

DIFFERENT PARTS OF THE SAME ELEPHANT: A ROADMAP TO DISENTANGLE AND CONNECT DIFFERENT PERSPECTIVES ON PROSODIC PROMINENCE

Petra Wagner¹, Antonio Origlia², Cinzia Avesani³, George Christodoulides⁴, Francesco Cutugno², Mariapaola D’Imperio⁵, David Escudero Mancebo⁶, Barbara Gili Fivela⁷, Anne Lacheret⁸, Bogdan Ludusan⁹, Helena Moniz¹⁰, Ailbhe Ní Chasaide¹¹, Oliver Niebuhr¹², Lucie Rousier-Vercruyssen^{8,13}, Anne-Catherine Simon⁴, Juraj Šimko¹⁴, Fabio Tesser³, Martti Vainio¹⁴

¹Bielefeld University, ²University of Naples Federico II, ³CNR-ISTC, ⁴Université catholique de Louvain, ⁵Aix-Marseille University, ⁶Universidad de Valladolid, ⁷University of Lecce, ⁸Université Paris Ouest, ⁹École Normale Supérieure, ¹⁰FLUL/INESC-ID, ¹¹Trinity College Dublin, ¹²University of Southern Denmark, ¹³Université de Neuchâtel, ¹⁴University of Helsinki
petra.wagner@uni-bielefeld.de

ABSTRACT

Prosodic prominence is an umbrella term encompassing various related but conceptually and functionally different phenomena such as phonological stress, paralinguistic emphasis, lexical, syntactic, semantic or pragmatic salience, to mention a few. Due to the high interest prominence has received from various disciplines, it has been studied from multiple perspectives (functional, physical, cognitive). It also has been operationalised and annotated across different descriptive levels (syllable, word), based on different scales (categorical, multi-level, continuous), and measured across a large variety of signal domains (acoustic, articulatory, gestural). The present paper offers an overview of the various perspectives involved and defines a preliminary roadmap for a better and more unified understanding of this multi-faceted phenomenon.

Keywords: prominence, annotation, methodology, definition, universals

1. INTRODUCTION

Researchers working on language and speech are notorious for disagreeing on definitions even of fundamental concepts such as what constitutes a word, a syllable or a sentence. Such controversies often reveal differences in theoretical schools, but can also simply be an indication of radically different perspectives on one and the same *definiendum*. This problem is nicely illustrated by the well known Indian parable about a group of blind men trying to

describe what an elephant is, each of them examining different parts of the animal by touch. Their individual descriptions will necessarily be in disagreement. In prosodic prominence studies, different narrow perspectives have also led to diverse conceptualisations and methodological approaches and, hence, controversial and incompatible findings.

A continuing debate about what is the ideal perspective on prosodic prominence does not appear very fruitful — arguing on the basis of the elephant metaphor, it would simply be a debate about whether the trunk, tail or leg of an elephant provides the best description of the elephant itself. Rather, the main challenge lies in finding out how the various puzzle pieces (or animal parts) connect, interact or depend on each other. As a first step towards solving this puzzle, a first multinational workshop took place in Capri¹, in course of which phoneticians, phonologists and engineers engaged in discussions on perspectives and methodological approaches in the study of prosodic prominence. This paper describes their preliminary results and defines a roadmap paving the way for a unified account of the subject matter. In Section 2, we describe the main perspectives on the notion of prominence. In Section 3, we provide some methodological guidelines for how these perspectives can be identified and connected. Finally, Section 4 provides a roadmap on how to reach a more comprehensive future understanding of prosodic prominence.

2. PERSPECTIVES ON PROMINENCE

Naturally and often necessarily so, our research perspectives are constrained by our research fields and issues. Only by narrowing our focus of attention accordingly, we are able to limit our models' parameter space and degrees of freedom to such an extent that we find ourselves in a position where we can develop experimental designs, (formal) models and theories.

If we choose not to do so, our investigations are limited to anecdotal findings, *ad hoc* explanations or at best qualitative data analyses that forbid generalization. These limitations have led to different choices of model parameters in work on prosodic prominence, caused by the various perspectives and research interests. While it cannot be a practical solution to broaden our individual perspectives due to a risk of overgeneralization, a too narrow perspective runs the risk to be overly specific and having little predictive power. Speaking within the analogy of the elephant, even if we are mostly interested in the elephant's tail, we need to take into account the tail's connection to the elephant's body and the elephant's environment in order to fully understand its position, shape and movements. In the following, we will sketch the three main perspectives on prosodic prominence and their interdependencies.

2.1. The functional perspective

A common perspective focuses on communicative, often on core linguistic functions of prominence, e.g. its realisation being indicative of information structure, contextual givenness, phrasal stress, word order or lexical class [2, 3, 7, 28, 31]. Such a perspective constitutes a large body of research on prominence and lends itself to a categorical classification, with a particular type of functional prominence being either present or not. [13] mentions two ways of how to model this phonological layer, a unitary one, in which the various functional prominence types add up to a global prominence impression, which can be captured by the column height in a metrical grid, and a way in which the different functional levels of stress (e.g. levels of the metrical grid) are treated autonomously.

A functional perspective may also encompass aspects often termed *paralinguistic*. These are the effects of emotions or attitudes, which may affect the same signal parameters as linguistic prominence. That is, emphatic speech may affect segmental duration [21], emotion expression can be modeled with the same voice source parameters as prominence [10, 11, 4], sadness has been found to co-vary with

a reduced pitch range, a situation also found in post-nuclear linguistic prominences [3]. Paralinguistics may therefore significantly contribute to the pertaining impressions of both prominence placement and strength and may be confounded with the linguistic core functions of prosodic prominence.

2.2. The physical perspective

Among the established physical correlates of prominence are fundamental frequency (excursion, shape), duration, voice source features including spectral tilt, open quotient, excitation strength/loudness, hyper-articulation, intensity and multimodal cues (eyebrow movements, head movements, manual co-speech gestures) [22, 27, 15, 20, 4]. A purely signal-based analysis will treat prominence as a continuous rather than a categorical phenomenon, similar to a psycho-acoustic scale, and may stand in opposition to a categorical-functional point of view. While most contemporary research takes into account some aspects of the signal correlates of prominence, the majority of the studies relies on individual rather than a large set of possible signal correlates, focusing on the interplay and interchangeability of correlates in particular contexts. A signal perspective treating prominence as a psycho-acoustic rather than a communicative event, might overlook function-related specifications of certain signal correlates. The interaction of the various cues signalling prominence in its different functions is hitherto not well understood, especially, as they are shaped by context effects [6, 14].

2.3. The cognitive perspective

A cognitive perspective on prominence usually focuses on perceptual processing, i.e. it studies the low-level neural pathways and psycho-acoustic processing mechanisms that contribute to higher-level cognitive processing [19, 18]. Such high-level processes are known to be strongly shaped by linguistic knowledge including linguistic and paralinguistic functions as well as situation-specific expectations [5, 30, 17]. So far, the interactions of these low- and high-level processes are not well understood.

Obviously, research carried out from a cognitive perspective heavily relies on the functional perspective (for defining prominence relevant functions and units shaping cognitive processes involving prominence such as attention, grouping and memory) and the physical perspective (for connecting cognitive processes with the relevant signal correlates).

3. CONNECTING PERSPECTIVES: THE ELEPHANT'S BODY

It is immediately obvious that each perspective taken to prominence will only lead to a narrow picture of its nature. A physical signal perspective not taking into account function related categorical judgements or function-related contextual embeddings will fail to model what makes prominence “prosodic”. This is enhanced by findings showing that prominence perception is driven largely by top-down expectations and at least partly independent of signal correlates, e.g. prominence may be perceived because “it belongs there”. Likewise, a purely functional perspective may miss out on signal related aspects if it relies on simple 1:1 function-signal mappings or on overly simplistic signal correlates. It may also fail to acknowledge the additional impact of processing constraints influencing prominence perception such as rhythmic expectations or attentional processes which may even lend themselves to be exploited by a language’s phonological system. Likewise, the cognitive perspective relies on linguistic categories as these either constrain its models or explain its data-driven results. Given that we know very little about the interactions between the various perspectives, we suggest the following four strategies to a) identify the individual research perspectives and b) better understand the complex relationship between functions, processes and signals: clarifying definitions (3.1), typological investigations (3.2), comparing annotations (3.3), and building technical models (3.4).

3.1. Prominence definitions

Given the various perspectives on prominence, it comes as no surprise that hitherto, no consensus on a definition of the term “prominence” has been reached. This is probably why the very generic definition of prominence provided by [25] appears to be the most widely accepted, containing no prior assumptions about signal correlates, function or cognitive processes²:

<We> say that a <linguistic entity> is prosodically prominent when it <stands out> from <its environment> (by virtue of <its prosodic characteristics>).

Terken’s generic definition is applicable to a large body of work due to its underspecification. When reviewing other, possibly narrower definitions – or rather usages – of the term prominence, it becomes

evident that they tend to be somewhat overloaded in both function and form: the term prominence is often being used synonymously with the terms emphasis, lexical stress, nuclear accent, prosodic focus, pitch accent, intensity peak, lengthening, to mention a few. Given the wide range of functions, forms, and research perspectives, neither an overly generic definition nor a too narrow one (ruling out potential alternative usages) appears to be helpful. A currently more fruitful approach seems to be *a set* of definitions which clarify the way the term prominence is used individually. For this purpose, Terken’s definition may serve as a template for a further specification, i.e. a clarification of what is meant by “we”, “linguistic entity”, “prosody”, “standing out” and (if not contained in these two) what is considered as the relevant environment in the given research. Notice that the link to the physical correlates (by virtue of . . .) is not considered obligatory. An example for a narrower definition building on this template would be

In this study, we say that a word is prosodically prominent when it is perceived as the focus exponent within a sentence by virtue of a pitch accent.

This need to explicate will assist the reader and, perhaps more importantly, the author of a study to discern what parts and aspects of the elephant’s body are under investigation, and from what perspective.

3.2. Language-specific and universal correlates

To this day, only very few typological studies on prominence expression or perception have been carried out [1, 23]. This is surprising given that we know very little about the language-specificity or universality of prominence and its correlates. A comparison of how prominence can be described at signal level across typologically diverse languages is likely to reveal more on the universality or language-specificity of certain cues, their relative impact and their interaction with other linguistic features, e.g. tonality, information structure marking, or syntactic structure. Furthermore, the comparison of listeners with diverse native languages may provide a useful diagnostic for language-specific expectations with regards to prominence.

3.3. Prominence annotations

A useful diagnostic lies in the identification of how prominence is treated in various annotation schemes: The usage of continuous or quasi continuous multi-level scales seems to show a more

signal-driven perspective treating prominence like a psycho-acoustic phenomenon that somehow interacts with linguistic categories [9], while a more autonomous functional perspective has led to categorical approaches [5]. Other ideas fall in between the two suggestions, probably due to a unitary functional approach, treating the various types of linguistic prominence in a cumulative, metrical grid approach [16] that can be expressed by a limited number of functionally relevant prominence levels. A systematic comparison of our annotation schemes would help us finding out whether the annotations describe *qualitatively different parts of the same elephant*, or whether they reveal *more or less fine-grained impressions of its same part*. It furthermore appears crucial to take the various independent contributions to prominence perception and processing seriously, namely the physical signal, the linguistic and contextually shaped expectations, the ways in which our cognitive system constrains, connects and weighs these factors both in a top-down and bottom-up fashion. This enables us to understand whether our various models account for these interactions.

3.4. Building technical models

Technological applications are by necessity fully explicit regarding their mechanisms, inputs and limitations. Therefore, the agreement between automatic and human prominence annotations helps clarify the cognitive mechanism of assessing prominence markings by human annotators, and also the relevant physical and functional aspects underlying the usage of prominence in discourse, e.g. by suggesting the importance of hierarchical aspects of prosody [29] or by providing a tool for analyzing and improving annotation strategies [8]. Likewise, technological approaches to prominence detection and exploitation may significantly contribute to an improved understanding of the subject matter: both rule-based and machine learning approaches for prominence detection have been showed to play a significant role in the understanding of signal weightings, function-signal interactions and the question of universality, e.g. [12, 24, 23]. More specifically, past work on automatic prominence detection using machine learning has highlighted that sequences of non-prominent syllables may exhibit self-consistent acoustic patterns that help predicting the occurrence of a prominent syllable [6]. Recent results [24] furthermore indicate that the relationship between acoustic measures in sequences of non-prominent syllables and manual labels is non-linear in nature. Generally, technological applications such as TTS provide challenging fields in which prominence

models can be tested [26].

4. CONCLUSION: A ROADMAP FOR FUTURE WORK

We have shown that despite prominence being a popular field of interdisciplinary research, our results often fail to contribute to a comprehensive understanding of the concept. Due to vastly different perspectives, conceptualisations and operationalisations our various insights often forbid comparison and integration. We believe that these problems can be fixed if more care is taken to specify which “elephant part” we are working on and how it connects to the others. A full set of guidelines how this can be achieved is the object of long-term discussions within our interdisciplinary and multilingual initiative. Our key results so far are summarized in the following testimony that can simultaneously be read as a preliminary set of methodological recommendations.

1. Establish from which perspective prominence is treated in your research.
2. Have a clear distinction between the various prominence correlates (bottom-up signal, top-down context/linguistics) and processing constraints influencing prominence perception (e.g. attention focus, psycho-acoustic constraints).
3. Precisely define your usage of the term prominence, taking [25] as a generic template.
4. Establish whether your research investigates prominence from a language-specific or universal perspective.
5. Clearly describe the relationship between your prominence concept and its extensional domain (e.g. acoustic, multimodal, perceptual...).
6. Say how your obtained findings relate to other perspectives on prominence. More specifically:
 - (a) Establish how your annotations and annotation schemes can be related to annotations obtained from other perspectives/on other languages.
 - (b) Establish how your examined domains of prominence expression (acoustics, articulatory, multimodal, linguistic expectations...) relate to other domains.

5. REFERENCES

- [1] Andreeva, B., Barry, W., Wolska, M. 2012. Language differences in the perceptual weight of prominence-lending properties. *Proceedings of Interspeech 2012* Portland, USA. 1–4.
- [2] Baumann, S., Roth, A. 2014. Prominence and

- coreference – on the perceptual relevance of f0 movement, duration and intensity. *Proceedings of Speech Prosody 2014*.
- [3] Bocci, G., Avesani, C. 2011. Phrasal prominences do not need pitch movements: Postfocal phrasal heads in Italian. *Proceedings of Interspeech 2011* Florence, Italy. 1357–1360.
- [4] Chasaide, A. N., Yanushevskaya, I., Cane, J., Gobl, C. 2013. The voice prominence hypothesis. *Proceedings of INTERSPEECH 2013* Lyon, France. 3527–3531.
- [5] Cole, J., Mo, Y., Hasegawa-Johnson, M. 2010. Signal-based and expectation based factors in the perception of prosodic prominence. *Journal of Laboratory Phonology* volume 1 425–452.
- [6] Cutugno, F., Leone, E., Ludusan, B., Origlia, A. 2012. Investigating syllabic prominence with conditional random fields and latent-dynamic conditional random fields. *Proceedings of Interspeech 2012* Florence, Italy. 2402–2405.
- [7] D’Imperio, M. 1998. Acoustic-perceptual correlates of sentence prominence in Italian questions and statements. *Journal of The Acoustical Society of America* 104.
- [8] Escudero, D., Aguilar, L., del Mar Vanrell, M., Prieto, P. 2012. Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labeling system. *Speech Communication* 54(4), 566 – 582.
- [9] Fant, G., Kruckenberg, A. 1989. Preliminaries to the study of Swedish prose reading and reading style. *STL-QPSR* 30(2), 1–80.
- [10] Fant, G., Kruckenberg, A., Liljencrants, J., Hertzgård, S. 2000. Acoustic-phonetic studies of prominence in Swedish. *Speech Transmission Laboratory—Quarterly Progress and Status Report* (2–3), 1–51.
- [11] Gobl, C., Chasaide, A. N. 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication* 40(1-2), 189–212.
- [12] Goldman, J. P., Avanzi, M., Simon, A. C., Lacheret, A., Auchlin, A. 2007. A methodology for the automatic detection of perceived prominent syllables in spoken French. *Proceedings of INTERSPEECH 2007* Antwerp, Belgium. 98–101.
- [13] Gussenhoven, C. 2011. Sentential prominence in English. In: Ostendorp, M., Ewen, C., Rice, K., (eds), *The Blackwell Companion to Phonology*. Wiley-Blackwell, Malden M.A. & Oxford 2780–2806.
- [14] Gussenhoven, C., Rietveld, T. 1998. On the speaker-dependence of the perceived prominence in f0 peaks. *Journal of Phonetics* 26, 371–380.
- [15] de Jong, K. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97, 491–504.
- [16] Kohler, K. J. 1997. Parametric control of prosodic variables by symbolic input in TTS synthesis. In: van Santen, J., Sproat, R., Olive, J., Hirschberg, J., (eds), *Progress in Speech Synthesis*. Springer, New York 459–475.
- [17] Lacheret, A., Simon, A.-C., Goldman, J.-P., Avanzi, M. 2013. Prominence perception and accent detection in French: from phonetic processing to grammatical analysis. *Language Sciences* 39, 95–106.
- [18] Ludusan, B., Origlia, A., Cutugno, F. 2011. On the use of the rhythmogram for automatic syllabic prominence detection. *Proceedings of Interspeech 2011* Florence. 2413–2416.
- [19] McAngus Todd, N. P., Brown, G. J. 1996. Visualization of rhythm and meter. *Artificial Intelligence Review* 10, 253–73.
- [20] Moubayed, S. A., Beskow, J., Granström, B. 2010. Auditory visual prominence. *Journal of Multimodal User Interfaces* 3, 299–309.
- [21] Niebuhr, O. 2010. On the phonetics of intensifying emphasis in German. *Phonetica* 66, 95–112.
- [22] Rietveld, T., Gussenhoven, C. 1983. On the relation between pitch excursion size and prominence. *Journal of Phonetics* 13, 299.
- [23] Rosenberg, A., Cooper, E., Levitan, R., Hirschberg, J. 2012. Cross-language prominence detection. *Speech Prosody*.
- [24] Tamburini, F., Bertini, C., Bertinetto, P. M. 2014. Prosodic prominence detection in Italian continuous speech using probabilistic graphical models. *Proceedings of Speech Prosody 2014* 285–289.
- [25] Terken, J. 1991. Fundamental frequency and perceived prominence. *Journal of the Acoustical Society of America* 89(4), 1768–1776.
- [26] Tesser, F., Somnavilla, G., Paci, G., Cosí, P. 2013. Experiments with signal-driven symbolic prosody for statistical parametric speech synthesis. *Proceedings of the 8th Speech Synthesis Workshop (SSW 8)* Barcelona, Spain. 183–187.
- [27] Turk, A. E., Sawusch, J. R. 1996. The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America* 99, 3782–3790.
- [28] Vainio, M., Järvikivi, J. 2006. Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics* 34(3), 319–342.
- [29] Vainio, M., Suni, A., Aalto, D., others, 2013. Continuous wavelet transform for analysis of speech prosody. *TRASP 2013-Tools and Resources for the Analysis of Speech Prosody, An Interspeech 2013 satellite event, August 30, 2013, Laboratoire Parole et Langage, Aix-en-Provence, France, Proceedings*.
- [30] Wagner, P. 2005. Great Expectations - Introspective vs. Perceptual Prominence Ratings and their Acoustic Correlates. *Proceedings of Interspeech 2005* Lisbon, Portugal. 2381–2384.
- [31] Watson, D. G., Arnold, J., Tanenhaus, M. 2008. Tic tac toe: Effects of predictability and importance on acoustic prominence in language production. *Cognition* 106, 1548–1557.

¹ <http://www.smile.unina.it/capriworkshop/index.html>

² Brackets and font modifications added