MODULARITY IN THE CHANNEL: THE LINK BETWEEN SEPARABILITY OF FEATURES AND LEARNABILITY OF DEPENDENCIES BETWEEN THEM

Vsevolod Kapatsinski

University of Oregon, USA
vkapatsi@uoregon.edu

ABSTRACT

Moreton [10] argued for a distinction between analytic bias and channel bias in language learning. Analytic bias is defined as a set of cognitive predispositions for certain types of generalizations that constrains the learner but does not influence perception and production. Channel bias is defined as 'phonetically systematic errors in transmission between speaker and hearer'. Recent studies [6, 10] document "modularity bias", whereby dependencies between consonant features and dependencies between vowel features are easier to learn than dependencies involving a vowel feature and a consonant feature. Moreton [10, 11] claims that modularity bias is an analytic bias. In this paper, I employ the Garner interference paradigm [4] to show that the modularity bias can be caused by how acoustic cues are parsed into cognitive representations in perception [2] p.151-153, and therefore should be regarded as a channel bias.

Keywords: inductive bias, phonology, speech perception, learning, modularity

1. INTRODUCTION

Channel bias can be defined as "the effect of systematic errors in transmission between speaker and listener" [10] p.83. Moreton [10] distinguishes several subtypes of channel bias. Channel bias includes differences in magnitudes between phonetic precursors of various phonological patterns. Moreton [10] argues that this subtype of channel bias is unable to account for his data. For the purposes of the present argument, we will assume that this claim is correct[1] (see also [6] for a replication of [10] with speakers of Mandarin and Southern Min). This still leaves one other possible type of channel bias that can account for the results: "cognitive biases... in how acoustic cues are parsed into phonological representations [2] p.151-153".

In order to count as an analytic bias, the bias must be localized in the evaluation of perceived evidence for alternative generalizations, i.e., in determining the probability of a generalization being true given the data, where the data have been filtered through the perceptual system [10, 11]. Perception is assumed in [10] to be part of the "channel". Thus in order to show that a certain cognitive predisposition is an analytic bias, the bias should *only* manifest itself when the learner generalizes across multiple perceived stimuli, and not during the perception of a single stimulus. In this paper, I argue that the modularity bias does not satisfy these criteria and is more profitably considered a bias in how acoustic cues are parsed during perception.

It is uncontroversial to say that every learner is biased in favor of certain hypotheses, since every set of observations is consistent with a vast, possibly infinite, set of hypotheses [9] and since the probability of a hypothesis given the data is proportional to the probability of the data given the hypothesis *times the prior probability of the hypothesis* (Bayes' Theorem). This paper is therefore not arguing that analytic bias does not exist but rather suggests that before attributing a learnability difference to analytic bias, perceptual explanations have to be ruled out. Even if the acoustic precursors for two patterns (in the present case, vowel raising before voiced consonants and height harmony) are equally robust, one still needs to account for the perceptual (encoding) processes within the learner. In the remainder of the paper, I report two experiments showing that the features linked by easy-to-learn dependencies are less perceptually separable [4] than the features linked by hard-to-learn dependencies.

We test perceptual separability using the Garner interference paradigm [4]. In this paradigm, degree of perceptual separability between a pair of stimulus dimensions is measured by testing whether random variation on one dimension adversely influences the speed and/or accuracy of categorization of the stimuli along the other dimension. If perceivers are slower in classifying stimuli along dimension X when the stimuli vary randomly along dimension Y than when dimension Y is held constant, X and Y are said to be perceptually integral. That is, listeners have trouble selectively attending to X alone.[2]

Why should separability of dimensions influence learnability of correlations between the values of the dimensions? One possible mechanism is suggested by Warker et al. [14]., who implement modularity of phonology by mapping signal dimensions (like consonant features and speech rate) onto nodes in a neural network that contains indirect trainable connections between phonological dimensions but not between phonological and extralinguistic dimensions.

As it stands, Warker et al.'s account cannot capture the results in [6, 10] and the present study since they all involve differences in learnability among dependencies between pairs of phonological dimensions. The complete lack of connections between phonological and extralinguistic dimensions is also inconsistent with research on sociophonetics and is not biologically plausible. However, a more gradient neural-network explanation, which we shall call the Neural Distance Hypothesis, is possible.

We define Neural Distance as the number of synaptic connections (or hidden layers in an artificial neural network model) on the most direct path between two areas involved in representing two types of stimuli. Spreading activation dissipates, thus the greater the neural distance between two representations, the less irrelevant activation of one representation will tend to influence the activation of the other representation. Therefore, we should observe that dimensions whose representations are neurally distant should manifest less mutual interference in the Garner paradigm. In addition, learning a dependency between neurally-distant dimensions involves modifying a greater number of connections than learning a dependency between neurally proximate dimensions. As [14] shows, the time to learn a relationship between two representation grows with neural distance. This can intuitively be understood as trying to attribute an effect to a relatively distant cause manifesting itself through many intermediaries vs. a direct influence. Thus learning dependencies between neurally-distant dimensions should also be more difficult and time-consuming than learning relationships between neurally-proximate dimensions.

2. METHODS

The working hypotheses in the present study are that, for a CVCV stimulus, 1) random variation in the height of the second vowel will slow down categorization of the height of the first vowel more than would random variation in the voicing of the second consonant, and 2) random variation in the voicing of the first consonant will slow down categorization of the voicing of the second consonant more than would random variation in the height of the first vowel. Thus, voicing features of different segments should be integrated in perception, as should height features, while voicing should remain relatively separable from height.

The stimuli used for the present study are identical to those used in [10]. In order to avoid coarticulation between non-adjacent segments, they were synthesized using the MBROLA diphone concatenative synthesizer. For details, see [10]. Stimuli were presented over headphones.

Since the present study is concerned with perceptual integrality of consonant and vowel features, rather than response competition, the stimuli were constrained so that the consonants of each CVCV word differed in place of articulation, and vowels differed along the front/back dimension. The consonants were chosen from the set {[t], [d], [k], [g]}. The vowels were chosen from the sent {[i], [æ], [u], [o]}. These are the same vowel and consonant sets used in [10].

In Experiment I, the listeners were asked to categorize the first vowel in each word as either the vowel in ‘beet’ ([i]) or the vowel in ‘bat’ ([æ]). There were several blocks of trials. In all blocks, the first consonant in a word could be either [t], [d], [k], or [g]. In some blocks, the vowel variation blocks, the second consonant was held constant while the second vowel varied between [u] and [o]. In consonant variation blocks the second vowel was held constant while the second consonant varied between a voiced consonant and a voiceless consonant with the same place of articulation. The sequence of trials within a block was randomized separately for each subject. The sequence of blocks was counterbalanced across subjects. The counterbalancing ensured that each position in the block sequence hosts consonant variation and vowel variation blocks equally often, thus if a difference between consonant variation and vowel variation blocks is found, it cannot be attributed to the block sequence. In Experiment II, listeners were asked to categorize the second consonant in each word as either [k] or [g]. The first vowel was either fixed (in consonant variation blocks) or varied between [æ] and [i].[3] The first consonant could either be fixed or varied between [t] and [d].

Each block in each experiment consisted of 64 stimuli. Subjects were allowed to take a break between blocks. The entire experiment lasted 10-15 minutes depending on the individual subjects' reaction times. The subjects were allowed up to 10 seconds to respond. To respond, the subject pressed a button on a button box. The buttons were labeled with the segments they corresponded to and an example word for each segment. The subjects were instructed to respond as fast as possible without sacrificing accuracy. Response and reaction time were recorded. Incorrect responses and responses with reaction times that were further than three standard deviations away from the grand mean were excluded. Overall, 1.8% of responses were excluded from the analysis.

The participants in both experiments were introductory psychology students who reported being native English speakers with no history of speech, language, or hearing impairments. The participants received course credit for participation.

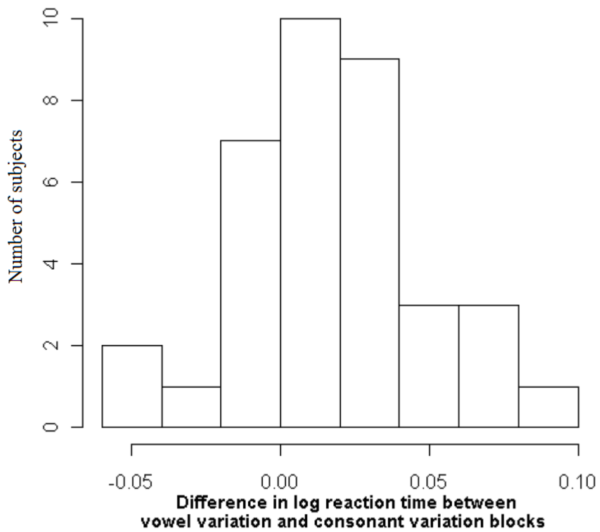
3. RESULTS AND DISCUSSION

Participants in Experiment I were slower at categorizing V_1 in a $C_1V_1C_2V_2$ as [i] vs. [æ] when V_2 varied between [u] and [o] than when the second consonant varied between [k] and [g] or [t] and [d] (single-sample Wilcoxon signed rank test on the differences between mean reaction times for consonant variation vs. vowel variation blocks, p=.0007). The distribution of reaction times is shown in Figure 1. Thus, listeners do not just find dependencies between height features of non-adjacent vowels easier to learn than dependencies between the height of a vowel and the voicing of the following consonant [6, 10]. They also find irrelevant variation in the height of V_2 more difficult to ignore than irrelevant variation in voicing of C_2, when identifying V_1. The former effect is predicted by both analytic bias and the Neural Distance Hypothesis. The latter is explained by the Neural Distance Hypothesis but not analytic bias.

In Experiment II, subjects asked to discriminate between /k/ and /g/ showed a significant reaction time effect (Wilcoxon signed rank test, $p<.05$) in the expected direction: the second consonant was classified faster when the preceding vowel varied than when the preceding consonant varied. Thus, listeners do not only find voicing harmony between C_1 and C_2 in $C_1V_1C_2V_2$ easier to learn than dependencies between V_1 height and C_2 voicing (in [6, 10]). They also find that irrelevant variation in C_1 voicing interferes with C_2 identification more than does irrelevant variation in V_1 height. The former effect is predicted by both analytic bias and the Neural Distance Hypothesis. The latter is explained by the Neural Distance Hypothesis but not analytic bias.

The weaker effect in Experiment II is consistent with the finding that the preference for consonant harmony over height-voicing dependencies is weaker than the preference for vowel harmony [10].

Figure 1: The distribution of reaction time differences between blocks with V_2 variation and blocks with C_2 variation for subjects classifying V_1 difference. A positive difference indicates that, as predicted, variation in V_2 slowed down categorization of V_1 more than did variation in C_2.[4]



4. CONCLUSION

The present results suggest that, in speech perception, consonant features interact with each other somewhat less than vowel features but more than they interact with vowel features (at least as far as height and voicing are concerned). This finding suggests that vowel and consonant features may be mapped onto somewhat different parts of the brain, making it possible to selectively attend to vowels or consonants. This is, of course, not a new idea in phonology: separate vowel and consonant tiers are the standard way to account for non-concatenative morphology [8]. The idea is also supported by fMRI data in [15], which shows that some areas of the human auditory cortex are particularly active when the subject is classifying vowels, while others are particularly active when the subject is classifying consonants.

Building on [14], we propose that learning associations between values of perceptual dimensions that map onto neurally proximate areas of the brain is easier than learning associations between values of perceptual dimensions that are neurally distant (the Neural Distance Hypothesis). Moreton (in [10]) identifies this kind of explanation as a channel bias, involving "how acoustic cues are parsed into phonological representations" [10] p.87. The Neural Distance Hypothesis can account for the results in [6, 10, 14, and 15] as well as the present data. On the other hand, the present data, and those in [13], are problematic for the hypothesis that the relative difficulty of learning vowel-consonant dependencies is due to a purely analytic bias because, by definition, analytic bias has effects only on hypothesis evaluation during learning and does not influence perception.

5. REFERENCES

[1] Ashby, F.G., Townsend, J.T. 1986. Varieties of perceptual independence. *Psychological Review* 91, 154-179.

[2] Blevins, J. 2004. *Evolutionary Phonology*. Cambridge: Cambridge University Press.

[3] Eimas, P.D., Tartter, V.C., Miller, J.L. 1981. Dependency relations during the processing of speech. In Eimas, P.D., Miller, J.L. (eds.), *Perspectives on the Study of Speech*. Hillsdale, NJ: Lawrence Erlbaum Associates, 283-309.

[4] Garner, W.R. 1974. *The Processing of Information and Structure.* New York: Wiley.

[5] Kingston, J., Macmillan, N.A. 1995. Integrality of nasalization and f1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *The Journal of the Acoustical Society of America* 97, 1261-1285.

[6] Lin, Y. 2009. Tests of analytic bias in native Mandarin speakers and native Southern Min speakers. *Proceedings of NAACL-21*, 81-92.

[7] Maddox, W.T. 1992. Perceptual and decisional separability. In Ashby, F.G. (ed.), *Multidimensional Models of Perception and Cognition*. Hillsdale, NJ: Erlbaum, 147-180.

[8] McCarthy, J. 1981. A prosodic theory of non-concatenative morphology. *Linguistic Inquiry* 12, 373-418.

[9] Mitchell, T.M. 1980. *The Need for Biases in Learning Generalizations*. Rutgers University: CBM-TR-117.

[10] Moreton, E. 2008a. Analytic bias and phonological typology. *Phonology* 25, 83-127.

[11] Moreton, E. 2008b. Modelling modularity bias in phonological pattern learning. *Proceedings of WCCFL 27*, 1-16.

[12] Mullenix, J.W., Pisoni, D.B. 1990. Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics* 47, 379-390.

[13] Obleser, J., Leaver, A.M., VanMeter, J., Rauscheker, J.P. 2010. Segregation of vowels and consonants in human auditory cortex: Evidence for distributed hierarchical organization. *Frontiers in Psychology* 2, doi: 10.3389/fpsyg.2010.00232.

[14] Warker, J.A., Dell, G.S., Whalen, C.A., Gereg, S. 2008. Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34, 1289-1295.

---

[1] This assumption means that in natural speech consonant voicing provides as much information about the height of the preceding vowel as does vowel height. Thus, if vowel height is judged on a partially syntagmatic basis, we still need to explain why heights of adjacent vowels are automatically compensated for but consonant voicing is not. The explanation could be analytic bias or channel bias.

[2] The Garner paradigm [4, 5, 12] was previously criticized on the grounds that it fails to distinguish between perceptual and decisional separability [7]. However, given that the learner purported to be subject to the modularity bias in [11] operates on the output of phonemic categorization, which involves making decisions about phoneme identity, this is not an issue for the present argument. In fact, the Garner paradigm is uniquely suited for the present study because it allows us to investigate processing of stimuli that are noise-free. Unlike alternative paradigms based on Signal Detection Theory ([1, 7]), the Garner paradigm allows us to infer separability from reaction time, rather than accuracy. Using Signal Detection Theory would require reducing identification accuracy and therefore introducing noise into the stimuli, which would make the stimuli in the present study not identical to, and therefore difficult to compare with [10].

[3] Subjects were asked to attend to the *second* consonant in CVCV because consonant voicing is supposed in [10] to influence the *preceding* vowel, and the second consonant is the only one that has a preceding vowel.

[4] We cannot be certain where on the horizontal axis a given subject falls. Therefore it is not fair to say that the behavior of subjects whose scores fall below zero contradicts the hypothesis: no individual scores are significantly below zero. Rather, as is typical with inferential statistics, we take individual subject scores to be imperfect / noisy reflections of typical processing dynamics.