Feature-specific perceptual sensitivity to within-category variations

Kuniko Nielsen[1], Rebecca Scarborough[2]

[1]Oakland University (USA), [2]University of Colorado Boulder (USA)
nielsen@oakland.edu, rebecca.scarborough@colorado.edu

ABSTRACT

This study investigates perceptual sensitivity to types of within-category variation that are produced systematically by speakers. The systematic subcategorical increase of certain features is sometimes hypothesized to be targeted at perceptual benefit. We ask, then, whether there is an asymmetry with respect to the perceptual salience of the increase vs. the decrease of phonetic features. In particular, we examine the within-category perception of voice onset time (VOT) and coarticulatory vowel nasality in English. Two perceptual experiments, an AX discrimination task and an AXB perceptual similarity task, with words with manipulated VOT and nasality were conducted. Results revealed a perceptual asymmetry of within-category variations for VOT: participants were more accurate in discriminating stimulus pairs containing increased (vs. decreased) VOT, and were less likely to identify increased VOT stimuli as similar to a natural VOT token. On the other hand, no such asymmetries were observed for coarticulatory vowel nasality in either experiment.

Keywords: perceptual asymmetry, within-category sensitivity, VOT, nasal coarticulation

1. INTRODUCTION

Speakers produce systematic within-category phonetic variation motivated by lexical and contextual factors that may relate to communicative success. For example, potentially confusable words such as those with many phonological neighbors are produced with more peripheral vowel qualities and greater nasal coarticulation (in appropriate contexts) than words with fewer neighbors [1, 2, 3]. Similarly, voiceless stop-initial words with minimal pair neighbors (e.g., pat, with minimal pair bat) are produced with longer VOT than words with no minimal pair (e.g., pal, with no minimal pair bal) [4, 5]. These adjustments resemble adaptations in pronunciation that are made in communicative contexts that explicitly demand clarity. For example, speakers produce hyperarticulated vowels and more released final stops in “clear” speech [6].

In the case of clear speech, it is generally assumed that the adjustments in pronunciation are made with the purpose of improving clarity, i.e., to result in perceptual benefit. And it has been suggested that the adjustments made in cases of lexical confusability or competition may have the same purpose: hyperarticulated vowels or increased VOT in hard words or contexts could serve to make those words more perceptible. Such effects can only be useful, though, if listeners are sensitive to the variation being produced.

In generating these effects, speakers are manipulating phonetic properties at a subcategorical level of detail. For example, a stop with a 60ms VOT and one with a 100ms VOT are both equally voiceless under the classical assumption that listeners perceive categorically. In categorical perception, while listeners are highly sensitive to differences between categories, listeners are insensitive to differences within a category (e.g., the difference between 60ms and 100ms VOT) [7, 8]. However, growing evidence indicates that, in certain tasks, listeners can hear and respond to within-category detail [e.g., 9, 10, 11]. Furthermore, McMurray et al. [10] showed that fine-grained phonetic details within phonemic categories are preserved in memory and have effects on lexical access. Crucially, such sensitivity could allow listeners to make use of the subcategorical variation produced by speakers in clear speech or confusable words.

Note, though, that it is an increase of vowel peripherality or VOT or coarticulatory nasality that is produced by speakers in harder contexts. Thus, it is the increase of these features, rather than their decrease, that could lead to perceptual benefit. It is possible that listeners might be sensitive only to within-category variation of the kind that contributes to perceptual benefit, but not to kinds of variation that do not. We ask, then, 1) to what extent listeners are sensitive to within-category variation, and 2) whether they are more sensitive to differences that may be useful than those that are not.

Both VOT and vowel nasality play a role in perception at a subcategorical level: the presence of nasality can contribute to lexical perception by predicting an upcoming nasal consonant [11], and the presence of greater VOT leads to perceptual rating of a better instance of a voiceless stop [12]. However, these features also differ in that VOT is an explicitly contrastive feature for English, serving as the primary cue to stop voicing, while vowel nasality is generally non-contrastive for English (despite the fact that it can help to cue the contrastive nasality of an adjacent nasal consonant).

Nielsen and Scarborough [13] showed that the increased presence of coarticulatory vowel nasality and VOT were more accurately discriminated in an AX (same–different) discrimination task than decreased nasality and VOT. However, that analysis included only the correct response rate for the “different” pairs, which fails to take into account response bias, rendering the results less interpretable. The current study aims to examine within-category perceptual sensitivity in VOT and vowel nasality by extending [13] in two respects: 1) by applying signal-detection analysis [14] for the AX discrimination data focused on differences, and 2) by conducting an additional experiment with an AXB task that examines perceptual similarity.

We predict that listeners should be sensitive to subcategorical variation for both of these features and that they may be more sensitive to variation where these features are increased rather than decreased. Thus, we investigate within-category perceptual sensitivity at two positions within a phonological category: where there is more of a feature (cf., hyperarticulation) and where there is less (cf., hypoarticulation). Further, since phonological contrast influences category structure, we are interested in whether these two features, which differ in phonological status, may show different patterns of perceptual sensitivity.

2. EXPERIMENT 1

2.1. Participants

Twenty-four native speakers of American English participated in a two-alternative forced-choice (AX) discrimination task. All were undergraduates and received course credit for their participation.

2.2. Stimulus selection and construction

Stimuli included 18 monosyllabic words with nasal codas and 19 monosyllabic words with onset /p/. Each set of words was recorded by a phonetically trained male native speaker of American English. Three versions of each word (i.e., more nasality/VOT, less nasality/VOT, and natural) were created. Degree of vowel nasality was increased or decreased by additively combining the waveform of the vowel in each nasal test item (e.g., ban) with the waveform of a more nasal vowel (e.g., from man) or a less nasal/oral vowel (e.g., from bad) respectively in varying ratios by formula using Praat. This process yielded tokens with intermediate spectral characteristics. Test tokens were selected to have approximately equal change in nasality, measured as ±2.5dB A1-P0. (See [15] for nasality measurement details.) VOT was increased or decreased by 40ms by splicing in aspiration from hyper-aspirated tokens or by deleting aspiration from original tokens.

2.3. Procedures

In each trial of the AX task, stimuli were presented in pairs comprising either different stimuli (e.g., moreVOT - naturalVOT) or the same stimulus presented twice. Participants indicated “different” or “same” as quickly as possible. There were 260 trials in each session, equally distributed between same and different.

2.4. Analysis

Signal detection sensitivity, measured as d-prime (d’) [14], was calculated for each condition/type for each participant. Statistical analysis of the data was based on mixed-effects modeling using the lmer function in the lme4 package for R [16], with Condition (more vs. less) and Type (nasal vs. VOT) as fixed factors, and random intercepts by Subject.

2.5. Results

Overall average d-primes (across more and less conditions) for VOT and nasal were 0.94 and 0.95, respectively, indicating that the degree of manipulation for the two types of stimuli in terms of their discriminability were comparable.

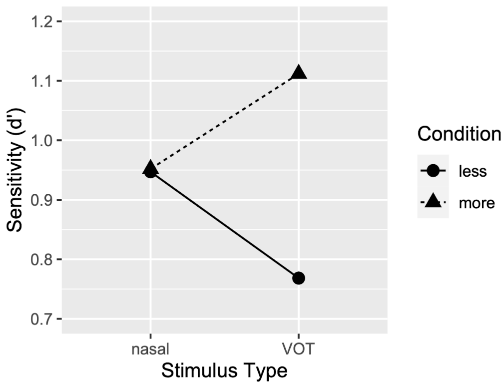

Figure 1: Summary of Experiment 1 (AX discrimination). d-prime is presented by stimulus type (Nasality vs. VOT) and condition (more vs. less).

Figure 1 summarizes the results of Experiment 1 and plots the d’ separately for the two types of stimuli and two conditions. There is a clear asymmetry between the two conditions for VOT, with a d’ of 1.11 for the more_VOT condition and a d’ of 0.77 for the less_VOT condition. However, there is no such difference for nasality, with d’ of 0.95 for both the more_nasal and less_nasal conditions. These patterns of d' were statistically assessed using linear mixed effect regressions. While there is no significant main effect observed for either Type ($t=-1.56$, $p>0.1$) or Condition ($t<1$, $p>0.1$), the interaction between Type and Condition was significant ($t=2.096$, $p<0.05$), confirming that the asymmetry between the two conditions is dependent on the stimulus type; in particular, only VOT shows a within-category asymmetry.

3. EXPERIMENT 2

3.1. Participants

Twenty-six native speakers of American English (different from those in Experiment 1) participated in an AXB perceptual similarity task. All were university undergraduates and received course credit for their participation.

3.2. Stimuli and procedures

The same stimuli from Experiment 1 were used. In each trial of the AXB task, participants heard three repetitions of the same word: an unchanged/natural token (X) and tokens with increased (more) or decreased (less) nasality or VOT (A and B). Participants indicated which of the two flanking items (A or B) sounded more similar to the middle item (X) as quickly as possible. The interstimulus interval (ISI) within triples was either 50ms or 500ms to explore effects of task difficulty.

3.3. Analysis

Responses (coded as *more* or *less*) were analyzed using Generalized Linear Mixed Effects regression (glmer function in the lme4 package for R [16]), with Type (VOT or nasality), Order ("more" as A vs. B), and ISI (50ms vs. 500ms) as fixed factors, and by-participant and by-item random intercepts.

3.4. Results

Figure 2 summarizes the results of Experiment 2 and plots the rate of "more" responses in the AXB task separately for the two types of stimuli. Overall, the "less" item and unchanged/natural item were judged as more similar in 52.6% of responses (vs. 47.4% for "more"). A significant intercept in an intercept-only model (est.=0.66, $z=-3.12$) confirms that the differences between "more" and unchanged/natural tokens were more perceptible than the differences between "less" and unchanged/natural tokens.

With respect to stimulus type, it can be seen that nasality responses are at 51% (cf., chance at 50%) with respect to 'more' (vs. 'less') responses. VOT, on the other hand, appears to show a 'less' bias, with only 42% 'more' responses (vs. 50% chance). A generalized linear mixed effect regression indicates, however, that the Type difference (nasality vs. VOT) was not significant ($z= -1.05$, $p>0.1$).

The effect of Order was significant, where A was chosen as more similar to X more often than B ($z=6.84$, $p<0.001$), but there was no main effect of ISI ($z= -0.31$, $p>0.1$) nor any significant interactions among the fixed effects ($p>0.1$).

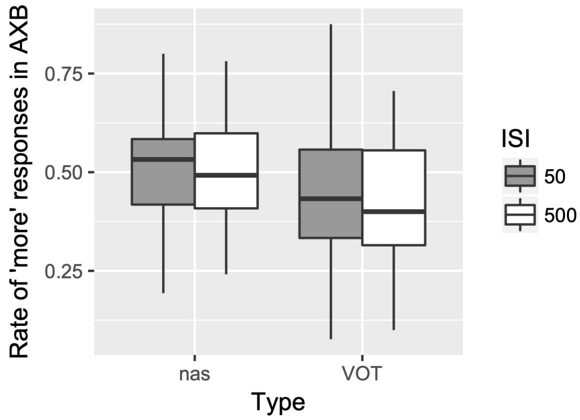

Figure 2: Summary of Experiment 2 (AXB similarity). Rate of 'more' responses (where the token with increased feature was selected as more similar to the natural token), presented by stimulus type (Nasality vs. VOT).

4. DISCUSSION

We were interested in exploring the potential perceptual consequences of subcategorical variation of the type produced systematically by listeners in various "hard" contexts. Speakers tend to hyperarticulate in such contexts, and it is generally suggested that this hyperarticulation has perceptual benefit for listeners. However, this perceptual benefit could not occur unless listeners have the ability to perceive the relevant subcategorical variation. To test this, we investigated listeners' sensitivity to variation in two phonologically relevant features (VOT and nasality) and explored a possible perceptual asymmetry between increased vs. decreased degrees of these features using an AX discrimination test (Experiment 1) and an AXB similarity test (Experiment 2).

Experiment 1 showed that listeners were more accurate in discriminating stimulus pairs containing tokens with increased VOT than pairs containing a decreased VOT token, while no difference in discrimination was observed for increased vs. decreased nasality pairs. Contrary to Nielsen and Scarborough [13], which showed that nasal tokens were discriminated more accurately than VOT tokens in the AX task, the overall discriminability for the two feature types were comparable in our data.

Experiment 2 showed that participants judged decreased-feature stimuli as more similar to the natural stimuli in both stimulus types. This suggests that an increase in these features (VOT and vowel nasality) is perceptually more salient than a decrease.

Both experiments showed listener sensitivity to within-category variation in that listeners could discriminate tokens with subcategorical differences and/or make systematic comparisons of tokens differing subcategorically. These findings contribute to a body of evidence that, despite what categorical perception would predict, listeners do have access to low-level phonetic distinctions within categories.

Both experiments also illustrated a bias in sensitivity, with greater sensitivity to variation in parts of a category that exhibit “more” rather than “less” of a feature, at least for some features. Since in at least the cases examined here, and probably more generally as well, it is an increase in a feature that is realized in contexts where clarity is required, this greater sensitivity to “more” suggests a link between within-category sensitivity and communicative benefit.

However, the asymmetrical perceptual sensitivity between “more” and “less” is shown only for VOT and not for nasality in Experiment 1 (with a non-significant difference in Experiment 2 that trends in the same direction). We consider several factors that might contribute to this feature-conditioned difference in result. In other words, why do listeners show evidence of more structured patterns of sensitivity to VOT than to vowel nasality?

An important difference between VOT and nasality lies in the phonological status of the two features. VOT is an explicitly contrastive feature in English, capturing the contrast between voiced and voiceless stops. In fact, VOT is the primary cue conveying this contrast. Vowel nasality, on the other hand, is generally not contrastive for English; rather it is the coarticulatory consequence of an adjacent nasal consonant. To the extent that vowel nasality carries information in English, then, it is information about the adjacent nasal (or oral) consonant. And there are cues other than vowel nasality that signal the nasal/oral contrast on the following consonant.

One might imagine that the contrastive status of VOT would cause the feature to be more categorical and therefore would lead to reduced perceptual sensitivity. In fact, although listeners were clearly able to attend to within-category variation in VOT, there are parts of the VOT category that listeners were not as sensitive to (namely, the reduced VOT parts of the category). With respect to nasality, listeners' sensitivity is distributed more broadly across the category.

We note an additional (probably related) difference between VOT and nasality. VOT is systematically structured in its variation, which is relatively predictable by speaker [17]. Degree of nasality, on the other hand, is quite variable across speakers, and even within speakers, as measured acoustically [18]. Not only are some speakers simply more nasal than others at their baseline, but the acoustic complexity of nasality leaves it vulnerable to influence from other properties of the signal (e.g., vowel quality or voice quality). This structured variation may contribute directly to a more structured perceptual sensitivity for VOT, which can be matched to our manipulation of increased and decreased VOT. The general variability in nasality, on the other hand, which is reflected in our manipulated tokens, may make it harder for listeners to assess what is more nasality vs. less.

The fact that this fine-grained featural difference in sensitivity is more evident in Experiment 1 (AX discrimination) than Experiment 2 (AXB similarity judgment) is unsurprising, given the much greater difficulty of the AXB task. While AX discrimination requires a single comparison (“same” or “different” between the members of a pair), AXB tasks require the listener to compare both A and B to X, and then to compare those two comparisons gradiently. Previous research has shown that such differences in task difficulty lead to perceptual difference limens that are double or more when calculated from AXB tasks, relative to AX tasks [18], indicating greater perceptual sensitivity in the AX task. Thus, the AX task allows listeners to better respond to low-level within-category acoustic differences.

5. CONCLUSION

Taken together, our results demonstrate that listeners are sensitive to within-category differences in both VOT and vowel nasality. Further, both experiments reveal an asymmetry in this subcategorical perceptual sensitivity: stimuli with increased VOT were perceived more accurately than the stimuli with decreased VOT. For nasality, however, the “more” vs. “less” asymmetry in sensitivity is weaker or absent. In other words, neither feature shows the flat within-category sensitivity that Categorical Perception would suggest; rather, listeners are sensitive to some within-category differences. They are especially sensitive to those differences where a feature is increased, rather than decreased, mirroring the systematic variation produced by speakers in hyperarticulation contexts. In other words, listeners have greatest perceptual sensitivity to those phonetic details that may be more communicatively useful.

6. REFERENCES

[1] R. Wright, "Factors of lexical competition in vowel articulation," in *Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 26–50, 2004.

[2] B. Munson and N. P. Solomon, "The effect of phonological neighborhood density on vowel articulation," *Journal of Speech, Language and Hearing Research* 47(5), 1048–1058, 2004.

[3] R. Scarborough, "Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation," *Journal of Phonetics* 41(6), 491–508, 2013.

[4] M. Baese-Berk and M. Goldrick, "Mechanisms of interaction in speech production," *Language and Cognitive Processes* 24(4), 527–554, 2009.

[5] N. P. Fox, M. Reilly., and S. E. Blumstein, "Phonological neighborhood competition affects spoken word production irrespective of sentential context," *JML* 83, 97-117, 2015.

[6] M. Picheny, N. Durlach, and L. Braida, "Speaking clearly for the hard-of-hearing II: Acoustic characteristics of clear and conversational speech," *JSLHR* 29, 434–446, 1986.

[7] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The discrimination of speech sounds within and across phoneme boundaries," *Journal of experimental psychology*, 54(5), 1957.

[8] S. Harnad, "Psychophysical and cognitive aspects of categorical perception: A critical overview," In *Categorical perception: The groundwork of cognition,* 1-52, 1987.

[9] P. K. Kuhl, "Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not." *Perception & psychophysics* 50.2, 93-107, 1991.

[10] B. McMurray, M. K. Tanenhaus, and R. N. Aslin, "Gradient effects of within-category phonetic variation on lexical access," *Cognition*, 86(2), B33-B42, 2002.

[11] P. S. Beddor, K.B. McGowan, J. E. Boland, A. W. Coetzee, and A. Brasher, "The time course of perception of coarticulation," *JASA* 133(4), 2350-2366, 2013.

[12] J. S. Allen, & J. L. Miller, "Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate," *Perception & Psychophysics,* 63(5), 798-810, 2001.

[13] K. Y. Nielsen & R. Scarborough, "Perceptual asymmetry between greater and lesser vowel nasality and VOT," Proc. of ICPhS, 2015.

[14] N. A. MacMillan and C. D. Creelman, "*Detection theory: A users guide,*" Cambridge, England: Cambridge University Press, 1991.

[15] M. Y. Chen, "Acoustic correlates of English and French nasalized vowels," JASA, 102, 2360–2370, 1997.

[16] D. Bates, M. Mächler, B. Bolker B, and S. Walker S, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software,* 67(1), 1–48, 2015.

[17] E. Chodroff and C. Wilson, "Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English," *Journal of Phonetics,* 61, 30-47, 2017.

[18] W. Styler, "On the acoustical features of vowel nasality in English and French," *JASA*, 142(4), 2469-2482, 2017.

[19] W. A. Rosenblith and K. N. Stevens, "On the DL for frequency", *JASA,* 25, 980-985, 1953.