# Patterns of generalization of perceptual learning on phonetic representations

Holger Mitterer
University of Malta
holger.mitterer@um.edu.mt

Sahyang Kim
Hongik University
sahyang@hongik.ac.kr

Taehong Cho
Hanyang University
tcho@hanyang.ac.kr

## ABSTRACT

Recent studies on perceptual learning have indicated that listeners use intermediate units between the acoustic input and lexical representations of words. The same paradigm may also reveal the nature of these intermediate units based on patterns of generalization of learning. We here test whether learning generalizes to other units of the same underlying or surface representation. This was achieved by exposing listeners to tensified Korean stops (i.e., underlying plain stops produced as tense due to a phonological process) and testing the consequences for later presented underlying tense or plain stops. Our results show that learning generalizes to underlying tense stops, while generalization to underlying plain stops could not be found. This indicates that the difference in the underlying phonological representation as tense or plain do not hinder learning as long as there is phonetic similarity on the surface.

**Keywords**: phonological units, perceptual learning

## 1. INTRODUCTION

One of the most fundamental questions in speech science is what type of units listeners use for speech perception. The challenge provided by proponents of so-called episodic models of lexical access [4] has led to research that showed that listeners make active use of pre-lexical units in generalization of learning [12]. When hearing a word that ends on /s/ but is produced with an ambiguous phonetic form between /s/ and /f/ (e.g., maʊ[s/f]), listeners not only recognize the ambiguous sound as /s/ (due to the lexical bias, since *mouse* is a word but *mouf* is not) but also learn—through multiple exposures to such ambiguous segments in unambiguous positions—that this speaker produces /s/ in an unusual way. They are then able to generalize this learning to any other word containing /s/, showing that they have learned something about pre-lexical information across words.

While generalization over words has been found repeatedly ([13, 16]), generalization to other phonological contexts has been difficult to find. For example, learning does not generalize to other allophones of the same unit (e.g, [14]), and learning about place of articulation does not generalize across different manners of articulation [15]. In fact, learning appears to be extremely specific, as learning does not generalize from one vowel context to another [15].

These results are problematic for theories that assume that speech processing is based on phonological features, often described in articulatory terms [3, 5]. These theories assume that the input is analysed as being decomposable into independent features, so that learning about one feature (e.g., place of articulation) should generalize to a new situation that involves the feature, independent of whether or not the new situation shares other features (such as manner) with the situation in which learning has taken place.

The aforementioned studies on perceptual learning, however, indicate that generalization does not easily occur when there are differences in the surface (phonetic) representation between the segment with which learning has taken place and the new segment to which learning may be generalizable. In the present study, we ask the opposite question: whether generalization is also constrained by differences in the underlying representation—i.e., whether learning can be generalizable to a phonetically same, but underlyingly different segment. Alongside addressing this question, the present study will also allow us to consider another question regarding whether learning occurs on a featural base. If learning occurs on a featural base, it may be generalized to other segments as long as they share the same feature of the segment based on which learning has taken place.

Our experiments exploit a feature of [tense] in Korean. Korean distinguishes three "laryngeal settings" in stops, so that a stop can be plain (lax or lenis, /p,t,k/), tense (fortis, /p*,t*,k*/), or aspirated (/$p^h$,$t^h$,$k^h$/) [1]. Crucially, a word-initial plain stop is realized as tense if the preceding word ends on an obstruent, a phonological process known post-obstruent tensification [9] as illustrated in (1).

(1) /tʃuŋku**k**/+/**p**atʃi/ → [tʃuŋkuk**p***atʃi]
'Chinese' 'pants'

While tensification may not be phonetically complete when the trigging segment is across a phrase boundary, no acoustic-phonetic difference between an underlying tense stop and a phonologically tensified stop has been observed in a phrase-medial environment—i.e., when the triggering segment occurs across a (phrase-internal) prosodic word

boundary [7]. That is, it is unlikely that listeners are able to reliably indicate whether a given stop is tensified by the tensing rule or it is underlyingly tense (in the absence of lexical information) especially when tensification takes place phrase-internally, the precise environment that is considered in the present study.

The tensification case in Korean therefore creates an opportunity to test the role of underlying versus surface representation in perceptual learning—i.e., whether perception learning that takes place on the surface forms make reference to their underlying representations, so that the learning effect may be generalizable to the underlying representations. Testing this is made possible by examining perceptual learning that involves ambiguous tensified stops in terms of place of articulation (henceforth transcribed as [$^{p*}/_{t*}$]), and whether learning about the [place] feature with tensified stops is generalizable to the underlyingly tense stops (that share the same surface (phonetic) forms) and to the underlying plain stops (that share the same underlying (phonological) forms).

Listeners were hence exposed to words with underlying lenis stops presented after the adjective /tʃuŋkuk/ (Engl., *Chinese*), which provides a tensifying context. One group of listeners hears ambiguous stops in words with underlying /p/ (i.e., [tʃuŋkuk{$^{p*}/_{t*}$}atʃi], where the ambiguous sound (transcribed as {$^{p*}/_{t*}$}) can only be interpreted as /p/, since /patʃi/ means *pants* while /tatʃi/ is a nonword). That is, this group learns to interpret the ambiguous stop as labial. The other group hears the ambiguous sound in an environment in which only an interpretation as /t/ is likely (e.g. [tʃuŋkukp{$^{p*}/_{t*}$}oma], where /toma/ means *cutting board* while /poma/ is a nonword in Korean). This group hence learns to interpret the ambiguous stop as alveolar.

One potential problem with this exposure is that the critical sounds are word initial, which may inhibit learning [6]. Therefore, we used, instead of the typical lexical decision task [12], a picture verification task. Participants first saw a picture and then heard a phrase (i.e., [tʃuŋkuk....]). They had to indicate whether the pictured object was mentioned in the phrase that they had heard. In this way, participants had a guide on how to interpret an ambiguous sound while hearing that sound, which is deemed critical for perceptual learning to occur [2].

After an exposure to a number of such items, both groups categorized place-of-articulation continua from a labial to an alveolar stop. We should expect that the first group is more likely to label an ambiguous stop on such a continuum as labial while the second group should label the same ambiguous stop as alveolar. This was tested with three different continua (three different conditions) as in (2):

(2) Three different conditions
  a. Baseline condition:
     [**p***antʃi]–[**t***antʃi] (in tensifying contexts)
     with underlying plain but tensified stops
  b. Generalization condition 1 (in Exp. 1):
     [**t***aŋtʃaŋsa]-[**p***aŋtʃaŋsa] (in isolation)
     with underlying tense stops
  c. Generalization condition 2 (in Exp. 2):
     [**p**antʃi]–[**t**antʃi] (in isolation)
     with underlying plain stops

With these continua, we conducted two experiments. In both experiments, participants were exposed to two test conditions after the same exposure (with tensified stops that were ambiguous in terms of their place of articulation, {$^{p*}/_{t*}$}). The first test condition was the baseline condition (2a) in which participants were tested on a continuum with tensified stops along with a preceding tensifying context. This was to examine whether perceptual learning on ambiguous sounds in terms of place indeed had taken place in both experiments.

The two experiments then differed in terms of which generalization condition was tested. In Exp. 1, the generalization condition used phonologically tense stops (2b); in Exp. 2, the generalization condition used phonologically plain stops (2c). (Note that in both generalization conditions, the words were presented in isolation, so that their phonetic forms are faithful to their underlying representations at least in terms of their tenseness.) Exp. 1 therefore tests whether learning can generalize to underlying tense forms. In Exp. 1, the baseline and the generalization conditions share the same phonetic representation ([+tense]) but differ in their underlying representations (underlying plain stops *vs.* underlying tense stops, respectively). In contrast, Exp. 2 tests whether learning can generalize to underlying plain stops. In this case, the baseline and the generalization conditions share the same underlying representation (plain, i.e., [-tense]), but they differ in their surface phonetic forms (tensified *vs.* underlyingly plain, respectively).

## 2. METHOD

### 2.1 Participants and materials

96 university students (in Seoul, Korea) participated in the study for pay. They were all native speakers of Korean. 48 subjects each participated in Experiment 1 and Experiment 2.

We identified 24 /p/-initial and 24 /t/-initial concrete nouns in Korean that were picturable and were nonwords if produced with the other stop (i.e., /p/ for /t/-initial words) as critical exposure items. We also generated 96 filler trials, among which 24 had matching and 72 had non-matching auditory word and picture. Among the non-matching cases, 12 /p/- and 12 /t/-initial words were visually presented, so that the presence of a picture for a /p/ or /t/ initial word was not a reliable cue that the target will be present.

These nouns were recorded by a native speaker of Korean in the context of the preceding adjective [tʃuŋkuk] (*Chinese*), which leads to the tensification of plain stops. Additionally, for the critical exposure items, the nonwords were recorded that arise by exchanging the initial /p/ or /t/ with a /t/ or /p/, respectively. Continua were generated from these word-nonword "minimal" pairs using the STRAIGHT auditory morphing algorithm [8], using the time-aligned version on the basis of hand generated phonetic segmentations of these stimuli. Five different native speakers then judged which of these sounded maximally ambiguous and those steps were used for exposure as ambiguous items.

For the picture verification task, we used Google image search to find suitable pictures, which were selected with the help of two native speakers. For the test phase, we recorded three minimal pairs: [tʃuŋkukp*antʃi]-[tʃuŋkukt*antʃi](tensifying context, *Chinese ring-Chinese pot*); [t*aŋtʃaŋsa]-[p*aŋtʃaŋsa] (in isolation, *landseller- breadseller*); [pantʃi]-[tantʃi] (in isolation, *ring-pot*). From these recordings, we generated 10-step continua, again using the STRAIGHT algorithm and selected a seven steps around the most ambiguous token for the test phase. The tokens had to be identified as starting with a labial or alveolar stop, using pictures of the respective words as response options.

### 2.2 Procedure

Participants were seated in front of a computer screen and instructed that there were two different tasks to perform. In the first task, they would hear a phrase and see a picture and indicate whether the phrase matched the picture. Each participant went through the 144 experimental trials, which were presented in a different random order for each participant. Random orders were constrained so that after each critical item, there had to be one filler item and no critical item was presented in the first three trials.

After completing these 144 exposure trials, participants performed a phonetic-categorization task in the test phase. Each participant was presented with the baseline-test and one generalization condition (which differed between Experiments, see above). The different continua were presented in a mixed fashion. Each of the 14 stimuli (two continua with 7 members) was presented 16 times, so that the test phase consisted of 224 trials in total.

## 3. RESULTS

### 3.1 Exposure

Note that the exposure phase was identical in both experiments. Participants overall accepted the items with ambiguous stops (as matched with the picture) (/p/-items: 91.9%, /t/-items: 95.1%), even though to a slightly lesser degree when the items were unambiguous (/p/-items: 93.4%, /t/-items: 98.3%). However, each participant accepted at least more than 75% of the critical items, so that each participant should have had the opportunity to learn. (Note that learning effects can decrease as participants reject more critical items.)

### 3.2 Test phase Experiment 1

Figure 1 shows the results from test phase in Experiment 1, in which participants were tested on a generalization continuum that was phonetically identical to the training items (= tensified stops) but phonologically different—i.e., underlyingly plain stops in the training items but underlyingly tense stops in the test items. The figure indicates a difference between the groups for both continua, that is, listeners generalized the learning from the tensified (underlyngly plain) stops to underlyingly tense stops. The statistical significance of the patterns in Figure 1 was tested using linear mixed-effect models in R (v3.1.2, using lme4, v1.01). Models were using a maximal random effects structure. We first tested whether there was a learning effect for each of the continua separately, which turned out to be the case (baseline: $b_{\text{Exposure}} = 1.29$, SE = 0.35, z = 3.64, $p < 0.001$; generalization: $b_{\text{Exposure}} = 1.34$, SE = 0.35, z = 3.85, $p < 0.001$). When analysed together, there was no interaction between the continuum and the learning effect (model comparison with and without interaction: $p > 0.2$).

### 3.3 Test Phase Experiment 2

Figure 2 shows that the Groups differ much less in the test phase of Experiment 2. There is only a small difference for the baseline continuum unlike the case in Experiment 1which showed a larger effect with the exact same training items. And identification functions for the generalization continuum are virtually of no difference between the ambiguous-to-alveolar group and the ambiguous-to-labial group.

Statistical testing revealed only a marginal learning effect for the baseline continuum ($b_{\text{Exposure}} = 0.60$, SE = 0.35, z = 1.73, $p < 0.1$) and no effect for

the generalization continuum ($b_{\text{Exposure}}$ = 0.25, SE = 0.34, z = 0.75, $p$ > 0.2). A combined analysis revealed no hint of an interaction of the learning effect with continuum (p> 0.2).

**Figure 1**: Results from Experiment 1 in terms of proportion "alveolar" responses, showing a group difference for both the baseline continuum (left panel and the generalization continuum (right panel).

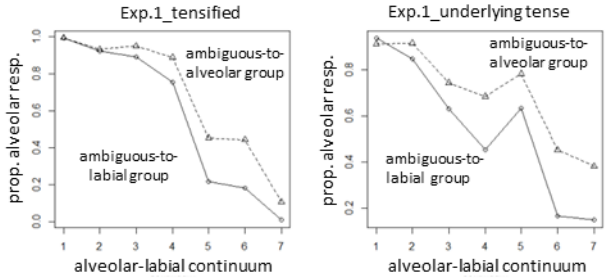
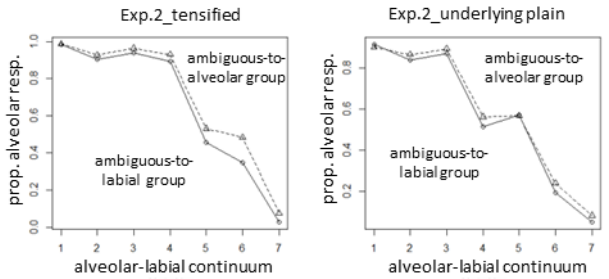



**Figure 2**: Results from Experiment 2 in terms of proportion "alveolar" responses, showing a group difference for both the baseline continuum (left panel and the generalization continuum (right panel).




## 4. DISCUSSION

The results show that perceptual learning regarding the stop's place of articulation indeed occurs even when the phonetic forms used for learning are derived as a consequence of a phonological rule (i.e., post-obstruent tensification in Korean). The results also show that such perceptual learning can generalize beyond the trained contrast. As found in Exp. 1, generalization occurs when the test items contain pivotal sounds that are phonetically the same as the ones in the exposure (training) items, even though they are different in their underlying phonological representations. However, as found in Exp. 2, generalization fails when the test items are different from the training items in terms of their phonetic forms (due to tensification on the training items), although they share the same underlying representation. These results therefore suggest that generalization does not make reference to abstract phonological representations, but rather it is strongly constrained by phonetic similarities between the learning and the test items. In other words, the difference in the underlying phonological representation as tense or plain do not hinder learning as long as there is a phonetic similarity between the learning and test items. This implies that perceptual learning takes place based on phonetic, rather than phonological, representations.

The results in Exp. 2 have also implications for generalizability of perceptual learning across features. As mentioned in the introduction, previous studies have shown that learning hardly generalizes across segments with different features. Reinsich et al. [15], for example, showed that learning about place does not generalize across segments with underlyingly different manners of articulation (e.g., oral vs. nasal). The results in Exp. 2 in the present study may be seen as showing that learning about place of a stop with a derived [+tense] feature is not generalized to a stop with no such derived feature, although the learning and test items share the same [-tense] feature in their underlying representations. This takes support away from the view that perceptual learning takes place on a featural base, but rather it implies that learning hardly generalizes across segments that are different in terms of other features, regardless of whether the other features are underlyingly specified or derived.

One caveat about the failure of generalization in Exp. 2 is that there was apparently no learning in the baseline condition of Exp. 2. This is puzzling, especially because exactly the same learning phase triggered learning for this continuum in Exp. 1. This was presumably because both the baseline condition and the generalization condition employed a continuum using the same minimal pair [pantʃi] – [tantʃi], thus potentially interfering with learning. We are currently testing this possibility in an additional experiment, and therefore our conclusions regarding the role of phonological features should be taken as interim.

Our results, subject to further corroboration, add to the literature a cautionary note on the use of features as basic units in speech perception. Features are often viewed as primary in both linguistics and neuroscience [3, 11]. Our results, however, imply that, in perception, listeners do not seem to abstract away from segments in which those features occur. This resonates with the classic problem of features, that is, their acoustic implementation is just too variable to be useful in perception, as already noted by Klatt [10].

## 5. REFERENCES


[1] Cho, T. et al. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, 30, 193–228.

[2] Cutler, A. et al. 2010. How abstract phonemic categories are necessary for coping with speaker-related variation. *Laboratory phonology 10*. C. Fougeron et al., eds. de Gruyter. 91–111.

[3] Embick, D. and Poeppel, D. 2014. Towards a computational (ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics. *Language, Cognition and Neuroscience*, 1–10.

[4] Goldinger, S.D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 22, 1166–1183.

[5] Goldstein, L. and Fowler, C.A. 2003. Articulatory phonology: A phonology for public language use. *Phonetics and phonology in language comprehension and production: Differences and similarities*. N.O. Schiller and A. Meyer, eds. Mouton de Gruyter. 159–207.

[6] Jesse, A. and McQueen, J.M. 2011. Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*. 18, 943–950.

[7] Jun, S.-A. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology*. 15, 189–226.

[8] Kawahara, H. et al. 1999. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction. *Speech Communication*. 27, 187–207.

[9] Kim-Renaud, Y.-K. 1974. *Korean consonantal phonology [PhD dissertation]*. University of Hawaii.

[10] Klatt, D. 1989. Review of selected models of speech perception. *Lexical representation and process*. W.D. Marslen- Wilson, ed. MIT Press. 169–226.

[11] Lahiri, A. and Reetz, H. 2010. Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*. 38, 44–59.

[12] McQueen, J.M. et al. 2006. Phonological abstraction in the mental lexicon. *Cognitive Science*. 30, 1113–1126.

[13] Mitterer, H. et al. 2011. Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science*. 35, 184–197.

[14] Mitterer, H. et al. 2013. Phonological abstraction without phonemes in speech perception. *Cognition*. 129, 356–361.

[15] Reinisch, E. et al. 2014. Phonetic category recalibration: What are the categories? *Journal of Phonetics*. 45, (Jul. 2014), 91–105.

[16] Sjerps, M.J. and McQueen, J.M. 2010. The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*. 36, 195–211.