SEGMENTAL VS. SUPRASEGMENTAL PROCESSING INTERACTIONS REVISITED

Satsuki Nakai & Alice Turk

Linguistics & English Language, University of Edinburgh
satsuki@ling.ed.ac.uk, turk@ling.ed.ac.uk

ABSTRACT

This paper investigates processing interactions between segmental (stop place) vs. suprasegmental (prosodic boundary) information in English using a two-choice speeded classification procedure. The results suggest that due to the presence of the boundary tonal contour, intonational phrase-boundary information and stop-place information can be processed more independently than phrase-internal, word-boundary information and stop-place information can. Possible mechanisms underlying the observed separability of the two processes are discussed.

Keywords: duration, f0, place of articulation, prosodic boundary, speech perception.

1. INTRODUCTION

Several previous studies [3, 5, 6, 7, 8] investigated processing interactions between segmental vs. suprasegmental information using the two-choice speeded classification procedure [4], which requires the listener to pay attention to one dimension (either segmental or suprasegmental) of the stimuli while ignoring the other. If the processing of the response dimension is slowed down by the variation in the irrelevant dimension, the processing of the response dimension is said to be contingent on the other dimension.

One of the early focuses of the researchers was on the relationship between the type of segmental information and patterns of processing interactions in speakers of English. Wood [8] investigated processing of stop place (/b/ vs. /g/) and pitch (high vs. low) and found unidirectional processing interaction. Specifically, variations in pitch slowed down the processing of stop place, but not *vice versa.* In contrast, Miller [6] found mutual processing interactions between vowel quality (/a/ vs. /æ/) and pitch (high vs. low) as well as loudness (soft vs. loud), and postulates that patterns of processing interactions between segmental and suprasegmental information may differ for consonants and vowels. However, Carrell et al. [3] used a range of vowel quality and pitch as response dimensions, and report that variation in the more discriminable dimension slowed down the processing of the less discriminable dimension more, suggesting that the pattern of processing interaction is not determined by the phonetic class of segmental information *per se*.

Other researchers examined possible effects of the linguistic status of suprasegmental information on the patterns of processing interactions. Repp & Lin [7] tested English and Mandarin speakers on two kinds of suprasegmental information: Mandarin tones (high vs. falling) and non-Mandarin tones (low vs. rising-falling). Lee & Nusbaum [5] also tested English and Mandarin speakers on two types of suprasegmental information: Mandarin tones (low-rising vs. falling) and pitch (high vs. low). Repp & Lin found mutual processing interactions between segmental and suprasegmental information for both Mandarin and non-Mandarin tones for both language groups. Similarly, Lee & Nusbaum's Mandarin and English speakers both exhibited mutual interactions for Mandarin tones. On the other hand, the Mandarin speakers exhibited mutual interactions for high vs. low pitch, while the English speakers exhibited unidirectional interaction, replicating Wood [8]. Based on these results, Lee & Nusbaum propose that Mandarin speakers always process f0 information integrally with segmental information because it is lexically important in Mandarin, while English speakers only do so when f0 is dynamic and thus potentially contains linguistically relevant information.

This study investigates interactions between English speakers' processing of segmental information (stop place) and suprasegmental cues that mark prosodic constituent boundaries (word and intonational phrase). If linguistic relevance of suprasegmental cues leads to mutual processing interactions, mutual interactions should be found for all response dimensions. If relative discriminability of compared dimensions determines the direction of processing interactions, systematic manipulations of the discriminability of response dimensions should affect the patterns of processing interactions.

2. EXPERIMENT 1

2.1. Stimuli

Three sets of nonce-word sequences were designed (cf. Table 1). Discriminability of segmental and suprasegmental information was varied across the stimulus sets. Stimulus sets DGW(ord) and DGP(hrase) contrasted the less discriminable segmental information (/d/ vs. /g/), and BGP(phrase), the more discriminable information (/b/ vs. /g/). Furthermore, DGW contrasted the less discriminable suprasegmental information (no boundary vs. word boundary), and DGP and BGP, the more discriminable information (no boundary vs. intonational phrase boundary).

Table 1: Experiment 1 stimulus sets.

Set	Stimuli
DGW	/gʌdlɪdʒ/, /gʌglɪdʒ/, /gʌd#lɪdʒ/, /gʌg#lɪdʒ/
DGP	/gʌdlɪdʒ/, /gʌglɪdʒ/, /gʌd%lɪdʒ/, /gʌg%lɪdʒ/
BGP	/gʌblɪdʒ/, /gʌglɪdʒ/, /gʌb%lɪdʒ/, /gʌg%lɪdʒ/

Five female speakers of Southern British English read the stimuli embedded in carrier sentences (cf. Table 2). The carrier sentences elicited contrastive stress on the second syllable of the nonce sequence for all boundary types.

Table 2: Example carrier sentences with stimuli. (Stimuli are underlined.)

Boundary type	Carrier sentences
No boundary	Jack's wife and kids live in GugLOO, and only Jack lives in GugLIDGE.
Word boundary	Jack's wife and kids live in Gug, CHAD, and only Jack lives in Gug, LIDGE.
Intonational phrase boundary	Jack has left Gug, though Jill still lives in Gug. LIDGE is where Jack lives now.

The stimuli were acoustically analysed using Praat [2] and excised from the carrier sentence (from the burst of the initial /g/ to the end of frication for /dʒ/). One token of each type of stimulus was selected on the following bases:

- All stimuli in each set were spoken by the same speaker;
- A majority of 5 native-speaker judges perceived the excised stimulus as the speaker had intended; and
- Final lengthening was present on the initial vowel /ʌ/ of the two-word stimuli (so that the cues for stop place and prosodic boundaries started around the same time).

The selected stimuli were appended to the same token of 'in' (and the following closure for /g/) taken from a stimulus that was spoken by the same speaker but was not selected to serve in the experiment.

2.2. Participants

Twelve native speakers of Southern British English each were tested on the three sets of stimuli ($N = 36$). All the participants reported no history of speech or hearing disorders.

2.3. Procedure

Each set of stimuli were grouped for two pairs of baseline blocks and two orthogonal blocks. In each pair of baseline blocks, the participants heard two stimuli that varied along one dimension (segmental or suprasegmental) and classified the stimuli along that dimension. In each orthogonal block, the participants heard all four stimuli, but classified the stimuli along one dimension (segmental or suprasegmental), ignoring the other. For instance, participants for the DGW stimulus set responded whether the stimulus they heard was 'gud…' or 'gug…' in two baseline blocks and one orthogonal block, and responded whether the stimulus was e.g. 'gudLIDGE' (one word) or e.g. 'gud LIDGE' (two words) in two baseline blocks and one orthogonal block (cf. Table 3).

Table 3: Stimulus grouping for DGW.

Response dimension: Segmental (d or g)	
Baseline 1	/gʌdlɪdʒ/ vs. /gʌglɪdʒ/
Baseline 2	/gʌd#lɪdʒ/ vs. /gʌg#lɪdʒ/
Orthogonal	/gʌdlɪdʒ/ or /gʌd#lɪdʒ/ vs. /gʌglɪdʒ/ or /gʌg#lɪdʒ/
Response dimension: Suprasegmental (1 word or 2 words)	
Baseline 1	/gʌdlɪdʒ/ vs. /gʌd#lɪdʒ/
Baseline 2	/gʌglɪdʒ/ vs. /gʌg#lɪdʒ/
Orthogonal	/gʌdlɪdʒ/ or /gʌglɪdʒ/ vs. /gʌd#lɪdʒ/ or /gʌg#lɪdʒ/

In each block, the stimuli were played 16 times each in random order. Half of the participants responded to the segmental information first, and the other half, the suprasegmental information first. The order of baseline and orthogonal blocks was counterbalanced within each response dimension across participants. The participants were instructed to respond as accurately and quickly as possible. A practice session was administered at the beginning of each block, which terminated when the error rate of less than 10 % was achieved, calculated over 10 repetitions of each stimulus.

2.4. Results

Mean Reaction Times (RTs; measured from the stimulus onset) to correct responses were calculated for each type of block and response dimension for each stimulus set (cf. Table 4). The error rates (calculated in the same way) were generally low (≤ 5.3 %). Repeated Measures ANOVAs were run on the mean RTs with Dimension (segmental vs. suprasegmental) and Block (baseline vs. orthogonal) as within-subject factors for each stimulus set. The effect of Block on RTs was significant for DGW only ($F(1,11) = 14.8$, $p = .003$). No other main effects or interactions were significant at $p < .05$.

Table 4: Mean RTs from Experiment 1 in ms (*SD*s in brackets). * indicates that the orthogonal RTs were significantly greater than the baseline RTs at $p < .05$. Mean error rates are given in percentage below the *SD*s.

Stimuli	Segmental		Suprasegmental	
	Base-line	Orthog-onal	Base-line	Orthog-onal
DGW	793 (80) 1.6 %	879* (96) 3.0 %	836 (106) 2.7 %	934* (192) 5.3 %
DGP	825 (102) 2.7 %	850 (83) 2.5 %	786 (125) 4.2 %	809 (124) 5.2 %
BGP	766 (90) 1.3 %	798 (100) 2.0 %	749 (136) 3.4 %	761 (125) 4.7 %

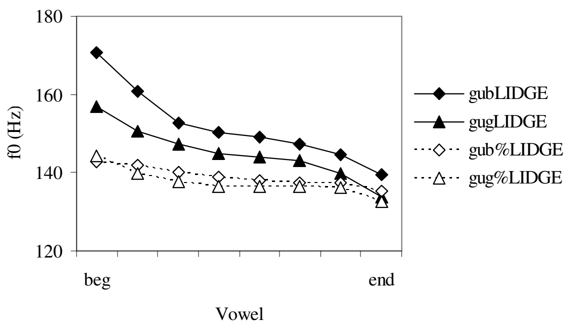
Thus, reliable processing interactions were evident for DGW but not for DGP or BGP, which was not predicted by the linguistic relevance of the response dimensions nor by the relative discriminability of compared dimensions. Pair-wise comparisons performed on DGW vs. DGP, and DGW vs. BGP, with Stimulus Set as a between-subject factor, revealed that the Stimulus Set * Block interactions were significant in both comparisons, confirming greater processing interactions for DGW than DGP or BGP. Further analyses indicated that the patterns of error rates were not significantly different between DGW vs. DGP, or DGW vs. BGP, eliminating the possibility that different speed-accuracy trade-offs were responsible for the observed difference.

3. EXPERIENT 2

3.1. Stimuli

One obvious difference between word-boundary vs. intonational phrase-boundary stimuli was that the vowels before the intonational phrase boundary were perceptibly lower in pitch than their no-boundary counterparts due to the boundary tonal contour (cf. Figure 1). In order to test whether this was responsible for the lack of significant processing interactions observed for DGP and BGP, the intonational phrase-boundary stimuli (/gʌb%lɪdʒ/, /gʌd%lɪdʒ/, /gʌg%lɪdʒ/) were resynthesised using Praat's PSOLA method [2], so that the initial vowel /ʌ/ of these stimuli had similar f0 contours to the no-boundary stimuli in the same stimulus set. As the duration of this portion of the no-boundary stimulus and that of the intonational phrase-boundary stimulus differed significantly (due to final lengthening), f0 contours of the vowel /ʌ/ of the no-boundary stimuli were stretched in time to fit the durations of the relevant portions of intonational phrase-boundary stimuli. In other words, the segmental durations of the modified intonational phrase-boundary stimuli were kept intact.

Figure 1: Time-normalised f0 contours of the vowel /ʌ/ of BGP stimuli.



3.2. Participants & Procedure

New groups of twelve native speakers of Southern British English were tested on each of the two sets of stimuli: DGPmod and BGPmod (N = 24). The procedure was identical to Experiment 1.

3.3. Results

Table 5 presents mean RTs to correct responses and error rates for each block type and response dimension. Again, the error rates were low (≤ 4.8 %). Repeated Measures ANOVAs were run on the mean RTs with Dimension and Block as within-subject factors for each stimulus set. The effect of Dimension on RTs was significant for DGPmod ($F(1,11) = 10.8$, $p = .007$), suggesting that the RTs to the suprasegmental dimension were significantly faster than those to the segmental dimension. More importantly, the effect of Block was significant for both DGPmod and BGPmod ($F(1,11) = 37.4$, $p < .001$; $F(1,11) = 15.2$, $p = .002$). No other main effects or interactions were significant. An additional mixed design ANOVA performed on RTs for DGP, BGP, DGPmod, and BGPmod confirmed that the processing interactions were significantly greater for the resynthesised stimuli (DGPmod, BGPmod) than the original intonational phrase-boundary stimuli (DGP, BGP). There were no significant differences in the error rates between the original and resynthesised stimulus sets.

Table 5: Mean RTs from Experiment 2 in ms (*SD*s in brackets). * indicates that the orthogonal RTs were significantly greater than the baseline RTs at $p < .05$. Mean error rates are given below the *SD*s.

Stimuli	Segmental		Suprasegmental	
	Base-line	Orthog-onal	Base-line	Orthog-onal
DGPmod	819 (90) 2.3 %	890* (118) 3.3 %	741 (82) 4.8 %	823* (82) 3.9 %
BGPmod	732 (67) 0.8 %	796* (68) 2.2 %	737 (95) 3.0 %	804* (149) 3.6 %

Thus, removal of the boundary tonal contours from the intonational phrase-boundary stimuli led to significant mutual processing interactions between segmental and suprasegmental dimensions. That is, the presence of the boundary tonal contours is likely to have been responsible for the lack of significant processing interactions for DGP and BGP in Experiment 1.

4. DISCUSSION

Taken together, our results suggest that information regarding stop place and prosodic boundary can be processed more independently when the boundary information is in part carried by pitch - a pattern of results that was not predicted by any of the previous studies employing the two-choice speeded classification procedure. Given that presence of boundary tonal contour and vowel durations were correlated in the no-boundary vs. intonational phrase-boundary stimulus sets, it is conceivable that the listeners used the f0 information to anticipate when the target consonants would appear in orthogonal, segmental blocks, offsetting the interference arising from durational variation. If so, however, it is unclear why processing of f0 information did not interfere with that of stop place, considering previous findings of processing interactions (unidirectional or mutual) between pitch and stop place information [5, 7, 8]. Can some types of f0 information (e.g. sentence intonation) be processed separately from segmental information? Could this be related to greater right hemisphere involvement in the processing of sentence intonation suggested for speakers of non-tonal languages (e.g. [1])? Or, do pitch and duration, only when combined, bring about the separability of segmental vs. suprasegmental processes? Future studies are being planned to address these questions.

5. REFERENCES

[1] Blumstein, S., Cooper, W. 1974. Hemispheric processing of intonation contours. *Cortex* 10, 146-158.

[2] Boersma, P., Weenink, D. Praat: Doing phonetics by computer (version 4.3.19). http://www.praat.org visited 1-Jun-05

[3] Carrell, T., Smith, L., Pisoni, D. 1981. Some perceptual dependencies in speeded classification of vowel color and pitch. *Percept. Psychophys.* 29, 1-10.

[4] Garner, W. 1974. *The processing of information and structure.* Maryland: Lawrence Erlbaum.

[5] Lee, L., Nusbaum, H. 1993. Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Percept. Psychophys.* 53 (2), 157-165.

[6] Miller, J. 1978. Interactions in processing segmental and suprasegmental features of speech. *Percept. Psychophys.* 24 (2), 175-180.

[7] Repp, B., Lin, H.-B. 1990. Integration of segmental and tonal information in speech perception: a cross-linguistic study. *J. Phon.* 18, 481-495.

[8] Wood, C. 1974. Parallel processing of auditory and phonetic information in speech discrimination. *Percept. Psychophys.* 15 (3), 501-508.