Sound of mind: electrophysiological and behavioural evidence for the role of context, variation and informativity in human speech processing

Nixon, J.S.

Author: Nixon, Jessie Sophia
Title: Sound of mind: electrophysiological and behavioural evidence for the role of context, variation and informativity in human speech processing
Issue Date: 2014-10-14

Summary

This thesis presents evidence obtained through a variety of empirical and statistical methods to investigate how healthy adult native speakers process the inherently noisy acoustic information that spoken communication consists of. While psycholinguistic models (especially of speech production and reading aloud) have generally measured phonological processing in terms of sequences of phonemes, this thesis has demonstrated several types of subphonemic processing in a variety of tasks. One of the main focuses of this dissertation is how speakers and listeners process phonetic variation. Importantly, a distinction is made between informative, regular variation, as in allophonic variants, and, in contrast, random noise, which reduces the informativity of acoustic cues. In addition, we also show that, even in a language where no information about speech contrasts is represented in the orthographic script, both overt production and visual processing of written words involve processing of contrastive speech sound categories.

Chapter 2 made use of the picture-word interference paradigm to investigate how tonal variants are processed during Mandarin speech production. In Beijing Mandarin, when Tone 3 is followed by another Tone 3 character (third tone sandhi), it sounds like Tone 2. This aspect of the tonal system means that sandhi words phonologically related to both Tone 3 and Tone 2 words, which allowed us to manipulate two kinds of phonological information: the phonological tone category and the acoustic tonal contour. Chapter 2 investigated which of these two types of phonological relatedness is important during speech production. While there

is a wealth of psycholinguistic research showing that processing is facilitated by pre-activation of congruent phonological information, the exact nature of this phonological information is not yet clear. Does speech production involve retrieval of the speech sound category? Or is it activation of the actual acoustic realisation that is important? Chapter 2 (see also Nixon et al., 2014) showed that speech production involves multilevel phonological processing. Production of sandhi picture names was significantly faster when distractor and target picture matched in tone category, but had different overt realisations and when target and distractor matched in overt realisation, but mismatched in tone category (compared to control distractors, which mismatched the target in both the toneme and the contour). Interestingly, there were differences in the time course of effects. The tonal contour facilitated production both when presented simultaneously with the target picture and when it was delayed by 83 milliseconds. The tone category, in contrast, only facilitated production with simultaneous presentation. This suggests that the speech category is activated early (perhaps during word retrieval), but that in later stages of processing (during speech preparation, for instance), the actual acoustic realisation becomes more important. There are two possible explanations for this pattern of results. It may be that the realisation contour remains activated for longer than the tone category during overt production. Another possibility is that, while both the contour and the category remain activated, as the task shifts from lexical retrieval to articulation preparation, only the articulatory/acoustic congruency benefits production. Further research is required to tease these two possibilities apart. A second question raised by the results of Experiment 1, is whether processing of the context-specific tonal contour is automatic, or whether it occurs only with overt production of the speech variants, in preparation for articulation. To test what kind of phonological information is activated when sandhi words are processed but not overtly produced, prime and target conditions were reversed in Experiment 2.

In Experiment 2, primes were visually presented sandhi tonal variants or control distractor words, and targets were Tone 3 or Tone 2 pictures. As with Experiment 1, results showed facilitation of both the tone category targets and the tonal contour targets with sandhi distractors compared to controls. This indicates that processing of the context-specific instantiation is automatic and is not simply a result of articulation preparation. Interestingly, the time course of activation differed from Experiment 1. When the tonal variants were processed visually, rather than overtly produced, it was the tone category that was facilitated both with simultaneous (0 ms) and delayed (83 ms) presentation of the distractor word. Facilitation from the tonal contour occurred only with simultaneous presentation. This suggests that during visual processing of tone sandhi variants, activation of the context-specific contour takes time relative to the speech category. When the distractor is presented simultaneously with the target, the contour is activated in time to benefit production. However, when presentation is delayed, it no longer benefits production of the contour target.

In an event related potential (ERP) study, Chapter 3 investigated processing of sub-phonemic segmental information in a typologically different language. Very little is known about sub-phonemic processing in reading aloud. This study used masked priming to investigate whether and when phonetic features are processed in Dutch reading aloud. In Dutch, the sound pairs t-d and p-b are produced at the same place of articulation (alveolar and bilabial, respectively), while the pairs t-p and d-b match in voicing (voiceless and voiced, respectively). EEG and reaction times were measured as participants read aloud real Dutch words (e.g. huid 'skin'). Each target word was preceded by a brief presentation of a masked non-word prime in which the final sound matched in voicing (*huib*), place of articulation (*huit*) or mismatched in both voicing and place (control condition, *huip*). Reaction times, analysed using linear mixed effects regression models, were significantly faster following voice-congruent primes, compared to control primes. Consistent with this, there was also reduced early negativity in the voice-match condition, compared to the control condition. The results indicate rapid processing of sub-phonemic voicing information in Dutch reading aloud. The facilitation cannot be due to orthographic or phoneme-level phonological information, since in match and mismatch conditions the relation between prime and target was identical in terms of both letters and phonemes: all primes differed from the target by exactly one phoneme and one letter. Only at the sub-phonemic feature level was the overlap greater in the match conditions, compared to controls. These results also have implications for the way sub-phonemic features are represented. Due to 'final devoicing', Dutch voiced stops have two realisations. In word-initial position, voiced (e.g. /d/ and /b/) and voiceless stops (/t/ and /p/) are distinguished by voice onset time (VOT). But in word-final position, the VOT values of voiced and voiceless stops are very similar. Therefore, at the level of articulation, both prime types in this study are voiceless. However, facilitation in voice-match condition indicates that the voicing distinction is retained and processed during reading aloud. This suggests that, although the overt realisation is similar, voiceless and devoiced stops are processed as separate categories. This is consistent with data presented in Chapter 2 showing both category-level and context-specific processing of speech variants during speech production and visual word processing. Although this experiment did not specifically test for activation of a context-specific variant, it is consistent with the possibility that multi-level processing also occurs at the feature level.

Chapter 4 used ERP measures to address the question of how context constrains phonological processing during reading aloud. ERPs and reaction times were recorded as native Beijing Mandarin speakers read aloud two-character words, preceded by masked primes. The initial character of all critical primes was a Tone 3 character, and initial characters were identical between priming conditions. Only the second character differed between conditions. The phonetic context created by the tone of the second character determined whether the first character had a rising (sandhi) contour or the canonical, low contour. Critical targets were words beginning with Tone 2, which also has a rising contour. Therefore all primes mismatched in tone category, but the sandhi primes matched in tonal contour. The best-fit generalised additive mixed model (GAMM) included full random effects structure, and significant predictors of prime condition, prime frequency and target frequency over time, and their interactions.

In the mismatch condition, effects of prime and target frequency were relatively minor. However, in the contour match condition, the tonal contour did not discriminate between prime and target creating a conflict for the participant response. In this condition, the a priori probabilities of the prime and target came into play. When target frequency was low and prime frequency high, ERP amplitudes were greater, suggesting increased processing effort due to resonse conflict. The differential effects in the match compared to the match condition provide evidence that top-down processing from information provided by the phonetic context constrains activation of the alternative realisations of allophonic variants. Identical initial characters (between conditions) led to different amplitude in the EEG, depending on the tonal context of the following character. The finding that target word frequency interacts with other predictors in explaining neural activity suggests that it is informative to include item information in ERP studies of language processing.

While the previous chapters investigated processing of speech contrasts, contrastive features and regular variation, Chapter 5 investigated effects of the degree of acoustic noise on processing of speech contrasts. Eye movements were recorded as native Cantonese listeners saw pictures corresponding aspirated and unaspirated word pairs and heard acoustic stimuli that contained either a relatively large amount of variation (the wide distribution condition) or relatively little variation (the narrow distribution condition). Analysis using generalised additive mixed modelling (GAMM) allowed complex, non-linear effects and interactions to be modelled, without the need to discretise continuous measures, such as time and frequency. The best-fit model revealed a clear shape of the distribution in the narrow (low-variability) condition. There was differential looking behaviour at category means, boundaries and peripheries. In contrast, in the wide (high-variability) condition, the distribution was flatter, particularly in the latter part of the trial. In other words, the effect of VOT was weak in the wide condition, especially after 600 ms, when the distribution appeared very flat across all VOTs. These results show that during online perception of speech contrasts, previous experience with the distribution an acoustic cue can affect the degree to which it is used to predict a linguistic outcome. That is, subtle differences in the degree of variation of a particular acoustic cue affect the way that is perceived and utilised to discriminate speech contrasts. In sum, this thesis provides new insights into how human listeners process contrastive and non-contrastive acoustic information during a variety of language processing tasks. It seems that speakers are able to extract several types of regularity from speech, whether it is contrastive speech categories, regular allophonic variants, or sub-phonemic feature information. These regularities are processed during overt production, visual processing of written words that are not overtly produced and during reading aloud. Top-down processing of information provided by the phonetic context constrains activation. Finally, the degree to which acoustic information is relied on as a cue to discrimination of sound contrasts during speech perception depends on the informativity–that is, the shape of the statistical distribution–of the acoustic cues.