Methods and models for quantitative assessment of speech intelligibility in cross-language communication

Sander J. van Wijngaarden, Herman J.M. Steeneken and Tammo Houtgast

Department of Perception
TNO Human Factors, PO Box 23, 3769 ZG Soesterberg, The Netherlands
{vanWijngaarden, Steeneken, Houtgast}@tm.tno.nl

Abstract

To deal with the effects of nonnative speech communication on speech intelligibility, one must know the magnitude of these effects. To measure this magnitude, suitable test methods must be available. Many of the methods used in cross-language speech communication research are not very suitable for this, since these methods are designed to investigate specific effects regarding speech perception and production, rather than quantifying overall intelligibility. In this paper, a simple model of cross-language speech intelligibility is shown that helps in selecting experimental methods to assess speech intelligibility. Based on this model, and practical observations regarding assessment of cross-language speech intelligibility, a multi-lingual version of the Speech Reception Threshold method was implemented as a suitable method for the quantification of cross-language speech intelligibility. The performance of this method is illustrated by means of experimental results.

1. Introduction

Most reported experiments concerning nonnative speech intelligibility have been designed to obtain a better insight into the details of the speech perception and production process. Researchers in the field of second-language speech production and perception usually aim to test very specific hypotheses. Which experimental method is the most efficient depends on the tested hypothesis.

Apart from research on the basics of human speech communication an increasing need is felt for a more applied approach, aiming at the overall effect on speech intelligibility. Cross-language speech communication, in which one or more parties engaged in a conversation depend on second-language skills, is an increasingly common phenomenon. The efficiency of cross-language speech communication is quite often experienced to be lower than 'fully native' communication. For many of those situations, it would be helpful to be able to assess the magnitude of the effect on speech intelligibility. Applications that could benefit from such knowledge would be, for example, the design of public address and communications systems, and prediction models in room acoustics. By knowing the extent to which speech intelligibility is reduced, better design criteria can be established.

Wanting to know the *extent* to which speech intelligibility is influenced means that quantitative methods for measuring speech intelligibility are needed. This is different from the hypothesis-driven methodology preferred for investigating the principles of nonnative speech communication; instead of looking for *reasons,* we are quantifying the *consequences.*

To illustrate this approach, consider the following situation. Suppose that an auditorium in a Dutch school is equipped with an air-conditioning system, which produces a known level of background noise. In 'normal' (native) situations, the intelligibility of the public address system in the auditorium is generally acceptable, despite the background noise. What if a native English talker addresses the Dutch students (in English), who have an average experience with the English language of 2 years? What if the average experience of the students is 5 years, or what if the native language of the talker is German? What reduction of the background noise level is necessary to obtain a certain minimum speech intelligibility?

When using suitable methods, it is possible to answer all these questions, if populations of talkers and listeners are properly defined. Not all of the *reasons* behind the differences in intelligibility have to be known. These reasons may be very complex, involving better analysis of the speech signal into phonetic units, larger vocabulary, better understanding of the grammar, etc. Regardless of the reasons, the effects are interesting enough in their own right.

In this paper, we will present a simplified model of nonnative speech communication. The aim of this model is to serve as a tool, which helps in choosing the proper methods to quantify the effects on intelligibility. Based on this model, we will describe a multi-lingual speech intelligibility evaluation method that is suitable for application to cross-language speech communication

2. Model of cross-language speech communication

2.1. Types of cross-language speech communication

Describing a specific cross-language conversation unambiguously takes a little consideration. As the number of people engaged in a conversation increases, the complexity of a proper description of the situation increases accordingly.

All situations can be broken down into variants of straightforward two-way communication, in which case only one person is talking, and only one other person is listening. This involves influences from up to three languages: the native language of the talker, the native language of the listener, and the language that is currently being spoken. The relations between these three languages will partly determine the speech communication process. Comparative studies of the involved languages could theoretically shed light on this; analyses of the existence of phonetic contrasts and inspection of the (sound-based) lexicon of a specific language could help understand its relation with other languages, provided this same information is also known for these other languages. Rather than trying to find a general model for language-related influences on cross-language communication, we will treat each combination of languages as a unique case.

It has become convention to denote native talkers and listeners as 'L1', and nonnative (second-language) talkers and listeners as 'L2'. Based on this notation, one could (for example) indicate that a native listener is listening to a non-native talker by writing 'L2>L1'. This notation works if the number of languages involved is no more than two. The situation 'L2>L2' could mean that a Dutch listener is speaking English to a German listener; it could also mean that a Dutch listener is speaking English to another Dutch listener. The difference may be important, since the common native language between talker and listener may influence their use of the second language (in our example English).

To avoid confusion, we will use the following notation throughout his paper:

Dutch > (English) > German

meaning that a Dutch talker is talking English to a German listener. We will generally abbreviate this to D>(E)>G.

2.2. Defining populations of talkers and listeners

Considering nonnative speech intelligibility separately for each individual that comes our way would become a very laborious process. By defining meaningful populations of talkers and listeners, we can collect more generally applicable quantitative results. First, we decide what populations we need to have quantitative data on; then we recruit subjects from these populations, and carry out experiments. Experiments may involve subjects selected from one single population, or may use talkers from one population and listeners from another.

In order to define a population, one should be able to describe it in terms of the determining factors for nonnative speech intelligibility. The description of the population starts with the native language of the subjects; preferably, details concerning regional accents (if any) should also be known.

A very important factor is the average experience of subjects within the population with the target (second) language (eg. ([1,2]). Age of acquisition of the second language is also of great importance. (eg. [3,4,5]).

Second-language experience and age of acquisition combine into second language *proficiency*, a term we will use rather loosely to indicate the underlying dimension explaining differences in nonnative speech intelligibility. Despite the fact that second-language proficiency comprises many different abilities (related to phonetic discrimination, vocabulary, grammar, etc.), subjects are able to rate their own proficiency with a sometimes impressive accuracy [6].

Possible other factors to consider could be more general descriptors of the population, such as age and gender. It seems fair to consider the influence of these variables on cross-language communication higher-order effects, but it is only prudent to keep variables like these in mind as well when selecting subjects for experiments.

Even when the populations of talkers and listeners are fully defined, the resulting speech intelligibility may still vary according to numerous other variables, most of which also apply to fully *native* communication, such as speaking rate and speaking style. These variables are not really related to the characteristics of the talkers and listeners, but rather to their *mode* of communication. One aspect related to this is worth mentioning. For nonnative talkers, the distinction between *read* speech and *spontaneous* speech is potentially of far greater importance than for native talkers. Nonnative talkers are likely to limit their effective vocabulary to easier and more familiar words when speaking spontaneously, while they are more likely to produce pronunciation errors when asked to read a certain text aloud. In the latter case, they are not only likely to mispronounce unfamiliar words, but a poor understanding of context may also lead to an impaired intonation of sentences.

2.3. Conditions for speech communication

Native as well as nonnative speech can be affected by adverse conditions, such as background babble, ambient noise, bandwidth limiting, or reverberation. However, the degrading influence on cross-language speech communication tends to be greater [5,7,8,9,10].

Measuring speech intelligibility under clear, undegraded, conditions is often not very effective. The effects of nonnativeness on intelligibility may be relatively small, whereas problems in practice are expected when degrading circumstances *are* present. By conducting experiments under conditions that represent a controlled degree of speech signal degradation, the effect of this degradation on cross-language speech communication may be assessed systematically.

Perhaps the easiest way to reduce speech intelligibility in a controlled manner, is by adding stationary noise with a known spectrum. For fully native speech communication, intelligibility in this case is a relatively stable and well-known function of the speech-to-noise ratio. For nonnative speech communication similar relations are found [10,11], which clearly show that noise is capable of affecting cross-language communication more profoundly than native speech communication.

2.4. Levels of analysis

Our approach towards the assessment of nonnative speech intelligibility needs a model that describes cross-language speech communication in such a way, that the proper characteristics to quantify intelligibility can be chosen.

In practice, this means that a description is needed of the determining factors for speech intelligibility (which we will call intelligibility cues), and an indication of where to find these. More specifically, we need to find out about intelligibility cues that are especially important when considering *cross-language* speech communication.

Speech intelligibility can be studied at various levels of analysis; the most basic analysis would involve studying the speech signal on an allophone-by-allophone basis. Perhaps the highest thinkable level would be to consider an entire story, where the amount of relevant information in the story that was transferred could be studied.

There are reasons to assume that the level of individual words takes an important position in the process of learning a second language [12]; it seems likely that one initially learns a second language mainly by collecting a sound-based representation of its lexicon. For this reason, and because of practical considerations, we will distinguish three levels of analysis: speech units smaller than words (allophones), words, and speech units larger than words (sentences).

Besides the level of analysis, intelligibility cues can also be separated depending on whether they can be found in the speech signal ('acoustic' cues) or somewhere else. As an example of the difference: the intelligibility of sentences (as compared to the intelligibility of the individual words of which they consist) is enhanced by means of intonation. Intonation (or more generally, prosody) is present in the speech signal, and can therefore be called an 'acoustic' intelligibility-enhancing factor. The semantic and syntactic redundancy contained in a sentence also increases its intelligibility relative to the individual words of which it consists. However, these factors can not be traced back to the speech signal; they improve intelligibility by aiding the listener in his cognitive processing of the message.

Table 1 illustrates the distinction between acoustic and non-acoustic intelligibility cues at the three defined levels of analysis.

Table I. Levels of analysis in nonnative speech communication

Level of analysis	Examples of affected intelligibility cues	
	Acoustic	Non-acoustic
Supra-word level (sentence level)	Prosody	Syntactic constraints Semantic constraints
Word level	Lexical dissimilarity	Word familiarity
Sub-word level (allophone level)	Phoneme inventory	

This distinction between acoustic and non-acoustic factors is not helpful at the sub-word level. For the non-acoustic factors at this level (such as the individual phoneme space representation that a listener uses to categorize L2 allophones) can hardly be tested without involving acoustic allophone realizations.

Table 1 can be used to decide which *characteristic* of cross-language speech intelligibility is the most appropriate in a specific case, for instance phoneme recognition versus sentence intelligibility. Only *after* deciding which is the most appropriate characteristic can we design a proper experiment.

For example, one may wish to quantify the intelligibility of a group of (nonnative) German actors, playing before an audience of native English listeners, in the English language (G>(E)>E). The non-acoustic intelligibility cues do not require special attention in this case, since only the talkers are nonnative, and their vocabulary and sentence construction are 'programmed' by the play they are acting out. Hence, all deviations from fully native communications can be found in the speech signal. At the very least, one may expect that the actors' allophone realizations will deviate from native English speech. A phoneme-based intelligibility test will be a suitable choice to quantify this effect. However, this may not be the *most* suitable intelligibility test. Unless the actors are thoroughly trained by a native English director or language coach, their intonation will also deviate from the authentic English patterns. In that case, a (sentence-based) intelligibility test that is sensitive to differences in prosody is a better choice.

As another example, consider the reverse situation (the actors are now English and the audience is German; E>(E)>G). Since the German audience is now the only nonnative factor, the speech signal is not at all affected. Still, the resulting speech intelligibility may be reduced considerably; partly because the nonnative listeners are not as good at identifying individual speech sounds, but also for reasons related to vocabulary and the less effective use of word context [11]. In this case, the average L2 linguistic development of the German audience is an important variable. Besides a speech intelligibility test using sentences (to include the effects of word context), it may be useful to include a separate test to quantify vocabulary and context-effects separately.

3. Speech intelligibility assessment methods

3.1. Practical considerations

A pragmatic approach toward measuring nonnative speech intelligibility is simply to adopt one of many proven experimental methods designed for native speech. Inevitably, some modifications to these proven methods will be necessary, if only for practical reasons.

Several intelligibility test methods are based on one-syllable nonsense words. These tests are generally quite efficient at measuring speech intelligibility phoneme level. Subjects participating in such tests must somehow communicate perceived nonsense syllables in response to the auditory stimuli. With L2 listeners, typing these responses should be ruled out as an option. Differences in orthographic representations of sounds between L1 and L2 will confuse the subject. Even highly proficient subjects, who are aware of differences in orthography between L1 and L2, are likely to produce errors, especially when working under time pressure. Collecting multiple-choice responses will partly solve this problem, especially if no 'confusing' alternatives are presented. In any case, proper subject instruction with regard to this issue is vital.

Some additional complications surrounding experiments with non-natives have to do with the recruiting of subjects. The definition of the population from which to draw subjects is much narrower than usual in speech intelligibility testing. Accordingly, subjects will be harder to find. Experimental methods can be designed or adapted to help cope with this issue. Methods that require special sound-insulated rooms or heavy equipment require subjects to travel to a certain location. By adapting these methods so that they can be implemented in a portable device (such as a notebook computer) hard-to-reach subjects (unwilling to travel in order to take part in a test) can be tested at remote locations.

The available time per subject may also be shortened. When tests run over longer periods of time, a smaller percentage of the population of potential subjects will be willing to participate. By shortening the duration of the experiment (by making tests more efficient, or by spreading the load over a slightly larger number of subjects) the number of available subjects may be increased.

3.2. Types of speech stimuli

Various types of speech stimuli are used in speech intelligibility tests. Generally, the length of each single stimulus determines which level of analysis (table I) is addressed by the test method.

The most fitting speech stimuli corresponding to the different levels indicated in table I would appear to be sentences, words and phonemes. However, individual phonemes are hard to test without the context of a word or syllable; hence the frequent use of nonsense syllables that was mentioned in the previous section. The individual recognition of phonemes is also difficult to test using *meaningful* words, since the word context will be of some influence on the probability of correct recognition.

Higher-then-word level effects are expected for most thinkable cross-language conversations. In principle, sentence intelligibility tests also include effects at lower (word and phoneme) levels, since all sentences are constructed from these smaller units of speech. If only one type of speech stimuli can be chosen, it makes sense to choose sentences. On the other hand, it should be noted that (nonsense) word tests will be more sensitive to effects at lower levels of analysis.

When comparing native and nonnative talkers, specific choices must be made before recording any speech stimuli. Speaking rate and speaking style are likely to vary between native and nonnative talkers. Nonnative talkers usually tend to (consciously or unconsciously) compensate for the effects of their accent on intelligibility by adjusting their speaking rate or speaking style [6]. This is a legitimate effect, which can also be observed in cross-language conversations in practice – it is in some ways similar to the Lombard-effect, which lets talkers automatically increase their vocal effort in the presence of background noise. One may choose to include this effect in the test, or force native and nonnative talkers into similar speaking styles (by giving suitable instructions, monitoring recordings, and pacing their speaking rate).

3.3. Multi-lingual test methods

One step further than nonnative speech intelligibility testing is multi-lingual intelligibility testing. Multi-lingual tests can involve either native or nonnative subjects, but must also be implemented in multiple languages. Obtaining equivalent implementations of the same test in various languages poses an additional difficulty. True equivalence across languages is hard to reach.

Whatever speech stimuli are used, these stimuli must somehow be matched across languages. When working with phoneme tests, the tested phonemes could be balanced to represent the mean frequency of occurrence in the corresponding language. Despite the fact that different phoneme sets must be tested for each language, these are equivalent in the sense that they represent a 'natural' distribution of phonemes for each languages.

When the test stimuli are isolated words, then on top of phonetic balancing the frequency distribution of the test vocabulary (measured frequencies of occurrence in representative texts) should be controlled. Where available, the appropriate information could be taken from (multi-lingual) lexical databases.

When using sentences, the main things that should be matched across sentences are the complexity of the sentences, and the *domain* from which the sentences are taken. The source of the sentences largely determines the domain (newspaper, radio, everyday conversation, etc.), making this variable relatively easy to control. The complexity can be controlled by adopting certain constraints for the selection of sentences; at least the length (number of syllables) of the sentences, and the length of the individual words in the sentences, should match pre-defined criteria.

When sentences are properly selected, phonetic balancing becomes of lesser importance. Each sentence consists of a certain mix of phonemes; when each condition is tested with multiple sentences, there is a more or less implicit phonetic balancing for the domain from which the sentences are taken.

An additional complicating factor when designing multi-lingual tests is the fact that the relative importance of different levels of analysis (table I) may vary between languages. Phoneme identification may be more difficult in some languages than others, simply because the number of existing phonemes differs (eg. English vowels versus Spanish vowels). Contextual information that is available in one language, for instance by the use of case and word gender, may be absent in other languages.

A pragmatic approach to the design of multi-lingual test is to simply try out the implementations in different languages on native subjects. If the native scores are the same across languages, then it seems fair to assume that the method performs equivalently.

3.4. Multi-lingual Speech Reception Threshold method

An example of a multi-lingual implementation of an existing intelligibility test method is the multi-lingual Speech Reception Threshold (SRT) method. The SRT method is widely used as a diagnostic tool in the field of audiology [13], and has been proven useful to evaluate speech intelligibility of talkers, listeners, and communication systems.

3.4.1. Test procedure

The SRT test gives a robust measure for sentence intelligibility in noise, corresponding to the speech-to-noise ratio that gives 50% correct response of short redundant sentences. In the SRT testing procedure, masking noise is added to test sentences in order to obtain speech at a known speech-to-noise ratio. The masking noise spectrum is equal to the long-term average spectrum of the test sentences. After presentation of each sentence, the subject responds by orally repeating the sentence to an experimenter. The experimenter compares the response with the actual sentence. If every word in the responded sentence is correct, the noise level for the next sentence is increased by 2 dB; after an incorrect response, the noise level is decreased by 2 dB. The first sentence of a list of 13 sentences is repeated until it is responded correctly, using 4 dB steps. This is done to quickly converge to the 50% intelligibility threshold. By taking the average speech-to-noise ratio over the last 10 sentences, the 50% sentence intelligibility threshold (SRT) is obtained.

3.4.2. Interpretation of SRT results

The score resulting from an SRT test ('the SRT' for the corresponding condition) is a speech-to-noise ratio (SNR); at this SNR, 50% of the sentences are repeated correctly by the listeners. At better (higher) SNRs, more than 50% will be intelligible, at more adverse (lower) SNRs, less than 50%. A lower SRT means better intelligibility: more noise can be allowed to reach 50% recognition of sentences.

The percentage of correctly recognized sentences is a (psychometric) function of SNR, often modeled as a cumulative normal distribution. The SRT is the adaptively estimated mean of this distribution, which is the best single parameter to characterize the whole curve. A logical second parameter to estimate would be the variance of the distribution, reflected by the slope of the psychometric curve. To estimate this slope (or even the full psychometric curve), one could use alternative testing paradigms using the same

SRT sentences. The description of such methods is beyond the scope of this paper.

3.4.3. Creating a multi-lingual version

The 'original' [13] Dutch SRT sentences describe common, everyday situations in simple wording. Based on these original sentences, the following constraints were defined for 'translation' of the sentence material:

- Sentence length 7-9 syllables
- No words longer than 3 syllables
- No more than one three-syllable word per sentence.
- Sentence content is of an everyday life nature
- Sentences of approximately equal redundancy (or predictability, perplexity) as the original sentences

3.4.4. Software implementation

A computer program was developed for maintaining multi-lingual databases of recorded SRT sentences and using these in intelligibility tests. This program also features a module for recording new material. In combination with a notebook computer and a high-quality sound card, a small, flexible setup is create which can be used to record and test talkers nonnative and listeners at any location that is sufficiently silent.

3.4.5. Speech recordings

Traditionally, talkers used in SRT tests for audiological purposes are trained professionals, speaking very clearly. The SRT scores obtained with these recordings are hard to reach for most ordinary talkers, especially under representative conditions.
Multi-lingual SRT talkers are not selected according to a strict regime, or following specific criteria. The talkers are simply verified not to exhibit any speaking disorders, and instructed to speak with a clear 'reading voice'. This makes it easier to recruit talkers, and quickly build up speech databases.
To prevent large differences in speaking rate, the speaking rate is paced by means of a 'progress bar'. Talkers have to pronounce each sentence within a 2.5-second timeframe, which is visually indicated on the computer screen.

3.4.6. Applications of the multi-lingual SRT

It should be noted that the application of the SRT method (and similar methods) to cross-language research is not new (eg. [5,14]). What is new about our current multi-lingual SRT implementation, is the effort to construct a coherent test in as many languages as possible. At the moment this paper was written, 'translations' of the sentences (text) were available in at least 8 different languages; a multi-speaker test speech database had been collected for at least 5 of these languages.
Sofar, the English, German and Dutch versions of the test were succesfully used to quantify cross-language speech intelligibility [6,11]. Apart from this, the English, French, German and Dutch version were used with solely native subjects (talkers and listeners) to measure the language dependency of voice coding systems [15].

4. Examples of experimental data

4.1. Nonnative listeners

For a population of Dutch university students, cross-language intelligibility-effects (in terms of SRT) were measured when listening to English and German [11]. Almost all Dutch university students have been taught English and German during secondary education, German at a slightly later age and for a shorter period than English. Also because of the more frequent use of English (university classes, textbooks, television and other media) the L2 proficiency tends to be much higher in English than in German. Figure 1 shows native and non-native SRT results related to this population.

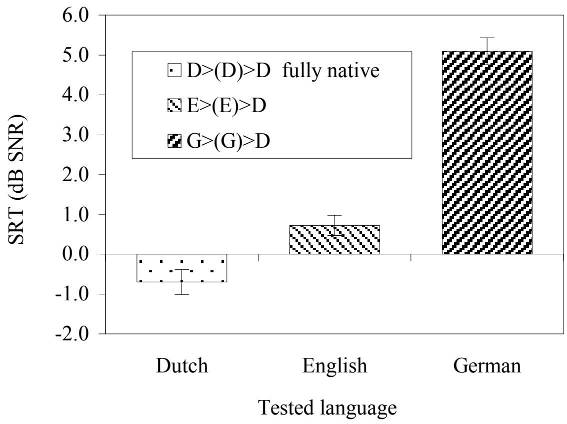

Figure 1. Mean SRT scores and standard errors measured for 9 Dutch university students when listening to three languages (3 talkers per language, N=27). (underlying data previously published, [11]).

The difference between the effects of listening to English and German is considerable; all differences in figure 1 are statistically significant. Despite the fact that the listeners were selected to be highly proficient in English, the effect of being nonnative listeners on the resulting intelligibility is clearly noticeable.

4.2. Nonnative talkers

Similar results as presented in figure 1 can be obtained for nonnative talkers. In that case, the population of listeners consists of 'average natives', and the talkers are recruited to match a certain desired profile.
Figure 2 shows results of in experiment aimed at measuring the effect of perceived foreign accent on intelligibility. For this experiment, 15 talkers were recruited who could all speak Dutch, differing in degree of foreign accent. These talkers were from 5 language backgrounds: Dutch, English, German, Polish and Chinese.
To measure the 'degree of perceived accent', a pairwise comparison experiment was conducted with native Dutch listeners. From this experiment, subjective foreign accent ratings were calculated. The relation between SRT scores and these ratings are shown in figure 2.

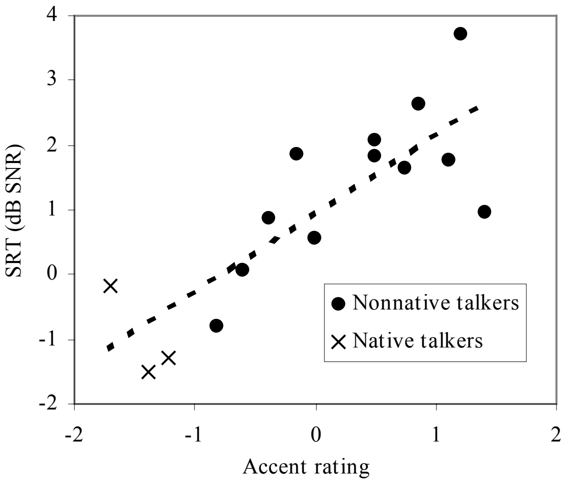

Figure 2. Relation between subjective accent ratings (N=39) and SRT scores (10 native listeners for each data point). R^2=0.68. Data previously presented [6].

Following from the correlation in figure 2, 68% of the variance in SRT scores could be explained by the perceived accent ratings.

4.3. Multi-lingual comparison

For some goals, multi-lingual speech intelligibility tests are useful even when no cross-language factors are directly involved. For example, to measure the language dependency of vocoders, one needs to test (native) speech intelligibility in a number of languages. The performance of a multi-lingual test should be closely matched across languages, otherwise the language dependency of the tested vocoders will be confounded with the language dependency of the test method [15]. The easiest way to verify if the results are sufficiently closely matched across languages, is by measuring the same (relatively undistorted) conditions in several languages. Results of such an experiment are shown in table II.

Table II. Mean native SRT scores and standard errors for four languages (3 talkers per language, 10 listeners). All speech was bandwidth limited (50-4000 Hz).

SRT	English	French	German	Dutch
mean	0.7 dB	1.0 dB	0.3 dB	0.4 dB
S.E. (*N*=10)	0.4 dB	0.8 dB	0.4 dB	0.5 dB

The results are closely matched (also note the magnitude of the effects shown in figures 1 and 2). None of the differences between languages are statistically significant.

5. Discussion and conclusion

The pragmatic model of cross-language speech communication presented in this paper was used to select the multi-lingual SRT method as a suitable tool for measuring nonnative speech intelligibility. As the examples in section 4 show, the method is effective in collecting quantitative data for nonnative talkers as well as listeners. The coherent performance across-languages makes the method suitable for various multi-lingual applications.

6. References

[1] Flege, J.E. "The intelligibility of English vowels spoken by British and Dutch talkers," In *Intelligibility in Speech Disorders,* edited by R.D. Kent, John Benjamins publishing company, Amsterdam, 1992.

[2] Strange, W. "Cross-Language Studies of Speech Perception; A historical review," In *Speech Perception and Linguistic Experience,* edited by W. Strange, York Press, Baltimore, 1995.

[3] Flege, J.E. "Second-language speech learning: theory, findings, and problems," In *Speech Perception and Linguistic Experience,* edited by W. Strange, York Press, Baltimore, 1995.

[4] Flege, J.E., Bohn, O-S., and Jang, S. "Effects of experience on nonnative speakers' production and perception of English vowels," *J. Phonetics*, Vol 25: 437-470, 1997.

[5] Mayo, L.H., Florentine, M. and Buus, S. "Age of second-language acquisition and perception of Speech in noise," *J. Speech Lang. Hear.Res.* Vol 40, 686-693, 1997.

[6] Wijngaarden, S.J. van, Steeneken, H.J.M. and Houtgast, T. (2001). The effect of a non-native accent in Dutch on speech intelligibility. *J. Acoust. Soc. Am.* Vol. 109 (5) Pt 2: page 2473, 2001.

[7] Lane, H. "Foreign accent and speech distortion," *J. Acoust. Soc. Am*. Vol. 35: 451-453, 1963.

[8] Gat, I.N. and Keith, R.W. "An Effect of Linguistic Experience; Auditory Word Discrimination by Native and Nonnative Speakers of English," *Audiology* Vol. 17: 339-345,1978.

[9] Náběl̆ek, A.K. and Donahue, A.M. "Perception of consonants in reverberation by native and nonnative listeners," *J. Acoust. Soc. Am*. Vol. 75:632-634, 1984

[10] Wijngaarden, S.J. van. "Speech intelligibility of native and non-native Dutch speech". *Speech Communication*, Vol. 35:103-113, 2001.

[11] Wijngaarden, S.J. van and Steeneken, H.J.M. "The Intelligibility of German and English Speech to Dutch Listeners". *Proceedings of the International Conference on Spoken Language Processing (ICSLP2000),* 2000.

[12] Bradlow, A.R. and Pisoni, D.B. "Recognition of spoken words by native and nonnative listeners: talker-, listener- and item-related factors," *J. Acoust. Soc. Am*. Vol. 106: 2074-2085, 1999.

[13] Plomp, R. and Mimpen, A.M. "Improving the Reliability of Testing the Speech Reception Threshold for Sentences", *Audiology*, Vol. 18: 43-52, 1979.

[14] Buus, S., Florentine, M., Scharf, B. and Canevet, G. "Native, French listeners' perception of American-English in noise." *Proceedings of Internoise 86*, 1986, pages. 895-898.

[15] Wijngaarden, S.J. van and Steeneken, H.J.M. "A Proposed Method for Measuring Language Dependency of Narrow Band Voice Coders". *Proceedings Eurospeech 2001, Aalborg, Denmark,* 2001.

[16] Wijngaarden, S.J. van. *Multilingual SRT 2.41. Database acquisition and speech reception threshold measuring program*. Soesterberg, TNO Human Factors, 2001.