Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 55 (2013) 486–507

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Perceived accentedness and intelligibility: The relative contributions of F0 and duration

Stephen Winters [a], Mary Grantham O'Brien [b,*]

[a] Department of Linguistics, University of Calgary, Social Sciences 814, 2500 University Dr. NW, Calgary, Canada AB T2N 1N4
[b] Department of Germanic, Slavic and East Asian Studies, University of Calgary, Craigie Hall C208, 2500 University Dr. NW, Calgary, Canada AB T2N 1N4

Received 12 February 2012; received in revised form 4 December 2012; accepted 11 December 2012
Available online 20 December 2012

Abstract

The current study sought to determine the relative contributions of suprasegmental and segmental features to the perception of foreign accent and intelligibility in both first language (L1) and second language (L2) German and English speech. Suprasegmental and segmental features were manipulated independently by transferring (1) native intonation contours and/or syllable durations onto non-native segments and (2) non-native intonation contours and/or syllable durations onto native segments in both English and German. These resynthesized stimuli were then presented, in an intelligibility task, to native speakers of German and English who were proficient in both languages. Both of these groups of speakers and monolingual native speakers of English also rated the foreign accentedness of the manipulated stimuli. In general, tokens became more accented and less intelligible, the more they were manipulated. Tokens were also less accented and more intelligible when produced by speakers of (and in) the listeners' L1. Nonetheless, in certain L2 productions, there was both a reduction in perceived accentedness and decreased intelligibility for tokens in which native prosody was applied to non-native segments, indicating a disconnect between the perceptual processing of intelligibility and accent.

Keywords: Foreign accent; Intelligibility; Second language; Perception; Prosody

1. Introduction

Whereas children learn the sound system of their first language (L1) with relative ease, research into second language (L2) acquisition has shown that the acquisition of a non-native phonological system and its phonetic realizations are among the most problematic areas for learners, both in perception and production. For this reason, non-native productions of a second language often deviate from native norms and can therefore sound "accented", or may even be difficult for native listeners to understand. But what is the phonetic source of a "foreign accent"? Which phonetic features of L2 speech make it "accented", or potentially difficult to understand?

This study sought to investigate the relative contributions of two suprasegmental features—intonation and syllable duration—and also segmental phonetic features to the intelligibility and perceived accentedness of both L1 and L2 speech. Accent rating tasks—generally using native speakers as raters—have shown that various segmental and suprasegmental features contribute to the relative accentedness of L2 speech (Holm, 2008; Jilka, 2007; Riney et al., 2000; Trofimovich and Baker, 2006). Other studies have independently shown that L2 speech which is perceived as "accented" may nonetheless be relatively easy to understand, or intelligible (Sidaras et al., 2009). However, the results of several recent studies that have attempted to determine the relative contributions of segmental vs. suprasegmental features in perceived accentedness and intelligibility have proven inconclusive (Anderson-Hsieh et al., 1992; Boula de Mareüil and Vieru-Dimulescu, 2006; Derwing and Munro, 1997; Hahn, 2004; Jilka, 2000; Munro and Derwing, 2005). This study expands on the results of

* Corresponding author. Tel.: +1 (403) 210 8542; fax: +1 (403) 284 3810.
E-mail addresses: swinters@ucalgary.ca (S. Winters), mgobrien@ucalgary.ca (M.G. O'Brien).

0167-6393/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2012.12.006

this earlier work by investigating the effects that segmental and suprasegmental cues have on the perception of accentedness in *both* L1 and L2 speech. Unlike previous studies, we seek to determine:

(a) the perceptual effects of mapping native prosody (intonation and duration) onto non-native segments; and
(b) the perceptual effects of mapping non-native prosody (intonation and duration) onto native segments.

Intuitively, it seems likely that the transplantation of native prosody onto a segmentally non-native production of a sentence would *reduce* that production's perceived accentedness, by making it more closely match the prosodic norms for that language. Conversely, the transplantation of non-native prosody onto a segmentally native production of a sentence would likely *increase* that production's perceived accentedness. Holm (2008), who manipulated both intonation and duration cues in the production of non-native sentences of Norwegian, has provided evidence for the perceptual benefits of (a), but not (b). Pinet and Iverson (2010) performed a combined (intonation and duration) PSOLA resynthesis in both directions in English and French and found that the benefits of (a) and (b) varied according to the L1 and L2 experience of the listener in question. Boula de Mareüil and Vieru-Dimulescu (2006) also transplanted prosodic cues across languages in a study wherein speakers produced the same sentences (segmentally) in two different languages (Spanish and Italian), but the listeners only identified the language category of the speech sample (Spanish, Spanish with an Italian accent, Italian with a Spanish accent, or Italian). With this experimental paradigm, we aim to more accurately quantify the perceptual effects—in both potential directions—of prosodic and segmental cues on the accentedness of L1 and L2 speech.

In this study, the durations and F0 contours of German and English sentences (as produced by both native and non-native speakers) were merged with the segmental content of the same sentences (as produced by a different speaker) and then presented to German-L1 and English-L1 listeners in both a sentence transcription task and an accent rating task. These two perceptual tasks shared the same test materials and—to a certain extent—the same set of listeners. Their procedures and results will be reported separately—as "Task 1" (accent rating) and "Task 2" (intelligibility)—for the sake of clarity of presentation.

2. Methods

2.1. Materials

Natural speech tokens were produced by four speakers, who read 24 sentences each in English and German. Two of these speakers were native English speakers who were highly proficient in German; the other two were native German speakers who were highly proficient in English. All speakers spoke standard varieties of their first languages. One speaker from each group was female, and the other speaker in each group was male. Information about the speakers is provided in Table 1 below.

The sentences these speakers produced consisted (in each language) of 12 declarative sentences, 6 yes/no questions and 6 wh-questions. With 24 sentences in each language, the speakers produced a total of 48 sentences for the study. Both declarative and interrogative sentences were included in order to incorporate some intonational variability in the speakers' original productions. A complete listing of all the sentences used for the study may be found in the Appendices.

All speakers were recorded as they read these sentences, at a normal pace in both languages, in a sound-attenuated booth in the phonetics laboratory at a university in western Canada. Speakers read from a printed script into a Sennheiser AX condenser microphone placed approximately 18 inches in front of their lips. The talkers' productions of each sentence were automatically digitized at a sampling rate of 48 kHz by an Edirol Hi-Speed USB UA-101 Audio Capture device, attached to a Rain computer, running Adobe Audition (version 2.0). Each speaker's recordings were later spliced, using Praat (Boersma and Weenink, 2007) into individual files for each sentence, and then segmented and labeled—by syllable—prior to further processing. Syllable boundaries were generally placed at the beginning of the first sound in each syllable (whether consonant or vowel); in cases of apparent double consonants across syllable boundaries (e.g., "allein"), the syllable

Table 1
Speaker details.

L1	English		German	
Gender	Female	Male	Female	Male
Age	24	21	26	30
Place of birth	Western Canada	Western Canada	Paderborn, Germany	Kassel, Germany
Age of learning L2	5	18	10	12
CEFR level[a]	C1	C1	C2	C1

[a] Native speakers of English completed the online proficiency test from the Goethe Institut (Einstufungstest [Placement test] and 2004, 2004), and native speakers of German completed the Oxford Online Placement Test (University Press. Oxford Online Placement Test, 2009).