Behavioural and ERP evidence of prosodic processing in the cognitive representation of multimorphemic words

Zhongpei Zhang*
Laboratory Models, Dynamics, Corpora, CNRS, University Paris Nanterre, France

***Corresponding author: Zhongpei Zhang**
UMR 7114, CNRS
Laboratory MoDyCo (Models, Dynamics, Corpora)
Paris Nanterre University,
92001, Nanterre, France.
Phone : +33 667676159
E-mail :zhongpei.zh@hotmail.com;

Abstract

In the auditory modality, the recognition and identification of multimorphemic words involve the accessing of prosody information that contained in the lexical structure. Both the behavioral study and event-related potential research were constructed by the scholars, through the manipulations of acoustic cues and suprasegmental information (e.g., pitch contour, stress pattern, and syllable duration, etc.) of target words, both the behavioral results (e.g., reaction times) and brain responses of participants are analyzed to better understand the encoding procedure in the lexical-semantic representation. This article reviews a series of selected research, our goal is to provide evidence to account for how the prosody structure of multimorphemic words is processed in word recognition, and to what extent the deviant metrical structure of words can disrupt the lexical access.

Keywords: prosody, multimorphemic words, auditory processing, ERP research
DOI Number: 10.48047/nq.2023.21.6.nq23174 NeuroQuantology2023;21(6): 1754-1767

1 Introduction

The representation of spoken words involves a procedure to activate a correct word among various potential lexical candidates[1, 2]. Prosodic structure of multimorphemic words, which has been revealed to constrain the lexical activation in the early stage, can assist the activation of correct representation among several candidates[3-7]. In the tone languages, suprasegmental information contained in the tones can convey variant semantic meanings, listeners must process the prosodic information in the auditory input to realize the word recognition[8-10]. Speech signal provides both segmental and suprasegmental information, such as pitch (F0)and syllable duration [11, 12]. Suprasegmental information is accessed by listeners in the spoken word recognition, for example, listeners were revealed to be sensitive to some small subtle differences between the stressed and the non-stressed segments, and these contrasts are revealed to constrain the procedure of the lexical access[13, 14].The different syllables in a word may have different phonetic realizations in the variant prosodic contexts, e.g., the same acoustic segment would be longer and higher when it is stressed. In this context, thephonetic variability will be processed in the prelexical processing stage, in which the specific prosodic feature is extracted to entail the abstract stress pattern, which could facilitate the listeners to cope

with the auditory processing in word recognition.

2 Evidence from behavioral research

Some behavioral studies examined the role of prosodic structure in the early stage of word identification. To encode the online auditory processing, one of the mostly used experimentation tasks is lexical decision, this method is to measure the participant's response time when they decide whether or not the target stimulus is a real word [3, 15, 16]. In some of the previous behavioral research, by manipulating the prosodic structure of stimulus, researchers want to observe the effects of the ill-formed word's prosody on the auditory processing, and their goal is to understand how prosody structure influences the word recognition procedure[5, 17]. To understand the role of prosody in the cognitive representation of multimorphemic words, most behavioral research has concerned mis-stressed prosodic patterns and mismatched syllable length in word recognition.

2.1 Mis-stressed prosody pattern

A main effect of mis-stressing was found in the stress languages, such as Dutch and English. In Dutch, the metrical structure of words is stress-based, the strong stressed syllables mark the lexical segmentation in speech streams[18]. When the words were deliberately mis-stressed, their recognition process is delayed[19]. Certain types of stress misplacement tend to impair the word recognition. Cutler and Clifton [20]switched the position of stressed syllables within disyllabic compounds, e.g., *canTEEN* and *TURbine*. Behavioral results showed that misplaced substitutions impaired word identification of native listeners, andthe application of suprasegmental information in the stress-pattern processing helped listeners to solve the word embedding and segmentation problems. Goodglass and Wingfield [21]conducted the research with aphasia patients, prosody was found to prime the phonological access.

By using the cross-modal word fragment priming tasks, the word prosody was reported to modulate the behavioral response's latency[13, 14]. Participants' responses to target words in the stress matching prime–target pairs (e.g., muS–muSEUM) are faster than to the target words in the stress mismatching prime–target pairs (e.g., muS–MUsic). Soto-Faraco, Sebastián-Gallés [14] found that mis-stressing targets inhibit the behavior results. In a Spanish study, by using disyllabic fragments, slower responses were observed for the targets with mismatching prosodypatterns (e.g., princi–PRINcipe) than the completely unrelated targets (e.g., mosqui–PRINcipe),results indicated that the representation of partially mismatching targets receives inhibition from the completely matching candidates (e.g., prinCIpio [beginning]).

The evidence of the functions of stress in word recognition has been provided in previous studies, e.g., in Dutch [22], Italian[23], and Spanish [14]. Listeners have been found to be sensitive to the sub-lexical unit stress information in Dutch, suprasegmental stress is applied to resolve the competition problem, and they can distinguish *voor* taken from the word *VOORnaam* (first name) and *voorNAAM* (respectable) [13], Dutch listeners were found to use the stress information in the early stage of lexical recognition, before the delivery of the specific segmental disambiguation information. For example, in this eye-tracking study, *oktober* (October, stress on the second syllable) and *octopus* (octopus, stress on the first syllable) are accessed by listeners before the occurrence of the distinctive consonants /b/ and /p/ in speech signal [22]. The recognition of visually presented words is facilitated by prior auditory primes with the same first two syllables when they were appropriately stressed, e.g., *OKTOBER* can be primed by *okTO-*, in contrast to the inappropriate stressed primes (e.g., by *Octo-*), which produced inhibition in the recognition process [24].

Previous studies in Dutch also confirmed that listeners use stress patternsto resolve inter-word competition problems. For example, in Dutch, *zee* was detected more quickly in the word *luzee* than in the word *muzee* (which activates the Dutch word

museum). However, when *muzee* is pronounced with a stress on the initial position as *MUsee*, the stress pattern is no longer matched with the word *muSEum,* it no longer inhibits the detection of the word *zee*[13].Italian listeners relied on the lexical stress cues in online recognition. By using a cross-modal priming paradigm, listeners were tested with a lexical decision task on visual targets, results showed that the reaction time could be facilitated by these spoken bi-syllabic primes when the targets and primes are in the matched stress pattern, e.g., the target word *GOmito* ('elbow') is facilitated by the prime *GOmi*[23]. Similar results were found for Spanish listeners by using the same cross-modal paradigm in Soto-Faraco, Sebastián-Gallés [14], lexical stress is used by Spanish listeners in the same manner as segmental information to constrain word access.

2.2 Mismatched syllable duration

Most of the spoken word recognition models claimed that the acoustic signal is decoded into a string of phonemes to be accessed in the mental lexicon. The effect of syllable duration on word recognition has been observed in previous studies. During lexical processing, the extraction of the suprasegmental cues is adopted by listeners to recognize the syllabic structure of a word or between the words to solve the lexical-embedding and syllabic segmentation problems. In some cases, different durations of the segments canbe applied to disambiguate the similar-sound words. When two compounds are constructed by identical segments, a lexical competition problem might arise due to the variant competitive candidates in the mental lexicon, and the duration of segments facilitates the recognition of correct segments and determine the word boundary in the continuous input of speech signals.

In English, when a word is monosyllabic, its duration is longer than when it's embedded within a polysyllabic word, e.g., *cap* is longer than in *captain*[25]. And there are some ambiguous sequences, e.g., *tulips* and *two lips*,the /l/arefound to be longer on the initial position than on the non-initial position, listeners use this acoustic cue to recognize the ambiguous sequences [26]. The subphonemic durational difference across the multiple segments works as a recognition cue for listeners to process the ambiguous two-word sequences in Dutch. Different lengths of consonant /s/ within *een spot* (a spotlight) and *eens pot* (once jar) are used by listeners in the spoken word recognition[27]. The study ofQuené and Kager [28]examined the sequence pairs like *diep in* (deep in) and *die pin* (that pin), Dutch listeners use the durational differences of consonant /p/ to segment word pairs. In French, the duration of consonant /r/ is extracted and employed for the distinction of liaison, the results from the cross-modal priming experiments revealed that /r/ in the liaison environment '*dernier oignon'*is shorter than consonants in the word-initial position '*dernier rognon'*[29]. And the duration of the sub-lexical unit is also employed as an acoustic cue by Italian listeners to segment speech and access the lexical sequences which have the same initial syllables. For instance, with respect to the duration of /l/ in the final positionof the first syllable like in *sil.vestre* is quite different from /l/ in the initial position of second syllable like in*si.lencio*, listeners could disambiguate the syllables between the minimal pairs without hearing the following segments[30].

Evidence from some languages suggests that the information on syllable duration can be effectively exploited in word recognition [31]. Salverda, Dahan [32]conducted an object-naming study by monitoring participants' eye movement. Participants were instructed to click on the object which was mentioned in the phrase, and the length of the first syllable of the object's name was manipulated by replacing them with monosyllabic words.

3 Evidence from Event-related potential (ERP)research

Electroencephalogram (EEG) is the electrophysiological recordings of the oscillations of brain electric potentials preserved from the electrodes on the human scalp. Scalp EEG recording is an important tool to investigate the correlations between potential recorded measurements and underlying cognitive processing[33]. Previous

psycholinguistic research revealeda correlation between EEG signals and human cognitive processes, such as mental calculation, working memory, and selective attention. Event-related potential (ERP)refers to the electrical potentials which are related/time locked with the exogenous or endogenous events (stimuli). Some exogenous stimuli, e.g., a light flash or a piece of auditory recording, can be used to measure event-related potentials. ERPs provide a direct temporal measurement for the state-dependent brain processing of these stimuli, the small changes in the electrical activities of the human brain are recorded from the scalp, the activities change rapidly and have a spatially extended field. Averaged ERPs can measure electrical potentials due to the postsynaptic activities of the neural ensembles, which are generated in the extracellular fluid flowing across the neurons[34].These sensory evoked potentials involve the auditory, somatosensory, and visual modalities in speechprocessing[35].

Event related potential (ERP)couldprovide the detailed quantitative measurements in the research of diagnosis and treatment of central nervous diseases, and an accurate temporal solution is important to study the brain activities from one millisecond to the next[34].As ERPs provide the temporal dynamics of neural activities in response to stimuli which evoked them, this makes it an ideal tool as to observe the effect of a certain experimental manipulation. Because the processing of speech devoted to different linguistic information occur quickly, ERPs could record the neutral activities to explore the natural language processing, as well as the prediction research, because ERPs can provide a continuous monitor of brain responses that cover the full course of a discourse[35]. The peaks of stimuli-locked ERP waveforms, i.e., the delay between the stimuli onset and the time at which the maximum electrical response (maxima) is observed, allow us to investigate in real time how the participant's cognitive system responds to the manipulations of the stimulus. The auditory stimuli can produce a potential variation in the midbrain, which can be measured on the scalp with suitable techniques (EEG). Previous research concerning event-related potentials (ERPs) revealed that both the Romance, Germanic languages and tone languages listeners use the stress information contained in the speech signals to realizeword recognition. Some aspects of prosodic information were analyzed in the auditoryprocessing.At the word level, researchers focused on the sub-phonemic and supra-phonemic feature of syllables, such as syllable duration, pitch and fundamental frequency (F0), stress pattern of compounds; at the sentence level, research focused on the prosodic congruency of compounds in the whole sentence (Table 1).

Table 1. Summary of ERP research concerning the prosodic manipulation of target words.

Prosodic manipulationsof targets	**Research**	**ERP effects**
Suprasegmental information of sub-phonemic units		
Pitch contour	Friedrich, Alter [36]	P200
	Friedrich, Kotz [37]	P350
Syllable duration	Magne, Astésano [38]	N400-like effect
Congruency of consonant	Ashby, Sanders [39]	N200-like effect
	Ashby and Martin [40]	N200-like effect
	Luo et al. (2006)	MMNs
Congruency of syllable structure	Ashby [41]	N100
	Grainger, Kiyonaga [42]	N250
Congruency of tones	Tsang et al. (2011)	MMNs, P3a
	Luo et al. (2016)	N400
Tone & intonation	Ren et al. (2009)	MMN, N2b, P300
Stress pattern of compound words		
Stress pattern of Dutch words	Böcker, Bastiaansen [43]	N325

Congruency of Mandarin lexical tone and pitch accent	Wang et al. (2005)	MMNs, P3a
	Luo and Zhou [44]	P300-like, N400
	Li, Yang [45]	N400
Prosody patternsof Turkish word	Domahs, Genc [46]	N400, P300
Rhythmic violations of English compounds	Henrich, Alter [47]	N400
	Magne, Gordon [48]	N400
Accentuation of Mandarin words	Li, Yang [45]	N400
	Li and Ren [49]	N400
Prosodic congruency in the sentential level		
Omission of sounds	Rüsseler et al. (2001)	MMNs
	Yabe et al. (1997)	MMNs
Prosody incongruency in phrases	Steinhauer, Alter [50]	P600
	Eckstein and Friederici [51]	N400, P600
	Schmidt-Kassow and Kotz [52]	P600
	McCauley, Hestvik [53]	N400, P600

3.1 Suprasegmental features of sub-phonemic units

Linguistic function of prosody in the processing of multimorphemic words was analyzed in several acoustic parameters at the word level, such as the syllable duration andpitch (F0). With regard to the syllable duration of individual constituent, by using French trisyllabic spoken words, Magne, Astésano [38]manipulated the syllable duration as to investigate the influences of rhythmic property on the semantic processing. French is described as a language with fixed accents at the end of rhythmic units, which are characterized by lengthening the last syllable of words [54]. Magne and his collaborators used the cross-splicing procedure to apply the duration of the last syllable (lengthening) to the penultimate syllable (non-accented and normal duration) within trisyllabic words. This experimental manipulation aimed to create a 'metrical violated condition', in which target words have misplaced stress accent and unnatural syllable duration. For example, the word *canaPÉ* (sofa) is normally accented on the last syllable, while a metrically violated word **caNApé*is accented with incongruous lengthening penultimate syllable *na*. The semantic congruency of target wordswas manipulated to create a semantic violated condition at the sentence level. Sentences with the targets placed at the last position are aurally presented to participants. Results revealedan N400-like effect on the targets with misplaced stress accents, and semantically incongruous words elicited a larger N400 effect than semantically congruous words in the metrical judgment tasks.Researchers interpreted the enhanced N400 as a reflection of disrupted access in response to the semantic integration caused by the abnormal metrical pattern.

Considering the manipulation of pitch(F0), Friedrich, Alter [36]found that one of the ERP components, P200 is sensitive topitch contour processing, the initially unstressed words elicited a larger P200 effect than the initially stressed words. Based on these findings, Friedrich, Kotz [37] suggested that the pitch contour of words can modulate the amplitude of the P350 component, which is an ERP component that links to prosodic information processing in word identification. To examine tonal manipulation in auditory lexical processing, one of the primary experimental paradigms adopted by researchers is the passive auditory oddball paradigm. In some experiments, participants are reading a book or watching a silent movie while simultaneously they are presented auditorily with two different categories of stimuli, both frequent standard stimulus and infrequent deviant stimulus. Event-related potentials (ERPs) are recorded to investigate the neural responses to standard and deviant stimuli. Tsang et al. (2011) investigated the pre-attentive processing of Cantonese tones with a passive oddball paradigm,standard and deviant stimuli in the experiments differed in their pitch levels and pitch contours, and ERP was recorded when the native listeners were watching the silent movie. Results revealed

that tones were processed pre-attentively, MMN and P3a were elicited by all the types of deviant tones, and the modification in pitch level and pitch contour were respectively in relation to the amplitude and latency of MMN and P3a. Luo, Ni [55] examined the pre-attentive processing of Mandarin lexical tones with an auditory oddball paradigm. Native listeners were presented with stimuli in the consonant-vowel structure, and these stimuli were either manipulated in their tone patterns or their initial consonants. Results indicated that a more negative MMN effect was elicited by the stimulus with the deviant tone patterns in relation to the standard stimulus. Similar research has been conducted by Renet al. (2009), which analyzed the pre-attentive processing of Mandarin lexical tones and intonation. Ren et al. (2009) reported a larger amplitude of MMNs associated with deviant tones compared to deviant intonation. Additionally, N2b and P300 were known to reflect the cognitive procedure which will take place in the attentive processing stage during the auditory stimulus perception in an explicit oddball paradigm. N2b is increased negativity that follows MMNs in time and peaks at 200–300ms from the stimulus onset.Zheng, Minett [56]showed that N2b and P300 are sensitive to the detection of deviant Mandarin tones in the auditory modality, and larger amplitudes of the two components were elicited by a cross-category stimulus (e.g., two tones from the same category of Tone 2) than within-category stimulus (e.g., two tones respectively from the categories of Tone 2 and Tone 4).

Some of the research has combined eye-tracking and ERP to investigate lexical access. Luo et al. (2016) analyzed the application of phonological knowledge during the reading of Chinese scripts, especially the processing of neutral tones during the silent reading. Both disyllabic full tone and neutral tone compound words were selected, their second characters possess the same syllable structure with either a full tone or a neutral tone, a more-negative amplitude of N400 effect during the readings of neutral tone compounds in relation to the readings of full tone compounds. Native readers read the sentences with the two types of target compounds embedded within them, and the eye-tracking technique was applied to record the eye movements of participants in the readings. Results indicated shorter viewing durations, fewer refixations, and fewer regressive saccades on target words that contain a neutral tone. A similar result was observed by the eye-movement study of Yan et al. (2014), shorter viewing durations are observed forthe neutral tone compounds than forthe full tone compounds, and readers processed an upcoming parafoveal word less effectively whenneutral tone compounds were fixated than fixations on fulltone compounds.

Prelexical processing of sub-phonemic and supra-phonemic features was investigated by a series of works by Ashby and his colleagues. Ashby, Sanders [39]analyzed the congruency of the consonant pattern by applying the masked priming paradigm. In this study, pseudoword primes with a voiced or an unvoiced final consonant (e.g., fap, or faz) were followed by the words with either a congruent or an incongruent final consonant articulation (fap–fat [congruent] vs. fap–fad [incongruent]). Results showed that early ERP components were modulated by the congruency of consonants contained in the prime–targetpairs, the congruent condition has yielded less-negative amplitudes within the time window of 250-350ms post-target.Ashby and Martin [40] analyzed the processing ofsub-phonemic information in the visual word recognitionin English, and smaller negative amplitudes were elicited in the 250-350ms post-target when the first syllablesof prime and target were matched (pi### of pilot) than when they were mismatched (pil##### of pilot). Ashby [41]observed an earlier effect of syllable congruency, the amplitudes of N100 component (within the 100-120ms time window) were reduced when targets were preceded by the syllable-congruent primes.

3.2 Stress pattern of multimorphemic words

Event-related potentials have been applied to explore how the deviant prosody pattern can hinder lexical processing, the metrical function of the prosody pattern is firstexamined by Böcker, Bastiaansen

[43]through the manipulation of rhythmic property (metrical stress) of Dutch spoken words. Participants either listen passively or discriminately to sequences of four disyllabic words, which end with a word that had either the same or a different stress pattern (strong-initial / weak-initial) as the rest of words in the list. Results showed that the weak-initial words, which were much less common in the Dutch lexicon (about 12%), canelicit several larger negative components than the strong initial words. An enhanced left-frontal distributed negativityN325 at around 325ms post-stimuli is reported when weak-initial words are followed by a sequence of strong-initial words. N325islargerfor the participants who have good performance in both the metrical discrimination task and the passive listening task. Authors suggested that N325 is a neural correlate of metrical stress perception, which manifests the extraction of metrical stress from the acoustic signals.

Mandarin readers are revealed tousethe features of prosodic structure in the lexical access. Wang, Friedman [57] appliedan auditory oddball paradigm to investigate whether the prosodic salience (e.g., stressed syllables in disyllabic speech sound) can serve as the cues to capture listener's attention inspeech processing. Disyllabic standard and deviant input, which differ in prosodic and temporal status, are presented to participants. Two categories of stimuli with prosodic salience in different positions (***BAga and baGA***) are chosen as standards. According to the temporal position and prosodic status, voiced consonants are replaced by unvoiced consonants (e.g., /b/→/p/). The four types of deviants are manipulated and produced: (1) stressed initial: ***PAga***; (2) unstressed-noninitial: ***BAka***; (3)stressed-noninitial: ***baKA***; (4) unstressed-initial: ***paGA***. Listeners are tested while watching silent movies. Behavioral EEG results showthat prosodic salient deviance captures attention in unattended sounds, and MMNsare elicited by deviants in all the four conditions in the experiments, indicating the discrimination between the standard and the deviant input. P3a is observed when the deviance is placed on a stressed syllable (e.g., ***PAga*** and ***baKA***). The listener's sensitivity of P3a to the syllabic prosodic manipulation can be viewed as strong evidence for the hypothesis that prosodic salience induces the attentional resources during the involuntary speech sound perception. Stressed syllables work as attentional cues to facilitate speech perception, results suggest that the cognitive system of listeners might give priority to prosodically salient information irrespective of its temporal position in the speech. This finding proposes that prosodic salience can capturethe listeners' attention, and the segmentation is achieved through increased attention to stressed syllables rather than the word onsets.

To investigate to what extent the Mandarin prosodic information can constrain the silent reading, Luo and Zhou [44]manipulated the number of syllables and semantic congruency of verbs within the compound combinations in the Mandarin lexical combinations.E.g., the word in the verb-noun combination *种蒜*（plant + garlic）is a normal [1+1] rhythmic pattern, compared to *种植蒜*（plant + garlic）, which is an abnormal [2+1] rhythmic pattern. The critical words are embedded in the same position of phase, and both EEG signals and behavioral results of listeners are recorded. In the results, abnormal [2+1] rhythmic pattern is revealed to evoke both a posterior positivity P300-like effect (in the 300-600ms time-window) and afrontocentralN400-like effect in the semantically congruent sentences, this P300-like effect is interpreted to reflect the detection of rhythmic violation between [1+1] and [2+1]. The observation of theN400-like effect suggests that a mismatch between the incoming prosodic information and stored rhythmic pattern can affect the recognition of this current word. The results offer us evidence that abnormal rhythmic patterns can evoke the N400-like effect, and the additional semantic incongruency can enhance this effect. Moreover, Li, Yang [45] conducted ERP research to examine the congruency of pitch accent and lexical tone in the Mandarin spoken word processing. Results show that both inconsistent pitch accents and inconsistent lexical tone scan elicittheN400 effect in semantically congruent sentences,

and the negativities evoked by the inconsistent pitch accents had presented the same topographical distribution as the negativities evoked by inconsistent lexical tone violations.

Some research concerned the role of pitch accentuation in speech processing. Li, Yang [45] analyzed the role played by the accented speech information in the semantic analysis of Mandarin discourse, the amplitude of N400 component is revealed to be larger for the accented than for the deaccented wordsin the semantically congruent sentences. Moreover, Li and Ren [49] investigated how the analysis of prosody information can be influenced by the pitch accentuation in Mandarin spoken language processing, an increased N400 effect is revealed for the accented words.

Considering the role of violated metrical prosody structure in the lexical access, Magne, Gordon [48]examined metrically unexpected English words in the visual processing. Researchers manipulate the prosody pattern of the last target word with a contrastive stress pattern in relation to the others in the sequence). Results indicate that the frequent regular (standard) stress pattern can elicit expectancy about the stress pattern of the incoming word. N400 is observed in response to the processing of prosodically-deviant target words. This result supports the idea that incorrect rhythmic patterns would hinder the semantic processing in the lexical representation. In addition, Domahs, Genc [46]conducted an ERP studyon the predictability of prosody patterns in Turkish. To this end, the stress pattern of trisyllabic words is manipulated either with a standard final stress pattern (e.g., mɪ'sɪr,*''corn''*) or with an expected non final stress pattern (e.g., 'mɪsɪr, *''Egypt''*). Results reveal that asymmetrical ERP responses were revealed for thesetwo different stress violations:the final stress produced an N400 effect whereas stress violations with nonfinal stress produced a P300 effect. Henrich, Alter [47] investigated how rhythmic predictability and violation influence lexical processing, an alternation of stressed and unstressed syllables is manipulated for the minimal pairs of English noun-noun compounds, for example, the pairs of compounds *champàgne* (cocktail, stress on the final position) and *chàmpagne* (cocktail, stress shifted on the initial position). Results showed that rhythmic regularity is relevant for speech processing,the deviations contained in the rhythmic pattern evoke different ERP components, and an N400 is revealed for the items with the deviant stress patterns, which reflects a higher cost in lexical processing due to theprosody deviation.

3.3 Prosodic congruency at the sentence level

Some research investigated the automatic auditory discrimination. The mismatch negativity (MMN) is thought to reflect the processing of auditory changes contained in the input. This kind of experiment often has both two categories of stimuli (deviant and standard), and MMNs are viewed as an index to estimate the detections of the auditory changes in the incoming input. The response and its magnetic counterpart, i.e., MMNeffect is stronger when the acoustic differences between the deviants and the standards are much greater (e.g., the deviant stimulus and standard stimulus aren't in the same language-specific category). At the sentence level, the MMNeffect was revealed to be elicited by the omission of an element [58, 59].

Moreover, at the sentential level, through the manipulation of the prosody pattern of target words, researchers examined how the prosody incongruency of individual words can modulate the speaker's brain response. For example, Steinhauer, Alter [50]observed a P600 effect in the well-formed sentences which contained incongruous prosodic structure, authors interpreted this effect as a reflection of the cost of both syntactic and prosodic revisions which were evoked by targets.

Eckstein and Friederici [51]manipulated the critical noun in two different positions in the German phrases (both penultimate and final positions), the prosodic congruency of the targets is controlled to investigate whether the two levels (prosodic and syntactic) are interactive or independent from each other in language processing. Researchers observed

N400 in syntactically correct phrases, and prosodic incongruity enhanced this effect, which provides evidence for the reflection of a lexical integration cost to prosodically unexpected targets. P600 is observed in the syntactically correct sentences that contained a prosodic mismatch at the final word, which is interpreted as a reflection of intonation contour violation. There is an over-additivity toP600 in combined prosodic-syntactic violation condition in relation to pure syntactic/prosodic condition. Schmidt-Kassow and Kotz [52] explored the perceptual regularity of disyllabic trochee prosody pattern words in German. In the trochaic sentences, the words are disyllabic and their first syllable isstressed. Target words in the sentences are manipulated with metrical and/or syntactic violation, bothmetrical homogeneity task and grammaticality judgment task are applied to participants. E.g., 'Gina hatte Norbert gestern abend *kussen/*be'lohnen* sollen' (Gina should have *kissed/rewarded* Norbert yesterday evening), target wordsare manipulated with stress pattern deviations. Results show that brain is sensitive to stress violations when participants focus on the metrical structure in the metric tasks, even the subtle metric deviation of targets elicitsa P600, which reveals that metric and syntactic cues interact in the late integrational stage indexed by the P600,but this effect isn't observed in the grammatic tasks, when participants focus their attention in the syntactic structure. McCauley, Hestvik [53]observed an N400-like effect in response to the rhythmic violations during the silent reading. A picture/word matching task is used to examine the role of compound/phrasal stress distinction in online language processing. E.g., the picture (content: a green-colored house) is matched with the compound/phrasal pairs *greenhouse* (incongruent) and *green house* (congruent). Results reveal N400 effects in response to the violated stress patterns, which result from the incongruity between the content of the image and the semantic representation of the targets. Unexpected phrasal stress can elicit a P600 effect in the semantic incongruent condition, researcher interpreted the P600 effect as a reflection of the syntactic integration difficulty resulting from the unexpected stress pattern.

4 Discussion

The goal of this study is to address the role of prosodic information in the auditory lexical processing of multimorphemic words. In the first part, by reviewing the behavioral research, we focus on the mis-stressed prosody patterns and the mismatched syllable duration of target stimuli in word recognition. In the second part, a board review is constructed on previous ERP research, we focus on the evidence from three types of prosodic manipulations of target stimuli in the ERP research, including the suprasegmental information of sub-morphemic unit, the stress pattern of compounds, and prosodic congruency at the sentence level.

In the behavioral research, we first focused on the manipulated prosody patterns of the target words. The incorrect prosody pattern of target word can impair listener's word identification, when targets have mismatched stressed syllables, participant's response time to manipulated words was significantly shorter than to correct words, to this end, lexical stress plays an important role in the speech perception, the irregular or violated stress pattern of targets was revealed to inhibit the word's phonological representation [20]. And the stress pattern of sub-lexical unitsis found to be detected by listeners, in some experiments which applied the priming paradigm, mismatched stress pattern of words or the fragments of words is revealed to constrain lexical access [14]. Secondly, we concentrated on the mismatched syllable duration in the auditory lexical processing. As one of the suprasegmental cues, the length of each syllable can be extracted and accessed in the processing, previous research found that the syllable duration plays a role in the detection of word boundary in a continuum, and incorrect syllable duration could hinder the lexical representation [14]. Correct syllable duration can solve the lexical competition problem, particularly, when listener's cognitive system must select the correct representation form among various

candidates, the correct word would be represented due to the distinction of syllable length between similar-sound words [13]. To realize word recognition, the syllable length of words would be processed in the early stage of phonological representation.

In the ERP research, firstly, by analyzing the listeners' neural correlates, previous research has demonstrated that the suprasegmental information contained in the sub-lexical units would be encoded and processed in the lexical recognition. By manipulating the syllable duration of target compounds in the French sentences, the target words carrying an irregular lengthening syllable can evoke a larger N400 than the prosodically congruous words in semantically-incongruent sentence, the lexical-semantic representation was disrupted and hindered by the abnormal syllable duration [38]. With regard to fundamental frequency(F0) and tone patterns, deviant tones and violated pitch patterns can evoke N2b [60], P200 [36], P3a [57] and P350 [37] in the word identification. And by using the pre-attentive paradigm in the experiments, researcher found that violated suprasegmental information contained in the target words can elicit the MMNs, which is viewed as a reflection of the brain responses to the deviant prosody information [55]. Secondly, violated stressed patterns of target words can evoke the N325 [43], P3a [61], P300-like and N400 effect[44] in the auditory processing, when the prosody violations are in the semantically-incongruous sentence, the deviant stressed patterns can evoke the N400 effect, which reflects the difficulties of semantic integration in the lexical access. Finally, in the sentence level [53], prosody incongruencies, such as the deviant stress patterns and suprasegmental information of target words, can evoke N400 and P600 effects in the lexical-semantic processing of these sentences.

5 Conclusion

Taken together, prosodic information, which is contained not only in the multimorphemic words but also in their sub-morphemic units, affects the word recognition in the auditory modality, and the encoding of suprasegmental information will contribute to the lexical representation. The prosody incongruency, such as the violations of stress pattern of words, can modulate and disrupt the phonological representation, which is reported by both the behavioral and ERP studies. By reviewing the previous research, we conclude that the role played by prosody pattern has been well analyzed in the auditory lexical representation, and the listener's sensitivity has been demonstrated to the prosody information contained in the multimorphemic words, these findings can facilitate the future investigations of auditory lexical access.

Credit authorship contribution statement

Acknowledgment

This research was conducted with all the supports of colleagues in Laboratory MoDyCo of Paris Nanterre University. I am grateful to the colleagueswho helped in the writings. Authors also thank the anonymous reviewers and their suggestions which help a lot to improve the manuscript.

Funding

The author received no financial support in this article.

ORCID iD

Zhongpei ZHANG, https://orcid.org/0000-0002-8341-238X

Reference

1. Cutler, A., *B. Butterworth (Ed.), Language Production, Vol. 2: Development, Writing and other Language Processes*. 1984
2. Cutler, A., *Lexical stress.* 2005
3. Cutler, A., *Spoken word recognition and production.* 1995
4. Cutler, A. and S. Butterfield, *Rhythmic cues to speech segmentation: Evidence from juncture misperception.* Journal of memory and language, 1992. **31**(2): p. 218-236
5. Cutler, A., D. Dahan, and W. Van Donselaar, *Prosody in the comprehension of spoken language: A literature review.* Language and speech, 1997. **40**(2): p. 141-201.https://doi.org/10.1177/002383099704000203

6. Grosjean, F. and J.P. Gee, *Prosodic structure and spoken word recognition.* Cognition, 1987. **25**(1-2): p. 135-155
7. Levelt, W.J., *Accessing words in speech production: Stages, processes and representations.* Cognition, 1992. **42**(1-3): p. 1-22
8. Packard, J.L., *Lexical access in Chinese speech comprehension and production.* Brain and Language, 1999. **68**(1-2): p. 89-94.https://doi.org/10.1006/brln.1999.2102
9. Perfetti, C.A. and S. Zhang, *Very early phonological activation in Chinese reading.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 1995. **21**(1): p. 24.https://doi.org/10.1037/0278-7393.21.1.24
10. Zhou and Marslen-Wilson, *The nature of sublexical processing in reading Chinese characters.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 1999. **25**(4): p. 819.https://doi.org/10.1037/0278-7393.25.4.819
11. Fry, D.B., *Experiments in the perception of stress.* Language and speech, 1958. **1**(2): p. 126-152
12. Lehiste, I., *Suprasegmentals, Cambridge, Massachusetts & London, UK*. 1970, MIT Press.
13. Cutler, A. and W. Van Donselaar, *Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch.* Language and speech, 2001. **44**(2): p. 171-195.https://doi.org/10.1177/00238309010440020301
14. Soto-Faraco, S., N. Sebastián-Gallés, and A. Cutler, *Segmental and suprasegmental mismatch in lexical access.* Journal of memory and language, 2001. **45**: p. 412-432. https://doi.org/10.1006/jmla.2000.2783
15. Norris, D., et al., *Phonological and conceptual activation in speech comprehension.* Cognitive Psychology, 2006. **53**(2): p. 146-193.https://doi.org/10.1016/j.cogpsych.2006.03.001
16. McQueen, J.M. and A. Cutler, *Cognitive processes in speech perception*, in *The handbook of phonetic sciences*. 2010, Blackwell. p. 489-520.
17. Cutler, A. and C. Clifton, *Comprehending spoken language: A blueprint of the listener.* The neurocognition of language, 1999: p. 123-166
18. Cutler, A. and D. Norris, *The role of strong syllables in segmentation for lexical access.* Journal of Experimental Psychology: Human perception and performance, 1988. **14**(1): p. 113
19. van Heuven, V.J. and L. Menert, *Why stress position bias?* The Journal of the Acoustical Society of America, 1996. **100**(4): p. 2439-2451
20. Cutler, A. and C. Clifton, *The use of prosodic information in word recognition.* 1984
21. Goodglass, H. and A. Wingfield, *Anomia: Neuroanatomical and cognitive correlates*. 1997: Academic Press
22. Reinisch, E., A. Jesse, and J.M. McQueen, *Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately.* Quarterly Journal of Experimental Psychology, 2010. **63**(4): p. 772-783.https://doi.org/10.1080/17470210903104412
23. Tagliapietra, L. and P. Tabossi. *Lexical stress effects in Italian spoken word recognition*. in *Proceedings of the XXVII annual conference of the cognitive science society*. 2005. Stresa, Italy: Lawrence Erlbaum.
24. Van Donselaar, W., M. Koster, and A. Cutler, *Exploring the role of lexical stress in lexical recognition.* The Quarterly Journal of Experimental Psychology Section A, 2005. **58**(2): p. 251-273.https://doi.org/10.1080/02724980343000927

25. Lehiste, I., *The timing of utterances and linguistic boundaries.* The Journal of the Acoustical Society of America, 1972. **51**(6B): p. 2018-2024
26. Gow, D.W. and P.C. Gordon, *Lexical and prelexical influences on word segmentation: evidence from priming.* Journal of Experimental Psychology: Human perception and performance, 1995. **21**(2): p. 344
27. Shatzman, K.B. and J.M. McQueen, *Segment duration as a cue to word boundaries in spoken-word recognition.* Perception & Psychophysics, 2006. **68**(1): p. 1-16.https://doi.org/10.3758/bf03193651
28. Quené, H. and R. Kager, *The derivation of prosody for text-to-speech from prosodic sentence structure.* Computer Speech & Language, 1992. **6**(1): p. 77-98
29. Spinelli, E., J.M. McQueen, and A. Cutler, *Processing resyllabified words in French.* Journal of memory and language, 2003. **48**(2): p. 233-254.https://doi.org/10.1016/S0749-596X(02)00513-2
30. Tabossi, P., et al., *Syllables in the processing of spoken Italian.* Journal of Experimental Psychology: Human perception and performance, 2000. **26**(2): p. 758.https://doi.org/10.1037/0096-1523.26.2.758
31. Zhou, X. and W. Marslen-Wilson, *Words, morphemes and syllables in the Chinese mental lexicon.* Language and Cognitive Processes, 1994. **9**(3): p. 393-422.https://doi.org/10.1080/01690969408402125
32. Salverda, A.P., D. Dahan, and J.M. McQueen, *The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension.* Cognition, 2003. **90**(1): p. 51-89.https://doi.org/10.1016/S0010-0277(03)00139-2
33. Luck, S.J., *Ten simple rules for designing and interpreting ERP experiments.* Event-related potentials: A methods handbook, 2005. **4**
34. Kutas, M. and A. Dale, *Electrical and magnetic readings of mental functions*. Cognitive neuroscience. 2013: Psychology press. 197-242.https://doi.org/10.4324/9780203765401
35. Nunez, P.L. and R. Srinivasan, *Electric fields of the brain: the neurophysics of EEG*. 2006: Oxford University Press, USA
36. Friedrich, C.K., K. Alter, and S.A. Kotz, *An electrophysiological response to different pitch contours in words.* NeuroReport, 2001. **12**(15): p. 3189-3191.https://doi.org/10.1097/00001756-200110290-00009
37. Friedrich, C.K., et al., *Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence.* Cognitive Brain Research, 2004. **20**(2): p. 300-308.https://doi.org/10.1016/j.cogbrainres.2004.03.007
38. Magne, C., et al., *Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence.* Cerebral cortex, 2007. **17**(11): p. 2659-2668.https://doi.org/10.1093/cercor/bhl174
39. Ashby, J., L.D. Sanders, and J. Kingston, *Skilled readers begin processing sub-phonemic features by 80 ms during visual word recognition: Evidence from ERPs.* Biological psychology, 2009. **80**(1): p. 84-94.https://doi.org/10.1016/j.biopsycho.2008.03.009
40. Ashby, J. and A.E. Martin, *Prosodic phonological representations early in visual word recognition.* Journal of Experimental Psychology: Human perception and performance, 2008. **34**(1): p. 224. https://doi.org/10.1037/0096-1523.34.1.224
41. Ashby, J., *Phonology is fundamental in skilled reading: Evidence from ERPs.* Psychonomic Bulletin & Review, 2010.

17(1): p. 95-100.https://doi.org/10.3758/PBR.17.1.95

42. Grainger, J., K. Kiyonaga, and P.J. Holcomb, *The time course of orthographic and phonological code activation.* Psychological science, 2006. **17**(12): p. 1021-1026.https://doi.org/10.1111/j.1467-9280.2006.01821.x
43. Böcker, K.B., et al., *An ERP correlate of metrical stress in spoken word recognition.* Psychophysiology, 1999. **36**(6): p. 706-720. https://doi.org/10.1111/1469-8986.3660706
44. Luo, Y. and X. Zhou, *ERP evidence for the online processing of rhythmic pattern during Chinese sentence reading.* NeuroImage, 2010. **49**(3): p. 2836-2849.https://doi.org/10.1016/j.neuroimage.2009.10.008
45. Li, X., Y. Yang, and P. Hagoort, *Pitch accent and lexical tone processing in Chinese discourse comprehension: An ERP study.* Brain Research, 2008. **1222**: p. 192-200.https://doi.org/10.1016/j.brainres.2008.05.031
46. Domahs, U., et al., *Processing (un-) predictable word stress: ERP evidence from Turkish.* Language and Cognitive Processes, 2013. **28**(3): p. 335-354.https://doi.org/10.1080/01690965.2011.634590
47. Henrich, K., et al., *The relevance of rhythmical alternation in language processing: An ERP study on English compounds.* Brain and language, 2014. **136**: p. 19-30.https://doi.org/10.1016/j.bandl.2014.07.003
48. Magne, C., R.L. Gordon, and S. Midha. *Influence of metrical expectancy on reading words: An ERP study*. in *Speech Prosody 2010-Fifth International Conference*. 2010.
49. Li, X.-q. and G.-q. Ren, *How and when accentuation influences temporally selective attention and subsequent semantic processing during on-line spoken language comprehension: An ERP study.* Neuropsychologia, 2012. **50**(8): p. 1882-1894.https://doi.org/10.1016/j.neuropsychologia.2012.04.013
50. Steinhauer, K., K. Alter, and A.D. Friederici, *Brain potentials indicate immediate use of prosodic cues in natural speech processing.* Nature neuroscience, 1999. **2**(2): p. 191-196. https://doi.org/10.1038/5757
51. Eckstein, K. and A.D. Friederici, *Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs.* Cognitive Brain Research, 2005. **25**(1): p. 130-143.https://doi.org/10.1016/j.cogbrainres.2005.05.003
52. Schmidt-Kassow, M. and S.A. Kotz, *Event-related brain potentials suggest a late interaction of meter and syntax in the P600.* Journal of Cognitive Neuroscience, 2009. **21**(9): p. 1693-1708.https://doi.org/10.1162/jocn.2008.21153
53. McCauley, S.M., A. Hestvik, and I. Vogel, *Perception and bias in the processing of compound versus phrasal stress: Evidence from event-related brain potentials.* Language and speech, 2013. **56**(1): p. 23-44.https://doi.org/10.1177/0023830911434277
54. Delattre, P., *Les dix intonations de base du français.* French review, 1966: p. 1-14
55. Luo, H., et al., *Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants.* Proceedings of the National Academy of Sciences, 2006. **103**(51): p. 19558-19563.https://doi.org/10.1073/pnas.0607065104
56. Zheng, H.-Y., et al., *The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study.* Language and Cognitive Processes, 2012. **27**(2): p. 184-

209.https://doi.org/10.1080/01690965.2010.520493

57. Wang, J., et al., *ERP correlates of involuntary attention capture by prosodic salience in speech.* Psychophysiology, 2005. **42**(1): p. 43-55. https://doi.org/10.1111/j.1469-8986.2005.00260.x
58. Yabe, H., et al., *Temporal window of integration revealed by MMN to sound omission.* Neuroreport, 1997. **8**(8): p. 1971-1974
59. Rüsseler, J., et al., *Event-related brain potentials to sound omissions differ in musicians and non-musicians.* Neuroscience letters, 2001. **308**(1): p. 33-36.https://doi.org/10.1016/S0304-3940(01)01977-2
60. Ren, G.-Q., Y. Yang, and X. Li, *Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity.* Neuroscience, 2009. **162**(1): p. 87-95.https://doi.org/10.1016/j.neuroscience.2009.04.021
61. Tsang, Y.-K., et al., *ERP correlates of pre-attentive processing of Cantonese lexical tones: The effects of pitch contour and pitch height.* Neuroscience Letters, 2011. **487**(3): p. 268-272.https://doi.org/10.1016/j.neulet.2010.10.035