Journal of Phonetics 68 (2018) 69–84

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

The role of segments and prosody in the identification of a speaker's dialect

Adrian Leemann [a,*], Marie-José Kolly [b], Francis Nolan [c], Yang Li [c]

[a] Department of Linguistics and English Language, County South, Lancaster University, LA1 4YL, United Kingdom
[b] Neue Zürcher Zeitung, Falkenstrasse 11, 8021 Zürich, Switzerland
[c] Phonetics Laboratory, Department of Theoretical and Applied Linguistics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, United Kingdom

ARTICLE INFO

Article history:
Received 4 August 2016
Received in revised form 7 February 2018
Accepted 12 February 2018

Keywords:
Dialect identification
Segments
Rhythm
Intonation
Prosody morphing
Swiss German

ABSTRACT

The objective of this study is to investigate the role of segments, rhythm, and rhythm combined with intonation in the identification of a speaker's dialect. In a between-subjects design using three conditions, we tested 62 listeners (Zurich Swiss German) in a two-alternative-forced choice dialect identification experiment: in condition one, 21 listeners were asked to identify two dialects (Valais and Bern Swiss German) in unmorphed form. In condition two, 20 different listeners had to identify the same two dialects but with swapped speech rhythm, and in condition three, 21 different listeners had to identify the same dialects with swapped speech rhythm and intonation. The experiment showed that exchanging speech rhythm alone or speech rhythm combined with intonation had very little effect on the listeners' dialect identification performance: listeners appear to use primarily segmental information in the identification process. Further results revealed that (a) superimposing the prosodic structure of one dialect (Bern Swiss German) onto another (Valais Swiss German) caused greater variability across some listeners than the other way around and that (b) identification performance varies as a function of sentence material used, i.e. how the sentences differ in segmental and prosodic make-up. We discuss implications for forensic phonetics, language and cognition, and automatic speech recognition.

1. Introduction

Dialect entails regional variation on various linguistic levels including phonetics, phonology, morphology, syntax, and the lexicon (Petyt, 1980). It is an indexical property that is encoded in everyday language situations. Upon making a new acquaintance in German-speaking Switzerland, for instance, it is not unusual that the first topic of small talk involves dialects: 'Judging by your dialect, you come from Bern – right?'. The speech stream consists of segmental and prosodic features, both of which carry diagnostic information for dialect identification. If, for example, Hans pronounces the word 'milk' as [mˈɪʊɣ̞] instead of [mˈɪlɣ̞], he will likely be placed in Western Switzerland, given that the vocalization of /l/ is a typical feature of this area. Swiss German (SwG) listeners perform well at dialect identification: Leemann and Siebenhaar (2008) as well as Guntern (2011) have shown that naïve listeners can accurately identify a speaker's dialect much above chance. Identification performance varies by language: Williams, Garrett, and Coupland (1999) report successful dialect identification of varieties of English in Wales. For dialects of Dutch and dialects of British English, Van Bezooijen and Gooskens (1999) found identification rates much above chance level – the same is valid for dialects of American English, as shown by Clopper and Pisoni (2004). In recent research, Bent, Atagi, Akbik, and Bonifield (2016) conducted a large-scale study using regional dialects and non-native accents of English to examine how they are perceptually organized on the part of the listeners. In a free classification task, listeners exhibited sensitivity to distinguishing between the 24 native and non-native accents, providing insight into listeners' representations of varieties of English. Overall, these findings reveal that naïve listeners are aware of dialectal variation and can perform an identification task with different degrees of accuracy, depending on the language.

Recent research has revealed that naïve listeners can successfully identify dialects using prosodic features alone. Leemann and Siebenhaar (2008) report that naïve SwG listeners are able to identify dialects in delexicalized speech (low-pass filtered <250 Hz) above chance level. Dialect identification based on prosodic features alone has also been documented

* Corresponding author.
E-mail address: a.leemann@lancaster.ac.uk (A. Leemann).

https://doi.org/10.1016/j.wocn.2018.02.001
0095-4470/

for German (Gilles, Peters, Auer, & Selting, 2001; Peters, Gilles, Auer, & Selting, 2002; Schaeffler & Summers, 1999) as well as for American English dialects (Vicenik, 2011; Vicenik & Sundara, 2013). These studies used delexicalized speech with different types of signal manipulations. Vicenik and Sundara (2013), for example, who studied the role of temporal and f0 information in the identification of two varieties of English (American and Australian) created three conditions: low-pass filtered speech, *sasasa* speech (Ramus & Mehler, 1999), i.e. rhythm only, and a resynthesis of the f0 contour onto a steady /a/ sound, i.e. intonation only. In a 2 AFC design, they found that American English listeners were able to identify the dialects in the first two conditions, but not in the intonation-only condition. Working on dialect identification of Dutch and British English, Van Bezooijen and Gooskens (1999) monotonized f0 in one condition – retaining segmental information – and applied a low-pass filter (350 Hz) – removing segmental information – in the other condition for eight dialects of Dutch and six dialects of British English. As a control, listeners judged unmanipulated speech. The authors report that dialect identification scores decreased the less segmental information there was in the signal. It is uncertain, however, whether a low-pass filter of 350 Hz effectively strips segmental information from the speech signal, when vowels such as /i/ and /u/ produced by males, for example, have first formants smaller than 350 Hz (cf. Peterson & Barney, 1952). Further, low-pass filtered speech contains multiple layers of acoustic information (rhythm, intonation, and loudness) which makes the actual role of segments, rhythm, intonation, and intensity difficult to assess. Fuchs (2015) examined the cues listeners of Indian and British English use when asked to identify the two varieties. To disentangle prosodic and segmental information, Fuchs created stimuli with multi-dimensional combinations of Indian and British English segmental and prosodic information (108 unique types of combinations consisting of different permutations of monotone f0, low-pass filtering, and swapping consonantal and vocalic interval durations). Preliminary results indicated that segmental cues were most diagnostic for the identification of the variety, followed by f0 and rhythmic information. The role of segmental and prosodic cues in dialect identification is ultimately one of saliency, i.e. the diagnostic accessibility of linguistic features (phonetic, morphological, syntactic, or lexical). Lenz (2010) conceives of saliency as the cognitive conspicuousness of a linguistic feature: a linguistic element stands out from a given context and is thus cognitively more quickly accessible than non-salient features. Guntern (2011) provides a qualitative approach to examining the saliency of different linguistic cues for identifying SwG dialects. While taking part in an 8 AFC dialect identification experiment, listeners were asked to write down features that they perceive as most salient for the individual SwG dialects. What was striking in the notes of the participants was that most of them mentioned segmental features as the most salient diagnostic cue to identifying a speaker's dialect, such as dialect-specific realizations of /r/ or the presence or absence of /l/-vocalization. As containing equally important diagnostic cues, the subjects mentioned dialect-specific lexical items. Prosodic features such as intonation, rhythm, or speaking rate were not considered as carrying much diagnostic power. By implication, Guntern's (2011) study suggests that listeners particularly need diagnostic segmental information in the sentence material to make judgments about the origins of a speaker, e.g. if there is material where /r/s and /l/s are lacking, it may be more difficult for a listener of Swiss German to identify the speaker's regional origin.

To examine the individual role of segments and prosody in the identification of a speaker's dialect even further, segmental and prosodic information can be disentangled, which can be achieved by swapping the two levels (cf. Vaissière & Boula de Mareüil, 2004). This so-called prosody transplantation or prosody morphing paradigm has gained much attention in second language research: a number of studies have attempted to show which features – segmental or prosodic – are more important for accentedness and intelligibility in second language speakers (Boula de Mareüil & Vieru-Dimulescu, 2006; Derwing & Munro, 2005; Holm, 2008; Ulbrich, 2013; Ulbrich & Mennen, 2015; Vieru-Dimulescu & Mareüil, 2005; Winters & O'Brien, 2013). Typically, prosodic features of one language variety are morphed onto the segments of another variety and vice versa, involving a form of intelligibility, accentedness, or accent rating task conducted by native listeners.

In the present study, we employ this methodology on a dialect identification task through a set of perception experiments where we separate segmental and prosodic features from one another and play them off against each other: we manipulate the speech signal in a way that prosodic features of dialect X are morphed onto the segments of dialect Y and vice versa. These manipulated stimuli are then played to naïve listeners of dialect Z (familiar with both dialects X and Y) who are then asked to indicate whether the stimulus heard is from a speaker of dialect X or dialect Y. We will do so by pursuing the following specific research questions:

RQ1. What is the role of segments and prosody (rhythm alone and rhythm combined with intonation) in the identification of a speaker's dialect?

RQ2. How is dialect identification contingent on the sentence material used?

These research questions will be studied in the context of the above-mentioned prosody morphing paradigm: to answer RQ1 we will use material from two SwG dialects: in the first condition, listeners judge unmorphed speech, in the second condition different listeners judge rhythm morphed speech, and in the third condition, different listeners judge speech that is morphed in rhythm combined with intonation. To answer RQ2, we will look at sentence material individually and study how identification performance varies as a function of the different segmental and prosodic make-up of the sentences.

Given the literature survey presented above, for RQ1 we predict that segmental features will carry more diagnostic weight than prosodic features. While this may be the case in this study, this expectation – ultimately – will depend on which dialects are studied and, more specifically, on the (dis)similarities in the segmental and prosodic domain. It is likely that there are dialects that are very different in the prosodic domain and possibly more similar in the segmental domain; and there are dialects where this relationship is diametrically opposite. The predictions we make for RQ1 are thus only valid for the dialects examined in the current study. As for RQ2, we expect listeners' dialect identification performance to vary