DE GRUYTER MOUTON
DOI 10.1515/lp-2013-0010 — Laboratory Phonology 2013; 4(2): 259–285

Alejandrina Cristia, Jeff Mielke, Robert Daland and Sharon Peperkamp

Similarity in the generalization of implicitly learned sound patterns

Abstract: It is likely that generalization of implicitly learned sound patterns to novel words and sounds is structured by a similarity metric, but how may this metric best be captured? We report on an experiment where participants were exposed to an artificial phonology, and frequency ratings were used to probe implicit abstraction of onset statistics. Non-words bearing an onset that was presented during initial exposure were subsequently rated most frequent, indicating that participants generalized onset statistics to new non-words. Participants also rated non-words with untrained onsets as somewhat frequent, indicating generalization to onsets that had not been used during the exposure phase. While generalization could be accounted for in terms of featural distance, it was insensitive to natural class structure. Generalization to untrained sounds was predicted better by models requiring prior linguistic knowledge (either traditional distinctive features or articulatory phonetic information) than by a model based on a linguistically naïve measure of acoustic similarity.

Alejandrina Cristia: Neurobiology of Language, Max Planck Institute. Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS, ENS-DEC, CNRS, Paris. E-mail: alecristia@gmail.com
Jeff Mielke: Department of Linguistics, University of Ottawa. E-mail: jeff.mielke@gmail.com
Robert Daland: Department of Linguistics, University of California at Los Angeles. E-mail: r.daland@gmail.com
Sharon Peperkamp: Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS, ENS-DEC, CNRS, Paris. E-mail: sharon.peperkamp@ens.fr

1 Introduction

The ability to make generalizations is a fundamental aspect of cognition and a crucial property of human language use. More specifically, some type of generalization may be implicated in cases of language change whereby a sound pattern that applied to one set of sounds at one point in the history of a given language (the *original set*) is extended to additional sounds at a subsequent point in history (the *extension set*). It is unclear to what extent cognitive constructs, present

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

within individual speakers' mental phonology, shape such generalization processes. Indeed, historical changes could reflect cognitive factors (Kiparsky 2006; Labov 2011), but they are also affected by a myriad of other variables (phonetic pressures, social changes, chance, etc.; see, e.g., Ohala 1983; Labov 1994, 2001; Blevins 2004), which makes it difficult to evaluate the extent to which patterns observable in languages reflect cognitive constructs.

Laboratory learning of artificial grammars has begun to provide evidence on cognitive biases shaping phonological acquisition (see Moreton and Pater 2012 for a recent summary). In that line of work, participants are first exposed to some kind of sound pattern, designed by the experimenter on the basis of current phonological knowledge. At a second stage, participants are tested to determine how/whether they learned that pattern. Importantly, generalization can also be gauged, by withholding critical cases from the initial exposure, and presenting them at test (Wilson 2006; Finley and Badecker 2009; Becker, Nevins, and Levine 2012). In the present paper, we report on an experiment that utilized this paradigm in order to shed light on a central question within phonological generalizations, regarding the nature of similarity between the original and extension sets.

We illustrate the importance of this question with an example of generalization taken from a natural language sound change: /o/-lowering in Northeastern Swiss German-speaking communities around Schaffhausen (Keel 1982; Janda and Joseph 2003). In the city of Schaffhausen, lowering is triggered by /r m n ŋ/, whereas in some nearby villages it is triggered by /r t d ʦ s z ʃ ʒ/. A plausible explanation for this state of affairs is that the original phonetic trigger of lowering was /r/ (since only /r/ triggers lowering in all local dialects), and that different communities generalized the sound change on the basis of differing aspects of the original trigger (see Janda and Joseph 2003; Mielke 2008, section 5.2.2, for a more in-depth discussion). This case of generalization illustrates two important points. First, generalization seems to be constrained by similarity, since the pattern did not transfer from /r/ to, say, /ts/, without also applying to an array of consonants that are more similar to /r/ than /ts/ is. Second, similarity must admit multiple dimensions of comparison, since the same original set led to two different generalizations in different populations. But, in both instances, what is 'similarity'?

Although no previous artificial grammar work has evaluated the question of similarity specifically, one study investigated the kinds of representations that allow generalization of newly-learned patterns. Bernard, Onishi, and Seidl (submitted) explored whether listeners encode sound patterns affecting dimensions that are phonemic, as compared to allophonic, in the learners' native language. In their experiment, Quebec French and American English listeners were presented with non-words where nasal vowels were consistently followed by frica-

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

tives and oral vowels by stops (or vice versa). Most of the non-words contained three specific vowel categories; a fourth vowel was presented rarely and in co-occurrence with both stops and fricatives. In the test phase, participants heard the rare items once more and were asked to rate their frequency during the initial exposure phase. Although in these rare items the vowels had occurred equally frequently with stops and fricatives, French listeners rated non-words that followed the general pattern (nasal+fricative, oral+stop) as having been presented more frequently than non-words that did not follow that regularity. In contrast, English listeners did not exhibit this behavior, suggesting that generalization occurs more readily along phonologically important dimensions. In other words, phonological knowledge from the participants' native language shapes generalization evidenced in the laboratory (see also Pajak and Levy 2011 for potentially relevant evidence); or, put differently, similarity calculations are based at least in part on one's native phonological grammar.

The present study investigates more closely what kinds of similarity metrics operate in phonological generalization. Previous artificial grammar research has assumed that the relevant metric involves phonological features (Finley and Badecker 2009; Finley 2011, submitted). In traditional generative phonology, distinctive features are fundamental building blocks of phonological rules and representations (Chomsky and Halle 1968; Kenstowicz and Kisseberth 1979). This is motivated in part by the observation that phonological patterns often involve familiar natural classes of phonetically similar sounds, which are analyzed as sharing one or more feature values (Jakobson, Fant, and Halle 1963; Chomsky and Halle 1968; Mielke 2008). For example, the English voiceless obstruents /p t k ʧ f θ s ʃ/ form a natural class, because they share phonetic properties such as being produced with a substantial oral constriction, with increased intraoral pressure, and without vocal fold vibration. Accordingly, they are analyzed as being [–vocalic], [–sonorant], and [–voiced]. English voiceless obstruents also display common behaviors, such as triggering voicing assimilation in a following genitive or possessive affix.

Because traditional natural classes are based on features, and phonological features are typically given articulatory and/or acoustic definitions, phonological, phonetic, and physical (i.e., independent from language-specific perception) notions of similarity could all explain generalizations that are observed in natural languages. Using an artificial grammar allows us to tease apart some of these factors. To investigate how similarity is organized in the adult perceiver, we consider four hypotheses:

– *Natural Class Generalization*. Listeners generalize within traditional natural classes.
– *Featural Distance Generalization*. Listeners generalize to featural neighbors.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

- *Phonetic Distance Generalization*. Listeners generalize to phonetically similar items, with phonetic similarity depending on language-specific experience (for example, using articulatory information).
- *Acoustic Distance Generalization*. Listeners generalize to acoustically similar items (using a naïve measure of raw acoustic similarity that does not depend on language experience, and could potentially be found even in non-human animals).

In the following subsections we motivate each hypothesis and describe its specific predictions.

1.1 Natural class generalization

A core empirical observation of phonological theory is that speech sounds pattern together in structured ways that can be related to shared phonetic properties and/or shared phonological behaviors (Chomsky and Halle 1968; Frisch 1996; Mielke 2008). Quantitative typological studies have shown that the great majority of sound patterns involve featurally and phonetically natural classes as the targets and/or triggers. Based on a survey of sound patterns in several hundred languages (Mielke 2008), Mielke, Magloughlin, and Hume (2011) report that typical distinctive feature theories capture 70–73% of phonologically active classes. One way of accounting for this distribution is to posit a bias toward generalizing along natural class lines – this is the Natural Class Generalization hypothesis.

As it stands, the Natural Class Generalization hypothesis is underspecified (at least from the learner's point of view), owing to the fact that natural classes overlap. In particular, some natural classes stand in a subset-superset relationship, e.g., voiceless oral stops are a subset of oral stops, which are a subset of obstruents. When observing a pattern involving /p t k/, the learner could consider any of these three levels of abstraction to encode it (and several more). The classical linguistic solution in such cases is to invoke the Subset Principle (e.g., Hale and Reiss 2003), which states that learners make the most restrictive generalization that is compatible with the data and formalizable within the grammar. In probabilistic terms, learning may be viewed as allocating probability over competing hypotheses rather than selecting a single one. In this framework, a mild generalization of the Subset Principle emerges as the statistically optimal outcome of learning under uncertainty (Hayes and Wilson 2008): the learner may assign some probability to less restrictive hypotheses, but crucially should assign more probability to the most restrictive hypotheses. Thus, Natural Class General-

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

ization predicts that if a learner were exposed to a sound pattern involving /d g v z ʒ/, they would be more likely to generalize the pattern to /b/ than to /k/, since /b/ is a member of the smallest natural class that contains all the participating sounds (voiced obstruents), whereas /k/ is not. It is important to note that natural classes could pattern together often in extant language patterns because the members of the class are likely to be affected by similar phonetic pressures. Therefore, generalization may not be *necessary* to explain the prevalence of natural classes in phonological patterns found cross-linguistically (see also Section 5.2). Nonetheless, it remains plausible that natural classes structure generalization if and when this process occurs.

1.2 Featural distance generalization

In certain phonological theories, distinctive features are the representational primitives of phonological processes, such that phonological rules/constraints refer to features, rather than to natural classes (e.g., Hall 2001). Featural Distance Generalization is the hypothesis that generalization of a sound pattern will be strongest for sounds that are featurally similar to one or more sounds already in the sound pattern. Thus, generalization need not respect natural class boundaries. This may be illustrated with the same example from before (exposure to a pattern involving /d g v z ʒ/). According to the Featural Distance Generalization, the learner should assign similar probabilities to /b/ and /k/ participating in that pattern: both sounds are one feature away from one or more sounds experienced in that pattern (e.g., /b/ differs from /g/ in place of articulation, /k/ differs from /g/ in voicing). (Please note that to distinguish voiced and voiceless stops and fricatives at three places of articulation, we use the traditional features [voice] and [continuant] and a ternary place feature.) In contrast, lower probability would be assigned to /p/, which is two features away (place and voicing).

1.3 Phonetic distance generalization

Generalization could be dependent on a language-specific phonetic distance measure; for instance, based on articulatory experience or on sophisticated acoustic similarity metrics that allocate attention to linguistically relevant information only. A recent study has gathered articulatory and acoustic measurements from 12 English obstruents (among other sounds; Mielke 2012; see Section 2.1.4 for further details on a similar set of measures). Figure 1 shows the distance among these obstruents as a function of a few articulatory and acoustic

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

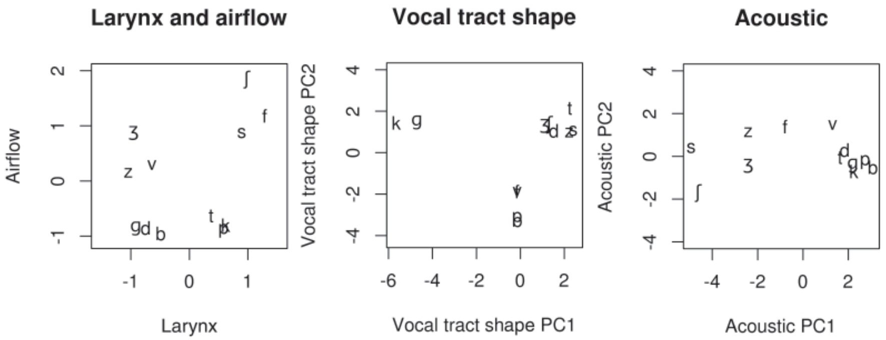

Fig. 1: Phonetic similarity between 12 obstruents, according to a variety of articulatory and acoustic measures (Mielke 2012). See the main text for details.

measurements. Within our recurrent example, a similarity matrix based on vocal tract shape would predict equal generalization to /p/ and /b/, since they are just as distant from the voiced obstruents associated with the exposure set (/d g v z ʒ/).

1.4 Acoustic distance generalization

Sound pattern generalization may be conditioned by physical similarity between the original and extension items. The rightmost panel in Figure 1 represents a linguistically uninformed acoustic similarity measure (based on spectral shape; see Section 2.1.4 for further details on a similar set of measures). A learner relying on this metric to guide her generalization judgments would respond similarly to /b/, /p/, and /k/, all of which are at similar distances from the exposure set /d g v z ʒ/. We call this measure ‘uninformed’ because no previous language experience is necessary, and it can be calculated directly from information that is physically present in the speech the learner hears.

1.5 Summary of motivation

Since generalization of sound patterns can be studied *in vitro* using artificial grammars, we adopt this paradigm to test predictions from four hypotheses. The first states that natural classes themselves structure and limit generalization (Natural Class Generalization); the second is that the likelihood of generalization declines as a function of discrete featural distances (Featural Distance Gen-

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

eralization); the third is that the likelihood of generalization is dependent on a linguistically-informed measure of dissimilarity across the relevant items (Phonetic Distance Generalization); finally, the fourth is that raw acoustic similarity can account for observed patterns (Acoustic Distance Generalization).

2 Experiment

A subjective frequency task was used to gauge listeners' encoding and implicit extension of static sound patterns (as in Bernard et al. submitted). In the exposure phase, participants heard non-words whose onsets are drawn from a set of obstruents sharing voicing, e.g., /d g v z ʒ/, and performed an irrelevant task (i.e., rated their overall well-formedness). During the pre-test phase, which was a continuation of the exposure phase with no break and the same instructions, participants were exposed to each test item once. During the test phase proper, participants were again presented with the test items, but this time they were asked to give their subjective impression of how frequently they had heard the item before. There were four types of test items, corresponding to how the onset of the test word differed from the onsets of the exposure set. The exposure and test stimuli were counterbalanced across participants, as shown in Figure 2 and explained in more detail below. Table 1 illustrates the relationship between training and test stimuli using the onsets that were presented to participants in Exposure Condition 1.

Participants in the example condition shown in Table 1 were exposed to nonwords whose initial onset was drawn from the list of all voiced obstruents *except* /b/, as shown on the top panel. At test, they were presented with new non-words bearing one of the 4 onsets shown on the bottom panel. These onsets differed on whether they belonged to the same natural class as the trained ones in terms of the minimum featural distance from the trained set; in the minimum distance they spanned along several articulatory dimensions (representing the linguistically informed phonetic dimensions), and in the minimum raw spectral distance (representing the uninformed phonetic dimensions).

The first type of test item is represented in Table 1 by /g/, and it will be referred to as Exposure. This first type is constituted by novel non-words whose onset was among the set presented during the initial exposure phase. These items allow us to ascertain that the paradigm was able to capture differences in subjective frequency ratings that were due to encoding of the onset of the test items. If participants were indeed affected by the frequency of onsets in their responses, they should assign the highest subjective frequency ratings to Exposure items.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

	Voiced						Voiceless					
Exposure	Stops			Fricatives			Stops			Fricatives		
condition	b	d	g	v	z	ʒ	p	t	k	f	s	ʃ
1	W	x	E	x	x	x	F		N			
2	x	W	x	x	E	x		F			N	
3	x	x	W	x	x	E			F			N
4	x	E	x	W	x	x		N		F		
5	x	x	x	E	W	x				N	F	
6	E	x	x	x	x	W	N					F
7	F		N				W	x	E	x	x	x
8		F			N		x	W	x	x	E	x
9			F			N	x	x	W	x	x	E
10		N		F			x	E	x	W	x	x
11				N	F		x	x	x	E	W	x
12	N					F	E	x	x	x	x	W

Fig. 2: Counterbalancing orders used in Experiment 1. Each row is an exposure condition (there were two participants per condition); 12 conditions were generated in order to fully counterbalance the assignment of segments to types across participants. This assignment is represented with the following codes: x indicates that the onset in that column was part of the exposure set only; E that the onset was used in both exposure and test (so it was an Exposure *onset*, but recall that different non-words were used in the two phases). All others were presented only at test: W is a Within onset (within the narrowest natural class to which all the members of the exposure set belonged); N is an onset that is Near (different voicing than every member of the exposure set, but sharing manner and place with an exposure onset), and F the Far onset (differing in both voicing and place).

The second test item type is called Within, represented in this example by /b/. We will refer to /b/ as the Within segment, because it is within the Subset class (the narrowest natural class that contains all the exposure onsets). This Subset is predicted to bound generalization by the Natural Class hypothesis. A mixed model with regressors for each type of trial was used to test the Natural Class prediction, by which Within items should get higher familiarity ratings than the other untrained types.

Like the Within onset, the Near onset /k/ differs in one feature from an Exposure onset; but unlike the Within onset, the Near onset does not belong to the Subset class. That is, /k/ differs only in one feature from /g/, but it is voiceless. Predictions made from the Featural distance hypothesis can be tested with these items, which should receive ratings comparable to those of the Within items, but higher than items whose onset is two features away from the set used in exposure.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

Table 1: Example of the mapping of individual obstruents to each test type, and value for the regressors of interest, in one counterbalancing condition. For the purposes of illustration in this table, both articulatory and acoustic measures are taken from Mielke (2012) (the units are standard deviations along a first principal component). For more details, see the main text and the Procedure section.

EXPOSURE PHASE							
Onsets /d g v z ʒ/							
Natural class voiced obstruents (/b/ held out)							
TEST PHASE							
Name	**Onset**	**Narrowest natural class?**	**Minimum featural distance?**	**Minimum vocal tract distance?**	**Minimul larynx distance?**	**Minimum airflow distance?**	**Minimum acoustic distance?**
Exposure	/g/	yes	0	0	0	0	0
Within	/b/	yes	1	1.68	0.15	0.09	0.97
Near	/k/	no	1	0.96	1.26	0.06	0.7
Far	/p/	no	2	1.63	1.23	0.03	0.82

The baseline in ratings is thus provided by items whose onset is at least two features from any of the onsets used during initial exposure, called Far onsets. In the example in Table 1, the Far onset (/p/) differs from one of the exposure onsets (e.g., /g/) in both voicing and place.

Additionally, the distance between each of the test onsets and those in the training set can be measured phonetically along a number of dimensions. In Table 1, we show distances derived from the articulatory and acoustic measurements plotted in Figure 1, which are calculated from data reported in Mielke (2012). Naturally, the Exposure onset has a distance of zero, because it was in the training set. For the distances derived from vocal tract shape, Near /k/ is fairly close to /g/, which is part of the training set, while Within /b/ and Far /p/ are in fact further from any exposure onset than Near /k/ was to /g/. Thus, a learner relying on this dimension in her calculation of similarity should be most inclined to a extend a sound pattern observed on /d g v z ʒ/ to Near /k/, and just as likely to extend it to Within /b/ and Far /p/. For the distance measure derived from larynx activity, only /b/ is close to the (voiced) consonants in the training set. Not surprisingly, the distance measure derived from airflow does not help distinguish among the test onsets.

Additionally, we illustrate the different predictions made from an uninformed phonetic measurement, in this case a raw spectral distance. Along this

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

dimension, the Near onset /k/ is closest to the trained onset /g/, and the Far onset /p/ happens to be somewhat closer than the Within onset /b/. A learner relying on this measure should extend the pattern to Near /k/ more than Within /b/ or Far /p/ onsets.

Each of these distance measurements, introduced in more detail below (Section 2.1.4), could potentially predict participants' generalization patterns. Importantly, whereas predictions made from the Natural Class and Featural Distance hypotheses are mutually incompatible, either of them is potentially compatible with some amount of generalization being related to the other distances. Therefore, we analyze the predictive value of the phonetic measures in a separate section, where linear mixed models are fitted using a variety of phonetic measures as regressors. For each of these models, we computed proportion of variance explained (a measure of predictive value) and model fit (a measure of predictive value taking into account model complexity). Further details are provided in Section 2.1.5.

2.1 Methods

2.1.1 Participants

Twenty-four native, monolingual French speakers were tested. Participants had volunteered for perceptual studies after hearing about the Laboratoire de Sciences Cognitives et Psycholinguistique through fliers, ads on websites, and word of mouth. They were paid 5€ for their participation. All procedures were performed in compliance with relevant laws and institutional guidelines.

2.1.2 Procedure

Participants were tested one at a time in a sound-attenuated booth. They sat in front of a computer, wore headphones, and responded through a buttonbox, whose buttons were labeled 1 through 5. They were told that: (1) they would make judgments on non-words, and we would use these judgments in future studies focusing on how children learn language; (2) they would hear one item at a time over their headphones, and they would have to answer questions that would show on the screen; (3) the first question would focus on well-formedness, but there would be other questions later on. They were not explicitly told what the other questions were, nor how many sections the study had. In the initial instruc-

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

tions, they were asked to respond quickly. Thus, no reference was made to artificial grammars.

The experiment consisted of three phases. During the first phase (the exposure), *training* items were presented in a random order, and participants had to judge each item's well-formedness (yes/no answer). In the second phase (the pre-test), the *test* items were presented in a random order with the same well-formedness instructions. This phase followed the previous one without interruption; from the participants' point of view, the first two phases were one and the same. In order to prevent participants from attempting to memorize the individual items, or to repeat them overtly or covertly during the well-formedness portion, a reminder written in red was displayed each time they took longer than 500 ms to answer. Finally, during the third phase (the test proper), the test items were presented twice again, interspersed with the training items, and participants were asked to rate how frequently they had heard each item before, on a scale from 1 (= very seldomly) to 5 (= very frequently). No feedback (on the response or the response time) was given at this stage. Thus, two judgments were collected for each test item. The presence of the training items in this test phase served to help maintain the statistical patterns heard in the initial exposure; responses to these items were not analyzed. Note also that all test items were presented exactly once before these frequency instructions were given, so that variation in the ratings cannot possibly reflect variation in the item's true frequency. Finally, item effects could not reflect their frequency in French (since no item was a real French word), and, due to counterbalancing, an effect of trial type could not respond to potential differences in frequency of the item's diphones, triphones, etc. The experiment lasted about 30 minutes.

It may be relevant to ponder a moment how participants may have approached the task, and particularly the test phase, when they were asked to provide subjective frequency scores. Naturally, any difference in ratings across items is factually incorrect, and a 'perfect listener' performing this task should rate all test items as 'very infrequent'. We nonetheless expected differences in ratings because human memory is susceptible to false alarms and intrusions (Deese 1959). This property of human cognition to overestimate the incidence of certain past events has been exploited in several lines of psychological research. For example, 'false recall' has been used to shed light on the semantic and phonological structure of the lexicon (McDermott 1996), and to assess whether shape and color are encoded together when they occur in a list of visual objects (Deese 1998). Tasks relying on factually incorrect responses have already been used to tap naïve participants' intuitions implicitly in artificial grammar learning studies applied to phonology (Wilson 2003). Previous readers of this manuscript have proposed a range of alternative interpretations for how participants could approach the task.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

However, no alternative interpretation offered to us to date has been able to explain why participants would consistently rate Exposure and Far items differently. Moreover, the inclusion of these two types facilitates the interpretation of the 'intermediate' Near and Within items. The only way in which these can be intermediate in terms of ratings is if raters have generalized the subjective frequency of the consonants that were part of the training set.[1]

2.1.3 Stimuli

Stimuli were designed so as to maximize the variability participants would experience (which facilitates abstraction; Pierrehumbert, Beckman, and Ladd 2001) while still maintaining control of possible confounds, such as the frequency of individual sounds and diphones. Items were of the form CV_1NV_2, where C was an obstruent in /p t k f s ʃ b d g v z ʒ/, V_1 in /a e i o u/, N in /m n l/, and V_2 in /a i u/. Every combination of 12 onsets and the 45 'frames' ($5\ V_1 \times 3\ N \times 3\ V_2$) (for a total of 540 items) was generated. The 540 items were distributed into 5 arbitrary lists (which had 9 items per onset), balanced in diphone frequency.

The stimuli were designed to eventually allow for a cross-linguistic comparison between English and French, since obstruent voicing classes are different phonologically (e.g., Jansen 2007) and phonetically (e.g., Keating 1984) in those two languages. In this paper we report data only from French participants. A sonorant coda (/l/ for words where *N* was /m/, /m/ otherwise) was added to all items that were actual words in French or English, and to all of its place and voicing counterparts. For example, /demi/ is a real word in French meaning "half", so it received a final coda to become /demil/, as did its voicing (/temil/), manner (/zemil/), and voicing+manner (/semil/) counterparts. Final codas were added in other quadruplets to balance their frequency across the lists to an average of 39% of closed final syllables. All of the items were recorded by a single native French female speaker.

Of the 5 lists, one was held out for the test, so that (a) all participants were faced with novel items at test; and (b) all participants were tested with subsets of the same list, so that any difference across conditions had to be due to initial exposure. The remaining four lists were used for the exposure phase. Since there were 5 onsets × 4 lists × 9 items per onset per list in the exposure phase of the study, participants heard a total of 180 different non-words during this phase. At

1 It goes without saying that the question of how results ensuing from the present paradigm bear on 'real' language is just as open for this task in particular as it is for any other experimental paradigm.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

test, they were presented with 9 items of each of 4 onsets. Twelve exposure conditions (represented in Figure 2) were designed such that each of the 12 consonants served as Exposure, Within, Near, and Far the same number of times across participants. By virtue of this complete counterbalancing, we ensured that effects could never be reduced to differences in the frequency of onsets or sequences in the participants' native language, since every test item was presented mapped onto the Exposure type for one quarter of the participants who heard it at test; onto Within for another quarter; onto Near for a third quarter; and onto Far for the fourth quarter.

2.1.4 Distance measures

Being conservative in which models were evaluated could have meant that we were missing the dimension that learners actually relied on, which would have led to an inaccurate statement about how much one or another metric influenced generalization. Therefore, we considered 20 possibilities spanning the four hypotheses set out above. We would like to be the first to point out that this constituted rather extreme repeated testing, and thus results, particularly for the phonetic dimensions, should best be viewed as tentative, to be corroborated by future work with more focused hypotheses. Nonetheless, we believe that it is important to report this initial exploration, as it shows to what extent this avenue of research is promising, and should be followed up through ad hoc investigation.

2.1.4.1 Natural class distances

The strongest version of this hypothesis states that Exposure and Within items would have a null distance, whereas the other two types would have a non-null distance; a weaker version holds that Within will have a smaller distance than both Near and Far.

2.1.4.2 Distinctive features distance measures

Three types of featural distance were considered. First, distance was measured as explained above as the number of feature changes needed to convert one onset into another. Thus, this measurement could have the ordinal values 0 (Exposure), 1 (Within and Near), and 2 (Far). By treating these levels as ordinals, a linear fit is imposed (that is, two features away should be the same as two times one feature). Second, fit was calculated for a non-linear feature distance, where the three levels are viewed as three independent levels. The third variant took into account

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

interpolation between places of articulation (e.g., /t/ is 0 features away from /p k/, whereas /p/ is 1 feature away from /t k/; see Wilson 2003; Finley and Badecker 2009).

2.1.4.3 Linguistically informed distance measures

It is difficult to decide how to define a linguistically informed distance matrix, since language-specific phonetic space is multidimensional. As a first approximation, we used distance measures derived from articulatory data collected for an independent study, reported in Mielke (2012). In that study, four phonetically trained American English native talkers produced common sounds of the world's languages in three vocalic contexts (a_a, i_i, u_u). These measurements are not ideal since they were taken from a different set of speakers, who had a different native language, and were producing the sounds in a different context. Nevertheless, salient articulatory differences between the consonants in question (different places of articulation, more vocal fold vibration in voiced consonants, more airflow during the constriction phase of fricatives, etc.) are expected to hold up. For the present analyses, articulatory phonetic representations for the relevant consonants were generated by isolating the instances of /p t k f s ʃ b d g v z ʒ/ and performing separate Principal Component Analyses on measures of vocal tract shape, airflow, and larynx activity. The first two principal components of vocal tract shape (derived from mid-sagittal ultrasound images) were retained. Oral and nasal airflow were highly correlated for these consonants (which are all oral); therefore, only the first principal component of the two airflow measurements was used. Similarly, the first principal component of electroglottograph signal amplitude and larynx height was used to represent the larynx data. More details on the methods can be found in the source article (Mielke 2012). For the present study, we calculated a minimum distance (between the test onset and the closest training onset) and an average distance (between a given test onset and the training onsets). For comparison with other regressors which collapse across different dimensions (e.g., the one representing the Featural Distance hypothesis), it was reasonable to calculate an additional set, defined as the sum over all of the articulatory dimensions. This procedure yielded 10 regressors (average/minimum × (4 dimensions + sum)).

2.1.4.4 An uninformed distance measure

A set of acoustic distances was calculated directly from the stimuli used in the present study. This process went in several steps, as follows. The first step was to extract a psychoacoustically motivated spectral representation, Mel-Frequency

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

Cepstral Coefficients (MFCC), for each sound file. Secondly, the spectral distance S_{ij} between sound files *i* and *j* was computed using Praat's Dynamic Time Warping (DTW) method (further details can be found in the Praat manual; Boersma and Weenink 2005). Since this DTW algorithm abstracts away from duration, the temporal distance T_{ij} between sound files *i* and *j* was also calculated, defined as the absolute magnitude of the difference in duration. This resulted in two distance matrices: a spectral one, and a temporal one. To calculate a *single* measure of distance, in the third step these two matrices were each submitted to Principal Components Analysis (PCA), a linear statistical method that identifies coordinates for each item in a low-dimensional subspace representing an orthogonalization of the greatest dimensions of variance. An abstract four-dimensional coordinate was assigned to each stimulus item by concatenating the first three components of the spectral PCA (p^S_1, p^S_2, p^S_3) and the first component of the temporal PCA (p^T_1). Finally, the Euclidean distance between every two tokens was calculated. We considered several measurements within the umbrella of uninformed acoustic distances: (1) the average and minimum distance between the item under consideration and all of the exposure items; (2) those distances between the item under consideration and the centroid of the exposure distribution in the four-dimensional space. All three distances were estimated also when the initial sound (and not the whole sound file) was considered. This procedure yielded 6 additional regressors (average/minimum/centroid × onset/word).

2.1.5 Statistical analyses

A linear mixed model was used to predict subjective frequency rating, declaring participant and item as random effects. Statistical analyses were carried out in R (R Development Core Team 2011), with lmer, part of the lme4 package (Bates and Maechler 2009); significance was estimated with pvals.fnc in the languageR package (Baayen 2008b). Similar methods have been used in previous laboratory phonology work (e.g., Moreton 2008; Daland et al. 2011). It is important to point out that we have over-tested these data by fitting many distance models. The results of the comparison of different distance measures should be replicated in a different dataset.

3 Results

Before analyzing results to explore the representations allowing for generalization, we first checked that the paradigm led to reliable variation in subjective

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

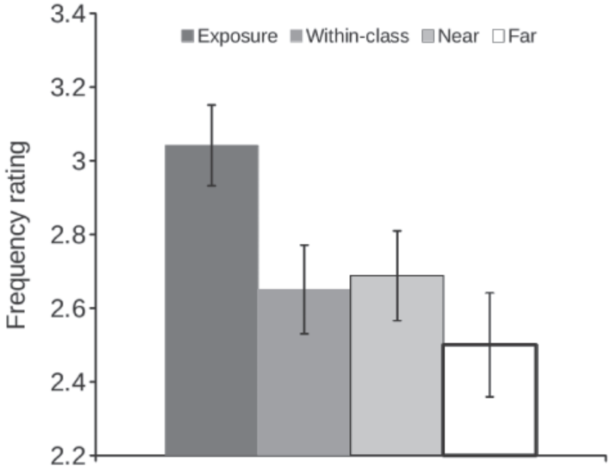

Fig. 3: Ratings by item type; error bars indicate standard errors.

frequency ratings. Average ratings by Test Item Type (Exposure, Within, Near, Far) are represented in Figure 3.

We used Far as the baseline, since it is easier to assess whether the task worked at all: if there is any learning, at least one other type should differ significantly from Far items. Indeed, the estimate for Exposure items was positive and highly significant ($\beta = .542$, $SE = .067$; $t = 8.09$, $p < .05$; confidence interval estimated with Highest Posterior Densities [HPD]: 0.3944–0.6621). Additionally, participants tended to estimate the frequency of Within ($\beta = .150$, $SE = .067$; $t = 2.25$, $p < .05$) and Near ($\beta = .188$, $SE = .067$; $t = 2.80$, $p < .05$) items as higher than that of Far items. This confirmed that the paradigm could tap implicit learning and generalization: variation in subjective frequency among *novel* non-words must have been caused by variation in the frequency of occurrence of their onsets and similar onsets. We now turn to the specific predictions made from each hypothesis.

One prediction made from the Natural Class Hypothesis was that generalization would be bound by the Subset Principle to Within, and not extend to Near. Since participants were being asked to rate subjective frequency, highest ratings were expected for Exposure, and lowest for Far. Declaring Near as the baseline level makes it easier to see whether Near ratings differed from those for Within. This prediction was not met, as the estimate for Within did not differ from that of the Near baseline ($\beta = -.037$, $SE = .067$; $t = 0.55$, $p > .05$). Thus, the Natural Class Hypothesis was not supported.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

The Featural Distance Hypothesis stated that generalization would respond to featural distance between the onsets in the exposure phase and the test onsets. Indeed, a model with featural distance as a regressor with three levels (Exposure 0, Within and Near 1, Far 2 features away) reveals that the estimates for both 1 and 2 depart from that of 0-distance (1: $\beta = -0.373$, $SE = .058$; $t = 6.43$, $p < .05$; 2: $\beta = -0.542$, $SE = .067$; $t = 8.09$, $p < .05$; treating this regressor as continuous naturally also yields a significant estimate for each unit of distance: $\beta = -0.270$, $SE = .033$; $t = 8.08$, $p < .05$). The estimate is negative, in keeping with the idea that listeners rated as less frequent items that were at a greater featural distance from the training ones. These results clearly support the Featural Distance Hypothesis over the Natural Class Hypothesis.

Finally, we investigated to what extent different distance models predicted listeners' frequency ratings. Given that regressors may be somewhat correlated, and in view of the sheer number of regressors considered, it was not ideal to incorporate them all into a single model. Therefore, each regressor was entered into a separate model, and models were compared in terms of the proportion of variance explained (the square of the correlation between fitted and real values) and the model fit as gauged by Bayes Information Criterion (BIC) (Pinheiro and Bates 2000; Baayen 2008a; Johnson 2008). Results from the top three models are shown in Table 2. The best model is one in which the minimum distance summed across articulatory dimensions was used to predict measures of similarity, which can explain approximately the same percentage of variance as any other model with good performance albeit at a lower BIC (or model cost). In contrast, the models including the uninformed phonetic measurements (raw acoustics) of the stimuli do little to improve the model fit over a baseline model with no fixed effects. For a full list of model performances, please refer to the

Table 2: Summary of variance explained and BIC (a measure of goodness of fit) of some representative models. In all cases, a linear mixed model was used to predict subjective frequency rating, declaring participant and item as random effects. The 'test type' model is that where the four levels of trial type (Exposure, Within, Near, Far) are entered as fixed effects.

Regressor	Variance	BIC
Random effects only	26.2	5072
Test type	29.5	5025
Linear feature-distance	29.5	5018
Sum of minimum articulatory distances	29.9	5008
Minimum acoustic distance, only onset	26.8	5066

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

Supplementary Materials, located at https://sites.google.com/site/acrsta/talks/generalizingphonotactics.

4 Discussion

In this study, participants were first presented with non-words whose onsets were restricted to five obstruents sharing voicing, and then asked to rate how frequently a non-word had been presented before. Crucially, all non-words had been presented the same number of times. Nonetheless, listeners gave lower frequency ratings to non-words whose onset differed in both voicing and place from all the word-initial sounds used during exposure than to non-words whose onset shared voicing and manner, or place and manner, with one of the exposure onsets. This effect confirms that the current design leads to the implicit acquisition of sound patterns which are reflected in the dependent measure used. Therefore, this paradigm can capture the variability in ratings that can help us answer the questions motivating the present experiment, as follows.

First, listeners' ratings of novel onsets that shared either voicing and manner or place and manner with one of the exposure onsets were higher than those attributed to onsets who shared neither place nor voicing. This study extends previous findings on adult learners' generalization of sound patterns to untrained consonants (Finley and Badecker 2009; Finley 2011, submitted), using a more implicit design (as in Bernard et al. submitted), thus providing clear laboratory evidence of the process of generalization. We observed that participants implicitly generalized the statistics of a group of obstruents to *both* place *and* voicing analogues. Such diverse generalization patterns may constitute the basis for diverging generalization patterns different language users make when faced with a similar change in progress. In the introduction, we discussed one such example: the triggers of /o/-lowering appear to have been extended from only /r/ to other non-lateral sonorants in the city of Schaffhausen, and to all coronals in some neighboring communities.

This brings us to our second question. While not all phonological patterns involve phonetically or featurally natural classes exclusively, a statistical majority of phonologically active classes are indeed natural according to traditional distinctive feature theories (Mielke 2008). While this observation may be largely attributable to the fact that sound change operates on phonetically-defined groups of sounds, it is reasonable for language users to encode productive phonological patterns in terms of featurally- or phonetically-defined classes of sounds. From this perspective, we would expect listeners to encode the observation that '/d g v z ʒ/ occur word-initially' using the Subset class, the minimal natural class

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

that contains these segments, to yield the more parsimonious constraint ‘voiced obstruents occur word-initially.’ If this was indeed the representation that listeners used in the present experiment, we should have observed higher ratings for the Within onsets (which belong to the Subset class) than for the Near ones (which do not). In the absence of such a difference, we are inclined to conclude that generalization in the current study operated not through a more abstract representation, but rather as a spread of the characteristics associated with an individual sound to sounds that are similar to it.

As for how similarity is measured, a feature-based classification of the test items explained a great deal more of the ratings than an uninformed distance dimension, based on wholesale acoustic measurements. However, using somewhat indirect articulatory measures produced as good a fit as a feature-based distance calculation. The articulatory measures were indirect because they were collected from a completely different set of stimuli. This leads to two important considerations. First, one would expect that more direct articulatory measures may even outperform the categorical feature-based distance measures. We hope future work may explore this possibility. Second, the fact that such distal articulatory measures fit responses better than direct acoustic measurements suggests to us that participants brought their own mental representations to the task. That is, they were not simply responding to proximity in the physical stimuli presented to them, as a machine or perhaps a non-human animal would have. On the contrary, they responded using knowledge, evoked by the stimuli, but crucially derived from their previous experience. The latter consideration reinforces findings of one’s native language shaping generalization in the lab (Bernard et al. submitted).

5 Implications

Language is characteristically productive. At the level of phonology, this is evidenced by a rich implicit knowledge of the sound patterns present in our native language, which allows differential processing of novel wordforms depending on the extent to which they conform to those patterns. In this paper, we investigated how listeners represent newly extracted patterns, through the way this knowledge is reflected in subjective frequency ratings. We documented robust implicit learning of the frequency of occurrence of individual sounds and investigated some of the factors that govern the spreading of subjective frequency to similar sounds. Our results bear more generally both on the units of representation that allow comparison between sounds, and the relevance of sound classes for the description of phonologies in language and the language user’s mind. We discuss each of these topics in turn.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

5.1 The dimensions and units of phonological similarity

Our results indicated that similarity is related to categorical features and/or related informed phonetic parameters, but did not seem to reflect uninformed phonetic distances. The success of models employing distance metrics based on the former factors (relative to models based on the latter factor) indicates that generalization is mediated by linguistic knowledge. We can also speculate that a more sophisticated acoustic measure targeting particular acoustic cues (such as presence of low-frequency or a voicing bar) would also be more successful than our naïve acoustic measure.

Naturally, conclusions may be different for other tasks. For example, a recent study reports that the well-established phonological similarity effects in verbal working memory reflect primarily acoustic interference at the stage of recall for a purely perceptual task, and articulatory similarity when production is involved (Schweppe, Grice, and Rummer 2011). Similar (partial) dissociations in the way adults represent subsegmental similarity have been evidenced in second-language learners by de Jong and colleagues (de Jong, Silbert, and Park 2009; de Jong, Hao, and Park 2010). In terms of perception, learners were asked to identify consonants varying in place, manner, and voicing in their second language, and identification was scored for each feature separately. If individual variation in perception is due to different individuals being better at detecting a given feature (all else being equal), one would expect accuracy for voicing among labial stops to be correlated with accuracy for voicing in coronal stops. In fact, participants were more internally consistent across place of articulation for manner than for voicing. A similar design was pursued in a production study, which revealed that internal correlations across place of articulation were higher for voicing than for manner. Thus, complementary patterns of internal consistency were documented for perception and production. Modeling work suggests that this diversity in the behavior of phonological features is not due exclusively to cognitive biases in humans, since it is clearly represented in the phonetic signal and can be captured instrumentally. Indeed, Lin and Mielke (2008) found that manner features could be extracted easily from the acoustic signal, whereas features representing place of articulation were hard to extract automatically based only on acoustic measurements. In contrast, place could be easily captured through articulatory measurements of vocal tract shape, while manner was more elusive in this type of signal.

Therefore, different strands of the literature on phonological representations (our artificial grammar generalization study above, perception and production data from L2, and modeling work) begin to converge in the suggestion that simi-

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

larity is not unidimensional. Instead, the sound patterns evidenced in language are likely the effect of both diachronic perceptual and articulatory pressures and, perhaps to a more limited extent, cognitive biases emergent from online calculations of similarity along articulatory and perceptual dimensions. There remain three outstanding questions now facing this literature.

First, there are important gaps in our empirical knowledge, particularly with respect to how adult listeners/speakers come to represent similarity the way they do. While cross-linguistic studies demonstrate that similarity measures for identical stimuli differ across language backgrounds (e.g., Nishi et al. 2008), the developmental timeline of such effects remains to be documented. This developmental timeline could shed light on the relative importance of perceptual and articulatory dimensions affecting similarity effects.

Second, it is important to determine whether there is any effect of an additional dimension of similarity, namely similarity in functional (phonological and lexical) properties. An example of phonological similarity is the following: imagine two languages having identical phonetic inventories, but the sounds X and Y pattern together (i.e., can occur in the same positions, trigger similar phonological processes) in only one of them. If functional behavior is a third dimension affecting similarity, one would predict greater perceptual similarity between X and Y in the first language, and for this to be evidenced only by native speakers of that language. Recent work documents the impact of functional behavior on similarity judgments through the comparison of linguistic populations in which a given pair of sounds is only weakly contrastive (Boomershine et al. 2008; Johnson and Babel 2010). For example, Johnson and Babel (2010) argue that Dutch listeners rate identical tokens of [s] and [ʃ] as more similar and have a harder time discriminating them than English listeners do because in Dutch [ʃ] occurs either as the surface realization of /s/ before /j/ or in some loanwords. These results suggest that phonological and lexical experience can affect cognitive representations through the increase of functional pressures to maintain or lose a distinction.[2] Therefore, it would be of interest to test how functional properties from one's native language constrain generalization. Bernard et al.'s (submitted) report that phonemic experience is crucial to the generalization of newly learned sound patterns fits in with the idea that functional properties also play an important role in structuring similarity.

2 Of course, this conclusion rests on the assumption that the sounds under study are acoustically equally discriminable across the languages being compared, and that only their functional roles differ. Boomershine et al. (2008) and Johnson and Babel (2010) get around this problem by testing all linguistic groups with a single set of stimuli.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

Finally, these insights should be integrated into models that quantify the extent to which these different factors explain language processing and phonological patterns. In particular, we would like to tease apart the effect of historical pressures and cognitive biases, a goal that at present can only be achieved through computational and statistical models.

Once the dimensions along which similarity is computed have been established, the next step is to determine how space is structured along each dimension. There are few studies in which parametric variations have been implemented, but in this sparse literature there seems to be some disagreement concerning the units of similarity. Informal inspection of our results suggests that most of the effect is brought about by the first feature, whereas the second feature seemed to yield a smaller effect (although the precise size difference should be studied directly in a different design). In contrast, White and Morgan (2008) document that the effects of distance along featural dimensions are linear in infants' word recognition. Even within artificial grammar studies on adults' learning of alternations, the metrics of similarity are unclear. Peperkamp, Skoruppa, and Dupoux (2006) and Peperkamp and Dupoux (2007) trained participants on an alternation between two sounds that differed in either only one feature (/p/ turning into /b/), or three features (/p/ turning into /ʒ/, where place, manner, and voicing change). In a perception task, participants succeeded in learning both the one-feature and the three-feature changes (Peperkamp and Dupoux 2007), while only the one-feature alternation was learnable in a production task (Peperkamp et al. 2006). Extending these results, Skoruppa, Lambrechts, and Peperkamp (2011) showed that talkers can quickly acquire a one-feature change in a production task with feedback, and that performance reaches an asymptote more quickly for these minimal changes than for a two- or three-feature change, with no difference among the latter two conditions. Evidently, this is a matter deserving further investigation. An equally important question, on which there is little research, is whether similarity is asymmetric (Chang, Plauche, and Ohala 2001; Garrett and Johnson 2013; McGuire and Babel 2012).

5.2 Natural classes

There appears to be a mismatch between our findings (no bias for Subset generalization) and observed phonological patterns (typically involving sound classes). If natural class–based encoding is not an automatic consequence of exposure to a phonological pattern, as the results of the present experiment suggest, then why are natural classes so prevalent in phonology? We put forward three, not mutually exclusive, explanations.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

First, natural-class based patterns could emerge as the consequence of language use, since similar sounds face similar phonetic pressures. As Mielke (2008: 90) puts it, "phonetic similarity [may be] relevant for the initiation of the parallel sound changes rather than in the extension of the result of one sound change to a larger class".

Second, there could be additional cognitive pressures that we did not target in the present study, but which would bias language users toward natural classes. One clear candidate involves greater learnability of multi-sound patterns when they share many phonetic characteristics. The present study cannot speak to this question, because *all* exposure conditions were based on natural classes. However, many other studies have documented that adults find it easier to learn patterns involving a natural class than to learn patterns involving an arbitrary set of sounds (Pycha et al. 2003; Wilson 2003; Moreton 2008; Endress and Mehler 2010; Skoruppa and Peperkamp 2011). For example, participants in the natural class condition in Wilson (2003) were capable of learning that the onsets of the second and third syllables agreed in nasality: either both were nasal *(dumena)* or neither was *(tukola, sutola)*. Participants in the arbitrary condition, in contrast, were unable to learn that, e.g., /m/ and /t/ were followed by /l/, whereas /k/ was followed by /n/ *(dumela, tukona, sutola)*. Thus, if a language has a sound pattern that affects a random set of sounds, even a small difference in ease of learning should translate to a higher chance of the pattern being lost over the course of several generations.

Third, learners' acquisition of patterns may be restricted by natural class boundaries only during early first language acquisition. Numerous studies have documented that young infants readily learn and generalize sound patterns to within-class sounds, although the ability to generalize may decline by about 14 months of age (see Cristia, Seidl, and Francis 2011 for a recent review). There is a strong version of the Subset Learning Hypothesis, which predicts that responses for the Within onsets should be equivalent to those for Exposure onsets, since both of them fulfill the represented pattern based on the natural class to the same extent. As evident in the experiment reported here, and every other comparable study, this is clearly not the case for adults, for whom there seems to be a cost in generalization to untrained sounds. In contrast, Cristia and Peperkamp (2012) report that this precise pattern of results obtains in 6-month-olds, who encode the sound class rather than the specific sounds. During familiarization, infants heard many different non-words with three different onsets (e.g., /b d ʒ/). At test, half of the infants were presented with new items having the three exposure onsets and items with three untrained, but within-class, obstruents (i.e., /g v ʒ/). These infants showed no preference, as if unable to detect the novelty in the Within trials. The second half of the

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

infants, who were given the choice between Within and Near (i.e., /k f s/) items, showed a robust novelty preference for the Near items. Even though the methods used with infants and adults are clearly not the same, it is intriguing that with similar stimuli, 6-month-olds and adults appear to encode sound patterns in very different ways. If the behavior recorded for infants in this artificial grammar learning study replicates their learning of the sound patterns found in their native language, then the prevalence of class-based patterns in language would not be at all surprising, as infants would automatically code patterns in terms of the subset class involved.

6 Conclusion

In this article, we sought to shed light on the factors affecting generalization of newly learned sound patterns to untrained non-words and untrained consonants. Our results suggest that generalization to untrained non-words is robust. When generalization to untrained consonants occurs, it does not seem to be constrained by the Subset Principle, because generalization targets are not limited to members of the narrowest natural class encompassing all sounds with similar phonological behavior. Instead, generalization to untrained sounds follows from pairwise similarity between consonants present in the exposure and the target consonants. This similarity is better captured through dimensions that rely on preexisting phonetic and phonological knowledge, whereas uninformed measures of acoustic similarity contribute little to shaping listeners' judgments. Further research should continue to explore the dimensions and units structuring similarity matrices, a crucial factor shaping phonological generalization.

Acknowledgments: AC and SP designed the study and gathered the data; AC and JM carried out the acoustic analyses; AC and RD carried out the statistical analyses; all authors contributed to the writing. We thank Laura Robotham, Isabelle Brunet, Ellenor Shoemaker, Amanda Seidl, and Tanya Awad for help with the stimuli; as well as Isabelle Brunet, and Anne-Caroline Fievet for gathering some of the data. Our recognition goes also to the audiences of the Manchester Phonology Meeting in May 2010 and the Acoustical Society of America meeting in November 2010 for helpful comments; Roger Levy for advice on the statistical analyses; Dan Dediu and particularly Emmanuel Dupoux for many shrewd discussions; and two reviewers for their very helpful comments. AC has been supported by Ecole de Neurosciences de Paris and Fondation Fyssen; JM by SSHRC (410-2007-0735); SP by the Agence Nationale de la Recherche (ANR-2010-BLAN-1901).

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

References

Baayen, R. Harald. 2008a. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R. Harald. 2008b. Package languageR version 1.2: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics". http://cran.r-project.org/web/packages/languageR/index.html

Bates, David, & Martin Maechler. 2009. Package lme4 version 0.999375-40: linear mixed-effects models using S4 classes. http://cran.r-project.org/web/packages/lme4/index.html

Becker, Michael, Andrew Nevins, & Jonathan Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88. 231–268.

Bernard, Amelie, Kristine Onishi, & Amanda Seidl. Submitted. Constraints on phonotactic learning in adulthood: Phonemic vs. allophonic variation.

Blevins, Juliette. 2004. *Evolutionary Phonology*. Cambridge: Cambridge University Press.

Boersma, Paul, & David Weenink. 2005. Praat: Doing phonetics by computer version 5.0.09. Retrieved April 4, 2009, from www.praat.org

Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume, & Keith Johnson. 2008. The impact of allophony vs. contrast on speech perception. In Peter Avery, Elan Dresher, & Keren Rice (eds.), *Contrast in Phonology,* 143–172. Berlin: Mouton de Gruyter.

Chang, Steve S., Madelaine C. Plauche, & John J. Ohala. 2001. Markedness and consonant confusion asymmetries. In Elizabeth Hume & Keith Johnson (eds.), *The Role of Speech Perception in Phonology*, 79–101. San Diego, CA: Academic Press.

Chomsky, Noam, & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Cristia, Alejandrina, & Sharon Peperkamp. 2012. Generalizing without encoding specifics: Infants infer phonotactic patterns on sound classes. In Alia K. Biller, Esther Y. Chung, & Amelia E. Kimball (eds.), *36th Boston University Conference on Language Development Proceedings*, 126–138. Somerville, MA: Cascadilla Press.

Cristia, Alejandrina, Amanda Seidl, & Alexander Francis. 2011. Phonological features in infancy. In G. Nock Clements & Rachid Ridouane (eds.), *Where Do Phonological Contrasts Come From? Cognitive, Physical and Developmental Bases of Phonological Features*, 303–326. Amsterdam: John Benjamins.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis, & Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28. 197–234.

de Jong, Kenneth J., Yen-Chen Hao, & Hanyong Park. 2010. Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent. *Journal of Phonetics* 37. 357–373.

de Jong, Kenneth J., Noah H. Silbert, & Hanyong Park. 2009. Generalization across segments in second language consonant identification. *Language Learning* 59. 1–31.

Deese, James. 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology* 58. 17–22.

Deese, James. 1998. Remembering visual feature conjunctions: Visual memory for shape-colour associations is object-based. *Visual Cognition* 5. 409–455.

Endress, Ansgar, & Jacques Mehler. 2010. Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and Performance* 36. 235–250.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

Finley, Sara. 2011. Generalization to novel consonants in artificial grammar learning. In Laura H. Carlson, Christophe Hölscher, & Thomas F. Shipley (eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 318–323. Austin, TX: Cognitive Science Society.

Finley, Sara. submitted. Generalization to novel consonants: Place vs. voice.

Finley, Sara, & William Badecker. 2009. Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61. 423–437.

Frisch, S. 1996. Similarity and frequency in phonology. Ph.D. thesis, Northwestern University.

Garrett, Andrew, & Keith Johnson. 2013. Phonetic bias in sound change. In Alan C. L. Yu (ed.), *Origins of Sound Change: Approaches to Phonologization*. Oxford: Oxford University Press.

Hale, Mike, & Charles Reiss. 2003. The subset principle in phonology: why the tabula can't be rasa. *Journal of Linguistics* 39. 219–244.

Hall, T. Alan. 2001. *Distinctive Feature Theory.* Berlin and New York: Mouton de Gruyter.

Hayes, Bruce, & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440.

Jakobson, Roman, Gunner Fant, & Morris Halle. 1963. *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates*. Cambridge, MA: MIT Press.

Janda, Richard D., & Brian D. Joseph. 2003. Reconsidering the canons of sound-change: Towards a big bang theory. In Barry J. Blake & Kate Burridge (eds.), *Historical Linguistics 2001*, 205–219. Amsterdam/Philadelphia: John Benjamins.

Jansen, Wouter. 2007. Phonological voicing, phonetic voicing, and assimilation in English. *Language Sciences* 29. 270–293.

Johnson, Keith. 2008. *Quantitative Methods in Linguistics*. Oxford: Blackwell.

Johnson, Keith, & Molly Babel. 2010. On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics* 38. 127–136.

Keating, Patricia. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60. 286–319.

Keel, William. 1982. *Atomic Phonology and Phonological Variation*. Tubingen: Gunter Narr.

Kenstowicz, Michael, & Charles Kisseberth. 1979. *Generative Phonology*. San Diego, CA: Academic Press.

Kiparsky, Paul. 2006. The amphichronic program vs. evolutionary phonology. *Theoretical Linguistics* 32. 217–236.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

Labov, William. 2001. *Principles of Linguistic Change: Social Factors*. Oxford: Blackwell.

Labov, William. 2011. *Principles of Linguistic Change: Cognitive and Cultural Factors*. Oxford: Blackwell.

Lin, Ying, & Jeff Mielke. 2008. Discovering place and manner features – what can be learned from acoustic and articulatory data? *University of Pennsylvania Working Papers in Linguistics* 14(1). Article 19. http://repository.upenn.edu/pwpl/vol14/iss1/19/

McDermott, Kathleen B. 1996. The persistence of false memories in list recall. *Journal of Memory and Language* 35. 212–230.

McGuire, Grant L., & Molly Babel. 2012. A cross-modal account for synchronic and diachronic patterns of /f/ and /θ/. *Laboratory Phonology* 3. 251–272.

Mielke, Jeff. 2008. *The Emergence of Distinctive Features.* New York: Oxford University Press.

Mielke, Jeff. 2012. A phonetically-based metric of sound similarity. *Lingua* 1222. 145–163.

Mielke, Jeff., Lyra Magloughlin, & Elizabeth Hume. 2011. Evaluating the effectiveness of Unified Feature Theory and three other feature systems. In John Goldsmith, Elizabeth Hume, &

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

Leo Wetzels (eds.), *Tones and Features: in Honor of G. Nick Clements,* 223–263. Berlin: Mouton de Gruyter.
Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25. 83–127.
Moreton, Elliot, & Joe Pater. 2012. Structure and substance in artificial-phonology learning, part 1: structure. *Language and Linguistics Compass* 6. 686–701.
Nishi Kanae, Winifred Strange, Reiko Akahane-Yamada, Rieko Kubo, & Sonja A. Trent-Brown. 2008. Acoustic and perceptual similarity of Japanese and American English vowels. *Journal of the Acoustical Society of America*, 124. 576–588.
Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The Production of Speech*, 189–216. Heidelberg: Springer-Verlag.
Pajak, Bozena, & Roger Levy. 2011. Phonological generalization from distributional evidence. In L. H. Carlson, C. Hölscher, & T. Shipley (eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 2673–2678. Austin, TX: Cognitive Science Society.
Peperkamp, Sharon, & Emmanuel Dupoux. 2007. Learning the mapping from surface to underlying representations in an artificial language. In Jennifer Cole & Jose-Ignacio Hualde (eds.), *Laboratory Phonology IX*, 315–338. Mouton de Gruyter, Berlin.
Peperkamp, Sharon, Katrin Skoruppa, & Emmanuel Dupoux. 2006. The role of phonetic naturalness in phonological rule acquisition. In David Bamman, Tatiana Magnitskaia, & Colleen Zaller (eds.), *30th Boston University Conference on Language Development Proceedings*, 464–475. Somerville, MA: Cascadilla Press.
Pierrehumbert, Janet, Mary E. Beckman, & Robert D. Ladd. 2001. Conceptual foundations of phonology as a laboratory science. In Noel Burton-Roberts, Philip Carr, & Gerard Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*, 79–101. Academic Press.
Pinheiro, Jose C., & Douglas M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
Pycha, Anne, Pawel Nowak, Eustad Shin, & Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In G. Garding & M. Tsujimura (eds.), *West Coast Conference on Formal Linguistics 22 Proceedings*, 101–114. Somerville, MA: Cascadilla Press.
R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
Schweppe, Judith, Martine Grice, & Ralf Rummer. 2011. What models of verbal working memory can learn from phonological theory: Decomposing the phonological similarity effect. *Journal of Memory and Language* 64. 256–269.
Skoruppa, Katrin, Anna Lambrechts, & Sharon Peperkamp. 2011. The role of phonetic distance in the acquisition of phonological alternations. *Proceedings of North East Linguistic Society* 39. 464–475.
Skoruppa, Katrin, & Sharon Peperkamp. 2011. Adaptation to novel accents: Feature-based learning of context-sensitive phonological regularities. *Cognitive Science* 35. 348–366.
White, Katherine S., & James L. Morgan. 2008. Sub-segmental detail in early lexical representations. *Journal of Memory and Language* 59. 114–132.
Wilson, Colin. 2003. Experimental investigation on phonological naturalness. In G. Garding & M. Tsujimura (eds.), *WCCFL 22 Proceedings*, 533–546. Somerville, MA: Cascadilla Press.
Wilson, Colin. 2006. Learning phonology with a substantial bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM

Authenticated | sharon.peperkamp@ens.fr author's copy
Download Date | 12/24/13 9:54 AM