, I ran a modified cue weighting task in which English listeners categorized auditory stimuli in which there was strong evidence that either pitch or breathiness alone is informative for the contrast, and crucially *no* evidence that the other of these cues is informative. They were then tested on test stimuli that change only in pitch but had a constant breathiness, and on stimuli that change only in breathiness but had a constant pitch. Listeners were expected to attend mostly to the dimension they were trained on. However, if listeners associate one cue to another, then we would also expect them to attend somewhat to the other cue, despite there being no evidence for it in the input. Thus, the asymmetric dependency hypothesis would predict specifically that listeners trained on breathiness would pay more attention to the uninformative pitch cue than listeners trained on pitch to the uninformative breathiness cue.

In the previous experiments, listeners all received the same amount of training on a specific distribution before being tested. In this experiment, I additionally controlled for the amount of training listeners had to see whether this would impact their cue weighting. Listeners with less training should be less certain about the relative importance of the two cues, and thus should still be hesitant to down-weight the non-distinctive cue. Listeners with more training should be fairly certain that they can categorize the sounds using only the distinctive cue, and thus should be able to give a (near-)zero weight to the non-distinctive cue. However, if the non-distinctive cue is perceptually inseparable from the distinctive cue, then we would expect listeners to have difficulty down-weighting the non-distinctive cue even with increased training. We can, thus, make a further prediction about listeners' weighting of non-distinctive pitch and breathiness. Of the listeners learning to categorize on the Distinctive Pitch distribution, those who receive less training should give higher weights to the non-distinctive breathiness cue than those who receive more training and have learned to down-weight that cue. On the other hand, listeners who are learning to categorize on the Distinctive Breathiness distribution should have difficulty down-weighting the pitch, regardless of how much training they receive.

The following section describes the details of the methodology.

4.3.1. Methods

4.3.1.1. Participants

148 undergraduate students were recruited from the Psych Subject Pool at UCLA and participated in the study for course credit. All were native speakers of English. 34 of these participants were excluded for speaking an additional language fluently or natively. Five

participant were excluded for not completing the study. Participants did not have any known hearing impairments.

4.3.1.2. Stimuli

The distributions of the auditory stimuli used in this study were based on the stimuli used in Chapter 2. Whereas those stimuli had one very informative, distinctive dimension and one weakly informative, non-distinctive dimension, the stimuli in the current experiment remove any evidence for categories along the non-distinctive dimension by neutralizing the category difference for that cue. Thus, if listeners assign a non-zero weight to the non-distinctive cue, then it could not have come from what they have learned from the input. The distributions of these stimuli are shown in Figure 4.7.

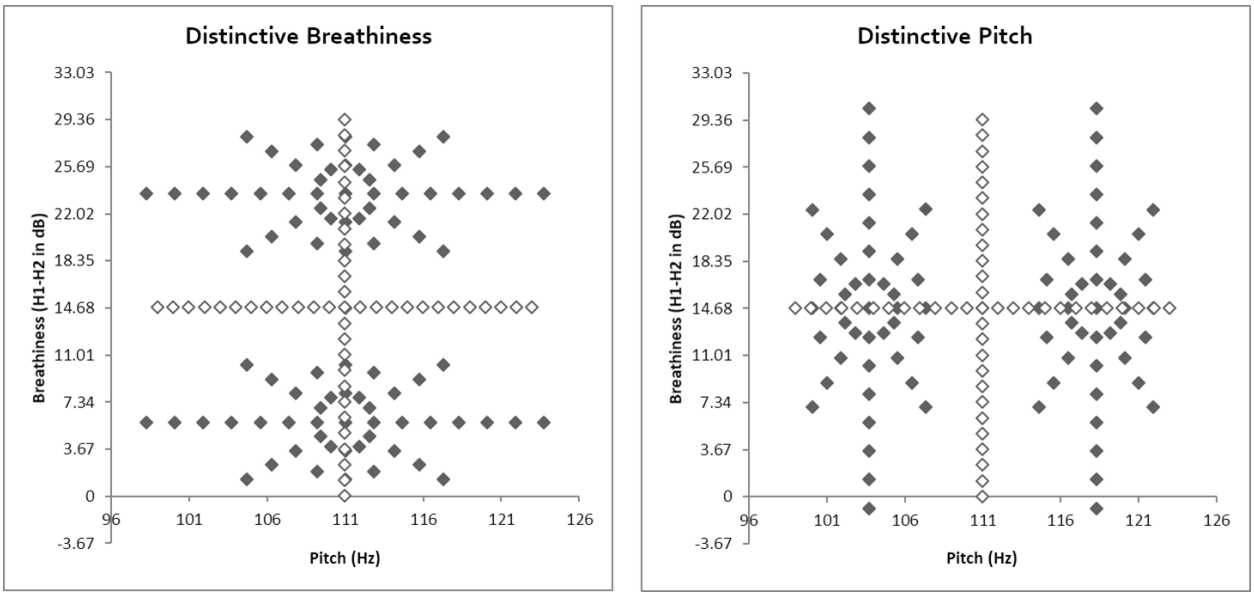

Figure 4.7. Training stimuli (black) and test stimuli (white) in Distinctive Breathiness (left) and Distinctive Pitch (right) distributions. Each stimulus has a breathiness (H1-H2 in dB) value and a pitch (Hz) value in the two-dimensional space.

The scaling of the two dimensions remains the same as the experiments in Chapter 2, but the values of the 86 stimuli in each distribution are modified to suite the purposes of this experiment.

In the Distinctive Breathiness distribution, the two categories, two clusters of test stimuli (black), are separated only along the Breathiness dimension (Distinctive) where the distance between category means is 4.8 JNDs (17.6 dB) and the within-category variance is 2.4 JNDs (8.8 dB). Along the Pitch dimension (Non-Distinctive) in this distribution, the categories overlap completely such that for every stimulus in one category, there is a stimulus in the other category with the same pitch value. In the Distinctive Pitch distribution, the categories are separated only along the Pitch dimension where the distance between means is 4.7 JNDs (14 Hz) and the within-category variance is 2.3 JNDs (7 Hz). Along the Breathiness dimension in this distribution, the categories, again, overlap completely such that for every stimulus in one category, there is a stimulus in the other category with the same breathiness value. The test stimuli (white) in both distributions, composed of 50 stimuli, are identical to those used in earlier experiments. 25 of these were held constant at the average breathiness (14.68 dB) but varied along the pitch dimension, while the other 25 were held constant at the average pitch (111 Hz) but varied along the breathiness dimension. See Chapter 2 for scaling and synthesis of the stimuli.

4.3.1.3. Procedure

Since I am only concerned about the cue weights assigned to each cue within a single distribution, and not with the amount of cue shifting, listeners in this experiment were only trained and tested on one distribution. This is shown schematically in Figure 4.8.

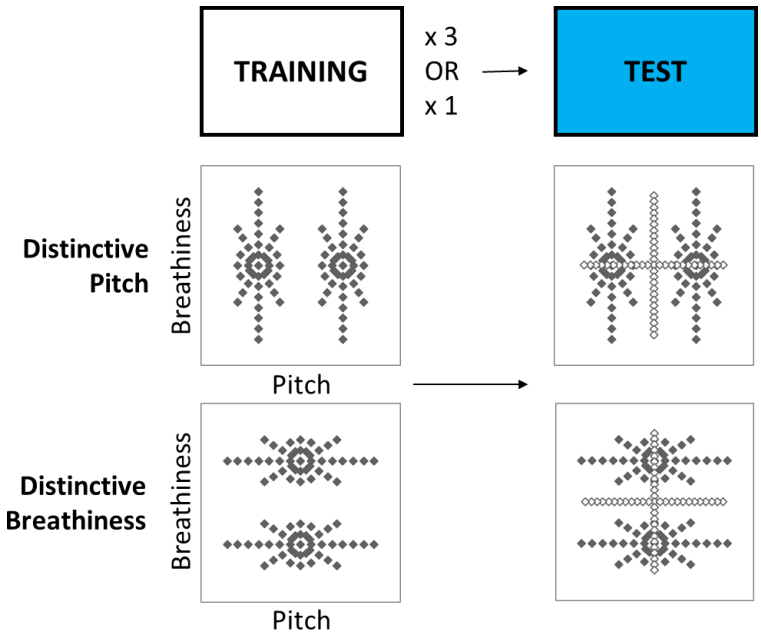

Figure 4.8. Overall Procedure: Participants completed training blocks (either 1 or 3) and the test block in order from left to right. Stimuli presented in each block are displayed for each condition under each block type (black points = training stimuli, white points = test stimuli).

An added manipulation in this experiment was how much training, in blocks, participants were exposed to before being tested. Half of the participants were given one training block before the test block and the other half were given three training blocks before the test block. Within each training condition, half of the participants learned Pitch as the Distinctive cue with Breathiness as the Non-Distinctive cue, while the other half learned Breathiness as the Distinctive cue with Pitch as the Non-Distinctive cue.

The procedure during each trial was identical to the earlier experiments. Listeners heard a stimulus token, made a binary choice on the keyboard (Category A or Category B), and received visual feedback (Correct or Incorrect for training trials, or uninformative blue triangle for test trials).

4.3.1.4. Analysis

The same performance threshold was applied in this experiment to exclude participants who were not learning to categorize the stimuli using the distinctive cue. 15 additional participants were excluded based on this criteria. A total of 92 participants were included in the final analysis, 23 in the Distinctive Pitch – 3 Blocks Training group, 25 in the Distinctive Breathiness – 3 Blocks Training group, 22 in the Distinctive Pitch – 1 Block Training group, and 22 in the Distinctive Breathiness – 1 Block Training group.

Cue weights were obtained in the same way as in the previous experiments, using a logit binomial regression model implemented in R (R Development Core Team, 2015) using the built-in glm function. The model had Category Choice on the test trials as its dependent variable and the Pitch and Breathiness values from the test trial as independent predictors. Since I am interested in the actual weights of the Non-Distinctive cue rather than the relative cue weight of the Distinctive and Non-Distinctive cues, the coefficients obtained from the logit models were not normalized. Instead, the raw coefficients were used in further analysis. The cue weights from each condition were compared using simple linear regressions implemented in R using the built-in *lm* function. Planned comparisons on these models were carried out using the *glht* function in the *multcomp* package (Hothorn et al., 2008).

4.3.2. Results

The weights of the Distinctive cue and the Non-Distinctive cue are presented separately. The Distinctive cues were analyzed first to make sure that listeners were weighting them equally across conditions. These data are shown in Figure 4.9.

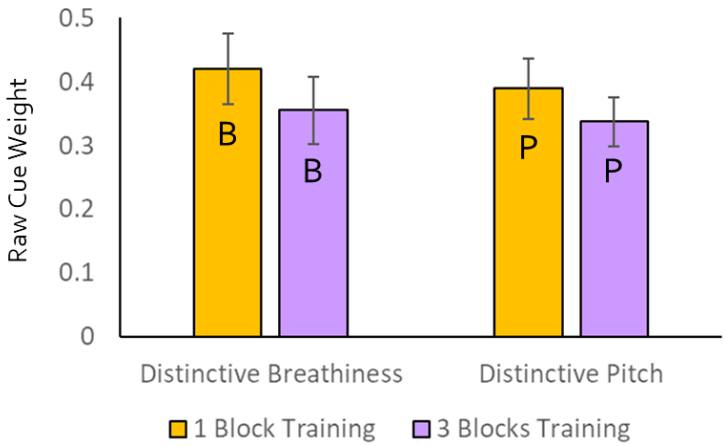

Figure 4.9. Distinctive cue weights by Distribution (Distinctive Breathiness vs. Distinctive Pitch) and Training (1 Block vs. 3 Blocks)

The weights of the distinctive cues are not different across conditions. Though the cues from the conditions with 3 blocks of training seem slightly lower, this is not supported by results from a linear regression with Raw Cue Weights as the dependent variable. The model included the fixed effects Distribution (Distinctive Breathiness vs. Distinctive Pitch) and Training (1 Block vs. 3 Blocks), and their two-way interaction. Results are shown in Table 4.3.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.42	.05	8.53	<.001	***
Distribution = *Distinctive Pitch*	-0.31	.07	-0.44	.663	
Training = *3 Blocks*	-0.07	.07	-0.93	.357	
Distribution × Training = *Distinctive Pitch & 3 Blocks*	0.13	.10	0.13	.894	

Table 4.3. Lmer results from Hani listeners' performance on Distinctive cue weights.

There were no significant fixed effects or interaction from the model. Given that the differences are small to begin with, I additionally ran several planned comparisons to ensure that each pair of cue weights within each condition was not different. There was no significant difference between the two Training conditions when Breathiness was distinctive (β = 0.07, p = .357), the two

Training conditions when Pitch was distinctive (β = 0.05, p = .475), the two Distribution conditions when there was 1 block of training (β = 0.03, p = .663), or the two Distribution conditions when there were 3 blocks of training (β = 0.02, p = .81). Thus, I failed to find a difference between the cue weights given to the Distinctive cue in any of the conditions.

The weights of the Non-Distinctive cues were analyzed next. These data are shown in Figure 4.10.

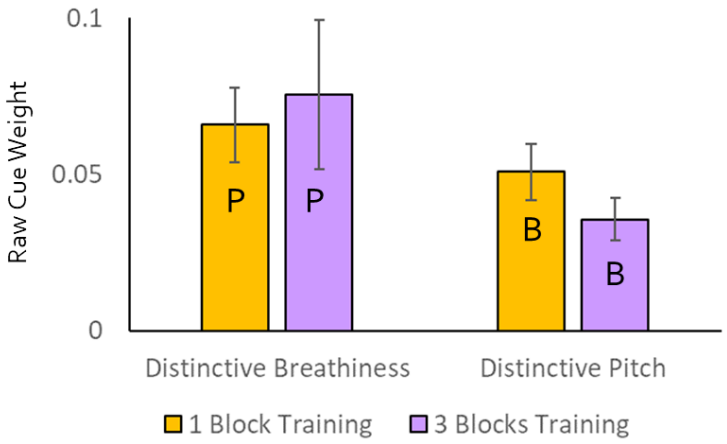

Figure 4.10. Non-Distinctive cue weights by Distribution (Distinctive Breathiness vs. Distinctive Pitch) and Training (1 Block vs. 3 Blocks)

The weights of the Non-Distinctive cues are notably smaller than the Distinctive cues. (Note the change in scale for Raw Cue Weight.) Yet, they all have non-zero weights, as shown by a one-sample t-test comparing each set of cue weights to a zero. These results are given in Table 4.4.

Distribution	Training	t-value	df	p-value
Distinctive Pitch	1 Block	5.49	24	<.001
	3 Block	3.18	21	.002
Distinctive Breathiness	1 Block	5.58	22	<.001
	3 Block	5.23	21	<.001

Table 4.4. Results from on sample t-tests comparing each set of cue weights to zero.

Figure 4.10. also shows that weights for Non-Distinctive cues were not consistent across conditions. Namely, Non-Distinctive Pitch seems to be weighted higher than Non-Distinctive Breathiness overall, and this difference is greater when listeners got three blocks of training as

opposed to one. These data were analyzed in a linear regression model with the fixed effects Distribution (Distinctive Breathiness vs. Distinctive Pitch) and Training (1 Block vs. 3 Blocks), and their two-way interaction. Results are shown in Table 4.5.

Predictor	Estimate	SE	t-value	p-value	sig.
(Intercept)	0.07	.01	4.82	<.001	***
Distribution = *Distinctive Pitch*	-0.02	.02	-0.76	.447	
Training = *3 Blocks*	0.01	.02	0.49	.629	
Distribution × Training = *Distinctive Pitch & 3 Blocks*	-0.02	.03	-0.87	.387	

Table 4.5. Lmer results from Hani listeners' performance on Non-Distinctive cue weights.

This model also does not have significant main effects or significant interaction. To probe it further, I ran another set of planned comparisons. Two of these tested for a difference between the Non-Distinctive Pitch and Breathiness cue weights in each of the Training conditions. These tests showed that there was no difference between Non-Distinctive Pitch and Non-Distinctive Breathiness for listeners with less training (β = 0.02, p = .447), but that Non-Distinctive Pitch was weighted marginally higher than Non-Distinctive Breathiness for listeners with more training (β = 0.04, p = .056). While the p-value is not significant, the effect size is small to medium (Cohen's d = .41). Since the sample size is quite small, this difference would likely become significant with more participants. The other two tested for differences between Non-Distinctive cues after 1 block of training and 3 blocks of training for each Distribution condition. These tests showed that there was no difference between the weights of Non-Distinctive Pitch after 1 training block and after 3 training blocks (β = 0.01, p = .629), nor is there a difference between the weights of Non-Distinctive Breathiness after 1 training block and after 3 training blocks (β = 0.02, p = .46).

4.3.3. Discussion

Experiment II was conducted to test whether the asymmetry in cue shifting observed cross-linguistically in previous studies was caused by an asymmetric perceptual dependency between pitch and breathiness. Specifically, I proposed that pitch is perceived independently of breathiness, but breathiness is not perceived independently of pitch. This makes the prediction that listeners learning to use breathiness for categorization would give some weight to pitch, even in the absence of any evidence from the signal that pitch is informative. However, listeners learning to use pitch for categorization should be able to give a zero weight to breathiness if the latter cue is neutralized across categories. In addition to this hypothesis, I also tested how the weights of the non-distinctive cue would change with increased training. Here, I predicted that with less exposure, listeners learning from both distributions would attribute some weight to the non-distinctive cue, and that listeners with more training would learn to down-weight non-distinctive breathiness but *not* non-distinctive pitch. Thus, I expected the difference between non-distinctive pitch and breathiness to be smaller in the 1 Training Block condition compared to the 3 Training Block condition. These predictions were partially borne out.

First, though it was predicted that the weight of the non-distinctive breathiness cue would be zero after extensive training, all weights were non-zero. Recall that there is no evidence in the signal that the non-distinctive dimension is at all informative to the categorical contrast. Thus, given that pitch is not dependent on breathiness, listeners who are learning to categorize using pitch should not attend to breathiness at all. The fact that the cue weights were not down-weighted to zero after three training blocks could have been due to insufficient training. That is, after a few more training blocks, listeners might eventually stop attending to this cue entirely.

However, let's reconsider the design of the stimuli. The distribution was such that the two categories overlap completely along the non-distinctive dimension and not at all along the distinctive dimension such that one could easily draw a decision boundary perpendicular to the line connecting the two category means (e.g. McKinley & Nosofsky, 1996). However, the within-category variance along the non-distinctive dimension is also wide, covering almost the entire range of possible values within the delimited acoustic space. This large amount of variation along one dimension is likely difficult for listeners to ignore completely, even if the changes do not help them to better categorize the stimuli. Indeed, it has been shown that listeners are more likely to attend to cues that have a wider range of values compared to those that have a narrower range (Lutfi, 1993), all else being equal. If the stimuli had been designed such that there was no categorical difference along one dimension and *also* no variance along that same dimension, listeners would not attend to that cue at all. This is unrealistic though, since in natural speech, we expect there to be variation along all dimensions of a speech signal. Thus, in a natural setting, and also in the experiment conducted here, we may not observe zero weights for any cue.

Second, the weights of the non-distinctive cues were not the same across all conditions. Specifically, while the non-distinctive pitch and breathiness weights were not different after one block of training, the difference between them was greater and nearing significance after three blocks of training. This is in line with the predictions made given the asymmetric perceptual dependency between pitch and breathiness. With less training, listeners would be less willing to down-weight the non-distinctive cue, whether they were learning pitch and breathiness was non-distinctive, or they were learning breathiness and pitch was non-distinctive. After three blocks of training however, listeners weighting of the two non-distinctive cues was predicted to differ.

Those learning to categorize based on pitch gave a higher weight to non-distinctive breathiness than those learning to categorize based on breathiness gave to non-distinctive pitch.

Overall, the results lend support to the proposal that pitch is perceived independently of breathiness, but breathiness is not perceived independently of pitch. Such an asymmetry in the perceptual dependence between these two cues would explain the cross-linguistic asymmetry in listeners' ability to shift attention from one cue to the other when they were in a non-enhancing relationship. That is, listeners had difficulty shifting from breathiness onto pitch, as predicted by Auditory Enhancement, but exhibited normal shifting from pitch to breathiness, an unexpected result. If the asymmetry is rooted in the perceptual dependency between pitch and breathiness, then the cue shifting results could be explained as follows: Listeners shifting their attention from breathiness to pitch experience interference when the mapping reverses the enhancing relationship between the cues, but listeners shifting from pitch to breathiness are able to treat the latter as a novel cue independent of the initial distinctive cue.

4.4. Summary and conclusion

In this chapter I first asked whether the asymmetry in listeners' cue shifting between pitch and breathiness was language-specific or language-general. To answer this question, I conducted the same cue shifting experiment on two additional groups of listeners, a tone group that used pitch phonemically and a phonation group that used breathiness phonemically. If language experience was the driving factor, then we expected the tone group to exhibit the same behaviour as English and Hani listeners, as all three groups use pitch more extensively than breathiness, but not the phonation group, who use breathiness more extensively than pitch. I found that both the tone and phonation listeners mirrored the English and Hani listeners in their

behaviour: They were able to shift cue weight from pitch to breathiness but not vice versa. The uniform result across all four language groups strongly suggests that the asymmetry is language-general. I thus proposed that it was due to an asymmetric perceptual dependency between the two cues. Specifically, these results could be accounted for if listeners perceptually associate changes in breathiness to changes in pitch, but perceive pitch as a relatively independent acoustic dimension.

I thus test this hypothesis in a follow-up experiment where listeners learned to categorize stimuli that, again, varied along the pitch and breathiness dimensions. This time however, the small difference between categories along the less informative dimension is neutralized, leaving no evidence that the cue is useful at all for distinguishing the contrast. If listeners continue to pay attention to the uninformative cue while learning the contrast using the informative cue, then we have evidence that they are associating the two cues. I hypothesized that i) listeners being trained to use breathiness would also give some weight to pitch, but listeners being trained to use pitch would not give weight to breathiness, and ii) this discrepancy would increase with training as listeners trained on pitch become more certain that the breathiness dimension is uninformative. These predictions were borne out, but marginally, lending some support to the asymmetric dependency hypothesis.

Together, the results from Chapter 2 and the experiments from this chapter support the claim that pitch and breathiness have perceptual dependencies. However, these experiments also highlight an asymmetry in the relationship between these cues that cannot be accounted for if we assume they form a single intermediate perceptual property (IPP) and are perceived as a whole, per the Auditory Enhancement theory. The fact that cue weights for the uninformative dimension never equals that of the informative dimension during learning clearly shows that the cues are

perceptually separable. But, the dependency of breathiness to pitch and the independence of pitch from breathiness is also not negligible.

Though the evidence presented in these chapters goes against the symmetrical dependency predicted by Auditory Enhancement, other findings are congruent with aspects of the theory. Most importantly, listeners clearly treated the cue pair that share auditory properties, pitch and breathiness, differently from a pair that does not share any such similarities, pitch and vowel duration. This supports the idea that cues can be perceptually privileged, even in the absence of, or despite, experience with co-variation. It is also important the point out that the evidence for the dependency of breathiness to pitch was strongest in the cue shifting task that overtly required listeners to associate both cues in a pair to arbitrary category labels, which force the listeners to use the cues at a phonological level. The relevance of enhancing cues to phonological contrast is also emphasized under Auditory Enhancement theory. Lastly, since these effects are not dependent on experience, they are predicted to be language-general. In this chapter, I provided limited experimental evidence showing that this might be true. However, if what we have observed in these artificial environments is reflective of basic traits of the human perceptual system, it should also be evidenced in the way languages pattern cross-linguistically.

Thus, in Chapter 5, I conduct a typological survey of the patterns of cue co-variation and diachronic contrast shift. I focus on the two sets of cues I have chosen for my experiments, pitch and breathiness and pitch and vowel duration, in an attempt to see whether the experimental patterns we observed are reflected in typology.

CHAPTER 5: TYPOLOGY

5.1. Experimental results as typological predictions

In Chapters 2, 3, and 4, I presented experiments testing listeners' perception of two sets of cues: the enhancing cues, pitch and breathiness, which are perceptually integrated and contribute to the same auditory effect, and the non-enhancing cues, pitch and vowel duration, which are perceptually integrated but do not contribute to the same auditory effect. Listeners were made to shift attention from one acoustic cue to another while categorizing stimuli into two categories with meaning differences. Results for the non-enhancing cues showed that the degree to which listeners were able to shift attention between cues was entirely dependent on their language experience with the co-variation between cues. Listeners experienced with one kind of co-variation had difficulty shifting attention between cues if this learned co-variation was reversed experimentally. Listeners without any experience were able to shift attention between cues regardless of how they were mapped onto each other experimentally. In contrast, results for the auditorily enhancing cues showed asymmetric enhancement effects. When the enhancing relation between the cues was experimentally reversed, listeners who learned to use pitch first were able to shift attention onto breathiness, but listeners who learned to use breathiness first were unable to shift attention onto pitch. I argued that this enhancement effect is rooted in the asymmetric perceptual dependency between the two cues. That is, pitch can be perceived independently of breathiness, but not vice versa.

These experiments highlight a key difference between perceptually integrated and enhancing cues: While even in the absence of experience listeners naturally associate enhancing cues to phonological categories, this is not true for non-enhancing cues. If this holds true as a generalization, then we should also see differences in the way enhancing and non-enhancing cue

pairs pattern cross-linguistically. Thus, in this chapter, I survey the world's languages and provide an overview of how the specific enhancing and non-enhancing cues chosen for the experiments in the earlier chapters pattern as acoustic cues both synchronically and diachronically. Since there are relatively few languages that contrast breathy phonation with modal phonation, I extended the scope of the survey on enhancing cues to also include voicing contrasts. This extension seems natural given that i) voicing, breathiness, and low pitch contribute to the same auditory effect of strong low frequency energy (Kingston, 2011), and ii) all three of these cues are involved in the same kinds of diachronic processes (see Section 5.3.1).

Synchronically, we expect both sets of cues to co-vary across languages since they are perceptually integrated. Furthermore, if a correlation is found between pitch and duration, they should bear a negative relation. That is, lower pitch should co-vary with longer duration and higher pitch should co-vary with shorter duration. Pitch and breathiness (or voicing) are also predicted to have a negative co-variation, with breathiness and voicing associated with lower pitch, and modal phonation and voicelessness associated with higher pitch.

On the other hand, since only enhancing cue pairs are perceptually associated with the phonological categories they distinguish, only secondary enhancing cues should be able to replace the primary cue in signalling the same contrast. To test this, I examined contrast shifts, a diachronic phenomenon that closely mirrors the experimental design from earlier chapters on a larger time scale. This is a process where a phonological contrast that was signalled by one acoustic cue becomes signalled by a different acoustic cue over time. Initially, these two cues would simply co-vary, with one cue primarily bearing the categorical distinction. Over time, due to language-internal reasons or language contact, the contrast becomes exaggerated along the secondary dimension and reduced along the primary dimension, resulting in a shift in which the

secondary cue becomes primary. This is schematized in Figure 5.1., from Kang (2014), which illustrates a contrast shift from VOT to f0 (where Stage I and Stage 5 are extremely unlikely).

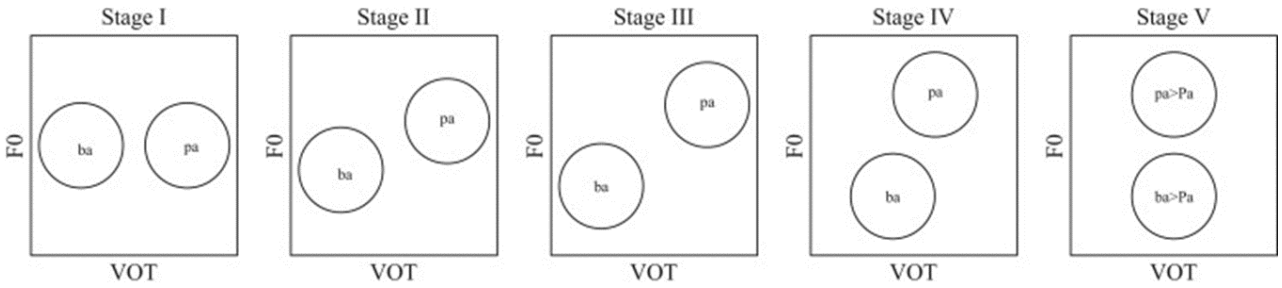

Figure 5.1. Stages of contrast shift between VOT and f0, from Kang (2014). Two circles represent the phonological categories in each distribution.

In Figure 5.1, each distribution is defined along two acoustic dimensions, voice onset time (VOT) and fundamental frequency (f0). Within this acoustic space, two phonological categories are represented by two circles, one labeled *pa* and the other *ba*. In the initial stage, the categories are distinguished by VOT and completely overlapping on the f0 dimension. In Stage II, the two categories are distinguished by VOT, but also differentiated slightly along the f0 dimension such that it is positively correlated with the VOT cue. At this stage, the (historically) voiced category has distinctly smaller VOT values and somewhat lower f0 values than the voiceless category. In Stage III, the secondary f0 cue has become exaggerated to the point that it could also be used to reliably distinguish between the *pa* and *ba* categories, though VOT remains informative as well. In Stages IV and V, the VOT difference between the categories reduces, then merges completely, leaving f0 to be the sole cue to the contrast.

The predictions for diachronic contrast transfers are as follows: Non-enhancing cues such as pitch and vowel duration are not expected to participate in contrast shift. That is, we do not expect to find cases in which a pitch contrast becomes a length contrast or vice versa. Conversely, enhancing cues such as pitch, breathiness, and voicing are expected to participate in

contrast shift. Given the asymmetry observed in the experimental setting, we also expect the direction of contrast transfer to be asymmetric. Since pitch is proposed to be perceived independently of breathiness (and by extension, voicing), pitch contrasts are not expected to transfer onto the other cues. However, voicing and breathiness contrasts are predicted to transfer onto pitch frequently.

Before beginning with the typology, a caveat: the survey is limited by the availability of published studies on the specific sets of cues I am interested in. For example, though the majority of the world's languages have voicing contrasts, only a subset of these languages has been studied with the interest of knowing whether pitch is a correlate of voicing contrasts. Thus, the basis for my generalizations is not whether some significant proportion of languages exhibit a given pattern, but rather whether the pattern can be observed in a typologically diverse set of languages. This is also true when comparing the prevalence of one pattern versus another pattern (e.g. contrast shift from pitch to voicing vs. voicing to pitch).

This chapter is split into two main sections. Section 5.2 addresses the predictions made about synchronic co-variation between pitch, breathiness, and voicing on one hand, and pitch and vowel duration on the other hand. Section 5.3 then examines cases of diachronic contrast transfer between these two sets of cues.

5.2. Synchronic cue co-variation

In this section, I present the languages in which the enhancing cues – pitch, breathiness, and voicing – and non-enhancing cues – pitch and vowel duration – co-vary. Since both sets of cues are perceptually integral, I predict that their co-variation should be frequently attested cross-linguistically.

5.2.1. Pitch, breathiness, and voicing

This section will discuss the co-variation between each cue pair separately. Section 5.2.1.1 will focus on the co-variation between voicing and pitch, first discussing languages in which pitch co-varies with a voicing contrast, then moving onto cases where voicing is affected by lexical pitch (a.k.a. tone). Section 5.2.1.2 will focus on breathiness and pitch as a cue pair. I begin with a survey of languages in which pitch co-varies with a phonation contrast, then discuss languages in which breathiness co-varies with a pitch contrast.

5.2.1.1. Voicing and pitch

House and Fairbanks (1953) conducted the first study documenting a consistent difference between the f0 of vowels following phonologically voiced vs. voiceless consonants. Since then, this result has been replicated in many other studies on English. Some of these studies are listed below in Table 5.1.

Eng. Dialect	Segment class – f0 (Hz)	Measurement	Source
American	Voiceless cons. – 126.46 Voiced cons. – 121.99	Average across vowel	House & Fairbanks, 1953
American	Voiceless stops – 175.67 Voiced stops – 163.67	Peak f0	Lehiste & Peterson, 1961
American	Voiceless stops – 130.53 Voiced stops – 124.97	?	Mohr, 1968
L2 (L1 = Chinese, Russian, German)	Voiceless obstr. – 134.5 Voiced obstr. – 123.1	Vowel-initial f0	Mohr, 1971
unspecified	Voiceless stops – 165* Voiced stops – 155*	Peak f0	Lea, 1973
American	Voiceless stops – 136* Voiced stops – 119*	Vowel-initial f0	Hombert, 1978
unspecified	Voiceless stops – 134* Voiced stops – 122*	Peak f0	Umeda, 1981
American	Voiceless stops – 135* Voiced stop – 107*	Vowel-initial f0	Ohde, 1984
American	Voiceless stops – 220* Voiced stops – 165*	Vowel-initial f0 (male, high-pitch environment, early in utterance)	Hanson, 2009
American	T > D by 2 semitones*	Vowel-initial f0	Dmitrieva et al., 2015

Table 5.1. Pitch difference for phonologically voiced and voiceless stops in English. *Estimates from figure.

Though the studies in Table 5.1. are all on English, the ranges of values given vary depending on the specific experimental design and context in which they were recorded. While some studies tended to elicit nonce words in isolation or small carrier phrases (e.g. House & Fairbanks, 1953; Lehiste & Peterson, 1961; Hombert, 1978), others took measurements from more naturalistic speech (e.g. Umeda, 1981; Hanson, 2009) where segments were taken from real words embedded in longer sentences with varied prosodies. These studies also used different measures of f0, indicated in Table 5.1., including average f0 across the vowel, peak f0, and vowel-initial f0 following the voiced or voiceless consonant onset. Despite the differences in experimental setting and type of f0 measures taken, the relation between vowel f0 and the

voicing of the preceding consonant is clear and consistent in English: f0 is higher after (phonologically) voiced consonants and lower after (phonologically) voiceless consonants. Data from studies in which the time course of pitch change was tracked through the vowel (e.g. Hombert 1978; Hanson, 2009) clearly showed that the largest f0 differences are observed at onset, but a significant difference persists even 100 ms after the onset of the vowel. This is likely why even studies that took an average pitch over the entire duration of the vowel (e.g. House & Fairbanks, 1953) found consistent differences. Furthermore, the pattern is observed even for (highly proficient) L2 speakers of English (Mohr, 1971), whose productions are influenced by their native phonologies and phonetic productions of similar consonants.

Consistent with the post-consonantal pitch differences found in the studies on English, numerous instrumental studies have found the same pattern across languages from different language families and geographic regions. These studies are summarized in Table 5.2.

Language	Family	Segment class – f0 (Hz)	Measurement	Source
Swedish	Germanic	T cons. – 174.17 D cons. – 156.83	Peak f0	Löfqvist, 1975
Dutch	Germanic	T obstr. – 176 D obstr. – 160	Vowel-initial f0	van Alphen & Smits, 2004
German	Germanic	T – 169 D – 161	Vowel-initial f0	Jessen, 2001
Spanish	Romance	T > D by 1.3 semitones*	Vowel-initial f0	Dmitrieva et al., 2015
Italian	Romance	T > D by 0.5-0.8 semitones*	Vowel-initial f0	Kirby & Ladd, 2016
French	Romance	T > D by 0.3-1 semitones*	Vowel-initial f0	Kirby & Ladd, 2016
Hindi	Indo-Aryan	T – 188 T^h – 178 D – 154 $D^ɦ$ – 120	Vowel-initial f0	Kagaya & Hirose, 1975
Persian	Indo-Iranian	T obstr. – 153* D obstr. – 142*	Vowel-initial f0	Bijankhan & Nourbakhsh, 2009
Yoruba	Niger-Congo	K – 169* G – 132*	Vowel-initial f0	Hombert, 1976
Khmer	Mon-Khmer	T^h > T by ~1 semitone* T > D by 1-2 semitones*	Vowel-initial f0	Kirby, 2018
Vietnamese	Mon-Khmer	T^h > T by 0-0.5 semitones* T > D by 0.5-1 semitones*	Vowel-initial f0	Kirby, 2018
Thai	Tai-Kadai	T^h > T, D by ~2 semitones* (only after high-falling tone in isolation)	Vowel-initial f0	Kirby, 2018
Shanghainese	Sinitic	T > T^h > $D^†$	Vowel-initial f0	Chen, 2011
Madurese	Austronesian	T^h – 251 (F), 162 (M) T – 248 (F), 163 (M) D – 237 (F), 153 (M)	Vowel-initial f0	Misnadin, 2016
Burmese	Tibeto-Burman	T^h – 186.8 T – 187.7 D – 167	Vowel-initial f0	Shimizu, 1989
Japanese	Japonic	T – 248.5 D – 213.8	Vowel-initial f0	Shimizu, 1989

Table 5.2. Pitch of vowels following onset consonants in languages with a phonological voicing distinction other than English. "T" = voiceless unaspirated stops, "T^h" = voiceless aspirated stops, "D" = voiced unaspirated stops, "D^h" = voiced aspirated stops, unless manner is otherwise indicated. For Yoruba, "K" represents voiceless velar stops and "G" represents voiced velar stops. *Estimates from figures. †Values given in normalized z-scores.

The 15 languages in Table 5.2. represent 11 large language families from Europe, Africa, and Asia. Eight of these languages have a two-way voicing contrast, six languages have an

additional contrast between aspirated and unaspirated voiceless stops, and one language has a four-way stop contrast characterized by voicing and aspiration.

Of the languages that are included in Table 5.2., I only consider the way in which f0 patterns with phonologically voiced vs. voiceless consonants. Therefore, languages like Mandarin (Xu & Xu, 2003), Cantonese (Francis et al., 2006), and Taiwanese (Lai et al., 2009) were excluded because their stops are distinguished by aspiration (e.g. /t/ vs /t^h/) rather than by voicing. It should be noted that languages with voiced stops, voiceless unaspirated stops, and voiceless aspirated stops are commonly thought to have a three-way voicing contrast in which these sounds are differentiated along a VOT continuum. On the surface, this may suggest that f0 on the following vowel may co-vary with VOT across all three categories of consonants such that it is the lowest after voiced stops and highest after voiceless aspirated stops. However, this is typologically untrue since aspiration lowers f0 in some languages but raises f0 in other languages, and results differ even in multiple investigations of the same language (see Kirby, 2018 for full review).

The relationship between voicing and f0 on the following vowel is consistent across all the languages in Table 5.2. In languages with a two-way voicing contrast, f0 is lower after voiced consonants than after voiceless consonants. Note that this is the case for both languages with "true" voiced consonants (e.g. French, Spanish) where voiced consonants are characterized by prevoicing and voiceless consonants are characterized by a lag in VOT, and for languages in which phonologically voiced and voiceless consonants are realized phonetically with short- vs. long-lag VOT (e.g. English). In languages that have a three-way stop contrast, f0 is lower after the voiced stops than both of the voiceless stops. In Hindi, which has a full four-way contrast

between aspirated and unaspirated and voiced and voiceless stops, f0 is also lower after the voiced consonants than their voiceless counterparts.

Thus far, I have not come across any instrumental study that has found the opposite effect of voicing on vowel f0. So, we not only find that voicing and f0 co-vary frequently across languages, but that there is a remarkable consistency in the direction in which these cues co-vary. As has been claimed by Kingston et al., the cross-linguistic prevalence of this pattern suggests that the co-variation between f0 and voicing is not accidental, but rather driven by the fact that they are auditorily enhancing.

However, this theory, which assumes equal dependence between two enhancing cues, also predicts that voicing should co-vary with pitch contrasts just as frequently. Interestingly, the typological evidence is not consistent with this claim. While there are many languages in which f0 co-varies with a voicing contrast, there are few convincing cases in which voicing co-varies with a pitch contrast, all of which are from anecdotal rather than instrumental studies. The first is the case of the Ohũhũ dialect of Igbo in which glottal fricatives are partially voiced before low tones (Dunstan & Igwe, 1966 cited in Maddieson, 1974). Since the original source can't be located, no further information – description or data – can be obtained. However, if this statement on the dialect is true, it would constitute a good case of voicing co-varying with a tonal contrast. The second is a phonological process in Jingpho by which a voiceless stop at the end of a stem with low tone becomes voiced when followed by a final particle with low tone (Maddieson, 1974; cf. Hyman, 1976). The pattern in both of these languages is consistent with the direction in which these cues co-vary in the languages already discussed.

Maddieson (1974) discusses a few more examples of tones affecting consonants, but these are less convincing (see Hyman, 1976 for counterarguments). As far as I am aware, there

also have been no instrumental studies showing that degree of voicing is correlated with pitch levels in tone contrasts. Thus, it is clear that the overwhelming majority of cases in which voicing and pitch co-vary are those where pitch is recruited as a secondary cue to a voicing contrast. Though the experiments in Chapters 2 and 4 focused on breathiness, this typological asymmetry between the voicing and pitch cue also echoes the experimental asymmetry found between breathiness and pitch; pitch can vary independently, but breathiness (and voicing by extension) are perceptually tied to pitch. I turn next to the cross-linguistic co-variation between breathiness and pitch.

5.2.1.2. Breathiness and pitch

In this section, I begin by surveying languages in which breathiness on the vowel is the main cue to a contrast, and pitch is a secondary cue. I then move on to languages in which pitch is the primary cue (i.e. tone languages), and breathiness is a secondary cue.

Phonemic use of non-modal phonation on vowels is typologically rare (Gordon, 1998). Thus, this survey includes both languages in which there is a phonation contrast as well as those which have register contrasts for which breathiness is the main cue. Languages with contrasts between modal and non-modal phonations other than breathiness were excluded, as were languages that have consonantal phonation contrasts.

Language	Family	Contrast	Segment class – f0 (Hz)	Measurement	Sig.?	Source
Jalapa Mazatec	Otomanguean	Phonation	Modal – 173* Breathy – 165*	Vowel onset	No	Garellek & Keating, 2011
Itunyoso Trique	Otomanguean	Phonation (phonetic)	Modal > Breathy	Vowel offset	Yes	DiCanio, 2012
Gujarati	Indic	Phonation	Modal > Breathy in front vowels Breathy > Modal in back vowels	Average across vowel	Yes	Khan, 2012
Jingpho	Tibeto-Burman	Register	Tense – 157 Lax – 145	Vowel onset	Yes	Maddieson & Hess, 1986; 1987
Hani	Tibeto-Burman	Register	Tense > Lax	Average across vowel	?	Maddieson & Ladefoged, 1985 (c.f. Kuang & Keating, 2012)
Bo	Tibeto-Burman	Register	Tense > Lax	?	Yes	Kuang, 2011
Javanese	Malayo-Polynesian	Phonation	Modal > Breathy by 7.8-17.5 Hz	Vowel onset	?	Wayland et al., 1995
Eastern Cham	Chamic	Register	High – 150* (m), 250* (f) Low – 120* (m), 210* (f)	Vowel onset	Yes	Brunelle, 2005
Wa	Mon-Khmer	Register	Modal – 166.8 Breathy – 158.9	Vowel onset	Yes	Watkins, 2002 (cf. Maddieson & Hess, 1987)
Takhian Thong Chong	Mon-Khmer	Register	Breathy > Modal by 2 semitones	Vowel onset	Yes	DiCanio, 2009

Nyah Kur	Mon-Khmer	Register	Modal > Breathy	?	?	Thongkum, 1987
Khmu' Rawk	Mon-Khmer	Register	Modal > Breathy by 3.27 semitones	Average across vowel	Yes	Abramson et al., 2007
Chanthaburi Khmer	Mon-Khmer	Register	Modal – 153 Breathy – 157	Average across vowel	Yes, for 2/5 speakers	Wayland & Jongman, 2003
Kuai Suai	Mon-Khmer	Register	Modal > Breathy by 2.1 semitones*	Vowel onset	Yes	Abramson et al., 2004
Mon	Mon-Khmer	Register	Modal > Breathy by 9 Hz	?	No	Abramson et al., 2015

Table 5.3. Pitch as a correlate in modal-breathy phonation contrasts or contrasts mainly distinguished by breathy vs. modal phonation. *Estimates from figures.

Since contrasts between breathy and modal vowels are comparatively less common than voicing and tone contrasts, the languages listed in Table 5.3. are noticeably less diverse. Only two languages, Jalapa Mazatec and Itunyoso Trique, are spoken in the Americas. All other languages are spoken in Asia, primarily in Southeast Asia, and the majority of these belong to the Mon-Khmer language family which is known for having register contrasts where phonation is the main correlate.

The way in which pitch co-varies with breathiness in these languages is less systematic than the way in which pitch co-varies with voicing. Nevertheless, there is a definite trend in the direction in which the two cues relate. The breathier category ("breathy" or "lax" in Table 5.3.) is associated with lower pitch in 11 out of the 14 languages included here, though the magnitude of this difference varies. In languages like Jalapa Mazatec and Mon, the measured pitch in the modal category was not significantly higher than the breathy category. However, in other languages like the Kuai dialect of Suai and Khmu', the authors report that pitch may have usurped phonation as the primary cue as these languages slowly become tonal (see Section 5.4.1. for discussion on contrast transfer from phonation to pitch). Itunyoso Trique is included here as a special case. Though it does not have a phonemic phonation contrast, DiCanio (2012) demonstrates that breathiness on the vowel from coarticulation with a final /h/ has the effect of lowering pitch. Moreover, the degree of pitch lowering can be predicted by the degree and duration of breathiness on the vowel. The phonetic relation between breathiness and pitch in this language provide further support for the overall pattern observed.

Of the languages that do not exhibit this pattern, two (varieties of Khmer and Chong) show the opposite pattern, where the breathier register actually has higher pitch than the modal register. Wayland and Jongman (2003) acknowledge that their finding is unexpected given the

universal trend, but offer no further discussion or explanation for why this might have come about. The case of Chong is complicated by the fact that it actually contrasts four phonation types, including modal (clear), breathy, tense, and breathy-tense (DiCanio, 2009). The overall pitch of the latter two phonation categories, which have some amount of vocal fold tension, is higher than that of both the modal and breathy categories. However, DiCanio notes that there is no direct correlation between measures of phonation such as H1-H2 and measures of pitch. Thus, the two cues seem to co-vary independently to signal differences between the phonological categories. This is perhaps also true of Gujarati, the last language listed, that deviates from the main pattern. Khan (2012) finds that pitch has a different relation with phonation depending on vowel quality, though this study was not well-controlled for potential prosodic effects across test words. In sum, while there are some outliers, the dominant trend is for breathy phonation to be associated with pitch lowering.

I now turn to the question of how breathiness patterns with pitch in tone languages. These languages are summarized in Table 5.4. Since there are many measures of breathiness across and within studies, making values difficult to compare, I merely state the tone that the breathy phonation is observed on without reporting the measured values.

Language	Family	Contrast	Pattern	Source
Northern Vietnamese	Vietic	6 tones	Low-falling tone is breathy	Edmondson & Lợi, 1997
Green Mong	Hmong-Mien	7 tones	Highest of low-falling tones is breathy	Andruski & Ratliff, 2000
Black Miao	Hmong-Mien	8 tones	Mid-level tone is breathy	Kuang, 2013b
White Hmong	Hmong-Mien	7 tones	Mid-falling tone is breathy	Esposito, 2012
Tamang	Tibeto-Burman	4 tones	2 lower tones are breathy	Mazaudon & Michaud, 2008
Yi	Tibeto-Burman	3 tones (2 registers)	Mid tone (both registers) is breathy	Kuang & Keating, 2014
Santa Ana Del Valle Zapotec	Otomanguean	2 tones, 3 phonations	Breathy vowels only realized with lower f0	Esposito, 2010

Table 5.4. Breathy phonation as a correlate of tone contrasts.

The seven languages in Table 5.4. are, once again, fairly homogenous in terms of the areas in which they are spoken and the large number of tones they tend to have in their inventories. This is unsurprising since a common question among the researchers who engage in studies of this type is how tone contrasts are maintained in a crowded pitch space. This would naturally lead them to investigate languages which have fairly large inventories, typical in Asia and the Americas, as opposed to languages with comparatively simple two-tone systems, typical in Africa. Since the focus of this survey is on how a particular type of phonation, breathiness, patterns with pitch, the scope once again narrows to exclude studies on languages that employ other phonation types such as creakiness or tenseness, typical of languages spoken in the Americas (Kingston, 2011). Santa Ana Del Valle Zapotec is included here as a special case. This is a language with three contrastive phonation types (modal, breathy, creaky), and a two-way tone contrast that is restricted to modal vowels. Interestingly, Esposito (2010) also finds that phonation differences are only realized when the f0 of the speaker is in the mid to low range. When the f0 is high, such as in initial or focused positions, all vowels are modal.

The fact that the use of breathy phonation is not investigated in smaller tone systems does not necessarily mean that it never co-varies with pitch in those languages. However, the representative bias in the literature certainly suggests that phonation cues are less important or not necessary when there are fewer pitch distinctions to be made. That is, when pitch is sufficiently distinctive as a cue, phonation is not recruited (cf. Yu, 2011).

This notion is corroborated by the way in which breathiness patterns with tonal categories in the languages in Table 5.4. First, there seems to be a trade-off relation between breathiness and pitch as cues in tone systems, meaning that breathiness is not used simultaneously with pitch to signal categorical differences. In Tamang, the two tones with the largest open quotient differences are also the tones with the smallest pitch differences (Mazaudon & Michaud, 2008). In Green Mong, breathiness is characteristic of one of the three low-falling tones that have similar pitch contours; the other two of these tones are characterized by modal and creaky voice respectively (Andruski & Ratliff, 2000). In Black Miao, which has five level tones, breathiness is again recruited as a tone dispersion mechanism to differentiate /33/ from /44/ and /22/, which have similar pitches in the middle of the pitch range (Kuang, 2013). Relatedly, breathiness seems to characterize tones in the middle of the pitch range more often than it characterizes tones at the higher or lower end of the pitch range. This is the case in Green Mong, White Hmong, Black Miao, and Yi, though this could be an areal feature. A perception study on listeners of White Hmong additionally shows that phonation cues are important for the identification of the breathy tone in this language (Garellek et al., 2013). Once again, the association of breathiness to mid-range tones suggests that breathiness does not necessarily co-vary with pitch in tone contrasts in the same way that pitch co-varies with breathiness in phonation contrasts. In fact, if the pitch cue is independent of breathiness, and breathy phonation is only recruited as a tone dispersion

mechanism, then the most effective use of it would be on a tone in the middle of the range. Mid tones are less distinct on the pitch dimension from the higher and lower tones than are the higher and lower tones from each other. Thus secondary cues on mid tones sets them apart from tones on either side.

Kuang (2013b) comes to the same conclusion about breathiness based on her results from Black Miao. She differentiates two types of non-modal phonations: first, those that are associated with certain regions of the pitch range (e.g. tenseness and falsetto with the higher pitch range, and creakiness with the lower pitch range), and second, other non-modal phonation types (e.g. breathiness) that vary independently of pitch, but can instead be recruited as an additional acoustic cue to aid the perceptual categorization of tones with similar pitch contours.

Overall, the relationship between breathiness and pitch as cues to phonological contrasts looks to be typologically asymmetric. While pitch co-varies with breathiness in phonation and register contrasts, breathiness does not co-vary with pitch in tone contrasts. These findings are congruent with the experimental results from Chapters 2 and 4, which suggest that these cues are asymmetrically dependent in perception. While pitch can vary and be perceived independently of breathiness, the percept of breathiness is tied to low pitch.

In the next section, I turn my attention to the synchronic co-variation between pitch and vowel duration, which, recall, are non-enhancing but perceptually integral cues.

5.2.2. Pitch and vowel duration

Pitch and vowel duration have long been noted to co-vary cross-linguistically. I begin this section by discussing the ways in which duration correlates with lexical tone contrasts, then move onto how pitch correlates with length contrasts.

Much of the earlier work on the relation between these cues focused on the relation between duration and pitch contour. In tone languages, the generalization is that rising tones have longer durations that falling tones (Gandour, 1974) and contour tones have longer durations than level tones (Yu, 2010), with the exception of falling tones which can be shorter than level tones (e.g. Ho, 1976 for Standard Mandarin; Khan, 2017 for Pahari). This is intuitive from a production perspective since it requires more effort and time to raise pitch than to lower it, and it requires more time to realize multiple pitch targets than to realize a single pitch target.

More relevant to our understanding of the perceptual relation between pitch and duration cues is whether and how duration is correlated with different pitch levels. With this question in mind, Faytak and Yu (2011) conducted a survey of 26 tone languages with level tones in their tone systems. After excluding the languages where one or more lexically level tones has a phonetic contour, and those languages for which only an impressionistic description is available, we are left with 10 languages. These languages from Faytak and Yu (2011) are listed in Table 5.5. along with two languages added from my own survey.

Language	Family	Tone System	Correlation	Pitch Difference	Dur. Difference	Source
Navajo	Na-Dené	H vs. L	Positive	~74 Hz (in stems with L tone on the ultimate syllable)	~40 ms (in stems with L tone on the ultimate syllable)	McDonough, 1999
Tahltan	Na-Dené	H vs. L	Negative	7.4 Hz (short V) 15.8 Hz (long V)	10 ms (short V) 52.3 ms (long V)	Alderete, 2005
Mixtec Chalcatongo	Otomanguean	H, M, L	Negative	HM: 17.5 Hz ML: 34.4 Hz	HM: 17.3 ms ML: 14.7 ms	Meacham, 1991
Dinka	Nilo-Saharan	H, L, contour tones	Negative	HL: 37 Hz	HL: 5 ms	Remijsen & Ladd, 2008
Zulu	Niger-Congo	H vs. L	Negative	H > L	10-20 ms	Russell, 2000
Bai	Sino-Tibetan	4 level, 4 contour	Negative	Higher > Lower	Lower > Higher	Faytak & Yu, 2011
Hani	Sino-Tibetan	Register	Negative	Tense > Lax	Tense < Lax	Maddieson & Ladefoged, 1985
Naxi	Sino-Tibetan	H, M, L	Negative	HM: ~30Hz ML: ~20Hz	HM: ~25 ms ML: ~13 ms	Michaud et al., 2015
Hu	Mon-Khmer	H vs. L	Negative	47 Hz	54 ms	Svantesson, 1991
Khmu' Rawk	Mon-Khmer	Register	Negative	H > L	22 ms	Abramson et al., 2007
Gaoba Dong	Tai-Kadai	5 level, 3 contour	Negative	T5-T4: 47.0 Hz T4-T3: 38.4 Hz T3-T2: 18.0 Hz T2-T1: 8.0 Hz	T5-T4: 136 ms T4-T3: 15 ms T3-T2: 15 ms T2-T1: 30 ms	Shi et al., 1987
Thai	Tai-Kadai	H, M, L, contour tones	Negative	Higher > Lower	Lower > Higher	Faytak & Yu, 2011
Eastern Cham	Chamic	Register	Negative	30-50 Hz	L > H	Brunelle, 2005

Table 5.5. Languages in which level tones are correlated with vowel duration.[4] Higher tone/register has shorter duration than lower tone/register. *Numerical values not available in reference.

[4] Values given under the Pitch Difference and Dur. Difference columns may be calculated using measurements reported in the original citations or estimates based on graphic representations of the data where no numerical values were reported.

Though modest in number, the languages in Table 5.5. represent nine language families from the Americas, Asia, and Africa. These languages also have diverse tone systems, ranging from simple two-tone and two-register systems, to systems where multiple level tones contrast with other contour tones, as in Gaoba Dong. Thus, the co-variation between pitch and duration seems not to be idiosyncratic to a specific language group or to a specific type of tone system, but is rather widespread.

In all of the tone languages listed, pitch is the primary cue to the contrast, with adjacent tonal categories having f0 differences of 15-75 Hz, at the least, in some contexts. The durational difference between tonal categories is much more variable, ranging from 5 ms, which is likely imperceptible, to 136 ms between the highest two tones in Gaoba Dong. While it is unclear whether all the durational differences between categories are statistically significant, and it is likely that a number of them are not large enough for listeners to use as perceptual cues, the relation between duration and pitch is remarkably consistent. In all but one language in Table 5.5., there is a negative correlation between the relative height of tone/register categories and vowel duration. That is to say, the categories with lower level pitch uniformly have longer durations and categories with higher level pitch have shorter durations. This relation is most striking in Gaoba Dong, a Tai-Kadai language spoken primarily in Southern China, in which the tone system comprises five level tones in addition to three contour tones. For each successively lower level tone, the vowel duration increases. Between the two lowest tones, where the pitch difference is fairly small, the durational difference is notably greater than the two intervals above. It should also be noted, that in Tahltan, a Na-Dené language of the Athabaskan family, the relation between pitch and duration is likely rooted in a different linguistic factor. Since the high tone is unmarked and the low tone is marked, low tones may be lengthened as a way of

preserving the marked category. As we will see, this is consistent with the pattern observed in Navajo, the other Na-Dené language in this list.

Three register languages, Hani, Cham, and Khmu' are also included in this table since pitch and duration are part of the constellation of cues that characterize these contrasts. In these languages as well, the two cues share a negative relation consistent with the other languages discussed above, where the register with higher pitch also has shorter duration.

Navajo, a Na-Dené language spoken primarily in Southwestern United States, is the only language in this list with a positive correlation between pitch and duration. High tones are realized with longer duration than low tones, particularly when they co-occur with another low tone in the same stem. However, as noted by McDonough (1999), this is confounded with tone markedness. Since, in Navajo, low tones are unmarked and more prevalent while high tones are marked and distributionally much rarer, lengthening syllables with high tones is likely a markedness preservation strategy. Recall that this is consistent with Tahltan, in which the low tone is marked, instead of the high tone.

Overall, the co-variation of duration to pitch levels in tone contrasts seems to be attested cross-linguistically and the correlation observed overwhelmingly aligns with predictions based on the way in which these cues are perceptually integrated.

Next, I turn to cases where pitch is a cue to phonemic length contrasts. However, as we will see, these cases are markedly different from those where duration co-varies with tone. In some languages with length contrasts on vowels, longer vowels have been known to be cued by pitch. Such is the case in several Finno-Ugric languages such as Estonian (Lippus et al., 2013), Finnish (Järvikivi et al., 2007; Vainio et al., 2010), and Livonian (Lehiste et al., 2007), and also in Japanese (Lehnert-LeHouillier, 2010; Takiguchi et al., 2010). Unlike in the tone and register

languages discussed above, where duration has a direct relationship to pitch levels in the tone system, the relationship of pitch to duration in the quantity systems of these languages is less straightforward. In Estonian for example, the three-way length contrast is phonologically restricted to stressed syllables (Asu & Teras, 2009), but is phonetically realized over the entire disyllabic foot via foot isochrony. That is, length differences on the (initial) stressed syllable are compensated by the (final) unstressed syllable such that all feet are of comparable duration. When the stressed syllable is produced with the H*+L pitch accent, it is the location of the pitch peak relative to the end of the accented syllable that is used to differentiate between long and super-long syllables (Lippus et al., 2013). In short and long syllables, the peak occurs at or just before the end of the accented syllable, whereas it occurs much earlier in super-long syllables, often in the first half. In Finnish, the binary length distinction is characterized by differences in pitch contour, rather than pitch alignment. Long (bimoraic) syllables are produced with a falling contour whereas short (monomoraic) syllables are produced with steady high pitch (Vainio et al., 2010). Moreover, differences in pitch contour clearly condition Finnish listeners' categorization of words as having short or long stressed syllables (Järvikivi et al., 2007). Similar effects are observed for Japanese listeners (Lehnert-LeHouillier, 2010; Takiguchi et al., 2010), who are more likely to identify a word as having a long syllable if the pitch on that syllable has a falling contour.

Given the patterns cited, we can call into question whether this is co-variation at all. In these languages, pitch cues only come into play in positions of prominence (e.g. stressed syllables that bear a pitch accent), but not in other positions. Additionally, the type of pitch cue attended to by listeners is not the pitch level, but rather pitch target alignment or pitch shape. Taken together, these facts suggest that the recruitment of pitch as another cue to length contrasts

is i) fairly restricted typologically, and ii) does not show that pitch bears any relation to duration in the way that was exemplified by the way that duration patterns with pitch contrasts.

Overall, pitch and duration do seem to co-vary cross-linguistically, but they do so asymmetrically. Duration co-varies with pitch in tone and register contrasts, where lower tones and registers have longer duration than higher tones and registers. However, pitch does not co-vary with duration in languages with length contrasts.

5.3. Diachronic contrast transfer

As described above, diachronic contrast shift is the process by which a phonological contrast signalled by one acoustic cue becomes signalled by another acoustic cue. Given that only pitch and breathiness/voicing are enhancing cues, I predict that only this set of cues would participate in contrast shift. That is, since pitch, breathiness, and voicing share a common auditory effect, one of these cues may replace the other as the primary cue to a phonological contrast. However, since breathiness and voicing are perceptually dependent on pitch but not vice versa, the direction of shifting should be predominantly from voicing and phonation to pitch. Conversely, contrast shifts between pitch and duration are predicted to be rare, since they do not contribute to the same auditory effect.

5.3.1. Pitch, breathiness, and voicing

In this section, I survey languages in which contrast shift occurred between pitch and breathiness/voicing. Once again, I am extending the scope of the survey beyond just pitch and breathiness to also include voicing because according to auditory enhancement theory, all three contribute to the same auditory effect of low frequency energy (e.g. Kingston, 2011).

Furthermore, there is general consensus that a common path for the development of contrastive pitch on the vowel is the loss of a voicing distinction in a preceding consonant (e.g. Hombert, 1978 and references therein), with voice quality playing a key role as an intermediate stage in these historical developments (Thurgood, 2002). Thus, all three cues are often involved at different stages of the same contrast transfer phenomenon and are included in this survey.

I begin in Section 5.3.1.1 by discussing languages in which contrastive tone has resulted from a historic voicing or phonation contrast. Then, in Section 5.3.1.2, I will discuss the singular case of a phonation contrast developing from a tone contrast.

5.3.1.1. Tones from voicing and breathiness

Contrast transfer resulting in contrastive tones is common and well-documented. These languages are listed in Table 5.6. These include languages that were atonal and became tonal as a result, as well as those in which the number of tones effectively doubled due to this process. In some of the languages listed, the reanalysis of a consonantal laryngeal contrast as a pitch contrast happened multiple times to produce the tone systems we observe today. In many languages, laryngeal contrasts other than voicing and breathy vs. modal phonation also conditioned tone (e.g. Kingston, 2011 on glottalization in Athabaskan), but these are beyond the scope of this survey and were thus excluded.

Language	Family	Contrast Shift	Source
Chinese	Sinitic	T → higher-pitched reflex of 4 tones D → lower-pitched reflex of 4 tones	Haudricourt, 1954
Eastern Kayah	Karenic	T → V → high-pitched reflex of 3 tones D → V̰ → low-pitched reflex of 3 tones	Kauffman, 1993

Sgaw Karen	Karenic	T → high-pitched reflex of 2 tones D → low-pitched reflex of 2 tones	Haudricourt, 1972
Tamang	Tibeto-Burman	T → higher-pitched, clear reflex of 2 tones D → lower-pitched, breathy reflex of 2 tones	Mazaudon & Michaud, 2008
Kurtöp	Tibeto-Burman	T, N̥ → high tone D, N → low tone	Hyslop, 2009
Dzongkha	Tibeto-Burman	NTV, NNV → TV́, NV́ NDV, NV → DV̀, NV̀	Hyslop, 2010
Eastern Cham	Chamic	T, N → V → 2 high tones D → V̤ → 2 low tones	Thurgood, 1996
Western Cham	Chamic	T → Register 1 (modal, higher pitch) D → Register 2 (breathy, lower pitch)	Thurgood, 1996
Tsat	Chamic	T, N → V → 3 high tones D → V̤ → 3 low tones	Thurgood, 1996
Utsat	Chamic	T (final) → higher 3 tones D (final) → lower 2 tones	Thurgood, 1992
Yabem	Malayo-Polynesian	T → low tone D → high tone	Ross, 1993
Khmu' Rawk	Mon-Khmer	Modal register → high tone Breathy register → low tone	Abramson et al., 2007
Northern and Western Kammu	Mon-Khmer	T, N̥ → high tone D, N → low tone	Svantesson & House, 2006
Vietnamese	Vietic	T → higher-pitched, clear reflex of 3 tones D → lower-pitched, breathy reflex of 3 tones	Thurgood, 2002
Southern Thai	Tai-Kadai	Tʰ → high tones ʔ → mid tones D → low tones	Brown, 1975
Kera	Chadic	T → high tone D → low tone	Pearce, 2005
Punjabi	Indo-Aryan	V{Dʱ, ɦ} → V̀{Dʱ, ɦ} {Dʱ, ɦ}V → {Dʱ, ɦ}V́	Bhatia, 1975
Afrikaans	Germanic	T → high tone D → low tone	Coetzee et al., 2018

Table 5.6. Languages in which tones have resulted from a historic voicing or phonation contrast. T = voiceless obstruent, D = voiced obstruent, N̥ = voiceless sonorant, N = voiced sonorant, V = vowel.

The languages from Table 5.6. are representative of several large language families. The majority of these – including Sino-Tibetan (e.g. Chinese, Karen, Dzongkha), Austronesian (e.g. Cham, Yabem), Austroasiatic (e.g. Vietnamese, Kammu), and Tai-Kadai – are spoken in Asia. Others, including Kera (East Chadic) and Afrikaans (Germanic), are spoken in Africa. Finally, Punjabi, spoken in the Indian sub-continent, is the only language representing the Indo-Aryan language family. A number of the languages listed are representative of numerous languages or dialects. The tone splitting that occurred in Chinese, for example, predates the off-shoot of most modern Chinese dialects. Thus, this single entry can be considered to encompass all of the sub-languages that developed afterward. Tamang and the other languages of the Tamangish branch of Tibeto-Burman – Gurung, Thakali, and Manangke – have similar synchronic tone systems, and all four languages underwent the same tonal division that split two tones into four (Mazaudon, 2005). Likewise, tone in other Chadic languages developed in a similar manner to Kera (Wolff, 1987). Finally, a number languages related to Punjabi are also tonal (see Baart, 2014), but since the path to tonogenesis has not been discussed explicitly for these other languages, I do not include them here. Baart (2014) does suggest, however, that there is a link between the development of tone in these languages and the loss of a series of breathy consonants.

Of course, there are obvious variations between the way in which tone developed in these languages. As mentioned previously, the shift caused some atonal languages to become tonal while the same process caused tonal languages to double their tonal inventory. In some languages, only obstruents were involved in the contrast shift, while in others, sonorants were also involved. The role of voice quality in the contrast shift is more apparent in some languages (e.g. Western Cham, Khmu' Rawk) than others. However, regardless of the minor variations from language to language, there is an undeniable consistency in the way that older voicing and

phonation contrasts transferred onto pitch. The emerging pattern is that as voicing categories merged, voiced categories became breathy and/or low pitched while voiceless categories became modal and/or high pitched. This pattern of contrast shift is predicted if voicing, breathiness, and pitch bear a perceptually enhancing relationship. Given that voicing, breathy phonation, and low pitch enhance the same auditory percept, we should expect listeners to associate low pitch and breathiness with the voiced category.

Note that some of these contrast shifts are still in progress (e.g. Khmu' Rawk, Kurtöp), while others were completed almost a millennium ago (e.g. Chinese, Vietnamese). The modern tonal reflexes in languages in which tonogenesis and tone splitting occurred early may no longer reflect the realizations of the tones immediately after the contrast shift (as can be seen in the vastly diverse tone systems in modern Chinese dialects). For these languages, the connection between modern tones and earlier consonant onsets was mainly established through reconstruction, though some vestiges of the link may still be found in their synchronic phonologies. For example, in Wuyi, a Chinese Wu dialect, voiced onset obstruents are devoiced if a high tone spreads across it (e.g. Yip, 2002). In Yabem (Ross, 1993), the onset of an underlyingly high-toned prefix becomes voiced if a low tone spreads to it from a root.

In the languages that have recently undergone or are currently undergoing contrast shift onto pitch, the connection between pitch level and consonant voicing is clearer. Yet, these languages are characterized by different traits indicating that the shift is not yet complete. In Afrikaans, there is a generational difference in the way that the contrast is being produced and perceived. While older speakers produce both a VOT and a pitch difference, younger speakers have more or less merged the two categories along the VOT dimension and are now relying on pitch to differentiate the two phonological categories (Coetzee et al., 2018). Similarly, in Khmu'

Rawk, Abramson et al. (2007) find that pitch has become the dominant cue to what was a register contrast, save in more conservative populations that continue to rely somewhat on the breathiness cue. In Northern and Western Kammu, the partiality of the tonogenetic process is evidenced in the fact that tones are only fully contrastive after sonorants and fricatives; after stops, tone remains predictable based on stop voicing.

Due to the scope of this survey, I am unable to discuss the developments in individual languages in more detail. However, the picture is much more complex than I have been able to convey. There are many language-internal and language-external factors that induced the sound changes described. These include, but are not limited to, monosyllabification (e.g. Chamic languages), onset consonant cluster reduction (e.g. Kurtöp), and language contact (e.g. Chamic languages). Additionally, as I have briefly mentioned, the final realizations of the tones and tone systems in many of these languages were also influenced by aspirates, fricatives, glottals, etc. occurring post-vocalically. The reader is referred to the sources cited for further details.

Finally, the universality of the "voiced-low" pattern has not gone unchallenged. Specifically, there are languages that show high tone reflexes after previously voiced consonants. However, the developmental patterns in these languages seem to be less direct. For example, in Shan, a Thai language spoken in Myanmar, the higher reflexes of three tones seem to have originated from syllables with voiced stops and the lower reflexes of three tones from syllables with voiceless stops. However, the voiced stops first developed into sonorants and voiceless stops, and the voiceless stops also first developed into sonorants (Kingston, 2011). These intervening developments therefore obscure the relation between tone and voicing. It has also been argued that the some of these atypical surface relations are a result of changes to the realization of tones after the contrast transfer occurred (cf. Kingston 2011 and references

therein). That is, the contrast transfer followed the normal cross-linguistic patterns, but tones later changed to obscure the historic relation.

Regardless, there is general consensus that the typical pattern for contrast shift from voicing to pitch is that low tones come from prevocalically voiced consonants and breathy phonation, and high tones come from prevocalically voiceless consonants and modal phonation. This typological observation is consistent with our understanding of how voicing, breathiness, and pitch co-vary synchronically, and both the diachronic and synchronic patterns discussed can be predicted by the fact that these three cues are perceptually enhancing.

5.3.1.2. Register or voicing from tone

There are, to my knowledge, no reports on contrast shift from pitch onto voicing. That is, no language is claimed to have developed a voicing contrast in place of a tone contrast.

It could be argued that some of the languages listed in Table 5.4. could have undergone contrast transfer from pitch onto phonation. Recall that these languages have a large number of tones. With crowded tone spaces, some of these contrasts are supported not by differences along the pitch dimension but by differences in phonation instead. However, it is difficult to discern whether these contrasts were fully distinguishable by pitch alone at any point in the language history, or whether the tones were accompanied by some phonatory trait from the beginning.

Perhaps a clearer case of contrast transfer from pitch to phonation occurred in the Otomanguean language Quiaviní Zapotec (Uchihara, 2016). Synchronically, this language has four tones (low, high, rising, falling), and four phonation categories (modal, breathy, creaky, and interrupted, which is characterized as modal with an intervening glottal stop), though tone and phonation have co-occurrence restrictions and are thus not fully crossed. A closely related, but

more conservative language, Güilá Zapotec, has the same tone contrasts, and almost the same phonation contrasts except that it lacks breathy phonation. Uchihara (2016) shows evidence that the modal vowel with low tone and the breathy vowel with low tone in Quiaviní Zapotec correspond to the modal rising tone and the modal low tone in Güilá Zapotec respectively. Thus, this is a case where an earlier pitch contrast between rising and low contours is now realized through phonation.

Assuming that these are true cases of contrast transfer from pitch to phonation, they are still characteristically different from the transfer from voicing and phonation onto pitch. In the latter, the merging of a voicing or phonation distinction might cause tones to arise in a limited context (e.g. only after stop consonants), but the effect generalizes until it causes a bifurcation of the entire vowel system. However, the instances of transfer from pitch to phonation discussed here have an overall limited effect on the system. Phonation contrasts may arise between two tones when their pitch differences are minimized, but the use of phonation does not generalize to other tones. That is, contrast transfer from voicing or breathiness onto pitch is much more frequent than the reverse.

5.3.2. Pitch and vowel duration

The focus of this section is on contrast transfer between pitch and vowel duration. Ratliff (2015) lists three languages that have developed pitch contrasts from length contrasts. One of these is Estonian, a Finno-Ugric language in which a rise-falling pitch contour now distinguishes the long and super-long length categories (Lippus et al., 2013). However, as discussed briefly in Section 5.2.2, this is not a case in which length distinctions were transferred to pitch levels. In Cem, an Austronesian language spoken in New Caledonia, comparative evidence shows that a

low tone unique to this language sometimes corresponds to long [a] in neighbouring languages (Rivierre, 1993). But this contrast shift has a very limited context, and there are other clearer sources for the low tone, such as from a following fricative or aspirated consonant. The most convincing case for contrast transfer from duration to pitch is from the Mon-Khmer language Hu. This language has a synchronic contrast between a high and a low tone that are modern reflexes of an earlier long-short vowel contrast (Svantesson, 1991).

As far as I am aware, there are no languages in which a pitch contrast has become a duration contrast, even though duration is frequently correlated with pitch levels. Given the sparsity of the data points we have, it is difficult to say whether there is an asymmetry in the direction of contrast transfer between these cues like there was between the enhancing cues.

Overall, it would seem that contrast transfer between pitch and duration occurs much less frequently than contrast transfer between voicing, phonation, and pitch. This has been noted repeated in the literature, particularly by those who work on tonogenesis. Transfer from consonant voicing (and phonation) onto pitch are described to be "attested extensively" (Hombert, 1978: 78) and "widespread" (Ratliff, 2015: 250), whereas transfer from length to pitch are "unorthodox" (Svantesson, 1991) and "less frequent" (Ratliff, 2015: 253). The implications of this bias will be discussed in Section 5.3.3.

5.3.3. Summary of diachronic contrast transfer

The main observations about contrast transfer between pitch, phonation, and voicing are summarized in (1) and contrast transfer between pitch and vowel duration in (2).

(1) Contrast transfer between pitch, breathiness and voicing

   a. Voicing/breathiness to pitch:
      - frequent, common path for tone development cross-linguistically
      - broad effect on the phonological system
      - transfer pattern reflects enhancing cue relations
   b. Pitch to voicing/breathiness:
      - infrequent
      - limited effect on the phonological system
      - arbitrary transfer of pitch category onto breathy phonation

(2) Contrast transfer between pitch and vowel duration
   a. Duration to pitch
      - Infrequent
   b. Pitch to duration
      - Unattested

Overall, transfer between enhancing cues occurs more frequently than non-enhancing cues, and transfer from breathiness (or voicing) onto pitch occurs more frequently than the reverse. These findings align with the experimental results from Chapters 2, 3, and 4. Listeners in these experiments naturally associated the enhancing cues, pitch and breathiness, which caused them to experience interference when shifting attention from breathiness onto pitch. Listeners did not, however, experience interference when shifting attention from pitch to breathiness, leading to the hypothesis that these two enhancing cues are asymmetrically dependent in perception. As for pitch and vowel duration, listeners required experience with their co-variation to show the same interference effects when shifting cue weights, indicating that these two cues, though perceptually integral, are not naturally associated with phonological contrast.

5.4. Conclusion

The purpose of this chapter was to check the hypotheses developed from the experimental findings against typological evidence. Since both enhancing and non-enhancing cues are perceptually integral, I predicted that co-variation between both sets of cues would be

attested cross-linguistically. This was true, though only asymmetrically in both cases. Pitch is used systematically as a cue to both voicing and breathiness, but not vice versa, and duration is used systematically as a cue to pitch, but not vice versa.

Given that only enhancing cues contribute to the same auditory percept, I hypothesized that this would allow for these cues to be involved in diachronic contrast transfers to the exclusion of non-enhancing cues. Additionally, I predicted that contrast transfers from breathiness (and voicing) onto pitch would be more frequent than the reverse given the hypothesis that breathiness is perceptually dependent on pitch but pitch is independent of breathiness. Both of these predictions were also supported. Thus, the typology of both synchronic cue co-variation and diachronic contrast transfer reflect the behavioural observations made in the experimental setting.

CHAPTER 6: CONCLUSION

6.1. Summary of findings

This dissertation has reported on the results of four cue weighting experiments designed to test for language-general auditory enhancement effects in speech perception, and one typological survey aimed at determining whether these perceptual effects influence cross-linguistic sound inventories.

The first two experiments used a novel cue weighting paradigm to compare experienced and inexperienced listeners' perception of a pair of enhancing cues (Chapter 2), pitch and breathiness, to their perception of a pair of non-enhancing cues (Chapter 3), pitch and vowel duration. Experienced listeners were speakers of Hani, who use the cues in both cue pairs to signal the tense-lax contrast. Inexperienced listeners were speakers of English, who do not use any of the selected cues phonemically. Listeners were made to shift attention between the cues in one of the two cue pairs. The mapping relation between the enhancing cues was manipulated such that it was either positive or negative, negative being congruent with the enhancing correlation and consistent with Hani listeners' experience. The mapping relation between the non-enhancing cues was similarly manipulated such that it was either positive or negative, negative being consistent with Hani listeners' experience.

My predictions were based on the assumption that all listeners associate enhancing cues naturally because of their shared auditory properties, without the need for exposure to cue co-variation through language experience. On the other hand, listeners only associate non-enhancing cues if they have language experience with the cue co-variation. Thus, Hani listeners shifting attention between both sets of cues were expected to have difficulty with the task when the mapping relation between the cues was the reverse of their linguistic experience. English

listeners' ability to shift attention between enhancing cues versus non-enhancing cue pairs was expected to differ. For enhancing cues, English listeners were predicted to show the same interference effects when the mapping relation was incongruent with the enhancing correlation between the cues, even though they have no experience with their co-variation. For non-enhancing cues, English listeners were predicted to have no difficulty shifting attention given either mapping condition.

The results from the first two experiments confirmed some of these predictions but deviated from them as well. Both Hani and English listeners' attentional shift between the non-enhancing cues could be predicted based on their experience (or inexperience) with the cue pair. Hani listeners successfully shifted attention from one non-enhancing cue to another when the mapping relation was consistent with their language experience, but were unsuccessful at shifting attention between the cues when the relation between the cues was inconsistent with their language experience. English listeners, who are not constrained by language experience, were able to shift attention between non-enhancing cues regardless of the mapping relation.

In contrast, as predicted, Hani and English listeners had difficulty shifting attention between the enhancing cues when the mapping relation was the reverse of the enhancing relation between pitch and breathiness. Note that this relation is also opposite of the co-variation observed in Hani. However, this was only true for listeners shifting attention from breathiness to pitch but not from pitch to breathiness. This directional asymmetry was not predicted.

Two more experiments were conducted to explore this asymmetry. The first of these addressed the question of whether the enhancement asymmetry could be due to unequal use of one of the two enhancing cues by English and Hani listeners. I extended the experiment on enhancing cues to two more groups of listeners who differ in their phonemic use of these cues.

These were tone listeners who use pitch phonemically and phonation listeners who use breathiness phonemically. If Hani and English listeners' asymmetric attentional shift was due to their heavier use of the pitch cue, then the tone listeners were predicted to exhibit the same asymmetry, but the phonation listeners were predicted to exhibit the opposite asymmetry. This was not the case. Both the tone and phonation listeners showed the same directional asymmetry as the English and Hani groups before, showing interference when shifting attention from breathiness to pitch but not vice versa. This was an indication that the enhancement asymmetry was not language-specific and experience-dependent, but language-general.

In the last experiment, I tested whether the directional asymmetry could be caused by an asymmetric perceptual dependency between pitch and breathiness. I proposed that listeners naturally associate changes in breathiness with changes in pitch, but pitch could be perceived as a relatively independently varying cue. Thus, interference in the cue shifting experiments was observed only in the direction of breathiness to pitch. To test for dependence asymmetries, I conducted an additional cue weighting experiment with the two enhancing cues where English listeners learned categories that were distinguished entirely by one of the cues and not by the other. Cue weights on the uninformative cue were compared. Based on the hypothesis, listeners who were trained on breathiness were expected to spontaneously increase the weight on the uninformative pitch cue, even in the absence of evidence, whereas the listeners who were trained on pitch were not expected to do so for the uninformative breathiness cue. The difference between these was expected to increase with training as listeners trained on pitch become more certain that breathiness could be down-weighted but listeners trained on breathiness continue to attend to pitch. This prediction was somewhat confirmed. There was no difference in cue weights to the uninformative cues for listeners with less training. For listeners with more training, cue

weights for uninformative pitch were marginally greater than cue weights for uninformative breathiness.

To summarize the results of the experimental portion of this dissertation, I provide evidence that non-enhancing cue association is learned through experience with co-variation, while enhancing cue association is natural, independent of experience, and language-general. I also provide evidence that the dependency between the specific enhancing cues selected, pitch and breathiness, is asymmetric, and propose that the percept of breathiness is dependent on pitch but pitch is independent of breathiness.

To further validate the experimental findings, I conducted a cross-linguistic survey of synchronic cue co-variation and diachronic cue transfer involving the enhancing and non-enhancing cue pairs studied. Overall, while both sets of cues co-vary synchronically to signal phonemic contrasts, only enhancing cues participate in diachronic contrast transfer, and overwhelmingly in the direction of voicing/phonation contrasts onto pitch contrasts.

6.2. Implications

The goal of this dissertation was to test the auditory enhancement theory of speech perception, which claims that listeners privilege cues that share the same auditory effect. The results described above provide evidence in favour of auditory enhancement in the following ways: First, I show that experience is not needed for listeners to perceptually associate two enhancing cues. This conclusion follows from the result that even inexperienced English listeners had difficulty shifting attention from breathiness to pitch when the enhancing relationship between the cues was reversed.

Second, listeners treat cue pairs that share auditory properties differently from cue pairs that perceptually integrate but do not contribute to the same auditory property. This difference was highlighted in the novel experimental paradigm that required listeners to attend to a new cue for the same contrast. English listeners showed perceptual interference when shifting attention between enhancing cues but not between non-enhancing cues, though they had no experience with either cue pair. If interference is an indication that listeners perceptually associate a pair of cues, then these results indicate that experience is not required for listeners to associate enhancing cues, but it is required for listeners to associate non-enhancing cues. One can also interpret these results to mean that enhancing cues are more prone to perceptual integration by listeners, for the obvious reason that they are auditorily more similar than other cue pairs. While much of the research in this area has equated auditory enhancement with perceptual integration, the results from these experiments showed that they are distinct.

Lastly, the typological observations made in this dissertation bolster the proposal by Kingston and colleagues that the privileged status of enhancing cues should be reflected in cross-linguistic patterns. While synchronic patterns show that both enhancing and non-enhancing cues co-vary, a survey of diachronic contrast transfer revealed that only enhancing secondary cues are able to replace the primary cue in signalling a contrast. This suggests that only those cues that contribute to the same auditory property trade off in signalling the same contrast.

The findings in this dissertation do not, however, support the claim that (all) enhancing cues should form a single intermediate perceptual property (IPP). In its strongest form, this claim predicts that listeners should not be able to perceptually tease apart two enhancing cues, and any interference between them should be completely symmetric. The experiments reported here show that listeners are able to separate pitch and breathiness, giving them differential cue

weights depending on their informativeness. If they were perceived as a single property, the cue weights for pitch and breathiness should have been the same in all conditions. Furthermore, the main, and somewhat surprising, result from the cue shifting experiments showed that the dependency between these two cues is asymmetric. Specifically, evidence from two experiments and the cross-linguistic typology suggest that breathiness is dependent on pitch but not vice versa.

This dissertation has addressed the question of cue dependency using perceptual evidence. But of course, any perceptual dependencies observed between the enhancing cues in the experiments could be rooted in either perception or articulation given that both are laryngeal cues. On one hand, we could argue that listeners associate breathiness to pitch because a greater amount of breathiness creates the same auditory percept as low pitch. On the other hand, these dependency effects would obtain if the same articulations that increase breathiness also lower pitch. One argument could be made in support of the former. Recall that the predictions made by the articulatory account are somewhat unclear. Depending on the exact mechanisms that come into play during voicing, producing breathiness could either lower or raise pitch (see Chapter 2, Section 2.1.2). For English speakers, the correlation between pitch and breathiness in production seems to be positive (Kreiman et al., 2007; Iseli et al., 2007), the exact opposite of the auditorily enhancing relationship between these cues. English speakers produce more breathy voice with higher pitch, but they perceptually associate breathy voice with lower pitch. If the dependency effect observed in the perception experiments described in this dissertation were rooted in articulation, we would expect the results to be congruent with the correlation in production. That is, English listeners in my experiments should have had shown interference when the cues were in a negative, enhancing mapping relation, rather than when the cues were in a positive, non-

enhancing mapping relation. Since the production data and the perception results from the experiments in this dissertation are at odds, we can more confidently say that the effects observed in the experiments were rooted in perception.

Overall, the findings suggest that models of speech perception should allow for cues to have different baseline propensities to be perceptually associated, which determine how much experience is needed for listeners to associate them. Cues that converge on the same auditory effect would be modeled with a higher propensity and require less experience with co-variation than those that do not. The fact that some non-enhancing cues are integrated but not others also suggests that there should be finer distinctions beyond the coarse division shown in this dissertation. More nuanced degrees of perceptual association likely extend to enhancing cues as well (Kingston, p.c.), but quantifying degrees of perceptual association will require more work beyond the scope of this dissertation. Lastly, associations between cues that form naturally should not have to be symmetric. Thus, the model should allow for asymmetries in cue dependencies for those cues with higher propensities to associate that require little experience. However, associations that are learned through experience with co-variation are necessarily symmetric, mirroring the distribution of two cues in the signal itself.

6.3. Future directions

There are of course, many limitations to the work presented in this dissertation. I list some of them here to motivate directions for future research.

First, this dissertation studies just one pair of enhancing cues and one pair of non-enhancing cues. While experimental and typological evidence point in the same direction, it is difficult to generalize the findings beyond these particular cues until more cue pairs have been

studied. As an extension of this research to additional cues, I would like to test nasalization and breathiness as enhancing cues. Unlike pitch and breathiness, these two acoustic dimensions are not controlled by the same set of articulators. Thus, studying these cues would be able to inform us of whether the effects observed are rooted in perception or in articulation.

Also, the Hani listeners in this experiment had less uniform experience with the cue pairs than expected, which may have had a direct impact on their cue perception. Recall that Hani listeners, like English listeners, showed a cue shifting asymmetry between pitch and breathiness. However, for a listener group that has experience with co-variation between two cues, what we expect is symmetric interference. One explanation might be that the listeners tested simply did not have input that was consistent enough for them to learn the association symmetrically, making their experience more like that of English listeners who, in a way, also have insufficient experience with the cues. The impact of type (e.g. phonemic vs. phonetic) and extent of language experience on perceptual association of cues can only be explored by studying more listener groups and having tighter controls on their language background.

Finally, in the typological survey, it was clear that pitch and breathiness/voicing participate in diachronic contrast shift while pitch and vowel duration do not. In this dissertation, I attributed this difference to the fact that the former cues are enhancing and converge on the same auditory effect but the latter do not. However, other differences between these cues could have made contrast shift possible between one set but not the other. In particular, Kingston (p.c.) points out that one major difference between pitch and breathiness/voicing and pitch and vowel duration is the segments they're realized on. In the first case, the cues undergoing transfer are properties of different segments, pitch belonging to the vowel, and breathiness and voicing coming from the preceding consonant. This disjunction, whether by speech unit or in time, may

be the condition necessary for listeners to i) disassociate two integral cues enough to realize they can be manipulated separately. In the latter case, the two cues, pitch and vowel duration are properties of the same segment, the vowel. Contrast transfer between two such cues may be difficult if they are treating the vowel as a single unit. Cues that make up the properties of that vowel can thus co-vary robustly synchronically, but one cue could not replace another cue as the primary to a contrast within the same segment. To tease these two accounts apart, we would need to broaden the research on diachronic contrast shift, which has primarily focused on tonogenesis and registrogenesis from consonant contrasts. It would be informative to compare the current findings to, for example, the synchronic and diachronic patterns between non-enhancing cues that are realized on different segments and between enhancing cues that are realized on the same segment. This typological work, of course, would go hand-in-hand with the expansion of experimental work to more cue pairs.

Appendix A: Breathiness and Pitch values

Training Tokens

Distinctive Breathiness				
f0	H1-H2		f0	H1-H2
113	23.58		109	5.78
115	23.58		111	5.78
115	24.69		110	6.90
114	25.50		110	7.71
113	25.80		109	8.01
112	25.50		108	7.71
112	24.69		107	6.90
111	23.58		107	5.78
112	22.46		107	4.67
112	21.65		108	3.86
113	21.35		109	3.56
114	21.65		110	3.86
115	22.46		110	4.67
117	23.58		112	5.78
116	25.80		112	8.01
115	27.43		111	9.64
113	28.03		109	10.23
111	27.43		107	9.64
110	25.80		106	8.01
110	23.58		105	5.78
110	21.35		106	3.56
111	19.72		107	1.93
113	19.13		109	1.33
115	19.72		111	1.93
116	21.35		112	3.56
119	23.58		114	5.78
118	26.91		113	9.12
109	26.91		104	9.12
108	23.58		103	5.78
109	20.24		104	2.45
118	20.24		113	2.45
121	23.58		116	5.78
120	28.03		115	10.23
107	28.03		102	10.23
106	23.58		101	5.78
107	19.13		102	1.33
120	19.13		115	1.33
122	23.58		118	5.78
104	23.58		100	5.78
124	23.58		120	5.78
102	23.58		98	5.78
126	23.58		121	5.78
101	23.58		96	5.78

Distinctive Pitch				
f0	H1-H2		f0	H1-H2
118	17.46		104	11.90
120	17.46		106	11.90
120	18.57		105	13.01
119	19.39		105	13.83
118	19.68		104	14.12
117	19.39		103	13.83
117	18.57		102	13.01
116	17.46		102	11.90
117	16.35		102	10.79
117	15.53		103	9.97
118	15.24		104	9.68
119	15.53		105	9.97
120	16.35		105	10.79
122	17.46		107	11.90
121	19.68		107	14.12
120	21.31		106	15.75
118	21.91		104	16.35
116	21.31		102	15.75
115	19.68		101	14.12
115	17.46		100	11.90
115	15.24		101	9.68
116	13.61		102	8.05
118	13.01		104	7.45
120	13.61		106	8.05
121	15.24		107	9.68
121	23.24		106	17.68
118	24.13		104	18.57
116	23.24		101	17.68
116	11.68		101	6.12
118	10.79		104	5.23
121	11.68		106	6.12
122	25.17		107	19.60
118	26.36		104	20.80
115	25.17		100	19.60
115	9.76		100	4.19
118	8.56		104	3.00
122	9.76		107	4.19
118	28.58		104	23.02
118	6.34		104	0.78
118	30.81		104	25.25
118	4.11		104	-1.45
118	33.03		104	27.47
118	1.89		104	-3.67

Test Tokens

f0	H1-H2
111	29.36
111	28.14
111	26.91
111	25.69
111	24.47
111	23.24
111	22.02
111	20.80
111	19.57
111	18.35
111	17.13
111	15.90
111	14.68
111	13.46
111	12.23
111	11.01
111	9.79
111	8.56
111	7.34
111	6.12
111	4.89
111	3.67
111	2.45
111	1.22
111	0.00

f0	H1-H2
123	14.68
122	14.68
121	14.68
120	14.68
119	14.68
118	14.68
117	14.68
116	14.68
115	14.68
114	14.68
113	14.68
112	14.68
111	14.68
110	14.68
109	14.68
108	14.68
107	14.68
106	14.68
105	14.68
104	14.68
103	14.68
102	14.68
101	14.68
100	14.68
99	14.68

Appendix B: Vowel Duration and Pitch values

Training Tokens

Distinctive Vowel Duration				
f0	Dur		f0	Dur
113	324.2		109	193.9
115	324.2		111	193.9
115	334.8		110	200.3
114	342.7		110	205.0
113	345.7		109	206.8
112	342.7		108	205.0
112	334.8		107	200.3
111	324.2		107	193.9
112	313.9		107	187.8
112	306.6		108	183.4
113	304.0		109	181.9
114	306.6		110	183.4
115	313.9		110	187.8
117	324.2		112	193.9
116	345.7		112	206.8
115	362.3		111	216.8
113	368.6		109	220.5
111	362.3		107	216.8
110	345.7		106	206.8
110	324.2		105	193.9
110	304.0		106	181.9
111	290.1		107	173.5
113	285.1		109	170.6
115	290.1		111	173.5
116	304.0		112	181.9
119	324.2		114	193.9
118	357.0		113	213.5
109	357.0		104	213.5
108	324.2		103	193.9
109	294.4		104	176.1
118	294.4		113	176.1
121	324.2		116	193.9
120	368.6		115	220.5
107	368.6		102	220.5
106	324.2		101	193.9
107	285.1		102	170.6
120	285.1		115	170.6
122	324.2		118	193.9
104	324.2		100	193.9
124	324.2		120	193.9
102	324.2		98	193.9
126	324.2		121	193.9
101	324.2		96	193.9

Distinctive Pitch				
f0	Dur		f0	Dur
118	271.7		104	235.1
120	271.7		106	235.1
120	280.6		105	242.8
119	287.2		105	248.6
118	289.7		104	250.7
117	287.2		103	248.6
117	280.6		102	242.8
116	271.7		102	235.1
117	263.1		102	227.7
117	257.0		103	222.4
118	254.8		104	220.5
119	257.0		105	222.4
120	263.1		105	227.7
122	271.7		107	235.1
121	289.7		107	250.7
120	303.7		106	262.8
118	308.9		104	267.4
116	303.7		102	262.8
115	289.7		101	250.7
115	271.7		100	235.1
115	254.8		101	220.5
116	243.1		102	210.4
118	238.9		104	206.8
120	243.1		106	210.4
121	254.8		107	220.5
121	321.0		106	277.8
118	329.4		104	285.1
116	321.0		101	277.8
116	229.9		101	199.0
118	224.1		104	193.9
121	229.9		106	199.0
122	339.4		107	293.7
118	351.3		104	304.0
115	339.4		100	293.7
115	217.5		100	188.2
118	210.1		104	181.9
122	217.5		107	188.2
118	374.6		104	324.2
118	197.1		104	170.6
118	399.4		104	345.7
118	184.8		104	159.9
118	425.9		104	368.6
118	173.3		104	150.0

Test Tokens

f0	Dur
111	380.6
111	367.6
111	355.1
111	342.9
111	331.2
111	319.9
111	308.9
111	298.4
111	288.2
111	278.3
111	268.8
111	259.6
111	250.7
111	242.2
111	233.9
111	225.9
111	218.2
111	210.7
111	203.5
111	196.5
111	189.8
111	183.3
111	177.1
111	171.0
111	165.2

f0	Dur
123	250.7
122	250.7
121	250.7
120	250.7
119	250.7
118	250.7
117	250.7
116	250.7
115	250.7
114	250.7
113	250.7
112	250.7
111	250.7
110	250.7
109	250.7
108	250.7
107	250.7
106	250.7
105	250.7
104	250.7
103	250.7
102	250.7
101	250.7
100	250.7
99	250.7

REFERENCES

Abramson, A. S. (1962) The vowels and tones of standard Thai: acoustical measurements and experiments. (Publication No. 20, Indiana University Research Center in Anthropology, Folklore and Linguistics). *International Journal of American Linguistics 28*(2): Part II.

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 25-33.

Abramson, A. S., Nye, P. W., & Luangthongkum, T. (2007). Voice register in Khmu': Experiments in production and perception. *Phonetica*, *64*(2-3), 80-104.

Abramson, A. S., Theraphan, L., & Nye, P. W. (2004). Voice register in Suai (Kuai): An analysis of perceptual and acoustic data. *Phonetica*, *61*(2-3), 147-171.

Abramson, A. S., Tiede, M. K., & Luangthongkum, T. (2015). Voice register in Mon: Acoustics and electroglottography. *Phonetica*, *72*(4), 237-256.

Alderete, J. (2005). On tone and length in Tahltan (Northern Athabaskan). *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, *269*, 185.

Alwan, A., Jiang, J., & Chen, W. (2011). Perception of place of articulation for plosives and fricatives in noise. *Speech Communication*, *53*(2), 195-209.

Andruski, J. E. (2006). Tone clarity in mixed pitch/phonation-type tones. *Journal of Phonetics*, *34*(3), 388-404.

Andruski, J. E., & Ratliff, M. (2000). Phonation types in production of phonological tone: The case of Green Mong. *Journal of the International Phonetic Association*, *30*(1-2), 37-61.

Antoñanzas-Barroso, N., Kreiman, J., & Gerratt, B. (2006) Voice Synthesis. [computer software].

Asu, E. L., & Teras, P. (2009). Estonian. *Journal of the International Phonetic Association*, *39*(3), 367-372.

Baart, J. (2014). Tone and stress in north-west Indo-Aryan. In Caspers Johanneke, Yiya Chen, Willemijn Heeren,Jos Pacilly, Niels O. Schiller & Ellen van Zanten (eds.), *Above and beyond the segments: Experimental linguistics and Phonetics*, *Amsterdam: John Benjamins*, 1-13.

Babel, M., McGuire, G., & King, J. (2014). Towards a more nuanced view of vocal attractiveness. *PloS one*, *9*(2), e88616.

Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-28. <http://lme4.rforge.r-project.org/>.

Bhatia, T. K. (1975). The evolution of tones in Punjabi. *Studies in the Linguistic Sciences 5*(2), 12-24.

Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. *Speech communication group working papers, 1,* 71-81.

Bijankhan, M., & Nourbakhsh, M. (2009). Voice onset time in Persian initial and intervocalic stop production. *Journal of the International Phonetic Association, 39*(3), 335-364.

Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics, 30*(2), 163-191.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics, 18*(1), 37-49.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America, 66*(4), 1001-1017.

Bochner, J. H., Snell, K. B., & MacKenzie, D. J. (1988). Duration discrimination of speech and tonal complex stimuli by normally hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 84*(2), 493-500.

Boersma, P. & Weenink, D. (2016). Praat: doing phonetics by computer [computer software].

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Attention, Perception, & Psychophysics*, 61(5), 977-985.

Brigner, W. L. (1988). Perceived duration as a function of pitch. *Perceptual and motor skills, 67*(1), 301-302.

Brown, J. M. (1975). The great tone split: Did it work in two opposite ways. *Studies in Tai linguistics in honor of William J. Gedney*, 33-48.

Brunelle, M. (2005). *Register in Eastern Cham: Phonological, phonetic and sociolinguistic approaches*. Doctoral dissertation, Cornell University.

Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics, 37*(1), 79-96.

Brunelle, M. (2012). Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham. *The Journal of the Acoustical Society of America 131*(4), 3088-3102.

Brunelle, M., & Kirby, J. (2016). Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass, 10*(4), 191-207.

Carignan, C. (2017). Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels. *Journal of Phonetics, 63*, 87-105.

Carignan, C., Shosted, R., Shih, C., & Rong, P. (2011). Compensatory articulation in American English nasalized vowels. *Journal of Phonetics*, *39*(4), 668-682.

Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction?. *Journal of Phonetics*, *39*(4), 612-625.

Cho, T., Jun, S. A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, *30*(2), 193-228.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804-809.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, *66*, 185-216.

DiCanio, C. T. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association*, *39*(2), 162-188.

DiCanio, C. T. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics*, *40*(1), 162-176.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, *27*(4), 769-773.

Diehl, R. L., Castleman, W. A., & Kingston, J. (1995). On the internal perceptual structure of phonological features: The [voice] distinction. *The Journal of the Acoustical Society of America*, *97*(5), 3333-3334.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*(2), 121-144.

Diehl, R. L., & Molis, M. R. (1995). Effect of Fundamental Frequency on Medial [+ Voice]/[– Voice] Judgments. *Phonetica*, *52*(3), 188-195.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77-95.

Dunstan, E. & Igwe G.E. (1966) Two views of the phonology of the Ohuhu dialect of Igbo. *Journal of West African Languages 3.2.*, 71-75.

Edmondson, J., & Lợi, N. V. (1997). Tones and voice quality in modern northern Vietnamese: instrumental case studies. *Mon-Khmer Studies*, *28*, 1-18.

Esposito, C. M. (2006) *The effects of linguistic experience on the perception of phonation*. Doctoral dissertation, UCLA.

Esposito, C. M. (2010). Variation in contrastive phonation in Santa Ana del Valle Zapotec. *Journal of the International Phonetic Association*, *40*(2), 181-198.

Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics*, *40*(3), 466-476.

Faytak, M., & Yu, A. C. L. (2011). A Typological Study of the Interaction between Level Tones and Duration. In *ICPhS* (pp. 659-662).Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops?. *The Journal of the Acoustical Society of America*, *120*(5), 2884-2895.

Fischer-Jørgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics*, 28, 71-139.

Franich, K. H. (2016). Internal and contextual cues to tone perception in Medʉmba. *The Journal of the Acoustical Society of America*, *140*(1), EL107-EL112.

Freese, J., & Maynard, D. W. (1998). Prosodic features of bad news and good news in conversation. *Language in Society*, *27*(2), 195-219.

Gandour, J. (1974). Consonant types and tone in Siamese. *University of California Working Papers in Phonetics*, *27*, 92-117.

Gao, J., & Hallé, P. A. (2013). Duration as a secondary cue for perception of voicing and tone in Shanghai Chinese. In *INTERSPEECH* (pp. 3157-3161).

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, *41*(2), 185-205.

Garellek, M., Keating, P., & Esposito, C. M. (2012). Relative importance of phonation cues in White Hmong tone perception. In *Annual Meeting of the Berkeley Linguistics Society 38*, (pp. 179-188).

Garellek, M., Keating, P., Esposito, C. M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America*, *133*(2), 1078-1089.

Garellek, M., Ritchart, A., & Kuang, J. (2016). Breathy voice during nasality: A cross-linguistic study. *Journal of Phonetics*, *59*, 110-121.

Garellek, M., Samlan, R., Gerratt, B. R., & Kreiman, J. (2016). Modeling the voice source in terms of spectral slopes. *The Journal of the Acoustical Society of America*, *139*(3), 1404-1410.

Garner, W. R. (1974). *The Processing of Information Structure* (Erlbaum Associates, Potomac, MD).

Gordon, M. (1998). The phonetics and phonology of non-modal vowels: a cross-linguistic perspective. In *Annual Meeting of the Berkeley Linguistics Society 24*(1), pp. 93-105.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383-406.

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, *25*(1), 1-42.

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

Gradin, D. (1966). Consonantal tone in Jeh phonemics. *Mon-Khmer Studies*, *2*, 41-53.

Gregerson, K. J. (1976). Tongue-root and register in Mon-Khmer. *Oceanic Linguistics Special Publications*, (13), 323-369.

Gussenhoven, C., & Zhou, W. (2013). Revisiting pitch slope and height effects on perceived duration. In *INTERSPEECH* (pp. 1365-1369).

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America*, *125*(1), 425-441.

Haudricourt, A. G. (1954). Comment reconstruire le chinois archaique. *Word* 10, 351-364.

Haudricourt, A. G. (1972). Two-way and three-way splitting of tonal systems in some Far Eastern languages. *Tai Phonetics and Phonology*, 58-86.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360-378.

Henderson, E. J. (1952). The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies*, *14*(1), 149-174.

Henderson, E. J. (1979). Bwe Karen as a two-tone Language. *South-east Asian Linguistic Studies*, *3*, 301-326.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, *37*(4), 769-778.

Hirano, M. (1981). Clinical Examination of Voice (Springer-Verlag, New York).

Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, *33*(5), 353-367.

Hombert, J. M. (1976). Consonant types, vowel height and tone in Yoruba. *UCLA Working Papers in Phonetics*, *33*, 40-54.

Hombert, J. M. (1978). Consonant types, vowel quality, and tone. In V.A. Fromkin (Ed.), *Tone: A Linguistic Survey*, (pp. 77-112).

Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, 37-58.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059-3071.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America 109*, 764-774.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *50*(3), 346-363.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, *25*(1), 105-113.

Hyman, L. (1976). On some controversial questions in the study of consonant types and tone. *UCLA Working Papers in Phonetics: Studies on Perception and Production of Tone*, 90-98.

Hyslop, G. (2009). Kurtop tone: A tonogenetic case study. *Lingua, 119*, pp. 827-845.

Hyslop, G. (2010). Tone and tonogenesis in Bhutan: degrees of tonality? *University of British Columbia Working Papers in Linguistics, 45*, 114-24.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939.

Idemaru, K., & Holt, L. L. (2014). Specificity of dimension-based statistical learning in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 1009.

Iseli, M, Shue, Y-L. & Alwan A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America 121*, 2283–2295.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267-3278.

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

Järvikivi, J., Aalto, D., Aulanko, R., & Vainio, M. (2007). Perception of vowel length: Tonality cues categorization even in a quantity language. In *Proceedings of ICPhS XVI* (pp. 693-696).

Jensen, M. K. (1958). Recognition of word tones in whispered speech. *Word*, *14*(2-3), 187-196.

Jessen, M. (2001). Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants. In *Distinctive Feature Theory*, *2*, 237.

Kagaya, R., & Hirose, H. (1975). Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 9*, 27-46.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, *45*, 76-90.

Kauffman, W. G. (1993). *The Great Tone Split and Central Karen*. Master's thesis, University of North Dakota.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 286-319.

Keating, P., Esposito, C., Garellek, M., Khan, S., and Kuang, J. (2011). Phonation contrasts across languages, in *Proceedings of ICPhS XVII* (pp. 1046-1049).

Keating, P., Kuang, J., Esposito, C., Garellek, M., and Khan, S. (2012). Multi-dimensional phonetic space for phonation contrasts, Poster presented in *LabPhon 13* (Stuttgart, Germany).

Khan, A. Q. (2017). The tonal system of Pahari. *Acta Linguistica Academica. An International Journal of Linguistics (Until 2016 Acta Linguistica Hungarica)*, *64*(2), 313-324.

Khan, S. U. D. (2012). The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics*, *40*(6), 780-795.

Kingston, J. (1992). The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech*, *35*(1-2), 99-113.

Kingston, J. (2007). Segmental influences on F0: Automatic or controlled?. *Tones and Tunes*, *2*, 171-210.

Kingston, J. (2011). Tonogenesis. *Companion to Phonology*. Malden, MA: Wiley-Blackwell, 2304-2333.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419-454.

Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, *36*(1), 28-54.

Kingston, J., & Macmillan, N. A. (1995). Integrality of nasalization and F1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *The Journal of the Acoustical Society of America*, *97*(2), 1261-1285.

Kingston, J., Macmillan, N. A., Dickey, L. W., Thorburn, R., & Bartels, C. (1997). Integrality in the perception of tongue root position and voice quality in vowels. *The Journal of the Acoustical Society of America*, *101*(3), 1696-1709.

Kirby, J. P. (2018). Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics*, *71*, 326-354.

Kirby, J. P., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel F 0: Evidence from "true voicing" languages. *The Journal of the Acoustical Society of America*, *140*(4), 2400-2411.

Klatt, D. H., & Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In *Structure and Process in Speech Perception* (pp. 69-89). Springer, Berlin, Heidelberg.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, *87*(2), 820-857.

Kohler, K. J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica*, *41*(3), 150-174.

Kollmeier, B., Brand, T., & Meyer, B. (2008). Perception of speech and sound. In *Springer Handbook of Speech Processing* (pp. 61-82). Springer Berlin Heidelberg.

Kreiman, J., & Gerratt, B. R. (2010). Perceptual sensitivity to first harmonic amplitude in the voice source. *The Journal of the Acoustical Society of America*, *128*(4), 2085-2089.

Kreiman, J., Gerratt B. R. & Antonanzas-Barroso, N. (2007). Measures of the glottal source spectrum. *Journal of Speech, Language, and Hearing Research* 50, 595–610.

Kuang, J. (2011). Phonation Contrast in Two Register Contrast Languages and Its Influence on Vowel Quality and Tone. In *Proceedings of ICPhS XVII* (pp. 1146-1149).

Kuang, J. (2013). *Phonation in tonal contrasts*. Doctoral dissertation, UCLA.

Kuang, J. (2013). The tonal space of contrastive five level tones. *Phonetica*, 70(1-2), 1-23.

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America*, *142*(3), 1693-1706.

Kuang, J., & Keating, P. (2012). Glottal articulations of phonation contrasts and their acoustic and perceptual consequences. *UCLA Working Papers in Phonetics*, *111*, 123-161.

Kuang, J., & Keating, P. (2014). Vocal fold vibratory patterns in tense versus lax phonation contrasts. *The Journal of the Acoustical Society of America*, *136*(5), 2784-2797.

Kuang, J., & Liberman, M. (2015). Influence of spectral cues on the perception of pitch height. *Proceeding of ICPhS XVIII*.

Lai, Y., Huff, C., Sereno, J., & Jongman, A. (2009). The raising effect of aspirated prevocalic consonants on F0 in Taiwanese. In *Proceedings of the 2nd International Conference on East Asian Linguistics*. Simon Fraser University Vancouver, Canada.

Lea, W. A. (1973). Segmental and suprasegmental influences on fundamental frequency contours. In L.M. Hyman (Ed.), *Consonant Types and Tones*, *1*, (pp. 15-70).

Lee, S., & Katz, J. (2016). Perceptual integration of acoustic cues to laryngeal contrasts in Korean fricatives. *The Journal of the Acoustical Society of America*, *139*(2), 605-611.

Lefkowitz, L. M. (2017). *Maxent Harmonic Grammars and Phonetic Duration*. Doctoral dissertation, UCLA.

Lehet, M., & Holt, L. L. (2016). Adaptation to accent is proportionate to the prevalence of accented speech. *The Journal of the Acoustical Society of America*, *139*(4), 2164.

Lehiste, I. (1976). Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, *4*(2), 113-117.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America*, *33*(4), 419-425.

Lehiste, I., Teras, P., Pajusalu, K., & Tuisk, T. (2007). Quantity in Livonian: preliminary results. *Linguistica Uralica*, *43*(1), 29-44.

Lehnert-LeHouillier, H. (2010). A cross-linguistic investigation of cues to vowel length perception. *Journal of Phonetics*, *38*(3), 472-482.

Lenth, R.V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software 69* (1), 1–33.

Lew, S., & Gruber, J. (2016). An acoustic analysis of tone and register in Louma Oeshi. *Proceedings of the Linguistic Society of America*, *1*, 33-1.

Li, X. and Pastore, R.E. (1995). Perceptual constancy of a global spectral property: Spectral slope discrimination. *The Journal of the Acoustical Society of America 98* (4), 1956-1968.

Lippus, P., Asu, E. L., Teras, P., & Tuisk, T. (2013). Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian. *Journal of Phonetics*, *41*(1), 17-28.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling/b/versus/p/in trochees. *Language and Speech*, *29*(1), 3-11.

Lisker, L. (1978). In qualified defense of VOT. *Language and Speech*, *21*(4), 375-383.

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1783.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and speech*, *47*(2), 109-138.

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

Löfqvist, A. (1975). Intrinsic and extrinsic Fo variations in Swedish tonal accents. *Phonetica*, *31*(3-4), 228-247.

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

Lutfi, R. A. (1993). A model of auditory pattern analysis based on component-relative-entropy. *The Journal of the Acoustical Society of America*, *94*(2), 748-758.

Macaulay, M. A. (1996). *A Grammar of Chalcatongo Mixtec* (Vol. 127). University of California Press.

Maddieson, I. (1974). A note on tone and consonants. *The Tone Tome: Studies on Tone from the UCLA Tone Project*, 18-27.

Maddieson, I., & Hess, S. (1986). 'Tense' and 'lax' revisited: more on phonation type and pitch in minority languages in China. *UCLA Working Papers in Phonetics, 63*, 103-109.

Maddieson, I., & Ladefoged, P. (1985). 'Tense' and 'lax' in four minority languages of China. *Journal of Phonetics 13*, 433–454.

Maeda, S. (1993). Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In *Nasals, Nasalization, and the Velum* (pp. 147-167). Academic Press.

Mayo, C., Clark, R. A., & King, S. (2011). Listeners' weighting of acoustic cues to synthetic speech naturalness: A multidimensional scaling analysis. *Speech Communication, 53*(3), 311-326.

Mazaudon, M. (2005). On tone in Tamang and neighbouring languages: synchrony and diachrony. In Proceedings of *Cross-linguistic studies of tonal phenomena*, 79-96.

Mazaudon, M. & Michaud, A. (2008) Tonal contrasts and initial consonants: A case study of Tamang, a 'missing link' in tonogenesis. *Phonetica*, *65*(4), 231-256.

McDonough, J. (1999). Tone in Navajo. *Anthropological Linguistics 41*(4), 503–540.

Meacham, M. (1991). The phonetics of tone in Chalcatongo Mixtec couplets. In *Papers from the American Indian Languages Conference, University of California, Santa Cruz, July and August* (pp. 156-167).

Michaud, A., Vaissière, J., & Nguyễn, M. C. (2015). Phonetic insights into a simple level-tone system: 'careful' vs. 'impatient' realizations of Naxi High, Mid and Low tones. In *Proceedings of ICPhS XVIII*.

McKinley, S. C., & Nosofsky, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human perception and performance*, *22*(2), 294.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46, 505–512.

Misnadin, M. (2016). *Phonetics and phonology of the three-way laryngeal contrast in Madurese*. Doctoral dissertation, University of Edinburgh.

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

Mohr, B. (1968). Intrinsic fundamental frequency variation. *Monthly Internal Memorandum. University of California, Berkeley*.

Mohr, B. (1971). Intrinsic variations in the speech signal. *Phonetica*, *23*(2), 65-93.

Moisik, S. R., Czaykowska-Higgins, E., & Esling, J. H. (2019). Phonological potentials and the lower vocal tract. *Journal of the International Phonetic Association*, 1-35.

Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *The Journal of the Acoustical Society of America*, *98*(6), 3080-3095.

Ohala, J. J. (1972). How is pitch lowered?. *The Journal of the Acoustical Society of America*, *52*(1A), 124-124.

Ohala, J. J. (1978). The production of tone. In V.A. Fromkin (Ed.), *Tone: A linguistic survey*, Academic Press, New York, pp. 5-39.

Ohala, J. J., & Ewan, W. G. (1973). Speed of pitch change. *The Journal of the Acoustical Society of America*, *53*(1), 345-345.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America*, *75*(1), 224-230.

Pearce, M. (2005). Kera tone and voicing. *University College London Working Papers in Linguistics*, *17*, 61.

Pisoni, D. B. (1976). Fundamental frequency and perceived vowel duration. *The Journal of the Acoustical Society of America*, *59*(S1), S39-S39.

Polka, L., & Strange, W. (1985). Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *The Journal of the Acoustical Society of America*, *78*(4), 1187-1197.

R Development Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.

Ratliff, M. (2015). Tonoexodus, tonogenesis, and tone change. *The Oxford Handbook of Historical Phonology*, 245-261.

Remijsen, B. & Ladd, D. (2008). The tone system of the Luanyjang dialect of Dinka. *Journal of African Languages and Linguistics*, 29(2), pp. 173-213.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81.

Rivierre, J. C. (1993). Tonogenesis in New Caledonia. *Oceanic Linguistics Special Publications*, (24), 155-173.

Robbins, F. E. (1968). *Quiotepec Chinantec Grammar* (Vol. 4). Museo Nacional de Antropología.

Ross, M. D. (1993). Tonogenesis in the North Huon Gulf chain. *Oceanic Linguistics Special Publications*, (24), 133-153.

Russell, M. (2000). Phonetic Aspects of Tone Displacement in Zulu. *CLS 36: The Main Session*, 427– 439.

Sandor, A. (2004). *Perceptual interactions of duration with pitch and rate of change in pitch: Implications for sonification*. Master's thesis, Rice University.

Shi, F., Shi, L., Liao, R. (1987). An experimental analysis of the five level tones of the Gaoba Dong language. *Journal of Chinese Linguistics 15*, 335–361.

Shimizu, K. (1989). A cross-language study of voicing contrasts of stops. *Studia Phonologica, 23*, pp. 1-12.

Shosted, R., Carignan, C., & Rong, P. (2012). Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi. *The Journal of the Acoustical Society of America*, *131*(1), 455-465.

Shosted, R., Smith, J., & Ihsane, T. (2015). Nasal vowels are not [+ nasal] oral vowels. In *Romance Linguistics 2012: Selected papers from the 42nd Linguistic Symposium on Romance Languages (LSRL)* (pp. 63-76). Amsterdam: Jon Benjamins.

Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011) VoiceSauce: A program for voice analysis. In *Proceedings of the ICPhS XVII* (pp. 1846-1849).

Silverman, D. (1997). Laryngeal complexity in Otomanguean vowels. *Phonology 14*(2), 235-261.

Silverman, D. (2003). Pitch discrimination during breathy versus modal phonation. *Phonetic Interpretation: Papers in Laboratory Phonology VI*, 293–304.

Stevens, K. N., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language 65*(1), 81-106.

Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*(1), 10-19.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3), 185-190.

Svantesson, J. O. (1991). Hu–a language with unorthodox tonogenesis. *Austroasiatic Languages, Essays in honour of HL Shorto*, 67-80.

Svantesson, J. O., & House, D. (2006). Tone production, tone perception and Kammu tonogenesis. *Phonology*, *23*(2), 309-333.

Takiguchi, I., Takeyasu, H., & Giriko, M. (2010). Effects of a dynamic F0 on the perceived vowel duration in Japanese. In *Speech Prosody 2010-Fifth International Conference*.

Tehrani, H. (2015). Appsobabble. [Computer Software].

Thongkum, T. L. (1987). Phonation types in Mon-Khmer languages. *UCLA Working Papers in Phonetics*, *67*, 29-48.

Thongkum, T. L. (1988). Phonation types in Mon-Khmer languages. *Voice production: Mechanisms and Functions*, 319-333.

Thurgood, G. (1992). From atonal to tonal in Utsat (a Chamic language of Hainan). In *Annual Meeting of the Berkeley Linguistics Society 18*(2), pp. 145-156.

Thurgood, G. (1996). Language contact and the directionality of internal drift: the development of tones and registers in Chamic. *Language*, 1-31.

Thurgood, G. (2002). Vietnamese and tonogenesis. *Diachronica* 19, 333-363.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive science*, *34*(3), 434-464.

Uchihara, H. (2016). Tone and registrogenesis in Quiaviní Zapotec. *Diachronica*, *33*(2), 220-254.

Umeda, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech. *The Journal of the Acoustical Society of America*, *70*(2), 350-355.

Vainio, M., Järvikivi, J., Aalto, D., & Suni, A. (2010). Phonetic tone signals phonological quantity and word structure. *The Journal of the Acoustical Society of America*, *128*(3), 1313-1321.

Van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics*, *32*(4), 455-491.

Watkins, J. (2002). *The phonetics of Wa: Experimental phonetics, phonology, orthography and sociolinguistics*. Pacific linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Wayland, R., Gargash, S., & Longman, A. (1995). Acoustic and perceptual investigation of breathy voice. *The Journal of the Acoustical Society of America*, *97*(5), 3364-3364.

Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics*, *31*(2), 181-201.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F 0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America*, *93*(4), 2152-2159.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*(1), 25-47.

Wells, J. C., & Wells, J. C. (1982). *Accents of English* (Vol. 1). Cambridge University Press.

Wolff, H. E. (1987). Consonant-tone interference and current theories on verbal aspect systems in Chadic languages. In *Proceedings of the 4th International Hamito-Semitic Congress. Amsterdam & Philadelphia*, 475-496.

Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, *33*(2), 165-181.

Yip, M. (2002). *Tone*. Cambridge University Press.

Yu, A. C. L. (2010). Tonal effects on perceived vowel duration. *Laboratory Phonology*, *10*(4), 151-168.

Yu, K. M. (2011). *The learnability of tones from the speech signal*. Doctoral dissertation, UCLA.

Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, *136*(3), 1320-1333.