Title
Cue-shifting between acoustic cues: Evidence for directional asymmetry

Authors
Yang, Meng
Sundara, Megha

Publication Date
2019-07-01

DOI
10.1016/j.wocn.2019.04.002

CUE-SHIFTING BETWEEN ACOUSTIC CUES: EVIDENCE FOR DIRECTIONAL ASYMMETRY

Meng Yang

Megha Sundara

University of California Los Angeles

Running title: *Cue-shifting between acoustic cues*

Abstract

Previous research shows that experience with co-varying cues is neither sufficient nor necessary for listeners to integrate them perceptually. Auditory Enhancement theorists explain this by positing that listeners integrate two cues more readily if the cues enhance each other's percept. To isolate the role of enhancement from that of experience, we forced English adult listeners to shift attention between two enhancing cues that they do not use phonemically, pitch and breathiness, by reversing the informativeness of the two cues in a cue weighting experiment. Listeners were able to shift attention from pitch to breathiness and vice versa if the two cues were in an enhancing relation. When this relationship was reversed, listeners could shift attention from pitch to breathiness but not in the opposite direction. Clearly, both the change in informativeness and the enhancing properties of the cues influenced the listeners' re-weighting of these cues. However, the directional asymmetry was not predicted. Moreover, the same asymmetry was observed in two new groups of listeners who have native language experience with either pitch or breathiness. We discuss the consequences of such asymmetric enhancement effects, rising from either processing limitations or articulatory contingencies, for language change.

I. INTRODUCTION

Speech sound categories are multidimensional and often signaled by two or more acoustic cues. For example, consonant voicing co-varies with both voice onset time (VOT) and initial fundamental frequency (f0) on the following segment (e.g. Abramson & Lisker, 1985), English vowel tenseness is signaled by both vowel duration and formant height (e.g. Hillenbrand et al., 2000), and obstruent place information is realized as differences in both formant transitions and energy distribution in the release burst (e.g. Francis et al., 2000).

Not surprisingly, speech perception research shows that listeners also attend to multiple cues when distinguishing speech sound categories. For example, it has been demonstrated that listeners attend to cues other than voice onset time (VOT) when making a voicing distinction. Studies have shown that listeners are more likely to identify CV syllables with ambiguous onset voicing as having a voiced onset if the f0 on the following vowel is low, but are more likely to identify the same syllables as having a voiceless onset if the f0 on the following vowel is high (e.g. Chistovich, 1969; Fujimura, 1971; Gruenenfelder & Pisoni, 1980; Castleman & Diehl, 1996).

Not only do listeners attend to multiple cues when distinguishing speech sound categories, they are also sensitive to the co-variation between them. Listeners' categorization boundary along one cue dimension can be shifted by changing the value along another cue dimension (e.g. Chang, 2013). In fact, it has been argued that experience with cue co-variation can affect perception to such an extent that sometimes listeners are unable to respond to individual cues separately (e.g. Brunelle, 2012; Lee & Katz, 2016).

A. In two experiments we sought to determine whether some cue pairs are privileged because they enhance each other's percept (e.g. Kingston & Diehl, 1994). Specifically, we wanted

to test whether the effect of enhancement is independent from that of language experience, as has been claimed previously. In the following sections, we lay out what it means for two cues to be enhancing (Section IA), how experience with multidimensional input shapes our perception (Section IB), in order to motivate the design of our two experiments (Section IC). In Section II, we describe our first experiment which tests the effects of enhancement on listeners who have no language experience with either of the cues selected. Section III describes a follow-up experiment that addresses a perceptual asymmetry found in Experiment I. Finally, we discuss the findings from the two experiments and their implications in Section IV. Privileged cue pairs

Some co-varying cues have been argued to be special. Proponents of the Auditory Enhancement account claim that some cue pairs are privileged because they have a mutually enhancing auditory effect (e.g. Diehl & Kluender, 1989; Kingston & Diehl, 1994; Diehl et al., 1995; Diehl & Molis, 1995). For example, Kingston and Diehl (1994) propose that voiced obstruents are characterized by a band of low-frequency energy during the closure, and the percept of this low-frequency energy is enhanced by low f0 on either side of the obstruent. In a language like English where phonologically voiced obstruents do not canonically have vocal fold vibration during the closure, it has been observed that the f0 at the onset of the following vowel is also low (Kingston & Diehl, 1994), thus still contributing to the percept that there is low-frequency energy in the stop region. Kingston and colleagues have argued that voicing co-varies with low f0 because the two cues jointly contribute to the percept of low frequency energy (Diehl & Molis, 1995; Diehl et al., 1995; Kingston & Diehl, 1994; Kingston, 2011).

Cue pairs such as these are thus *enhancing*, a specialized term Kingston and colleagues define to refer to cues 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. *Enhancement* as defined by Kingston et al. is distinct from featural enhancement (e.g. Stevens & Keyser, 1989) or gestural enhancement (e.g. Stevens & Keyser, 2010) where either a secondary feature or a secondary gesture, respectively, are coupled with a primary distinctive feature to increasing the acoustic distance between two sounds. In the rest of the paper, we will use the term *enhancing* in the Kingstonian sense – for cue pairs that converge on a single IPP – and we will use the term *integral* to describe the fact that these cues cannot be perceived independently.

There is some evidence that experience with the enhancing cue pair is neither *sufficient* nor *necessary* for listeners to integrate them. Kingston et al. (2008) use the Garner paradigm (Garner, 1974) to determine which of the following cues – F1, f0, closure duration, and voicing into a stop – are integral. In English, voiced stops have shorter closure durations and are followed by vowels with lower f0 and F1 at the onset, while voiceless stops have longer closure durations and are followed by vowels with higher f0 and F1 at the onset. So English listeners have experience with the co-variation between all these cue pairs. Kingston et al. asked English listeners to categorize stimuli drawn from four quadrants of a two-dimensional acoustic space for every cue pair. In the Garner Paradigm, discriminability between pairs of stimuli from different quadrants is one measure of perceptual distance between them. Equal perceptual distance between all four quadrants indicates that listeners perceive the two cues that delimit the space independently. In contrast, a greater perceptual distance between positively co-varying stimuli (top-right and bottom-left quadrants) or negatively co-varying stimuli (top-left and bottom-right quadrants) indicates that listeners do not perceive the cues independently. Their results showed that F1 and

f0 are perceived independent of each other, but neither is perceived independent of the continuation of voicing into a stop. Further, F1, f0 and voicing into a stop are also perceived independently from closure duration. These findings indicate that listeners do not treat all co-varying cue pairs similarly: listeners only integrated 2 pairs of cues – F1 and continuation of voicing into a stop, and f0 and continuation of voicing into a stop. Note these are the only two cue pairs that Kingston et al. claim are enhancing. 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.

There is also some evidence that listeners may integrate cues even without experience with their co-variation. The contrast between the Korean lenis coronal fricative [s] and its fortis counterpart [s*] in medial position is signaled by the presence or absence of voicing during frication respectively, but never by f0 differences on the following vowel (Cho et al., 2002; Chang, 2013). However, Chang (2013) shows that raising f0 on the following vowel increases listeners' likelihood of identifying ambiguous tokens as the fortis [s*]. Lee and Katz (2016) further demonstrate that listeners are more sensitive to the [s*] ~ [s] contrast when the relation between voicing and f0 is in the enhancing direction, that is, when lower f0 is paired with voicing during frication and higher f0 is paired with no voicing during frication. Together, these results show that even in the absence of experience with co-variation, listeners still perceptually integrate enhancing cues like f0 and voicing. However, as Lee and Katz (2016) themselves point out, 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 stops.

Overall, the role of experience in determining which cue pairs are integral is unclear. In Kingston et al,'s experiments with English listeners one could argue that the extent to which cue pairs co-vary predicts whether those cue pairs are integrated. In Lee and Katz, listeners experience with stops could explain why they integrated f0 and VOT in fricatives. We therefore designed two experiments to isolate the role of enhancement from that of language experience. We used a cue-weighting paradigm to do so.

B. Cue weighting of multidimensional stimuli and its relation to the input

Given that the co-variation between cue pairs differs across languages, it must be learned to some extent. Research shows that given multidimensional stimuli, listeners rely on some cues more than others, a phenomenon referred to as *cue weighting* (e.g. Holt & Lotto, 2006; Mayo et al., 2011). 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 & 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). The primary

cue then is the most distinctive acoustic correlate, while secondary cues are simply less distinctive based on their input distributions.

Secondary cues can play a more crucial role in categorization when the primary cue is obscured. For example, Liu and Samuel (2004) showed that Mandarin listeners are able to categorize tones using duration and phonation cues when pitch information is removed from portions of the stimuli. Similarly, Alwan and Jiang (2011) demonstrate that secondary cues to the perception of labial/alveolar distinctions (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 established cue weights are not impervious to change; listeners can re-weight cues as a result of changes in the speech signal itself.

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 as a result of exposure to synthesized (Iverson et al., 2005) and natural stimuli (e.g. Hazan et al., 2006; 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; 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).

While weight adjustments can be rapid and temporary (Idemaru & Holt, 2011; 2014; Liu & Holt, 2015; Lehet & Holt, 2016), these adjustments may also be long-lasting. When there are long-term variations in speech production, such as when a secondary cue is exaggerated over several generations of speakers, listeners must also learn these changes. For instance, Kirby (2013), under the assumption that cue dimensions are separable, simulated an ongoing sound change in Seoul Korean in which a VOT contrast in stop consonants is becoming a laryngeal contrast signalled (in part) by f0. He found that f0 eventually emerged as the most informative cue to the contrast as a result of reducing the distinctiveness of the primary cue, VOT. Results from Kirby's sound change simulation are thus consistent with an account where cue weighting and cue shifting are solely determined by distributional properties of cues in the signal (Holt et al., 2001).

In fact, Holt et al. (2001) claim that any co-variation between cues can be learned through experience. They trained Japanese quail on stimuli in which VOT and onset f0 had a positive correlation, as in most languages with a stop voicing contrast, or on stimuli in which the two cues had an "unnatural", negative correlation, or on stimuli in which the two cues were not correlated. Compared to the condition where the two cues were uncorrelated, birds trained on the naturally correlated as well as unnaturally correlated stimuli learned the co-variation they were trained on.

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. We used cue

weights as a tool to probe whether listeners draw inferences about enhancing cue pairs even when they are not supported by input distributions.

C. The present study

Recall that there is ample evidence that cue weights assigned by listeners in category learning tasks are sensitive to the co-variation between two cues in the input. The study on Japanese quail also demonstrates that any co-variation between two cues can be learned, even when the enhancing correlation between them is reversed. In contrast, research by Kingston and others (e.g. Kingston et al., 2011; Lee & Katz, 2016) shows that not all cues that co-vary are enhancing, and that listeners integrate enhancing cues even when they do not co-vary in the input.

In two experiments, we sought to isolate the role of enhancement (if any) from that of language experience. We chose pitch and breathiness as the two cues in our experiments – pitch being the percept of change in f0, and breathiness being a voice quality characterized by a larger difference in amplitude between the first and second harmonics. These cues were chosen crucially because Kingston (2011) claims this cue pair is enhancing, and thus not perceived independently. That is, breathiness and lower f0 both strengthen the percept of low frequency energy, so lower pitch is associated with more breathiness, and higher pitch is associated with less breathiness. Evidence of their enhancing relationship comes from research which shows that (a) listeners' perception of spectral slope (a correlate of voice quality) is affected by changes in pitch (Li & Pastore, 1995), (b) listeners' perception of pitch is affected by changes in spectral shape – another correlate of voice quality (Silverman, 2003; Kuang & Liberman, 2015), and (c) pitch and voice quality interfere with each other (i.e. are not fully distinguishable) at the sensory/perceptual level, as shown using a Garner paradigm (Brunelle, 2012).

We controlled for listeners' experience with these cues in two ways. First, we recruited three groups of listeners who did not have long term experience with the co-variation between pitch and breathiness as cues to a phonemic contrast in their native language. That is, the co-variation between pitch and breathiness was not linguistically relevant for any of the three groups of listeners. We then experimentally controlled their experience with the distribution of stimuli across these two cues in a cue weighting paradigm, first teaching them to weight one cue higher, then changing the 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. We then compared whether shifts in cue weights were proportional to listeners' experience in the training paradigm.

Since our listeners had no long-term contrast-relevant experience with any co-variation between these cues, the same amount of shift in cue weight is expected for all conditions given the same degree of experience. Of particular interest were conditions where shifts in cue weights were not proportional to the listeners' experience. If listeners perceive pitch and breathiness as integral even in the absence of experience with the cue pair, we expected the extent of shift in cue weights to differ based on whether the enhancing relationship between the two cues was maintained or reversed.

II. EXPERIMENT I

We used a cue-weighting paradigm to isolate the role of Auditory Enhancement from language experience. We first trained participants to categorize a set of stimuli from "Language 1", in which the distribution of stimuli along two cue dimensions favored higher weights to one of the cues. Following Nosofsky (1986), we expected participants to selectively attend to the cue

that optimized categorization. Then, participants were exposed to a set of stimuli from "Language 2" in which the distributional informativeness of the two cues was reversed, favoring the other cue. Participants were asked to categorize test stimuli after training with Language 1 as well as Language 2. Using two artificial languages, rather than using listeners' language background alone in place of Language 1, allowed us to control for the relative informativeness of the two cues in both languages.

In order to do well on the categorization task for Language 2, participants had to shift attention away from the primary cue in Language 1. While the *amount* of distributional change between Language 1 and 2 was kept constant, the relationship between category labels in Language 1 and Language 2 was designed such that the enhancing correlation between the two cues was either (a) preserved or (b) reversed (see Methods belows). We then evaluated the extent to which participants re-weighted cues.

In Experiment I, English listeners were chosen as the test group, since neither pitch nor breathiness is used to signal differences in meaning. Thus, these listeners were expected not to have an advantage with either of these cues. Additionally, they had no long-term experience with the co-variation between the two cues in signaling a single phonemic contrast.

Based on experience alone, listeners' performance on the categorization task in Language 2 should be the same regardless of the relationship between the two cues; it should only track the amount of distributional change – which is held constant by design. If enhancement is independent of experience, then listeners' performance on the categorization task in Language 2 should be facilitated when they are able to exploit the enhancing relationship between the two cues while cue-shifting, compared to when this relationship is reversed.

A. Methods

1. ***Participants***

150 undergraduate participants (age 18-31) were recruited from the Subject Pool at a North American University. 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 thus excluded.

2. ***Stimuli***

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

a. Perceptual scaling. 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). 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.

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, to match the range for breathiness. 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 linear below 500 Hz (Stevens et al., 1937).

b. Stimuli distribution. The experimental paradigm, adapted from Holt and 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). We adopt this particular distribution, given in Figure 1, since it has been demonstrated to give the listeners the best chance to learn a higher weight for the more informative cue (Holt & Lotto, 2006).

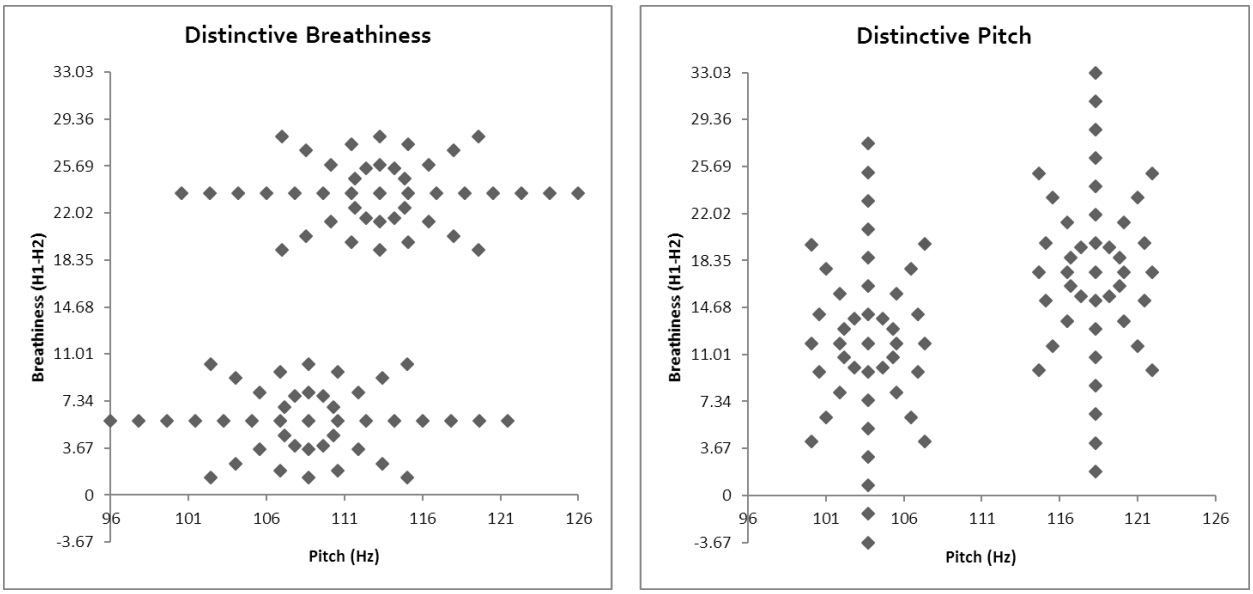

Figure 1. Training stimuli: Distinctive Breathiness (left) and Distinctive Pitch (right) distributions. Each training stimulus has a breathiness (H1-H2) value and a pitch (Hz) value, represented by a black point in the two-dimensional space.

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

The stimuli in each training set were designed to cause participants to favor one cue over the other (i.e. give a higher weight to one cue than the other). For the Distinctive Breathiness stimuli (Fig. 1, left), optimal categorization would be obtained by attending more to the breathiness cue, and for the Distinctive Pitch stimuli (Fig. 2, right), optimal categorization would be obtained by attending more 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, 67 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, 70 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.

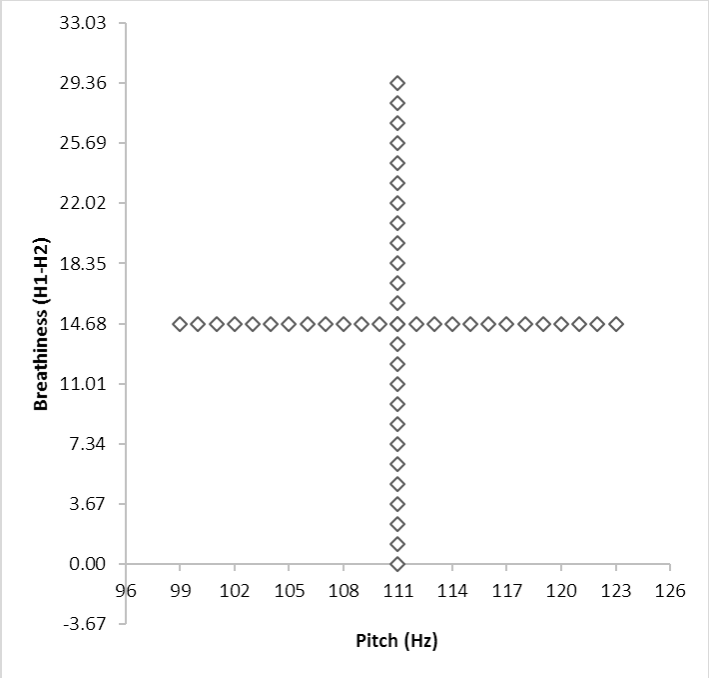

Figure 2. Test Stimuli for all conditions. Each test stimulus is represented by a point in the two-dimensional space. Vertically arranged points have the same pitch (111 Hz) but vary in breathiness. Horizontally arranged points have the same breathiness (14.68 dB) but vary in pitch.

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.

c. 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 Voice Synthesis (Antoñanzas-Barroso, Kreiman, and Gerratt, 2006). First, a natural voice sample was inverse-filtered to obtain the harmonic part of the glottal source. Inharmonic information (e.g. noise, vocal tremors, jitter and shimmer, and formant frequencies and bandwidths) were then reintroduced to approximate the original voice. 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. After the vowel was manipulated, a [t] was spliced onto each token to form the syllable [tɑ]. Figure 3 shows the resulting spectrum for two test stimuli that have the same pitch (111Hz) but are at either ends of the Breathiness continuum.

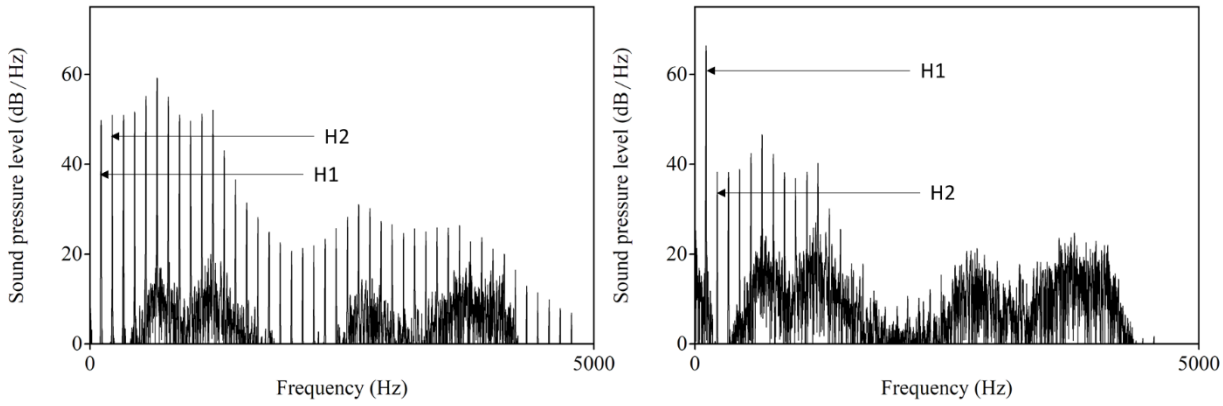

Figure 3. Examples of synthesized stimuli. Modal (left): H1-H2 = 0 dB. Breathy (right), H1-H2 = 29.36 dB.

Note that integration between cues has been reported to be specific to the cues as well as 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.

3. ***Procedure***

Presentation of the stimuli was done using the online Appsobabble platform (Tehrani, 2015). Participants listened to the stimuli on 3M Peltor HTB79A-02 headphones and responded on a QWERTY keyboard. The study was conducted in a quiet room in a university research lab. A schematic of the procedure is given in Figure 4.

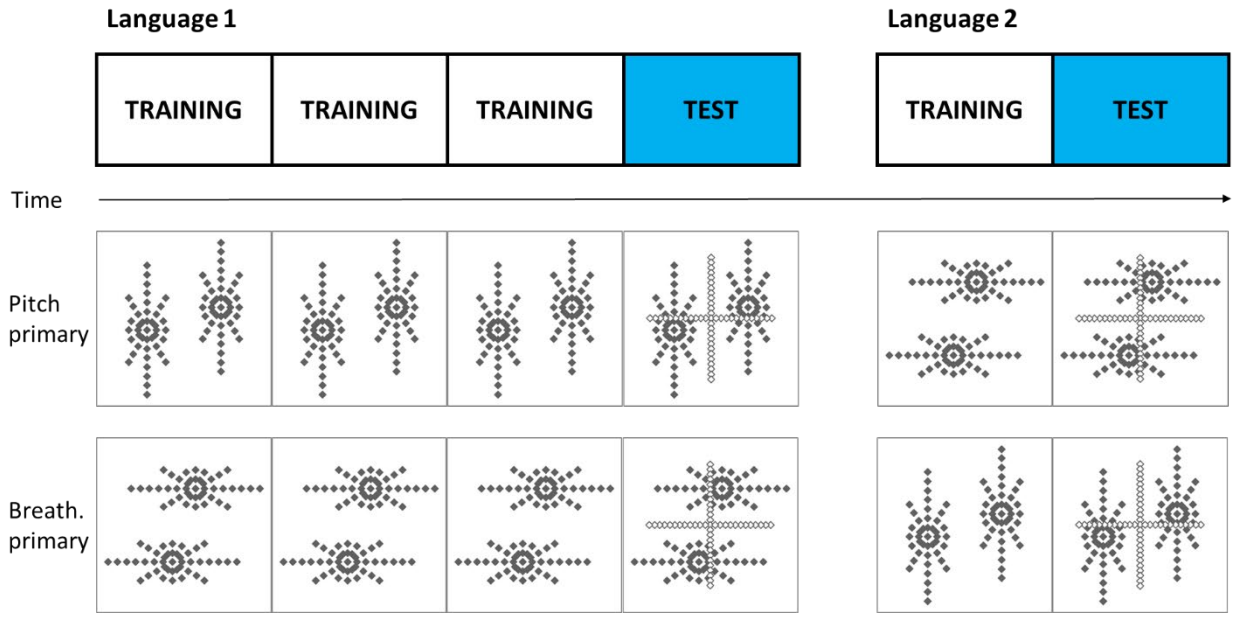

Figure 4. Design of the experiment: Participants completed training blocks and test blocks in order from left to right (L1: 3 training, 1 test; L2: 1 training, 1 test). Pitch-primary participants heard the Distinctive Pitch stimuli in L1 and Distinctive Breathiness stimuli in L2. Breathiness-primary 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).

All participants were trained on a Language 1 and a Language 2. We counterbalanced the direction of the shift (Pitch to Breathiness or Breathiness to Pitch) such that half of the participants (Pitch-primary) were trained and tested on the *Distinctive Pitch* stimulus set as their Language 1 and the *Distinctive Breathiness* set as their Language 2, while the other half

(Breathiness-primary) were trained and tested on the *Distinctive Breathiness* set as their Language 1 and the *Distinctive Pitch* set as their Language 2. In Language 1, all participants heard three blocks of training stimuli each consisting of 86 randomized trials (labeled "Training" in Fig. 4), then one test block (labeled "Test" in Fig. 4) which included 136 randomized trials consisting of both training and test stimuli. The purpose for including training stimuli in the test block was to maintain learning. In Language 2, participants heard one block of new training stimuli, then one block with the same training stimuli plus the test stimuli.

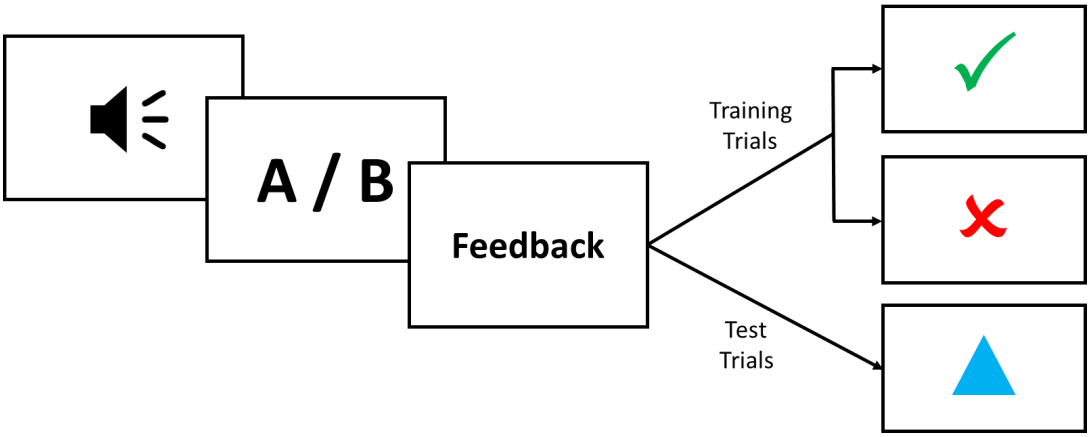

Figure 5. Procedure during each trial: Participants heard a sound, made a choice between Category A and Category B, then received feedback: Correct (green check mark), Incorrect (red ex), Unknown (blue triangle).

The sequence of events per trial is given schematically in Figure 5. On each trial, participants listened to a single stimulus token and decided whether the "foreign" word they heard was 'sea' or 'land' and pressed either the S key (Category A) or the L key (Category B) on the keyboard to indicate their choice. (These will henceforth only be referred to as Category A and Category B, or simply A and B). 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 Language 1 and Language 2, participants continued to receive informative feedback on the training trials, but feedback was an uninformative blue triangle for the novel test trials. After completing the study, participants filled out a Language Background form.

4. ***Conditions***

In addition to the direction of the shift between Language 1 and Language 2 (i.e. Pitch-primary or Breathiness-primary) being counterbalanced, the mapping of categories to category labels was crucially manipulated to test for enhancement effects. The two resulting conditions are the *Enhancing Relation* condition, in which the change in category labels from Language 1 to Language 2 respects the enhancing co-variation between Pitch and Breathiness, and the *Non-Enhancing Relation* condition, in which the labeling reverses the enhancing co-variation. This manipulation will be further described below. A schematic of the stimuli and category labels for all conditions is given in Figure 6.

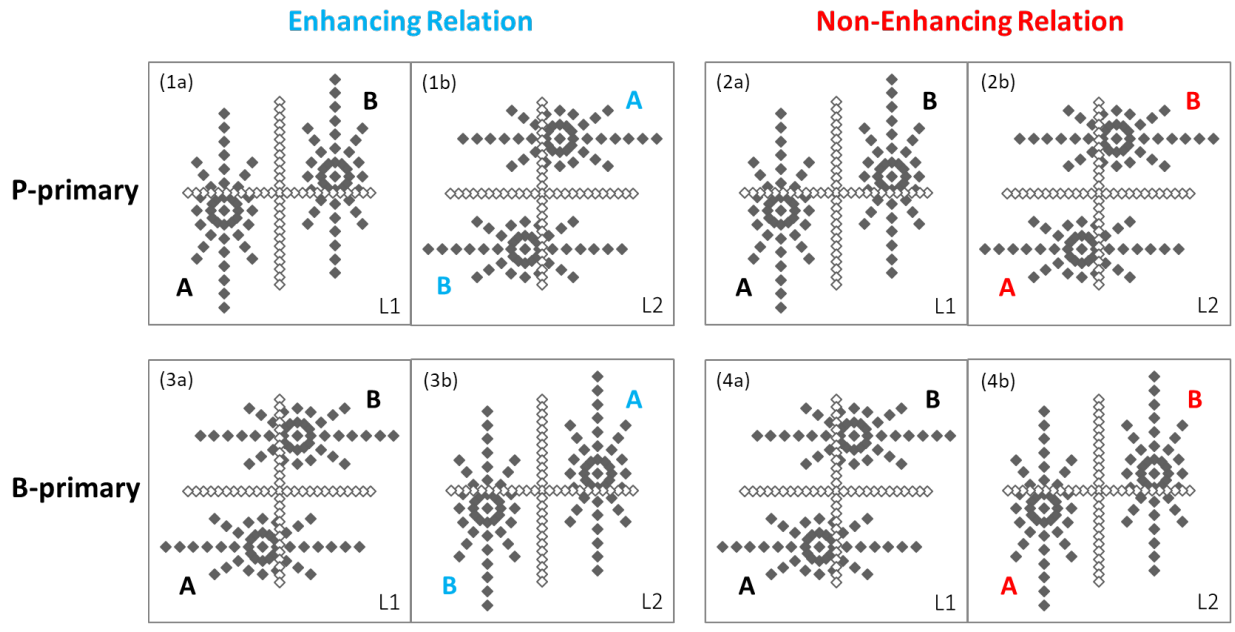

Figure 6. Experiment Conditions: Direction (Pitch-primary, upper panels vs. Breathiness-primary, lower panels) × Cue Relation (Enhancing, left panels vs. Non-Enhancing, right panels). Category labels (A or B) are labeled for each set of training stimuli in each panel.

The two stimulus sets described above, P(itch)-primary and B(reathiness)-primary, were each presented in the Enhancing (blue) and the Non-Enhancing conditions (red), giving four conditions total. In all four conditions, the category labels were the same for Language 1, the left-most stimulus set in each pair. That is, for Language 1 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, Language 1 is identical in the Enhancing and Non-Enhancing conditions, providing a built-in replication of our results.

The Enhancing and Non-Enhancing conditions differ only in the labels assigned to the distributions in Language 2. In Language 2 of the *Enhancing* conditions, following the enhancing relation between f0 and H1-H2, 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 Language 2 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 this manipulation was as follows: Participants learn to attribute more weight to the more distinctive cue in Language 1, and then are forced to transfer the weight onto a different cue in Language 2. 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. If Pitch and Breathiness are integral, the

participant will expect the category with lower pitch in Language 2 to have the same label, B, as the breathier category in Language 1, since lower pitch and breathiness enhance each other's percept. Similarly, they will expect the category with higher pitch in Language 2 to have the same label, A, as the category with less breathiness in Language 1. 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-primary – Enhancing, Breathiness-primary – Enhancing, Pitch-primary – Non-Enhancing, and Breathiness-primary – Non-Enhancing. Participants were randomly assigned to one of these four conditions when they came in for the experiment.

5. ***Analysis***

Since we were interested in the magnitude of change in cue weights between the test blocks in Language 1 and Language 2, participants were excluded if they were clearly not using the primary cue to categorize in Language 1. To this end, we excluded all participants who did not perform above chance on the training trials in the test block in Language 1 (maximally 86 trials if the participant responded to all training trials). For these trials, we performed a sign test comparing the observed number of correct responses to the hypothetical number of correct responses given a binomial choice ($p = .05$). If the probability of the observed number was less than .05, then we considered the participant's performance to be above threshold and their data was included in the analysis. For example, a participant who responded to all the training trials had to respond correctly at least 53 times out of 86 trials, that is, a minimum of 62% correct, to be considered performing above chance. 14 participants were excluded for performing below this threshold. Including the 4 participants who were excluded because of their language background

and 9 participants who did not complete the study, there were 27 exclusions. In the final analysis, there were 30 participants in the Breathiness-Primary – Enhancing group, 30 participants in the Pitch-Primary – Enhancing group, 32 participants in the Breathiness-Primary – Non-Enhancing group, and 31 participants in the Pitch-Primary – Non-Enhancing group, totalling 123 participants.

We obtained two pairs of cue weights for each participant: one weight for each cue, Pitch and Breathiness, from Language 1, and one weight for each cue from Language 2. 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), we ran a logit binomial regression 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.

Because raw weights are very noisy, following studies (e.g. Berg, 1989; Christensen & Humes, 1996; Doherty & Turner, 1996; Lutfi, 1992; including Holt & Lotto, 2006), we normalized the absolute values of the coefficients to sum to one. Note that the normalization of weights does not take into account the accuracy of listeners' categorization, but better captures the *relative* contribution of each cue for each listener. These normalized cue weights were the dependent variable in all 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 when more than 2 levels were compared. P-values from these tests are adjusted for multiple comparisons.

B. Results

The normalized cue weights for the Distinctive and Non-Distinctive cues from the test block of Language 1 and 2 are given in Figure 7, grouped by condition.

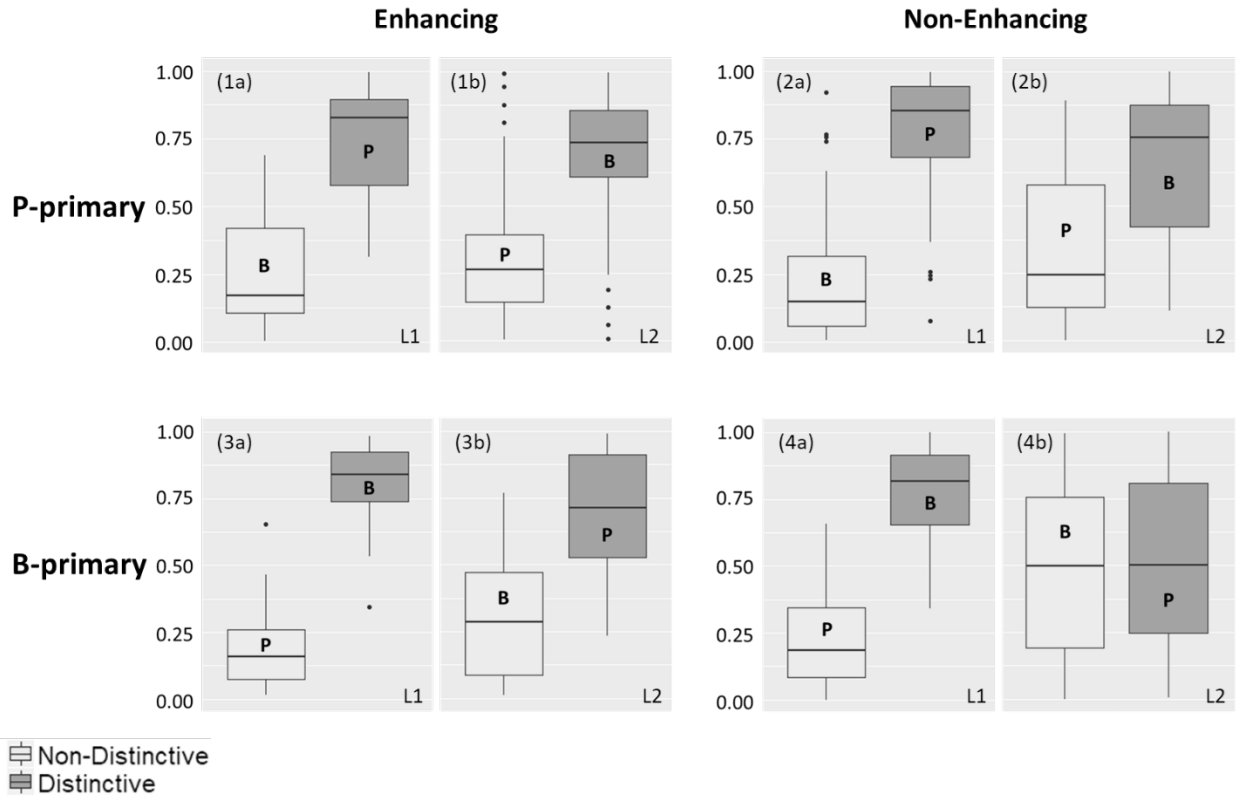

Figure 7. Normalized cue weights by Direction (Pitch-primary, upper panels vs. Breathiness-primary, lower panels) and Cue Relation (Enhancing and Non-Enhancing). There are two cue weights for each language (L1 and L2) in each condition, one for Pitch and one for Breathiness.

*1. **Language 1***

Overall, in Language 1 (Fig. 7: panels 1a, 2a, 3a, and 4a), participants learned to weight the distributionally Distinctive cue higher than the Non-Distinctive cue, as indicated by the difference in cue weights between these two cues. This was true regardless of whether the distinctive cue was Pitch or Breathiness. Recall that Language 1 was identical for the Enhancing and Non-Enhancing conditions, and so, results from these conditions were not different either.

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 (P-primary vs. B-primary) and Cue Relation (Enhancing vs. Non-Enhancing), and the within-subjects variable Distinctiveness (Distinctive vs. Non-Distinctive). This was the highest level of random effects structure that converged. We also included all 2- and 3-way interactions. The model results are given in Table 1.

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

Table 1. Lmer results from English listeners' performance on Language 1 in Experiment I

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 B-Primary – 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 (P-primary – Enhancing, $\beta = 0.49$, $p < .001$; B-primary – Enhancing, $\beta = 0.61$, $p < .001$, P-primary – Non-Enhancing, $\beta = 0.51$, $p < .001$, B-primary – Non-Enhancing, $\beta = 0.52$, $p < .001$). We additionally failed to find a significant difference between the Distinctive cue weights in different conditions and between the Non-Distinctive cues in different conditions (p-values ~ 1.0). Thus, subjects in all 4 conditions learned to weight the Distinctive cue higher than the Non-Distinctive cue in Language 1.

2. ***Language 2***

In Language 2 (Fig. 7: panels 1b, 2b, 3b, and 4b), the distinctiveness of the Pitch and Breathiness cues was switched. Breathiness was now the Distinctive cue in the Pitch-primary condition, and Pitch the new Distinctive cue in the Breathiness-primary condition. If participants successfully shifted cue weight onto the new Distinctive cue, then cue weights from the test trials should show a higher weight for Breathiness and a lower weight for Pitch in the Pitch-primary conditions (panels 1b and 2b), and the opposite weighting in the Breathiness-primary conditions (panels 3b and 4b). This was the case for every condition except the Breathiness-primary Non-Enhancing condition (panel 4b), 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 (P-primary vs. B-primary) and Cue Relation (Enhancing vs. Non-Enhancing), and the within-

subjects variable Distinctiveness (Distinctive vs. Non-Distinctive). We also included all 2- and 3-way interactions[1]. The model results are in Table 2.

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

Table 2. Lmer results from English listeners' performance on Language 2 in Experiment I

There was a significant interaction of Direction × Cue Relation × Distinctiveness. This was driven by an effect that is unique to the B-primary – Non-Enhancing 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-primary – Enhancing condition ($\beta = 0.32$, $p < .001$), the Breathiness-primary – Enhancing condition ($\beta = 0.38$, $p < .001$), and the Pitch-primary– Non-Enhancing condition ($\beta = 0.31$, $p < .001$), but not for the Breathiness-primary – Non-Enhancing condition ($\beta = 0.00$, $p = 1$).

C. Discussion

[1] The pattern of lmer results was unchanged when cue weight differences in L1 were included as a covariate to account for differences in participant success in L1. This was expected since there were no significant differences between conditions in L1 for English listeners.

To summarize, results from Language 1 showed that listeners learned cue weights equally well when either Pitch or Breathiness was the primary cue and assigned higher weights to the respective cue that had a more informative distribution. Thus, any differences after learning Language 2 could not be attributed to baseline differences in participants' learning of the primary cue in Language 1. In Language 2, the relative distinctiveness of the primary and secondary cue was switched for all conditions, and the category labels either matched listener expectations of the enhancing relation between cues (Enhancing Relation) or not (Non-Enhancing Relation). Results from Language 2 showed that participants did not shift cue weights to the same extent in the four conditions. Specifically, when shifting from Breathiness to Pitch in the absence of the enhancing relationship, the resulting weights of the Distinctive cue and the Non-Distinctive cue were not different in Language 2, whereas in all other conditions the Distinctive cue was weighted higher. Similar cue weights in Language 2 could either mean that listeners promoted the secondary cue from Language 1 but failed to demote the primary cue, or that listeners demoted the primary cue from Language 1 but failed to promote the secondary cue. In either case, it is clear that in the Non-Enhancing condition, the listeners had difficulty transferring cue weights when they first learned Breathiness as a primary cue, but were able to promote Breathiness to primary cue when they first learned Pitch.

These results show that the shifts in listeners' cue weights were not proportional to their experience in only one Non-Enhancing condition. Consistent with the Auditory Enhancement Account, listeners' shifts in cue weights were not predicted by the input distributions in the non-Enhancing condition because they **failed** to promote Pitch over Breathiness. However, they were successful at shifting weights from Pitch to Breathiness in the Non-Enhancing condition, despite the reversal of the enhancing relationship between the cues. Thus, Auditory Enhancement effects

accounted for categorization results in one but not the other Non-Enhancing condition. That is, we saw an asymmetric enhancement effect.

III. EXPERIMENT II

Recall that we chose Pitch and Breathiness because English listeners do not use these as primary cues to signal word-level meaning differences. However, one could argue that English listeners do in fact use Pitch to signal phrase-level meaning differences (i.e. intonation). English speakers also use breathiness to cue 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. As such, English listeners have more linguistic experience with Pitch than with Breathiness. Then we cannot rule out that this unequal experience with the two cues is the cause of the asymmetric enhancement effect observed in English listeners in Experiment I.

In Experiment II, we compared the categorization performance of two groups with distinct differences in their exposure to Pitch and Breathiness at the word level. These included i) a Tone language group, with listeners for whom pitch is used as the primary cue to contrast word meanings but breathiness is not, and ii) a Phonation language group, with listeners for whom breathiness is used as the primary cue to word contrasts, but pitch is not.

Since in Experiment I, the differences between the B- and P-primary conditions were observed only in the Non-Enhancing condition, in Experiment II we tested listeners from these new language groups only on the Non-Enhancing conditions. If the asymmetric enhancement effect was caused by listeners' language experience, we expected the Tone language group to

perform like English listeners, given their extensive experience with pitch, whereas the Phonation language group was expected to have the opposite directional asymmetry.

A. Methods

1. Participants

44 participants (age 18-39) were recruited at a North American university 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[2]. The same exclusion criteria were used in Experiment II as in the earlier experiment. From the Tone group, 5 subjects were excluded for being non-fluent speakers of the tone language they cited on the Language Background Questionnaire, 7 subjects were excluded for not completing the study, and one subject was excluded for performing below threshold. Of the 31 remaining participants, 16 were in the Pitch-primary condition and 15 were in the Breathiness-primary conditions.

32 participants (age 18-27) were recruited at two north American universities 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 technically difficulties occurred during the experiment, and 7 more were excluded for performing below threshold. Of the remaining 24 participants, 12 were in the Pitch-primary condition and 12 were in the Breathiness-primary condition.

[2] 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).

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

2. *Stimuli and procedure*

The stimuli and procedure are the same as Experiment I.

3. *Conditions*

Participants in Experiment II were only assigned to the Non-Enhancing condition. Half of the participants in each language group were in the Pitch-primary condition and the other half were in the Breathiness-primary condition.

4. *Analysis*

Cue weights were computed in the same way as Experiment I. These data were analyzed using mixed-effects regression models, and significant interactions were probed using Tukey's HSD post-hoc tests.

B. Results

Figure 8 shows the normalized cue weights in all conditions, separated by Group.

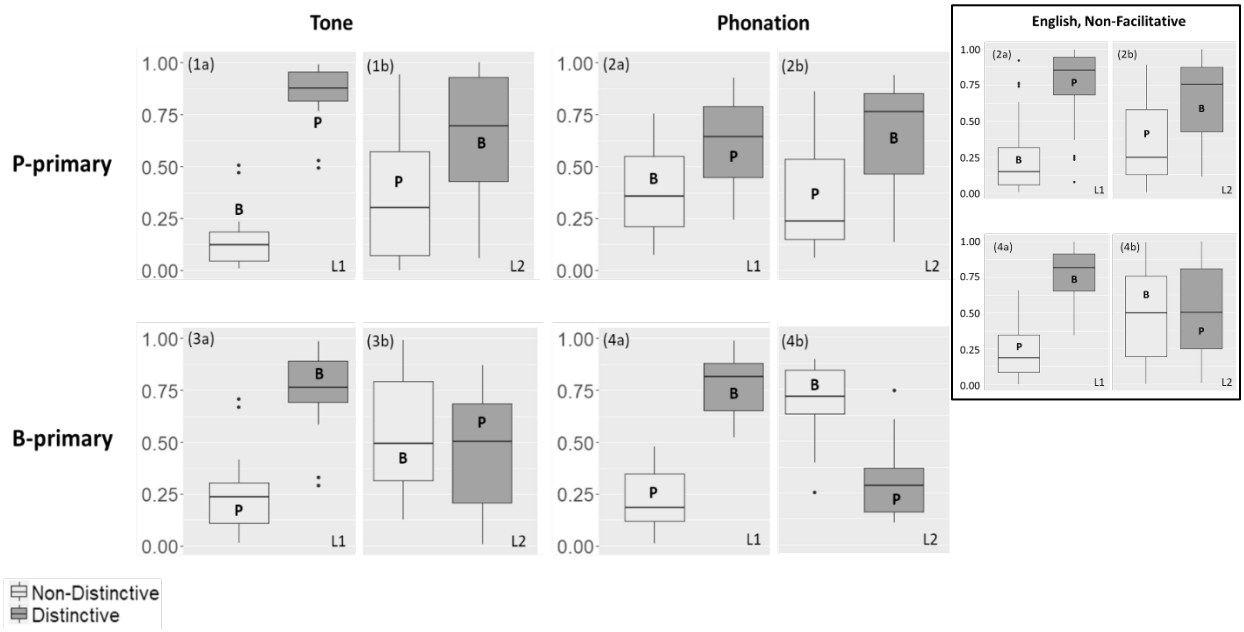

Figure 8. Normalized cue weights by Direction (Pitch-Primary, upper panels vs. Breathiness-Primary, lower panels) and Language Group (Tone group vs. Phonation group). All participants were tested on the Non-Enhancing Condition. There are two cue weights for each language (L1 and L2) in each condition, one for Pitch and one for Breathiness. Cue weights from the English, Non-Facilitative conditions analyzed in Experiment I are provided in the inset for reference.

1. ***Language 1***

We ran a mixed effects model on Language 1 (Fig. 8: panels 1a, 2a, 3a, and 4a), which included the random intercept of Subject (the highest random effects structure to converge), as well as the between-subjects fixed effects of Direction (P-primary vs. B-primary) and Language Group (Tone language vs. Phonation language), and the within-subject effect of Distinctiveness (Distinctive vs. Non-Distinctive). All 2- and 3-way interactions were included. The results are in Table 3.

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-primary*	0.17	.07	2.22	.027	*
Direction × Group = *P-primary & Tone*	-0.27	.10	-2.76	.006	**
Direction × Distinct. = *P-primary & Distinct.*	-0.33	.11	-3.14	.002	**
Group × Distinct. = *Tone & Distinct.*	-0.09	.10	-0.89	.375	
Direction × Distinct. × Group = *P-primary & Distinct. & Tone*	0.55	.14	3.90	<.001	***

Table 3. Lmer results from Tone and Phonation group listeners' performance on Language 1 in Experiment II

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 co-variate in the mixed effects model for Language 2 (see below), to control for initial differences in the learning of L1.

2. ***Language 2***

In Language 2 as well (Fig. 8: panels 1b, 2b, 3b, and 4b), successful learning of the new distribution is indicated by a higher cue weight for the Distinctive cue and a lower cue weight for the Non-Distinctive cue. This was observed in the two language groups when Pitch was primary (panels 1b and 2b), but not when Breathiness was primary (panels 3b and 4b). 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 (P-primary vs. B-primary), 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.

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 =	-0.14	.11	-1.32	.188	

Tone					
Direction = *P-primary*	-0.32	.12	-2.67	.008	**
Direction × Group = *P-primary & Tone*	0.17	.16	1.10	.292	
Direction × Distinct. = *P-primary & Distinct.*	0.63	.16	3.87	<.001	***
Group × Distinct. = *Tone & Distinct.*	0.29	.15	1.86	.062	
Direction × Distinct. × Group = *P-primary & Distinct. & Tone*	-0.34	.22	-1.54	.123	

Table 4. Lmer results from Tone and Phonation group listeners' performance on Language 2 in Experiment II

Results from the tone and phonation groups were similar to results from the English listeners. There was a significant interaction between Direction and Distinctiveness, $\beta = 0.63$, $p < .001$. A pairwise Tukey's HSD test showed that in the P-primary conditions, the Distinctive cue was weighted higher than the Non-Distinctive cue ($\beta = 0.26$, $p = .004$), but that in the B-primary conditions, the Distinctive cue was weighted *lower* than the Non-Distinctive cue ($\beta = 0.20$, $p = .041$), 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, $\beta = -0.34$, $p = 0.123$. Since Language Group was also not significant as a main effect, and its interaction with Distinctiveness is only marginal, there is no statistical evidence that participants from the two groups behaved differently in Language 2.[3]

[3] We re-ran the model without the single outlier in the Phonation group in the B-primary condition (two points in Figure 8, panel 4b), to confirm out results, Crucially, in this model, the interaction between Direction and Distinctiveness was still significant, $\beta = 0.71$, $p < .001$, confirming a 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 B-primary condition.

C. Discussion

Experiment 2 replicated the results from Experiment 1. Like English 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 P-primary condition. Then, like English participants, participants in the two groups also failed to shift cue weights from Breathiness to Pitch, when the enhancing relationship was reversed. Thus, the inability to promote Pitch to a primary cue in the absence of an enhancing relationship with Breathiness was independent of participants' language experience. This is consistent with an Auditory Enhancement account.

What is problematic for the Auditory enhancement account is that, again, listeners' cue shifting in the non-Enhancing P-primary condition was commensurate with the evidence in the input. In other words, the lack of enhancing relationship between cues did not affect listeners in the P-primary condition.

We note that although not significantly different, Gujarati listeners nevertheless seemed to behave differently from the English and Tone groups when shifting cue weights from breathiness to pitch with the enhancing relationship reversed. English 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.

IV. GENERAL DISCUSSION

In this study, we set out to determine the extent to which auditory enhancement effects are independent of language experience. Two categorization studies were conducted in which listeners learned to weight Pitch and Breathiness by training on auditory stimuli that distributionally biased them towards using one of these cues, then training on a different set of stimuli that distributionally biased them towards using the other cue. The intention was for listeners to learn one cue as primary and the other cue as secondary in Language 1, then test the extent to which they reweighted the two cues under various conditions in Language 2. We did this so as to completely control the extent and nature of the experience that listeners had with the co-variation between the two cues.

In Experiment I, we asked specifically whether English listeners shift cue weights commensurate with the input distribution when the status of enhancing relationship between the cues was manipulated. Additionally, the direction of cue-shifting was counterbalanced such that half of the listeners were shifting from Pitch to Breathiness and the other half were shifting from Breathiness to Pitch. If listeners are sensitive to the privileged status of enhancing cues, then learning of the cue weights in L2 was expected to be different in the enhancing compared to the non-enhancing conditions.

English listeners were able to promote either breathiness or pitch to a primary cue when the category relations were enhancing. When the category relations were non-enhancing, English listeners successfully promoted breathiness, just as well as in the enhancing condition, but failed in promoting pitch above breathiness as the primary cue, although the distributional evidence was the same in both conditions. Therefore, the *Auditory Enhancement Theory* was able to account for the difference in performance between the enhancing and non-enhancing conditions when breathiness was the primary cue in Language 1. However, English listeners succeeded at

promoting breathiness when pitch was the primary cue in Language 1, even in the non-enhancing condition. Thus, we observed an asymmetry in the enhancement effect based on whether listeners were exposed to breathiness or pitch first.

We designed Experiment II to rule out if more linguistic experience with either pitch or breathiness could explain the asymmetry. For this, we tested two additional groups of listeners, speakers of lexical tone languages and speakers of Gujarati, a language that uses breathiness contrastively. These two groups showed the same pattern of results; in the non-enhancing condition, listeners in both groups failed to promote pitch, but not breathiness.

Given that listeners in the Tone language group had more experience with pitch than breathiness, like English listeners, they were expected to pattern in the same way. Our results confirmed this. However, Gujarati listeners who rely on breathiness to distinguish a native contrast, also showed the same asymmetry. The identical pattern of asymmetry in the three groups 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.

Alternatively, 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 English or Gujarati listeners.

Listeners in all three groups successfully shifted cue weights from pitch to breathiness but failed to shift weights from breathiness to pitch in the Non-Enhancing condition. We take this as evidence that (a) there is an asymmetry in the enhancement effect and (b) it is independent of language experience. Because asymmetric enhancement is observed cross-linguistically, it is likely to be reflected in typology and/or diachrony.

We see that this is indeed the case in diachronic contrast transfer, whereby a phonemic contrast signaled by one cue becomes signaled by another cue over time. For example, there are many cases of tonogenesis in which a voicing contrast became a tone contrast. Languages that underwent this process at some point in their development include Vietnamese (Thurgood, 2002), Western Kammu (Kingston, 2011), Yabem (Kingston, 2011), Eastern Cham (Phu et al. 1992), Chinese (Hombert, 1978), Karen (Hombert, 1978), Tamang (Mazaudon & Michaud, 2008), and Hottentot languages in South Africa (Beach, 1938). In contrast, there are no known languages in which a tone contrast has become a voicing contrast, and only one language for which transfer of a tone contrast onto a phonation contrast has been claimed (Uchihara, 2016). It has been

proposed that instances of transfer from a voicing contrast of initial consonants onto a pitch contrast (i.e. tone) may have undergone an intermediate stage where the voicing contrast was first realized as a phonation difference on the vowel, which then became a pitch contrast (Thurgood, 2002; anticipated by e.g. Haudricourt, 1965; Egerod, 1970; Pulleyblank, 1978; cf. Coetzee et al., 2018). However, it is unclear whether this middle stage occurred for all languages. In all of these languages, the transfer of the contrast from one cue to the other respects the enhancing relationship between the cues.

Contemporary synchronic sound systems also show effects consistent with asymmetric enhancement effects. There are tone languages in which pitch is a primary cue with breathiness additionally distinguishing between tone categories of similar pitch (e.g. Kuang, 2013 on Black Miao; Garellek, et al. 2013 on White Hmong; Brunelle, 2009 on Northern Vietnamese). However, comparatively uncommon are phonation languages in which breathiness is a primary cue with pitch additionally distinguishing similar phonation categories (Silverman, 1997; 2003; e.g. Mazaudon & Michaud, 2008 on Tamang; Edmondson et al., 2001 on Yi and Bai).

It is worth noting that the typological asymmetries discussed above may have a basis in articulation since both pitch and voice quality are controlled at the larynx. If an articulatory gesture producing breathiness necessarily lowers pitch, but pitch can be altered using articulatory gestures that do not always change voice quality, this could result in asymmetric enhancement effects.

Asymmetric enhancement can also be rooted in perception. In their experiments on the relative cue weighting of two non-speech dimensions - central frequency (CF) and modular frequency (MF), Holt and Lotto (2006) also found an asymmetry. Even when the cues were perceptually equated and the distributional informativeness of CF was equal to or less than that

of MF, listeners still attributed higher weights to CF than to MF. It was only when they increased the within category variance of CF and reduced the within category variance of MF that the weights were reversed. Holt and Lotto postulate that this asymmetry could be caused by listeners having a default higher weighting for CF as a result of either an innate predisposition or experience with CF being a more reliable cue.

We think it is unlikely that the asymmetry in our study could have occurred because listeners simply have a default higher weighting for breathiness as opposed to pitch. First, English listeners had no overall preference for either cue in Language 1; they learned the pitch distribution just as well as the breathiness distribution. Second, English listeners did not show any preference for breathiness in Language 2 in the enhancing condition either.

Alternately, asymmetric enhancement effects could 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. That is, 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. 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.

This kind of perceptual dependency can be modeled as crosstalk between two processing channels (Pomerantz et al., 1989; Melara & Marks, 1990), in this case between a channel for pitch and a channel for breathiness. Interference effects are observed if information from one

channel flows to the other channel and inhibits some level of processing in the second channel. Importantly, this model allows for crosstalk to be unidirectional. In such a model, there would be a path for information from the pitch channel to interfere with the breathiness channel, without interference in the reverse direction. Thus, when listeners first attend to the pitch channel, then switch to the breathiness channel, information from the pitch channel disrupts listeners' ability to use breathiness in categorization. However, when listeners first attend to the breathiness channel, then switch to pitch, listeners can attend to the latter channel independently. While such a model can account for an asymmetry, it cannot account for the enhancement effect itself since it does not address the particular correlation the cues must have for there to be an interference effect. To tease apart the articulatory account and a perceptual one, we would need to test listeners on a cue pair that is enhancing but could not be produced by the same gestural mechanism. If enhancement effects are still observed, then we can be more confident that they are rooted in perception.

V. CONCLUSION

In this study we used a cue weighting paradigm to isolate the effect of enhancement from that of language experience. We tested English listeners as well as native speakers of tone and phonation languages on their ability to learn categories based on these two acoustic dimensions. We found that, when listeners were unable to use the enhancing relationship between the two cues, they failed to promote pitch (but not breathiness) from a secondary to a primary cue. Future research is needed to confirm whether the asymmetric enhancement observed has articulatory or perceptual roots.

Acknowledgments

This project was funded by the UCLA Department of Linguistics Student Research Support Committee. We would like to express our gratitude to Patricia Keating and Jody Kreiman for their comments and discussion which have both challenged and improved this work. We would also like to thank Norma Antoñanzas-Barroso and Henry Tehrani for their assistance in stimuli synthesis and experiment set-up, and our undergraduate research assistants who made data collection possible. Last but not least, thank you to the audiences of the UCLA Phonetics Seminar, Marc Garellek, and Jianjing Kuang for their feedback and support of this project.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press, Edinburgh.

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.) *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25-33). New York: Academic Press.

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

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

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/>.

Beach, D. M. (1938). *The phonetics of the Hottentot language*. W. Heffer & Sons, ltd.

Berg, B. G. (1989). Analysis of weights in multiple observation tasks. *The Journal of the Acoustical Society of America*, *86*(5), 1743-1746.

Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. *Speech Communication Group Working Papers, Research Laboratory of Electronics* (pp. 73-93). Boston: MIT.

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.

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.

Castleman, W. A., & Diehl, R. L. (1996). Effects of fundamental frequency on medial and final [voice] judgments. *Journal of Phonetics*, *24*(4); 383-398.

Chang, C. (2013). The production and perception of coronal fricatives in Seoul Korean. *Korean Linguistics 15*(1), 7-49.

Chistovich, L. A. (1969). Variations of the fundamental voice pitch as a discriminatory cue for consonants. *Soviet Physics-Acoustics*, *14*(8), 372-378.

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

Christensen, L. A., & Humes, L. E. (1996). Identification of multidimensional complex sounds having parallel dimension structure. *The Journal of the Acoustical Society of America*, *99*(4), 2307-2315.

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.

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.

Doherty, K. A., & Turner, C. W. (1996). Use of a correlational method to estimate a listener's weighting function for speech. *The Journal of the Acoustical Society of America*, *100*(6), 3769-3773.

Edmondson, J.A., Ziwo, L., Esling, J.H., Harris, J.G. and Li Shaoni. (2001). The aryepiglottic folds and voice quality in the Yi and Bai languages: Laryngoscopic case studies. *Mon-Khmer Studies 31*, 83–100.

Egerod, S. (1971). Phonation types in Chinese and South East Asian languages. *Acta Linguistica Hafniensia 13* (2), 159–171.

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

Fairbanks, G. (1960). *Voice and articulation drillbook*. Harper and Row, New York.

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

Francis, A. L. Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception Psychophysics 62*, 1668-1680.

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

Fujimura, O. (1971). Remarks on stop consonants: Synthesis experiments and acoustic cues. *Form and substance: phonetic and linguistic papers presented to Eli Fischer-Jørgensen. Odense: Akademisk Forlag*, 221-232.

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., 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.

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.

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

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

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

Gruenenfelder, T. M., & Pisoni, D. B. (1980). Fundamental frequency as a cue to postvocalic consonantal voicing: Some data from speech perception and production. *Perception & psychophysics*, *28*(6), 514-520.

Haudricourt, A.-G. (1965). Les mutations consonantiques des occlusives initiales en mon-khmer. *Bulletin de la Société de Linguistique de Paris 60* (1), 160–172.

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.

Hillenbrand, J. M., Clark, M. J., & Houde, R. R. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America* 108, 3013-3022.

Holt, L. L., & Lotto, A. J., (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *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.

Hombert, J. M. (1978). Consonant types, vowel quality, and tone. *Tone: A linguistic survey*, 77, 112.

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.

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.

Keating, P., Kreiman, J., & Alwan. A. (2018) *The UCLA Speaker Variability Database*.Ttalk at the 19eme Colloque d'avril sur l'anglais oral de Villetaneuse, Paris, France.

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

Kingston, J., and Diehl, R. L. (1994). Phonetic knowledge. *Language*, 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., 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. (2013). The role of probabilistic enhancement in phonologization. In A. Yu (Ed.) *Origins of Sound Change: Approaches to Phonologization*, Oxford, UK: Oxford University Press. pp. 228-246.

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.

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 a. *The Journal of the Acoustical Society of America*, 128(4), 2085-2089.

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

Kuang, J., & Liberman, M. (2015). Influence of spectral cues on the perception of pitch height. *Proceeding of ICPH*, 18.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press, Cambridge.

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.

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.

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

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.

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.

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. (1992). Informational processing of complex sound. III: Interference. *The Journal of the Acoustical Society of America*, *91*(6), 3391-3401.

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.

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

Melara, R. D., & Marks, L. E. (1990). Dimensional interactions in language processing: Investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 539.

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.

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.

Nosofsky, R. M. (1986). Attention, similarity, and the identification--categorization relationship. *Journal of Experimental Psychology. General*, *115*(1), 39.

Phu, V. H., Edmondson, J., & Gregerson, K. (1997). Eastern Cham as a tone language. *MATS*, *20*, 31-43.

Pomerantz, J. R., Pristach, E. A., & Carson, C. E. (1989). Attention and object perception. In *Object Perception: Structure and Process*, Eds. B. E. Shepp and S. Ballesteros, Laurence Erlbaum Associates, Hillsdale, 53-90.

Pulleyblank, E.G. (1978). The nature of Middle Chinese tones and their development. *Journal of Chinese Linguistics 6* (2), 173–203.

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>.

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*, 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.

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

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

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

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.

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