Journal of Phonetics 48 (2015) 13–28

Contents lists available at ScienceDirect

Journal of Phonetics

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

The recognition of read and spontaneous speech in local vernacular: The case of Zurich German

Volker Dellwo [a,*], Adrian Leemann [b], Marie-José Kolly [a]

[a] University of Zurich, Phonetics Laboratory, Department of Comparative Linguistics, Plattenstrasse 54, 8032 Zurich, Switzerland
[b] Department of Theoretical and Applied Linguistics, University of Cambridge, UK

ARTICLE INFO

Article history:
Received 27 November 2013
Received in revised form
8 October 2014
Accepted 29 October 2014
Available online 17 December 2014

Keywords:
Speech perception
Speaking style
Read and spontaneous speech
Speaker-specific characteristics
Listener-specific characteristics

ABSTRACT

Listeners are typically able to identify speech as either being produced spontaneously or read from a transcript. In the present research we investigated whether this is true in vernacular speech when typical cues to read and spontaneous speech are either missing and/or ambiguous. In addition it was investigated what the acoustic cues for listeners' identification ability were. 26 listeners of Zurich German participated in two perception experiments. In Experiment I, listeners judged 128 stimuli (64 spontaneous, 64 read) in a two-alternative identification task as either spontaneous or read. Results revealed that overall listener performance was well above chance (mean $A' = 0.82$) while there was no bias for either read or spontaneous speech (mean $B''_D = 0$). There were significant effects of speaker and listener in A' and B''_D. In Experiment II, the same 26 listeners rated the same 64 read speech stimuli from Experiment I as to whether they sounded more or less read. Results revealed that there was considerable within-category variability as a function of speaker. From eight acoustic prosodic parameters only articulation rate explained listener behavior to some degree in Experiments I and II. Overall the study suggested that read and spontaneous speech can be recognized based on very subtle cues to these speaking styles.

1. Introduction

The present study was directly motivated from our previous research on speaker-individual speech rhythmic characteristics in which we investigated the variability of speech rhythm within and between speakers based on the *Temporal Voice Idiosyncrasy Corpus* (henceforth: TEVOID Corpus; cf. Dellwo, Leemann, & Kolly 2012; Leemann, Kolly, & Dellwo 2014). . In this speech corpus we created acoustic within-speaker variability by asking speakers to produce spontaneous and read speech in Zurich German, two production modes of speech that are generally referred to as *speaking styles* in the literature (Llisterri, 1992; Laan, 1997). Both tasks were performed in our laboratory. We first recorded speakers in an interview situation. From these recordings we extracted a number of sentences that were free from grammatical errors, semantically coherent and fluently produced, and created a written transcript for these sentences. In a second recording session (a few months later), each speaker read all sentences from the transcript (his/her own sentences as well as those previously produced by his/her peers). We carried out a large variety of durational measurements that are typically referred to as rhythm measures (Ramus, Nespor, & Mehler, 1999; Grabe & Low, 2002). Results from measurements of these variables consistently showed strong between-speaker but little or inconsistent within-speaker effects between spontaneous and read speech (Dellwo et al., 2012; Leemann et al., 2014). We thus argued that the rhythmic durational characteristics we observed do not vary between read and spontaneous speech. There is, however, an alternative interpretation possible: the speech produced under the strict laboratory settings might be poor examples of the categories *read* and *spontaneous* and might thus contain poor cues to these speaking styles. In the present research we followed up this interpretation and hypothesized that if this was the case then listeners should show a poor performance in identifying the speech from the TEVOID corpus as either read or spontaneous.

What are the reasons to assume that the speech elicited under the conditions described above might not contain sufficient cues to speaking style? It has often been argued that read and spontaneous speech are not members of two distinct categories but that there

* Corresponding author. Tel.: +41 44 634 29 95.
E-mail addresses: volker.dellwo@uzh.ch (V. Dellwo), al764@cam.ac.uk (A. Leemann), marie-jose.kolly@uzh.ch (M.-J. Kolly).

0095-4470/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2014.10.011

is a continuum reaching from canonical spontaneous speech, e.g. a casual conversation in the street, to canonical read speech, e.g. the speech of a news reader (see review in Laan, 1997). In terms of production, numerous mechanisms may lead to strong variability within the read and spontaneous categories. It might be, for example, that speakers take on different styles in their reading (news style, theatrical style, etc.) or read speech might also vary depending on who is addressed with it (Niebuhr et al., 2010). The same seems plausible for spontaneous speech. In addition, the labels “read” and “spontaneous” are not well defined categories but are often subcategorized. Labov (1972), for example, distinguishes, amongst others, between reading style, careful speech, casual speech and spontaneous speech. Smiljanic and Bradlow (2009) further distinguish between casual and clear speech, which, in laboratory settings, are typically both variants of read speech. Clear speech is intended to be particularly intelligible to listeners with listening impairments or other speech comprehension problems. Even though the labels “spontaneous” and “read” might seem applicable for the speech in our TEVOID corpus – we know it for a fact that read speech was read and spontaneous speech was produced spontaneously – it seems that our speaking styles are not very far apart on a possible read-spontaneous continuum. We argue that this is mainly the result of (a) a lack of acoustic cues peculiar to read/spontaneous speech due to our data-collection method and (b) ambiguous acoustic cues to read and spontaneous speech due to the use of vernacular speech. In the following we give support for both reasons:

Re (a): Our spontaneous speech was recorded in a rather formal recording situation (recording booth, semi-structured interview) and our sentence selection criteria of grammatical correctness, and semantic coherence and fluency (see above) may have removed important cues to this style such as false starts, rhythmic irregularity, phrase-final voicing/devoicing or timing phenomena (see discussion in Laan, 1997: 44). Such a drastic reduction of cues to spontaneity is typically not present in previous studies. In Mixdorff and Pfitzinger (2005) listeners identified speech as either read or spontaneous. Like in our method Mixdorff and Pfitzinger asked listeners to read scripted utterances they previously produced spontaneously. However, their spontaneous speech contained characteristics such as filled and unfilled pauses and they reported that in particular filled pauses such as [E:m] facilitated its identification (p. 317). This means that the most important cues to spontaneous speech in Mixdorff and Pfitzinger's study were not present in our data. The same is true for numerous other cues reported in the literature: Cucchiarini, Strik, and Bou (2002) used mainly different characteristics of silent pauses and hesitation markers in quantifying fluency cues in read and spontaneous speech. Liu and Zeng (2006) found that the insertion of silent gaps into conversational speech increases its intelligibility and thus makes it more appear like clear speech. Hirschberg (2000) found that about 20% of English spontaneous sentences were characterized by disfluencies which might be used for their identification. Pettinato and Hazan (2013) and Hazan and Pettinato (2014) studied the use of pausing behavior in speakers' attempts to make their speech sound more clear and found that it was used both by adults and children (even though to different degrees). Finally, Levin, Schaffer, and Snow (1982) found that listeners could identify read and spontaneous speech with an accuracy of 84% mainly based on pause characteristics, false starts, wording and grammar. Given the lack of such prominent indicators in our non-disfluent spontaneous speech it is unclear whether listeners are still able to distinguish it from read speech.

Re (b): Another peculiarity in our data lies in the ambiguity of cues to spontaneity/reading resulting from the use of vernacular speech. We asked Swiss German speakers from Zurich to read in their local vernacular of Zurich German[1]. In Swiss German two forms of the same language – a high variety and a low variety – co-exist. This situation has been described as one of “medial diglossia” (Kolde, 1981): Standard German is mainly used in written form while Swiss German is used for verbal communication. Swiss German, however, is viewed as more prestigious than Standard German (Sieber & Sitta, 1986) even though it does not feature a formally-defined spelling system. Writing in Swiss German is largely restricted to recent technological inventions, including e-mailing, mobile text messaging, social media, and chat-rooms (Christen, 1999; Siebenhaar, 2003). Younger speakers are therefore used to some degree to reading texts in vernacular but typically in situations that are very related to settings in which spontaneous speech would occur. Swiss Germans typically acquire the vernacular variety first in life – and this was the case with all our speakers – the standard variety of German is later acquired at school. To some degree speakers become acquainted to it prior to school through the media. So Standard German has strong similarities with an L2 for Swiss speakers (Häcki Buhofer & Burger, 1998). Given this situation, Swiss German speakers might be particularly prone to read texts written in Swiss German in a spontaneous way. In addition, and possibly more importantly, when Swiss German listeners come across vernacular speech they might be strongly biased towards identifying this speech as spontaneous. This might be additionally evoked by lexical markers to spontaneous speech such as the adverb *dänn* (closest English translation: 'then') in the sentence *Ich bin wäge Spraachwüsseschaft dänn usegheit* (English: 'I then failed because of linguistics') which is not typically used in formal writing, or the modal particle *ja* (closest English translation: 'as you know') in *Chasch ja nöd nöime andersch go studiere mit Erasmus* ('You cannot go studying somewhere else with Erasmus'). It might also have been evoked by syntactical constructions that are typically not found in written language. Further, we might have introduced ambiguity in style cues through our reading instructions. Cucchiarini et al. (2002) reported strong differences between the fluency of speakers reading speech without prior rehearsal. In our case all speakers practiced reading the sentences prior to the recordings to be able to read them fluently. This means that in particular temporal variability present in

[1] Even though in Swiss German the vernacular varieties are generally referred to as 'dialects' they reveal more strongly the characteristics of a vernacular as they particularly lack the typical continuous variability between dialectal variety and standard.