Direction of Priming and Phonetic Prototypicality in VOT Specificity Effects

Danyuan Ho[1] *& James Sneed German*[2]

[1]Division of Linguistics & Multilingual Studies, Nanyang Technological University, Singapore
[2]Aix-Marseille University, CNRS, LPL UMR 7039, 13100, Aix-en-Provence, France

hoda003@e.ntu.edu.sg, james.german@lpl-aix.fr

ABSTRACT

While previous studies have observed that the specificity effect of voice-onset time (VOT) is mediated by VOT length [1], the role of the direction of mismatch at prime versus test has not been directly explored. This study addresses this issue through a long-term repetition priming experiment that simultaneously manipulated both VOT length at test (unmodified vs. reduced) and the VOT match status (matched vs. mismatched). The results show that having an unmodified VOT and matching the VOT of the study prime were both significantly correlated with shorter reaction times at test, though unmodified VOTs were identified faster overall regardless of match status. These findings corroborate the importance of the role played by fine-grained phonetic information in word representations, and we argue that the dominance of VOT length can be explained if the malleability of word-level representations depends on the density of speech experiences across the phonetic space.

Keywords: voice onset time; specificity; phonetic detail; learning

1. INTRODUCTION

Specificity refers to the phenomenon whereby retrieval is more effective when information present at retrieval aligns with the input stimuli that gave rise to existing representations [2]. In Hintzman, Block and Inskeep [3], this was shown to extend to visual word recognition, in that words were better recognised when presented in a constant typography during test and study as compared to words with different typographies, suggesting that surface details associated to the font type is retained in the orthographic representation of words. Goldinger [4] examined the effect of speaker voice on spoken word identification and found that words produced by the same speakers are more readily recognised than those produced by different speakers, leading the author to conclude that extra-linguistic information is preserved in the representation of spoken words. McLennan, Luce & Charles-Luce [5] explored the effect of allophonic variability on sublexical ambiguity resolution and found that lexical items with matching allophonic details (e.g., [ɾ] → [ɾ] or [t] → [t] as opposed to [ɾ] → [t] or [t] → [ɾ]) resulted in better priming that their unmatched counterparts. Ju and Luce [1] further explored the role of subphonemic detail by studying how words with artificially reduced VOTs affected lexical decision during a long-term repetition priming experiment. Participants listened to primes in various levels of reduction (namely: intact, -1/3, -2/3) where half of the items at test matched those at prime in terms of VOT and the other half did not match. It was found that the specificity effect of VOT was affected by the degree of VOT modification since only the intact and -2/3 reduced tokens showed a priming advantage for their corresponding targets. In a second experiment within the same study, it was found that VOT length also affects the processing speed of the word tokens, with intact VOT being the fastest and the -2/3 VOT with the slowest processing time.

Since VOT categories are both language and dialect-specific, one aim of the present study is to corroborate the key finding of Ju and Luce [1] in a novel linguistic context, namely Singaporean English. The study also seeks, however, to directly assess whether and how the direction of mismatch interacts with VOT length at test. Understanding how the size of the specificity effect varies across these conditions is important for understanding the structure of the word-level representations themselves. More importantly, being able to estimate the robustness of the VOT specificity effect under various conditions can be useful for constructing more sophisticated follow-up studies that are specifically designed to assess the role of para-linguistic or non-linguistic factors in the word-level encoding of phonetic detail.

2. METHODOLOGY

2.1. Participants

28 Singaporeans took part in the study. The participants received a remuneration of S$10 for

their time. All were native speakers of Singapore English and reported no auditory impairments.

2.2. Stimuli

128 auditory word tokens were used for this study. These consisted of 32 /t/-initial target words, 32 filler words, and 64 non-words. Each set of targets, fillers and non-words were divided equally into bisyllabic and monosyllabic words. /t/ did not occur in any non-targets (i.e., fillers or non-words). Furthermore, since specificity for sub-phonemic detail can transfer through sub-lexical features [6], other voiceless plosives (/k/, /p/) did not appear in the target words. Target words were further divided equally according to lexical frequency (16 high, 16 low) based on their Cobuild frequency as reported in CELEX2 [7]. The mean frequency for high and low lexical frequency words are 147.13 and 1972.50 respectively.

Ju and Luce found that tokens with VOTs reduced by 2/3 of their original length were not affected by perceptual magnet effect [8] where the discrimination of the speech sounds is reduced or otherwise affected by their proximity to the prototypical value. In this study therefore, only unmodified targets and targets reduced by 2/3 of their original length were used.

A female speaker who speaks a standard variety of Singapore English provided the recording of the word and non-word targets. The stimuli were recorded with a Shure SM81 microphone into a computer using and were digitalised at sampling rate of 44.1 KHz. To create word targets with reduced VOTs, the middle 2/3 portion of the VOT interval was manually removed from the intact words with Praat [9].

2.3. Design

The experiment consisted of two phases: the study phase and the test phase. Both the study and test phases included 32 target words, half of which had reduced VOTs (*Reduced*) and half of which had unreduced VOTs (*Normal*). In the test phase, half of the target words matched those in the study phase in terms of VOT length (*Concordant*), and half did not match (*Non-concordant*).

The participants were therefore exposed to 4 experimental conditions, as shown in Table 1. In the *Normal-Concordant* condition, the VOT of the intact target words in the testing phase matched those in the study phase. In the *Normal-Non-concordant* condition, the VOT of the intact target words in the testing phase did not match those the study phase. In the *Reduced-Concordant* condition, the VOT of the shortened target words in the testing phase matched those in the study phase. Finally, in the *Reduced-Non-concordant* condition, the VOT of the shortened target words in the testing phase did not match those in the study phase. Thus, there were two different ways in which a mismatch could occur.

Table 1: Relationship of targets in the Study phase versus Test phase across the four experimental conditions.

Study block	Test block	Match status
Normal	Normal	Concordant
Normal	Reduced	Non-concordant
Reduced	Reduced	Concordant
Reduced	Normal	Non-concordant

2.4. Procedures

The experiment took place in a sound-attenuated room. The participants sat in front of a computer screen while wearing headphones. Presentation of the stimuli was controlled using the E-Prime 2.0 software [10], and participants gave their responses using a serial response box (SRBox 200A). All trials were randomised using the built-in function in E-Prime.

In the study phase, the participants completed a series of lexical decision tasks. Each block consisted of 128 trials. Each trial proceeded as follows: the symbol "***" appeared on the screen, and 500ms later audio playback of the target began. The participants responded by pressing one button for 'word' and another for 'non-word'. Reaction time was taken from the completion of the stimulus presentation to the moment when the participant made a keyboard response. If no response was given within 5000ms, the trial terminated automatically. Trials with reaction times greater than 2000ms were removed from analysis.

During the study phase, the same block was repeated four times in order to ensure ample exposure to the primes. After the completion of the study phase, the participants were instructed to take a 10-minute break during which they were supposed to solve Sudoku puzzles. The procedure in the test phase was identical to that in the study phase, with the exception that there was no repetition of blocks.

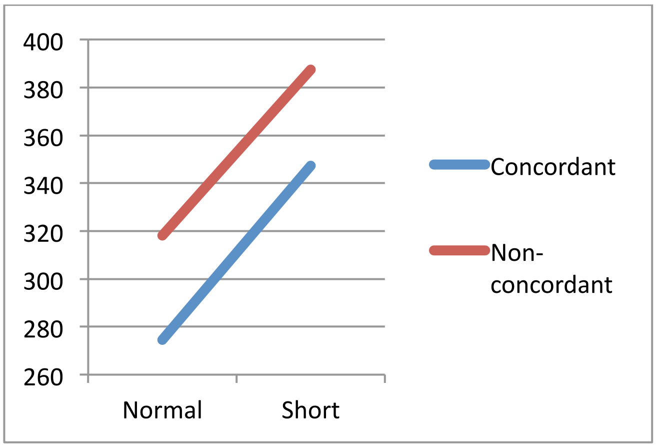
3. RESULTS

The results are summarized in Table 2. Overall, participants responded faster for items in the two Concordant conditions as compared to the Non-concordant conditions (309ms vs 354ms). Also, the participants responded faster to targets with normal VOTs compared to those with reduced VOTs (296ms vs 368ms). A visual inspection of Figure 1 reveals that concordance has a weaker effect on reaction time than VOT status. Figure 1 also suggests that the effects of the two factors are largely independent, with little or no interaction between them.

Table 2: Mean reaction time of all trials in each of the four conditions

VOT in study	VOT in training	Match status	Reaction time
Normal	Normal	Concordant	275ms (SD: 275)
Normal	Short	Non-concordant	318ms (SD: 305)
Short	Short	Concordant	347ms (SD: 277)
Short	Normal	Non-concordant	387ms (SD: 338)

Figure 1: Reaction time (ms) as a function of VOT length and match status (concordance) between study and test.



The data were analysed using linear mixed effects modelling in R [11] with the package *lme4* [12] treating subjects and items as random effects. The model with the best fit included both VOT status ($\chi2(1) = 13.87$, $p < 0.00001$) and Concordance ($\chi2(1) = 5.82$, $p < 0.05$). Including the contrast of these factors did not improve the fit of the model, confirming the descriptive observation that they represent independent effects. Syllable length also significantly improved the fit of the model ($\chi2(1) = 5.74$, $p < 0.05$) though word frequency did not ($\chi2(1) = 0.63$).

4. DISCUSSION

These results clearly reveal that both concordance and VOT length at test are important determinants of lexical access speed during lexical decision. This corroborates earlier findings [1] related to VOT specificity effects, and further shows that the size of the mismatch effect is relatively constant regardless of whether VOT length is modified or unmodified at test – in other words, the direction of priming does not seem to matter.

The role of VOT status was clearly dominant in this study, since unmodified targets showed faster reaction times than shortened targets regardless of concordance. The smaller size of the concordance effect can potentially be explained by a model in which (i) word-level representations are built from individual memories of speech events (exemplars), and (ii) word identification proceeds by phonetically matching the input to activation-weighted aggregates of those exemplars [13, 14, 15] such that the retrieval of a lexical item occurs if it is associated with both more and more highly activated exemplars than all other words. Under this view, any advantage of concordant trials over non-concordant trials was due to the fact that exposure during study had added new exemplars of a word to a specific region of the phonetic space, thereby changing their distribution over that space. If the study involved shortened tokens of a word, then these were being added to a region with relatively few exemplars, which had the effect of shifting the distribution slightly leftward. If the study phase for that word involved unmodified tokens, then these were being added to a phonetic region already dense with exemplars. A shortened word at test activated the target word more readily if it was also shortened at study (concordant), because it was a better match for the new distribution than if it was unmodified at study (non-concordant). A similar explanation applies to words that were unshortened at test.

Overall, however, since word selection is based on both the strength and *number* of activated exemplars, tokens with shortened VOTs at test were at a disadvantage compared to tokens with unmodified VOTs. Considering that VOT for voiceless plosives in English follows an approximately normal distribution [16], it follows that there are more exemplars in a listener's total experience that are close to the center of that distribution (like the unmodified tokens) than there are further from the center of that distribution (like the shortened tokens). Even though shortened tokens in the study phase served to add new exemplars to the lower tail of this distribution, this was a very small change compared to the hundreds or thousands (depending on a word's frequency) of exemplars in the listener's entire experience with the word. In short, being phonetically similar to the bulk of exemplars in the listener's experience with a word is expected to matter much more than being phonetically similar to a small number of exemplars that were added very recently.

Note that activation due to recency alone cannot explain the concordance effect, if it is assumed that selection is based on the weighted activation of *all* exemplars for a word. In that case, prior activation of exemplars of a word in *any* part of the phonetic space should contribute equally to the selection (and speed of identification) of that word. Instead, concordance effects must be due to a shift in the distribution of exemplars for that word, effectively changing what is prototypical for that word. It may well be that the recency of exemplars affects how strongly they contribute to the activation-weighting (due to, e.g., enhanced responsiveness), though to our knowledge, no existing models provide for such a mechanism. The converse is true for reduced VOTs where the relative paucity of such representations slowed down the recognition of such targets.

Overall, this study not only showed that listeners are sensitive to fined-grained details of speech information, but also that the relative dominance of the VOT length effect compared to concordance is consistent with the involvement of episodic detail in the representation of the wordform. Crucially, this study provides a baseline for the relative robustness of specificity effects across different conditions. By knowing more about the strength and behaviour of the effect, the VOT modification paradigm in a lexical decision task can be usefully extended to studies that explore the connection during encoding between phonetic detail and non-linguistic information like noise or visual information.

5. REFERENCES

[1] Ju, M. & Luce, P. 2006. Representational specificity of within-category phonetic variation in the long-term mental lexicon. *Journal of Experimental Phonology: Human Perception and Performance*, 32(1), 120-138.

[2] Tulving, E. & Thomson, D. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352–373

[3] Hintzman, D. L., Block, R., & Inskeep, N. 1972. Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior,* 11, 741-749.

[4] Goldinger, S. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166-1183.

[5] McLenan, C. T., Luce, P. A., & Charles-Luce, J. 2003. Representation of lexical form. *Journal of Exeperimental Psychology: Learning, Memory and Cognition*, 29, 539-553.

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

[7] Baayen, R., Piepenbrock, R. & Gulikers, L. 1995. CELEX2 LDC96L14. Web Download. Philadelphia: Linguistic Data Consortium.

[8] Kuhl, P. K. 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysic*s, 50, 93-107.

[9] Boersma, Paul & Weenink, David. 2015. Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved 24 March 2015 from http://www.praat.org/

[10] Psychology Software Tools, Inc. [E-Prime 2.0]. 2012. Retrieved 24 March 2015 from http://www.pstnet.com/

[11] R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

[12] Bates, D. M., Maechler, M., & Bolker, B. 2012. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.

[13] Johnson, K. 2007 Decisions and Mechanisms in Exemplar-based Phonology. In Sole, M. J., Beddor, P. & Ohala, M. (eds) *Experimental Approaches to Phonology. In Honor of John Ohala*. Oxford: Oxford University Press, 25-40.

[14] Bybee, J. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In Barlow, M. & Kemmer, S. (eds.), *Usage-Based models of language*. Stanford: CSLI Publications, 65-85.

[15] Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In, Bybee, J. & Hopper, P (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 137-157.

[16] Lisker L., and Abramson A. S. 1967. Some effects of context on voice onset time in English stops. *Language and Speech*, 10, 1–28.