

Review of: "A Review of Prosody, Punctuation, and Dyslexia: Implications for the Use of Speech Technologies"

Gabrijela Kisicek

Potential competing interests: The author(s) declared that no potential competing interests exist.

This paper presents an interesting view on prosody in connection to speech technologies with a clear purpose to help students with dyslexia. However, the view on prosody is quite simplified. It has been well known that "certain features of oral communication have always been difficult to capture in writing, such as the changes in dynamics from loud to soft, the variations in pitch from high to deep etc. Fahnestock, 2011, p. 204). Prosody is far more than just a pause (or comma, colon, semi-colon or dash as writing equivalents). Although authors write: "This paper considers all aspects of prosody" there has been very little attention paid to the specific elaboration of prosodic features. Authors mention difference between implicit and explicit prosody which is manifested in reading silently (implicit) or prosody produced in speech (explicit). However, puzzling part of this paper is: "Specifically, according to Fodor's (2002) Implicit Prosody Hypothesis, the prosodic structures generated during silent reading match the prosodic structures produced when reading aloud. Therefore, implicit prosody is language dependent and reflects the accent spoken by the reader (Filik & Barber, 2011; Jun, 2010). This has implications for the development of reading, which might not rely exclusively on decoding, but also on prosody (Godde et al., 2020; Groen et al., 2019)." Structures of implicit and explicit prosody match each other and the conclusion (marked by therefore) is that implicit prosody is language dependent and reflects the accent spoken by reader. I do not understand what authors wanted to say. Because many different prosodic features depend on the speaker: voice quality might be the first that comes to mind. What do authors mean by accent is another ambiguity. Do they consider accent as a sign of regional variety? Is it sign of language specificities (for instance, English with Russian or French accent). Or do they refer to idiosyncratic features of a specific speaker which depends on pronunciation of certain vowels or consonants? And in which prosodic feature would that idiosyncratic dimension be manifested? In my opinion, accent represents a distinctive pattern of pronunciation that is usually associated with a regional dialect, culture, co-culture, or socioeconomic group. It is the most important feature for recognizing different regional or dialectal background, and for recognizing a speaker of a foreign language. Accent is manifested through the pronunciation of vowels or certain consonants. It is well known that every language in addition to its Standard variety has different dialectal and regional varieties. However, we need to distinguish dialects which include pronunciation, grammar, vocabulary, and idioms from accents for which only pronunciation and intonation are important. It has to be emphasized that accents contribute to the perception not only of the regional origin, but also of the social status and education of speakers. I think that in this paper authors should have emphasized specific prosodic features which are important as an aid to students with dyslexia and which could be somehow marked in a written text. Prosodic features refer to both voice and speech cues of the speaker. They include features such as pitch, temporal structure, loudness and voice quality, emphasis and accentuation, and also (non)fluencies of the speaker.

Prosodic features which might be significant for this topic, are, in my opinion:

1. **Pitch** is often explained with its acoustical correlate, the fundamental frequency, which is measured in Hertz. Pitch is informally identified as the highness or lowness of a sound and it is always connected to overall voice quality. Pitch can also rise or be lowered depending on the emotional state of the speaker. And, it can have grammatical function. Knapp, Hall and Horgan (2014, p. 324) explain: we manipulate vocal pitch to indicate the end of a declarative sentence (by lowering it) or a question (by raising it). Sometimes we consciously manipulate our tone to contradict the verbal message, as in sarcasm.
2. **Intonation** is the “tune” of the sentence, (i.e., variations in pitch) which usually differs between a statement and a question. However, intonation might also give as a clue on the emotional state of a speaker, depending on intonation endings which can be rising or falling, fall-rise or rise-falling. So, rising intonations can reflect excitement and joyfulness, while falling intonations can signal sadness. In the connection to intonation, staccato and legato rate of articulation have to be mentioned because they can also indicate emotional state of the speaker. Staccato is characterized by sharp transitions between sound and silence, and it is connected with the perception of dramatic, intensive speech together with the perception of order. On the other hand, legato speech is the complete opposite, characterized by smooth transitions between sound and silence. It is perceived as soft and gentle speech connected with sensitive personality.
3. **Pauses** contribute to discontinuous speech, and can have grammatical, lexical-semantic function (corresponding to the interpunction of the written text), and nongrammatical, i.e. stylistic function (when a speaker wants to emphasize a certain part of a speech). Of course, they can also be the result of hesitation.
4. **Emphasis** reflects the hierarchical structure of the words in a sentence. Together with loudness and intonation, it indicates the most important word in a sentence depending on what the speaker wants to communicate. For instance, in the sentence *John is reading a book*, the speaker is giving information about *who* is reading a book. It is not Bob, or Michael, but John. Depending on the context, a speaker could say *John is reading a book*, if someone perhaps thought that John decided to write a book. And the third possibility is that *John is reading a book*; if someone is surprised that John is not reading magazines or newspapers, it has to be emphasized that John is reading a book.

Other prosodic features such as voice quality (each person’s voice has a specific quality which results from anatomical and physiological characteristics e.g. nasal, breathy, hoarse voice) or speech rate or loudness/intensity or rhythm may have important role in communication but are hard to capture on paper and would not assist people with dyslexia to overcome their difficulties.

In conclusion, I believe that the paper discusses the important topic and that connecting prosodic features with speech technology is interesting. However, to be able to cover this topic, one should be more familiar with the research of prosodic features and their communicative function.