Brain & Language 136 (2014) 31–43

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Phoneme-free prosodic representations are involved in pre-lexical and lexical neurobiological mechanisms underlying spoken word processing

Ulrike Schild [a,*], Angelika B.C. Becker [b], Claudia K. Friedrich [a]

[a] University of Tübingen, Developmental Psychology, Germany
[b] University of Hamburg, Biological Psychology and Neuropsychology, Germany

ARTICLE INFO

Article history:
Accepted 21 July 2014

Keywords:
Spoken word recognition
Lexical access
Lexical stress
ERPs
Word fragment priming
Lexical decision

ABSTRACT

Recently we reported that spoken stressed and unstressed primes differently modulate Event Related Potentials (ERPs) of spoken initially stressed targets. ERP stress priming was independent of prime–target phoneme overlap. Here we test whether phoneme-free ERP stress priming involves the lexicon. We used German target words with the same onset phonemes but different onset stress, such as *MANdel* (“almond”) and *manDAT* (“mandate”; capital letters indicate stress). First syllables of those words served as primes. We orthogonally varied prime–target overlap in stress and phonemes. ERP stress priming did neither interact with phoneme priming nor with the stress pattern of the targets. However, polarity of ERP stress priming was reversed to that previously obtained. The present results are evidence for phoneme-free prosodic processing at the lexical level. Together with the previous results they reveal that phoneme-free prosodic representations at the pre-lexical and lexical level are recruited by neurobiological spoken word recognition.

1. Introduction

Current modelling of spoken word recognition is largely determined by phonemes and their establishing features. Classical models converge in the assumptions that individual speech sounds are mapped onto pre-lexical phoneme representations and that word recognition is a function of the amount of overlapping representations at the pre-lexical phoneme level and the lexical word form level (e.g., Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994). How phonological characteristics beyond phoneme-relevant information, such as the words’ syllables with their specific stress pattern, contribute to spoken word recognition remains unspecified in those models. Here we propose that prosodic characteristics of the speech signal have their own phoneme-free representations, which are independent from phoneme representations. We base this assumption on our previous work on the role of syllable stress in German listeners’ spoken word recognition.

In stress-timed languages like German or English, typically a single syllable of a multisyllabic word is perceived to be more prominent than the remaining syllable or syllables. The prominent syllable is said to be stressed. For example, the first syllables of the words *FAther* or *MARket*, and the second syllables of the words *neON* and *musEUM* are stressed (capital letters indicate stress). Stressed syllables typically are longer, louder and marked by higher pitch than unstressed syllables (e.g., Fry, 1958). Next to those prosodic features, vowel identity might vary between stressed and unstressed syllables. While stressed syllables always contain a full vowel, unstressed syllables either contain a full vowel, such as the first syllable of *neON*, or they contain a reduced vowel, such as the second syllable of *FAther*. A confound results when stressed syllables and reduced unstressed syllables are compared. Those syllables do not only differ in their prosodic features, but also in the identity of their vowels. Therefore, we use stressed syllables and unstressed syllables with full vowels in the present experiment and focus on studies using stressed and unstressed syllables with full vowels when we review the literature on processing syllable prosody in the following paragraphs.

Previous behavioral research on the role of syllable stress in spoken word recognition focused on its function in differentiating phonemically ambiguous words such as *FORbear* and *forBEAR* (henceforth referred to as minimal word pairs), or in differentiating words with phonemically ambiguous word onsets such *MUsic* and *muSEUM* (henceforth referred to as minimal word onset pairs). Basically, this work reveals that syllable stress is used immediately to disambiguate phonemically ambiguous strings. Auditory repetition priming showed that minimal word pairs do not facilitate recognition of one-another (Cutler & van Donselaar, 2001; but see Cutler, 1986). Forced choice word completion indicated that listeners can correctly judge the respective carrier word given the onset

* Corresponding author. Address: University of Tübingen, Developmental Psychology, Schleichstraße 4, D-72076 Tübingen, Germany. Fax: +49 (0)7071 29 5219.
E-mail address: ulrike.schild@uni-tuebingen.de (U. Schild).

http://dx.doi.org/10.1016/j.bandl.2014.07.006
0093-934X/

of a minimal word onset pair member (Cooper, Cutler, & Wales, 2002; Mattys, 2000). Cross-modal visual–auditory priming revealed stronger facilitation exerted by the carrier word onset (*MUs-music*) as compared to the onset of a minimal word onset pair member (*muS-music*; Cooper et al., 2002; Soto-Faraco, Sebastián-Gallés, & Cutler, 2001; van Donselaar, Koster, & Cutler, 2005). Finally, eye tracking showed that Dutch listeners fixate the printed version of the word that a speaker intended to say (*OCtopus*), more frequently than they fixate the minimal word onset pair member already before they heard the end of the first syllable of the respective word (*ocTOber*; Reinisch, Jesse, & McQueen, 2010, 2011).

In the framework of pre-lexical phonological representations and lexical word form representations sketched by classical models of spoken word recognition, the facilitation effect exerted by syllable prosody might have at least two origins. Firstly, syllable stress might be tightly linked to phonemes both at the pre-lexical level and at the lexical level of representation. For example, the relatively long duration of /u/ in the initial syllable of *MUsic* might be mapped onto a pre-lexical representation coding for a long /u/. In turn, this pre-lexical representation is a better match for lexical representations with a long /u/ in the first syllable, such as *MUsic*, than for lexical representations with a short /u/ in the first syllable, such as *muSEUM*. Combined phoneme-prosody representations would not modulate the activation of word forms that are phonemically unrelated. Alternatively, syllable stress might be coded by phoneme-free prosodic representations. For example, the relatively long duration of the /u/ in the initial syllable of *MUSic* as well as the relatively long duration of the /o/ in the initial syllable of *OCtopus* might be mapped onto a pre-lexical representation coding for long vowels regardless of vowel identity. In turn, those abstract prosodic representations might be mapped onto lexical representations coding for a long vowel in their first syllable.

The architecture of neural auditory processing suggests that syllable prosody might not be that tightly linked with phonemes. Crucially, the different temporal availability of both types of information in the acoustic input is associated with specialized auditory processing networks respectively. Information that characterizes phonemes varies at a fast rate. Typically, rapid transitions ranging between 20 and 100 ms establish distinctive features, such as the voice onset time difference between /b/ and /p/. Information that characterizes syllable varies somewhat slower. Typically, features of pitch, loudness and duration ranging between 100 and 300 ms are relevant to distinguish between stressed and unstressed syllables such as *MUS* and *mus*. There is some neurocognitive evidence for lateralized specialization of auditory cortices to different temporal integration windows. Fast acoustic variation in the range of phoneme-relevant information appears to be pre-dominantly processed in the left hemisphere, slower acoustic variation in the range of syllable-relevant information appears to be pre-dominantly processed in the right hemisphere (e.g., Boemio, Fromm, Braun, & Poeppel, 2005; Giraud & Poeppel, 2012; Giraud et al., 2007; Luo & Poeppel, 2012; Zatorre & Belin, 2001). Yet, whether the initial separation of both types of information is maintained at higher language-specific processing levels has to be figured out.

Previous behavioral evidence for independent processing of syllable prosody along the spoken word recognition pathway is weak. In four auditory priming experiments, Slowiaczek, Soltano, and Bernstein (2006) failed to show pure stress priming. Neither lexical decision latencies nor shadowing differed for spoken target words that either were preceded by spoken words with the same stress pattern (*RAting – LIFEtime*) or by spoken words with a different stress pattern (*RAting – ciGAR*). That is, if there are some types of abstract prosodic representations, their activation might not be obligatorily reflected in response latencies obtained in auditory priming tasks.

Event-Related Potentials (ERPs) recorded in word onset priming previously revealed some evidence for independent processing of syllable prosody and phonemes. In a former study of us, we were selectively interested in the processing of pitch contours (Friedrich, Kotz, Friederici, & Alter, 2004). We extracted the first syllables of initially stressed German words, such as *KObold* (Engl. goblin), and of initially unstressed German words, such as *faSAN* (Engl. pheasant). We calculated the mean pitch contours of the stressed word onset syllables, such as *KO-*, and of the unstressed word onset syllables, such as *fa-*, and applied them to each individual syllable. This resulted in one version of each syllable with a stressed pitch contour and another version of the same syllable with an unstressed pitch contour. We used those syllables as primes. Primes were followed by written versions of the carrier words. Prime–target pairs varied in phoneme overlap, such as *KO-KObold* vs. *fa-Kobold*. Furthermore, primes varied in stress overlap. A stressed pitch contour preceding the written version of an initially stressed word as well as an unstressed pitch contour preceding the written version of an initially unstressed word were considered a stress match. The reversed pairings were considered a stress mismatch. ERPs reflected enhanced posterior negativity for stress mismatch compared to stress match. ERP stress priming did not interact with prime–target overlap in phonemes. This is evidence for abstract prosodic processing.

In a recently published study on literacy acquisition we found further evidence for independent processing of syllable stress and phonemes (Schild, Becker, & Friedrich, 2014). We presented spoken stressed and unstressed prime syllables followed by spoken German disyllabic target words. In order to make the words accessible for pre-schoolers, we presented only targets with stress on the first syllable, such as *MONster* (Engl. monster). We did not present words with stress on the second syllable, because they are not only less frequent in German, but they also are usually acquired later than initially stressed words. Spoken prime syllables were (i) the target words' first syllables, such as *MON-MONster*; (ii) unstressed versions of the target words' first syllables, such as *mon-MONster*; (iii) phonemically unrelated stressed syllables, such as *TEP-MONster*; or (iv) phonemically unrelated unstressed syllables, such as *tep-MONster*. Across pre-schoolers, beginning readers and adults we found comparable indices for independent processing of prosody and phonemes in the ERPs. However, in contrast to our former study (Friedrich, Kotz, Friederici, & Gunter, 2004; Friedrich, Kotz, Friederici, & Alter, 2004), stress match (conditions [i] and [iii]), elicited enhanced posterior negativity as compared to stress mismatch (conditions [ii] and [iv]). In addition there was enhanced frontal negativity for stress mismatch.

Although, both former priming studies revealed that prosodic processing is somewhat independent from phoneme processing, ERP stress priming remarkably differed in polarity between both studies. While there was enhanced posterior negativity for stress mismatch in the auditory–visual paradigm (Friedrich, Kotz, Friederici, & Alter, 2004; Friedrich, Kotz, Friederici, & Gunter, 2004), there was enhanced posterior negativity for stress match in the unimodal paradigm (Schild et al., 2014). Methodological differences between both studies might exert their influences here. On the one hand, targets were presented in different modalities. We used written target words in the auditory–visual study, but spoken target words in the unimodal study. Different target word modality might have modulated the ERP results. For example, the specific role that implicit prosody might play in visual word recognition (e.g., Ashby & Clifton, 2004; Ashby & Martin, 2008) might have driven the ERP stress priming effect in the cross-modal study.

On the other hand, the quick succession of spoken syllables together with the restriction to initially stressed target words might have elicited a unique response in the unimodal study (Schild et al., 2014). Two confounds could not be dissociated in