An asymmetric perceptual dependency between pitch and breathiness

Meng Yang

University of California, Los Angeles
mengyang@ucla.edu

ABSTRACT

A recent, cross-linguistic study shows that, under certain experimental conditions, listeners have trouble shifting perceptual attention from breathiness to pitch but not from pitch to breathiness. The hypothesis put forth to explain this asymmetry is that listeners are associating differences in breathiness with differences in pitch, but they are not associating differences in pitch with differences in breathiness.

The current study tests this hypothesis using a cue weighting task. English listeners categorized auditory stimuli in which there was strong evidence that one of the two cues, pitch or breathiness, is informative, and no evidence that the other cue is informative. If the hypothesis is correct, we would expect listeners learning to use breathiness to also attend to pitch, but we would not expect listeners learning to use pitch to also attend to breathiness. Results show a marginal effect in this direction, providing weak support for the hypothesis.

Keywords: perception, cue weighting, asymmetry.

1. INTRODUCTION

It has been shown that pitch, the percept of f0, and breathiness, a voice quality characterized acoustically by a larger difference between the amplitude of the first harmonic (H1) and the amplitude of the second harmonic (H2), have effects of perceptual interference (e.g. [3], [16]) in the sense of Garner [6].

Two dimensions may interfere because they are perceived holistically [7]. Based on this idea, proponents of auditory enhancement theory claim that some pairs of dimensions form *intermediate perceptual properties* (IPPs) that mediate between raw acoustics and phonological contrasts [13]. Since breathiness is characterized by a much stronger H1, the lowest harmonic, than higher harmonics, this voice quality, together with low pitch, contributes to the percept of low-frequency energy [12]. Because pitch and breathiness are perceived holistically, interference between these two dimensions should always be symmetric.

Alternatively, interference effects can stem from crosstalk between different processing channels [17]. Unlike auditory enhancement, this theory allows for asymmetric interference, which can result from unidirectional crosstalk between two dimensions that are interpreted at different processing levels (see [17] for example). However, given two dimensions that are interpreted at the same level, crosstalk between them is also expected to be bidirectional.

Results from a recent study pose a challenge to the symmetric interference predicted by both of these theories for pitch and breathiness. Listeners were trained to categorize auditory stimuli using either pitch or breathiness, then they were forced to shift their attention to the other of these two cues. In the incongruent condition, where the natural relation between pitch and breathiness is reversed, listeners who were trained on pitch could shift their attention to breathiness, but listeners who were trained on breathiness could not shift their attention to pitch [23]. This was true for English listeners, who do not have experience using these cues for any phonemic contrast, Chinese and Gujarati listeners, who use one of these cues for a phonemic contrast, and Hani listeners, who have a phonemic contrast that uses both of these cues.

These results show that interference between pitch and breathiness is asymmetric. If pitch and breathiness are perceived as a single IPP, then attentional shift between the two should be equally difficult in either direction when the cues are incongruently correlated. The crosstalk account correctly predicts asymmetric interference for Chinese and Gujarati listeners, where one of the cues is phonemic whereas the other is phonetic, and thus, in different channels. However, because pitch and breathiness are both phonetic or phonemic for English and Hani listeners respectively, any interference between the two dimensions is predicted to be bidirectional. However, listeners in all 4 groups had asymmetric interference. Thus, these results are not consistent with either account.

Yang and Sundara [23] offer a third account for the asymmetric interference between pitch and breathiness. They propose that listeners associate differences in breathiness with changes in pitch, but they do not associate differences in pitch with changes in breathiness. Support for the perceptual association of breathiness to pitch comes from studies showing that listeners can estimate the pitch range of talkers based on spectral tilt information (e.g. [10], [11]) If indeed the dependency between the two cues holds, then the results from Yang and Sundara [23] could be explained as follows: Listeners shifting their attention from breathiness to pitch experience interference, while listeners shifting from pitch to breathiness can treat the latter as a novel cue.

The current study aims to test the asymmetric dependency hypothesis more directly. Here, in a cue weighting task similar to the one in Yang and Sundara [23], English listeners categorized auditory stimuli in which there was strong evidence that either pitch or breathiness alone is informative for the contrast, and no evidence that the other of these cues is informative. They were then tested on critical stimuli that change only in pitch but had a constant breathiness, and on stimuli that change only in breathiness but had a constant pitch. Listeners were expected to attend mostly to the dimension they were trained on. However, if listeners associate one cue to another, then we would also expect them to attend somewhat to the other cue, despite there being no evidence for it in the input. Thus, the asymmetric dependency hypothesis would predict specifically that listeners trained on breathiness would pay more attention to the uninformative pitch cue than listeners trained on pitch to the uninformative breathiness cue. The following section describes the details of the study design.

2. METHOD

2.1. Participants

66 undergraduate students participated in the study for course credit. All were native speakers of English. 12 of these participants were excluded for speaking an additional language fluently or natively. One participant was excluded for not completing the study. Participants did not have any known hearing impediments.

2.2. Stimuli

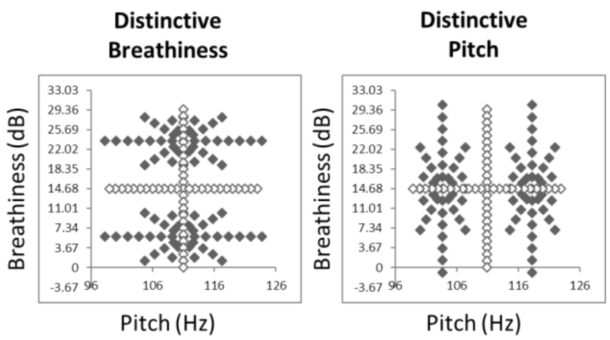
2.1.1. Stimuli in the current study

The distributions of the auditory stimuli used in this study were based on the stimuli used in the Yang and Sundara [23] study in which the asymmetry was found. Whereas their stimuli had one very informative, distinctive dimension and one weakly informative, non-distinctive dimension, the stimuli in the current study removes any evidence for categories along the non-distinctive dimension by neutralizing the category difference for that cue. Thus, if listeners assign a non-zero weight to the non-distinctive cue, then it could not have come from what they have learned from the input. The distributions of these stimuli are shown in Figure 1.

The acoustic parameter used to manipulate breathiness in this study was H1-H2, e.g. [4] and [8], whose values ranged from -3.67 to 33.03 dB. This range, 36.7 dB, is ten times the just-noticeable difference (JND) for English listeners, that is, 3.67 dB [15]. Though this is a larger range than what is typically measured in speech (e.g. [5]), two trained phoneticians judged the endpoints to sound natural given the auditory impressions of the stimuli. The purpose of extending the range was so that the breathiness scale matched the pitch scale in number of JNDs.

Pitch was manipulated by changing fundamental frequency (f0) measured in Hertz (Hz). The pitch dimension ranged from 96 to 126 Hz. The 30 Hz range was also set at ten times the JND for English listeners, approximately 3 Hz [14], to perceptually match the breathiness scale. The maximum and minimum points of the scale are within the normal range for the male human voice.

Figure 1: Stimuli distribution for the current study.



In Figure 1, the 86 black dots in each distribution represent training stimuli. These dots form two clusters, which correspond to two categories of speech sounds. The categories are distinct along one dimension (i.e. have differentiated means, and small within-category variance) and are not distinct along the other dimension (i.e. have the same mean, but also have large within-category variance).

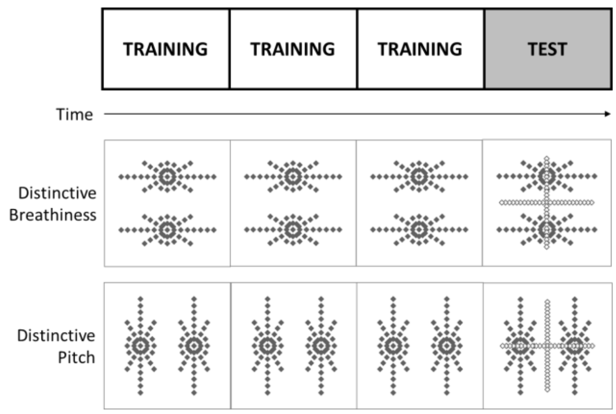
The white dots in Figure 1 represent the critical stimuli. These 50 stimuli are identical across the two stimuli distributions, Distinctive Breathiness and Distinctive Pitch. 25 of these stimuli are held at a constant pitch of 111 Hz which is ambiguous between the two categories along this dimension, but they vary along the breathiness dimension from 0 to 29.36 dB by increments of 1/3 JND (~1.22 dB). The other 25 test stimuli are held at a constant H1-H2 of 14.68 dB which is ambiguous between the two categories along this dimension, but they vary along the pitch dimension from 99 to 123 Hz by increments of 1/3 JND (1 Hz).

The 222 unique stimulus tokens – 86 training tokens for the Distinctive Breathiness training set, 86 training tokens for the Distinctive Pitch training set, and 50 critical tokens – were synthesized using Voice Synthesis [1]. Using an inverse-filtered male voice sample of the [ɑ] vowel as the base, pitch and breathiness were manipulated in that order. Pitch was controlled by changing the f0 parameter. Breathiness was changed by increasing or decreasing the amplitude of the first harmonic. After each vowel was synthesized with specific pitch and breathiness measures, a [t] was spliced onto each token in Praat [2] to form the syllable [tɑ].

2.3. Procedure

In this study, there's no shifting of attention from one cue dimension to another. Half of the participants were trained and tested on the Distinctive Breathiness stimulus set, and the other half were trained and tested on the Distinctive Pitch stimulus set. Each participant heard three training blocks followed by one test block, as shown in Figure 2.

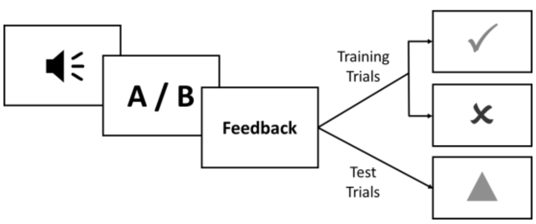
Figure 2: Study procedure for each participant.



In each of the three training blocks, participants heard the 86 training trials in randomized order. They then heard the test block which had the 86 unique training trials as well as the 50 critical trials.

For each trial, participants heard an auditory stimulus through headphones, made a binary choice on the keyboard, and were given visual feedback, as shown in Figure 3.

Figure 3: Study procedure for each participant.



2.4. Analysis

An additional 9 participants were excluded from analysis for performing below chance on the training trials in the test block, 3 from the Distinctive Breathiness group and 6 from the Distinctive Pitch group. A total of 44 participants were included in the final analysis, 22 in each group.

Cue weights for breathiness and pitch were obtained for each participant by running a logit binomial regression on the category choice (A or B) of critical trials. The equation is given in (1).

(1) logit(*choice*) = int. + ω*breathiness + ω*pitch,
Where ω is cue weight.

The pitch and breathiness values of critical stimuli were the model predictors. In the model output then, the coefficients of pitch and breathiness are a measure of how well changes in each dimension predict the listener's category choice. A higher coefficient indicates that increments in that dimension increase the log odds of choosing a category (e.g. A). Thus, the model coefficients were taken as a proxy for how much attention listeners paid to each cue.

Consistent with our a prior predictions, the cue weights on the non-distinctive cues in each condition were compared using a one-tailed independent samples t-test. Statistical comparisons were run in R [16] using the built-in t.test function.

3. RESULTS

The breathiness cue in the Distinctive Breathiness condition and the pitch cue in the Distinctive Pitch condition were compared first to ensure equal learning of the distinctive cues between the two groups. Their cue weights are given in Figure 4.

Figure 4: Distinctive cue weights. Breathiness (left) from the Distinctive Breathiness condition and pitch (right) from the Distinctive Pitch condition.

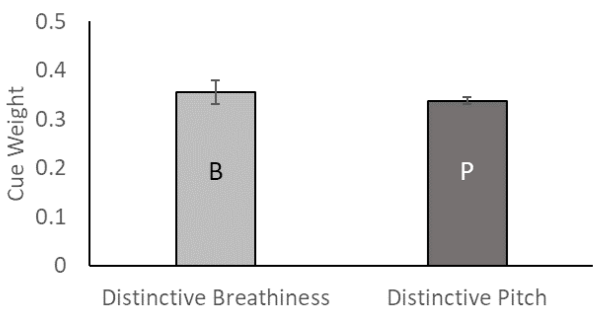
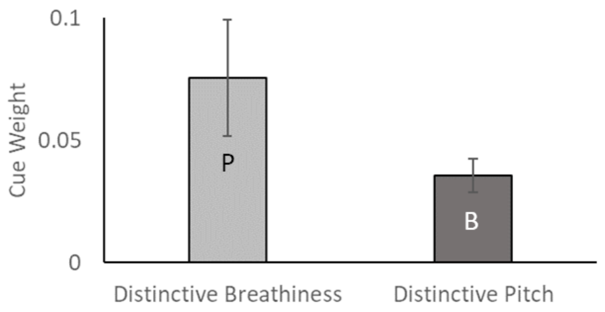


Figure 4 shows that listeners weighted the distinctive cues the same, regardless of whether they were learning breathiness or pitch, $t(41) = .262$, $p = .795$.

Having established the equal learning between the two groups, we now turn to the non-distinctive cues, pitch in the Distinctive Breathiness condition and breathiness in the Distinctive Pitch condition. Their cue weights are shown in Figure 5.

Figure 5: Non-distinctive cue weights. Pitch (left) from the Distinctive Breathiness condition and breathiness (right) from the Distinctive Pitch condition.



The weights of the non-distinctive cues are clearly lower than those of the distinctive cues, as is to be expected. However, the prediction was for pitch in the Distinctive Breathiness condition to be weighted higher than breathiness in the Distinctive Pitch condition. The two sets of non-distinctive cue weights were thus subjected to a one-tailed independent groups t-test where equal variance was not assumed,

which shows that the cue weight of pitch is marginally higher than the cue weight, $t(42) = 1.61$, $p = .057$.

4. DISCUSSION AND CONCLUSION

In this study, I tested a hypothesis that breathiness and pitch are asymmetrically dependent; listeners associate differences in breathiness with higher or lower pitch, but they do not associate changes in pitch to different degrees of breathiness. Listeners in the study were trained to categorize sounds in which one of these cues was informative to the contrast, but the other cue was uninformative, giving them no reason to use the other cue at all. The cue weights of listeners learning to use breathiness for categorization were compared to the cue weights of listeners learning to use pitch. In particular, I was interested in whether the uninformative cues were differentially weighted.

A strong version of the enhancement theory, in which pitch and breathiness are perceived as a single IPP, would predict that the use of one dimension for categorization should generalize fully onto the other dimension. This was clearly not the case as here, the cue weights for the non-distinctive dimensions were lower. The crosstalk theory predicts that equally little weight would be assigned to both non-distinctive cues. This was also not supported. If the asymmetric dependency hypothesis is correct, then we would expect listeners' categorization using the breathiness cue to generalize onto the pitch cue more than the reverse. This was marginally supported, lending some support to the asymmetric dependency hypothesis.

The asymmetry between pitch and breathiness was robust in the Yang and Sundara [23] experiments, while it was weak in the current experiment. This difference could be due to task effects; the asymmetry may be brought out when listeners must shift their attention from one cue to another. A natural place where this might be observed is in diachronic sound change involving contrast transfer, where a phonemic contrast cued by one acoustic dimension gradually becomes cued by another acoustic dimension. Given the directionality of the asymmetric dependency, we should expect there to be more contrast transfer from breathiness, or voice quality, onto pitch than the reverse. This seems to hold true typologically if we consider a change in voice quality to be an intermediate step in the process of a consonant voicing contrast becoming a tone contrast (see e.g. [21]). There have been more instances of transfer from consonant voicing and voice quality to a tonal contrast (e.g. Vietnamese [21], Chinese [9], Eastern Cham [18]), than from tone to a voice contrast, Quiaviní Zapotec being the only language for which such a claim has been made [22].

In sum, this study found weak evidence that pitch and breathiness are asymmetrically dependent, corroborating findings from previous studies as well as the typology of diachronic sound change involving these two acoustic dimensions.

7. REFERENCES

[1] Antoñanzas-Barroso, N., Kreiman, J., Gerratt, B. R. 2006. Voice Synthesis. [computer software]

[2] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved from http://www.praat.org/

[3] 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.

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

[5] Garellek, M., Samlan, R., Gerratt, B. R., Kreiman, J. 2016. Modeling the voice source in terms of spectral slopes. *J. Acoust. Soc. Am. 139*(3), 1404-1410.

[6] Garner, W. R. 1974. *The Processing of Information and Structure*. Psychology Press.

[7] Garner, W. R., Felfoldy, G. L. 1970. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*(3), 225-241.
[8] Gordon, M., Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383-406.
[9] Hombert, J. M. 1978. Consonant types, vowel quality, and tone. *Tone: A linguistic survey*, 77, 112.
[10] Honorof, D. N., Whalen, D. H. 2005. Perception of pitch location within a speaker's F0 range. *J. Acoust. Soc. Am*. *117*(4), 2193-2200.
[11] Kuang, J., Liberman, M. 2015. Influence of spectral cues on the perception of pitch height. In *ICPhS* Glaslow.
[12] Kingston, J. 2011. Tonogenesis. *Companion to Phonology*. Malden, MA: Wiley-Blackwell, 2304-2333.
[13] Kingston, J., Diehl, R. L. 1994. Phonetic knowledge. *Language*, 419-454.
[14] Kollmeier, B., Brand, T., Meyer, B. 2008. Perception of speech and sound. In *Springer Handbook of Speech Processing*, 61-82. Springer Berlin Heidelberg.
[15] Kreiman, J., Gerratt, B. R. 2010. Perceptual sensitivity to first harmonic amplitude in the voice source a. *J. Acoust. Soc. Am*. 128(4), 2085-2089.
[16] Li, X., Pastore, R. E. 1995. Perceptual constancy of a global spectral property: Spectral slope discrimination. *J. Acoust. Soc. Am*. *98*(4), 1956-1968.
[17] 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.
[18] Van Han, P., Edmondson, J., Gregerson, K. 1997. Eastern Cham as a tone language. *Mon-Khmer Studies*, *20*, 31-43.
[19] 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.
[20] Tehrani, H. 2015. Appsobabble: Online Applications Program.
[21] Thurgood, G. 2002. Vietnamese and tonogensis. *Diachronica* 19, 333-363.
[22] Uchihara, H. 2016. Tone and registrogenesis in Quiaviní Zapotec. *Diachronica*, 33(2), 220-254.
[23] Yang, M. & Sundara, M. Cue-shifting between acoustic cues: evidence for directional asymmetry. Under revision.