Asymmetric processing of durational differences – Electrophysiological investigations in Bengali

Adam C. Roberts*, Sandra Kotzor, Allison Wetterlin, Aditi Lahiri

University of Oxford, Faculty of Linguistics, Philology & Phonetics, Walton Street, Oxford OX1 2HG, UK

ARTICLE INFO

Article history:
Received 3 October 2013
Received in revised form
25 March 2014
Accepted 26 March 2014
Available online 12 April 2014

Keywords:
Duration
Semantic priming
Speech perception
Asymmetry

ABSTRACT

Duration is used contrastively in many languages to distinguish word meaning (e.g. in Bengali, [pata] ‘leaf’ vs. [pat-a] ‘whereabouts’). While there is a large body of research on other contrasts in speech perception (e.g. vowel contrasts and consonantal place features), little work has been done on how durational information is used in speech processing. In non-linguistic studies of low-level processing, such as visual and non-linguistic acoustic pop-out tasks, an asymmetry is found where additional information is more readily detected than missing information. In this study, event-related potentials were recorded during two cross-modal auditory-visual semantic priming studies, where nonword mispronunciations of spoken prime words were created by changing the duration of a medial consonant (real word [dana] ‘seed’ > nonword *[dan:a]). N400 amplitudes showed an opposite asymmetric pattern of results, where increases in consonantal duration were tolerated and led to priming of the visual target, but decreases in consonantal duration were not accepted. This asymmetrical pattern of acceptability is attributed to the fact that a longer consonant includes all essential information for the recognition of the original word with a short medial consonant (a possible default category) and any additional information can be ignored. However, when a consonant is shortened, it lacks the required durational information to activate the word with the original long consonant.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

1. Introduction

Humans are able to make comparative judgements about two auditory stimuli to a high degree of accuracy; discrimination of two tone stimuli can be made down to a 7 ms difference and empty intervals can be discriminated within 20 ms (Rammsayer & Lima, 1991). In speech, small durational differences can lead to categorical binary contrasts, e.g. VOT (voice onset time) differences, English pin (~ 30 ms VOT) – bin (~ 10 ms VOT). In these cases, listeners do not discriminate between stimuli with long or short VOTs but are matching the acoustic input to a stored phonological representation in memory. Longer durational differences can also be used to indicate contrast in both vowels and consonants. For example, vowel length contrasts are possible in German e.g. Stall ([stal] ‘stable’) vs. Stahl ([sta:l] ‘steel’), while consonant durations are contrastive in languages such as Italian, Finnish and Bengali (e.g. in Bengali [pata] ‘leaf’ vs. [pat-a] ‘whereabouts’, ~90 vs 180 ms). Generally, speech employs only two degrees of length in consonants or vowels to differentiate meaning between words.

Despite this binary contrast, duration in speech sounds is a relative measure, since a sound can only be classified as long if a corresponding short item exists. Duration is also highly variable – speech rate, prosodic contexts, as well as segmental contexts can all influence phonetic duration. For instance, accented vowels are generally produced longer than unaccented vowels (Reetz & Jongman, 2009: 211ff), and vowels before voiced consonants are longer than before unvoiced consonants (Klatt, 1973; e.g., [æ] in mud is longer than mat). Although duration plays an important role in language, very rarely are we asked to discriminate between two consonants and determine which is the shorter. Nevertheless, implicitly humans have to distinguish between these on a daily basis to make sense of auditory speech input and consequently the lexical processing system must be sensitive to durational differences. The question we ask is the following: when duration affects word meaning, how sensitive is the brain in registering these durational differences? If a sound is produced as long within a word instead of short (or vice versa), would this mispronounced word be accepted as the real word despite the change in duration? If this is the case, would this acceptability be symmetrical (both durations are accepted for the other) or asymmetrical with one durational change being more acceptable?

Many categorical discrimination tasks show asymmetries in processing. For example, in visual perception a diagonal line can be spotted in a field of vertical lines much more quickly than a
vertical line in a field of diagonals (Treisman & Gormican, 1988). Vertical lines are accepted as a standard input and a deviation away from this is easy to spot. Similarly, a Q “pops out” of a field of Os whereas in the inverse case a serial search has to be performed to find an O in a field of Qs. It has been claimed that this asymmetry is not caused by an increase of visual information, but rather an increase in complexity or change from a default, as introducing a half circle (instead of a Q) in a field of Os shows the same pop out effect.

Corresponding asymmetries have been found in acoustic stimuli, both for changes in complexity (frequency modulated vs. standard tones) and duration (Cusack & Carlyon, 2003). The finding here is that longer durations among short duration distractors are easier to spot in a target detection task than the corresponding short duration targets among long duration distractors. This effect has also been shown in rats measuring epipidal auditory cortex potentials (Nakamura et al., 2011). Aside from this, asymmetries of pure acoustic length have not yet been explored.

Neuropsychological investigations provide further evidence for the brain’s sensitivity to durational differences. Näätänen, Paavilainen, and Reinväinen (1989) have used Mismatch Negativity (MMN) studies to show that humans are capable of distinguishing differences in simple tone duration at a variety of different lengths, both when comparing short to long stimuli as well as long to short stimuli. This effect has also been shown in speech stimuli, for example duration-dependent MMNs appear when standard short consonants are compared to deviant long consonants in Finnish (Leppänen & Lyytinen, 1997), as well as in short compared to long vowels in Finnish (Tervaniemi et al., 2009) and Thai (Gandour et al., 2002). These studies all employed nonsense syllables and do not tap into lexical processing. However, Kirmse et al. (2008) have also demonstrated sensitivity in the perception of length where the contrasts are linguistically relevant. They found cross-linguistic asymmetries between native and non-native speakers’ MMN responses using Finnish vowel length contrasts. While non-native speakers treated vowel contrasts in the same way as contrasts between pure auditory tones, native speakers of Finnish did not. This is in contrast to Fox and Lehiste (1989), who showed no difference between linguistic and non-linguistic stimuli or native and non-native speakers performance in an attended vowel length discrimination task, implying that this sensitivity to linguistically relevant stimuli is pre-attentive in nature. Catts et al. (1995) found an asymmetrical effect in the processing of duration of acoustic tones. However, this effect is likely caused by an overestimation of the long-deviant MMN because of differential states of refractoriness of the recruited neural populations (Jacobsen & Schroger, 2003).

Asymmetry in the discrimination of vowel length in nonwords is found in Japanese infants (Mugitani et al., 2009). In a habituation switch task, 18-month-old infants were able to discriminate a change from a long to a short vowel, but not vice versa. However, neither 10-month-old infants nor adults showed the asymmetry in discrimination. The lack of asymmetry for the adults is attributed to regular phonemic processing, while the younger infants are said to be making a purely acoustic contrast. The asymmetric discrimination of the 18-month-old infants is claimed to reflect the transition of treating vowel length as an acoustic-phonetic property to treating it as a phonological contrast.

There is a dearth of experimental research on neural correlates of the processing of consonantal duration distinguishing between long (geminate) and short (singleton) consonants. The only aspect of duration differences that lead to consonantal contrast that has been examined is VOT. In English, /p, t, k/ have longer consonantal release duration than the corresponding /b, d, g/ (~35 ms vs. ~10 ms). However, although the phonemic contrast is binary, there is a great deal of acoustic variation across the categories. For example, the voiceless consonantal sounds can vary between 35 and 100 ms. However, the absolute acoustic difference plays no role; no phonemic category is distinguished between, for instance 60 ms and 105 ms. In a seminal paper, Phillips et al. (2000) show that an MMF response (magnetic mismatch field, a counterpart of the MMN) to /t/-/d/ variations is triggered only by duration differences marking the voiced/unvoiced phonological categories and not by acoustic phonetic differences alone. MMF differences were found across (e.g. 24–40 ms) and not within (e.g. 44–60 ms) phonological categories. While MMF responses were symmetric when compared across phonetic categories, an asymmetry was found when comparing these stimuli within their respective phonetic categories, with no significant difference between long VOT standards and deviants. However, the authors attribute this asymmetry to the lower amplitude MMF response leading to lower signal:noise ratio in the MEG signal in the long VOT condition.

The contrast between long and short consonant duration displays similar characteristics to VOT differences, as the phonological categories are binary despite acoustic variability within each category. The main acoustic correlate of consonant duration is the duration of closure before the release of the consonant. Like in most languages with phonemic consonant duration contrast, in Bengali, for instance, the difference between a short and a long consonant can be between 80 ms and 200 ms, whereas no phonemic categorical distinctions are made, for example, between 160 ms and 280 ms. In a perception study where the closure duration of 6 minimal pairs was varied incrementally in 10 ms steps ranging from 60 to 200 ms, Hankamer, Lahiri, and Koreman (1989) found that Bengali native speakers exhibited an S-shaped response curve with a categorical shift around 130–140 ms.

Switching closure duration between minimal pairs has been shown to switch categories in languages like Turkish (Lahiri & Hankamer, 1988). The question that arises is to what extent is the brain sensitive to the manipulation of phonemic consonant duration alone. Our goal is to examine how a decrease or increase of consonant duration affects processing and whether incorrect duration causes violations which hinder lexical access and semantic integration. Violations can evoke strong brain reactions and several components have been discussed in the literature as a signal of error detection. In this paper we will address detection of duration errors by investigating the N400 component during the auditory presentation of correct and manipulated consonant duration in natural words.

1.1. The present study

About half of the world’s languages use consonant duration to differentiate meaning. The language we chose to examine is Bengali, which has a very productive long (geminate) – short (singleton) consonant distinction across 24 phonemes. Our goal was to examine the brain’s sensitivity to manipulations of consonantal duration in word recognition and specifically whether the processing of manipulated duration would yield symmetric (e.g. any deviation in duration is rejected) or asymmetric results (e.g. the input is assumed to be short unless there is a specific match that is long).

For this purpose we used a cross-modal (auditory to visual) lexical decision task with semantic priming, examining the elicitation of the N400 ERP component, which reflects cognitive processing in word recognition and semantic integration. The prime is here considered the “context” into which the target must be integrated. Semantically anomalous words produce larger N400 responses than related words (Kutas & Hillyard, 1980), as do words that are preceded by nonword primes (Deacon, Dynowska, Ritter, & Grose-Fiffer, 2004). Semantic priming paradigms have often been
used to investigate lexical access. The facilitative effect (‘priming’) of semantic context on word recognition has been demonstrated in a large number of studies using reaction time (RT) measures. In general, a word (e.g. river) is recognised with greater speed and/or accuracy if it is preceded by a semantically related word (e.g. stream) than by an unrelated context (e.g. clock) (Collins & Loftus, 1975; Neely, 1977, 1991). Since accessing single words and integrating their meaning are operations fundamental to successful language comprehension, a number of studies use the same priming procedures to examine to what extent similar sounding nonwords or mispronounced words would activate real words in the lexicon by employing lexical decision (deciding whether a letter string is or is not a word) (Bolte & Coenen, 2000; Connine, Blasko, & Titone, 1993; Roberts, Wetterlin, & Lahiri, 2013). Here, the expectation is that if the mispronounced nonword *biver* activates river it should facilitate recognition of stream. Therefore a semantically related target following the mispronounced word will still be “primed”. In ERP studies, this can be seen as a reduction in the N400 component, in comparison to an semantically unrelated word that would generate a large N400 component. Previous studies have shown this effect in a variety of mispronunciations; alteration of word initial phonemes (*tomato* for tomato, Connine et al., 1993; Friedrich, Lahiri, & Eulitz, 2008) and word medial phonemes (*tonato* for tomato; Roberts et al., 2013) lead to semantic priming and an attenuation of the N400 component.

Here we apply this methodology for the first time to mispronunciations of duration. For Experiment 1, a cross-modal lexical decision task with semantic priming was used, where the duration of the singleton medial consonant was lengthened to produce a larger N400 component. Previous studies have shown this effect in a variety of mispronunciations; alteration of word initial phonemes (*tomato* for tomato, Connine et al., 1993; Friedrich, Lahiri, & Eulitz, 2008) and word medial phonemes (*tonato* for tomato; Roberts et al., 2013) lead to semantic priming and an attenuation of the N400 component.

### 2. Methods

#### 2.1. Participants

42 undergraduate students from Gokhale Memorial Girls’ College (Calcutta) participated in the study. All were native Bengali speakers aged between 18 and 23 and right handed (assessed using a modified version of the Edinburgh Handedness Inventory; Oldfield, 1971). All participants had normal or corrected-to-normal vision, no hearing or neurological impairments and were not dyslexic.

#### 2.2. Stimuli

In each experiment, the stimuli consisted of 36 real word targets that were associated with four sets of primes: 2 sets of semantically related test primes (i.e. 1 set of real words and 1 corresponding set of mispronounced nonwords) and 2 sets of unrelated control primes (i.e. 1 set of real words and 1 corresponding set of mispronounced nonwords). Participants were also presented with 36 nonword targets in each experiment with matching sets of real-word and mispronounced primes. In Experiment 1, all real word targets were semantically related to primes that had singleton medial consonants, whereas in Experiment 2 all real word targets were paired with primes that had geminate medial consonants (see Table 1).

All real-word primes were diyllabic monomorphemic Bengali words with initial stress containing a single medial consonant or geminate from the set /m, n, l, k, d, g, h, f, p, b, t/: Mispronounced nonword primes were created by changing the medial singleton consonant to a geminate, or the medial geminate consonant to a singleton. All mispronounced nonwords were phonotactically legal in Bengali and were matched across experiments for possible neighbours by assessing how many words could be generated by changing the initial consonant; the total number of possible neighbours for the mispronounced nonwords were 24 in Experiment 1, and 25 in Experiment 2. That is, two thirds of the nonwords in each experiment had one competitor, the rest had no competitors. Unrelated control primes were constructed in the same way as related test primes, but were not phonetically nor phonologically related to the targets. All primes and targets were common words in Bengali.

Real word targets were semantically related to the test prime, and whenever possible synonyms were chosen. Targets were between one and three syllables in length, equally matched across conditions. Pairs were evaluated for semantic relatedness on a seven point scale from “completely unrelated” (1) to “completely related” (7) by 30 native speakers of Bengali. Only pairs with a score greater than 6 for test items and below two for control items were included in the experiment. As no published frequency norms for Bengali exist, all words were evaluated and matched across conditions for frequency and familiarity using separate subjective rating tasks by 60 native speakers of Bengali (30 per task). Judgments of this type have been shown to correlate well with traditional measures of corpora world frequency (Segui, Mehler, Frauenfelder, & Morton, 1982; Balota, Pilotti, & Cortese, 2001). Frequency ratings were based on how often a participant encountered a word from 1 (never) to 7 (several times a day). Familiarity ratings scored from 1 (unfamiliar) to 5 (very familiar). All words that were used in the study had an average frequency rating above 5 and familiarity rating of 5.

#### 2.2.1. Stimulus recording

All real-word and mispronounced primes were natural spoken stimuli, recorded by a female native speaker of Bengali in a sound attenuated room using a Roland R26 digital recorder (sampling rate 44.1 kHz) and a professional quality microphone (Shure SM27). The speech stimuli were extracted and the volume equalised using the acoustic analysis programs Praat (Boersma & Weenink, 2011) and Audacity (Audacity Team, 2010), but no other acoustic manipulations (e.g. lengthening or shortening of the medial consonant) were performed. The mean duration of geminates was 207 ms (min 157 ms, max 265 ms) whereas the singletons had a mean duration of 86 ms (min 59 ms, max 112 ms). These correspond well with the categories for singletons and geminates shown by the cross-spliced study by Hankamer et al. (1989) and gating studies by Lahiri and Marslen-Wilson (1992) in Bengali.

#### 2.3. Procedure

The experiments were conducted in a quiet airconditioned, darkened room where participants were seated in a comfortable chair with a 15” laptop display approximately 1 m in front of them. Participants were presented with auditory primes via the headphones (e.g. {*n:o*}) followed by a visual target (e.g. [haʃi]). The visual target appeared on the screen in Bengali script 250 ms after the offset of the auditory prime and was presented for 300 ms. Participants were instructed to indicate whether the word they saw on the screen was an actual word in Bengali by pressing corresponding “Yes” or “No” buttons on a response pad with their thumbs. Participants undertook a training block of 10 trials, and then the actual experiments, each consisting of 4 blocks of 72 trials, with short breaks (5 s) every 10 words, and longer breaks (~2 min) between blocks to reduce fatigue. A block contained every target word once only, such that in each block a participant saw 36
real-word targets and 36 nonword targets, pseudorandomised across conditions, with related test word primes occurring in 12.5% of trials. Block presentation order within experiments was counterbalanced across all participants. Response mapping was changed for half of the participants, so that 50% pressed ‘yes’ with their dominant hand and 50% with their non-dominant hand to compensate for response preparation effects in the contralateral hemisphere (Kutas & Donchin, 1977). The importance of speed and accuracy was explained, with the greater emphasis on accuracy.

2.4. EEG recordings

EEG recordings were made using a Biosemi ActiveTwo amplifier from 64 sintered Ag/AgCl pin electrodes placed in a 10–10 montage, online referenced to the mastoids. EEG activity was measured using 4 facial electrodes (F1, F2, F3, F4) across the experimental data. All electrode offsets (in an active-electrode system this is comparable to impedance) were kept below 30 mV and signals were digitised at 2048 Hz.

2.5. Behavioural analysis

2.5.1. Reaction times

Reaction times for correct responses to word targets were analysed using a LME model for the fixed effects Prime-type (related test) and Word-status (word/nonword). Subjects were included in the model as a random effect.

2.5.2. Response accuracy

Effects on response accuracy were analysed using a logit Generalised Linear Model (treating response accuracy as a binomial distribution) again for the fixed effects of Prime-type and Word-status.

2.6. ERP analysis

The continuous EEG data were filtered with a 0.1 Hz High-pass and 25 Hz Low-pass filter. Pre-experimental eye movement data were used to capture characteristic scalp topographies of eye artifacts, which were then applied to an EEG correction algorithm (Proctor, Hegarty, Nijboer, & Hillyard, 2002) across the experimental data. To remove other sources of non-EEG noise, trials were rejected if they exceeded an amplitude of 90 μV or a gradient of 75 μV/division in a semi-automatic procedure. During trial rejection, any participant that showed too many EEG artifacts (>30% of trials rejected in a single condition) was excluded from further analysis. The EEG epochs were averaged with respect to the target stimulus onset for all targets that were correctly classified as words, with a pre-stimulus baseline period of 200 ms and a window of 1000 ms from target onset. The N400 response was set at 300–500 ms, taking the mean amplitude in the time window. Nine regions of interest (ROIs) were defined by dividing the electrode array into a 3 × 3 grid (Wilson et al., 2012), with the following electrode groupings: anterior-left (AF7, F3, F5 and F7), anterior-middle (Fp1, Fpz, Fp2), anterior-right (AF8, F4, F6 and F8), centro-left (FC5, FC3, T7, CP5 and CP3), centro-middle (FC1, FC2, C1, C2, CP1, CP2 and Cpz), centro-right (FC4, FC6, FT8, T8, C4, C6, CP4 and TP8), posterior-left (P3, P5, P7 and P9), posterior-middle (P1, P2, PO3, PO4, O1, O2 and Pz), and posterior-right (P4, P6, P8, PO6, PO8). Data from each electrode within a ROI were treated as repeated measures for display purposes (Fig. 3) using a point by point subtraction of the test items from the control response average within each participant. ERP data were analysed with a Linear Mixed Effects (LME) model (for a discussion, see for example Newman, Tremblay, Nichols, Neville and Ullman, 2012) with subject and electrode as crossed random effects and Prime-type, Word-status and ROI as fixed effects.

3. Results

3.1. Behavioural measures

3.1.1. Experiment 1: singleton real words

Nine participants with more than 10% errors were excluded from analysis. Reaction times were trimmed to exclude any responses ±2 standard deviations from the group mean. Mean reaction times and percentage of correct responses are given in Table 2. Faster reaction times were seen in the test conditions in comparison to the control conditions. The LME analysis showed a significant effect of Prime-type, F(1,4093) = 38.24, p < .0001 as well as a significant interaction effect between Prime-type and Word-status, F(1,4093) = 9.43, p = .002. Planned comparisons were performed, comparing control to test items within Word-status, which revealed significant priming between controls and test items for both words, t = 6.51, p < .0001 and nonwords, t = 2.21, p = .027. Control words elicited a greater number of incorrect responses than test words, with a significant effect of Prime-type only, χ² = 9.89, p = .002. Again, the planned comparison showed priming for both the responses to words t = 10.57, p < .001 and nonwords, t = 1.24, p < .0001.

3.1.2. Experiment 2: geminate real words

Seven participants made more than 10% errors and were excluded from further analyses. Reaction times were again trimmed to ±2 SD from the group mean. Mean reaction times and percentage of correct responses are given in Table 2. Faster reaction times were seen for test items, reflected in a significant effect of Prime-type only, F(1,4324) = 31.41, p < .0001. Both words, t = 4.12, p < .0001 and nonwords, t = 3.81, p < .0001 showed priming in the planned comparison test. In the analysis of response accuracy, there was a significant interaction effect between Prime-type and Word-status, χ² = 4.28, p = .039. With the planned comparison test showing a significant priming effect between control words and test words, t = 7.57, p < .006, but no difference between nonwords, t = 0.06, p = .803.

3.2. Evoked potentials

3.2.1. Experiment 1: singleton real words

Nine participants were removed from the sample for having greater than 30% rejected trials in any one condition, leaving 33 participants for analysis. Visual inspection of the ERP waveforms across conditions showed enhanced N400 responses for control words relative to test words in both the word and nonword conditions (cf. Fig. 1). The scalp distribution of this effect was maximal over midline and posterior electrodes. Difference waveforms (Fig. 3) show a highly similar N400 pattern for both the words and nonwords, peaking around 400 ms. The LME model is shown in Table 3, and shows separate effects for Prime-type and Word-status, but no interaction, confirming that there was no difference in N400 effects. Planned

| Table 2 | Mean reaction times in ms and accuracy in per cent (with standard error) for Experiment 1 (singleton real words) and Experiment 2 (geminate real words). |
|-----------------|-----------------|-----------------|-----------------|
| Prime-type      | Word-status     | RT (SE)         | Accuracy (SE)   |
| Experiment 1    |                 |                 |                 |
| Related test    | Singleton (word)| 637 (13.3)      | 95.7 (6)        |
|                 | Geminane (nonword)| 658 (13.3)      | 93.8 (7)        |
| Unrelated control| Singleton (word)| 675 (13.4)      | 92.4 (8)        |
|                 | Geminane (nonword)| 670 (13.3)      | 92.6 (8)        |
| Experiment 2    |                 |                 |                 |
| Related test    | Singleton (nonword)| 598 (13.2)      | 95.1 (6)        |
|                 | Geminane (word) | 603 (13.2)      | 94.5 (7)        |
| Unrelated control| Singleton (word)| 616 (13.2)      | 92.4 (8)        |
|                 | Geminane (nonword)| 620 (13.2)      | 94.7 (7)        |
comparisons were performed on this nonsignificant interaction, looking at the effect of Prime-type within Word-status (Table 4), and showed a strong effect of Prime-type for both words and nonwords. Further post-hoc tests were performed on the three way interaction of Prime-type × Word-status × ROI to confirm the scalp distribution seen in the difference waveforms. The post-hoc analysis showed significant effects of Prime-type for all central and parietal ROIs, consistent with a lack of semantic violation effects for both test words and mispronounced nonwords.

3.2.2. Experiment 2: geminate real words

Thirteen participants were removed from the sample for having greater than 30% rejected trials in any one condition, leaving 29
participants for analysis. More negative N400 responses were present for control words and nonwords as well as test nonwords in comparison to test words (cf. Fig. 2). Difference waveforms (Fig. 3) show a difference between the test words and nonwords, confirmed by the Prime-type × Word-status interaction effect (Table 5), indicating that these singleton forms of geminate words are not accepted as viable tokens. Planned comparisons of Prime-type within Word-status (Table 6) confirmed that while test words did not elicit an N400 in comparison to control words, test nonwords elicited an almost identical N400 response as control nonwords. The Prime-type × Word-status × ROI interaction reflected the greater Prime-type effect of words but not nonwords over posterior electrode sites. To confirm this we conducted a post-hoc analysis of Prime-type within Word-status for each ROI. This analysis revealed significant N400 effects in the word condition but not the nonword condition for parietal midline and right central and parietal ROIs, consistent with an N400 effect associated with semantic violations for the control words.

4. Discussion

The present study aimed to explore how durational differences, which represent a binary phonemic contrast, are processed at the lexical level in the brain. In two cross-modal semantic priming experiments, we presented spoken prime words that varied in medial consonant length in conjunction with either semantically related or unrelated visual targets. In the first experiment, the goal was to see whether a nonword geminate derived from a real word singleton would facilitate access to the real word’s semantic relation. In the second experiment, the reverse was tested. An asymmetry was found in the acceptance of these mispronounced nonwords in both N400 amplitude and response errors.

While the acoustic difference between long and short speech sounds can be easily perceived, previous evidence has been unclear on how durational information is employed during speech processing and lexical activation. Neurophysiological research on duration has focused primarily on the question of whether there is categorical perception in languages for vowel length and VOT contrasts. It has been argued that durational differences across phonemic categories are easier to discriminate than within categories for VOT (Phillips et al., 2000) but not necessarily for vowel length (Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006). For differences in VOT, across category differences for both positive (Phillips et al., 2000) and negative VOTs (Sharma & Dorman, 1999) triggered significantly stronger MMNs across categories than within categories. In Finnish, native speakers perception of within as well as across category differences triggered equivalent MMNs (Kirmse et al., 2008). However, native Finnish speakers showed greater MMNs than non-native speakers as they had better vowel prototypes. Our focus has been on investigating to what extent prototypical consonant duration differences are necessary for lexical access. We looked at phonemic contrasts not associated with VOT but rather with the duration of closure which marks geminate vs. singleton consonants in a large variety of languages across the world.

4.1. N400 amplitude

In our first experiment, no difference was found in the acceptability of the mispronounced nonwords compared to the real word primes. In Experiment 2, where we decreased the duration of geminate medial consonants, we found that these nonwords were not accepted as real word primes. Semantic activation of targets, as indexed by N400 component amplitude, differed for singleton and geminate mispronunciations; geminate mispronunciations of singleton words still caused semantic activation of the target while singleton mispronunciation of geminate real words were treated as controls. An N400 component was elicited in all cases of semantic violation (control conditions) in both experiments. The scalp distribution and latency of this effect was similar to previous studies of N400 elicitation in semantic priming (Kutas & Federmeier, 2011). While in Experiment 1, lengthened mispronounced related nonwords showed an attenuation of the N400 response and ERP waveforms patterned like related real words, in Experiment 2 the shortened mispronounced related nonwords showed a clear N400 response and their ERP waveforms patterned like unrelated control nonwords.

In Experiment 1, the N400 component was broadly distributed across central and parietal sites, whereas in Experiment 2 the N400 was strongest in the right-parietal region. The difference observed here while using identical tasks could indicate either different neural generators or reductions in generator activity resulting in a change in the orientation of the generator dipole. Experiments examining the N400 are split over whether a left/ right asymmetry is observed; around 60% of the studies show a greater right-parietal N400 (Kutas & Van Petten, 1994). It is difficult to deduce from these experiments why this difference has arisen, although the assumption is that this effect is task dependent. Whether our findings are directly related to the geminate – singleton contrast requires further investigation.

---

Table 3
ANOVA table for N400 amplitude in the 300–500 ms time window for Experiment 1.

<table>
<thead>
<tr>
<th>Factor</th>
<th>df</th>
<th>d(E)</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prime-type</td>
<td>1</td>
<td>6210</td>
<td>259.79</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Word-status</td>
<td>1</td>
<td>6210</td>
<td>292.80</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Prime-type × Word-status</td>
<td>1</td>
<td>54</td>
<td>1.76</td>
<td>.185</td>
</tr>
<tr>
<td>ROI</td>
<td>8</td>
<td>6210</td>
<td>2.07</td>
<td>.055</td>
</tr>
<tr>
<td>Prime-type × ROI</td>
<td>8</td>
<td>6210</td>
<td>7.94</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Word-status × ROI</td>
<td>8</td>
<td>6210</td>
<td>5.85</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Prime-type × Word-status × ROI</td>
<td>8</td>
<td>6210</td>
<td>.48</td>
<td>.870</td>
</tr>
</tbody>
</table>

Table 4
Contrast table for planned and post-hoc comparisons of Prime-type within Word-status for Experiment 1.

<table>
<thead>
<tr>
<th>Planned comparison, Prime-type × Word-status</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control word vs Test word</td>
<td>−12.34</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Control nonword vs Test nonword</td>
<td>−10.46</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Post hoc, Prime-type × Word-status × ROI</td>
<td>t</td>
<td>p</td>
</tr>
<tr>
<td>Anterior left Word</td>
<td>−1.78</td>
<td>.223</td>
</tr>
<tr>
<td>Anterior midline Word</td>
<td>−3.53</td>
<td>.004*</td>
</tr>
<tr>
<td>Anterior right Word</td>
<td>−2.31</td>
<td>.105</td>
</tr>
<tr>
<td>Anterior midline Nonword</td>
<td>−2.00</td>
<td>.184</td>
</tr>
<tr>
<td>Central left Word</td>
<td>−3.49</td>
<td>.004*</td>
</tr>
<tr>
<td>Central midline Nonword</td>
<td>−3.10</td>
<td>.011*</td>
</tr>
<tr>
<td>Central right Word</td>
<td>−5.38</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Central midline Nonword</td>
<td>−3.80</td>
<td>.001*</td>
</tr>
<tr>
<td>Parietal left Word</td>
<td>−3.16</td>
<td>.011*</td>
</tr>
<tr>
<td>Parietal midline Nonword</td>
<td>−4.74</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Parietal right Word</td>
<td>−7.93</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Parietal left Nonword</td>
<td>−8.90</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Parietal right Nonword</td>
<td>−5.98</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Parietal midline Nonword</td>
<td>−5.39</td>
<td>&lt;.0001*</td>
</tr>
</tbody>
</table>

*Post hoc p values were Holm–Bonferroni corrected for 18 comparisons.
4.2. Response accuracy

While reaction time analyses showed symmetric responses in Experiment 1 and 2, examination of the error rates in the lexical decision task revealed an asymmetric pattern of results identical to the ERP data. Participants showed a greater accuracy to word targets when primed with a related word. In Experiment 1, both related words and related nonwords showed an increase in accuracy of response in comparison to unrelated control words and nonwords. This is again an indication that the mispronounced (lengthened) nonwords cause semantic activation of the target. In Experiment 2, only related words showed an increase in accuracy of response, the related (shortened) nonwords did not differ from the control nonwords, indicating that they did not cause semantic activation of the target.

4.3. Asymmetries in speech

The asymmetry in acceptability of length deviants is not an inherent part of the auditory pathway. The MMN studies of
Näätänen and his colleagues (cf. Näätänen et al., 1989; Kaukaranta, Sams, Hari, Hämäläinen, & Näätänen, 1989) show that humans are capable of distinguishing both increases and decreases of simple tone duration pre-attentively, and this effect is replicated for speech stimuli (Leppänen & Lyytinen, 1997; Gandour et al., 2002; Fox & Lehiste, 1989). The latter experiments with adults, which show a sensitivity to increases in duration in speech stimuli, were not just pre-attentive where participants passively watched a movie, but employed additional tasks like lexical decision or categorisation.

Our claim is that where consonantal duration is concerned, although speakers are able to distinguish between long and short consonants (Lahiri & Hankamer, 1988), the asymmetry in our study shows that a long consonant subsumes a short one, but not vice versa. Thus, when a singleton is mispronounced as a geminate, native speakers can ignore the extra length since the longer duration includes the shorter one as well as all other necessary information to identify the consonant. However, when a geminate is incorrectly shortened, the shortening has a different effect that does not facilitate lexical access. Consequently, the N400 component in this case is no different from that of a control item.

Asymmetries in phoneme inventories in languages are not unusual. For instance, long vowel and consonantal phonemes imply that corresponding short ones exist. Grey shadowed areas mark the N400 time window used in statistical analyses.

Table 5
ANOVA table for N400 Amplitude in the 300–500 ms time window for Experiment 2.

<table>
<thead>
<tr>
<th>Factor</th>
<th>df</th>
<th>df(E)</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prime-type</td>
<td>1</td>
<td>5454</td>
<td>30.46</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Word-status</td>
<td>1</td>
<td>5454</td>
<td>35.96</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Prime-type x Word-status</td>
<td>1</td>
<td>54</td>
<td>25.47</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ROI</td>
<td>8</td>
<td>5454</td>
<td>3.21</td>
<td>.005</td>
</tr>
<tr>
<td>Prime-type x ROI</td>
<td>8</td>
<td>5454</td>
<td>4.21</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Word-status x ROI</td>
<td>8</td>
<td>5454</td>
<td>1.12</td>
<td>.344</td>
</tr>
<tr>
<td>Prime-type x Word-status x ROI</td>
<td>8</td>
<td>5454</td>
<td>2.20</td>
<td>.025</td>
</tr>
</tbody>
</table>

Fig. 3. ERP difference waveforms (control – test) for words and nonwords. (a) Experiment 1, singleton words and geminate nonwords and (b) Experiment 2, geminate words and singleton nonwords. Grey shadowed areas mark the N400 time window used in statistical analyses.
Consonants imply the existence of voiceless phonemes. Similarly, existence of front rounded vowels (e.g. [ü]) imply unrounded vowels (e.g. [i]). Processing asymmetries are also not unusual. Infants have been known to exhibit asymmetries in different linguistic dimensions, such as place of articulation; e.g., [l]/[r] in English and German (Polka & Bohn, 2003), [i]/[i] in English (Desjardins & Trainor, 1998). Similar results are put forward for vowels (e.g. [i]). Processing asymmetries are also not unusual. This incompressibility, or minimum length (Dmin) that segments cannot or otherwise, that can shorten a segment. However, there is a point of incompressibility, or minimum length (Dmin) that segments cannot fall under (Lehiste, 1972; Klatt, 1973; Port, 1980). This incompressibility is not a purely articulatory limitation, as geminate consonants have a Dmin that is still above the singleton/geminate category boundary, even when comparing geminates in fast speech to singletons in normally paced speech (Arvaniti, 1999, 2001). Our findings are consistent with Klatt's notion of minimum duration since singleton nonwords did not act as geminate real-word primes. The findings also suggest that there is no equivalent Dmax, i.e. a maximum duration beyond which segments cannot be produced/ perceived since geminate nonwords did act as singletons real-word primes. This is consistent with evidence from Japanese (Hirata & Whiton, 2005), where singletons in slow speech can be produced exceeding the durational categorical boundary.

### 4.5. Phonological constraints

Another possible explanation for the asymmetry seen here is an effect of phonotactic probability. Like all languages with a consonantal duration contrast, there are more singletons in Bengali than geminate consonants. This is particularly the case when historically the geminate consonants have been derived from singletons. In this study we tried to control for this by ensuring that our nonwords were equal in their cohort size for both the singleton and geminate categories and were controlled for word frequency and familiarity for both prime and target stimuli. We cannot eliminate this hypothesis, but there is a strong likelihood that the effect we are seeing is not purely based on phonotactics.

Duration in speech is not only phonological, since the length of individual segments is always relative. They are dependent on factors such as the length of the word, the surrounding sounds, position in a phrase, and the rate of speech. There are many factors, phonological or otherwise, that can shorten a segment. However, there is a point of incompressibility, or minimum length (Dmin) that segments cannot fall under (Lehiste, 1972; Klatt, 1973; Port, 1980). This incompressibility is not a purely articulatory limitation, as geminate consonants have a Dmin that is still above the singleton/geminate category boundary, even when comparing geminates in fast speech to singletons in normally paced speech (Arvaniti, 1999, 2001). Our findings are consistent with Klatt's notion of minimum duration since singleton nonwords did not act as geminate real-word primes. The findings also suggest that there is no equivalent Dmax, i.e. a maximum duration beyond which segments cannot be produced/ perceived since geminate nonwords did act as singletons real-word primes. This is consistent with evidence from Japanese (Hirata & Whiton, 2005), where singletons in slow speech can be produced exceeding the durational categorical boundary.

### 5. Conclusion

The current study shows that an asymmetry exists in the processing of duration of speech sounds during lexical access, when the consonantal duration is linguistically contrastive and distinguishes words. Consonants in all languages have some intrinsic duration, but only half of the word's languages use duration to mark a linguistic contrast. By employing real words and corresponding mispronounced nonwords as primes in a semantic priming experiment in Bengali, which productively uses singleton-geminate contrasts for all consonants, our results show that modifying short consonantal durations to make them longer does not hinder lexical access, as reflected by an attenuated N400 component in comparison to a control word. In contrast, shortening long consonantal durations does hinder lexical access, giving a clear N400 component that matches a response to a control word. We claim that these results were the consequence of asymmetries in the phonological representations of the singleton-geminate contrast, where long geminate consonants subsume singletons, but not vice versa.

Although consonant duration contrasts are not unusual, very little is known about their representation and processing across languages. Few neuropsychological studies examining consonantal duration exist, and none use semantic priming which is understood to tap into the lexical representation. To examine whether

### Table 6

<table>
<thead>
<tr>
<th>Planned comparison, Prime-type × Word-status</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control word vs Test word</td>
<td>.74</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>Control nonword vs Test nonword</td>
<td>.33</td>
<td>.738</td>
</tr>
<tr>
<td>Post hoc, Prime-type × Word-status × ROI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Post hoc p values were Holm–Bonferroni corrected for 18 comparisons.
contrastive duration is processed similarly in other languages, further research is required with speakers of (for example) Tamil, Estonian or Turkish, who also have productive singleton-geminate contrasts, and where we would predict a similar effect.

Acknowledgements

This work has been supported by the EC within the 7th framework programme under grant agreement no. FP7-IST-269670 (Advanced Investigator’s Grant, PI Lohiri). We would like to thank our three anonymous reviewers for their helpful comments on previous versions of our manuscript and Narmada Paul and Oindrila Bhattacharya for their help with EEG recording. The fieldwork this research is based on would not have been possible without the participation of the staff and students at Gokhale Memorial Girls’ College in Calcutta and in particular Dr Leena Sengupta’s patience and support.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuropsychologia.2014.03.015.

References


