Journal of Phonetics 56 (2016) 110–123

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

# What are the letters of speech? Testing the role of phonological specification and phonetic similarity in perceptual learning

Holger Mitterer [a], Taehong Cho [b,*], Sahyang Kim [c]

[a] Department of Cognitive Science, University of Malta, Msida, Malta
[b] Hanyang Phonetics and Psycholinguistics Lab, Department of English Language and Literature, Hanyang University, Seoul, Republic of Korea
[c] Department of English Education, Hongik University, Seoul, Republic of Korea

ARTICLE INFO

*Article history:*
Received 14 December 2015
Received in revised form
23 February 2016
Accepted 11 March 2016
Available online 2 April 2016

*Keywords:*
Speech perception
Perceptual learning
Prelexical processing
Phonetic similarity
Korean
Underlying representation

ABSTRACT

Recent studies on perceptual learning have indicated that listeners use some form of pre-lexical abstraction (an intermediate unit) between the acoustic input and lexical representations of words. Patterns of generalization of learning that can be observed with the perceptual learning paradigm have also been effectively examined for exploring the nature of these intermediate pre-lexical units. We here test whether perceptual learning generalizes to other sounds that share an underlying or a phonetic representation with the sounds based on which learning has taken place. This was achieved by exposing listeners to phonologically altered (tensified) plain (lax) stops in Korean (i.e., underlyingly plain stops are produced as tense due to a phonological process in Korean) with which listeners learned to recalibrate place of articulation in tensified plain stops. After the recalibration with tensified plain stops, Korean listeners generalized perceptual learning (1) to phonetically similar but underlyingly (phonemically) different stops (i.e., from tensified plain stops to underlyingly tense stops) and (2) to phonetically dissimilar but underlyingly (phonemically) same stops (i.e., from tensified plain stops to non-tensified ones) while generalization failed to phonetically dissimilar *and* underlyingly different consonants (aspirated stops and nasals) even though they share the same [place] feature. The results imply that pre-lexical units can be better understood in terms of phonetically-definable segments of granular size rather than phonological features, although perceptual learning appears to make some reference to the underlying (phonemic) representation of speech sounds based on which learning takes place.



## 1. Introduction

One of the most salient differences in the research on visual- and spoken-word recognition may be found in the issue regarding pre-lexical units—i.e., whether listeners use some form of abstraction in pre-lexical processing, and, if they do, what the nature of the units would be. In visual-word recognition, there is no controversy that letters are important units that help listeners to recognize words. Models of visual-word recognition usually involve some kind of letter units (Rastle, 2007), so that the difference between *lice* and *dice* at the lexical level, for example, is not the presence versus absence of a semicircle at the left edge with 'l', but the difference between the letter 'l' and the letter 'd' in the first position. However, it is precisely this kind of question that is controversial in the field of spoken-word recognition.

On the one end, it has been argued that there are no units at all in pre-lexical processing, as listeners store multiple "grainy spectrograms" for a word (Pierrehumbert, 2002) and the incoming input is compared to the multiple versions or the "episodes" for each word (Goldinger, 1998). This view provided a useful challenge for the existing models of spoken-word recognition, which tended to assume some kind of an intermediate unit between the acoustic input and the stored abstract mental representation in the mental lexicon (McClelland & Elman, 1986; Norris, 1994). As a consequence, subsequent studies explored this issue and provided some evidence that listeners indeed make active use of pre-lexical units in generalization of perceptual learning. For example, in a phonetic categorization study, confronted with 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

* Correspondence to: Hanyang Phonetics and Psycholinguistics Lab, Department of English Language and Literature, Hanyang University, 222 Wangsimni-ro, Seongdong-gu Seoul (133-791), Korea. Tel.: +82 2 2220 0746; fax: +82 2 2220 0741.
*E-mail addresses:* holger.mitterer@um.edu.mt (H. Mitterer), tcho@hanyang.ac.kr (T. Cho), sahyang@hongik.ac.kr (S. Kim).

http://dx.doi.org/10.1016/j.wocn.2016.03.001

not) but also learn, through multiple exposures, the speaker's idiosyncratic way of producing /s/. They are then able to generalize this learning to any other word containing /s/, showing that they have learned something about fine-grained pre-lexical information across words. This learning paradigm has hence led to the general acceptance of the assumption that some form of pre-lexical abstraction occurs between the acoustic output and the lexical representation in speech perception (Goldinger, 2007).

Part of the appeal of an episodic model may lie in the fact that it does not need to deal with the problems in delineating the fine-grained size of these units. At one point in the past, the selective-adaptation paradigm was thought to reveal pre-lexical units (Samuel, 1982), but it turned out that selective adaptation did not exclusively target pre-lexical representations and any perceptual decision on different levels could be influenced by selective adaptation (Remez, 1987). Similar problems arose with other paradigms, like the one using sub-categorical mismatches (Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999). Proponents of episodic models also pointed out that one could find evidence for any kind of unit in a selective-adaptation paradigm (Goldinger & Azuma, 2003). In summary, there never was any consensus on how the evidence from different types of procedures should be weighted with regard to the question of the grain size of pre-lexical units.

However, if the perceptual-learning paradigm is able to show the existence of some kind of pre-lexical units, it may also be useful to indicate what form these units have. This is especially so, because this paradigm reveals a unit that listeners are actively using in adapting to variability in speech, and as such, has some ecological validity. In an effort to explore the form of pre-lexical units with this paradigm, Mitterer, Scharenborg, and McQueen (2013) tested whether learning generalizes from one allophone of Dutch /r/ to another, with the reasoning that, if pre-lexical units are phonemic, learning should generalize over phonetic variants of a given phoneme, that is, over allophones. The results, however, showed no generalization from an approximant variant to a trilled variant of /r/, suggesting that learning does not generalize on a phonemic level. Based on this, Mitterer et al. argued that the pre-lexical units are not phonemic but are likely to be sub-phonemic in nature.

Generalization in the perceptual-learning paradigm was found, however, for the voicing contrast across place of articulation (Kraljic & Samuel, 2006). Participants, who learned that an ambiguous voice onset time (in terms of voiced versus voiceless) was indicative of a voiced /d/ rather than a voiceless /t/, generalized the learning to the voicing contrast in another place of articulation (e.g., between /g/ and /k/). This finding would indicate that the pre-lexical processing makes use of features (in this case, the [voice] feature) rather than phonemes. It is then useful to consider how a featural account might deal with the failure of generalization over allophones of /r/, because this may depend on what kind of features one would assume. If one assumes that the features refer to articulatory gestures, as assumed by Articulatory Phonology (Goldstein & Fowler, 2003), the failure of generalization is predicted, because completely different articulatory gestures are involved for a trilled variation and an approximant variation of /r/. If one, however, assumes the role of abstract phonological features (Lahiri & Reetz, 2010), and perceptual learning takes place on the [+rhotic] feature, generalization would be expected because both allophones share the feature.[1]

In fact, it appears that the two disciplines, psychology and linguistics that deal with spoken-word recognition have quite different ideas about the nature of pre-lexical units. Embick and Poeppel (2014) note that psychologists tend to think in terms of segments of some sort while linguistic accounts tend to argue for the role of phonological features. A recent study by Reinisch, Wozny, Mitterer, and Holt (2014), however, questioned the role of features. Reinisch et al. (2014) used a visually-guided learning paradigm (rather than a more commonly employed lexically-guided learning paradigm), in which an ambiguous sound is accompanied by an unambiguous visual signal. Whereas an ambiguous sound (e.g., a sound between /s/ and /f/ after *gira*…) may be disambiguated by lexical knowledge (because *giraffe* is a word and *girasse* is not), the paradigm makes use of a kind of the McGurk effect, so that an ambiguous syllable between /ba/ and /da/ is perceived as /ba/ if it is accompanied by a visual lip-closing gesture, the visual cue to /b/. If the same syllable is subsequently heard in an audio-only condition, it is more likely to be labeled as /ba/ than when the same syllable was previously accompanied by a visual gesture for /da/ without a lip-closing gesture (Bertelson, Vroomen, & de Gelder, 2003). Reinisch et al. (2014) tested how such learning or recalibration of phonetic categories may generalize, and found that it is context-specific. In contrast with the assumption that pre-lexical units are phonemic, they found that listeners did not generalize the visually-guided learning for /b/ in one vowel context (i.e., /aba/) to perception of the same phoneme /b/ in another vowel context (i.e., /ibi/). Furthermore, exploring the role of features, they also tested generalization from /aba/ to /ama/—i.e., whether the perceptual recalibration about place of articulation (the [place] feature for /b/ vs. /d/) may be generalizable to the /m/-/n/ contrast with the same [place] feature in question but across the manner of articulation (stops vs. nasals). Again they found no generalization. That is, the category boundaries regarding the [place] feature between /ama/ and /ana/, and between /aba/ and /ada/ were both amenable to be recalibrated if the visually-guided exposure contained the same manner of articulation (stops or nasals), but no generalization of the learning on the [place] feature occurred across different manners of articulation (from stops to nasals and vice versa).

Reinisch et al.'s results provide a useful challenge for theories that assume that speech processing is based on phonological features, independent of whether features are articulatory or acoustic. In a featural account, the input is generally analyzed 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 involves other features (such as manner features) which are not included in the situation in which learning has taken place. The aforementioned studies on perceptual learning, however, suggest that this is not the case. Instead, they show that generalization does not easily occur when there are phonetic differences in the surface (phonetic) representation between the segment with which learning has taken place (the learning condition) and the new segment to which learning may be generalizable (the generalization condition). This implies that perceptual recalibration

[1] Alternatively, one might argue that no learning should occur, since contextually specified features are not coded in the lexicon, and hence, lexically induced recalibration should not occur. But since learning was observed in the baseline condition by Mitterer et al. (2013), this alternative explanation can be ruled out as well.

is likely to be constrained by the difference in phonetic similarity between learning and generalization items, rather than by the difference in phonological features. More specifically, the fact that phonetic category recalibration fails to generalize not only across different manners of articulation, but also across different vowel contexts (e.g., Reinisch et al., 2014) leads to an assumption that generalization of perceptual learning is facilitated by 'phonetic similarity' or constrained by 'phonetic dissimilarity' in the surface representation between the learning context and the generalization context. In fact, Kraljic and Samuel (2006) also suggest that their finding of generalization may be due to the strong acoustic similarity of the critical cue in both conditions: the presence of some aspiration noise. If this turns out to be important, generalization should only occur to acoustically similar instances.

An important question that arises is then under which conditions perceptual learning can generalize from the exposure to a different test condition. This question is the point of departure for the present study, sparking a number of specific questions with a view to elucidating the nature of the difference (between the learning items and the generalization items) that constrains generalization of perceptual learning, which will eventually enhance our understanding of the nature of pre-lexical units.

The current study focuses on the role of surface and underlying representations and asks two different questions. The first specific question (to be explored in Experiment 1) is whether generalization is constrained by differences in the underlying representation even when the phonetic representations are the same—that is, whether learning can be generalizable to a phonetically same, but underlyingly different segment. This can be tested by examining how Korean listeners generalize lexically-guided perceptual learning over the labial-alveolar contrast to other phonetic forms. Korean has a post-obstruent tensification rule that makes a plain (lax) stop phonetically tense (tensified) when preceded by another obstruent (see below for further explanation on this process). This phonological process provides a testing ground for the question as to whether listeners generalize perceptual learning by making reference to underlying representations. Tensified plain stops due to a phonological process in Korean are phonetically the same as but underlyingly different from tense stops. So if generalization occurs purely on the basis of phonetic similarity between the segments in the learning and the generalization conditions, one might expect that generalization occurs over segments that are identical in terms of surface phonetic representation (from tensified plain stops to underlyingly tense stops), even if their phonological (underlying) representations are different (plain vs. tense). Alternatively, it is reasonable to assume that listeners may effectively restore underlying representations of phonologically altered sounds (tensified stops) to which they are exposed during learning. If so, listeners may generalize by making reference to underlying representations rather than to surface (phonetic) representations. In such a case, one might expect that generalization to underlyingly different sounds (i.e., from tensified 'plain' stops to 'tense' stops) may be constrained, even if they are phonetically the same.

Experiment 2 explores the mirror-reversed situation. The underlying representations of the segments used in exposure and test are the same, but the surface realization differs. This is achieved by testing whether learning that takes place based on phonologically derived (tensified) surface forms of plain stops generalizes to phonologically unaltered (non-tensified) plain stops (that are therefore phonetically different from, but underlyingly the same as tensified plain stops). Again if generalization is regulated purely by phonetic similarity, generalization from tensified plain stops to the non-tensified ones is likely to be constrained because the learning items and the generalization items are phonetically 'dissimilar.' This experiment hence may highlight a potential role of underlying representations in generalization of perceptual learning.

It is also conceivable that no generalization is found in both experiments, that is, from tensified stops to either tense stops (Experiment 1) or plain stops (Experiment 2). This prediction can be derived as follows. It has been argued that prosody plays an important role in segmental processing (Cho & McQueen, 2005; Cho, McQueen, & Cox, 2007; Kim & Cho, 2009, 2013; Kuzla, Ernestus, & Mitterer, 2010; Mitterer, Cho, & Kim, 2016). Kim & Cho (2013) found that listeners require a longer VOT to accept a stop as voiceless in English if the stop occurs after a major prosodic boundary. One possibility to account for this effect would be to argue that listeners have different pre-lexical representations for stops in different prosodic contexts, so that a given unit is only activated if the prosodic environment is appropriate for that unit. It would hence be possible to argue that pre-lexical representations include prosodic boundary information. If this is the case, the units recalibrated for tensified stops at a prosodic word boundary would be different from the units for underlyingly plain stops or the underlyingly tense stops after a larger prosodic boundary (since these words in the generalization condition were presented in isolation, thus forming an utterance preceded by a prosodic break).

To summarize, the experiments ask two questions. First, does generalization of learning occur when there is a difference in underlying representation but no difference in surface form? Second, does generalization of learning occur when there is a difference in surface form but no difference in underlying representation?

## 2. Experiment 1

In Experiment 1, we test whether learning can generalize to a phonetically similar but phonologically different segment. This is made possible by exploiting the feature of [tense] in Korean. Korean distinguishes three "laryngeal settings" in stops, so that a stop can be plain (lax, /p,t,k/), tense (fortis, /p*,t*,k*/), or aspirated (/pʰ,tʰ,kʰ/) (Cho, Jun, & Ladefoged, 2002). Crucially, a word-initial plain stop is phonetically realized as tense if the preceding word ends on an obstruent, a phonological process known as post-obstruent tensification (Jun, 1998; Kim-Renaud, 1974), as illustrated in (1).

(1) /tʃuŋku**k**/ + /**p**ɑtʃi/ → [tʃuŋku**kp***ɑtʃi][2]
'Chinese' 'pants'

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

The tensification in Korean therefore creates an opportunity to test the role of underlying vs. surface representations in perceptual learning—i.e., whether perceptual learning takes place based on the surface phonetic forms or whether it makes reference to their underlying representations. While this may be taken to be an explorative question, one might predict that a difference in underlying representation may constrain learning and prevent generalization, especially given that the extremely constrained learning found by Reinisch et al. (2014) suggests that any difference between the learning and the test items may be sufficient to block learning. As briefly mentioned earlier, Reinisch et al. also reported that the fine-grained phonetic detail constrains perceptual learning (e.g., learning on the particular phonetic form [b] due to the flanking vowel's coarticulatory influence in [ibi] did not generalize to [b] in another coarticulatory context [aba]). One might therefore predict that perceptual learning takes place purely based on the similarity in the phonetic forms. Given the specificity of learning found by Reinisch et al. (2014) it is also conceivable that pre-lexical representations are specific not only to their segmental environment (/i/ versus /a/) as context, but also specific to their prosodic environment.

Testing this was made possible by examining perceptual learning that involved (phonologically derived) tensified plain stops that were ambiguous in terms of place of articulation (henceforth transcribed as [$^{p*}/_{t*}$]), and whether learning about the [place] feature with tensified plain stops would be generalizable to the underlyingly tense stops that share the same surface (phonetic) forms, but occur in a different prosodic environment (i.e., in isolation)

Listeners were hence exposed to words with underlyingly plain stops presented after the adjective /tʃuŋkuk/ (Engl., *Chinese*) whose final obstruent (/k/) provided a tensifying context. One group of listeners heard ambiguous stops in words with underlying /p/ (i.e., [tʃuŋguk{$^{p*}/_{t*}$}adʒi], where the ambiguous sound (transcribed as {$^{p*}/_{t*}$}) could only be interpreted as /p/, since [padʒi] means *pants* while [tadʒi] is a nonword). That is, this group learned to interpret the ambiguous stop as labial (the *ambiguous-to-labial group* or the "Amb-to-LAB" group). The other group heard the ambiguous sound in an environment in which only an interpretation as /t/ is likely (e.g. [tʃuŋguk{$^{p*}/_{t*}$}oma], where /toma/ means *cutting board* while /poma/ is a nonword in Korean). This group hence learned to interpret the ambiguous stop as alveolar (the *ambiguous-to-alveolar group,* the "Amb-to-ALV" group).

One potential problem with this exposure was that the critical sounds were word-initial, which might inhibit learning (Jesse & McQueen, 2011). Therefore, we used a picture verification task, instead of the typical lexical decision task (McQueen, Cutler & Norris, 2006). Participants first saw a picture and then heard a phrase (i.e., [tʃuŋguk...]). 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 (Cutler, Eisner, McQueen, & Norris, 2010).

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

(2) Two different test conditions in Experiment 1

a. Baseline condition with tensified plain stops (same as in the learning phase):
[**p***andʒi]—[**t***andʒi] (*ring—pot*, in tensifying contexts)

b. Generalization condition with underlyingly tense stops:
**[t***ɑŋdʒaŋsa]–[**p***ɑŋdʒaŋsa] (*land-seller—bread-seller* in isolation)

## 2.1. Method

### 2.1.1. Participants

48 university students from Hanyang University (in Seoul, Korea) participated in the study for pay. They were all native speakers of Korean, free of hearing problems.

### 2.1.2. Materials and procedure

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 (e.g., /p/ for /t/-initial words) as critical exposure items (so that, for example, the initial ambiguous sound would be accepted as /p/ when it is lexically supported by an existing /p/-initial word, or as /t/ when there exists a /t/-initial word). Ten of the critical /p/-initial words were trisyllabic and the other fourteen were bisyllabic. For the critical /t/-initial words, twelve were trisyllabic and twelve were bisyllabic words. We also generated 96 filler trials, among which 24 had matching auditory words and pictures and 72 had non-matching ones. 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 with eleven stimuli were generated from these word-nonword "minimal" pairs using the STRAIGHT auditory morphing algorithm with the time-aligned version on the basis of hand generated

(*footnote continued*)
1998). That is, it is unlikely that listeners are able to reliably identity whether a given stop is tensified by the post-obstruent tensing rule or it is underlyingly tense (in the absence of lexical information) especially in a phrase-medial context, the precise environment that is employed by the present study.

phonetic segmentations of these stimuli. Five different native speakers then judged which of these stimuli 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 using the picture name as the search word, sometimes in conjunction with the word *Chinese*, since the items were presented after the adjective *Chinese*. This was necessary as, for instance, an AMTRAK train is not a good match for the phrase *Chinese train*. The initial selection was then reviewed by two native speakers and if they differed in their opinion about the prototypicality of a picture, a new one was searched for until both speakers agreed that the picture was sufficiently prototypical.

For the test phase, we recorded two minimal pairs: [tʃuŋgukp*andʒi] – [tʃuŋgukt*andʒi] (tensifying context, *Chinese ring-Chinese pot*); [t*aŋdʒaŋsa] – [p*aŋdʒaŋsa] (in isolation, *landseller–breadseller*). From these recordings, we generated 10-step continua, again using the STRAIGHT algorithm and selected seven steps around the most ambiguous token for the test phase. The tokens had to be identified as starting with a labial or an alveolar stop, using pictures of the respective words as response options.

### 2.1.3. Procedure

Participants were seated in front of a computer screen and were instructed that there were two different tasks to perform. In the first task, they would hear a phrase along with a picture on a computer screen. They were then supposed to 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 that no critical item was presented in the first three trials. Participants were divided into two groups to implement the critical between-participant manipulation of exposure condition. One group (*the ambiguous-to-labial group*, or the "Amb-to-LAB" group) heard the half of the critical items with an ambiguous sound that was lexically supported by a /p/-initial word, and the other half with an unambiguous (unaltered) /t/ matched with a /t/-initial word. The other group (*the ambiguous-to-alveolar group*, "Amb-to-ALV" group) heard ambiguous /t/-initial words and unambiguous /p/-initial words.

After completing these 144 exposure trials, participants performed a phonetic-categorization task in the test phase. Each participant was presented with the baseline condition (with the same tensified plain stops as used in the learning phase) and the generalization condition (with the new pivotal stops that are underlyingly tense). The different continua were presented intermixed. Previous studies (e.g., Mitterer & de Ruiter, 2008; Mitterer et al., 2013; Reinisch & Mitterer, 2016) found that the perceptual learning effect can dissipate during the learning phase, and by presenting the two continua intermixed, this dissipation may affect both continua to the same degree. This would not be the case for a blocked presentation, in which all items from one condition would first be presented, followed by all items from the other condition. 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.

## 2.2. Results

### 2.2.1. Exposure phase

In the exposure phase, as shown in Fig. 1a, participants overall accepted the items with ambiguous stops as matched with the picture (/p/-items: 92.2%, /t/-items: 95.3%), though to slightly lesser acceptance rates than when the items were unambiguous (/p/-items: 94.1%, /t/-items: 98.6%). (Note that for the sake of simplicity, Fig. 1 includes the acceptance rates for critical items for all the experiments of the present study.) The data show an interaction between item and group: there tend to be slightly lower acceptance rates in the ambiguous form than in the unambiguous form. In other words, the same ambiguous forms were perceived

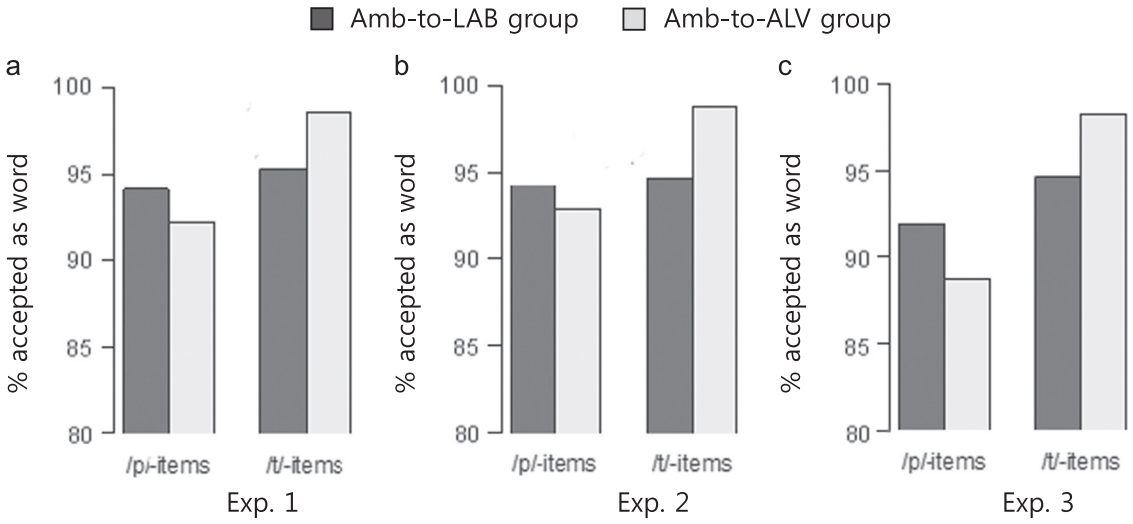


**Fig. 1.** Acceptance rates for critical items in the exposure phase of Experiments 1–3.The "Amb-to-LAB" group hears /p/ in the lexically guided ambiguous condition and /t/ in the unambiguous condition, and vice versa for the "Amb-to-ALV" group. (Note that in addition to Experiment 1, we carried out two more experiments which used the same critical items in the exposure phase. See the following sections for the detail of the other two experiments).

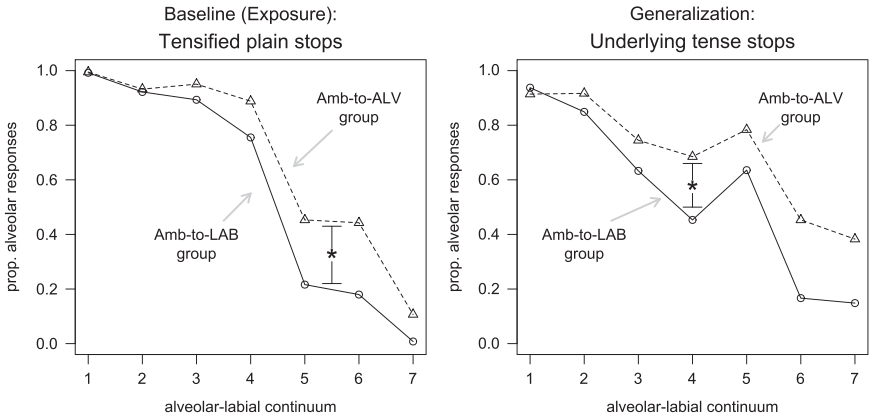


**Fig. 2.** Results from Experiment 1 in terms of proportion "alveolar" responses, showing a group difference for both the baseline continuum in the phonological tensifying context as was with the training (exposure) items (left panel) and the generalization continuum in the non-tensifying context (right panel).

as the /p/-items by the ambiguous-to-labial (Amb-to-LAB) group, but as the /t/-items by the ambiguous-to-alveolar (Amb-to-ALV) group, while the unambiguous forms were perceived more accurately (Recall that the Amb-to-LAB group heard unambiguous /t/-items and the Amb-to-ALV group heard unambiguous /p/-items.). Furthermore, each participant accepted at least more than 75% of the critical items, indicating that each participant had the opportunity to learn (Note that learning effects can decrease as participants reject more critical items.).

*2.2.1.1. Test phase.* Fig. 2 shows the results from the test phase, in which participants were tested on a generalization continuum that was phonetically identical to the training (exposure) items (tensified stops) but phonologically (underlyingly) different—i.e., listeners were exposed to underlyingly plain but phonetically tense stops in the training items and were tested with underlyingly tense stops in the generalization condition. The figure indicates a difference between the groups for both continua—i.e., listeners generalized the learning from the tensified (underlyingly plain) stops to underlyingly tense stops. The statistical significance of the patterns in Fig. 2 was tested using linear mixed-effect models in R (v3.1.2, using lme4, v1.01). Models were using a maximal random effects structure with the exception of a random slope of step over participants[3]. Predictors were Step, Exposure Condition (Amb-to-LAB vs. Amb-to-ALV), and Continuum (baseline versus generalization). In order to keep the model simple (and hence likely to converge) we did not specify all possible interactions, but only an interaction of Exposure Condition by Continuum (testing whether learning is the same across baseline and generalization condition) and a Step:Continuum condition, acknowledging the fact that the two continua may lead to differently steep categorization functions. The results (see Table 1) revealed a significant learning effect that was not modified by continuum, in line with what Fig. 2 suggests.

### 2.3. Discussion

The results showed 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). Crucially, the results showed that such perceptual learning can generalize beyond the trained contrast: generalization occurs when the test items contain pivotal sounds that are phonetically the same (i.e., tense) as the ones in the exposure (learning) items, but different in their underlyingly phonological representations (underlyingly plain vs. tense in the learning vs. the testing phase). These results therefore suggest that abstract phonological representations do not necessarily constrain generalization of perceptual learning, but that generalization is conditioned by phonetic similarities between the learning and the test items. In other words, the difference in the underlying phonological representation as plain (in the learning items) vs. tense (in the test items) does not hinder learning as long as there is a phonetic similarity on the surface between the learning and the test items. It also suggests that these pre-lexical representations are not specific to their prosodic environment, given the generalization from a prosodic word boundary context to an utterance context in which testing words were produced in isolation

In effect, Experiment 1 presents an extreme test of the assumption of context-specific phonological learning. It shows that minimal differences in terms of underlying phonological representation and prosodic environment are not sufficient to block the application of learning to new contexts. Experiment 2 now tests the opposite situation to Experiment 1. Does learning generalize when the underlying representations are the same but the surface representations are different? The data of Reinisch et al. (2014), as noted in the introduction, would suggest that generalization of learning is unlikely in this case, because even small phonetic differences in different coarticulatory context might be sufficient to block generalization of learning.

[3] We have often observed that adding random slopes for step leads to non-convergence which also was the case here. Note that our primary questions are neither concerned with the effect of step nor the interactions of step with other variables. The primary factor of interest is Exposure and its interaction with continuum. For the evaluation of these effects, only the included random slope for participant over continuum is necessary, since exposure is a between-participant variable.

**Table 1**
Results from the linear mixed-effect models for the test phase of Experiment 1. The critical regression weights are those for the Exposure Group, treatment coded with "Amb-to-ALV" (the ambiguous sound is alveolar) mapped on the intercept. For continuum, the baseline condition was mapped on the intercept, hence reflecting learning in this condition.

| Parameters | $\beta$ | SE ($\beta$) | $z$ |
|---|---|---|---|
| Step | −9.12 | 0.52 | −17.8*** |
| Generalization | −0.48 | 0.32 | −1.5 |
| Exposure = Amb-to-LAB | 1.15 | 0.34 | 3.4*** |
| Step:Generalization | 4.42 | 0.54 | 8.2*** |
| Amb-to-LAB:Generalization | 0.16 | 0.42 | 0.4 |

+ = $p<0.1$, * = $p<0.05$, *** = $p<0.001$.

## 3. Experiment 2

The tensification case in Korean also allows us to test whether the kind of perceptual learning that was observed in Experiment 1 (with tensified phonetic forms of underlyingly plain stops) might generalize to (non-tensified) underlyingly plain stops that are phonetically realized as plain on the surface. Whereas Experiment 1 focused on whether a difference in the underlying representation between the learning and the test items would constrain generalization across phonetically similar segments, Experiment 2 tests whether generalization would be constrained by phonetic discrepancy or distance between the learning and the test items which share the same underlying representation. Recall that earlier attempts to find such generalization have produced mixed results (Kraljic & Samuel, 2006; Reinisch et al., 2014), so addressing this question is explorative. If the phonetic similarity between learning and test items is a crucial factor that induces generalization of perceptual learning, there would be no generalization from the tensified stops to the underlyingly plain stops (because tensified plain stops and non-tensified ones are phonetically dissimilar).

### 3.1. Method

#### 3.1.1. Participants

48 university students from Hanyang University (in Seoul, Korea) participated in the study for pay. They were all native speakers of Korean, free of hearing problems. None of them had participated in Experiment 1.

#### 3.1.2. Materials and procedure

The materials and procedure for the exposure phase were the same as in Experiment 1. In the test phase, we used the same item as in Experiment 1 for the baseline condition, and used an additional minimal pair, [palgɨrim]–[talɨrim] for the generalization items (in isolation, 'painting of the foot'–'painting of the moon')[4], for which we also generated a 10-step continuum using the STRAIGHT algorithm. As for the other continua, seven steps around the most ambiguous token were selected for the test phase. The tokens (in both the baseline condition and the generalization condition) had to be identified as starting with a labial or an alveolar stop, using pictures of the respective words as response options. The two test conditions used in Experiment 2 are given in (3). The procedure was the same as in Experiment 1, with the only difference that the tokens for the generalization condition now were different.

(3) The two test conditions in Experiment 2

a. Baseline condition with *tensified* plain stops (same as in the learning phase):
[**p***andʒi]—[**t***andʒi] (*ring—pot*, in tensifying contexts)

b. Generalization condition with *non-tensified* plain stops:
[**p**algɨrim]—[**t**algɨrim] (*foot painting—moon painting*, in isolation)

### 3.2. Results

#### 3.2.1. Exposure phase

As was in Experiment 1, in the exposure phase, participants overall accepted the items with ambiguous stops (as matched with the picture) (/p/-items: 92.9%, /t/-items: 94.6%), but again to a slightly lesser degree than when the items were unambiguous (/p/-items: 94.3%, /t/-items: 98.8%) (see Fig. 1). However, each participant accepted at least more than 75% of the critical items, suggesting that each participant had the opportunity to learn.

#### 3.2.2. Test phase

Fig. 3 shows a clear group difference (Amb-to-LAB vs. Amb-to-ALV) for both baseline and generalization continua, showing generalization of perceptual learning to the test items with phonetically different but underlyingly same stops. This was confirmed by

[4] In the first attempt to run this experiment, we used [pandʒi] – [tandʒi] in the generalization condition, corresponding to the same words [p*andʒi] – [t*andʒi] (with tensified stops) in the baseline condition, and found no clear learning effect at all (not even in the baseline condition). While we can only guess at the reason for this, participants may find a long test phase more boring with just the same words in both conditions, leading to null effects.

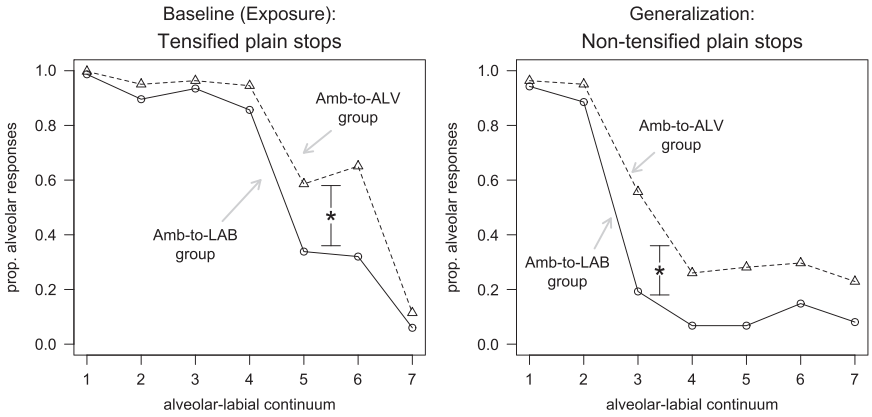


**Fig. 3.** Results from the test phase of Experiment 2 in terms of proportion "alveolar" responses, showing a group difference for both the baseline continuum in the phonologically tensifying condition as was with the training (exposure) items (left panel) and the generalization continuum with underlyingly plain stops in the non-tensifying content (right panel).

**Table 2**
Results from the linear mixed-effect models for the test phase of Experiment 2. The critical regression weights are those for the Exposure Group, treatment coded with "Amb-to-ALV" (the ambiguous sound is alveolar) mapped on the intercept. For continuum, the baseline condition was mapped on the intercept, hence reflecting learning in this condition.

| Parameters | $\beta$ | SE ($\beta$) | $z$ |
|---|---|---|---|
| Step | −9.09 | 0.43 | $-18.6^{***}$ |
| Generalization | −2.58 | 0.34 | $-7.5^{***}$ |
| Exposure =Amb-to-LAB | 1.32 | 0.31 | $4.2^{***}$ |
| Step:Generalization | 1.23 | 0.62 | $2.0^{*}$ |
| Amb-to-LAB:Generalization | −0.65 | 0.37 | $-1.7^{+}$ |

$^{+}=p<0.1$, $^{*}=p<0.05$, $^{***}=p<0.001$.

the statistical analysis as summarized in Table 2 (especially the significant group effect as reflected in the parameter, Exposure=Amb-to-LAB), which indicated that learning was present. The learning effect was only marginally modified by continuum (baseline versus generalization) as reflected in the parameter, Amb-to-LAB:Generalization, which was due to the tendency that learning was smaller in the generalization condition. Given this marginal interaction, we also investigated each condition separately, taking into account how the learning might change over the course of the test phase. For this, we used trial number as an additional predictor. Trial number was recoded to center around zero, ranging from −0.5 to 0.5, meaning that the model parameters reflect the overall mean over the course of the experiment.[5]

This revealed a learning effect for both continua (baseline, $\beta=1.21$, SE(β)=0.28, $z=4{,}26$, $p<0.001$; generalization, $\beta=0.76$, SE($\beta$)=0.20, z=3.71, $p<0.001$) and a decay of the learning effect over the course of the experiment that was only significant in the generalization condition (baseline, $\beta=-0.48$, SE($\beta$)=0.29, $z=-1{,}63$, $p=0.10$; generalization, $\beta=-1.02$, SE($\beta$)=0.27, $z=-3.89$, $p<0.001$). To confirm the statistical validity of this pattern, we ran an overall analysis with and without a three-way interaction of Exposure Condition, Trial Number, and Test Continuum, and found that the model with an interaction provided a better fit ($X^2(4)=104$, $p<0.001$), showing that there is a reliable difference between the two conditions.

### 3.3. Discussion

The results show that perceptual learning that takes place on tensified phonetic forms derived from underlyingly plain stops can generalize to new contrasts with (non-tensified) plain stops that are phonetically different from learning items. This is in contrast to what we expected on the basis of the findings of Reinisch et al. (2014).

The generalization observed here, however, is not complete. There was a significant decay of learning effects in the generalization condition (with plain stops) but not in the baseline condition (with tensified stops). The 'incomplete' generalization suggests that the phonetic difference between learning and test items still plays a role in generalization. In other words, the phonetic discrepancy in the surface form between tensified plain stops (in the learning items) and non-tensified ones (in the test items) could constrain the generalization, although the constraining effect is not as extreme as the null effect reported in Reinisch et al. (2014) which used a visually-guided learning paradigm (i.e., the phonetic discrepancy constrained generalization severely so that recalibration of phonetic category was not generalized at all to the same phonemes in the different vowel contexts, let alone to the different phonemes in the same phonetic context).

There are three possible explanations for the finding that generalization occurred in Experiment 2. First of all, listeners may make reference to underlying phonological representations during exposure with the tensified trials, so that generalization takes place when

[5] Were trial number entered as is, other model parameters would reflect the situation with trial number=0, that is, a trial number that was not even used in the experiment.

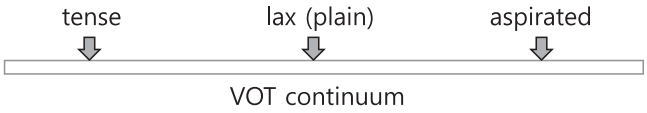


**Fig. 4.** The phonetic distance along the VOT continuum among three-way contrastive stops in Korean.

processing the plain trials. Alternatively, however, given that both the underlyingly plain stop (in the test items) and the tensified plain stop (in the learning items) share the same place feature (e.g., [labial] for [p*] and [p] or [alveolar] for [t*] and [t]), the result is consistent with the view that the perceptual recalibration operates on the feature (i.e., the place feature in this case) rather than on the underlying phonological (phonemic) representation. Even though the featural account was questioned by earlier work with the perceptual learning paradigm (Mitterer et al., 2013; Reinisch et al., 2014), these might be considered special cases. Given the centrality of this question, it would seem premature to make a firm conclusion based on just two studies. Finally, it may be the case that phonetic similarity has graded effects. That is, learning generalizes based on the degree of similarity between exposure and target items. In the case of, for instance, Mitterer et al. (2013), the phonetic difference between exposure and test items appears to be quite large. Listeners learned about approximant /ɹ/ and the generalization condition made use of trilled /r/. The presence of amplitude modulation in the latter case makes them psycho-acoustically quite distinct from the approximant, certainly more distinct than plain and tense stops in Korean, which are very difficult to distinguish for speakers of other languages.

Experiment 3 aims to distinguish between these accounts. To do so, it was tested whether learning regarding place of articulation in the same tensified stops generalizes to, first, aspirated stops and, second, nasals. The three different accounts for the results of Experiment 2 make differential predictions regarding generalizations to these two segments. If learning makes reference to underlying representations, a change in underlying representations from the tensified [p*] and [t*] used during exposure to different phonemes with different underlying representations (/pʰ, tʰ/ vs. /m, n/) at test should prevent generalization. In contrast, a feature account would predict that generalization should occur because all consonants involved share the same [place] feature. Finally, a phonetic-similarity account predicts that there will be differential degrees of generalization (in a gradient fashion) as a function of the phonetic similarity between learning and test items. For example, Korean has 'aspirated' stops in addition to plain and tense stops. Both nasals (/m, n/) and aspirated stops (/pʰ, tʰ/) share the same [place] feature with the tensified plain stops in the learning items, but the phonetic distance between the aspirated stops and the tensified plain stops are smaller than that between the nasals and the tensified plain stops (see below for further explanation). If the phonetic similarity is an important factor that constrains generalization, there would be difference in the degree of perceptual learning between the two conditions (with more robust generalization of learning to aspirated stops than to nasals). These possibilities are further tested in Experiment 3.

## 4. Experiment 3

In Experiment 3, we test whether the lexically-guided perceptual learning of the labial-alveolar contrast in tensified stops generalizes to different phonemes whose place features are the same as those of pivotal stops in the learning items. Experiments 1 and 2 revealed that perceptual recalibration of phonetic categorization with the tensified (plain) stops generalizes both to the underlyingly tense stops that are phonetically the same as the tensified stops and, though to a less degree, to the underlyingly plain stops that are phonetically different but underlyingly the same. As discussed before, these results were largely consistent with three accounts: a featural account, a phonetic-similarity account, and an underlying representation account. Experiment 3 therefore tests two more generalization conditions, which will help further disentangle the potential explanations of the results of Experiments 1 and 2.

The first generalization condition contains aspirated stops as pivotal consonants, so that the phonetic distance between the learning items and the test items (tense vs. aspirated) is only slightly greater than in Experiment 2 (tense vs. plain). The phonetic distance may be estimated based on how far two segments are pulled apart along the VOT continuum as shown in Fig. 4, especially given that VOT is one of the important acoustic-phonetic cues to the three-way contrastive stops in Korean (e.g., Cho et al., 2002): VOT is shortest for the tense, intermediate for the plain and longest for the aspirated in Korean.[6] Thus, the phonetic distance along the VOT continuum is greater between the tense and the aspirated stops (to be tested in Experiment 3) than between the tense and the plain stops (that were tested in Experiment 2). The second generalization condition contains nasals whose phonetic distance from the tense stops is assumed to be larger than the aspirated-tense distance. Nasals are acoustically different from oral stops in various aspects—e.g., in terms of the presence or absence of nasal murmur and voicing during the occlusion, which may be reflected by the difference in the manner features ([nasal] vs. [oral], or more broadly, [sonorant] vs. [obstruent]).

If perceptual recalibration of phonetic category for the labial–alveolar contrast operates on the [place] feature, the perceptual learning will generalize to both aspirated stops and nasals as they share the same place feature (e.g., from [p*] to [pʰ, m]). Alternatively, if recalibration of phonetic category occurs with reference to underlying representations, generalization will be non-applicable to both the aspirated stops and the nasals. Furthermore, if the phonetic distance between the pivotal consonants in the

[6] Note that although VOT has traditionally been considered as one of the primary cues to make a three-way distinction among Korean stops, recent studies (e.g., Kang, 2014; Silva, 2006) indicate that young Korean speakers produce plain and aspirated stops with substantial overlap in VOT, and rely more on F0 in the following vowel in distinguishing between plain and aspirated stops (lower for the plain and higher for the aspirated). However, stops used for our stimuli (which were produced by a relatively young Korean speaker (27 years old) show a clear three-way distinction in VOT (tense < plain < aspirated).

learning vs. the testing phase plays a role, perceptual learning is likely to occur with the phonetically similar aspirated stops, but not with the phonetically dissimilar nasals.

### 4.1. Method

#### 4.1.1. Participants

Seventy-six[7] university students from Hanyang University (in Seoul, Korea) participated in the study for pay. They were all native speakers of Korean, free of hearing problems. None of them had participated in Experiment 1 or 2.

#### 4.1.2. Materials and procedure

The materials and procedure for the exposure phase were the same as in Experiment 1 (i.e., with tensified plain stops in the learning items). For the test phase, we recorded two additional minimal pairs, [pʰalgɨrim]-[tʰalgɨrim] (with aspirated stops) and [magwi]-[nagwi] (with nasals) that share the same place feature with the tensified plain stops in the learning items. The aspirated stops in the generalization condition are phonetically closer to the tensified plain stops relative to the nasals (the "near distance" condition versus the "far distance" condition). The procedure was the same as in Experiment 1, with the only difference that the tokens for the generalization condition now were different as in (4). Note also that the baseline condition was not tested in Experiment 3, given the robust learning effects in the baseline condition with the same learning items were observed in Experiments 1 and 2.

(4) The two test conditions in Experiment 3

a. The near phonetic distance condition with the *aspirate*d stops:
[**p**ʰalgɨrim]–[**t**ʰalgɨrim] (*arm painting—mask painting*, in isolation)

b. The far phonetic distance condition with the *nasal* condition:
[**m**agwi]–[**n**agwi] (*devil—donkey*, in isolation)

### 4.2. Results

#### 4.2.1. Exposure phase

Two participants rejected more than 75% of the critical items and were hence excluded from further processing. As was the case with Experiments 1 and 2, for the remaining 74 participants the items with ambiguous stops were mostly accepted (as matched with the picture) (/p/-items: 88.7%, /t/-items: 94.6%), even though to a slightly lesser degree than when the items were unambiguous (/p/-items: 91.8%, /t/-items: 98.2%) (see Fig. 1).

#### 4.2.2. Test phase

Fig. 5 shows the results from the test conditions, with only small differences between the groups based on exposure ("Amb-to-LAB" vs. "Amb-to-ALV" groups). The statistical analysis, with the near distance condition (with aspirated stops) mapped on the intercept, confirms this. There is no overall effect of exposure and also no difference between the near distance and the far distance conditions (see Table 3, with no significant effect of Exposure and no interaction of Exposure with Continuum).

However, especially in the near-distance condition (with aspirated stops), there is an overall difference in Exposure Group in the expected direction—i.e., a trend towards generalization to aspirated stops. We therefore tested the learning (generalization) effect in the near-distance condition alone, which, however, turned to be non-significant ($B=0.46$, SE($B$)$=0.31$, $z=1.48$, $p=0.14$). To test whether the current finding is just a null result due to excessive error variance, we compared the learning effects here with those in the previous experiments. If there is excessive error variance in Experiment 3, no significant interaction should be observed. The analysis, however, revealed that the learning effect in Experiment 3 is significantly smaller than the learning effect in the earlier experiments ($\beta=-0.82$, SE($\beta$)$=0.31$, $z=-2.63$, $p<0.01$). This shows that the learning effect is at least significantly reduced.

### 4.3. Discussion

Experiment 3 reveals that perceptual learning as reflected in phonetic category recalibration that takes place during exposure to tensified (plain) stops does not generalize to either aspirated stops or nasals in the test phases, even though the exposure and the test items share the same place feature. This result therefore weakens the possibility that phonetic category recalibration for the labial-alveolar contrast operates on the place feature. The result also differs from the prediction made on the basis of a phonetic-similarity account. This account would have been strengthened by the finding that some learning occurs with the phonetically similar aspirated stops but not the phonetically dissimilar nasals. The result is hence most compatible with the underlying representation account—i.e., the reason why the aspirated stops and the nasals do not show generalization effects is that that they do not share the underlying representation with the tensified (plain) stops in the learning context. Furthermore, there is no significant difference in generalization between the aspirated stop (near phonetic distance) condition and the nasal (far phonetic distance) condition,

[7] We had originally tested 48 participants as in the other experiments, and since there was a visual trend towards a learning effect, we decided to add 24 more participants to make the results more solid.

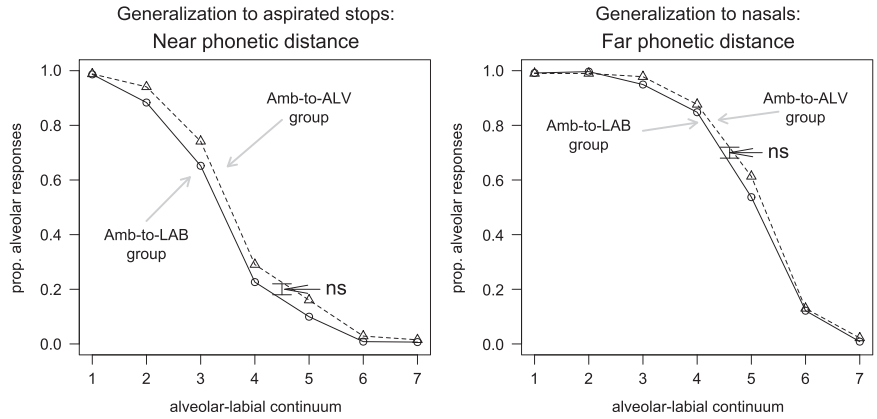


**Fig. 5.** Results from the test phase of Experiment 3 in terms of proportion "alveolar" responses, showing a small group difference for the near phonetic distance condition (left panel) and no difference for the far phonetic distance condition (right panel).

**Table 3**
Results from the linear mixed-effect models for the test phase of Experiment 3. The critical regression weights are those for the Exposure Group, treatment coded with "Amb-to-ALV"(the ambiguous sound is alveolar) mapped on the intercept. For continuum, the baseline condition was mapped on the intercept, hence reflecting learning in this condition.

| Parameters | $\beta$ | SE (β) | $z$ |
|---|---|---|---|
| Step | −2.08 | 0.04 | −54.42** |
| Near-Distance Continuum (aspirated) | −4.60 | 0.42 | −10.95*** |
| Amb-to-LAB (Exposure) | 0.39 | 0.26 | 1.51 |
| Step:Far Distance (nasal) | −0.30 | 0.07 | 4.46* |
| Amb-to-LAB: Far Distance Continuum (nasal) | −0.26 | 0.41 | −0.52 |

*$=p<0.05$, **$=p<0.1$, ***$=p<0.001$.

suggesting that the assumed difference in phonetic distance between the tensified (plain) stops (in the learning phrase) and the aspirated stops vs. the nasals (in the testing phrase) does not constrain the perception learning. This also weakens the phonetic similarity account, although it may still be possible that generalization fails when the phonetic distance between exposure and test stimuli exceeds a certain degree (or the maximal allowable phonetic distance), given that both aspirated stops and nasals are phonetically quite dissimilar from the tensified stops.

## 5. General discussion

With the three experiments reported above, we aimed at further exploring when perceptual learning generalizes to other contrast rather than the one exposed. The experiments addressed three questions. First, does perceptual recalibration of pre-lexical units in speech perception generalize to segments that are identical on the surface form but have a different underlying (phonological) representation? Second, does perceptual recalibration generalize to a segment that has the same underlying representation but a different surface form? Third, does perceptual recalibration occur across different phonemes that share the [place] feature with the learning items, and how does it matter that phonemes in the generalization conditions have differential degrees of phonetic similarity to the learning items?

Experiment 1 provided an answer to the first question, showing that perceptual learning is not constrained by underlying representations. When exposed to tensified but underlyingly plain stops that were ambiguous between a labial and an alveolar place of articulation, listeners learned to interpret these stops according to lexical information that was consistent with either labial or alveolar (i.e., *ambiguous-to-labial* vs. *ambiguous-to-alveolar*). During the test phase in which no disambiguating lexical information was provided, they perceived newly introduced ambiguous (underlying) tense stops as labial or alveolar, depending on the type of exposure they had—e.g., the same ambiguous stop was perceived as *labial* by listeners in the ambiguous-to-*labial* learning condition, but as *alveolar* by listeners in the ambiguous-to-*alveolar* learning condition. Crucially, the lexically-guided perceptual learning that took place with tensified but underlyingly plain stops generalized to underlyingly tense stops. If underlying representations constrained perceptual learning, the generalization would not have been possible. Instead, learning affected the perception of phonetically tense stops independent of whether the phonetic tense form was associated with an underlyingly tense stop or derived from an underlyingly lax stop due to a phonological process. This result thus implies that listeners generalize their learning by making reference to the phonetic representation rather than to the underlying representation. (No generalization from phonologically tensified stops to underlyingly tense stops would have been observed if perceptual learning was based on underlying (lax) representations of the tensified stops.)

Experiment 2 provided an answer to the second question by exploring whether perceptual learning that takes place based on the tensified plain stop would generalize to underlyingly plain stops that are not tensified. The result of Experiment 2 showed that perceptual learning *does* generalize to non-tensified plain stops which share the underlying representation with the tensified plain

stops. In other words, generalization occurred despite the phonetic differences between the learning and the test items (phonetically tense vs. phonetically plain). The result therefore suggests that the listeners do make reference to underlying representations when processing phonologically altered surface forms, and use the information in perceptual recalibration.

Given that both the tensified stops (in the learning condition) and the plain stop (in the generalization condition) share the same place feature, however, the result of Experiment 2 sparked another question: whether perceptual learning in Experiment 1 is due to learning based on the features or learning based on the underlying representation. Furthermore, the generalization effect with the plain stop in Experiment 2 was not as robust as that with the tense stop in Experiment 1, suggesting that the phonetic similarity of tense stops with tensified plain stops may have played an important role. It was therefore tested whether the degree of phonetic similarity between the learning items and the test items would influence perceptual learning in a gradient fashion. These questions were explored in Experiment 3 by testing generalization of perceptual learning (with tensified plain stops) to aspirated stops *vs.* nasals that are different from tensified plain stops in terms of both phonetic similarity and underlying representation. The aspirated stops and the nasals are also different from each other in terms of their relative phonetic similarity to tensified stops: aspirated stops are phonetically more similar to tensified plain stops than nasals are. We therefore asked whether this assumed relative difference in the degree of phonetic similarity would show differential degrees of generalization. The results, however, revealed that no generalization occurred to the two consonant types which were different from tensified plain stops in terms of both phonetic similarity and underlying representation. Given that this result was obtained despite the fact that aspirated stops and nasals share the [place] feature with tensified plain stops in the learning condition, these results ruled out the featural account. The result did not fully support the phonetic similarity account, either, especially given that the assumed differential degrees of phonetic similarity between the learning and the test items for aspirated stops vs. nasals did not turn out to be a significant influencer on perceptual learning. The result is therefore most consistent with the view that listeners make some reference to underlying phonological representations, and use them for generalization, which accounts for the observed generalization pattern in Experiment 2 (from tensified plain stops to non-tensified plain stops).

One might therefore argue that the current study at least found some generalization across items that share the same underlying representation (as was found in Experiment 2), whereas Reinisch et al. (2014) found no effect at all even across the same phonemes in different coarticulatory contexts (e.g., [ibi] vs. [aba]). This comparison, however, appears to confound stimulus variability during exposure with the acoustic similarity of exposure and the test stimuli in the generalization condition. One cannot entirely rule out the possibility that generalization of perceptual learning based on tensified stops to the plain stops is not because they share the underlying representations but because the plain stops are acoustic-phonetically more similar to the tensified stops than aspirated stops and nasals which showed no generalization. If so, the result is comparable with the finding of Kraljic and Samuel (2006), who found generalization for the voicing contrast over place. In this case, there also was an acoustic overlap (phonetic similarity) between the baseline and the generalization conditions (in the sense that for both, the presence of some aspiration noise was critical). Furthermore, the fact that the generalization effect was attenuated when generalization items with the non-tensified plain stops were phonetically dissimilar to the tensified plain stops in the learning phases suggests that phonetic similarity works possibly in a gradient fashion. However, it appears that perceptual learning fails to generalize when generalization items are phonetically far distant from the learning items, perhaps far beyond the minimal phonetic distance that is required for generalization, as reflected in the failure of generalization from tensified plain stops to aspirated stops and nasals in Experiment 3.

The results of three experiments, taken together, suggest that listeners make reference to both the phonetic similarity (between the learning and the generalization items) and the underlying phonological representations, but not necessarily to a particular phonological feature. In other words, the perceptual recalibration generalizes not only to other phonemes that are phonetically the same as phonologically altered pre-lexical units in the learning items, but also to the underlying representation of the phonologically altered pre-lexical units even when there are phonetic differences. We therefore interpret the results as implying that perceptual recalibration is modulated by an intricate interaction between the phonetic similarity and the underlying representation whose exact nature remains to be further explored.

Results of the present study also have some methodological implications as to whether learning generalizes more easily when learning is based on exposure to a range of words rather than just one stimulus, and to what extend learning can be generalized. Regarding the first question, the comparison of the results of Reinisch et al. (2014) (with only one exposure stimulus in a visually-guided perceptual learning paradigm) and the present study (with multiple exposure stimuli) suggests that this may not be an issue. Both studies tested whether learning about place (alveolar vs. labial) in stops generalizes to nasals, and both studies found that no generalization occurs. As such, the current results suggest that the failure of generalization in Reinisch et al. (2014) may not be a paradigm-specific artefact (see also Reinisch & Mitterer, 2016).

What do these results mean for the question as to whether pre-lexical processing makes use of segments of some sort or decomposes the signal into features? At least, the current results do not support the featural account (e.g., Lahiri & Reetz, 2010) which assumes that features are perceived independently, based on which one might predict generalization in all of our three generalization conditions (with plain stops, aspirated stops, and nasals in the test items which share the same [place] feature with the tensified stops in the learning items). Our results are therefore not interpretable in terms of the featural account and seem to argue for some kind of segmental representation on a pre-lexical level, as often assumed in models of spoken-word recognition (McClelland & Elman, 1986; Norris & McQueen, 2008).

There might be two related issues here. First of all, recent neuroscientific evidence using cortical surface recordings in humans seems to argue for features rather than segments (Mesgarani, Cheung, Johnson, & Chang, 2014). The authors find that single electrodes tend to code not only for one segment but for a variety of acoustically similar segments. Note, first of all, that these findings

are not in line with feature systems that make use of place features like [labial], since the response patterns do not show electrodes that respond to acoustically dissimilar segments that share a place of articulation. Secondly, it is far from clear that segments may not be an emerging property of this seemingly distributed neural net, as the properties of a neural net may be difficult to predict from local response properties. Finally, evidence for or against a given theory should not be weighted by the cost for that evidence, despite the cognitive bias to do so (Festinger, 1962). The importance of evidence for a given debate is related to the number and plausibility of the auxiliary assumptions that are necessary. The perceptual learning paradigm shows which types of units listeners use to make functional generalization and as such seems like the prime candidate to reveal such units.

A second problem for a segmental account is to explain those generalizations across segments that do occur. A tentative answer might be that the "segments" at a pre-lexical level may be quite different from the segments that we typically think of. Instead, fine-grained phonetic events such as "short noise burst" or "aspiration noise" may be pre-lexical units. If one considers units of such a granular size as segments, generalization can be said to operate on phonetically-definable segments. Other units may also encode for frequently occurring segment sequences (Poellmann, Bosker, McQueen, & Mitterer, 2014). This would actually be in line with the type of segments that are in use in speech recognizers, which are usually allowed to generate multiple allophones for each segment. While this is speculation, such phonetically-detailed pre-lexical representation in speech perception may also explain why the alphabetic principle for written language was discovered only once: Because it is in fact much more different from the segments used in speech processing than we typically think (see also Mitterer & Reinisch (2015)).

To summarize, the present study investigated when perceptual recalibration of pre-lexical processing generalizes to other contrasts. The results show that some generalization is possible in two cases: when there is a strong acoustic-phonetic similarity between the trained and the generalization contrast that arises due to a phonological process (e.g., the tensified plain stops vs. the underlyingly plain stops) or when generalization items share an underlying representation with the learning items. But in the latter case, generalization is not as robust as in the former case, and the learning appears to dissipate faster than for the learned contrast, which suggests that phonetic similarity comes into play: the non-tensified plain stops are phonetically different from the tensified plain stops. No generalization was found when pivotal consonants in the generalization condition were aspirated stops or nasals although they share the same [place] feature: They did not share the underlying representation with the learning items, nor were they similar acoustic-phonetically to the learning items. This indicates that pre-lexical processing employs phonetically-definable segments of granular size, operating based on phonetic similarity with some reference to underlying (phonemic) representations, rather than making use of independent features. The results of the present study warrant further investigation concerning the exact mechanism of the role of phonetic similarity and its intricate relationship with the underlying (phonological) representation in pre-lexical processing.

## Acknowledgment

We thank our graduate student assistants, Daejin Kim, Jiyoung Jang and Yuna Baek for assisting us with data acquisition. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013S1A2A2035410) to the corresponding author (T. Cho).

## References

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychological Science*, *14*, 592–597.
Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, *30*(2), 193–228.
Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33*(2), 121–157, http://dx.doi.org/10.1016/j.wocn.2005.01.001.
Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, *35*(2), 210–243, http://dx.doi.org/10.1016/j.wocn.2006.03.003.
Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology*, 10 (pp. 91–111). Berlin: de Gruyter.
Embick, D., & Poeppel, D. (2014). Towards a computational (ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics. *Language, Cognition and Neuroscience*, 1–10.
Festinger, L. (1962). *A theory of cognitive dissonance*, Vol. 2. Stanford, CA: Stanford University Press Retrieved from ⟨https://books.google.de/books?hl=de&lr=&id=voeQ-8CASacC&oi=fnd&pg=PA1&dq=Festinger+dissonance+reduction&ots=9x64Szvetv&sig=5TU_LqwoOck9Jp45SVQmlmq2jwo⟩.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279, http://dx.doi.org/10.1037//0033-295X.105.2.251.
Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th international congress of phonetic sciences* (pp. 49–54). Dudweiler, Germany: Pirrot.
Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *Journal of Phonetics*, *31*, 305–320, http://dx.doi.org/10.1016/S0095-4470(03)00030-5.
Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: a phonology for public language use. In N. O. Schiller, & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: differences and similarities* (pp. 159–207). Berlin: Mouton de Gruyter.
Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, *18*, 943–950, http://dx.doi.org/10.3758/s13423-011-0129-2.
Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, *15*(02), 189–226. ⟨http://doi.org/null⟩.
Kang, Y. (2014). Voice onset time merger and development of tonal contrast in Seoul Korean stops: a corpus study. *Journal of Phonetics*, *45*, 76–90.
Kim-Renaud, Y.-K. (1974). *Korean consonantal phonology(PhD Dissertation)*. Hawaii: University of Hawaii.
Kim, S., & Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean. *The Journal of the Acoustical Society of America*, *125*(5), 3373, http://dx.doi.org/10.1121/1.3097777.
Kim, S., & Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *The Journal of the Acoustical Society of America*, *134*(1), EL19–EL25.
Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, *13*, 262–268, http://dx.doi.org/10.3758/BF03193841.
Kuzla, C., Ernestus, M., & Mitterer, H. (2010). Compensation for assimilatory devoicing and prosodic structure in German fricative perception. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology*, 10 (pp. 731–758). Berlin: Mouton.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*(1), 44–59.
Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review*, *101*(4), 653–675, http://dx.doi.org/10.1037/0033-295X.101.4.653.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86, http://dx.doi.org/10.1016/0010-0285(86)90015-0.
McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*, 1113–1126, http://dx.doi.org/10.1207/s15516709cog0000_79.
McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1363–1389, http://dx.doi.org/10.1037/0096-1523.25.5.1363.
Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–101010, http://dx.doi.org/10.1126/science.1245994.
Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *Journal of Phonetics*, *54*, 68–79, http://dx.doi.org/10.1016/j.wocn.2015.09.002.
Mitterer, H., & de Ruiter, J. P. (2008). Recalibrating color categories using world knowledge. *Psychological Science*, *19*(7), 629–634, http://dx.doi.org/10.1111/j.1467-9280.2008.02133.x.
Mitterer, H., & Reinisch, E. (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language*, *85*, 116–134, http://dx.doi.org/10.1016/j.jml.2015.08.005.
Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, *129*(2), 356–361, http://dx.doi.org/10.1016/j.cognition.2013.07.011.
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234, http://dx.doi.org/10.1016/0010-0277(94)90043-4.
Norris, D., & McQueen, J. M. (2008). Shortlist B: a bayesian model of continuous speech recognition. *Psychological Review*, *115*, 357–395, http://dx.doi.org/10.1037/0033-295X.115.2.357.
Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory Phonology*, VII (pp. 101–139). Berlin: Mouton de Gruyter.
Poellmann, K., Bosker, H. R., McQueen, J. M., & Mitterer, H. (2014). Perceptual adaptation to segmental and syllabic reductions in continuous spoken Dutch. *Journal of Phonetics*, *46*, 101–127.
Rastle, K. (2007). Visual word recognition. *The Oxford Handbook of Psycholinguistics*, 71–87.
Reinisch, E., & Mitterer, H. (2016). Exposure modality, input variability and the categories of perceptual recalibration. *Journal of Phonetics*, *55*, 96–108, http://dx.doi.org/10.1016/j.wocn.2015.12.004.
Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: what are the categories?. *Journal of Phonetics*, *45*, 91–105, http://dx.doi.org/10.1016/j.wocn.2014.04.002.
Remez, R. E. (1987). Neural models of speech perception: a case history. In S. Harnad (Ed.), *Categorical perception: the groundwork of cognition* (pp. 199–225). Cambridge, Mass: Cambridge University Press.
Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, *31*(4), 307–314.
Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, *23*(02), 287–308.