Journal of Phonetics 70 (2018) 20–38

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

What makes a word prominent? Predicting untrained German listeners' perceptual judgments

Stefan Baumann [a,*], Bodo Winter [b]

[a] IfL Phonetik, University of Cologne, Germany
[b] Department of English Language and Applied Linguistics, University of Birmingham, United Kingdom

ARTICLE INFO

Article history:
Received 22 June 2017
Received in revised form 4 December 2017
Accepted 12 May 2018

Keywords:
Prominence
Intonation
Perception
Random forests
Individual differences
German

ABSTRACT

One important feature of linguistic communication is that some parts of utterances are more prominent than others. Prominence as a perceptual feature of spoken language is influenced by many different linguistic variables, but it is not clear how these variables interact in perception and which variables are most important for determining prominence. We report results from a prosody transcription task which assessed how untrained German listeners are simultaneously affected by gradient signal-based factors such as pitch, intensity and duration, as well as discrete prosodic factors (pitch accent type and placement) and non-prosodic factors (semantic-syntactic, lexical). All 17 linguistic variables tested were reliably associated with listeners' prominence judgments. We used random forests, a data mining algorithm, to uncover which variables are most important in determining the prominence judgments. This analysis showed that discrete prosodic variables relating to intonational phonology, specifically the type of pitch accent and its position, were most predictive of prominence. At the same time, prominence judgments were characterized by large individual differences. An exploratory cluster analysis suggests that some listeners pay more attention to pitch-related variables but less to semantic-syntactic and lexical variables, while others do the reverse. Our results paint a complex picture of prominence perception that is highly variable across listeners, but that assures robust communication of prominence information through the simultaneous perceptual cueing of many different linguistic variables.

1. Introduction

When speakers communicate with each other, not all information is equally important. Some parts of an utterance are intrinsically more informative, such as novel discourse topics and uncommon words, while some parts are actively highlighted by speakers as being important through prosodic and syntactic means. As a result of both semantic-pragmatic *importance* and prosodic and syntactic *highlighting* (Streefkerk, 2002), listeners perceive certain utterance parts as more or less prominent. Loosely defined, "perceptual prominence" refers to any aspect of speech that somehow "stands out" to the listener.

As an analogy for prominence in speech, we may consider a tree standing alone on an empty field. This tree is more prominent than a tree in a forest, since it differs in height, shape and color from its environment. What determines perceptual prominence in speech is much less well understood. Already at the level of language structure, there is a host of potential cues for prominence, including the speaker's choice of words, syntactic constructions, and pitch accents. Then, within the more circumscribed domain of prosody, many phonetic variables are associated with prominence, including pitch movement, loudness, duration and voice quality. These different cues may interact in complex ways, and they may have different effects on different listeners (Cole, Mo, & Baek, 2010; Cole, Mo, & Hasegawa-Johnson, 2010). Our paper investigates this multi-layered network of prominence cues at the level of the individual word within a sentence.

It is currently still unclear which linguistic variables have the strongest impact on the perception of prominence (but see Arnold, Wagner, & Baayen, 2013; Wagner, Tamburini, & Windmann, 2012; Wagner et al., 2015). For the sake of the present discussion, we distinguish between (1) continuous-

* Corresponding author at: IfL Phonetik, University of Cologne, Herbert-Lewin-Str. 6, 50931 Köln, Germany. Fax: +49 221 470 5938.
E-mail addresses: stefan.baumann@uni-koeln.de (S. Baumann), bodo@bodowinter.com (B. Winter).

https://doi.org/10.1016/j.wocn.2018.05.004
0095-4470/

valued prosodic parameters, (2) contrastive prosodic categories and (3) non-prosodic factors. Of course, we have to recognize that not all linguistic variables can neatly be categorized into one of these three groups.

By continuous-valued prosodic factors we understand those acoustic parameters that are signal-driven, such as intensity, fundamental frequency (F0) and duration. As contrastive prosodic factors, we classify those discrete and/or symbolic variables that relate to intonational phonology, such as the particular pitch accent types described in the German Tones and Break Indices system (GToBI; Grice, Baumann, & Benzmüller, 2005). As with other ToBI systems, GToBI characterizes pitch accents as discrete and abstract phonological elements that mediate between the actual phonetic elements they are composed of and their associated linguistic meanings (cf. Cangemi & Grice, 2016; Cole & Shattuck-Hufnagel, 2016). Finally, non-prosodic factors include semantic, syntactic and lexical variables. These relate to word choice or choice of syntactic structure (e.g., focus particles, part-of-speech differences, and word frequency).

Our goal in this study is to investigate the impact of these three classes of linguistic variables on prominence perception. We are furthermore assessing interrelations between the linguistic variables (i.e., which variable has the strongest influence on perceived prominence?) and potential differences in the perception strategies of listener groups (such as 'pitch listeners' versus 'spectral listeners'; Schneider & Wengenroth, 2009). Our study aims to contribute to the study of prominence both in terms of theory (descriptive and theoretical generalizations of prominence cues in German) and in terms of methodology (showing how multiple analytical techniques can be synthesized to get a more comprehensive picture of prominence perception).

2. Background

2.1. Prominence cues

The domain of investigation for most of our non-prosodic variables is the word. However, much of the past literature on prominence has focused on the level of the syllable within the word, in particular the large number of studies on lexical stress.[1] In fact, one of the aims of the present study is to examine whether what is known about syllable-level prominence also holds for word-level prominence within an utterance.

In 'stress-accent languages' (Beckman, 1986) such as English, syllables within a word are either strong (stressed) or weak (unstressed). Several correlates of stress in English and other Germanic languages have been identified. First, vowel quality and other segmental features in unstressed syllables are reduced compared to stressed syllables. The segments of stressed syllables and words generally tend to be hyperarticulated in order to enhance their perceptual clarity, at least in contrast to unstressed syllables (De Jong, 1995). Second, stressed syllables have more local pitch movement (Fry, 1958; Sluijter & van Heuven, 1995). Third, stressed syllables are longer in duration (Fry, 1955; Turk & Sawusch, 1996). Fourth, stressed syllables have overall higher intensity (Fry, 1955; Lea, 1977; Rietveld, 1984; Kochanski, Grabe, Coleman, & Rosner, 2005), which results in the perception of increased loudness. Fifth, stressed syllables have shifted spectral balance, with higher intensity in high-frequency components (Sluijter & van Heuven, 1996).

These studies on lexical stress show that several cues help the listener determine which syllable within a word is more prominent than another. In terms of perception, several of these cues have been shown to play a role, but to differing degrees. In particular, loudness and vowel quality have been shown to be relatively weak cues in earlier work on English (Fry, 1955, 1958, 1965), although Sluijter, van Heuven, and Pacilly (1997) find that in Dutch, it is not overall intensity (uniform across the frequency spectrum) but particularly intensity in high frequency components that matters for the perception of lexical stress. Beckman (1986) claims that duration and intensity do not act independently as correlates of prosodic prominence, both in production and perception. In a perception experiment on American English, she found that the most dominant cue for stress recognition was what she coined 'total amplitude', a factor that combines duration and intensity into a single acoustic category (also in line with the results of Kochanski et al., 2005).

With respect to phonological factors, we are particularly interested in pitch accent types as classified by the German Tones and Break Indices system (GToBI). This annotation scheme aims at describing 'Standard German' and, like other ToBI systems, has its roots in autosegmental-metrical phonology (see Beckman & Hirschberg, 1994; Beckman, Hirschberg, & Shattuck-Hufnagel, 2005; Ladd, 2008; online guidelines for American English ToBI: Veilleux, Shattuck-Hufnagel, & Brugos, 2006; for GToBI: Grice, Baumann, Ritter, & Röhr, 2017). Within the ToBI framework, a major distinction is made between pitch accents and boundary tones. These are classified according to two communicative functions: Whereas pitch accents, which are associated with stressed (metrically strong) syllables, serve to highlight relevant constituents, boundary tones, which are associated with phrase-final syllables, serve to delimit prosodic phrases. The tonal inventory of GToBI makes use of two basic levels, H(igh) and L(ow) tones – marked by a star (*) to show the association with a stressed syllable and by a minus (-) or percent sign (%) to indicate the association with a (minor or major) boundary. The actual tone values are relative to the pitch range that a speaker exploits.

Both the position and the type of pitch accent have been shown to influence prominence perception. With respect to accent position, it has long been claimed — both in the British tradition of intonation analysis (e.g., Halliday, 1967) and in the American generative tradition (e.g., Chomsky & Halle, 1968) — that the last stress or accent in an utterance is most prominent, i.e., the nuclear stress or accent. This structural or positional prominence is usually regarded as most important for an appropriate interpretation of the whole utterance, which is in turn the basis of many studies investigating the relation between accentuation and focus (e.g., Gussenhoven, 1984; Selkirk, 1984; Uhmann, 1991). Prenuclear accents have often been regarded as less important (see Büring's, 2007 notion of

[1] Strictly speaking, we have to talk of 'post-lexical stress', because we are dealing with *concrete* prominence at the utterance level. The term 'lexical stress' is often used when addressing abstract strength relations of syllables in words (see e.g., Ladd, 2008). Thus, the studies discussed here investigate post-lexical (acoustic) cues for the detection of lexical stress.