Generalizing rules over clicks

PEDRO TIAGO DA SILVA GONÇALVES MARTINS

Submitted in partial fulfillment of the requirements for the degree of

M.Sc. in Brain and Cognition

Department of Information and Communication Technologies
Universitat Pompeu Fabra

Supervisors:
Luca Bonatti & Juan Manuel Toro

September 2015

Universitat Pompeu Fabra Barcelona

Acknowledgments

I would like to thank my supervisors, Luca Bonatti and Juan Manuel Toro, whose guidance, expertise, helpfulness and certainly patience made this thesis possible.

Some of my fellow classmates offered their help at different stages, and I would like to thank them here: Marc Lluís Vives, Ryszard Cetnarski, Jasna Čakarun and Mara Rolke.

I also thank Cedric Boeckx for unfailing support.

Finally, I must express my gratitude to three people who in different but very important ways allowed me to get to this point: Jone Intxaurraga, João Veloso and Liza Lampič.

Abstract

Humans are able to distill regularities from speech data. Consonants and vowels, the basic units of speech, carry different kinds of information. Consonants carry mainly lexical information and are crucial for word identification, while vowels carry mainly prosodic information are associated with with syntactic structure, crucial for grammar learning. This has led to the idea that there is a division of labor that ascribes different functional roles to consonants and vowels, called the *CV hypothesis*. Indeed, humans are able to extract structural information and segment words from consonants, and extract generalizations from vowels, and the converse is true: humans cannot extract generalization rules from consonants nor segment words on the basis of vowels. A possibility is that the computation mechanisms underlying this division of labor are constrained by linguistic representation. This possibility was tested by presenting participants with a stream of CVCVCV syllables made up of vowels and click consonants, the latter following an ABA pattern. Clicks were chosen because they are not perceived as linguistic by speakers of languages that don't make use of them. Participants showed preference for items that did not follow the pattern. Linguistic representation (or lack thereof) influenced the computations that participants could perform over the clicks. Some possible explanations for the preference for nonrule-words are provided.

Keywords: *rule-learning*, *consonants and vowels*, *speech perception*, *click consonants*

Contents

1 Introduction

Humans are exposed to apparently chaotic speech sounds from the moment they are born, yet they are able to process and make (linguistic) sense of them, and eventually learn what becomes their native language. In linguistics, this impressive feat has been attributed to some kind of innate machinery or knowledge that allows infants to learn a language from impoverished and insufficient data (for recent discussion, see Berwick et al., 2011). While this might be true to an extent, recent investigations have shown that the data infants are exposed to does have some important richness to it, and that humans exploit certain features of speech sounds that allow them to extract different kinds of information which facilitates language learning. An important aspect lies in the informational and functional differences between consonants and vowels.

Consonants and vowels, the basic units that make up speech, have different brain representations and are processed differently(Caramazza et al., 2000; Carreiras and Price, 2008), even before a lexicon is acquired (Benavides-Varela et al., 2012; Poltrock and Nazzi, 2015). Consonants are usually associated with lexical information and are crucial for word identification (Cutler et al., 2000) — witness the Czech sentence "*Strč prst skrz krk*" ("stick your finger through your neck"), built entirely out of consonants — while vowels, the main carries of prosodic cues (vowel harmony, pitch, etc.), are associated with syntactic structure (Nespor and Vogel, 1986). These two kinds of information allow humans to learn a distill a lexicon (the words) and a grammar (the rules) from the signal. This has led to the idea that there is a division of labor that ascribes different functional roles to consonants and vowels, the *CV hypothesis* (Nespor et al., 2003). This hypothesis has been supported by studies showing that consonants are the target of statistical computation, which allows for word segmentation (Bonatti et al., 2005; Mehler et al., 2006), and a bad target for rule generalization (Toro et al., 2008a). The converse is also true: humans extract generalization rules but not statistical cues from vowels. (Bonatti et al., 2005; Pons and Toro, 2010).

Contra the CV Hypothesis, it could be argued that the asymmetric distribution of vowels and consonants in the world's languages accounts for this difference. If there are more vowels than consonants, consonants would necessarily be more informative. But this is not true of all languages. It has been shown that French speakers, for example, also process vowels and consonants differently, even though the vowel-consonant ratio in French is very balanced (New et al., 2008), thus showing that consonant distribution cannot account for the C-V functional distinction.

It could also be argued that the physical properties of vowels and consonants account for the distinction, since vowels carry more energy are more salient than consonants, a well-known fact (Ladefoged and Johnson, 2014). However, it has been shown that if energy in consonants is artificially increased, even beyond that of vowels, listeners still fail to extract generalization rules over them (Toro et al., 2008b). This goes to show that acoustic salience cannot explain the functional difference between consonants and vowels.

A question that still remains is what features of the signal account for this division of labor. An interesting possibility is that the lexical import of consonants goes hand-in-hand with a linguistic kind of representation (Bonatti et al., 2007). In other words, if consonants were represented in a non-linguistic way, perhaps structures could be generalized over them. This is in fact what happens in rats, which do not represent linguistically, and as such are “better” than humans at rule extraction (de la Mora and Toro, 2013). It is known that whether a stimulus is represented linguistically or as noise influences its processing. For instance, language-related brain areas are modulated differently by the same stimulus depending on whether or not they are perceived as speech (Möttönen et al., 2006), and audio-visual speech perception also relies on the stimuli being perceived as speech (Tuomainen et al., 2005).

A way in which this possibility could be explored is by using stimuli composed of clicks. Clicks are lingual ingressive consonants, that is, they are produced by articulating two places of contact in the mouth with the tongue, which forms an air pocket that is then rarefied by applying pressure with the tongue, which is then released. The result is a powerful, plosive sound. An interest feature of clicks is that though they are consonants, they are perceived as noise by speakers of languages who do not employ them (Best et al., 1988; Best and Avery, 1999), while speakers of Zulu or !Xhosa, languages with clicks, used them contrastively. Thus, clicks might be a good way of assessing whether the activation of linguistic representation blocks rule generalization over consonants, by testing the prediction that if consonants are not represented linguistically, then speakers will be able to generalize rules over them.

To test this prediction, for the present experiment a stream of click-vowel syllables was assembled, in which clicks followed a specific structural pattern. If linguistic representation is an important factor for ascribing different functional values to consonants and vowels, participants should be able to generalize rules over the clicks.

2 Methods

2.1 Participants

Participants were 20 Catalan speakers, all of whom also spoke at least Spanish, and some of whom a third one (10 females, 10 males; mean age = 26,95 years; range = 21–44). They each received €5 for their participation.

2.2 Stimuli

We created a stream of CV syllables made up of 12 CVCVCV nonce words, all consonants being clicks. For each word, the click tier followed one of six ABA patterns (that is, first and third clicks were identical), while the vowel tier followed one of two ABC patterns (all vowels were different from each other within and across patterns). See **Table 1** for a full list of stimuli. In order to avoid syllable repetition and/or different, accidental patterns, extra CV syllables were inserted in between the words, made up of the same vowels that composed the words ([a,e,ɛ,i,ɔ,u]), and non-click, and sonorant consonants [l,m,n] (thus yielding [la], [ma], [na], [le], [me], etc.). These randomly distributed syllables occurred in sequences of one to three, and made up about a third of the stream.

Noise syllables in the stream were synthesized with the MBROLA Brazilian Portuguese database BR4 (Dutoit et al., 1996). This database was chosen for the clarity of the vowels therein, deemed appropriate for experimentation with Catalan speakers. F0 was set to 240 Hz and phoneme duration to 120 ms. The main CVCVCV items were composed by concatenating natural clicks and vowels extracted from previously synthesized syllables. Clicks were obtained from !Xhosa sound files made available in (Ladefoged, 2006). The exact procedure was as follows: clicks were manually extracted from natural recordings, and the final set was chosen on the basis of their distinctiveness. Each click was kept to 120 ms, so as to fit in seamlessly with the rest of the stream. All segments were equalized in order to eliminate differences in intensity between the synthesized segments and the clicks, due to both their natural physical characteristics and the different intensities of all the sound files (and also to mitigate the natural perceptual bias towards vowels, which usually sound more prominent).

Test items were of two kinds: *nonrule-words* and *rule words* (see **Table 1**), and were composed following the procedure above. Rule-words displayed the same vowel sequences in the familiarization words and the ABA click structure, but with new clicks not present in the familiarization stream. Nonrule-words had either an AAB or ABB click structure, in equal measure. Thus, click structure is the only characteristic distinguishing rule-words from nonrule-words. Rule-words instantiate the ABA click structure, while nonrule-words violate it.

Familiarization words	Test words	
	Rule-words	Nonrule-words
ba**n**u**t**ɛ	**k**a**p**u**k**ɛ	**k**a**k**u**p**ɛ
ba**t**u**b**ɛ	**p**a**k**u**p**ɛ	**p**a**k**u**k**ɛ
na**b**u**n**ɛ	**k**i**p**e**k**ɔ	**k**i**k**e**p**ɔ
na**t**u**n**ɛ	**p**i**k**e**p**ɔ	**p**i**k**e**k**ɔ
ta**b**u**t**ɛ		
ta**n**u**t**ɛ		
bi**n**e**b**ɔ		
bi**t**e**b**ɔ		
ni**b**e**n**ɔ		
ni**t**e**n**ɔ		
ti**b**e**t**ɔ		
ti**n**e**t**ɔ		

Table 1 – Stimuli. For simplicity, click consonants are represented here and throughout this paper by letters in bold, which are not to be taken as phonetic symbols. Vowels are all IPA. For a breakdown and articulatory description of the clicks used in the stimuli, please refer to the **Appendix**.

2.3 Procedure

All participants were tested individually in a silent room with no distractions. The experiment was controlled by a Macbook Pro laptop running Psyscope X (Cohen et al., 1993), and the stimuli delivered through supra-aural headphones. After a familiarization period of 5 min, an auditory two-alternative forced-choice test followed, with test pairs of rule-words versus nonrule-words. Participants were prompted to choose the member of the pair that most resembled the familiarization stream, by pressing a key on the keyboard. Throughout the experiment, there was no mention of the stimuli being composed completely of human-language sounds. This test assessed the ability to extract a generalization rule on the basis of (click) consonant structure. 16 test trials were presented pseudo-randomly, so as to avoid consecutive pairs comparing the same patterns and in the same order. Test items in each trial had a 500 ms pause in between them.

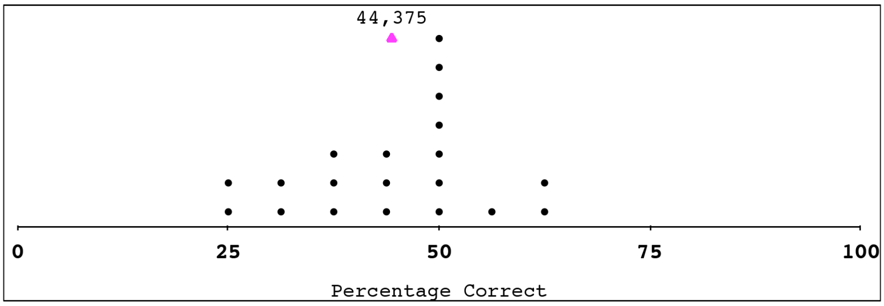

Figure 1 – Mean (triangle) and individual scores (black circles) in the test.

2.4 Results

Participants preferred nonrule-words to rule words, more so than expected by chance. ($M = 44.4\% \pm 10.9$), $t(19) = -2.308, p = 0.0324$ (see **Fig. 1**) This is a somewhat surprising result, but it shows that participants do distinguish between the rule and the non-rule patterns, something which is not true in stimuli composed of regular, linguistically represented consonants (Toro et al., 2008a). An ANOVA showed that the click-tier of nonrule-words did not have an effect in the results: $F_{(1,19)} = 0.21, p = 0.65$.

3 Discussion

As predicted by the CV hypothesis and as expected (Bonatti et al., 2007), participants were not indifferent to the clicks, given that they could not be represented linguistically by speakers of languages that do not employ them. If the clicks were processed linguistically, as "regular" consonants are, we would expect no preference over any kind of stimulus, rule- or nonrule-word, in line with previous results (Bonatti et al., 2005; Toro et al., 2008a). The fact that there was a preference also goes to show that the clicks were definitely distinguished by the speakers, despite their perception as "noisy", as some participants reported. Thus, our results corroborate the idea that linguistic representation mediates the way in which speech data is processed by the individual.

But the real intriguing question is indeed that the preference is towards nonrule-words. This is unexpected, and is a novel result. At face value, if a preference *is* displayed, rule-words would be the expected tendency, since the rule was fed to the participants in the familiarization, and the conclusion would be that clicks are as good a candidate for generalizing rules once linguistic representation is blocked as vowels in normal conditions. But this is not what happened. Participants did not prefer the ABA pattern they were exposed to in the familiarization. A possible explanation is that, since linguistic representation was not possible, participants generalized other patterns in the test items. This is much more likely than participants selectively preferring nonrule-words by excluding rule-words. Here, the pattern of the nonrule-words is very important. Nonrule-words followed two different kinds of patterns: AAB and ABB. Since linguistic representation was blocked, participants might have defaulted to a natural preference for adjacent reduplication. Several studies support this idea. For example, (Reber, 1967, 1969) shows that participants are able to generalize structures with repeated items to novel instances, yet this ability was absent if repetition was not present in the structure (Tunney and Altmann, 2001). This is true in manifestly non-linguistic items as well. For instance, adults can learn repetition-based grammars but not ordinal grammars implemented with tones (Endress et al., 2007), showing a clear preference for adjacent repetition. but not ordinal. In another study, it was shown that neonates prefer syllable sequences that contain adjacent repetition (ABB) than sequences with distant repetitions (ABA). Both instantiate a simple rule, but ABA got higher response when contrasted with an ABC sequence. Taken together, these studies lend support to the explanation that participants in our experiment preferred nonrule-words by virtue of their adjacent reduplication structure, when faced with items that could not be represented linguistically. An interesting adjustment in future experiments would be to expose participants to longer periods of familiarization and test, since our results could be an instance of ongoing learning. It could be that participants realized there was a rule, and in the test phase had not yet learned it, and latched on to the adjacent-repetition items. Given more exposure time, it could be that a tendency towards the ABA rule words would surface.

Another direction for future research would be to have speakers of click languages undergo the same experiment. The CV hypothesis would predict that they would fail to extract the generalization rules over the clicks, given the fact that they would represent them linguistically. But such an experiment would help refine the role of linguistic representation vs. physical features of items in the processing of speech and language learning.

4 Final Remarks

Humans process consonants and vowels differently, by computing statistics over consonants and generalizing rules over vowels. There is a very tight correlation between the kind of information consonants and vowels carry and the kinds of computations that can be performed over each of them, which seems to be mediated by differentiated linguistic representations. Our results follow a line of research that has yielded important discoveries regarding what these computations are and what kind of linguistic knowledge they facilitate. By using clicks, we have not only corroborated the CV hypothesis, but also gained insight on the computations humans can perform over non-linguistic material.

Bibliography

Benavides-Varela, S., Hochmann, J.-R., Macagno, F., Nespor, M., and Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences*, 109(44):17908–17913.

Berwick, R. C., Pietroski, P., Yankama, B., and Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science*, 35(7):1207–1242.

Best, C. T. and Avery, R. A. (1999). Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. *Psychological Science*, 10(1):65–70.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by english-speaking adults and infants. *Journal of Experimental Psychology: Human perception and performance*, 14(3):345.

Bonatti, L. L., Pena, M., Nespor, M., and Mehler, J. (2005). Linguistic constraints on statistical computations the role of consonants and vowels in continuous speech processing. *Psychological Science*, 16(6):451–459.

Bonatti, L. L., Peña, M., Nespor, M., and Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychological Science*, 18(10):924–925.

Caramazza, A., Chialant, D., Capasso, R., and Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, 403(6768):428–430.

Carreiras, M. and Price, C. J. (2008). Brain activation for consonants and vowels. *Cerebral Cortex*, 18(7):1727–1735.

Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). Psyscope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using macintosh computers. *Behavior Research Methods, Instruments, & Computers*, 25(2):257–271.

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., and Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & cognition*, 28(5):746–755.

de la Mora, D. M. and Toro, J. M. (2013). Rule learning over consonants and vowels in a non-human animal. *Cognition*, 126(2):307–312.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and Van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on Spoken Language Processing*, volume 3, pages 1393–1396. IEEE.

Endress, A. D., Dehaene-Lambertz, G., and Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, 105(3):577–614.

Ladefoged, P. (2006). UCLA Phonetics Lab Data. [Materials used in A Course in Phonetics 5th Ed., Table 6.3]: `hhttp://www.phonetics.ucla.edu/index.html`.

Ladefoged, P. and Johnson, K. (2014). *A course in phonetics [7th Ed.]*. Cengage learning, Belmont, CA.

Mehler, J., Peña, M., Nespor, M., and Bonatti, L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex*, 42(6):846–854.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., and Sams, M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage*, 30(2):563–569.

Nespor, M., Peña, M., and Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e linguaggio*, 2(2):203–230.

Nespor, M. and Vogel, I. (1986). *Prosodic phonology.* Foris, Dordrecht.

New, B., Araújo, V., and Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological Science*, 19(12):1223–1227.

Poltrock, S. and Nazzi, T. (2015). Consonant/vowel asymmetry in early word form recognition. *Journal of experimental child psychology*, 131:135–148.

Pons, F. and Toro, J. M. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition*, 116(3):361–367.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior*, 6(6):855–863.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81(1):115.

Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (2008a). Finding words and rules in a speech stream functional differences between vowels and consonants. *Psychological Science*, 19(2):137–144.

Toro, J. M., Shukla, M., Nespor, M., and Endress, A. D. (2008b). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception & psychophysics*, 70(8):1515–1525.

Tunney, R. J. and Altmann, G. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3):614.

Tuomainen, J., Andersen, T. S., Tiippana, K., and Sams, M. (2005). Audio–visual speech perception is special. *Cognition*, 96(1):B13–B22.

Appendix

Clicks used in the stimuli

key	description	IPA
t	dental voiceless unaspirated velar plosive	k|
n	dental voiceless aspirated velar plosive	k|ʰ
b	alveolopalatal voiceless aspirated velar plosive	k!ʰ
p	alveolopalatal voiced velar nasal	ŋ!
k	alveolar lateral voiceless aspirated velar plosive	ŋ||