SINGING YOUR ACCENT AWAY, AND WHY IT WORKS

Marinda Hagen[a], *Joop Kerkhoff*[b] *& Carlos Gussenhoven*[a,b]

[a]Centre for Language Studies, Radboud University Nijmegen, the Netherlands;
[b]Queen Mary University of London, UK
marinda.hagen@hotmail.nl; c.gussenhoven@let.ru.nl

ABSTRACT

Eleven Dutch secondary school children sang and spoke the lyrics of a number of popular English songs. Their pronunciation accuracy was rated significantly higher in the singing mode by a group of native speakers. Additional perception scores showed that the decrease in Dutch-accentedness must be due to the fact that their foreign accent could not be signaled in the f0 and segment durations during singing.

Keywords: foreign accent, singing, Dutch accented English

1. INTRODUCTION

Adults learning a foreign or second language are typically perceived as having an accent by native speakers of the language. A great number of studies have investigated the concept of *foreign accent* (e.g., [4, 5, 7, 10, 13]), many of which have dealt with the perception of foreign-accentedness, and the factors that affect it, like speaking rate [11, 12]. To the best of our knowledge, the effect of singing on the perception of foreign-accentedness has never been investigated. It is this factor which the present study addresses.

Sounds (and other units of speech) may differ from one another in quality, pitch, loudness and duration [9]. Non-native speakers are often recognized by native speakers because of deviations from native norms for those phonetic dimensions. Our concern is not to explain foreign-accentedness (e.g., [1, 2, 6, 8]), but to show that deviations from native norms during singing only concern quality and loudness and that as a result foreign-accentedness is reduced when non-native speakers sing.

A number of differences between speech and singing have been observed. Professional Western opera singers modify the quality of their vowels to avoid masking by the loud orchestral accompaniment they generally have to sing to. Male professional Western opera singers accomplish this by bringing F3, F4, and F5 closer together, increasing loudness and audibility. This is also characteristic of male Western concert singing [15, 16]. The modifications made by female Western opera singers are different from those described above; when F1 is lower than f0, F1 is brought closer to f0, increasing audibility [16]. More generally, the pitch contour in singing is determined by the composer [16] and the same is true for duration. Sundberg [17] points out another difference between singing and speech when he notes that pitch changes in singing are perceived differently from those in speech. While pitch change size in speech is perceived continuously, pitch changes in singing are categorized into musical intervals. Also, tone length differences in singing are perceived categorically, i.e., as note values [17]. But the most important and general difference between speech and singing for our purposes is that singing imposes duration and f0 patterns on speech. Consequently, the usual segmental durations and f0 variation of speech are not available in singing.

Anecdotally, people have reported that when they sing, they seem to have a less strong foreign accent than when they speak. This contribution investigates the effect of singing on the perception of Dutch-accentedness in English. We hypothesize, first, that Dutch speakers of English have a less strong foreign accent when they sing than when they speak. At first blush, a number of explanations might be put forward if this were found to be true. One is that the realization of segments is more authentic in singing than in speech. Another might be that listeners judge foreign-accentedness in singing less harshly than that in speech. A third candidate, one which we consider the most likely explanation, is that in singing, the syllable durations and f0 patterns of spoken texts are imposed by the rhythm and melody of the musical score. As a result, the conventional durations and f0 of spoken texts are no longer available, and thus neither is their contribution to the degree of foreign-accentedness.

2. METHOD

2.1. Speakers

Eleven Dutch secondary school students, two of whom were male, were recruited from secondary schools in Duiven and Arnhem (Gelderland, The Netherlands). Their ages ranged from fifteen through eighteen years. Six native speakers of English, two of whom were male, with ages between 23 and 47 years, served as a control group. Two of them grew up in Glasgow; one lived in Glasgow until the age of five and then moved to the Isle of Man; another is a bilingual who lived in the Netherlands until the age of twelve and then moved to the UK; the fifth grew up in the USA, but not in one particular State; and the last grew up in Bath, UK. Their education levels ranged from 'College' to 'PhD'. We selected adult native speakers of different varieties of English. We did not recruit RP speakers in order to avoid a possible ceiling effect in accentedness judgements.

2.2. Materials

In order to test the hypothesis that Dutch speakers of English have a less strong foreign accent when they sing than when they speak, we prepared eleven pairs of spoken and sung versions of the lyrics of four songs by Dutch speakers of English, as well as six pairs of spoken and sung versions produced by native speakers of English, to serve as a control. Using a Sony MZ-R55 MiniDisc recorder with external microphone, we recorded the eleven Dutch speakers of English, five of whom first read the lyric they had chosen and then sang it, while the other six did these in the opposite order. From the recordings by each participant, we selected two representative sets of sung and spoken text fragments, which gave us 22 stimuli. We produced comparable sets of spoken and sung stimuli by the six native speakers, giving another twelve stimuli. Text fragments ranged between 30 and 60 words in length.

In order to test the hypothesis that the absence of duration and f0 cues is responsible for any reduction in foreign-accentedness during singing, we required English utterances spoken by Dutch speakers of English in which segmental durations and f0 have been made unavailable for signalling foreign-accentedness. To this end, one sentence-long fragment was taken from the recordings of the spoken versions by each of the eleven Dutch participants. We produced (i) a *manipulated* set in which the durations of all segment durations in the experimental utterances were matched with durations of the equivalent segments in the utterances as produced by one native English speaker; (ii) a *monotonized original* set in which f0 was monotonized; (iii) a *monotonized manipulated* set in which the manipulated utterances were monotonized. Segmentations and manipulations were carried out with Praat [3]. We thus ended up with 33 manipulated sentence-long fragments in addition to the eleven originals, making a total of 44 stimuli to be used to test the hypothesis that any higher scores for the sung versions was to be attributed to the unavailability of duration and f0 to signal Dutch-accentedness.

2.3. Procedure

In order to hide the aim of the experiment from the judges, we created two tests: the 'Spoken Test', containing stimuli taken from the spoken versions of the recordings, and the 'Sung Test', which had stimuli taken from the sung versions. The general structure of the tests was otherwise the same (see Table 1). Eight Spoken Agreement stimuli and eight Sung Agreement stimuli were included to be able to check the agreement between the two groups of listeners. The Spoken Agreement stimuli and the manipulated spoken sentences were put at the end of the Sung Test, and the Sung Agreement stimuli were put at the end of the Spoken Test, in order not to reveal the aim of the experiment to the subjects before they had finished scoring the experimental stimuli.

Table 1: Structure of Spoken Test and Sung Test.

Spoken Test	Sung Test
8 spoken practice stimuli	8 sung practice stimuli
8 spoken Agreement stimuli	8 sung Agreement stimuli
34 spoken Experimental stimuli 11 original and 33 manipulated sentences	34 sung Experimental stimuli
8 sung Agreement stimuli	8 spoken Agreement stimuli
	11 original and 33 manipulated sentences

Twenty native English listeners, five of whom were male, were recruited in Norwich, UK, through a local intermediary. They were asked to listen to and judge the stimuli on native accent authenticity using a 7-point scale, with '1' meaning 'very strong foreign accent' and '7' meaning 'native'. Ten listeners judged the stimuli in the Spoken Test, the other ten those in the Sung

Test. Listeners' ages ranged from 30 to 59 years. The education levels ranged from 'secondary education' to 'BA'.

3. RESULTS

3.1. Sung vs. spoken stimuli

Agreement among the twenty judges was established on the basis of their scores on the 16 agreement stimuli. Cronbach's alpha was 0.925, which suggests that they interpreted the variation in the stimuli in similar ways, to a degree that we felt made it unnecessary to remove subjects. However, mean scores over all agreement passages were highly variable across subjects, who must have interpreted the scales very differently in absolute terms. Means varied from a low 2.88 to a high 6.63. We normalized the scores by carrying out 20 linear regressions with the mean scores over judges as the dependent variable and each subject's scores as the independent variable. Scores were subsequently corrected using the linear regression model, which yielded a mean score of 4.60 for all judges.

3.1.1. Experimental stimuli

In order to establish whether the degree of foreign-accentedness is lower in the sung texts than in the spoken texts in the Dutch group, mean corrected scores of the Dutch speakers were submitted to an independent samples t-test. For nine of the eleven subjects, the scores for the sung versions were significantly different from those for the spoken versions (see Table 2). The scores for the sung versions were higher than those for the spoken versions in all eleven cases.

Table 2: T-values and p-values for speakers who have significantly higher scores for the sung versions than for the spoken versions.

	t-value	*p*-value
speaker 1	3.429	.003
speaker 4	3.426	.003
speaker 5	6.394	.000
speaker 6	3.099	.007
speaker 7	4.220	.001
speaker 8	4.680	.000
speaker 9	2.434	.026
speaker 10	4.091	.001
speaker 11	3.946	.001

To test whether any difference was found in the native speaker group, the same procedure was applied to their mean normalized scores. This showed that for only one of the six native speakers was there a significant difference between singing and speaking, with the spoken versions being judged as more authentic ($t[13.615]=-2.261$, $p<.05$). Thus, the sung versions by the Dutch speakers are judged to be more authentic than the spoken versions, while no such difference is found in the native speaker group.

3.2. Manipulation stimuli

In order to test the further hypothesis that the results are explained by the unavailability of duration and f0 for signalling foreign accent in singing, a mixed-effects model (R) analysis was carried out on the raw scores for the three manipulation versions. Mean scores differed significantly ($F[2, 638.00]=3.784$, $p<.05$). Of the monotonized versions, those with English durations were judged to be more authentic than those with the original durations, and of the versions with the English durations, the *monotonized* set was judged to be more authentic than the set with the original intonation ($p<.05$).

4. CONCLUSIONS

Dutch speakers of English have a less strong foreign accent when they sing than when they speak. The results for all eleven subjects consistently showed higher native accent authenticity for the sung than for the spoken versions of the text fragments. A control group of six native speakers showed no consistent differences between their spoken and sung versions of the same text fragments.

There are a number of indications that these results are to be interpreted as being due to the unavailability of f0 and duration to signal one's foreign accent. The musical art form represented by the sung versions imposes its own duration and f0 variation on the text, leaving spectral variation and intensity variation as the only sources of accent information. We are not aware of work showing that intensity variation carries cues to foreign accent and assume they do not.

The first of these is the fact that the control group of native English speakers showed no effect of speech mode. This suggests that the f0 and duration variation in their spoken versions carried no foreign-accentedness information that could be hidden in their sung versions. Second, when original durations in the productions of each of the Dutch participants were replaced with durations taken from those of a native speaker, native speaker authenticity increased significantly. This shows that the Dutch speakers of English did in

fact produce foreign-accented segment durations in their spoken versions. Third, when we replaced the original f0 contour in the durationally manipulated utterances with consistent monotones, the native-accent authenticity increased. This shows that the Dutch speakers must have used f0 contours that signal foreign-accentedness to the native listeners. The durationally manipulated versions with monotones and those with original f0 contours only differed in the f0 contour, and both had comparable traces of the resynthesis procedure, and thus were comparable for the variable under consideration.

Thus, it is improbable that the increased authenticity in the sung versions of the Dutch listeners is to be attributed to a more lenient attitude of the judges listening to sung speech. Neither is it plausible that the speakers somehow improved the authenticity of their pronunciation in the sung mode. The intriguing question now is whether the improvement in accent authenticity in sung speech is specific to Dutch speakers of English. The two languages are closely related. Experiments with infants have shown that the rhythmic properties of Dutch and English are indistinguishable to new-born infants, as established with the help of low-pass filtered stimuli in which segmental information was removed (for a survey, see [14]). It would be reasonable to assume that language pairs that are typologically further apart show even greater effects, since their prosodic features are likely to differ more. However, the same can be said of the spectral (segmental) features. Our experiment is a first step towards establishing an easy test of the proportion of a foreign accent that is due to 'suprasegmentals'. Claims about such proportions are often made with no empirical data to back them up. Our experiment suggests that a singing test will provide a quick answer to such questions.

5. REFERENCES

[1] Best, C.T. 1995. A direct realist view of cross-language speech perception. In Strange, W. (ed.), *Speech Perception and Linguistic Experience*. Timonium, MD: York Press, 171-204.

[2] Best, C.T., Tyler, M.D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Munro, M.J., Bohn, O.S. (eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Amsterdam: Benjamins, 13-34.

[3] Boersma, P., Weenink, D. 2002-2009. Doing phonetics by computer. *http://www.praat.org*

[4] Flege, J.E. 1981. The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly* 15, 443-455.

[5] Flege, J.E. 1988. Factors affecting degree of perceived foreign accent in English sentences. *J. Acoust. Soc. Am.* 84, 70-79.

[6] Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In Strange, W. (ed.), *Speech Perception and Linguistic Experience*. Timonium, MD: York Press, 233-277.

[7] Flege, J.E., Eefting, W. 1988. Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *J. Acoust. Soc. Am.* 83, 729-740.

[8] Lai, E.Y.W. 2002. *Prosody and Prosodic Transfer in Foreign Language Acquisition: Cantonese and Japanese.* Muenchen: Lincom Europa.

[9] Laver, J. 1994. *Principles of Phonetics.* Cambridge: Cambridge University Press.

[10] McAllister, R., Flege, J.E., Piske, T. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30, 229-258.

[11] Munro, M.J., Derwing, T.M. 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48, 159-182.

[12] Munro, M.J., Derwing, T.M. 2001. Modelling perceptions of the accentedness and comprehensibility of L2 speech. *SSLA* 23, 451-468.

[13] Munro, M.J., Flege J.E., MacKay, I.R.A. 1996. The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics* 17, 313-334.

[14] Nazzi, T., Ramus, F. 2003. Perception and acquisition of linguistic rhythm by infants. *Speech Communication* 41, 233-243.

[15] Sundberg, J. 1974. Articulatory interpretation of the 'singing' formant. *J. Acoust. Soc. Am.* 55, 838-844.

[16] Sundberg, J. 1987. *The Science of the Singing Voice.* DeKalb, IL: Northern Illinois University Press.

[17] Sundberg, J. 1994. Perceptual aspects of singing. *Journal of Voice* 8, 106-122.