

How abstract are phonological representations? Evidence from distributional perceptual learning

Bożena Pająk and Roger Levy
University of California, San Diego

1 Introduction

Phonological representations are typically thought of as abstract: for example, phonemes are argued to be comprised of abstract subphonemic units of some sort, whether distinctive features (e.g., Chomsky & Halle 1968), articulatory gestures (e.g., Browman & Goldstein 1989), or acoustic-phonetic dimensions (e.g., Pierrehumbert 2000). There is direct evidence that people are sensitive to these phonemic and subphonemic units. A good example comes from studying speech errors: as shown in (1) and (2), the phrase “big and fat” can sometimes be mispronounced by transposing whole segments (as in (1), where two initial stops are switched), but also just parts of segments (as in (2), where only the voicing appears to be switched between the initial stops; Fromkin 1973).

- (1) **big** and **fat** → **fig** and **bat**
- (2) **big** and **fat** → **pig** and **vat**

This kind of evidence suggests that in speakers’ mind not only segments are represented as single units, but that subphonemic properties such as voicing also have some psychological reality in that they can be abstracted away from individual segments, even in running speech. In this paper we are interested in investigating further the degree of abstractness of these subphonemic properties. Namely, are subphonemic properties represented as completely abstracted away from individual segments? Or are these representations mediated by the acoustic (or gestural) properties of the segments? For example, the classes of obstruents and sonorants are acoustically (and gesturally) very distinct, but laryngeal contrasts (e.g., *voiced* vs. *voiceless*) can apply to segments from either of these two classes. Do speakers, then, represent *voicing* as a single property that can be applied to any segment, or are there two (or more) distinct representations of *voicing*, each specific to a given class of segments?

A subphonemic property that is convenient for investigating this kind of question is contrastive *length*, since segmental length is a relatively salient acoustic-phonetic cue that cross-cuts a wide range of possible segments, as shown in (3a-d).

- (3) a. [taka-] vs. [takka] Finnish: *back* vs. *fireplace*
- b. [kisasi] vs. [kissaki] Japanese: *empress* vs. *point of a sword*
- c. [belo] vs. [bello] Italian: *I bleat* vs. *beautiful*
- d. [seki] vs. [se:ki] Japanese: *seat* vs. *century*

The acoustic correlates of length are different for different classes of segments. For example, length can be signaled by closure duration (stops), duration of friction noise (fricatives), or duration of voicing, as well as formant transitions, or intensity (sonorant consonants and vowels).

There have been many different proposals regarding formal means of representing length: using the distinctive feature $[\pm\text{long}]$ (Chomsky & Halle 1968), timing slots (Levin 1985, Selkirk 1991, Tranel 1991, Hume *et al.* 1997), moras (Hyman 1985, Hayes 1989, Davis 1999), or a combination of the latter two (Muller 2001). In this paper we make no attempt to differentiate among these different analyses. Instead, what is of interest to us is that all of these proposals share a commonality in that length is represented as abstracted across different segments, despite different raw acoustic cues that signal segmental length differences. The question we are asking here is whether this is a justified assumption when psychological representations are concerned. More specifically, is there a single psychological representation of length as independent from individual segments? Or are there distinct representations of length for different segment classes? We hope that addressing these questions will provide some additional insight into how length should be formally represented, as well as how phonological knowledge is represented psychologically.

One way of probing the abstractness of phonological representations is by looking at how novel phonological contrasts are learned and generalized. For example, we can expose adult learners to a new language, where there are novel length contrasts for some set of segments (e.g., sonorant consonants). Subsequently, we can test participants' categorization of short and long segments for (1) segments they have been trained on (e.g., sonorant consonants), and (2) novel segments (e.g., voiceless fricatives). If participants generalize—that is, if they categorize short and long segments in the same way for both trained and novel segments—then it would suggest that length is represented as abstracted across different segment classes (in this case, at least shared between sonorant consonants and voiceless fricatives). No generalization, on the other hand, would suggest that the representation of the length property is not completely independent from individual segments, but rather is more specific to segment classes, and perhaps mediated by between-segment phonetic similarity.

In this paper we present results of two experiments. In experiment 1 we expose participants to a new language with a novel length contrast for one class of segments (either sonorant consonants or voiceless fricatives), and subsequently test for learning and generalization to another class of segments. The main results from experiment 1 were originally reported in Pajak & Levy (2011), but here we provide an additional analysis of these data that was not discussed in the original paper. Experiment 2 is a follow-up to experiment 1, where we manipulate the stimuli so as to make the length cue more salient, with the goal to facilitate learning.

2 Experiment 1

In this experiment we exposed monolingual English speakers to evidence suggesting a novel phonological contrast between short and long segments in a new language. The exposure to the language was done through the distributional learning paradigm, as applied by Maye & Gerken (2001) in a study with adult participants.

Subsequently, we tested participant's categorization of short and long segments for trained and untrained segment classes (sonorant consonants and voiceless fricatives). For the main result, we predicted that participants would both learn the length contrast and generalize it from a trained class to an untrained class. As we already reported in Pająk & Levy (2011), this was indeed the case when the participants were trained on the sonorant class, but not when they were trained on the fricative class, in which case we saw no evidence of learning. Here, we report a more detailed analysis of the same data by looking at individual segments. It is possible that the previously reported results were driven by only a subset of segments. Therefore, in the current analysis we are checking whether the results were robust and held for each individual segment.

2.1 Method

2.1.1 Paradigm

We designed the experiment using the distributional learning paradigm (Maye & Gerken 2000, Maye & Gerken 2001, Maye *et al.* 2002), where the main idea is to provide experiment participants with exposure to a new language that is more akin to natural first language acquisition than previously used experimental paradigms that rely on explicit perceptual training (e.g., McClaskey *et al.* 1983). In natural languages, sounds vary along continuous dimensions. Thus, we can create sounds that vary gradually along a given dimension (in this case, length), and simply expose participants to sounds that come from this continuum. Crucially, we can vary the frequency with which individual tokens are presented, as illustrated in Fig. 1. For example, more frequent presentation of tokens from the endpoints of a continuum (in the case of length, either relatively short or relatively long segments) should lead participants to infer bimodal (or, two-peak) distribution of the data, while more frequent presentation of tokens from the middle of the continuum (i.e., medium-length segments) should lead them to infer unimodal (or, one-peak) data distribution. Bimodal data distribution suggests two underlying categories along the continuum (in this case, a contrast between short and long segments), while unimodal distribution suggests the existence of only one underlying category (i.e., no length contrasts). Following the paradigm, we can then present participants with sounds that are endpoints of the continuum, and ask whether they think the sounds are 'same' or 'different'. If the distributional training worked, we would expect participants trained in the bimodal condition to respond 'different' more often than participants trained in the unimodal condition. This paradigm lets us, then, look at learning and generalization by comparing two groups of participants: (1) participants trained with a bimodal distribution, and (2) participants trained with a unimodal distribution. Crucially, all participants are trained with novel sounds that come from the same exact continuum, which means that any differences between the bimodal and the unimodal conditions must be due to participants' interpretation of the novel sounds as influenced by training, and not just auditory sensitization.

2.1.2 Participants

48 undergraduate students at UC San Diego participated in the experiment for course credit. They were all monolingual speakers of English, in most cases with

some limited high school and/or college exposure to Spanish or French. Crucially, none of them had any exposure to any language that uses length contrastively. All participants reported no history of speech or hearing problems.

2.1.3 Materials

The materials consisted of nonce words recorded in a soundproof booth by a phonetically-trained native speaker of Polish. The critical length items included segments from two classes: sonorants ([j], [l], [m], [n]), and fricatives ([s], [f], [θ], [ʃ]). They were recorded as words with long consonants: [ajja], [illa], [amma], [inna], [assa], [iffa], [aθθa], [iʃʃa]. Subsequently, the consonant length in each word was manipulated to create length continua, each with eight tokens. There are several ways in which such continua could be created. One way would be to maintain natural between-segment duration differences (e.g., sonorant consonants are generally shorter than fricatives¹), but manipulate relative durations so that for each continuum the endpoints are always in the same duration ratio (cross-linguistically, the long-to-short consonant ratio varies between 1.5 to 3; Ladefoged & Maddieson 1996). Another way, which we adopted, is to use the same distribution on absolute durations for all segments (see the discussion section for more on the consequences of this choice). In the continua we created, durations of all consonants ranged from 100msec (short) to 205msec (long), and each adjacent token differed by 15msec. The fillers resembled the critical items, but different consonants were used: [ira], [iʔa], [aʒa], [aʦa], [atʃa], [idʒa], [itsa], [aba], [apa], [ida], [ita], [aga], [aka], [ixa], [ixʌ], [aʁa], [aŋa].

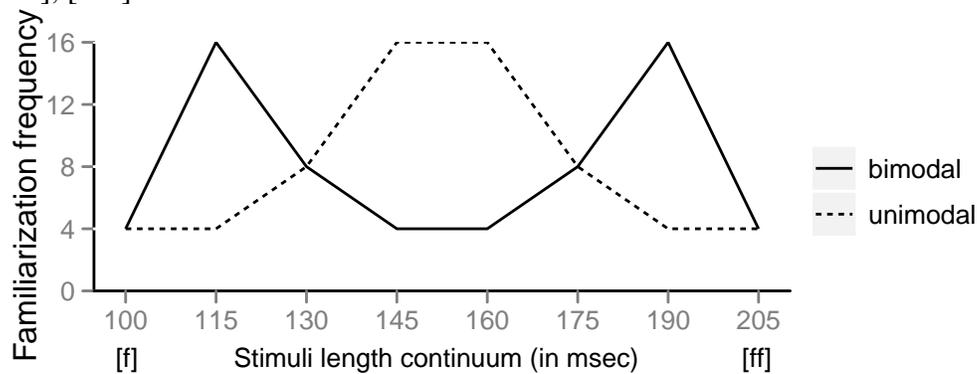


Figure 1: Critical training stimuli in experiment 1.

2.1.4 Procedure

The experiment adhered as closely as possible to the procedure used by Maye & Gerken (2001), and consisted of two main parts: training and testing.

¹The ranges of duration for English consonants that are equivalent to those used in the experiment are roughly the following (in msec): [j] 39-100, [l] 42-85, [m] 50-89, [n] 38-83, [θ] 46-90, [f] 56-119, [s] 61-126, [ʃ] 88-138 (based on the phonetically annotated portion of the Switchboard corpus, as described in ‘The Switchboard Transcription Project’ report by Steven Greenberg, 1996.)

Training: In training, participants listened to single words presented over headphones that were of one of two STIMULUS TYPES: *critical* or *filler*. Each participant was trained on critical items from one TRAINED SEGMENT CLASS (either *sonorants* or *fricatives*), and in one of two CONDITIONS: (1) *bimodal*, imitating a language with phonemic contrasts between short and long consonants, and (2) *unimodal*, imitating a language with no phonemic length contrasts (see Fig. 1). All participants were trained on the same filler items: the words [ira], [iʔa], [aɕa], [atʃa]. To maintain participants' attention on the experimental items, they were instructed to push a button after they heard each word. The response to a given stimulus triggered the presentation of the following stimulus with a delay of 1sec. Training consisted of a total of 384 words and lasted for about 10min. This included four repetitions of a training block, where each block had 64 critical items (16 tokens from each length continuum) and 32 filler items (8 different recordings of each item). Stimulus order was randomized for each participant, and there was a self-terminated break after each block.

Testing: The testing was identical for all participants, and consisted of an AX discrimination task. Participants listened to pairs of words, and were asked to judge whether these were two different words or two repetitions of the same word. For critical pairs, these were endpoints of each continuum, either 'different' (100msec–205msec, 205msec–100msec) or 'same' (100msec–100msec, 205msec–205msec). For filler 'different' pairs, these were two words that differed by one segment: the contrasts were either in voicing ([ɕ]–[tʃ], [ɕ]–[ts], [b]–[p], [d]–[t], [g]–[k]), in place of articulation ([x]–[χ], [β]–[ʁ]), or in both ([r]–[ʔ]). The 'same' pairs were always physically identical. The TESTED WORDS were of one of two types: *trained* (i.e., heard in training) or *untrained* (i.e., heard for the first time in testing). There was a total of 384 word pairs, which included 6 repetitions of a testing block. One block consisted of 32 critical pairs (16 'same' and 16 'different') and 32 filler pairs (16 'same' and 16 'different'). The words in each pair were separated by an interstimulus interval of 750ms. As with training, stimulus order was randomized for each participant, and there was a self-terminated break after each block. Participants responded by pushing a button on a gamepad. They were instructed to respond according to their intuition based on what they learned during the training period, and were assured that there were no strictly right or wrong answers. The instructions included a short practice with English words, where 'different' words were minimal pairs (e.g., *mass* – *miss*), and 'same' words were repetitions of the same word pronounced with different intonations. Testing lasted about 20min.

2.2 Results

In Pająk & Levy (2011) we predicted that successful distributional training should lead to a difference between the bimodal and the unimodal conditions on critical length trials: bimodal training resulting in more 'different' responses (since the training should suggest that short and long consonants are contrastive in this language), while unimodal training leading to fewer 'different' responses (because the training provided no evidence that short and long consonants belong to different categories). Furthermore, we predicted that participants would generalize the

relevance of length from trained to untrained words (reflected in no difference in performance on trained and untrained items), and that this generalization would be bidirectional (i.e., from sonorants to obstruents, and vice versa). This prediction was confirmed for participants trained on the sonorant class (significant main effect of CONDITION for both trained and untrained words), but not the fricative class (no significant difference between *bimodal* and *unimodal* conditions). In what follows we examine these results in more detail for each segment.

Performance was at ceiling on ‘same’ trials (>95% ‘same’ responses for each TYPE, CONDITION, TRAINED SEGMENT CLASS, and TESTED WORDS type), so we only analyzed the responses from ‘different’ trials.² The results for critical items from ‘different’ trials split by segment are shown in Fig. 2 (for participants trained on the sonorant class), and Fig. 3 (for participants trained on the fricative class).

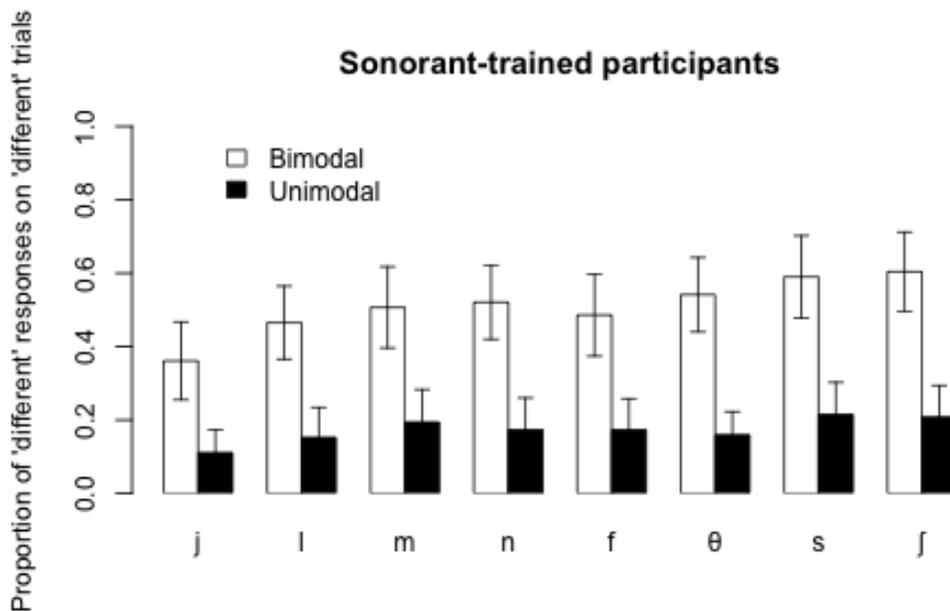


Figure 2: Performance by participants trained on the sonorant class. Trained segments: [j], [l], [m], [n]; untrained segments: [f], [θ], [s], [ʃ]. (Error bars are standard errors.)

In order to determine whether the overall learning and generalization effect holds for all of the tested segments, we performed t-tests for each segment comparing *bimodal* and *unimodal* conditions. For participants trained on the sonorant class, the t-tests revealed significant differences between the two conditions for all the segments ($ps < .05$)³, except for [j], where the effect was only marginal ($p = .054$). On the other hand, for participants trained on the fricatives, there was no significant difference between *bimodal* and *unimodal* condition for any of the

²The same results hold for an analysis using d-prime.

³[l]: $t(22) = 2.43; p < .05$; [m]: $t(22) = 2.21; p < .05$; [n]: $t(22) = 2.61; p < .05$; [f]: $t(22) = 2.23; p < .05$; [θ]: $t(22) = 3.21; p < .01$; [s]: $t(22) = 2.63; p < .05$; [ʃ]: $t(22) = 2.88; p < .01$.

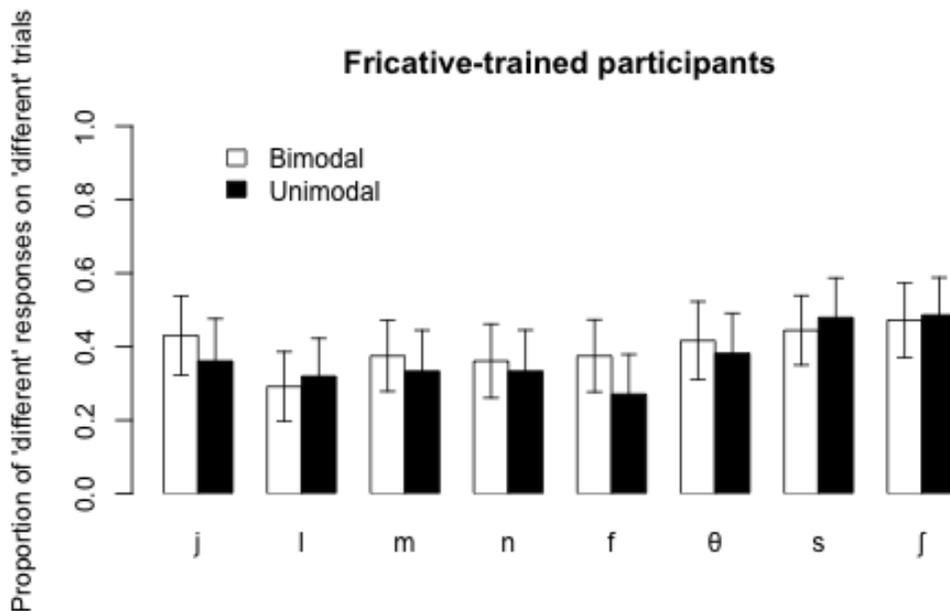


Figure 3: Performance by participants trained on the fricative class. Trained segments: [f], [θ], [s], [ʃ]; untrained segments: [j], [l], [m], [n]. (Error bars are standard errors.)

segments. Thus, the main result reported in Pająk & Levy (2011) was not driven by only a subset of segments, but rather it was true of all tested segments.

2.3 Discussion

As already reported in Pająk & Levy (2011), this study yielded two key results. First, monolingual speakers of English can be trained through purely distributional learning to recognize a phonological category distinction on a phonetic dimension (segmental length) which is never contrastive in their native language. Second, speakers generalized the relevance of length for sound categorization to a different set of consonants, fricatives.

In this paper we analyzed the data in more detail in order to see whether this main result holds for each individual short/long segment pair. The analysis revealed that this is indeed the case: the results look roughly the same for each segment pair. That is, for participants trained on the sonorant class, where distributional training was successful, participants responded ‘different’ more often in the bimodal than in the unimodal condition consistently for each of the short/long word pairs. Similarly, for participants trained on the fricative class, where distributional training failed, participants responded ‘different’ at the same rate in each condition and for each short/long word pair. This result provides further support for the claim that there is generalization of length across different segments in that learners are able to abstract the length cue away from individual segments and apply it to the same degree to all segment, whether familiar or novel.

An additional question that arose from this study concerns not generalization,

but learning from distributional information. Namely, we observed that learners were able to make better inferences about the data when they were trained on the sonorant segment class than when they were trained on the fricative segment class. This finding is surprising in light of Kawahara's (2007) result showing that intervocalic length contrasts are perceptually more salient for obstruents than sonorants. Furthermore, sonorant length contrasts are typologically less common than obstruent length contrasts, and the presence of a sonorant length contrast in a given language universally implies the existence of an obstruent length contrast in that language (Podesva 2002, Taylor 1985). Kawahara (2007) suggested that lesser perceptual saliency of sonorant length contrasts, perhaps also standing behind the typological data, might be due to the fact that sonorants are acoustically much more similar to vowels than obstruents are, which in turn makes it relatively harder for listeners to assess the length of an intervocalic sonorant than of an intervocalic obstruent.

On the other hand, the markedness of sonorant length contrasts (both perceptual and typological), might be the exact reason behind the obtained result: learning and aggressive generalization to fricative stimuli for sonorant-trained participants, but no learning for fricative-trained participants, even though the tested short/long fricative pairs were exactly the same for all participants. The possible explanation comes from the findings in child language acquisition and treatment of speech disorders indicating that training children on sounds of higher complexity (or, marked sounds) often results in the emergence of not just the complex sounds the children were trained on, but also other untrained sounds of lower complexity, while training on lower-complexity (or, unmarked) sounds alone is not always successful (Cataño *et al.* 2009, Dinnsen *et al.* 1990, Dinnsen 1992, Powell 1993, Tyler & Figurski 1994). Following the same logic, one could say that training on length contrasts for the marked sonorant class can lead to automatic learning of the same contrast for the unmarked fricatives, but the reverse training, on the unmarked fricatives, may lead to no learning whatsoever.

Furthermore, the relative difficulty of our participants to learn from the fricative stimuli could also be explained in terms of previous experience with long consonants from English. In English, length is not generally considered phonologically contrastive (i.e., there are no monomorphemic minimal pairs contrasted just by segmental length; e.g., Ladefoged 2001), and is instead used primarily as a prosodic cue (e.g., signaling stress or prosodic boundaries; Klatt 1976). Additionally, there is evidence that English-learning infants as young as 18 months old already process length contrasts differently from infants learning a language like Dutch or Japanese, where length is phonologically contrastive (Dietrich *et al.* 2007, Mugitani *et al.* 2008). However, long consonants are attested in English at morpheme junctures, at least for some speakers (e.g., *innate*, *vowelles*, or *big game*, Benus *et al.* 2003, Kaye 2005, Ladefoged 2001), and even some minimal pairs can be found (e.g., *unnamed* vs. *unaimed*, *wholly* vs. *holy*, *some more* vs. *some ore*).⁴ While not based on a quantitative analysis, a detailed review of possible examples of English long segments

⁴Kaye (2005) reports that the differences in durations between short and long consonants in English can indeed be significant. For example, for words spoken in isolation by 10 speakers, the *n* in *named* was on average 30msec long, while the *nn* in *unnamed* was 122msec long.

in Kaye (2005) suggests that long consonants in English are overwhelmingly sonorants (mostly using the prefixes ‘un-’ and ‘in-’) and—less frequently—stops, but only rarely fricatives (the relatively few examples include *dissatisfied*, *race sends* vs. *race ends*, ‘is she’ [ɪʃfi] or ‘is this’ [ɪzɪs] as a result of assimilation, or ‘bursts’ [bɔrɪs] when the cluster has been reduced; Bowen 1975, Kaye 2005, Spencer 1996, p. 225). The fact that this kind of asymmetry seems to exist between sonorants and fricatives in English—length used contrastively more often for sonorants than fricatives—might contribute to the relative difficulty of learning the length contrast for fricatives, based simply on people’s previous relative amount of exposure to length applied to both of these segment classes.

The final factor possibly contributing to the relative difficulty of making inferences from the fricative stimuli comes from looking at the durations of English consonants that correspond to the short/long consonants tested in this experiment. As mentioned in §2.1.3, English sonorants tend to be shorter than English fricatives (the relevant duration ranges are shown in Table 1), but, as you recall, we created the materials by using the same absolute durations for all segments (100msec-205msec). This meant that all the tokens from the sonorant continua were longer than their usual duration ranges in English, while for (most) fricatives these ranges partially overlapped. This might have led to difficulty in picking up on the distributional information for fricatives: participants may have heard the fricatives of around 200msec as unusually long, but still were inclined to interpret them as within reasonable English-like duration range, which consequently was not sufficient for bimodally-trained participants to infer contrastiveness of the length dimension.

| sonorants | fricatives |
|-----------|------------|
| j: 39-100 | θ: 46-90 |
| l: 42-85 | f: 56-119 |
| m: 50-89 | s: 61-126 |
| n: 38-83 | ʃ: 88-138 |

Table 1: The ranges of duration for English consonants that are equivalent to those used in experiment 1 (in msec; based on the phonetically annotated portion of the Switchboard corpus, as described in ‘The Switchboard Transcription Project’ report by Steven Greenberg, 1996.)

In order to further explore the reasons for the reported asymmetry between learning length contrasts for sonorants and for fricatives, we ran a follow-up experiment, reported in §3, in which we modified the fricative training stimuli so that the evidence for a length contrast would be more salient.

3 Experiment 2

In this experiment we exposed additional monolingual English speakers to evidence suggesting a novel phonological contrast between short and long fricatives in a new language. The experiment was identical to experiment 1 with the exception of part of the materials: here, we created longer continua for the fricative segment class (140-280msec) so that there was no overlap between the length of the crucial

stimuli segments and the usual duration ranges of these segments in English. This made for a more fair comparison between sonorant and fricative classes in terms of the saliency of the length cue, because durations were adjusted relative to natural durations of these two segment classes.

The goal of this experiment was to inform two crucial aspects of phonological acquisition: (1) learning from distributional information: specifically, investigating differences between learning from (comparable) distributional cues to length on distinct segment classes; and (2) generalization: looking at further evidence for cross-segment generalization of length contrasts. For (1), we predicted that a longer fricative continuum should be more conducive to making bimodally-trained participants infer a length contrast, as compared to experiment 1. Crucially, we were interested to see whether the learning would be as robust as for the sonorant class. For (2), based on results from experiment 1, we expected that the same response pattern should hold across all segments.

3.1 Method

3.1.1 Participants

24 undergraduate students at UC San Diego participated in the experiment for course credit. They were all monolingual speakers of English, in most cases with some limited high school and/or college exposure to Spanish or French. Crucially, none of them had any exposure to any language that uses length contrastively. All participants reported no history of speech or hearing problems, and none of them previously participated in experiment 1.

3.1.2 Materials

The materials were almost identical to those used in experiment 1. The only difference concerned the class of fricatives. We used the same recordings, but for each segment we created a new continuum, where durations of all fricative consonants ranged from 140msec (short) to 240msec (long), as shown in Fig. 4. The goal was to keep the same relative short-to-long duration ratios as in experiment 1, but increase the overall durations of the fricatives in order to make the length cue more salient. Crucially, this manipulation led to no overlap between usual English durations of fricatives and the stimuli, just as was the case for sonorants in experiment 1.

3.1.3 Procedure

The procedure was also almost identical to experiment 1. Unlike in experiment 1, however, where half of the participants were trained on the sonorant class and the other half on the fricative class, in this experiment all participants were trained on the fricative class with the newly created stimuli. The number of participants was equal to the number fricative-trained participants in experiment 1.

3.2 Results

As in experiment 1, we predicted that successful distributional training should lead to a difference between the bimodal and the unimodal conditions on critical length

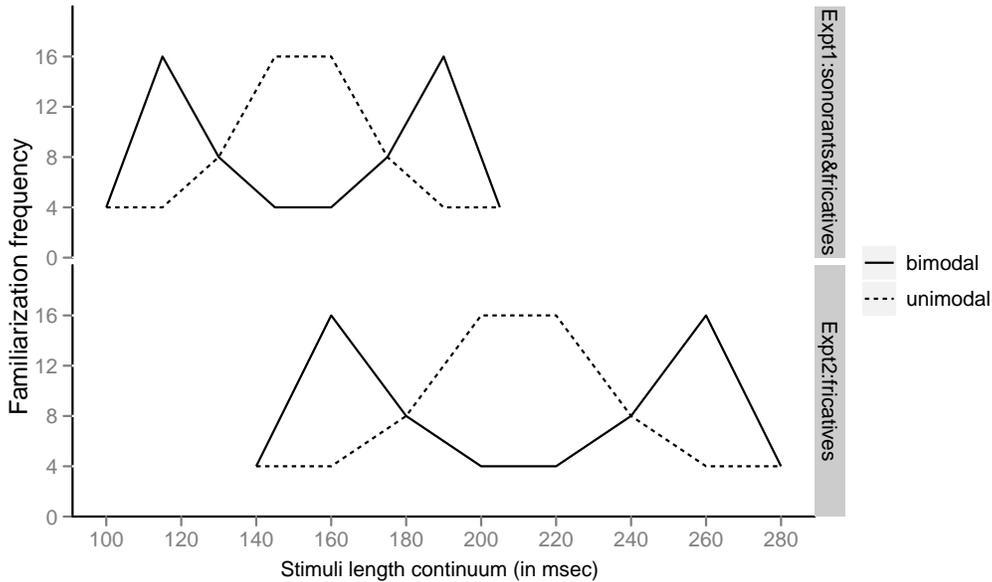


Figure 4: Critical training stimuli in experiments 1 and 2.

trials: bimodal training resulting in more ‘different’ responses, while unimodal training leading to fewer ‘different’ responses. Furthermore, in case of successful training, we predicted that participants would generalize the relevance of length to untrained words.

As in experiment 1, we only analyzed responses from ‘different’ trials because performance was at ceiling on ‘same’ trials (>97% ‘same’ responses). The results for critical items from ‘different’ trials are shown in Fig. 5.

First, we analyzed the results using a repeated measures ANOVA by participants with the within-participants factor TESTED WORDS (*trained* or *untrained*) and the between-participant factor CONDITION (*bimodal* or *unimodal*). While numerically there were more ‘different’ responses in the bimodal than in the unimodal condition (30% vs. 19%), there was no significant main effect of CONDITION ($F = 1.01; p = .32$), suggesting that bimodal vs. unimodal training did not significantly alter participants’ responses. The numerical bimodal vs. unimodal response pattern was similar for each segment class (fricatives: 36% vs. 24%; sonorants: 25% vs. 15%), and there was no significant interaction between CONDITION and TESTED WORDS ($F < 1$).

We also conducted a set of analyses focusing on segments rather than participants. No individual *t*-test for a specific segment comparing bimodal and unimodal conditions revealed a significant difference; the closest candidates were [l] and [θ] (both $ts(22) = 1.50; p = .14$). However, a repeated-measures by-segments ANOVA with the between-segment factor TESTED WORDS and the within-segment factor CONDITION revealed a highly significant effect of CONDITION ($F(1, 6) = 31.19; p < .01$). The reason for the discrepancy between by-participants and by-segments analyses is that participants in this study are highly variable; some par-

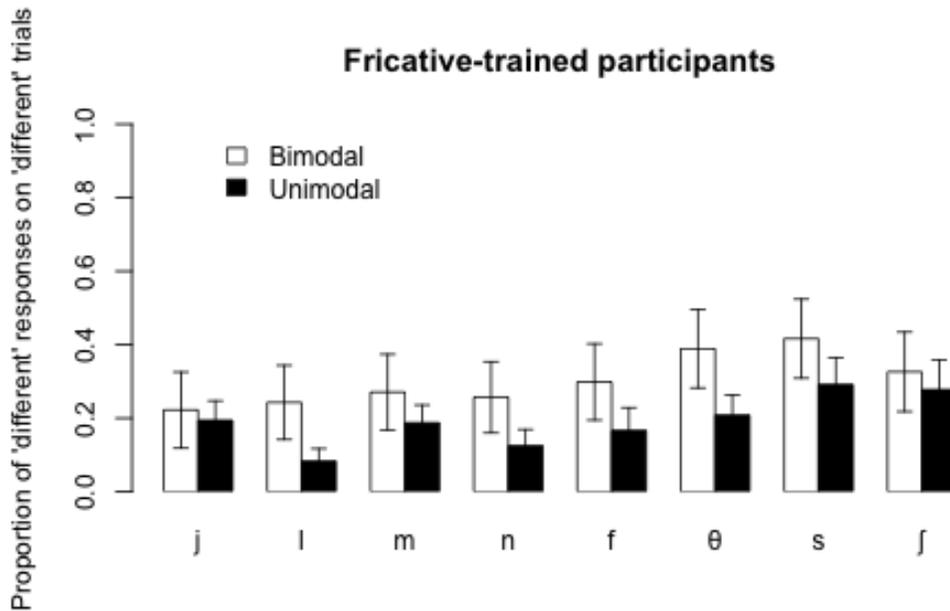


Figure 5: Performance by participants trained on the fricative class. Trained segments: [f], [θ], [s], [ʃ]; untrained segments: [j], [l], [m], [n]. (Error bars are standard errors.)

ticipants tend to classify “different” tokens as different close to 100% of the time, others close to 0% of the time. Because the CONDITION manipulation is between-participants, we are unable to factor out this large variation (as would be possible in a within-participants study), so that even though there is a clear numeric pattern suggesting that learning really was different in the unimodal versus bimodal conditions, we cannot reliably distinguish this pattern from participant-level noise in experiment 2.⁵

In order to compare experiments 1 and 2, we submitted the results from both experiments together to repeated measures ANOVAs with the between-participants factors EXPERIMENT (1 or 2) and CONDITION (*bimodal* or *unimodal*). First, we compared only fricative-trained participants from experiments 1 and 2. There was no significant interaction between EXPERIMENT and CONDITION ($F < 1$), suggesting that the two experiments did not differ significantly in terms of the response pattern in bimodal and unimodal conditions. Second, we compared sonorant-trained participants from experiment 1 to fricative-trained participants from experiment 2. Again, there was no significant interaction between EXPERIMENT and CONDITION, but the effect in this case was much closer to being marginal ($F(1, 44) = 1.81; p = .18$) than in the previous comparison. Once again, large variation in behavior participant across participants prevents us from drawing fine-grained inferences regarding differences between experiment 2 and both sonorant and fricative training from experiment 1, though our results suggest that experiment 2 fell somewhere in between sonorant-trained and fricative-trained results from experiment 1, perhaps closer to

⁵Mixed logit model analysis led to similar conclusions.

the pattern of sonorant-trained participants.

Overall, the results suggest that while the modified stimuli may have enhanced distributional learning from the fricative stimuli, the effect was not nearly as robust as learning from the sonorant stimuli in experiment 1.

3.3 Discussion

In this study we focused on two aspects of phonological acquisition: (1) learning from distributional information for different segment classes, and (2) generalization across segments. For the latter, just as in experiment 1, we observed evidence for generalization: there was a numerical difference between bimodal and unimodal conditions, and the numerical pattern was similar for both trained and untrained words, suggesting that the inferences made by participants for the fricatives during training generalized to sonorants.

For the learning aspect, we expected that the new fricative stimuli would lead to more effective distributional learning than in experiment 1, since the relative durations of fricatives were now comparable to sonorant durations. However, while there was a consistent numerical trend holding for each of the tested segments with more ‘different’ responses in the bimodal than in the unimodal condition, this difference was not significant in the by-participant analysis (but significant in the by-items analysis) due to large between-participant variation. Therefore, although experiment 2 does show some evidence for learning from fricatives, it seems that the effect is attenuated when compared with learning from sonorants. This suggests that the absolute durations of our stimuli, as relative to their corresponding segment durations in English, cannot by themselves explain the asymmetry in distributional learning between sonorants and fricatives. Instead, other factors—such as differential previous exposure to long sonorants and fricatives in English or relative ‘lesser complexity’ of long fricatives as compared to long sonorants—must also be in play causing learning of fricative length contrasts more difficult than learning similar contrasts for sonorant consonants. More research with other segments (e.g., stops, voiced fricatives, other sonorants) might provide better understanding of the reasons behind successful vs. failed learning from distributional information. Whatever these reasons are, however, they must be powerful enough to override the intrinsic higher perceptual saliency of intervocalic short/long fricatives as compared to short/long sonorants (Kawahara 2007).

4 Conclusion

The main question we asked in this paper was whether subphonemic phonetic properties, such as segmental length, are represented psychologically as independent from individual segments, or whether their representations are more constrained and depend on the segments’ acoustic-phonetic (or gestural) properties. We investigated this question by looking at how novel length contrasts are learned by adult participants, and whether and how they are generalized to additional segment classes. We found evidence that the property of contrastive length can be learned from distributional information alone—albeit the efficacy of learning may depend on the exact segment class—and can be immediately generalized across segment

classes that are acoustically very distinct (sonorant consonants and voiceless fricatives). This result suggests that length is abstracted away from individual segments, and has a shared representation for segments that have very different phonetic implementations.

5 Acknowledgments

For their helpful feedback the authors thank Eric Baković, Klinton Bicknell, Sarah Creel, Vic Ferreira, Tamar Gollan, Sharon Rose, and the audience of CLS 47. K. Michael Brooks, Annabelle Cadang, Rafi Feliciano, and Nicole Pyon helped with data collection. This research was supported by NIH Training Grant T32-DC000041 from the Center for Research in Language at UC San Diego to the first author.

References

- Benus, Stefan, Iris Smorodinsky, & Adamantios Gafos. 2003. Gestural coordination and the distribution of English ‘geminate’. In *Proceedings of the 27th Annual Penn Linguistics Colloquium*, 33–46, Philadelphia, PA. University of Pennsylvania.
- Bowen, J. Donald. 1975. *Patterns of English pronunciation*. Rowley, MA: Newbury House.
- Browman, Catherine P., & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6.201–251.
- Cataño, Lorena, Jessica A. Barlow, & María Irene Moyna. 2009. A retrospective study of phonetic inventory complexity in acquisition of Spanish: implications for phonological universals. *Clinical Linguistics & Phonetics* 23.446–472.
- Chomsky, Noam, & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
- Davis, Stuart. 1999. On the representation of initial geminates. *Phonology* 16.93–104.
- Dietrich, Christiane, Daniel Swingley, & Janet F. Werker. 2007. Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences* 104.16027–16031.
- Dinnsen, D. A. 1992. Variation in developing and fully developed phonologies. In *Phonological development: models, research, implications*, ed. by Charles Ferguson, Lise Menn, & Carol Stoel-Gammon, 191–210. Timonium, MD: York Press.
- , S. B. Chin, M. Elbert, & T. W. Powell. 1990. Some constraints on functionally disordered phonologies: phonetic inventories and phonotactics. *Journal of Speech and Hearing Research* 33.28–37.
- Fromkin, Victoria A. 1973. The non-anomalous nature of anomalous utterances. In *Speech errors as linguistic evidence*, ed. by Victoria A. Fromkin, 215–242. The Hague/Paris: Mouton.
- Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20.253–306.
- Hume, Elizabeth, Jennifer Muller, & Aone van Engelenhoven. 1997. Non-moraic geminates in Leti. *Phonology* 14.371–402.
- Hyman, Larry. 1985. *A theory of phonological weight*. Dordrecht: Foris.

- Kawahara, Shigeto. 2007. Sonorancy and geminacy. In *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality III*, 145–186. Amherst: GLSA.
- Kaye, Alan. 2005. Gemination in English. *English Today* 21.43–55.
- Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59.1208–1221.
- Ladefoged, Peter. 2001. *A course in phonetics*. Fort Worth, TX: Harcourt.
- , & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford, UK; Cambridge, MA: Blackwell.
- Levin, Juliette, 1985. *A metrical theory of syllabicity*. Massachusetts Institute of Technology dissertation.
- Maye, Jessica, & LouAnn Gerken. 2000. Learning phonemes without minimal pairs. In *Proceedings of the 24th Annual Boston University Conference on Language Development*, ed. by S. Catherine Howell, Sarah A. Fish, & Thea Keith-Lucas, 522–533. Somerville, MA: Cascadilla Press.
- , & ———. 2001. Learning phonemes: how far can the input take us? In *Proceedings of the 25th Annual Boston University Conference on Language Development*, ed. by A. H-J. Do, L. Domínguez, & A. Johansen, 480–490. Somerville, MA: Cascadilla Press.
- , Janet F. Werker, & LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82.B101–B111.
- McClaskey, Cynthia L., David B. Pisoni, & Thomas D. Carrell. 1983. Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics* 34.323–330.
- Mugitani, Ryoko, Ferran Pons, Laurel Fais, Janet F. Werker, & Shigeaki Amano. 2008. Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology* 45.236–247.
- Muller, Jennifer, 2001. *The phonology and phonetics of word-initial geminates*. Ohio State University dissertation.
- Pająk, Bożena, & Roger Levy. 2011. Phonological generalization from distributional evidence. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, ed. by L. Carlson, C. Hölscher, & T. Shipley, 2673–2678. Austin, TX: Cognitive Science Society.
- Pierrehumbert, Janet. 2000. The phonetic grounding of phonology. *Bulletin de la Communication Parlee* 5.7–23.
- Podesva, Robert J. 2002. Segmental constraints on geminates and their implications for typology. Handout from the 76th LSA Annual Meeting.
- Powell, T. W. 1993. Phonetic inventory constraints in young children: factors affecting acquisition patterns during treatment. *Clinical Linguistics & Phonetics* 7.45–57.
- Selkirk, Elizabeth. 1991. On the inalterability of geminates. In *Certamen Phonologicum II: Papers from the Second Annual Cortona Phonology Meeting*, ed. by Pier Marco Bertinetto, Michael Kenstowicz, & Michele Loporcaro, 187–210. Turin: Rosenberg & Sellier.
- Spencer, Andrew. 1996. *Phonology*. Oxford: Blackwell.
- Taylor, Martha. 1985. Some patterns of geminate consonants. *The University of Chicago Working Papers in Linguistics* 1.120–129.
- Tranel, Bernard. 1991. CVC light syllables, geminates and Moraic Theory. *Phonology* 8.113–136.
- Tyler, A. A., & G. R. Figurski. 1994. Phonetic inventory changes after treating distinctions along an implicational hierarchy. *Clinical Linguistics & Phonetics* 8.91–108.