Duration Approaches of Prosody

Subjects: Language & Linguistics
Contributor: Mercedes Muñetón-Ayala, Manuel de Vega, John Ochoa-Gómez, David Beltrán

Prosody is a complex aspect of communicative speech act that requires the successful integration of multiple acoustic parameters, such as fundamental frequency (F0), duration and intensity, whose perceptual correlates are pitch, timing and loudness, respectively, all of which contribute to the perception of the suprasegmental structure of sentences.

Keywords: prosody ; duration ; prosodic ; EEG

1. Introduction

Prosody is a complex aspect of communicative speech act that requires the successful integration of multiple acoustic parameters, such as fundamental frequency (F0), duration and intensity, whose perceptual correlates are pitch, timing and loudness, respectively, all of which contribute to the perception of the suprasegmental structure of sentences. In several studies related to Spanish language, the most studied parameter has been F0, which, according to some scholars, plays the principal role in marking prominent syllables in speech [1]. Yet, some recent studies have also shown the importance of duration and intensity [2][3][4][5][6]. Duration, is defined as the time taken to utter any part of the speech signal [4], for instance, of the syllable.

In general, languages have been described to fall into one of two rhythmic, mutually exclusive categories based on isochrony, depending on whether they have equal intervals of time for syllables or stress, called syllable-timed languages and stress-timed languages, respectively [7]. The former are languages with stable syllable duration, such as French, Italian or Spanish, and the latter are languages with stable duration for interstress interval, such as English, Dutch or Arabic [8][9]. However, some studies do not support the isochrony principle to differentiate languages. For example, in a cross-linguistic study (English, Thai, Spanish, Italian and Greek) in which informants had to read a passage from a contemporary novel translated to their own language, Dauer [10] showed that rhythmic grouping occurs more or less regularly not only in English (stress-timed language) but also in Spanish (syllable-timed language). Additionally, Dorta and Mora [11] analyzed the rhythmic characteristics of two dialectal varieties of spoken Spanish in the Canary Islands and Venezuela. They focused on the syllable as the rhythmic unity, studying the timing behavior of the syllabic nucleus, according to different segmental and suprasegmental factors. The results did not support the hypothesis of syllable-timed rhythm in Spanish because stressed syllables were longer than unstressed syllables. The same results have been found by other researchers in Spanish language studies [12][13][14][15].

Additionally, it is well known that vowel or syllabic duration has a phonological and a phonetic dimension with specific functional consequences. Its phonological dimension allows the distinction of one word from another in some languages. For example, in Japanese, there are pairs such as /isso/ ("rather") vs. /isso:/ ("more"); in Finnish, there are triplets such as /tule/ ("come") vs. /tule:/ ("comes") vs. /tu:le/ ("it blow") [4]; meanwhile, the phonetic dimension does not change the meaning of a word as in Spanish. However, as mentioned above, in this language people can differentiate between long (tonic syllable) and short vowels (adjacent syllables [16] (p. 55)). These characteristics make Spanish an interesting language for studying the influence of duration on the perception of sentences.

2. Duration Approaches

The relevance of duration features in prosody has been investigated throughout different approaches, such as neuropsychological [17][18] and electrophysiological [19][20][21][22][23]. Neuropsychological studies have shown that duration features in prosody can be used to distinguish lexical items both in speech perception [17] and production [18]. The results agree in pointing out that temporal information, such as syllable duration, is processed in the left hemisphere. For example, Van Lancker and Sidtis [17] found that left-hemisphere-damaged patients (LHD) and right-hemisphere-damaged patients (RHD) utilize acoustic cues differently to make a prosody judgment task. Namely, syllable duration variability was the principal cue used by RHD patients, while F0 variability was used by LHD patients, indicating that activation of the left hemisphere and right hemisphere is related to the durational cue and the F0 cue, respectively. In the same vein, Yang and

Van Lancker [18] investigated the production of idiomatic and literal expression in left- and right-hemisphere-damaged and in normal Korean speakers. The major finding was that distinguishing the two types of expressions relied on F0 changes in speakers with LHD, while duration was the main cue in speakers with RHD, confirming that the left hemisphere is specialized to process temporal cues, and the right hemisphere is specialized to process pitch cues. However, no conclusive information is available about this subject. A different study showed that the linguistic prosody process is as bilateral as the emotional prosody [24] or that prosody is lateralized to the right irrespective of the communicative function [25].

The event-related potential technique (ERP) has found a signature pattern of brain activity to index semantic or phonetic congruency during language comprehension [26]. N400 is one of the most important components in the research of language, with a negative polarity that peaks around 400 ms after the word onset. Specifically, it is a useful signature pattern for addressing questions on the integration of prosodic information in auditory processing [20], providing a measure of the time course of prosodic integration in semantic [22][23] and syntactic processes [20][27]. In these studies, researchers manipulated suprasegmental characteristics using words with incorrect lexical/metrical stress patterns [21][28][29][30], words with correct but unexpected stress patterns [22][23][31][32] or unexpected stress patterns in words or pseudowords [33]. The results indicate that subjects are aware of the stimuli changes, and their brain is sensitive to the subtle violation of rhythmical structure, which can influence the semantic encoding. For instance, Böcker et al. [31] investigated how listeners process words starting with the alternation of strong and weak syllables in a stress-timed language, such as Dutch. They found that initially weak words, as compared to initially strong words, elicited a negative brain response, probably related to stress discrimination, peaking at around 325 ms post-stimulus onset and maximum at the frontocentral scalp, likely reflecting the modulation of an anterior N400 component. An ERP study by Bohn et al. [32] that manipulated the German rhythm (alternation of stressed and unstressed syllable) in auditorily presented words found that where irregular but possible meter words involve semantic cost (i.e., enhanced N400 to unexpected stress change), those with regular meter do not. This suggests that regular meter is required to avoid additional cost in the semantic processing of words.

Furthermore, other studies that have manipulated syllable duration have also found modulations of both the N400 and a late positive component (LPC). For example, Magne et al. [21] used a design to examine the relationship between semantic and prosodic processing in spoken French. In this case, the prosodic violation was realized as a lengthened instance of the pretonic syllables (the second one) of the critical trisyllabic word. The result showed the on-line processing of the metric structure of the words. They found an N400-like negativity and a LPC (P600) in the prosodic judgment task but only the N400 effect in the semantic judgment task for the incongruent lengthening. This suggested the automaticity of metrical structure processing and demonstrated that violations of a word’s metric structure may hinder lexical access and word comprehension.

Similar LPC effects were reported by Astesano et al. [19]. They used semantically congruous and incongruous sentences, and sentences whose prosody matched or mismatched its syntactic form, by cross-shifting sentence beginning and ending. Therefore, they created four conditions: (1) both semantically and prosodically congruous; (2) semantically congruous and prosodically incongruous; (3) semantically incongruous and prosodically congruous; and finally, (4) both semantically and prosodically incongruous. Regarding prosody, they found a late positive component (P800) associated with prosodic mismatch. The late positivity was mediated by the task demand because it only emerged when prosody was in task focus. In the same vein, Paulmann et al. [34] compared the linguistic and emotional functions of prosody. The objective was to analyze whether the two prosodic functions engage a similar time course or not. To this aim, they merged a prosodically neutral head of a sentence to a second half of a sentence that differed in emotional and/or linguistic prosody. Consequently, the study consisted of: an emotional task in which participants judged whether the sentence that they had just heard was spoken in a neutral tone of voice or not; and a linguistic task in which participants decided whether the sentence was a declarative sentence or not. As was expected, the results reported a prosodic expectancy positivity irrespective of the task, but the latency for the linguistics prosody effect was later (~620 ms) than the latency for the emotional prosody violation (~470 ms).

In general, these ERP findings reflect the influence of prosody on comprehension taking into account its linguistic functions, such as lexical access/integration [21][35] or judgment of the sentence modality [19].

References

1. Kochanski, G.; Grabe, E.; Coleman, J.; Rosner, B. Loudness Predicts Prominence: Fundamental Frequency Lends Little. J. Acoust. Soc. Am. 2005, 118, 1038–1054.

2. Cabrera, F. Stress and Intonation in Spanish for Affirmative and Interrogative Sentences. In Proceedings of the EUROSPEECH'95, Fourth European Conference on Speech Communication and Technology, Madrid, Spain, 18–21 September 1995; pp. 2085–2088.
3. Candia, L.; Urrutia Cárdenas, S.; Fernández Ulloa, T. Rasgos Acústicos de la Prosodia Acentual del Español. Boletín Filol. 2006, 41, 11–44.
4. Fox, A. Prosody Features and Prosodic Structure; Oxford University Press: Oxford, UK, 2000; ISBN 0-191-823785-5.
5. Muñetón-Ayala, M.; Dorta, J. Estudio de la Duración En El Marco de la Entonación de las Principales Ciudades de Colombia . Estud. Fonética Exp. 2019, 28, 161–184.
6. Muñetón-Ayala, M. Asociación de la F0, Duración e Intensidad en el Habla de una Mujer de Medellín (Colombia) en Función de la Modalidad Oracional y sus Sintagmas
7. Varnet, L.; Ortiz-Barajas, M.C.; Erra, R.G.; Gervain, J.; Lorenzi, C. A Cross-Linguistic Study of Speech Modulation Spectra. J. Acoust. Soc. Am. 2017, 142, 1976–1989.
8. Abercrombie, D. Elements of General Phonetics; Edinburgh University Press: Edinburgh, UK, 1967; ISBN 9780852240281.
9. Pike, K.L. The Intonation of American English; University of Michigan Press: Ann Arbor, MI, USA, 1945; ISBN 0472087312.
10. Dauer, R.M. Stress-Timing and Syllable-Timing Reanalyzed. J. Phone 1983, 11, 51–62.
11. Dorta, J.; Mora, E. Patrones Temporales en dos Variedades del Español Hablado en Venezuela y Canarias. . Rev. Int. Lingüística Iberoam. 2011, 9, 91–100.
12. Canellada, M.J.; Madsen, J.K. Pronunciación del Español: Lengua Hablada y Literaria; Castalia: Madrid, Spain, 1987; ISBN 84-7039-483-5.
13. Clegg, J.H.; Fails, W.C. Structure of the Syllable and Syllable Length in Spanish. Proc. Deseret Lang. Linguist. Soc. Symp. 1987, 13, 47–54.
14. Muñetón Ayala, M. La F0, Duración e Intensidad de las Oraciones Interrogativas Absolutas en un Informante Varón de Medellín. Estud. Fonética Exp. 2016, 25, 167–192.
15. Ortega-Llebaria, M.; Prieto, P. Acoustic Correlates of Stress in Central Catalan and Castilian Spanish. Lang. Speech 2011, 54, 73–97.
16. Quilis, A.; Fernández, J. Curso de Fonética y Fonología Españolas; CSIC: Madrid, Spain, 1972; ISBN 8400070887.
17. Van Lancker, D.; Sidtis, J.J. The Identification of Affective-Prosodic Stimuli by Left- and Right- Hemisphere-Damaged Subjects: All Errors Are Not Created Equal. J. Speech Hear. Res. 1992, 35, 963–970.
18. Yang, S.; Van Lancker Sidtis, D. Production of Korean Idiomatic Utterances Following Left- and Right-Hemisphere Damage: Acoustic Studies. J. Speech Lang. Hear. Res. 2016, 59, 267–280.
19. Astésano, C.; Besson, M.; Alter, K. Brain Potentials during Semantic and Prosodic Processing in French. Cogn. Brain Res. 2004, 18, 172–184.
20. Eckstein, K.; Friederici, A.D. It's Early: Event-Related Potential Evidence for Initial Interaction of Syntax and Prosody in Speech Comprehension. J. Cogn. Neurosci. 2006, 18, 1696–1711.
21. Magne, C.; Astesano, C.; Aramaki, M.; Ystad, S.; Kronland-Martinet, R.; Besson, M. Influence of Syllabic Lengthening on Semantic Processing in Spoken French: Behavioral and Electrophysiological Evidence. Cereb. Cortex 2007, 17, 2659–2668.
22. Magne, C.; Jordan, D.K.; Gordon, R.L. Speech Rhythm Sensitivity and Musical Aptitude: ERPs and Individual Differences. Brain Lang. 2016, 153–154, 13–19.
23. Moon, H.; Magne, C. Noun/Verb Distinction in English Stress Homographs: An ERP Study. Neuroreport 2015, 26, 753–757.
24. Witteman, J.; Van Ijzendoorn, M.H.; Van de Velde, D.; Van Heuven, V.J.J.P.; Schiller, N.O. The Nature of Hemispheric Specialization for Linguistic and Emotional Prosodic Perception: A Meta-Analysis of the Lesion Literature. Neuropsychologia 2011, 49, 3722–3738.
25. Friederici, A.D.; Alter, K. Lateralization of Auditory Language Functions: A Dynamic Dual Pathway Model. Brain Lang. 2004, 89, 267–276.
26. Kutas, M.; Federmeier, K. Thirty years and counting: Finding meaning en the N400 component of the event related brain potential (ERP). Ann. Rev. Psychol. 2011, 62, 621–647.

27. Morgan, E.U.; van der Meer, A.; Vulchanova, M.; Blasi, D.E.; Baggio, G. Meaning before grammar: A review of ERP experiments on the neurodevelopmental origins of semantic processing. Psychon. Bull. Rev. 2020, 27, 441–464.
28. Domahs, U.; Bornkessel-schlesewsky, I.; Schlesewsky, M. The Processing of German Word Stress: Evidence for the Prosodic Hierarchy. Phonology 2008, 25, 1–36.
29. Marie, C.; Magne, C.; Besson, M. Musicians and the Metric Structure of Words. J. Cogn. Neurosci. 2011, 23, 294–305.
30. McCauley, S.M.; Hestvik, A.; Vogel, I. Perception and Bias in the Processing of Compound versus Phrasal Stress: Evidence from Event-Related Brain Potentials. Lang. Speech 2012, 56, 23–44.
31. Böcker, K.B.E.; Bastiaansen, M.C.M.; Vroomen, J.; Brunia, C.H.M.; De Gelder, B. An ERP Correlate of Metrical Stress in Spoken Word Recognition. Psychophysiology 1999, 36, 706–720.
32. Bohn, K.; Knaus, J.; Wiese, R.; Domahs, U. The Influence of Rhythmic (Ir)Regularities on Speech Processing: Evidence from an ERP Study on German Phrases. Neuropsychologia 2013, 51, 760–771.
33. Rothermich, K.; Schmidt-Kassow, M.; Schwartze, M.; Kotz, S.A. Event-Related Potential Responses to Metric Violations: Rules versus Meaning. Neuroreport 2010, 21, 580–584.
34. Paulmann, S.; Jessen, S.; Kotz, S.A. It’s Special the Way You Say It: An ERP Investigation on the Temporal Dynamics of Two Types of Prosody. Neuropsychologia 2012, 50, 1609–1620.
35. Eckstein, K.; Friederici, A. Late Interaction of Syntactic and Prosodic Processes in Sentence Comprehension as Revealed by ERPs. Brain Res. Cogn. Brain Res. 2005, 25, 130–143.