Journal of Phonetics 48 (2015) 1–12

Contents lists available at ScienceDirect

Journal of Phonetics

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

In defense of stylistic diversity in speech research

Petra Wagner [a,*], Jürgen Trouvain [b], Frank Zimmerer [b,c]

[a] Faculty of Linguistics and Literary Studies, Bielefeld University, Postbox 100131, D-33501 Bielefeld, Germany
[b] Department of Computational Linguistics and Phonetics, Saarland University, Postbox 151150, D-66041 Saarbrücken, Germany
[c] Institute of Phonetics, Goethe University, Postbox 170, 60054 Frankfurt, Germany

ARTICLE INFO

Available online 16 December 2014

Keywords:
Natural speech
Laboratory speech
Speaking styles
Methodology
Prominence
Regressive assimilation
Laughter

ABSTRACT

Recently, the debate about what kind of speech data is most appropriate for linguistic research has intensified. Generally, with 'laboratory speech' defenders on the one hand and 'natural speech' proponents on the other, two seemingly clearly distinct phonetic data types have been identified. In this article, this dichotomy is called into question. Results from previous studies on segmental phonetics, prosody and paralinguistics indicate that the data we are using may indeed have had an immense influence on our results. The research papers in the present Special Issue in Journal of Phonetics provide further evidence for the style-dependency of speech data and hence, on our theories and models. Importantly, they also show that some results remain stable independently of the speaking style under investigation. We claim that these findings do not point to an inherent superiority of one particular type of data used in phonetics research. Instead, we argue for a stronger methodological awareness in investigations of speech phenomena and more cautious interpretations of the findings that we make. We also believe that we need a much better understanding of the extent to which our methods and our ways of collecting speech data influence our results. A generally increased methodological awareness and a higher variety of investigated styles of speech will promote our research progress further than a continuing argument for or against using one particular type of speech data.

1. Introduction

Recent years have seen a debate between defenders of strictly controlled, experimental data collections based on 'lab speech' (e.g., Xu, 2010) and promoters of more 'naturalistic', 'authentic' or 'ecologically valid' data sets (e.g., Ernestus, 2000; Kohler, Pätzold, & Simpson, 1995; Oertel, Cummins, Edlund, Wagner, & Campbell, 2013). The main reason behind this discussion is our knowledge about style-specific phonetic variation, and different opinions about what data provides a better window to 'the truth'. Although 'speaking style' is a heterogenously used and multiple-dimensional term (cf. Barry, 1995; Blaauw, 1995; Eskénazi, 1993) we can consider it as accepted that the same sentence leads to a different phonetic output depending on whether it has been uttered in the style of a news broadcaster or in a conversation with a friend. On the perception side a news broadcaster speaking in a conversational style would be as unacceptable for listeners as a friend talking to you in a news broadcasting style. This Special Issue aims at providing further arguments for the impact of speaking style diversity on phonetic investigations. Rather than favouring a decision for either style in phonetic data collections, we aim to raise a stronger awareness for the data and methods we use in phonetic research, and for what we can or cannot conclude from the results based on the data that was used.

We agree with a lot of points raised in Xu's article 'In defense of lab speech' in this journal (Xu, 2010). Still, we would like to extend this discussion and add some further points. Researchers focusing on different aspects of phonetic sciences participated in our effort to bring together a broad basis of results indicating how crucial both methods and data are for our insights. Unlike earlier debates in this area, we do not want to argue for or against the superiority of one or another type of phonetic data. Instead, unifying the outcome of the different articles, one of the central take-home messages is the need to investigate *various* styles and their potential effects on the results. This Special Issue shows that style specificity of results is not limited to single areas or few aspects of phonetics

* Corresponding author. Tel.: +49 521 106 3510; fax: +49 521 106 2996.
E-mail addresses: petra.wagner@uni-bielefeld.de (P. Wagner), trouvain@coli.uni-saarland.de (J. Trouvain), zimmerer@coli.uni-saarland.de (F. Zimmerer).

0095-4470/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2014.11.001

(and phonology), a fact clearly reflected in the topics covered in the collected papers ranging from investigations on segments and suprasegmentals to work on paralinguistics.

Another aim of this Special Issue is to raise the sensitivity for the *conditions* under which the speech was recorded and studied. Usually, these manifold conditions are collected under the umbrella of 'style', a term for which exists no standard definition in phonetic research (but also not in linguistics where 'style' concurs with 'register', 'genre' and 'text type', see e.g. Biber & Conrad, 2009). Here, 'style' is considered as a broad cover term for speaking situations leading to phonetic variation where the *level of control* is the main descriptive factor, and is mirrored in labels such as read, scripted, careful, clear, formal versus spontaneous, unscripted, casual, conversational. These contrasts represent extremes on a continuum and elements of both extremes can be found across many situations. The term 'lab speech' is also a loosely used term that can have several meanings. In a narrow sense 'lab speech' is speech sampled under scrutinized experimental control, in a wide sense, all speech samples recorded in a lab (or similar acoustic conditions) can be considered as 'lab speech'. The latter would include the vast majority of speech data studied in the phonetic sciences, i.e. also speech elicited under conditions considered as more ecologically valid or 'natural' – with 'natural' being another highly problematic term that should be avoided (see Section 3).

Instead of using 'lab speech' and 'natural speech' as attributes of a given speaking style, the *level of control* should be described in more detail. This entails information on the recording context, which can be the phonological context one controls for but also the communicative situation. The latter reading of 'context' includes the communicative intent and the purpose of the speech production, which may have a strong impact on the way it is articulated, e.g. due to adaptation to the experimenter's expectation or the degree of acting (see Section 3). The diversity of speaking styles, its conditions and the level of control is also mirrored in the articles for this special issue where on the highly controlled end, speech was elicited by having subjects read sentences or single words or by shadowing speech. On the end of lesser control were narrations and picture-based conversations with or without any pre-selected topic. The diversity of styles we are looking for can also be observed by the fact that not a single style – or rather level of control – that is studied in the present Special Issue, appears twice.

Dellwo, Leemann, and Kolly (2015) investigate to what extent Swiss German listeners are able to detect whether excerpts were taken from read or spontaneous speech in Swiss German. The data for the experiments is drawn from the TEVOID corpus (Leemann, Dellwo, & Kolly, 2014), where speakers produced spontaneous speech in interviews, and later read their spontaneously produced interview sentences. Despite the lack of Swiss German as a read language in everyday life, i.e. read aloud speech is produced in standard rather than Swiss German, Swiss listeners are well able to distinguish between read and spontaneous speech. The article is able to replicate findings from English and (standard) German (e.g., Batliner, Kompe, Kießling, Nöth, & Niemann, 1995; Laan, 1997). Furthermore, the authors can show that there is variation concerning the extent to which read speech sounds more or less read. Dellwo and colleagues do not find clear evidence for a definite set of acoustic cues (apart from articulation rate) that may have contributed to listeners' ability to discriminate the two speaking styles. This lends further support to the importance of an increased awareness for style or control-driven effects.

This also becomes evident in de Ruiter's article (de Ruiter, 2015) comparing suprasegmental variation in read and spontaneously produced speech. She analyses how contextually given words (see her article for a definition of 'givenness') are realized with different intonation patterns depending on the level of control: In one condition participants tell a story based on pictures, while in the other condition, the pictures are additionally accompanied by a textual representation which has to be read aloud. Her results show that what is commonly *assumed* to be the standard way to indicate givenness in prosody is *not* what speakers actually do, unless they are reading. Furthermore, her results indicate that what usually is thought to be the standard way to indicate givenness prosodically, is in reality a particular form of learned, style-specific behavior. Thus, assumptions about underlying prosodic categories and patterns might need to be put to more thorough and systematic tests using different speaking styles.

DiCanio and colleagues compare cross-stylistic production data of the endangered language Yoloxóchitl Mixtec (DiCanio, Nam, Amith, Castillo García & Whalen, 2015). They investigate whether style and other factors have an impact on the speakers' vowel space dispersion (i.e., the produced vowel formants). Their results indicate that the vowel space is smaller in spontaneous than in elicited speech. Besides, spontaneous speech shows more variation. However, the vocalic differences remain systematic relative to the overall vowel space. The authors argue that vowel space can be reliably estimated based on less controlled (i.e., spontaneous) speech. Their work is of special interest to field research on endangered languages or other investigations, where highly controlled data may be harder to collect than spontaneous speech such as narrations or interviews.

The article by Ernestus, Hanique, and Verboom (2015) provides further evidence for the necessity to be careful with generalizations based on a single style. A comparison of Dutch corpus data differing in styles, ranging from spontaneous face-to-face conversations to lectures and read texts reveals notable differences concerning the amount of phonetic reductions. Studying the patterns of ten frequent words, their results suggest that the more formal the speech situation, the less massive are the phonetic reductions. The formality of a speech situation can be regarded as being directly linked to the level of control. Secondly, the authors investigate to what extent segment reduction patterns can be elicited with the help of a shadowing task, i.e. highly controlled speech. Their results show that the patterns observable in spontaneous speech differ from the ones produced via shadowing. In fact, participants rarely mimicked the reductions present in the speech to be shadowed. Thus, the methodological choice for data collections appears to be crucial when studying reduction patterns.

The paper by Sanchez and colleagues provides further evidence for the fact that phoneticians need to be very careful when creating experimental settings (Sanchez, Hay, & Nilson, 2015). In an extension of previous work where it has been shown that external distractors such as visible stuffed toys triggering a concept connected to a certain language variety have small but significant effects on the perception of vowels, the authors provide evidence that conceptual activation may also lead to changes in production.