



Cues to linguistic origin: The contribution of speech temporal information to foreign accent recognition



Marie-José Kolly *, Volker Dellwo

University of Zurich, Phonetics Laboratory, Department of General Linguistics, Plattenstrasse 54, 8032 Zurich, Switzerland

ARTICLE INFO

Article history:

Received 24 January 2013

Received in revised form

26 October 2013

Accepted 9 November 2013

Available online 30 November 2013

ABSTRACT

Foreign-accented speech typically contains information about speakers' linguistic origin, i.e., their native language. The present study explored the importance of different temporal and rhythmic prosodic characteristics for the recognition of French- and English-accented German. In perception experiments with Swiss German listeners, stimuli for accent recognition contained speech that was reduced artificially to convey temporal and rhythmic prosodic characteristics: (a) amplitude envelope durational information (by noise vocoding), (b) segment durations (by 1-bit requantisation) and (c) durations of voiced and voiceless intervals (by *sasasa*-delexicalisation). This preserved mainly time domain characteristics and different degrees of rudimentary information from the frequency domain. Results showed that listeners could recognise French- and English-accented German above chance even when their access to segmental and spectral cues was strongly reduced. Different types of temporal cues led to different recognition scores – segment durations were found to be the temporal cue most salient for accent recognition. Signal conditions that contained fewer segmental and spectral cues led to lower accent recognition scores.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Foreign-accented speech contains numerous cues about the native language (L1) of its speakers (Cunningham-Andersson & Engstrand, 1987). If, for example, we consider Swiss-German- or French-accented English, it is typically easy for listeners who are familiar with these varieties to recognise these two accents. What are the acoustic cues for this? On a segmental level, for example, consonants may be pronounced at a different place of articulation, in a different manner of articulation, or with different degrees of voicing (see Leemann, 2011; Schmid, 2012a): the consonant in English *the* is likely to be pronounced [z] in a prototypical French accent, [ɖ] in a prototypical Swiss German accent, thus differing from the English target [ð] in its place of articulation (French) or in place, manner and voicing (Swiss German). Similarly, /r/ in foreign-accented *random* is typically realised as a uvular trill [ʀ] or fricative [ʁ] by French speakers, as an alveolar trill [r] by Swiss German speakers. The first vowel in *random* would typically be nasalised ([ã]) by French and non-nasalised ([æ]) by Swiss Germans. Thus, segmental cues seem to play a large role for the recognition of these foreign accents (e.g. Cunningham-Andersson & Engstrand, 1987; Koster & Koet, 1993; Boula de Mareüil, Vieru-Dimulescu, Woehrling, & Adda-Decker, 2008; Park, 2013).

Apart from segmental cues there has also been a strong interest in prosodic phenomena of second language (L2) speech (Anderson-Hsieh, Johnson, & Koehler, 1992; Boula de Mareüil & Vieru-Dimulescu, 2006; Jilka, 2000; Magen, 1998; Munro, 1995; Munro, Derwing, & Burgess, 2010; Tajima, Port, & Dalby, 1997; Trouvain & Gut, 2007). This research typically deals with acoustic correlates of foreign accent degree, intelligibility, or foreign accent detection (temporal characteristics: Bent, Bradlow, & Smith, 2008; Dellwo, 2010; Holm, 2008; Munro & Derwing, 2001; Quené & van Delft, 2010; Tajima et al., 1997; Winters & O'Brien, 2013). However, the question whether particular foreign accents can be recognised based on specific prosodic cues has barely been tapped into. So far, it has been shown that speaker origin can be recognised in natural L2 speech (Derwing & Munro, 1997; Boula de Mareüil et al., 2008; Guntern, 2011; Kolly, 2013; Kumpf & King, 1997), in L2 speech with monotone intonation (Van Els & De Bot, 1987), in resynthesised L2 speech containing cues to intonation and segment durations only (Boula de Mareüil & Vieru-Dimulescu, 2006), but not in lowpass filtered L2 speech below 350 Hz (Van Els & De Bot, 1987). This body of research thus demonstrates that foreign accents can be recognised based on a variety of prosodic and segmental cues.

Little, however, is known about the role of time domain cues such as suprasegmental timing phenomena or speech rhythm in foreign accent recognition. Moreover, the lowpass filtering study by Van Els & De Bot (1987) suggests that after heavy reduction of frequency domain cues, foreign

* Corresponding author. Tel.: +41 44 634 59 48.

E-mail addresses: marie-jose.kolly@pholab.uzh.ch, mj.kolly@gmx.ch (M.-J. Kolly), volker.dellwo@uzh.ch (V. Dellwo).

accent recognition is no longer possible. Somehow contradictory evidence can be found in the domain of L1 dialect recognition where lowpass filtered speech with a cutoff frequency of 250 Hz allows for the recognition of Swiss German dialects (Leemann & Siebenhaar, 2008). The same is true for lowpass filtered speech with an unknown cutoff in recognising English dialects (Bush, 1967). Furthermore, temporal cues like durations of consonantal and vocalic intervals allow listeners to discriminate between English dialects (White, Mattys, & Wiget, 2012). We take this as an indication that temporal cues may also play a role in the recognition of foreign-accented speech. The principal aim for the present study is to explore whether temporal characteristics of foreign accented speech are perceptually salient, by investigating how the reduction of listeners' access to segmental and spectral content of speech affects their ability to recognise foreign accents.

Why temporal characteristics? It is widely acknowledged that languages (Abercrombie, 1967; Grabe & Low, 2002; Pike, 1945; Ramus, Nespor, & Mehler, 1999) and dialects (Ferragne & Pellegrino, 2004; Leemann, Dellwo, Kolly, & Schmid, 2012; Schmid, 2012b; White et al., 2012; White & Mattys, 2007b) differ in their suprasegmental temporal organisation, or speech rhythm. Whether and to what degree language-specific rhythm allows for a classification of languages into rhythmic classes is a matter of heavy debate in the literature (see Arvaniti, 2012). However, there is strong evidence that languages can be discriminated based on auditory rhythmic characteristics (Ramus & Mehler, 1999; Ramus et al., 1999). Such characteristics have been associated with the sound of a Morse-code signal for some languages (e.g. English, German, Dutch) and with the sound of a machine-gun for others (e.g. French, Italian, Spanish; Lloyd James, 1929), while the latter expresses more regular rhythmic timing – in French as opposed to English, for example. In fact, there is evidence that durational characteristics of consonantal and vocalic intervals are perceived as more regularly timed in French than they are in English (Dellwo, 2008). It was also found that Mandarin speakers produce more regularly timed speech when speaking in synchrony while such effects cannot be obtained for English (Cummins, Li, & Wang, 2013). In summary, there is evidence for some languages to be more regularly timed than others, in speech production as well as in speech perception research.

Are rhythmic characteristics transferred from L1 to L2 speech? The literature demonstrates that this is true for some L1/L2-pairs and some durational variables, but not for others. For example, the rate-normalised durational variability of vocalic intervals¹ locates L2 speech in between the native and the target language values for rhythmically regular Spanish vs. irregular English (Carter, 2005; Gutiérrez Díez, Dellwo, Gavalda, & Rosen, 2008; White & Mattys, 2007a). English and Dutch, which are both rhythmically irregular, show very similar values for native and target language as well as L2 speech (White & Mattys, 2007a). This points in favour of an L1-transfer hypothesis. However, other findings do not support such a hypothesis: Regarding the percentage over which speech is vocalic,² English learners of Spanish (White & Mattys, 2007a) as well as German learners of French and English (Dellwo, 2010) overshoot the values of their native as well as their target language. A high percentage over which speech is vocalic seems to be a general property of L2 speech. In fact, L2 speakers tend to lengthen the duration of vowels, particularly of unstressed vowels, giving the auditory impression of more regular speech timing (Adams & Munro, 1978; Taylor, 1981). Thus, L2 speech seems to be influenced by L1 durational characteristics for some variables and language pairs; other variables and language pairs, however, seem to reflect general properties of L2 speech rather than specific L1-transfer, as suggested by Taylor (1981) and Dellwo (2010). It therefore remains unclear, for L2 speech, which of the durational characteristics associated with speech rhythm are L1-specific, and which are a general feature of (L1-independent) L2 speech. It further remains widely unclear whether such acoustic variability between L2 accents is perceptually salient. While perceptually salient rhythmic differences between some languages have been empirically attested by different studies (Nazzi, Bertoni, & Mehler, 1998; Ramus & Mehler, 1999; Ramus, Dupoux, & Mehler, 2003), the idea that such characteristics also play a role in L2 speech has been investigated empirically only for speech production (Dellwo, 2010; White & Mattys, 2007a, 2007b).

Durational characteristics of foreign-accented speech may be perceptually salient typically if speakers were to transfer durational patterns from a rhythmically more regular L1 to a less regular L2. This can be tested, for example, with French- and English-accented German speech: two foreign accents that stem from two languages that have been shown to differ in time domain characteristics (French and English; Abercrombie, 1967; Dellwo, 2006; Grabe & Low, 2002; Pike, 1945; Ramus et al., 1999). A rationale for this is the following: English and German, in contrast to French, are characterised by vowel reduction, complex syllables and consonant clusters, high durational variability between stressed and unstressed syllables. In comparison, French has less vowel reduction, less complex syllables and consonant clusters as well as less durational variability between stressed and unstressed syllables (Dauer, 1983; Auer, 2001). The percept of rhythmic regularity in French may be a result of such phonological characteristics. If language-typical phonological characteristics were indeed transferred from L1 to L2 speech, one would expect French accented German to sound rhythmically more regular than English accented German.

Cues for the perception of speech rhythmic characteristics are assumed to lie in the more or less regular recurrence of perceptually salient speech intervals. Since durational patterns are encoded on many levels in the speech signal, different types of such speech intervals have been considered to be acoustic correlates of speech rhythm: interstress intervals and syllables (Pike, 1945; Abercrombie, 1967), consonantal and vocalic intervals (Ramus et al., 1999), voiced and voiceless intervals (Dellwo, Fourcin, & Abberton, 2007; Fourcin & Dellwo, 2009), intervals related to amplitude envelope timing (Lee & Todd, 2004; Dellwo, Leemann, & Kolly, 2012; Tilsen & Johnson, 2008) or to fundamental frequency (Kohler, 2009). In research on speech perception, a small number of speech intervals have been used to study language discrimination based on durational characteristics: It has been shown that listeners can discriminate a rhythmically regular from an irregular language based on monotone lowpass filtered speech below 180 Hz (den Os, 1988) and based on the durational variability of consonantal and vocalic intervals in monotone *sasasa*-speech³ (Ramus et al., 2003). Research on rhythm production and perception has thus mainly focused on temporal characteristics of vocalic and consonantal intervals. To test foreign accent recognition in conditions of heavily reduced frequency domain information, it thus seems reasonable to use different types of speech intervals to present time domain information to listeners. Durational characteristics of some speech intervals may contain more or less information about the L1 origin in L2 speech, which may lead to different accent recognition scores.

Based on the ideas presented above, we formulated the following research questions: To what degree can we reduce frequency domain characteristics of the speech signal such that listeners can still recognise two different foreign accents? And which type of temporal cue (i.e., which type of temporally structured speech interval) leads to higher accent recognition scores? To test this, Swiss German listeners were asked to recognise French- and English-accented German in signal-degraded speech containing primarily durational cues. In a between-subject design we used three different types of signal-degraded speech to provide listeners with different types of temporal cues. By doing this, we gain insight into the speech intervals that contribute more or less to accent recognition, i.e., the speech intervals which are (a) subject to durational L1-transfer

¹ A variable that has been shown to discriminate between hypothesised rhythm classes (Dellwo, 2006; White & Mattys, 2007a).

² This variable also discriminates between hypothesised rhythm classes (Ramus et al., 1999).

³ Ramus & Mehler (1999) developed this procedure for delexicalisation, where the original utterances are resynthesised: every consonantal speech interval is replaced with the same [s]- and every vocalic speech interval with the same [a]-phone.

Download English Version:

<https://daneshyari.com/en/article/1100846>

Download Persian Version:

<https://daneshyari.com/article/1100846>

[Daneshyari.com](https://daneshyari.com)