Separability of prosodic phrase boundary and phonemic information

Citation for published version:
Nakai, S & Turk, AE 2011, 'Separability of prosodic phrase boundary and phonemic information' Journal of the Acoustical Society of America, vol. 129, no. 2, pp. 966-976. DOI: 10.1121/1.3514419

Digital Object Identifier (DOI):
10.1121/1.3514419

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Journal of the Acoustical Society of America

Separability of prosodic phrase boundary and phonemic information[a)]

Satsuki Nakai[b)] and Alice E. Turk
Department of Linguistics and English Language, The University of Edinburgh, Dugald, Stewart Building, 3 Charles Street, Edinburgh, Scotland EH8 9AD, United Kingdom

(Received 5 February 2010; revised 10 August 2010; accepted 8 October 2010)

It was hypothesized that the retrieval of prosodic and phonemic information from the acoustic signal is facilitated when prosodic information is encoded by co-occurring suprasegmental cues. To test the hypothesis, two-choice speeded classification experiments were conducted, which examined processing interaction between prosodic phrase-boundary vs stop-place information in speakers of Southern British English. Results confirmed that the degree of interaction between boundary and stop-place information diminished when the pre-boundary vowel was signaled by duration and F_0, compared to when it was signaled by either duration or F_0 alone. It is argued that the relative ease of retrieval of prosodic and phonemic information arose from advantages of prosodic cue integration.
 [DOI: 10.1121/1.3514419]

PACS number(s): 43.71.Es, 43.71.Sy [MSS] Pages: 966–976

I. INTRODUCTION

Speech contains diverse types of linguistic (and non-linguistic) information, encoded by a limited set of temporally overlapping, shared acoustic cues. For example, the prosodic organization of speech is typically encoded by acoustic cues such as duration, F_0, and intensity (or spectral tilt), the so-called suprasegmentals. These cues also signal phonemic information even in languages like English that do not use these cues primarily for lexical contrast (Cole and Scott, 1974). A listener whose task is to decode a speaker's message is thus faced with a task of teasing apart different sources of linguistic information in the speech signal. To give a simplified example, a vowel can have a longer duration than other vowels in an utterance because it is a low vowel or because it is an utterance-final vowel. To arrive at the correct interpretation of the vowel duration, these two possibilities are likely to be weighed against each other in light of other evidence. A successful retrieval of phonemic and prosodic information is thus likely to require an interaction between the processing of the two types of information (Soli, 1980). This study asks whether the magnitudes of such processing interactions between phonemic and prosodic information can be, in part, determined by how prosodic information is encoded.

A behavioral measure often used to investigate processing interactions is the two-choice speeded classification task popularized by Garner (1974). The task was originally used to characterize perceptual dimensions of visual stimuli along whether they are separable (e.g., hue and shape) or integral (e.g., brightness and saturation). In this task, the participant classifies stimuli along one aspect of the stimuli (response dimension) in two conditions: baseline and orthogonal.[1] In the baseline condition, stimuli vary along the response dimension only; it is assumed that in this condition the participant only needs to attend to stimulus variation in the response dimension, as other aspects of the stimuli are kept constant. In the orthogonal condition, stimuli vary orthogonally along the response dimension and an additional dimension. Similar reaction times (RTs) to the response dimension in baseline and orthogonal conditions imply that the participant can ignore the stimulus variation in the additional dimension, and that the processing of the response dimension is separable from that of the additional stimulus dimension. In contrast, if the processing of the two stimulus dimensions interact with each other, the participant should be slower in the orthogonal condition because the additional variation in the irrelevant stimulus dimension adds cost to the processing speed.

Several past studies extended the use of this task to the investigation of processing interactions between phonemic information, such as stop place and vowel quality, and suprasegmental parameters, such as F_0 and intensity (e.g., Carrell *et al.*, 1981; Eimas *et al.*, 1981; Lee and Nusbaum, 1993; Miller, 1978; Pallier *et al.*, 1997; Repp and Lin, 1990; Wood, 1974). A general picture that emerges from these studies is that the processing of phonemic information interacts with that of suprasegmental information significantly where the two types of information overlap temporally. As discussed earlier, this is expected where two types of information are encoded by shared cues. Since the prosodic organization of an utterance is typically encoded by suprasegmentals, one would expect the processing of information about prosodic organization also to interact with the processing of phonemic information, where the two types of information overlap temporally.

At the same time, most of the above studies pitted the processing of individual suprasegmental parameters (e.g., F_0) against the processing of phonemic information (e.g., stop place), while the prosodic organization of an utterance is often encoded by a combination of suprasegmentals.

[a)]Preliminary interpretations of part of the present study were presented in "Segmental vs suprasegmental processing interactions revisited," *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, August 2007.
[b)]Author to whom correspondence should be addressed. Electronic mail: satsuki@ling.ed.ac.uk

 0001-4966/2011/129(2)/966/11/$30.00

This difference is potentially important, as it could affect the degree of processing interaction between phonemic and prosodic information, as we see in the following paragraph.

Combining information from multiple cues has been shown to produce perceptual advantages within and across sensory modalities such as audition, haptics, and vision. For example, a drawing of an object is recognized faster when appropriately colored (e.g., yellow bananas) than when it is monochrome (e.g., bananas in a black-and-white picture) (Tanaka *et al.*, 2001). In other words, object recognition is facilitated when multiple cues (shape and color in the above example) are available. We do not know of an analogous study in audition, but it seems reasonable to hypothesize that the same principle applies to audition, given theories of cue integration such as the Bayesian decision theory and the maximum likelihood estimation theory (e.g., Ernst, 2005). These theories state that learned associations between multiple cues are utilized by the perceiver to produce statistical (near-) optimality for object estimation, leading to perceptual advantages. It is plausible, then, that multiplicity of cues to prosodic information facilitates its identification and retrieval. This should also facilitate the retrieval of phonemic information, which needs to be teased apart from prosodic information.

A hint of this possibility is found in Pallier *et al.* (1997), who used the two-choice speeded classification task to study interference from variation in lexical stress position on phonemic processing. Compared to other studies that manipulated a single suprasegmental parameter, the amounts of processing interference in Pallier *et al.* (1997) appear small, suggesting the relative ease of retrieval of phonemic information in the face of varying lexical stress position. In that study, RT differences between baseline and orthogonal conditions (orthogonal RT interference) are only 9–15 ms in three out of four comparisons and 32 ms in one, though these differences were statistically significant. By contrast, other studies that manipulated a single suprasegmental parameter often report orthogonal RT interference on the order of 50–100 ms. Lexical stress information is typically encoded by multiple suprasegmental cues, including duration and intensity. The small amounts of interference from the variation in lexical stress position on phonemic processing in Pallier *et al.* (1997) might have resulted from the ease at which lexical stress position was identified, thanks to the integration of multiple cues, and phonemic information teased apart from stress information.

There is another possible reason for the small amounts of processing interference in Pallier *et al.* (1997), however. The amount of orthogonal RT interference in the speeded classification task can be affected by the relative discriminability of compared stimulus dimensions, as the stimulus dimension composed of a less discriminable contrast is easier to ignore (Carrell *et al.*, 1981; Garner and Felfoldy, 1970; but see, Eimas *et al.*, 1981). It is possible that lexical stress information was less discriminable than phonemic information in Pallier *et al.*'s (1997) stimuli, leading to small amounts of interference from the variation in lexical stress position on phonemic processing. If so, the amount of processing interference from phonemic information on lexical stress judgments would have been large. However, we do not know whether this was true, as Pallier *et al.* (1997) did not conduct stress judgments.

In the present study, we test the hypothesis that the concurrent use of multiple suprasegmental cues to signal prosodic organization facilitates the retrieval of temporally overlapping phonemic and prosodic information. To test the multiple-cue hypothesis, we used the amount of orthogonal RT interference in the two-choice speeded classification task as a diagnostic of the relative ease of retrieval of phonemic and prosodic information. Three experiments were conducted, all of which examined processing interactions between prosodic phrase-boundary and stop-place information in Southern British English. Experiment 1 used spoken nonce stimuli, in which the place information of the critical stop (e.g., *gu**d**lidge* vs *gu**g**lidge*) and boundary information (e.g., *gudlidge* vs *gud lidge*) became available around the same time (the vowel portion of *gud* and *gug*). The phrase boundary was signaled on the pre-boundary vowel with either duration alone (single cue) or a combination of duration and F_0 (multiple cues). As predicted, boundary and phonemic processing interacted significantly less in the multiple-boundary-cue than in the single-boundary-cue stimuli. Two follow-up experiments used resynthesized stimuli to verify our interpretation that the relatively small magnitudes of processing interactions found for the multiple-boundary-cue stimuli were due to the multiplicity of the cues that signaled the pre-boundary vowel.

II. EXPERIMENT 1

A. Method

1. Stimuli

Two sets of nonce-word sequences were designed (cf. Table I). Along the boundary dimension, DG# contrasted no boundary with a single-cue phrase boundary, signaled by duration (final lengthening) on the pre-boundary vowel. DG% contrasted no boundary with a multiple-cue phrase boundary, signaled by duration and F_0 (final lengthening and a boundary tonal contour) on the pre-boundary vowel. Along the phonemic dimension, both stimulus sets contrasted /d/ and /g/ before the boundary.

The test sequences were produced in carrier sentences (cf. Table II) and excised from the sentences (from the burst of the word-initial /g/ to the end of frication for /dʒ/ in *lidge*). The carrier sentences were designed to elicit contrastive

TABLE I. Stimulus sets in Experiment 1: # indicates a phrase boundary signaled by duration alone on the pre-boundary vowel; % indicates a phrase boundary signaled by duration and F_0 on the pre-boundary vowel.

Stimulus set	Boundary contrast	Phonemic contrast	Stimuli
DG#	No boundary vs #	/d/ vs /g/	/gʌ**d**lɪdʒ/, /gʌ**g**lɪdʒ/, /gʌ**d#**lɪdʒ/, /gʌ**g#**lɪdʒ/
DG%	No boundary vs %	/d/ vs /g/	/gʌ**d**lɪdʒ/, /gʌ**g**lɪdʒ/, /gʌ**d%**lɪdʒ/, /gʌ**g%**lɪdʒ/

TABLE II. Example carrier sentences. Stimuli (underlined) were excised from the carrier sentences. The words in bold capital letters carried contrastive phrasal stress.

Boundary type	Carrier sentence
No boundary	Jack's wife and kids live in Gug**LOO**, and only Jack lives in Gug**LIDGE**. /gʌglɪdʒ/
Boundary cued by duration	Jack's wife and kids live in Gug, **CHAD**, and only Jack lives in Gug, **LIDGE**. /gʌg#lɪdʒ/
Boundary cued by duration and F_0	Jack has left Gug, though Jill still lives in Gug. **LIDGE** is where Jack lives now. /gʌg%lɪdʒ/

phrasal stress on the second syllable of the test sequence, so that the presence/absence of a phrase boundary could not be guessed from stress placement (Cutler and Norris, 1988).

Five female speakers of Southern British English with no history of hearing or speech disorders read the test sequences embedded in carrier sentences four times each. The speakers were naïve to the purpose of the subsequent perception experiments and were paid for their time. The recordings were digitized at a sampling rate of 48 kHz and had 16 bit quantization.

Potential test sequences were acoustically analyzed using PRAAT (Boersma and Weenink, 2005). One token of each type of test sequence was selected for each stimulus set (DG# and DG%); all were spoken by the same speaker. These sequences had the following characteristics (cf. Fig. 2 and Table X in Appendix):

(1) In the boundary test sequence, final lengthening was present on the pre-boundary vowel (i.e., the vowel /ʌ/ was longer in the boundary than in the no-boundary test sequence).
(2) A phrase-boundary tonal contour (L%) was audible in the boundary test sequence in DG% (multiple-cue boundary stimuli).
(3) At least four out of five native speaker judges perceived the excised test sequence as intended in a self-paced forced-choice classification task.

The stimuli were created by appending the excised test sequence to a token of *in* (and the following closure for /g/) taken from the same speaker's utterance that did not contain a selected test sequence. This was done because the native speaker judges were slightly more accurate (by ca. 10%) in boundary classification when the preceding *in* was included, although the acoustic characteristics of *in* produced by each speaker were similar for no-boundary and boundary conditions.

Additional gated versions of stimuli were created for two-choice gated classification tasks. Gated classification tasks were administered alongside the speeded classification task to check if any findings from the speeded classification task were attributable to the timing difference in the processing of acoustic cues relevant to boundary and phonemic classification. The stimuli were gated in 25-ms increments, from 50 ms after the burst of the word-initial /g/ up to 550 ms into the stimulus. The 550-ms gate was long enough to include the release of the target stop of all stimuli and about half of the pause of the boundary stimuli in DG%.

2. Participants

Twelve participants were recruited for each of the two stimulus sets (24 participants in all).[2] The participants in this and the following experiments were native speakers of Southern British English with no history of speech or hearing difficulties. They were paid for their time. Their mean ages were 24 yr for DG% and 23 yr for DG#.

3. Procedure

In both speeded and gated classification tasks, the participants heard the stimuli, one at a time, and classified them into two categories along the response dimension (boundary or phoneme) in baseline and orthogonal blocks. The speeded and gated classification tasks differed mainly in two ways. In the speeded classification task, whole stimuli were presented, and both speed and accuracy were required from participants; in the gated classification task, fragments of the stimuli were presented, and participants had no time pressure.[3]

Stimuli in each set were grouped for two baseline blocks and one orthogonal block, for each response dimension (cf. Table III). In each baseline block, participants heard two of the four stimuli that varied along one (boundary or phoneme) dimension and classified the stimuli along that dimension. In the corresponding orthogonal block, the participants heard all four stimuli and classified them along the same dimension, ignoring the variation in the other dimension. For instance, when the response dimension was boundary, participants assigned to DG# responded whether the test sequence constituted one word (e.g., "GudLIDGE") or two words (e.g., "Gud, LIDGE") in two baseline blocks and one orthogonal block. When the response dimension was phoneme, the same participants responded whether the test sequence contained "Gu**d**" or "Gu**g**" in two baseline blocks and one orthogonal block. Each participant completed all six blocks [(two baseline + one orthogonal blocks) × (two response dimensions)] for the stimulus set they were assigned to. Participants assigned to DG% were told that no-boundary stimuli (e.g., gudLIDGE) and the first syllable (e.g., /gʌd/) of

TABLE III. Stimulus groupings for DG#.

Response dimension	Baseline 1	Baseline 2	Orthogonal
Boundary: One word (e.g., "GudLIDGE") vs two words (e.g., "Gud, LIDGE")	/gʌdlɪdʒ/ /gʌd#lɪdʒ/	/gʌglɪdʒ/ /gʌg#lɪdʒ/	/gʌdlɪdʒ/ /gʌglɪdʒ/ /gʌd#lɪdʒ/ /gʌg#lɪdʒ/
Phoneme: "Gu**d**…" vs "Gu**g**…"	/gʌ**d**lɪdʒ/ /gʌ**g**lɪdʒ/	/gʌ**d**#lɪdʒ/ /gʌ**g**#lɪdʒ/	/gʌ**d**lɪdʒ/ /gʌ**d**#lɪdʒ/ /gʌ**g**lɪdʒ/ /gʌ**g**#lɪdʒ/

boundary stimuli (e.g., gud%LIDGE) came from the end of a sentence, but the second syllable (/lɪdʒ/) of the boundary stimuli came from the next sentence. Those assigned to DG# were told that both no-boundary and boundary stimuli came from the end of a sentence.

Baseline and orthogonal blocks, each consisting of a speeded and a gated classification task, were grouped for two response (boundary and phoneme) dimensions. The order of the tasks, blocks, response dimensions, and the assignments of the answers to response keys were counterbalanced across participants within each stimulus set. For the speeded classification task, 16 repetitions of each stimulus pair/quadruplet were played in random order. For the gated classification task, duration-blocked stimuli were presented in random order, four times in total. To familiarize the participants with the stimuli and the task, a practice speeded classification task was given at the beginning of each block. Familiarization was accompanied by feedback and terminated when an overall error rate of less than 10% was achieved, calculated over ten repetitions of all stimuli in each block.

B. Results

1. Two-choice speeded classification task

The multiple-cue hypothesis predicts that the amounts of orthogonal RT interference in the speeded classification task are greater for DG# (single boundary cue) than DG% (multiple boundary cues) for both boundary and phonemic classification. Since we opted for a between-subjects design, a measure was taken to safeguard against drawing conclusions from data that were unduly influenced by one or two participants who produced extreme values. To this end, the difference in each participant's mean RTs to correct answers between baseline and orthogonal blocks (the amount of orthogonal RT interference) was calculated for each response dimension. For each stimulus set, participants who produced values that fell outside of ± 2 standard deviations (SDs) from the mean for one or both of the response dimensions were excluded from further analyses. Two participants from DG% (and none from DG#) were excluded. (Analyses including outlier participants were also conducted for all stimulus sets. Main conclusions that can be drawn from these additional analyses are the same.)

Table IV presents mean RTs (from the stimulus onset, including *in*) to correct responses and percent error rates in the speeded classification task. For both DG# and DG%, the mean orthogonal RTs were greater than the mean baseline RTs for both boundary and phonemic classification. At the same time, this difference was much smaller for DG% (multiple boundary cues) than for DG# (single boundary cue), for both response dimensions, consistent with the multiple-cue hypothesis. The error rates were also higher in the orthogonal than in the baseline blocks in most cases, though they were generally low ($\leq$5.3%).

A mixed-design analysis of variance (ANOVA) was run on individual participants' orthogonal–baseline mean RT differences, with response dimension as a within-subject factor and stimulus set as a between-subjects factor. The effect of stimulus set was significant: $F(1,20) = 5.6$, $p = 0.03$. Neither the effect of response dimension nor the response dimension $\times$ stimulus set interaction was significant [$F(1,20) < 1$], suggesting that the amounts of orthogonal RT interference in boundary and phonemic classification were similar for both DG# and DG%. These results are consistent with the prediction of the multiple-cue hypothesis.

The relatively small amounts of orthogonal RT interference for DG% compared to DG# do not seem to be accounted for by different speed–accuracy tradeoffs (e.g., Audley, 1960). If that were the case, the increase in error rates in the orthogonal as compared to the baseline block should be greater for DG% than for DG#. However, the mean difference in the orthogonal–baseline error rates was greater for DG# than DG% for both boundary and phonemic classification.

Different amounts of orthogonal RT interference between DG# and DG% are unlikely to be explained by differences in the relative discriminability of response dimensions (Carrell *et al.*, 1981). Had that been the case, perceptually more salient variation in the irrelevant dimension would have produced greater orthogonal interference. Assuming that the boundary information was more salient in DG% (multiple boundary cues) than in DG# (single boundary cue), the amount of orthogonal interference from boundary information on phonemic classification should be greater for DG% than for DG#, but the reverse was found.

In their study of processing interactions between phoneme and tone, Repp and Lin (1990) found a correlation between the amount of orthogonal RT interference and baseline RT duration and conclude that a greater amount of orthogonal RT interference does not necessarily indicate greater processing interactions. The observed difference in the amounts of orthogonal RT interference in this experiment does not seem to be a result of such a correlation. The

TABLE IV. The results of the two-choice speeded classification tasks in Experiment 1, excluding outliers. RTs are given in milliseconds; the standard errors of the mean are given in brackets.

			Boundary classification			Phonemic classification		
Stimulus set			Baseline	Orthogonal	Difference	Baseline	Orthogonal	Difference
DG# ($N = 12$)	RT	Mean	836	935	99	793	879	86
		(SEM)	(31)	(55)		(23)	(28)	
	% Error	Mean	1.6	3.0	1.4	2.7	5.3	2.6
DG% ($N = 10$)	RT	Mean	780	795	15	817	844	27
		(SEM)	(41)	(34)		(30)	(26)	
	% Error	Mean	1.5	1.5	0	2.3	3.2	0.9

Difference: (Mean orthogonal value) – (Mean baseline value).

baseline RTs for phonemic classification were, if anything, shorter for DG# than for DG%. Yet, orthogonal RT interference was larger for DG# in phonemic as well as boundary classification.

It is conceivable, however, that different amounts of orthogonal RT interference between DG# (single boundary cue) and DG% (multiple boundary cues) arose from earlier processing of boundary information in DG% as compared to DG#. F_0 contours of no-boundary and boundary stimuli in DG% were distinct at the onset of the critical vowel /ʌ/ (cf. Fig. 2 in Appendix), much earlier than formant transitions signaling the place of the following stop at the vowel offset. Possibly, participants in DG% used the F_0 cue available in the early part of the stimuli to predict the boundary type, before phonemic processing. If so, it follows that the relatively small amounts of orthogonal RT interference for DG% arose from the timing difference in the processing of the acoustic cues relevant to boundary and phonemic classification, rather than the multiplicity of boundary cues *per se*. We next examine whether this explanation is likely.

2. Two-choice gated classification task

For both DG# and DG%, the correct answer rates in the gated classification task increased as the gate duration increased from the shortest 50 ms (from the burst of the word-initial /g/) and reached an asymptote of ca. 90%–100% correct answer rates before the longest 550-ms gate. Table V presents the mean of each participant's "recognition point" for each block and response dimension, excluding data from the two outliers of the speeded classification task. The recognition point was defined as the first of the three consecutive gates for which the correct answer rate was 87.5% or above.[4] This point corresponded roughly to the onset of the asymptote of each participant's identification curve.

Table V shows that for both DG# and DG%, the mean recognition point is earlier for boundary than for phonemic classification in the baseline block, but this difference is greater for DG% than for DG# (42 vs 9 ms difference), possibly reflecting the early availability of F_0 boundary cues in DG%. A mixed-design ANOVA was run on individual participants' mean recognition points in the baseline block for DG# and DG%, with response dimension as a within-subject factor and stimulus set as a between-subjects factor. If the difference in the recognition points for boundary and phonemic classification is reliably larger for DG% than for DG#, we should find a significant response dimension × stimulus set interaction. However, this difference was not significant: $F(1,20)=1.4$, $p=0.26$.

TABLE V. Mean timing (in milliseconds relative to the release of the word-initial /g/) of recognition points in the gated classification task in Experiment 1, excluding outliers. The standard error is given in brackets.

	Boundary classification		Phonemic classification	
Stimulus set	Baseline	Orthogonal	Baseline	Orthogonal
DG# ($N=12$)	154	192	163	160
	(18)	(36)	(5)	(9)
DG% ($N=10$)	168	220	210	193
	(18)	(30)	(10)	(13)

Moreover, in the orthogonal block, the mean recognition point is later for boundary than phonemic classification for DG# and DG% by similar amounts (32 vs 27 ms), indicating that the smaller orthogonal RT interference found for DG% than for DG# in the speeded classification task is not likely to be due to the early arrival of the F_0 boundary cue in DG%.

C. Discussion

Overall, the results of Experiment 1 were consistent with the multiple-cue hypothesis that the amounts of processing interactions between boundary and phonemic information are smaller when the boundary information is signaled by co-occurring suprasegmental cues than when signaled by a single suprasegmental cue. Before we accept this interpretation, however, two alternative accounts must be rejected.

First, because the stimuli were spoken, the multiple- and single-boundary-cue stimuli were different from each other in aspects other than the presence/absence of F_0 boundary cues. Possibly, these uncontrolled differences in stimulus attributes, and not the presence of the F_0 boundary cue, led to the relatively small orthogonal RT interference observed for DG%. To test this possibility, we ran an identical experiment (Experiment 2) using another multiple-boundary-cue stimulus set BG%, and resynthesized versions of BG% and DG%, from which F_0 boundary cues were removed (single boundary cue). If the small amounts of orthogonal RT interference in DG% arose from the multiplicity of boundary cues, resynthesized stimulus sets with a single boundary cue should produce greater amounts of orthogonal RT interference than the original multiple-boundary-cue stimuli.

Though less likely (see, e.g., Lee and Nusbaum, 1993), another logical possibility is that small orthogonal RT interference in DG% is due to the presence of F_0 boundary cues itself, not to the multiplicity of boundary cues. If so, F_0 boundary cues should produce equally small amounts of orthogonal RT interference in the speeded classification task in the absence of duration cues. The multiple-cue hypothesis, on the other hand, predicts larger amounts of orthogonal RT interference when boundaries are cued by F_0 alone than when they are cued by a combination of F_0 and duration. This was tested in Experiment 3.

III. EXPERIMENT 2

A. Method

1. Stimuli

To test whether factors other than F_0 boundary cues led to the small amounts of orthogonal RT interference observed in DG%, we ran an identical experiment to Experiment 1 with three new stimulus sets: a multiple-boundary-cue stimulus set (BG%) and resynthesized single-boundary-cue stimulus sets (BG%mod and DG%mod). Just like DG% in Experiment 1, BG% consisted of no-boundary and boundary stimuli, signaled by duration and F_0 on the pre-boundary vowel. Unlike DG%, however, BG% contrasted /b/-/g/, instead of /d/-/g/, along the phonemic dimension (e.g., *gu**b**LIDGE* vs *gu**g**LIDGE*). The stimuli were produced by the same speaker who produced the stimuli in DG# and DG% in Experiment 1 and selected using the same criteria as before (cf. Table X and Fig. 2 in the Appendix).

BG%mod and DG%mod were created by removing the boundary tonal contours from BG% and DG%. PRAAT's PSOLA (pitch-synchronous overlap-and-add) method was used to resynthesize the F_0 contours of the stimuli, so that the pre-boundary vowel of all four stimuli had a similar F_0 contour to one of the no-boundary stimuli in each set. As the durations of the vowel /ʌ/ of the no-boundary and boundary stimuli differed considerably due to final lengthening on the latter, the F_0 contour of /ʌ/ of the no-boundary stimuli was stretched in time to fit the duration of /ʌ/ of the boundary stimuli, keeping the durations of the segments in the modified stimuli intact.

2. Participants

New groups of 12 participants each were tested on the three stimulus sets (36 participants in total). Their average age was 20 yr for BG%, 20 yr for BG%mod, and 22 yr for DG%mod.

3. Procedure

The procedure was identical to Experiment 1.

B. Results

1. Two-choice speeded classification task

As in Experiment 1, the amount of orthogonal RT interference in the speeded classification task was calculated for each participant for each response dimension, and outliers were identified. One participant was excluded from each set. Table VI presents the results of the speeded classification task for BG%, BG%mod, and DG%mod, along with those for DG% in Experiment 1, excluding outliers. Consistent with the multiple-cue hypothesis, for both boundary and phonemic classification the amounts of orthogonal RT interference were larger for BG%mod and DG%mod (single boundary cue) than for their multiple-cue counterparts (BG% and DG%).[5]

A mixed-design ANOVA was run on individual participants' orthogonal–baseline mean RT difference for DG% (from Experiment 1), BG%, BG%mod, and DG%mod, with response dimension as a within-subject factor, and F_0 (original and modified) and stop contrast (/b/-/g/ and /d/-/g/) as between-subjects factors. The effect of F_0 was significant [$F(1,39) = 11.8$, $p = 0.001$], confirming our observation that orthogonal RT interference was greater for the resynthesized, single-boundary-cue stimuli (BG%mod and DG%mod) than for the original, multiple-boundary-cue stimuli (BG% and DG%). No other main effects or interactions were significant [$F(1,39) = 1.3$, $p = 0.25$ for the effect of stop contrast; $F(1,39) < 1$ for the rest], suggesting that the amounts of increase in orthogonal RT interference were comparable for boundary and phonemic classification, and for BG%mod and DG%mod.

The increase in the orthogonal RT interference in the modified stimuli is unlikely to be due to different speed–accuracy tradeoffs between the original and modified stimulus sets. A mixed-design ANOVA (within-subject factor: response dimension; between-subjects factors: F_0 and stop contrast) run on the orthogonal–baseline difference in error rates revealed no significant effect of F_0: $F(1,39) < 1$. Neither are these results likely to be explained by differences in the relative discriminability of response dimensions between the original and modified stimulus sets. Had that been the case, we should have found smaller, not greater, orthogonal RT interference from boundary information on phonemic classification for the modified stimuli, assuming that the removal of the F_0 boundary cue had reduced the salience of boundary information in the modified stimuli. Finally, the findings do not appear to be explained by a simple correlation between baseline RT duration and the amount of orthogonal RT interference. The baseline RTs for BG%mod and DG%mod were in most cases slightly shorter than those for BG% and DG%.

In Sec. III B 2, we examine the results of the gated classification task to see whether the timing difference in the processing of acoustic cues relevant to boundary vs phonemic classification was significantly larger for the original, multiple-boundary-cue stimuli (BG% and DG%) than for the modified, single-boundary-cue stimuli (BG%mod and DG%mod), due to the presence of F_0 boundary cues in the multiple-boundary-cue stimuli. If so, the timing difference in the processing of boundary and phonemic information could explain the relatively small processing interactions found for the multiple-boundary-cue

TABLE VI. The results of the two-choice speeded classification tasks for BG%, BG%mod, DG%mod (Experiment 2), and DG% (Experiment 1), excluding outliers. RTs are given in milliseconds; the standard errors of the mean are in brackets.

Stimulus set			Boundary classification: Baseline	Boundary classification: Orthogonal	Boundary classification: Difference	Phonemic classification: Baseline	Phonemic classification: Orthogonal	Phonemic classification: Difference
BG% ($N = 11$)	RT	Mean	759	776	17	767	785	18
		(SEM)	(42)	(36)		(28)	(28)	
	% Error	Mean	1.4	2	0.6	1.9	2.5	0.6
BG%mod ($N = 11$)	RT	Mean	721	768	47	729	785	56
		(SEM)	(24)	(25)		(21)	(18)	
	% Error	Mean	0.5	1.5	0.5	1.8	2.2	0.4
DG% (Experiment 1) ($N = 10$)	RT	Mean	780	795	15	817	844	27
		(SEM)	(41)	(34)		(30)	(26)	
	% Error	Mean	1.5	1.5	0	2.3	3.2	0.9
DG%mod ($N = 11$)	RT	Mean	740	835	95	820	884	64
		(SEM)	(26)	(22)		(28)	(37)	
	% Error	Mean	1.4	2	0.6	3.2	2.6	–0.6

Difference: (Mean orthogonal value) – (Mean baseline value).

stimuli. Recall that the gated classification results of DG# and DG% in Experiment 1 did not suggest that the early arrival of F_0 cues was responsible for the relatively small amounts of orthogonal RT interference found for DG% as compared to DG#. However, the failure to find evidence for early processing of F_0 cues could be due to factors other than F_0 that differed between DG# and DG%, as the stimuli were spoken.

2. Two-choice gated classification task

As in Experiment 1, the recognition points of the stimuli in the gated classification tasks were calculated for each participant for each stimulus dimension, excluding the outlier participants in the speeded classification task. Table VII presents the mean recognition points for BG%, BG%mod, and DG%mod, with those for DG% from Experiment 1.

Despite the absence of F_0 boundary cues, the baseline recognition points for boundary classification were slightly (7–10 ms) earlier for BG%mod and DG%mod than for the original stimuli (BG% and DG%). A mixed-design ANOVA run on the baseline recognition points for the original and modified stimuli (within-subject factor: response dimension; between-subjects factor: F_0 and stop contrast) indicated a significant effect of response dimension, reflecting the overall earlier recognition points for boundary than for phonemic classification: $F(1,39) = 5.3$, $p = 0.03$. However, the $F_0 \times$ response dimension interaction was not significant [$F(1,39) < 1$], indicating that the timing difference in baseline recognition points for boundary and phonemic classification was not systematically different between the original and modified stimulus sets. It thus appears that the F_0 cues in BG% and DG% were not used early on for boundary classification in the gated classification task, even though the F_0 contours of the no-boundary and boundary stimuli in BG% and DG% were distinct from the onset of the critical vowel. Given the clear effect of F_0 cues in the speeded classification task, F_0 differences in no-boundary and boundary stimuli seem to have served as an effective boundary cue only when all or most of the contour was present.

Similarly, the mean recognition points in the orthogonal block do not exhibit any systematic differences between the original and modified stimulus sets that could explain the observed difference between them in the speeded classification task. In short, the gated classification results suggest that the relatively small orthogonal RT interference for the multiple-boundary-cue compared to the modified single-boundary-cue stimuli is unlikely to be due to the early arrival of the F_0 boundary cue.

TABLE VII. Mean timing (in milliseconds relative to the release of the word-initial /g/) of recognition points in the gated classification task in Experiment 2 (BG%, BG%mod, and DG%mod) and DG% in Experiment 1, excluding outliers. The standard errors of the mean are given in brackets.

	Boundary classification		Phonemic classification	
Stimulus set	Baseline	Orthogonal	Baseline	Orthogonal
BG%	160	161	177	186
($N = 11$)	(23)	(14)	(9)	(7)
BG%mod	150	180	166	166
($N = 11$)	(13)	(10)	(10)	(10)
DG% (Experiment 1)	168	220	210	193
($N = 10$)	(18)	(30)	(10)	(13)
DG%mod	161	180	186	182
($N = 11$)	(19)	(18)	(12)	(12)

IV. EXPERIMENT 3

In Experiment 3, stimuli that cued the phrase boundary with F_0 alone were used to test the possibility that the small amounts of orthogonal RT interference for BG% and DG% in the speeded classification task and hence the relative separability of boundary vs phoneme information in these stimulus sets had arisen from the presence of F_0 boundary cues, not the multiplicity of boundary cues. If the presence of F_0 cues were responsible for the relatively small orthogonal RT interference observed for BG% and DG%, the amounts of orthogonal RT interference for the new stimuli should also be small. The multiple-cue hypothesis, in contrast, predicts relatively large amounts of orthogonal RT interference in the speeded classification task for the new stimuli, compared to the multiple-boundary-cue stimuli, because the new stimuli have only a single cue to the boundary.

A. Method

1. Stimuli

Four stimuli (/gʌblɪdʒ/, /gʌdlɪdʒ/, /gʌb%lɪdʒ/, and /gʌd%lɪdʒ/; hereafter "BD%") were created from utterances of a different speaker from Experiments 1 and 2, who did not exhibit significant pre-boundary lengthening in the "multiple-boundary-cue" elicitation condition (cf. Table X in Appendix).[6] PRAAT's PSOLA method was used to resynthesize the F_0 contours of the stimuli, so that F_0 contours of no-boundary and boundary stimuli in BD% matched the mean F_0 contours of no-boundary and boundary stimuli in BG% and DG% in Experiments 1 and 2 (cf. Fig. 1). As in Experiment 2, the F_0 values were time-normalized to fit the length of each corresponding segment of the stimuli. Again, the test sequence was appended to a token of *in* (and the following closure for /g/) spoken by the speaker who produced the test sequence.

Reflecting the shorter durations of the boundary test sequences, the durational range of gates in the gated classification task was shorter than in Experiments 1 and 2. The longest gate was now 350 ms, which covered the first syllable of all the stimuli and roughly half of the pause of the boundary stimuli.

2. Participants

A new group of 12 participants were tested. Their mean age was 23 yr.

3. Procedure

The procedure was identical to Experiments 1 and 2.

B. Results

1. Two-choice speeded classification task

As in Experiments 1 and 2, the difference in each participant's mean RTs to correct answers between baseline and orthogonal blocks in the speeded classification task was calculated for each response dimension. No outliers were found. Table VIII

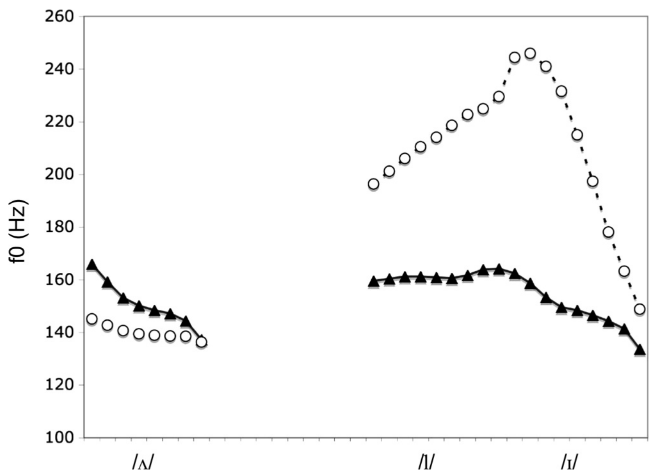

FIG. 1. Time-normalized mean F_0 contours of the /ʌ/ and /lɪ/ portions of no-boundary (▲) and boundary (○) stimuli in BG% and DG%.

presents the results of the speeded classification task. Unlike in Experiments 1 and 2, where mean baseline RTs for boundary and phonemic classification were more alike, in this experiment the mean baseline RT for boundary classification was much longer than that for phonemic classification. Nevertheless, orthogonal RT interference is relatively large for both response dimensions, consistent with the multiple-cue hypothesis.

Because there was no multiple-boundary-cue stimulus set directly comparable to BD% with respect to the stop contrast, two pairwise comparisons were performed between BD% (F_0 boundary cue) and each of the two multiple-boundary-cue stimulus sets (BG% and DG%) from Experiments 1 and 2. For each comparison, a mixed-design ANOVA was run on individual participants' orthogonal–baseline mean RT differences, with response dimension as a within-subject factor and stimulus set as a between-subjects factor. The effect of stimulus set was significant for the comparison between BD% (F_0 boundary cue) vs BG% (duration and F_0 cues) and just missed the 0.05 significance level for BD% vs DG% (duration and F_0 cues), reflecting larger orthogonal RT interference for F_0-only BD% compared to the multiple-boundary-cue stimuli: $F(1,21) = 5.6$, $p = 0.03$; $F(1,20) = 4.0$, $p = 0.06$, respectively. In neither comparison, was the response dimension × stimulus set interaction significant ($F < 1$ in both cases), suggesting comparable differences between F_0-only BD% and the multiple-boundary-cue stimuli for boundary and phonemic classification. These results are consistent with the multiple-cue hypothesis.

However, unlike the error rates in Experiments 1 and 2, error rates for BD% were generally lower in the orthogonal than in the baseline block (cf. Table VIII). Pairwise comparisons of differences in error rates between orthogonal and baseline blocks (within-subject factor: response dimension; between-subject factor: stimulus set) indicated a near trend for the effect of stimulus set for the comparison of BD% (F_0 boundary cue) vs BG% (duration and F_0 cues), with a smaller mean difference for BD% than for BG% (−0.3% vs 0.6%): $F(1,21) = 2.4$, $p = 0.13$. It is therefore possible that different speed–accuracy tradeoffs between BD% and BG% skewed the results of the RT analyses above in favor of the multiple-cue hypothesis.

To check if this is the case, a comparison of RT results between BD% and BG% was repeated, with a subset of the participants who made similar numbers of errors in the baseline and orthogonal blocks. The effect of stimulus set in a comparison of the orthogonal–baseline difference in error rates of these participants was not significant [$F(1,11) < 1$]. The reanalysis of RTs of these participants still indicated greater amounts of orthogonal RT interference for BD% (F_0 boundary cue) than for BG% (duration and F_0 cues) [89 vs 11 ms on average; the effect of stimulus set: $F(1,11) = 10.2$, $p = 0.008$]. Thus, different speed–accuracy tradeoffs do not appear to account for the greater amounts of orthogonal RT interference for BD% compared to BG%.

Again, these results are unlikely to be explained by the difference in the relative discriminability of the response dimensions between the stimulus sets. Had that been the case, the amount of orthogonal RT interference from the boundary information on the phonemic classification would have been smaller for BD% (F_0 boundary cue) than for BG% and DG% (duration and F_0 cues), assuming that boundary information was more salient in BG% and DG% than in BD%. Finally, a simple correlation between baseline RTs and the amount of orthogonal RT interference would not wholly explain the greater orthogonal RT interference found for BD% (F_0 boundary cue) compared to BG% and DG% (duration and F_0 cues). The mean baseline RT for the phonemic classification is, if anything, smaller for BD% than for DG%, but the amount of orthogonal RT interference on phonemic classification was greater for BD% than for DG%.

2. Two-choice gated classification task

As in Experiments 1 and 2, the recognition points of the stimuli in the gated classification task were calculated. As Table IX shows, the correct answer rates for boundary classification reached an asymptote much later (by ca. 150 ms) than for phonemic classification, even though F_0 differed between no-boundary and boundary stimuli from the onset of the critical vowel (cf. Fig. 1). A repeated-measures ANOVA (within-subject factors: block and response dimension) indicated that the recognition points for boundary and phonemic information

TABLE VIII. The results of the two-choice speeded classification task in Experiment 3. RTs are given in milliseconds; the standard errors of the mean are in brackets. $N = 12$.

		Boundary classification			Phonemic classification		
		Baseline	Orthogonal	Difference	Baseline	Orthogonal	Difference
RT	Mean	949	1022	73	787	839	52
	(SEM)	(37)	(51)		(23)	(34)	
% error	Mean	1.9	1.6	−0.3	0.8	0.7	−0.1

Difference: (Mean orthogonal value) − (Mean baseline value).

TABLE IX. Mean timing (in milliseconds relative to the release of the word-initial /g/) of recognition points in the gated classification task in Experiment 3. The standard errors of the mean are given in brackets. $N = 12$.

Boundary classification		Phonemic classification	
Baseline	Orthogonal	Baseline	Orthogonal
283	277	121	131
(19)	(15)	(6)	(14)

were significantly different [$F(1,11) = 123.4$, $p < 0.001$], which is mirrored in the RT difference between boundary and phonemic classification in the speeded classification task. Neither the effect of block [$F(1,11) < 1$] nor the block × response dimension interaction was significant [$F(1,11) = 1.5$, $p = 0.24$], suggesting that the recognition points for boundary and phonemic classification were similar between baseline and orthogonal blocks.

The mean recognition point of ca. 280 ms (from the burst for the /g/) for boundary classification suggests that listeners, instead of using the F_0 cues, waited till they heard part of the following pause in making the boundary judgment in the gated classification task (cf. Table X in Appendix). In other words, listeners do not seem to have used the "boundary" F_0 contour imposed on the pre-boundary vowel that is unaccompanied by final lengthening for boundary classification. It is likely that imposing the F_0 contour of a vowel on another vowel does not give the same perception of intonation when the durations of the two vowels are not comparable (see Streeter, 1978).

Despite the 150-ms difference in the recognition points for phonemic and boundary classification in the gated classification task, orthogonal RT interference was present for both types of classification in the speeded classification task. This is not surprising, however, considering the results of the study by Newman and Sawusch (1996) on rate normalization. In that study, the perceptual category boundary location between phonemic contrasts such as /ʃ/-/tʃ/ was affected by the duration of a segment within a stretch of up to around 300 ms, suggesting that speech segments within a temporal window of up to 300 ms can be analyzed together by the perception system.

V. CONCLUSION

In this study, we tested a hypothesis that the use of co-occurring suprasegmental cues to prosodic organization leads to the relative ease of retrieval of temporally overlapping prosodic phrase-boundary and phonemic information. To this end, we examined the degree of processing interactions between phrase-boundary and stop-place information in speakers of Southern British English, on the assumption that small processing interactions between the two types of information lead to an ease of their retrieval. Three experiments were conducted. Each experiment consisted of a two-choice speeded classification task, used to measure the amount of interaction between boundary and phonemic processing; and a gated classification task, used to estimate the timing at which acoustic information necessary for the speeded classification task became available in the stimuli.

Overall, the results of the three experiments were consistent with the multiple-cue hypothesis. Experiment 1 revealed that the processing of phrase-boundary information and the place information of a pre-boundary stop indeed interacted less when the pre-boundary vowel was cued by F_0 (a boundary tonal contour) and duration (final lengthening), compared to when the pre-boundary vowel was cued by duration alone. In Experiment 2 the removal of the F_0 cue from the multiple-cue boundary stimuli led to greater processing interactions between boundary and stop-place information. In Experiment 3, where the pre-boundary vowel was cued by F_0 alone, processing interactions were again greater than those found for multiple-boundary-cue stimuli. The observed differences between the multiple- and single-boundary-cue stimulus sets could not be attributed to known experimental artefacts, such as speed–accuracy tradeoffs, baseline RT duration, the relative discriminability of response dimensions, and the relative timing at which relevant acoustic cues were processed (as measured by gated classification tasks).

As discussed in the introduction, we think that the perceptual advantages observed for the multiple-boundary-cue stimuli in the orthogonal block of the speeded classification task arose from listener integration of multiple cues to the phrase boundary. The integration of multiple cues to the phrase boundary facilitates the identification of boundary information, in a similar way in which cue integration produces perceptual advantages in other sensory modalities such as vision. The ease of identification of boundary information is accompanied by an ease of retrieval of phonemic information, which needs to be teased apart from boundary information.

We note that the stimulus types used in this study are limited, and more work needs to be done to confirm the multiple-cue hypothesis. If we are correct, the retrieval of temporally overlapping phonemic and phrasal stress information, for example, should be slowed down by the removal of some of the acoustic cues to phrasal stress. We also note that not all the stimuli were ideally constructed, particularly those in Experiment 3, as we wanted to use spoken stimuli as far as possible. Because we opted for spoken stimuli, not all aspects of the stimulus attributes were systematically controlled. It is therefore possible that experimental artifacts such as discriminability of stimulus dimensions may have had some effects on the results where the effects were not evident. Future studies should vary these factors more systematically and separate possible effects arising from them.

Assuming for now that we are on the right track, we suggest that there are two different ways in which cue redundancy facilitates speech perception. One advantage is that it makes the speech signal robust in noise. A typical example of this is the cue redundancy resulting from co-articulation between an adjacent consonant and vowel. Gestural overlap in a vowel-stop sequence produces formant transitions out of the vowel and into the stop closure, providing an extra cue to the stop's place of articulation, in addition to the stop burst. This increases the chance for the listener to identify the stop place from the acoustic signal in the presence of noise, as one of the cues may survive the noise even if the other does not (Wright, 2004). Indeed, studies suggest that listeners can identify speech segments with high accuracy from small spectro-temporal regions where the speech signal is least affected by background noise. Cooke (2006), for instance, reports around 80% identification accuracy for the consonant in VCV

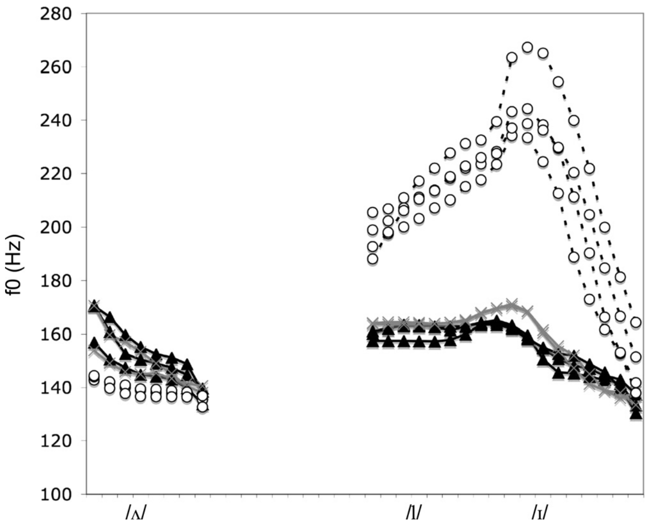

FIG. 2. Time-normalized F_0 contours of the vowel /ʌ/ and following /lɪ/ of spoken stimuli in Experiments 1 and 2. Solid lines with filled triangles represent F_0 contours of no-boundary stimuli; gray lines with crosses, single-cue boundary stimuli; and broken lines with open circles, multiple-cue boundary stimuli.

sequences from less than 25% of "clean" spectrotemporal regions (regions where the local signal-to-noise ratio exceeded 3 dB). It is likely that cue redundancy arising from co-articulation contributes to such high identification accuracy.

The perceptual advantage arising from cue redundancy dealt with in this study is different in kind. Co-occurring suprasegmental cues to prosodic phrase-boundary information seem to help the listener tease apart phrase-boundary information from phonemic information encoded together in the speech signal. If this perceptual advantage arises from the general principle of cue integration, that is, statistical (near-) optimality for object estimation produced by combining available evidence, it follows that the same kind of perceptual advantage should also arise from the multiplicity of cues to other types of linguistic information, including phonemic information. For example, all else being equal, short vs long vowel phonemes are likely to be retrieved more easily from the speech signal when they are associated with quality differences (e.g., many Swedish vowels) than when they are not (e.g., Finnish vowels, as described traditionally).

Does this mean that the multiplicity of cues facilitates the retrieval of different types of linguistic information from the speech signal in all circumstances? Probably not. For example, a Japanese study by Nakai and Turk (2010)[7] suggests that utterance-final lengthening can have a detrimental effect on the recoverability of phonemic length information, even in the presence of other acoustic cues including F_0 cues, when final lengthening leads to durational overlap between short and long phonemes across phrase-medial and utterance-final positions. This suggests that the retrieval of a given type of information is likely to be facilitated by an additional cue only when the additional cue makes the distribution of that information more distinct from other types of information in the acoustic space. In other words, whether the multiplicity of cues brings about perceptual advantages hinges on how the cues are exploited to encode other types of information.

ACKNOWLEDGMENTS

We thank Mitch Sommers and two anonymous reviewers for their very insightful comments on an earlier version of this article, and Ellen Bard for advice on statistical issues. We also acknowledge the Leverhulme Trust for funding this study (Grant No. F/00158/AF).

APPENDIX: ACOUSTIC MEASUREMENTS OF STIMULI

TABLE X. Durational measurements of spoken stimuli (in milliseconds).

Experiment	Stimulus	ʌ	C	Pause[a]	l	ɪ	d	ʒ
Experiment 1	gudLIDGE	114	90	—	99	141	108	212
	gugLIDGE	118	92	—	84	169	126	205
	gud#LIDGE	182	115	232	71	174	107	197
	gug#LIDGE	195	94	203	87	156	122	179
	gud%LIDGE	208	61	451	85	161	48	79
	gug%LIDGE	218	76	462	72	151	64	78
Experiment 2	gubLIDGE	106	78	—	87	145	93	191
	gugLIDGE (same as Experiment 1)	118	92	—	84	169	126	205
	gub%LIDGE	198	50	473	75	174	52	55
	gug%LIDGE	204	67	547	79	147	60	68
Experiment 3	gubLIDGE	130	73	—	66	134	81	138
	gudLIDGE	136	100	—	68	124	95	144
	gub%LIDGE	133	76	180	96	89	69	46
	gud%LIDGE	141	88	214	95	83	78	36

[a]Pause duration includes the frication after the stop closure.

[1]Earlier studies (e.g., Wood, 1974) included a third condition where two stimulus dimensions varied in a correlated manner. Due to the difficulty of interpretation, however, many later studies either do not fully rationalize the results of the correlated condition, or exclude this condition (e.g., Eimas *et al.*, 1978; Repp and Lin, 1990; Soli, 1980; Tomiak *et al.*, 1987).

[2]Different groups of randomly selected participants were used for each stimulus set, as the experiment lasted between 1 and 1.5 h per stimulus set, a large portion of which was spent on the gated classification task. The confound between the effect of stimulus set and that of participant group will be taken into consideration in the comparisons of the results for different stimulus sets.

[3]In the gated classification task, participants also gave confidence ratings after classifying each stimulus. These ratings generally mirrored the correct answer rates; we will not report the ratings for brevity.

[4]Each gate was presented only eight times in each block. Thus, 87.5% correct answer rate meant that the participant made just one mistake with a particular gate along the response dimension in a specified block.

[5]Interestingly, baseline RTs for boundary classification were shorter for the modified (BG%mod and DG%mod) than the original (BG% and DG%) stimuli. We do not have a full explanation for this, but suspect that the earlier "recognition points" of the modified stimuli (see below) are at least partially responsible.

[6]We could not use the same stop contrasts as in Experiments 1 and 2 (/b/-/g/ and /d/-/g/), partly because some of this speaker's utterances were not classified by the judges as the speaker had intended in boundary classification and partly because many of her pre-boundary vowels were creaky, which were avoided as they could confound the results.

[7]Nakai, S., and Turk, A. E. (2010). "Phonological word-final vowel shortening in Tokyo Japanese: a phonetic versus markedness account," Manuscript submitted for publication.

Audley, R. J. (**1960**). "A stochastic model for individual choice behavior," Psychol. Rev. **67**, 1–15.

Boersma, P., and Weenink, D. (**2005**). "PRAAT: Doing phonetics by computer [computer program]," http://www.fon.hum.uva.nl/praat/ (Last viewed February 3, 2010).

Carrell, T. D., Smith, L. B., and Pisoni, D. B. (**1981**). "Some perceptual dependencies in speeded classification of vowel color and pitch." Percept. Psychophys. **29**, 1–10.

Cole, R. A., and Scott, B. (**1974**). "Toward a theory of speech perception," Psychol. Rev. **81**, 348–374.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Cutler, A., and Norris, D. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

Eimas, P. D., Tartter, V. C., and Miller, J. L. (**1981**). "Dependency relations during the processing of speech," in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ), pp. 283–309.

Eimas, P. D., Tartter, V. C., Miller, J. L., and Keuthen, N. J. (**1978**). "Asymmetric dependencies in processing phonetic features," Percept. Psychophys. **23**, 12–20.

Ernst, M. O. (**2005**). "A Bayesian view on multimodal cue integration," in *Perception of the Human Body From the Inside Out*, edited by G. Knoblich, I. Thornton, M. Grosjean, and M. Shiffrar (Oxford University Press, Oxford), pp. 105–131.

Garner, W. R. (**1974**). *The Processing of Information and Structure* (Lawrence Erlbaum, Maryland), pp. 97–177.

Garner, W. R., and Felfoldy, G. L. (**1970**). "Integrality of stimulus dimensions in various types of information processing," Cognit. Psychol. **1**, 225–241.

Lee, L., and Nusbaum, H. C. (**1993**). "Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese," Percept. Psychophys. **53**, 157–165.

Miller, J. L. (**1978**). "Interactions in processing segmental and suprasegmental features of speech," Percept. Psychophys. **24**, 175–180.

Newman, R. S., and Sawusch, J. R. (**1996**). "Perceptual normalization for speaking rate: Effects of temporal distance," Percept. Psychophys. **58**, 540–560.

Pallier, C., Cutler, A., and Sebastian-Galles, N. (**1997**). "Prosodic structure and phonetic processing: A cross-linguistic study," in *Proceedings of Eurospeech*, ISCA, Rhodes, vol. **4**, pp. 2131–2134.

Repp, B. H., and Lin, H. B. (**1990**). "Integration of segmental and tonal information in speech perception: A cross-linguistic study," J. Phonetics **18**, 481–495.

Soli, S. D. (**1980**). "Some effects of acoustic attributes of speech on the processing of phonetic feature information," J. Exp. Psychol. Hum. Percept. Perform. **6**, 622–638.

Streeter, L. A. (**1978**). "Acoustic determinants of phrase boundary perception," J. Acoust. Soc. Am. **64**, 1582–1592.

Tanaka, J., Weiskopf, D., and Williams, P. (**2001**). "The role of color in high-level vision," Trends Cogn. Sci. **5**, 211–215.

Tomiak, G. R., Mullennix, J. W., and Sawusch, J. R. (**1987**). "Integral processing of phonemes: Evidence for a phonetic mode of perception," J. Acoust. Soc. Am. **81**, 755–764.

Wood, C. C. (**1974**). "Parallel processing of auditory and phonetic information in speech discrimination," Percept. Psychophys. **15**, 501–508.

Wright, R. (**2004**). "A review of perceptual cues and cue robustness," in *Phonetically Based Phonology*, edited by B. Hayes, R. M. Kirchner, and D. Steriade (CUP, Cambridge), pp. 34–57.