The Learning and Generalization of Contrasts Consistent or Inconsistent with Native Biases

Kyuwon Moon, Meghan Sumner

Department of Linguistics, Stanford University, Stanford, CA, USA

kyuwon@stanford.edu, sumner@stanford.edu

Abstract

This study investigates the process of generalizing a learned sub-lexical contrast across speakers of different non-native accents of English. We examine the generalization of a novel cue (voicing-cued release) that is non-contrastive in English, but contrastive in the manipulated speech of our L2 speakers of English, to a speaker with the same or different L1 as our training speakers (Exp. 1). We then examine performance when the learned contrastive cue is paired with native contrastive vowel duration (Exp. 2), and when the native contrast is present, but the learned contrast is not contrastive in the speech of the new speaker (Exp. 3). We find that learned cues are dominant enough to additively improve word recognition when paired with congruent native cues, although the performance is inhibited when native and learned cues conflict. These data illuminate the intricate balance between native contrasts and recently learned information, where learned information overrides, but is still affected by, native contrasts.

Index Terms: phonetic variation, accent perception, sub-lexical contrasts, learning and generalization

1. Introduction

The perception and recognition of L2 speech depends greatly on factors such as proficiency of the L2 speaker [1], number of talkers on which a listener was trained [2], and intelligibility of the speaker [3]. Grossly summarizing this literature, as intelligibility, proficiency, and variability in training decrease, difficulty increases (see also [4, 5]). While it is true that comprehension depends greatly on these factors, there are numerous other factors that may illuminate the ways in which speakers understand non-native accents, which is especially striking given our strong native phonetic biases [6, 7, 8].

These studies have informed our understanding of the role of variability in speech perception. One question that remains, though, is whether, once a new contrast is internalized, a listener uses this information across the board, applying it to all non-native speakers, or whether a speaker with a new L1 (as cued by indexical variation) prompts a reset to native biases. Additionally, whether generalization depends on new-speaker usage patterns – where a learned contrast may either be consistent or inconsistence with the production patterns of different novel speakers, may also shed light on how listeners use learned contrasts and the conditions that promote contrast maintenance.

In this study, we examine the generalization of novel contrasts at the sub-lexical level *and* the effect of a new speaker who uses either the trained consonantal contrast (contrastive voicing dependent on the presence/absence of release of final stops), or a native English vowel duration phonetic contrast [9]. Across three experiments, we investigate whether listeners who learn a novel contrast apply the learned mapping to novel speakers with different accents who either use the learned cue contrastively or non-contrastively, or whether indexical cues to a different accent prompt a reset to native perceptual biases.

2. Experiment 1

In Exp. 1, we trained English listeners to treat final release as contrastive (eg. bet/bed - [bɛt]/[bɛt˺]). After being trained with this new contrastive cue, we expected listeners to generalize the learned cue to a new speaker of the same language background (see [1, 2, 10, 11]). Following within-accent generalization, we include a second generalization test with a novel speaker with a different L2 accent. If a novel contrast is internalized and influences perception independent of speaker attributes, the accent of the second speaker should not hinder generalization. If, though, contrasts are accent-dependent, generalization should not persist across speakers with novel accents.

1.1. Methods

Participants. Fifty-four college students participated in this study for pay. All were monolingual speakers of American English. None reported any hearing problems. The participants were split into three groups of eighteen, corresponding to three experimental conditions (Group 1, 2, and Control; See *Design*).

Stimuli. The stimuli were recorded by four Korean speakers (one for pre/post-tests (K1); three for training (K2, K3, K4)) and one Arabic speaker (for tests of cross-accent generalization (A1)) in a sound attenuated booth. All five speakers were female with obvious accents, but fluent English speech. The speakers were chosen based on independent identification of words in isolation, ensuring that they produce reliable vowel differences before voiced and voiceless final stops, and also that potentially confusable sounds (e.g. [ɛ] and [æ]) were identified correctly. Additionally, they were all rated high on a 7-point scale of accentedness (5.9). These criteria enabled us to create a contrast that is believable as uttered by a non-native speaker, but also to avoid issues of comprehension that may muddy our results. The Arabic speaker was also female, and her speech met the same criteria as the Korean speakers. All speakers were naïve with respect to the purpose of the experiment.

The stimuli included 40 final stop minimal pairs (eg. bed/bet, cup/cup, wick/wig). Three productions of 80 pairs were recorded, from which 40 pairs were chosen to best balance word frequency within a voiced/voiceless pair, duration, and intensity. The voiceless tokens of each minimal pair corresponded to voiceless-final words in training (e.g. BET). From this voiceless token, we created an unreleased version (eg. [bɛt˺] from [bɛt]) by removing the release burst part of the voiceless stop. This set corresponded to voiced-final words in training (eg. BED). This manipulation isolated the contrast to release only, as it was the only difference in the voiceless-voiced pair. In addition to the critical items, 80 filler

items were recorded. The fillers included 40 coda-contrast items that varied by place (eg. dam/dan), 24 onset-contrast items (eg. coal/goal), and 16 vowel-contrast items (eg. ball/bell). For comparison, we have included the conditions and contrasts used across experiments in Table 1.

Design. A Minimal Pair Decision (MPD) task was used for all four phases of the experiment: Pre-test, Training, Post-test, and Generalization (Gen) test. In all phases, orthographic minimal pairs were presented on a monitor, and screen placement (left/right) was randomized on every trial.

The Pre-test included half of the critical items (20 pairs; 40 items) from the training, plus filler pairs (40 pairs; 80 items). The participants were tested only once with these 120 items. A single speaker presented the words (K1).

The Training phase included all the critical pairs (40 pairs; 80 items), along with immediate feedback. Stimuli from three speakers (K2, K3, K4) were used and repeated once, resulting in 480 minimal pair decisions made. While no blocks were obvious to the listener, the first set of 240 items was randomized and then presented, followed by the randomized second set. Fillers were not included in the training.

The Post-test was identical to the Pre-test. It was either presented immediately after training, or following the Gen test phase. The reason for counterbalancing the order was to ensure that any null effect potentially associated with cross-speaker generalization was not due to recency. The Gen test phase included the same items and design as the Post-test, but the stimuli used were those produced by the native Arabic speaker (A1). Feedback was not given during the Pre-test, Post-test, and Gen test phases.

Two groups of participants participated in the Pre-test and Training phase: 18 followed training by Post-test – Gen test (Group 1); 18 by Gen test – Post-test (Group 2). An additional group of 18 participants were used as a control for the A1 speaker, on par with the Pre-test by the K1 speaker, to establish a baseline for potential speaker-dependent responses.

Procedure. Participants were tested in a sound-attenuated booth individually. All stimuli were presented over headphones at a comfortable listening level that participants can adjust at the beginning of the experiment. Participants were presented with a sound (eg. [bɛt]) and were asked to make a decision from two lexical items presented on the screen (eg. bed/bet). Participants were instructed to respond as quickly and accurately as possible. If participants did not respond within three seconds, a new pair was presented. A new trial began one second after the responses had been made. The experiment lasted about 40-50 minutes per subject.

Table 1. Summary for Contrasts (and examples) used in Exps. 1-3

Phase	Pre-test	Training	Post-test	Gen test
Speaker	K 1	K 2,3,4	K 1	A 1
Exp. 1 Contrast	[Vt-Vt’] ([bet-bet’])	[Vt-Vt’] ([bet-bet’])	[Vt-Vt’] ([bet-bet’])	[Vt-Vt’] ([bet-bet’])
Exp. 2 Contrast	[Vt-Vt’] ([bet-bet’])	[Vt-Vt’] ([bet-bet’])	[Vt-Vt’] ([bet-bet’])	[Vt-V:t’] ([bet-be:t’])
Exp. 3 Contrast	[Vt-Vt’] ([bet-bet’])	[Vt-Vt’] ([bet-bet’])	[Vt-Vt’] ([bet-bet’])	[Vt’-V:t’] ([bet’-be:t’])

Results. We are generally interested in the extent to which listeners have learned to overcome the lack of vowel contrast and to rely instead on release as a contrastive cue to voicing. Therefore, we measured the proportion of *voiced* responses to *unreleased* stimuli (responding BED for [bɛt’]). Trials with responses faster than 400ms and longer than 1800ms were excluded from all the analyses. As a result, 2.17% of the responses were excluded. Mean proportion voiced responses are provided in Fig. 1. A Generalized Linear Mixed Model with phase (Pre-test, Post-test, Training, and Gen test) as a fixed effect and subject and sound token as random effects was performed on proportion voiced responses to unreleased tokens, to test the effect of training. Across conditions, responses increased from Pre- to Post-test, suggesting that the intended contrast was learned ($p < .0001$). The effect was also occurred between Pre-test and Gen test ($p < .0001$).

Effects of order of presentation in the generalization phase was examined using a Generalized Linear Mixed Model with the phase (Pre-test, Post-test, Training, and Gen test), order (Post-test first or Gen test first), and the interaction of the phase and order as fixed effects, and subject and sound token as random effects. The interaction between phase and order was not significant.

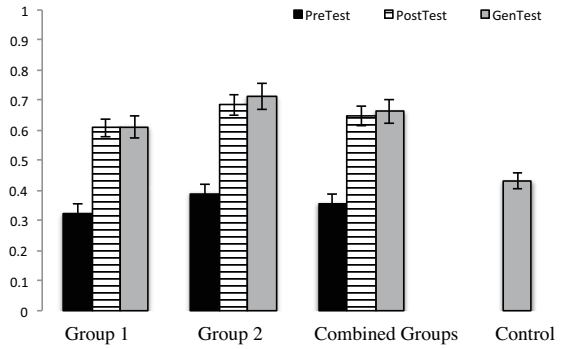

Figure 1. Proportion *voiced* responses to *unreleased* tokens across conditions. Group 1 received Post-test – Gen test; Group 2 received Gen test – Post-test.

To verify the main effect of training for the Gen test, planned comparisons compared the responses to the A1 control to those made in the Gen test phase. Performance was significantly higher ($p < .01$) than that of the control group. This shows that trained listeners were far better in the Gen test than untrained listeners.

In sum, the results of this first experiment support three findings. First, listeners generalize learned cues to novel speakers who share an L1 as the training talkers once a controlled, novel contrastive cue has been learned. This finding is supported by the significant shift from labeling words with unreleased stops as voiced from Pre-test to Post-test. After being trained with three different L1 Korean speakers, listeners, as expected, were able to generalize the release/non-release contrastive cue to a novel Korean speaker.

Second, listeners generalize a learned contrast across speakers independent of L1. The increase in *voiced* responses to unreleased stops in the generalization phase compared both to the K1 Pre-test and the A1 control support this interpretation of the data. Even though the speaker has a different non-native accent from the speakers in training, generalization occurs. At the very least, this suggests that learning has occurred at the sub-lexical level, as a phonetic contrast, and not at the word level, as the wide array of work highlighting specificity effects at the lexical level would lead us to expect some differences were the lexicon to be driving learning here [12, 13].

Finally, there is no obvious benefit associated with immediately following training. This is supported by the fact that no significant interaction was found between phase and post-training order. As an added control, though, we continue to counterbalance order in the remaining two experiments.

3. Experiment 2: Generalization with a native cue

In Exp. 1, we found that listeners generalize learned patterns robustly across speakers from different L1 backgrounds. In that experiment, both speakers produced the learned contrast, so one might say that the learned contrast was useful, in a sense, when confronting the novel speaker. In this experiment, we examine generalization from training on speakers K2, K3, and K4 to A1, but reinstate the native English vowel duration cue (see Table 1). The question changes from examining whether listeners generalize sub-lexical contrasts across speakers with different L1s to examining whether the native-biased conditions aid or hinder effects of a learned contrast. Specifically, we investigate whether listeners generalize a learned contrast across non-native speakers as a larger group, independent of the information provided by individual speaker patterns, and whether newly learned information has an additive effect when encountering a new speaker.

3.1. Methods

Participants. Fifty-four new participants participated in this experiment for pay or for course credit. They were again split into three groups of eighteen, corresponding to the three experimental conditions as in Exp. 1.

Stimuli. The stimuli for the Pre-test, Training, and Post-test were identical to Exp. 1. For the Gen test, we created long vowels from the unreleased tokens that were generated by removing the burst parts of voiceless tokens, as explained in Exp. 1. Vowel duration is a strong cue used by listeners to perceive word-final voicing in English [9, 14]. We used Praat's PSOLA [15] to extend the duration of each vowel on a token-by-token basis, so that the duration of the vowel was two times than that in the unreleased counterpart tokens. As an added control, editing was also done with the voiceless (short) counterparts: the vowel length was again increased and then shortened to the original duration for released tokens.

Design. The task conditions for the Exp. 2 are the same as Exp. 1. The subjects of the Main Group take the Pre-, Training, Post- and Gen tests, and the order between the post- and Gen tests was counterbalanced. The Gen test, which was given after or before the Post-test (depending on the order), had the same items as the Pre- and Post-tests, differed in the contrast used by the speaker. As noted above, the release/non-release contrast was paired with the vowel length cue in the Gen test, and was spoken by an Arabic speaker. As the control group was presented with the A1 speech without training, and this condition has a vowel-duration contrast that is present in English, we expect the score for A1 control to be relatively high (though, some argue that the consonant-vowel ratio is the best predictor of voicing, and there is no information about consonant duration in the unreleased forms).

Procedure. The procedure was identical to that of Exp. 1.

Results. A Generalized Linear Mixed Model with phase as a fixed effect and subject and item as random effects showed significant effects of training. Differences in *voiced* responses to unreleased stimuli again increased from Pre-test to Post-test ($p < .0001$), replicating the basic pattern found in Exp. 1.

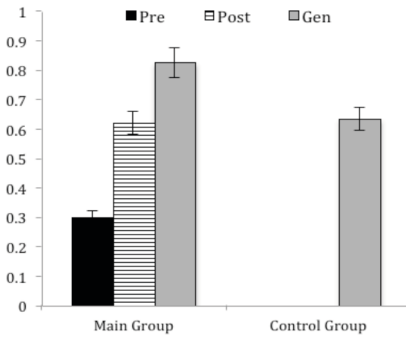

Figure 2. Proportion *voiced* responses to *unreleased* tokens across conditions, including the A1 control group.

Planned comparisons were performed between the Gen phase in Main and Control groups in Exp. 2, and also between the Gen phases of Exps. 1 and 2. Both were significant ($p < .01$, $p < .001$), showing that listeners performed better on the generalization task when they had both a native cue *and* a newly learned cue.

Taken with Exp. 1, the results of this experiment support two findings. First, effects of a newly learned cue are robust, extending across speakers even when a native cue is present. This is again not affected by order, suggesting that this is not only a matter of having been immediately trained. Second, and more interestingly, the significant difference in responding *voiced* to unreleased items between the Gen tests across experiments suggests that the existence of a native bias (vowel duration) helps listeners to distinguish the minimal pairs. Where there is vowel duration contrast paired with the learned contrast (Exp. 2), listeners identify the words with unreleased stops as voiced more easily than where there is no vowel duration contrast (Exp. 1). This significant increase in voiced responses suggests that listeners does not simply reset to their native bias, but make use of the learned contrast with novel speakers from a different L1 background than the training speakers. Since the newly learned cue is consistent with the native bias, it enhances the listeners' performance on generalization, despite being faced with indexical variation that cues not only a new speaker, but a new accent, as well.

4. Experiment 3: Native bias vs. recently learned contrast

In the previous experiments, we found that the listeners generalize across speakers, and use a newly learned cue robustly across speakers, whether or not native contrasts are present. In fact, newly learned cues appear to be additive, increasing the likelihood of a *voiced* response above and beyond the contribution of duration. So far, in the first two experiments, the novel speaker/accent always used the learned contrast. In this Experiment, we investigate the effects of a learned contrast when the contrast is not present in the speech of the A1 speaker. In this case, the training stays the same as before, and we are only interested in performance on the A1 generalization. We maintain the English voicing contrast, but reduce all final stops to unreleased. In this case, we are interested both in responses that are consistent with the English bias and training (words with unreleased stops preceded by a long vowel), but also how listeners handle words in which the native bias and learned contrast *conflict* (words with unreleased stops preceded by a short vowel).

4.1. Methods

Participants. The same number of speakers (fifty-four, mostly college students) as the Exps. 1 and 2 was used for the Exp. 3.

Stimuli. The stimuli for the Pre-test, Training, and Post-test were identical to the previous experiments. As in Exp.1, the unreleased stimuli were created from voiceless stops. Also, as in Exp. 2, vowel duration was manipulated to create the native contrast. However, unlike Exp. 2, only unreleased stimuli are used, resulting in a long-short vowel contrast, but no release contrast.

Design. The overall design remained the same as the two previous experiments. The main difference is that we analyze responses to unreleased forms, dependent on whether they had a long or a short vowel, in both the control condition (no training) and training.

In the Exp. 3, however, the release/non-release contrast was neutralized in the Gen test and the vowel length contrast distinguishes the voiceless and voiced minimal pairs.

Procedure. The procedure was identical to the previous experiments.

Results. The proportion of *voiced* responses to unreleased stops by vowel duration (long-short) was examined (means provided in Figure 3).

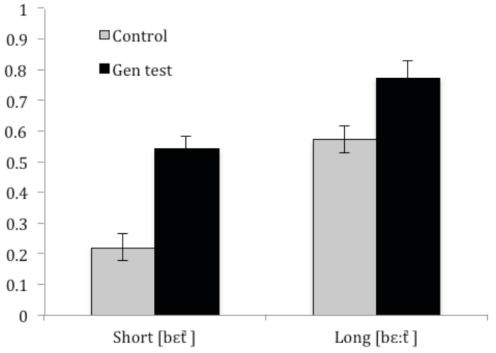

Figure 3. Proportion *voiced* responses to speaker A1 after training (left) and without training (right) by preceding vowel.

Planned comparisons were performed to examine differences between responses dependent on training for words with short vowels and long vowels ($p < .0001$, $p < .01$, respectively). Unsurprisingly, the long-vowel condition elicited more *voiced* responses after training, as this condition is consistent with training and English (a replication of Experiment 2). Interestingly, however, in the short-vowel condition, training still increases the likelihood of *voiced* responses significantly, with training than without.

These results not only show the effect of training but also show the interaction of native bias and the short-term training effect in two opposite directions. The generalization test itself contains the native bias – the vowel length – that is the most salient cue for native English speakers. With the presence of this salient native bias, listeners hear either the stimuli that are consistent with the training (voiceless – long vowel with unreleased stops) or the stimuli that are inconsistent with the training (voiced – short vowel with unreleased stops). When they are consistent, as in the long vowel condition, the result shows the significantly higher performance, which is in line with the result of the Exp. 2. However, even though they are inconsistent, as in the short vowel condition, listeners do show the effects of training. This seems to suggest that listeners prioritize the short-term training over the native bias, although they are certainly affected by the native bias, shown by the lower performance of the Gen test result in the short vowel condition than in the long vowel condition.

5. Discussion

In this study, we have investigated the learning of a sub-lexical contrast and how listeners' responses are influenced by that contrast when confronted with a new speaker with an L1 different from training. In Exp. 1, we found that listeners reliably learn a release contrast and generalize this contrast when listening to a new speaker independent of L1. In Exp. 2, we found that a learned cue is additive, increasing listener responses beyond the control. Finally, in Exp. 3, we found that listeners are sensitive to novel speaker cues, using learned cues not only when they are consistent with native cues, but also when they are in conflict with native cues.

These data raise a number of questions that reach beyond the scope of this paper. For example, one might ask whether there are processing differences associated with the perception of native and learned non-native cues. While it is certainly the case that both interact with each other, showing additive or inhibitory effects dependent on their consistency, additional research using a convergent task methodology is necessary to tease these two contrasts apart. Especially, a study on the long-term learning effects of non-native contrasting cues is needed, to complement the findings of the current study.

In conclusion, we have found that learned cues below the word-level are easily extended to speakers with different L1 background, are additive when consistent with native contrasts, and are inhibitory when they conflict with native contrasts. A deeper examination of these issues may lead us to a better understanding of contrast composition and contrast maintenance across language more generally.

6. Acknowledgements

This material is based in part upon work supported by the National Science Foundation under Grant Numbers 0720054 made to Meghan Sumner. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. We are grateful to Kara Altmann for help with data collection.

7. References

[1] Bradlow, A.R. and Bent, T., "Perceptual adaptation to non-nativespeech", Cognition, 106:707-729, 2008.

[2] Logan, J. S., Lively, S. E., and Pisoni, D. B., "Training Japanese listeners to identify English /r/ and /l/: A first report", J. Acoust. Soc. Am., 89:874-886, 1991.

[3] Flege, J.E., "Perception and production: The relevance of phonetic input to L2 phonological learning", in C. Ferguson and T. Huebner [Ed], Crosscurrents in Second Language Acquisition and Linguistic Theories, 249-290, John Benjamins, 1991.

[4] van Wijngaarden, S.J., "Intelligibility of native and non-native Dutch speech", Speech Commun., 35:103-113, 2001.

[5] Hayes-Harb, R., Smith, B.L., Bent, T., and Bradlow, A.R., "The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrast", J. Phonetics, 36:664-679, 2008.

[6] Best, C.T., McRoberts, G.W., and Goodell, E., "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system", J. Acoust. Soc. Am., 109:75-794, 2001.

[7] Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S.A., and Nishi, K., "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners", J. Acoust. Soc. Am., 109:1691-1704.
[8] Hallé, P.A., Best, C.T., and Levitt, A., "Phonetic versus phonological influences of French listeners' perception of American English approximants", J. Phonetics, 27:281-306, 1999.
[9] Port, R.F., and Dalby, J., "Consonant/vowel ratio is a cue for voicing in English", Perception & Psychophysics, 32:141-152, 1982.
[10] Bradlow, A.R., Pisoni, D.B., Yamada, R.A., and Tohkura, Y., "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production", J. Acoust. Soc. Am., 101(4):2299-2310, 1997.
[11] Weil, S.A., "Foreign-accented speech: Encoding and generalization", J. Acoust. Soc. Am., 109:2473, 2001.
[12] Nygaard, L.C., and Pisoni, D.B., "Talker-specific learning in speech perception", *Perception & Psychophysics*, 60:355-376, 1998.
[13] Nygaard, L.C., Sommers, M.S., Pisoni, D.B., "Speech perception as a talker contingent processes", *Psychological Science*, 5:42-46, 1994.
[14] Massaro, D.W., and Cohen, M.M., "Consonant/vowel ratio: An improbable cue in speech", *Perception & Psychophysics, 33*:501-505, 1983.
[15] Boersma, P., and David W., "Praat: doing phonetics by computer", http://www.praat.org, 2004.