QUANTIFYING THE INTERLANGUAGE SPEECH INTELLIGIBILITY BENEFIT

Wang Hongyan[*] and Vincent J. van Heuven

Phonetics Laboratory, Leiden University Centre for Linguistics ([*] now at Shenzhen University)
WANGhongyan0069@hotmail.com; V.J.J.P.van.Heuven@Let.LeidenUniv.NL

ABSTRACT

Generally, native listeners of a target language are better at understanding foreign-accented speech than any other type of listener, with one possible exception: if the listener speaks the same mother tongue as the speaker, e.g. when Chinese speakers and listeners communicate in English, the information transfer may be more successful than with a native English listener. We review literature data, and present results of our own in an attempt to come up with the optimal quantification of this so-called interlanguage speech intelligibility effect. We argue that the benefit is best quantified in relative terms, as the residual in a linear model that remains after the main effects of speaker and hearer language background have been included.

Keywords: interlanguage, foreign accent, speech intelligibility, Mandarin, English, Dutch.

1. INTRODUCTION

A lot of research has been done on the intelligibility of foreign-accented speech. The general result is that foreign-accented speech is less intelligible to native listeners of the target language than native speech. The reason for the loss of intelligibility is that the accented speech has segmental deviations from the norm such that speech sounds are off-target, so that they are more difficult to identify as tokens of the intended phoneme, or are even identified as a token of a different phoneme. In a classical study, word recognition by native listeners proved more difficult for Serbian-, Japanese- and Punjabi-accented English was some 36% poorer than for native English speech in a range of signal-to-noise ratios and filtering conditions [1]. More recently, it was shown that the word error rate of English spoken with a Mandarin accent was 11% against a mere 4% for native American control speakers, when in both cases the listeners were Americans ([2]). Using a different methodology, native-speaker superiority was measured in terms of the Speech Reception Threshold (SRT) by [3]. SRT was found to be at a 4-dB better signal-to noise ratio when the Dutch listeners responded to Dutch speakers, than when the speakers were British learners of Dutch.

By the same token, native listeners generally perform better at all sorts of speech recognition tasks than foreign listeners do. They are better at recognizing degraded speech (telephone speech, synthetic speech, speech in noise) than non-native speakers. For instance, Dutch listeners could recognize Dutch words from shorter onset portions than English learners of Dutch, even if the latter had resided in the Netherlands for twenty years [4].

Native listeners have a vast knowledge of the sound system and statistical structure of the lexicon, which allows them to optimally exploit the redundancy patterns in the language so as to compensate for any deficiencies in the speech input. From this one would predict that native listeners will always outperform non-native listeners. There is recent evidence, however, to suggest that this prediction, although largely correct, sometimes fails, namely when a non-native listener communicates with a non-native speaker with whom he shares the same language background. For instance, it may the case that Mandarin-accented English is more readily understood by a Mandarin listener than by a native listener of English. Here, the Mandarin learners of English, whether speaker or listener, both have access to the phonetics and phonology of Mandarin, so that they speak the same interlanguage, i.e. the type of non-native English that is influenced by Mandarin as the interfering source language.

This so-called interlanguage speech intelligibility benefit was addressed in earlier research [3, 5]. We argue, however, that the researchers did not quantify the interlanguage benefit in an optimally insightful fashion. A relative, rather than an absolute, quantification of the benefit will reveal the true extent of the effect. We will explain the procedure, then present results on the interlanguage benefit from our own experimental data,

quantified both in absolute and relative terms. We will show that interlanguage benefit generally does not show in absolute terms, but that it is pervasive when quantified in relative terms. At the end of the paper we will apply our relative measure to earlier work in the literature and show that it provides a better understanding of the effects.

2. MUTUAL INTELIGIBILITY DATA

We ran a large study on the mutual intelligibility of Dutch, Mandarin and American speakers of English. Speakers from these three different native-language backgrounds produced materials in English, i.e. (i) vowels in /hVd/ contexts, (ii) consonants and (iii) consonant clusters in inter-vocalic contexts, (iv) semantically unpredictable sentences (SUS), and (v) semantically meaningful sentences with final target words in unpredictable (non-pregnant) and (vi) predictable (pregnant) contexts [6, 7]. The materials of one representative male and one female speaker for each of the three language backgrounds were then offered for ident-ification (of vowels, consonants and clusters) or recognition (of words in sentences) to 20 listeners in each of three countries (20 learners of English at Changchun University (Mandarin language area), 20 learners of English at Leiden University (Netherlands) and 20 American native listeners at UCLA, Los Angeles), so that all nine possible combinations of speaker and listener backgrounds occurred equally often in the experiment.

Figure 1 presents the results of the vowel identification part of the study. Here percent correct vowel identification is plotted vertically broken down first by the language background of the listeners (Mandarin, Dutch, English) and broken down further by the language background of the speakers. The data were submitted to an Analysis of Variance (ANOVA) run on the mean percent correct scores for each listener with the first language (L1) of speaker and L1 of listener as fixed factors.

Across speaker groups, the Chinese listeners have the lowest vowel identification scores (29–34% correct, mean = 32%). Dutch listeners perform at an intermediate level (40–59% correct, mean = 53%), and the American listeners are the best (45–75% correct, mean = 60%). The effect of listener group was significant, $F(2, 315) = 204.9$ ($p < .001$). A post-hoc test (Scheffé procedure with $\alpha = .05$) indicates that all three speaker L1 backgrounds were different from each other.

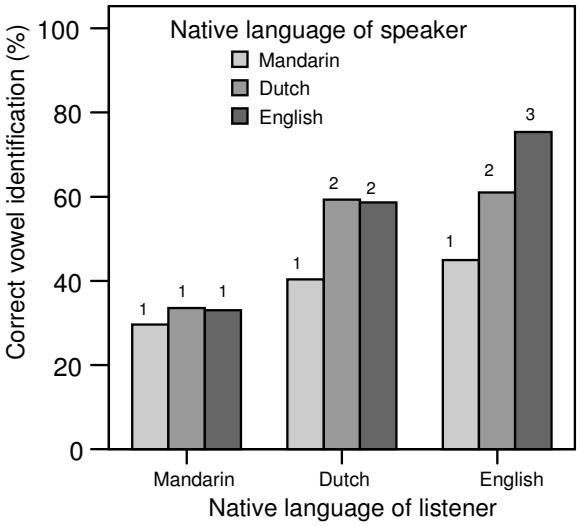

Figure 1. Correctly identified vowels (%) for Mandarin, Dutch and American listeners broken down by accent of speaker. Numbers above the bars indicate subgroup membership as determined by a Scheffé procedure.

Across listener groups, Chinese speakers obtained the lowest vowel identification scores (38%). The Dutch and American speakers' vowels were identified with 51% and 56% correct, respectively. The effect of speaker L1 is significant, $F(2, 315) = 77.7$ ($p < .001$). The Mandarin speakers are significantly poorer than the other two speaker groups, which do not differ from each other. We may note that the effect of listener L1 is almost three times larger than the effect of speaker L1.

Crucially, the interaction between listener and speaker L1 also reaches significance, $F(4, 315) = 17.0$ ($p < .001$). This implies that the mean scores obtained for specific combinations of listener and speaker L1 cannot be computed by simply adding or subtracting a term for each factor level. Specifically, it can be shown that listeners obtain higher vowel identification scores when responding to materials produced by speakers of their own native language. This can be shown by computing the expected scores for each of the nine possible combinations of listener and speaker language background and then comparing this ex-pected score with the observed score. Mean per-cent correct vowel identification equals 50. When the listeners are Chinese, Dutch and American, the expected score is –18, +3 and +10 below or above the mean; for the three speaker L1 backgrounds the mean should be corrected with –12, +1 and +6,

respectively. Expected and observed scores are listed in Table 1 together with the difference between the two (Δ, residual or prediction error).

Table 1. Expected vowel identification scores (% correct) on the basis of grand mean = 50% and main effects for Listener and Speaker L1. Observed scores (Obs.) and residuals (Δ) are indicated. Bolded delta's represent the interlanguage or native language benefit.

	Language background of				Exp.	Obs.	Δ
	Listener		Speaker				
1.	Mandarin	–18	Mandarin	–12	20	30	**+10**
2.	Mandarin	–18	Dutch	+1	33	34	+1
3.	Mandarin	–18	English	+6	38	34	–4
4.	Dutch	+3	Mandarin	–12	41	40	–1
5.	Dutch	+3	Dutch	+1	54	59	**+5**
6.	Dutch	+3	English	+6	59	59	0
7.	English	+10	Mandarin	–12	48	45	–3
8.	English	+10	Dutch	+1	61	61	0
9.	English	+10	English	+6	66	75	**+9**

Generally, the observed scores are correctly predicted or even underestimated by the linear addition of the two main effects. Only in three combinations of factor levels is the observed score substantially better than the prediction. These are precisely the conditions in which the listeners are confronted with vowel tokens spoken by their fellow countrymen (shaded rows in Table 1). This native or interlanguage benefit is 5 to 10 percentage points better than the expected score.

Table 2 lists the segment identification or word recognition scores for the six tests broken down by the nine combinations of listener and speaker language backgrounds. For each of 18 situations the listener group with the best score is bolded. Table 2 shows that the American listeners outperform the other two listener groups (Chinese, Dutch) in 15 out of 18 situations. Yet, in three test × speaker L1 conditions, the American listeners are outperformed, albeit only just, by Dutch listeners. Crucially, is each of these three exceptional situations, the Dutch listeners respond to Dutch speakers of English. These three exceptional conditions, then, are examples of what we might call absolute interlanguage benefit. However, the Chinese speaker-listener combinations do not seem to reap any benefit from their matched interlanguage. If we quantify the interlanguage benefit in the relative manner, as explained above, the matter takes on a different complexion.

Table 2. Summary of test results. Percent correct on each of six tests broken down by L1 of listener and broken down further by L1 of speaker. Each mean is based on 36 listeners. The listener group with the best performance is represented in bold face in a shaded cell.

L1 of		Tests					
List.	Speak.	Vow	Cons	Clus.	SUS	LP	HP
Mandarin	Mand.	29.7	57.2	52.8	39.3	19.4	16.7
	Dutch	33.5	46.8	36.9	39.0	38.9	37.8
	English	33.1	58.2	56.0	44.2	17.9	31.8
Dutch	Mand.	40.3	66.6	78.8	57.1	26.9	33.1
	Dutch	59.3	73.7	**87.8**	**86.2**	**81.3**	76.1
	English	58.6	80.6	89.1	90.5	77.8	84.9
English	Mand.	**44.9**	**72.5**	**82.5**	**59.5**	**39.4**	**57.8**
	Dutch	**61.0**	**76.1**	85.7	83.0	67.7	**99.4**
	English	**75.3**	**87.7**	**89.3**	**95.5**	**95.2**	**99.1**

The size of the relative interlanguage benefit is plotted in Figure 2 for each of the six segment identification or word recognition tests and broken down by the language background of the listeners.

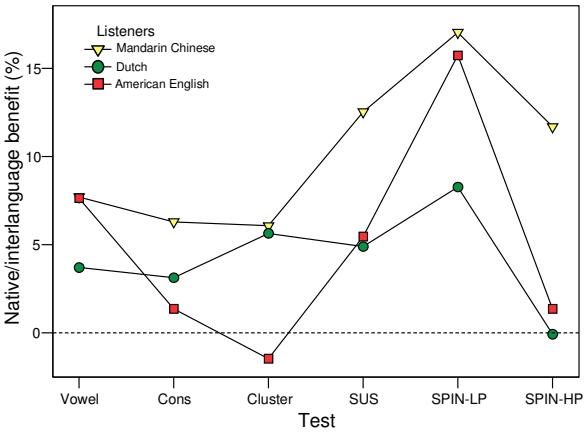

Figure 2. Native/interlanguage benefit (percentage points) for Chinese, Dutch and American speaker-hearers of English, for six tests (further see text).

In Figure 2 we have only plotted the benefit for combinations of speakers and listeners sharing the same native language. In 16 out of 18 conditions there is weak to strong interlanguage or (in the case of the American listeners) native-language speech intelligibility benefit. If we had also plotted the results for the speaker and listener groups not sharing the same native language, we would have seen the exact mirror image, i.e., negative residuals with just a few exceptions.

3. REANALYSIS OF BENT & BRADLOW

Bent & Bradlow [5] examined the interlanguage benefit in a database with mutual intelligibility scores in English obtained for three types of speakers: a high-proficiency Korean L2 speaker, a high-proficiency Chinese learner, and a native speaker of American English. Sentences produced by these three speakers were presented to four groups of listeners: Chinese, Korean, American and mixed-foreign backgrounds. Intelligibility scores were determined for all $3 \times 4 = 12$ combinations of speaker and hearer L1 backgrounds. The results showed that the L2 speakers were as intelligible as the native English speaker if the listeners shared the native language of the speaker. This then qualifies as a case of matched interlanguage intelligibility benefit. The authors even claim a non-matched interlanguage benefit. Here, a non-native speaker was more intelligible than the native American speaker even if the non-native listeners' L1 did not match that of the speaker. The magnitude of the benefit was not explicitly quantified in [5]. Therefore, we constructed table 4, which is analogous to our Table 1 above.

Table 4. Expected vowel identification scores (% correct transformed to rationalized arcsine units, RAU) on the basis of grand mean = 71% and main effects for Listener and Speaker L1 for each combination of factor levels. Further see Table 1.

	Language background of				Exp.	Obs.	Δ
	Listener		Speaker				
1.	Chinese	–9	Chinese	–5	57	64	**+7**
2.		–9	Korean	+4	66	66	0
3.		–9	American	+2	64	56	–8
4.	Korean	–6	Chinese	–5	60	60	0
5.		–6	Korean	+4	69	74	**+5**
6.		–6	American	+2	67	60	–7
7.	Other NN	–5	Chinese	–5	61	62	+1
8.		–5	Korean	+4	70	70	0
9.		–5	American	+2	68	67	–1
10.	American	+21	Chinese	–5	87	77	–10
11.		+21	Korean	+4	96	91	–5
12.		+21	American	+2	94	109	**+15**

There is a clear benefit of matched interlanguage between speaker and listener. Intelligibility is 7 RAU points better than predicted from linear combination of speaker and listener effects when speakers and listeners are both Chinese, 5 points better than expected when both are Korean and even 15 points better when both are American.

American listeners outperform all other types of listener in an absolute sense. In terms of our relative benefit measure, however, American listeners get much poorer scores than expected when listening to non-native talkers.

The American speaker is understood quite well, and much better than predicted when the listeners are fellow Americans, but the scores are poorer than predicted for L2 listeners. This could be called a native speaker handicap in the communication with a non-native listener.

Finally, our reanalysis reveals quite clearly that there is neither a benefit nor a handicap between non-native speakers and listeners if they have different native languages.

4. CONCLUSION

Analysis of our own data [6, 7] and a reanalysis of earlier data in [5] shows that the interlanguage speech intelligibility benefit is more insightfully quantified in relative than in absolute terms, as the residual, i.e. the prediction error, that remains after that the main effects of speaker and listener native language background have been included in a linear model.

5. REFERENCES

[1] Lane, H. (1963). Foreign accent and speech distortion. *J. Acoust. Soc. Am.* 35, 451–453.

[2] Munro, M. J. & Derwing, T. M. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Lang. Learning* 45, 73–97.

[3] Wijngaarden, S. J. van (2001). Intelligibility of native and non-native Dutch speech. *Speech Comm.* 35, 103–113.

[4] Nooteboom, S. G. & Truin, P. G. M. (1980). Word recognition from fragments of spoken words by native and non-native listeners. *IPO An. Progr. Rep.* 15, 42–47.

[5] Bent, T. & Bradlow, A.R. (2003). The interlanguage speech intelligibility benefit. *J. Acoust. Soc. Am.* 114, 1600–1610.

[6] Wang, H. & Heuven, V. J. van (2005). Mutual intelligibility of American Chinese and Dutch-accented English. *Proc. Interspeech 2005*, Lisbon, 2225–2228.

[7] Wang, H. (2007). *English as a lingua franca: Mutual intelligibility of Chinese, Dutch and American speakers of English.* LOT dissertation series 147. LOT, Utrecht. Also downloadable from http://www.lotpublications.nl/index3.html.