PERCEPTUAL ASYMMETRY BETWEEN GREATER AND LESSER VOWEL NASALITY AND VOT

Kuniko Nielsen[1] & Rebecca Scarborough[2]

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

ABSTRACT

This study explores whether there is an asymmetry with respect to the perceptual salience of an increase vs. a decrease of phonologically relevant features. A forced-choice discrimination experiment was conducted on stimulus pairs that included one member with unchanged phonetic features and the other with either increased or decreased degree of features (i.e., vowel nasality or VOT). The results revealed a perceptual asymmetry: participants were more accurate and in some cases faster in their responses to stimulus-pairs containing increased-feature stimuli, suggesting that greater presence of these phonetic features is perceptually more salient than decreased presence. These findings demonstrate sub-categorical sensitivity for features known to be perceived categorically (VOT), as well as for features not primary to phonological contrast (vowel nasality). They also support a possible salience explanation for previously observed asymmetries in phonetic imitation.

Keywords: perceptual asymmetry, sub-categorical sensitivity, VOT, nasal coarticulation

1. INTRODUCTION

Previous research in speech perception shows various examples of asymmetries in the perception of phonetic features. For example, with respect to voicing, confusion matrices by Wang & Bilger [18] show that [b] is confused as [p] more often than [p] is confused as [b]. Similarly, for place of articulation, Friedrich, Lahiri & Eulitz [8] observe ERP mismatch negativity for priming from [coronal] to [dorsal], but not for [dorsal] to [coronal]. VOT goodness ratings from Allen & Miller [2] show better ratings for members of the /p/ category with longer VOT than with shorter VOT. And Lahiri & Marslen-Wilson [11] found that while nasalized vowels may be perceived as oral, oral vowels are much less frequently perceived as nasal.

Asymmetrical response with respect to VOT and nasality can be seen in quite different contexts as well. In studies of spontaneous phonetic imitation, both increased VOT and increased vowel nasality in phonologically appropriate contexts are imitated to a greater extent than decreased VOT and decreased nasality [13, 19]. In the current study, we explore whether an asymmetry in perceptual salience might contribute to the explanation of these asymmetries in imitation. In particular, if the increased presence of a given feature (e.g., vowel nasality or aspiration) is more salient than its decreased presence, increased nasality/VOT tokens should be better perceived by listeners, and therefore would be more likely to be imitated.

By looking at both vowel nasality and VOT, we also aim to examine whether the phonemic/categorical status of a phonetic feature modulates its patterns of perceptual salience. VOT is a primary cue to voicing contrasts in stops in English, and like many contrastive features, it is known to be perceived categorically [12]. The phenomenon of categorical perception demonstrates that the phonemic categories possessed by speakers influence their perception along relevant dimensions, resulting in reduced within-category sensitivity. Although previous work has shown that listeners are sensitive to vowel nasality and can use it to detect an adjacent nasal consonant (e.g., [1, 4, 6, 14]), it is generally not considered to be a contrastive feature in English. For instance, whereas Hindi listeners, whose language is considered to have contrastive vowel nasality, perceive vowel nasality categorically (like VOT in English), English listeners perceive vowel nasality more continuously [15]. In other words, in the absence of a categorical boundary, non-contrastive features such as different degrees of vowel nasality in English should be easier to discriminate than contrastive within-category features such as VOT.

2. METHODS

2.1. Participants

Forty-three native speakers of American English participated in a two-alternative forced-choice discrimination task. All were undergraduates at the University of Colorado and received course credit for their participation.

2.2. Stimuli

Eighteen monosyllabic words with nasal codas and nineteen words with onset /p/ were used as stimuli. Each set of words was recorded by a phonetically-trained male speaker. 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, yielding tokens with intermediate spectral characteristics. Test tokens were selected to have approximately equal change in nasality, measured as ±2.5dB A1-P0. (See [7] 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. Procedure

Stimuli were presented in pairs comprising either different stimuli (e.g., moreVOT-naturalVOT) or same stimuli (e.g., lessNasal-lessNasal). Inter-stimulus interval (ISI) was either 50 ms (18 participants) or 500 ms (25 participants). Participants listened to the stimuli over headphones and were instructed to indicate whether the stimuli in each pair were the same or different as quickly as possible. Careful listening was emphasized. Response, response correctness, and reaction time (RT) were recorded for each trial. Stimulus presentation and data collection were controlled using PsychoPy.

2.4. Statistical Analysis

Response measures from the data were (1) the correct responses (binary) in “different” trials and (2) RTs on correctly responded “different” trials. Statistical analysis of correctness was based on generalized linear mixed-effects modeling using the `glmer` function in the `lme4` package for R; analysis of RTs was based on linear mixed-effects modeling using the `lmer` function in the same package [3]. For both models, Condition (more vs. less), Type (nasal vs. VOT), and ISI (50 ms vs. 500 ms) were included as fixed effects, along with by-participant and by-item random intercepts and slopes for Condition. Statistical significance was determined based on p-values calculated using the `summary` function in the `lmerTest` package in R [10].

3. RESULTS

3.1. Correct Response

Table 1 presents a summary of generalized linear mixed-effects modeling for response accuracy. There are significant main effects of Condition (more vs. less) [$p<0.05$], Type (nasal vs. VOT) [$p<0.01$], and ISI (50 ms vs. 500 ms) [$p<0.01$], which indicate that stimuli with increased features (“more” tokens) received more correct responses than stimuli with decreased features (“less” tokens), that nasal stimuli were discriminated more correctly than VOT stimuli, and that stimuli presented with shorter ISI were responded to more correctly. Significant interactions between Condition and ISI [$p<0.001$] and Type and ISI [$p<0.001$] were also observed, where stimuli with increased features were responded to more correctly than stimuli with decreased features in 500 ms ISI, and nasal stimuli were responded to more correctly than VOT stimuli in 500 ms ISI. Lastly, there was a 3-way interaction of Condition, Type, and ISI [$p<0.05$], which showed that nasal tokens were discriminated more correctly than VOT tokens in “more” trials, and the effect was greater in ISI 500.

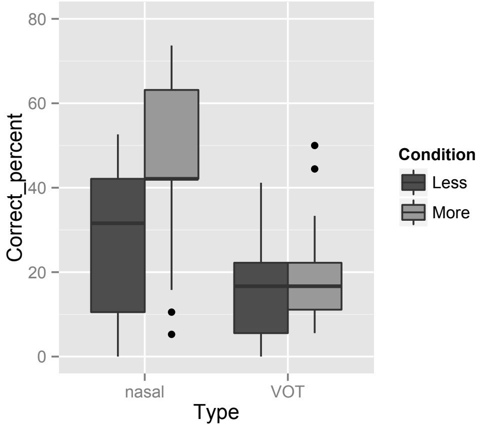
Table 1: Summary of generalized linear mixed-effects model for the correct response analysis.

	Estimate	Std.Err	z	p
(Intercept)	-0.874	0.328	-2.667	0.008
Cond=More	0.833	0.327	2.543	**0.011**
Type=VOT	-0.917	0.31	-2.959	**0.003**
ISI=500	-1.137	0.362	-3.138	**0.002**
More:VOT	-0.601	0.487	-1.236	0.217
More:500	1.072	0.285	3.759	**0.000**
VOT:500	1.263	0.291	4.347	**0.000**
More:VOT:500	-0.91	0.404	-2.253	**0.024**

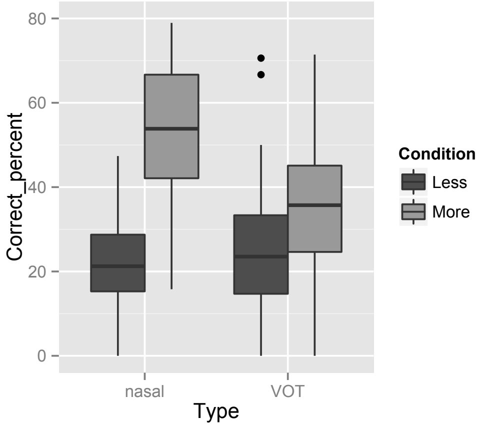
Figure 1 summarizes the correct response effects, with rate of correct response plotted by Condition (more vs. less) and Type (nasal vs. VOT). (Panel (a) shows ISI 50 ms, and panel (b) shows ISI 500 ms.) As can be seen, response accuracy was greater for “more” stimuli than for “less” stimuli in both ISIs, but the effect was greater for 500ms ISI.

Figure 1: Correct response rate by Condition (more vs. less), Type (nasal vs. VOT), and ISI (50 ms in panel (a) vs. 500 ms in panel (b)).

(a) 50ms ISI



(b) 500ms ISI



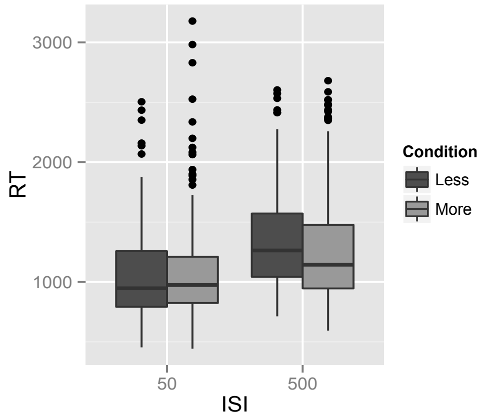
3.1. Reaction Times

Table 2 presents a summary of linear mixed-effects modeling for RTs. This model revealed a significant main effect of ISI (50 ms vs. 500 ms) [$p<0.01$], showing that participants responded faster when stimuli were presented with 50 ms ISI. Although there was no main effect of Condition, there was a significant interaction between Condition (more vs. less) and ISI [$p<0.01$], showing (as in the correctness analysis) a greater advantage for the “more” stimuli (=shorter RT, relative to “less” stimuli) at 500 ms ISI than at 50ms ISI. Figure 2 summarizes these RT effects, showing the faster RT for “more” stimuli over “less” stimuli just in the 500 ms ISI. Lastly, there was a 3-way interaction of Condition, Type, and ISI [$p<0.05$], showing that the interaction between Condition and ISI depends on Type: “more” was faster than “less” in 500 ms ISI for *nasal* words specifically. Again, this interaction pattern parallels the 3-way interaction found in the correct response analysis.

Table 2: Summary of linear mixed-effects modeling for the RT analysis.

	Estimate	Std.Err	df	t	p
(Intercept)	1107.2	57.73	62.3	19.18	0.000
Cond=More	42.6	66.34	78.3	0.642	0.523
ISI=500	294.1	72.71	84.7	4.044	**0.000**
Type=VOT	16.5	70.98	168.2	0.232	0.817
More:500	-244.4	75.98	84.6	-3.216	**0.002**
More:VOT	-30.5	100.77	216.4	-0.303	0.762
VOT:500	-160.4	85.31	590.9	-1.880	0.061
More:500: VOT	301.7	109.53	748.9	2.755	**0.006**

Figure 2: RT by Condition (more vs. less) and ISI (50 ms vs. 500 ms). Responses from 50 ms ISI were faster than those from 500 ms ISI, and the “more” advantage in RT was greater for 500 ms ISI.



4. DISCUSSION AND CONCLUSIONS

Our data showed that the increased presence of coarticulatory vowel nasality and VOT were more accurately discriminated (relative to natural) than decreased nasality and VOT. These results suggest that “more” of a phonologically relevant feature is more perceptually salient than “less” of that feature.

This asymmetrical pattern of salience makes sense in light of phonological and communicative considerations. Long lag VOT is the primary cue to the voicelessness of (initial) voiceless stops in English [12], which contrast with voiced stops, which have very short lag or 0 VOT. Similarly, although it is not a primary contrastive feature in English, vowel nasality is a useable cue to the nasality of an adjacent nasal stop (e.g., [1, 6]). Thus, in both cases, since the usefulness of the feature depends on its presence, as differentiated from its absence, sensitivity to increased presence would seem to allow for improved certainty about the contrast.

These patterns of sensitivity also mirror the patterns of realization seen in hyperspeech contexts. For instance, "clear speech" shows longer VOTs for voiceless stops, relative to conversational speech [15], and speech in a goal-oriented listener-directed task shows a greater degree of nasal coarticulation (vowel nasality) than citation speech [17]. The hypothesized communicative usefulness of these hyperspeech adjustments in production depends crucially on the perceptibility of the increased VOT or nasality. (In fact, the enhanced nasality tokens from the listener-directed context yielded better lexical decisions than tokens drawn from a context with comparable other properties but less nasality [17].)

Our results also showed that nasal tokens were discriminated more accurately than VOT tokens, and that the "more" vs. "less" asymmetry was especially present for the nasal stimuli in our data, which showed a bigger "more" advantage than VOT at 500ms ISI with respect to both accuracy and RT. Although we must interpret these differences cautiously, since the differences in degree in nasality and VOT were not calibrated relative to one another, it suggests possible greater sensitivity to differences in nasality than to VOT. If this difference is real, it would suggest that phonemic/categorical status of a phonetic feature might affect the pattern of its perceptual salience. In particular, it makes sense that there might be greater constraints on perceptual sensitivity where the feature is clearly contrastive, as VOT is, and even reported to be perceived categorically [12].

Lastly, ISI played a role in the patterning of our results: stimuli with shorter ISI were responded to faster and more accurately. This result seems to suggest that fine phonetic details may be more available in the short ISI condition. This is consistent with previous findings on the time course of phonemic or phonetic feature priming, which has shown larger priming effects in shorter ISI conditions than in longer ISI conditions [9, 16].

In the current study, though, ISI also participated in the three-way interactions for both correctness and RT: the perceptual advantage of increased features (the "more" advantage) was greater in 500ms ISI, in nasal tokens in particular. These results suggest that memory may play a role in the different patterns for more vs. less of a phonetic feature. Since there is lower within-category sensitivity for decreased features, as opposed to increased features, we expect that less detail is held in memory. Thus, if longer ISI contributes to memory decay of the first-heard item in each pair, its effect might be greater for "less" tokens that left less phonetic detail in memory to begin with, making discrimination of decreased features even more difficult after more time. Similarly, if there is lower within-category sensitivity for VOT differences than for nasality differences, we might expect similar interactions with ISI. If less VOT detail is held in memory, it might be more affected by memory decay over the longer ISI, making discrimination of VOT differences even more difficult after more time. (However, since we cannot directly compare the differences in degree of nasality and VOT, this possibility is somewhat speculative.)

Finally, the findings in the current study provide support for a possible salience explanation for the previously observed asymmetries in imitation [13, 19], in which increased VOT and increased nasality were more imitated than decreased VOT and nasality. If the increased presence of nasality or VOT is more salient to listeners, they are more likely to perceive these properties in the model speech, which could lead to greater likelihood of imitation.

5. REFERENCES

[1] Ali, L., Gallagher, T., Goldstein, J. & Daniloff, R. (1971). Perception of coarticulated nasality. *JASA* 70, 329–339.

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

[3] Bates, D. M. & Macheler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32.

[4] Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85, 785–821.

[5] Beddor, P. S. & Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *JASA* 71(6), 1551-1561.

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

[7] Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *JASA*, 102, 2360–2370.

[8] Friedrich, C. K. , Lahiri, A. & Eulitz, C. (2008). Neurophysiological Evidence for Underspecified Lexical Representations: Asymmetries With Word Initial Variations. *J. of Experimental Psych.: Human Perception and Performance*, 34(6), 1545–1559.

[9] Goldinger, S. D., Luce, P. A., Pisoni, D. B. & Marcario, J. K. (1992). Form-based priming in spoken

word recognition: the roles of competition and bias. *J. of Experimental Psych.: Learning, Memory, and Cognition*, 18(6), 1211.

[10] Kuznetsova, A., Christensen, R. H. B. & Brockhoff, P. B. (2013). “lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)”. In: R package version.

[11] Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38(3), 245-294.

[12] Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D. & Mandler, R. (1977). On pushing the voice-onset-time (VOT) boundary about. *Language and Speech*, 20(3), 209-216.

[13] Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132-142.

[14] Ohala, J. & Ohala, M. (1995). Speech perception and lexical representation: the role of vowel nasalization in Hindi and English. In B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology IV: Phonology and phonetic evidence.* Cambridge: Cambridge University Press, pp. 41–60..

[15] Picheny, M. A., Durlach, N. I. & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *JSLHR*, 29, 434–446.

[16] Radeau, M., Morais, J., & Segui, J. (1995). Phonological priming between monosyllabic spoken words. *J. of Experimental Psych.: Human Perception and Performance*, 21(6), 1297.

[17] Scarborough, R. & Zellou, G. (2013). Clarity in communication: “Clear” speech authenticity and lexical neighborhood density effects in speech production and perception. *JASA* 134, 3793-3807.

[18] Wang, M. D. & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *JASA*, *54*(5), 1248-1266.

[19] Zellou, G., Scarborough, R., & Nielsen, K. (2013). Imitability of contextual vowel nasalization and interactions with lexical neighborhood density. In *Proceedings of Meetings on Acoustics* (Vol. 19(1) p. 060083). Acoustical Society of America.