MEASURING THE SPEECH TRANSMISSION INDEX OF SYSTEMS FEATURING DIGITAL VOICE CODING

Jan A Verhave
Margriet Vlot
Sander J van Wijngaarden

Embedded Acoustics BV, Delft, The Netherlands
{jan,margriet,sander}@embeddedacoustics.com

1 INTRODUCTION

Digital voice coding algorithms, commonly used in private mobile radio systems, utilise speech production models to digitally compress speech signals at very low bitrates. The impact on speech intelligibility is substantial, but not easy to measure. Low-bitrate voice coding systems do not process acoustic test signals in a representative manner. In particular signals for measuring the Speech Transmission Index (such as the STIPA signal) can rarely be used. These signals consist of intensity-modulated noise carriers which do not match the voice patterns that low-bitrate voice coders are trained to reproduce. Because of the noise-like nature of the signals, they are often even suppressed actively.

This paper explores the potential of an alternative category of STIPA test signals to circumvent the issues described above. Instead of a noise carrier, these signals use sine complexes that are more easily recognized (and reproduced) by narrow-band voice coders. These signals are tested with TETRA (Terrestrial Trunked Radio) devices featuring the ACELP voice coder (ETSI-EN-300-395-1, 2004), and DMR (Digital Mobile Radio) devices featuring AMBE+2 coding (ETSI-EN-102-361-2).

The main challenge is to validate if the Speech Transmission Index results obtained with the new test signals are indeed accurate predictors of speech intelligibility. Inevitable, subjective listening tests (with human participants) are needed for such a validation effort. These listening tests are done with the Matrix Response test.

2 NOVEL STIPA SIGNALS AND TEST SETUP

2.1 Test signals with a deterministic carrier

To be able to utilize standard STIPA measuring devices in combination with a transmission channel that features speech coding, we needed to develop a new carrier signal that mimics the natural speech as much as possible without interfering with the modulation information. Within the direct STI method, modulated pink noise carriers are generally applied to measure the modulation transfer function between talkers and listeners. It was already recognized in earlier research that for some situation - such as algorithm testing - noise free modulated sine carriers could be used as an STI test signal. Similar test signals will be proposed in upcoming STI standards to test STI measuring algorithms. With the above in mind, we developed a new STIPA test signal for which the carrier signal mimics natural speech as much as possible, in combination with the modulation scheme as used by STIPA. Knowing that the energy carrying parts of speech mainly consist of tonal parts (vowels) and noisy parts (plosives and fricatives), a carrier signal was designed that combines formant spectra and pink noise bursts.
The new STIPA test signal for voice coders (which we will call STIPA-VC) was divided into 50 milliseconds time slots of which 30 % were randomly chosen to become noise carriers and 70 % formant carriers. As formant carrier signals, two individual ones with a fundamental frequency 10 Hz apart (120 and 130 Hz) were randomly chosen 50 % of the time. The formant harmonics were fitted within the ½ octave bandwidth at each octave band. Care was taken that equal energy was obtained per octave band for the different carrier types. The new carrier signal was subsequently modulated according to STIPA. Figure 1 shows the measured STI

value with the new carrier signal for different signal to noise ratios compared with the theoretical value using a generic STIPA measuring device. As shown, the results are in close agreement.

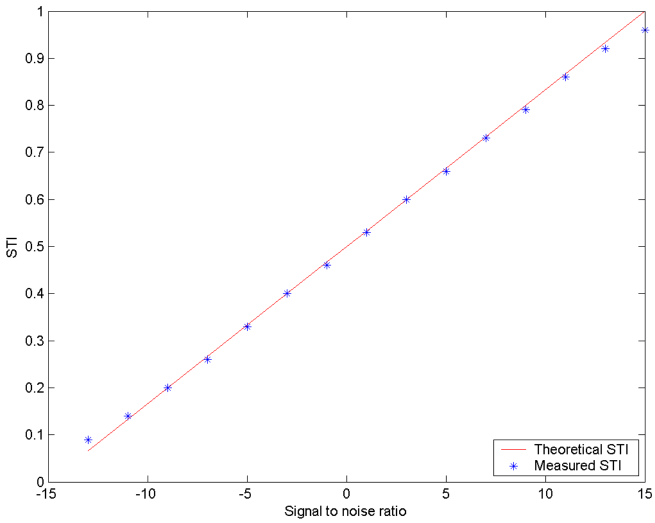

Figure 1. Measured STI values with the new carrier signal for different signal to noise ratios compared with the theoretical value

2.2 Tested TETRA and DMR systems

Two radios were used: one TETRA radio and one DMR radio. So as not to reveal information about third-party products that potentially may be commercially sensitive, we call these radios T1 and D1 for the TETRA radio and the DMR radio, respectively.

Both radios are products that are currently operationally used with first responders and police forces throughout Europe.

2.3 Experimental setup for STIPA measurements

The microphone of the radio was placed directly in front of the artificial mouth of a Head Acoustics HMSII.3 head-and-torso simulator. The distance between the microphone and the mouth reference plane was kept fixed throughout the tests at 50mm.

Before playback through the artificial mouth, the (modified) STIPA test signal was mixed with male speech noise noise at fixed signal-to-noise ratios. This way, different ambient noise situations were simulated with the need for a tightly controlled acoustic environment.

The frequency transfer function of artificial mouth was calibrated to yield a flat frequency transfer function for all frequencies between 80 Hz and 12.5 kHz, measured in 1/3 octave bands, accepting differences of no more than $\pm$ 1 dB.

3 VALIDATION METHOD

The proposed modified STIPA method differs from the standardized STIPA method in only one way: the carrier signal is different. A deterministic sinusoidal complex, interleaved with noise bursts, is used instead of continuous noise. Voice coders are “tricked” into treating the signal as speech.

This means that there is no need to completely revalidate the STIPA framework, since all parameters (frequency weightings, masking factors, etc) remain unchanged. All commercially available test equipment for STIPA will even work with the modified signal. In fact, it is fair to say that the original choice to use noise as a carrier signal for STI signals was, in part, arbitrary. The main reason for choosing noise as a carrier signal is that it offers a high spectral density, without favouring any frequency (within an octave band) of any other frequency. As longs as we restrict ourselves to use cases where this is acceptable, the results should not be affected by the change in carrier structure. In practice, this means that we need to avoid transmission channels in which the dominant causes of intelligibility reduction are related to room acoustics (reverberation/echoes), or in which we expect highly irregular frequency transfer functions and an interaction with the carrier signal.

In terms of validation, we then simply need to demonstrate that the measured STI values correlate well with subjective speech intelligibility of voice coding systems, yielding results similar to the correlation curves on which the original STI validation study was founded.

The most efficient way to evaluate speech intelligibility across a variety of test conditions and systems is through an adaptive sentence-based procedure, such as the Speech Reception Threshold (SRT) [1] test. The SRT is the signal-to-noise ratio at which (on average) 50% of sentences is correctly repeated by test subjects.

A modern variant of this test procedure is the Matrix test (also known as the OLSA test [2]). We chose the Matrix Test as our test procedure for the experiments described here.

3.1 Subjective test method

The matrix sentence test was originally developed by Hagerman in 1982, in Swedish [2]. It consists of short, grammatically correct but semantically unpredictable sentences. Every sentence consists of five words, each from a different category: name, verb, number, adjective and noun. For example (in Dutch): *Jan koopt vijf mooie bloemen* (Jan buys five beautiful flowers). A matrix has been constructed with 10 words for every one of the five categories: fifty words in total. For every sentence, the noise level is kept constant while the speech level is varied, through which the test aims to find the SRT. While developing the matrix test, co-articulation was taken into account. All words are recorded in a way that the last sound of one word already sounds at the beginning of the second word. This results in fluent, natural-sounding speech. The advantage of the matrix test is that it can be used multiple times with the same subject, since there is hardly any learning effect. Besides that, it makes use of sentences as we encounter in daily life. A drawback can be that the test is not very satisfying for subjects. This is due to the fact that the speech reception threshold (at 50% intelligibility) is searched, which results in hearing a lot of noise and not so much speech from the subject’s perspective. Moreover, it takes some time before the subject is familiar with the method of the test. The Dutch version of the matrix test is only available with female speech. We chose the Dutch-language version of the test for practical reasons (mainly the fact that native Dutch listeners are the easiest to find). The Dutch matrix test is composed according to Table I.

Table I. Structure of the Dutch-language Matrix test

	Name	**Verb**	**Number**	**Adjective**	**Name**
1	Anneke	wint	drie	grote	dozen
2	Monique	geeft	negen	nieuwe	schoenen
3	Sarah	kiest	acht	vuile	boeken
4	Christien	koopt	vier	kleine	munten
5	Heleen	tekent	twee	mooie	stenen
6	Jan	vroeg	vijf	goede	ringen
7	Pieter	vond	tien	zware	boten
8	Mark	telde	twaalf	dure	messen
9	Willem	maakte	achttien	oranje	bloemen
10	Tom	had	zes	groene	fietsen

The words used were selected as a cross-section of ordinary, every day Dutch speech. It was ensured that the occurrence of the phonemes in the matrix test mirrored that of standard Dutch. Even though in theory there are very many possible combinations (105), only 340 sentences are available. These sentences were selected in order to yield a balanced test result across different word transitions. In developing the Dutch matrix test, the sentences were also balanced in terms of word intelligibility. The sound volume of the words was sometimes attenuated, based on their individual SRT, in order to achieve similar intelligibility. The Dutch version of the matrix test has an SRT of (−8.4±0.7) dB as shown by Houben et al. [3]

3.2 Participants

Based on a power analysis on data from a pilot experiment with 5 participants, we determined that a minimum of 8 subjects was sufficient (accepting a 1% standard error on SRT results). A total of 13 participants took place, all students from a local university, who were not paid for their participation but given a €25 gift card as an incentive to take part. All participants were informally screened to have normal hearing.

3.3 Test conditions and reference conditions

By the vary nature of SRT-like methods such as the Matrix tests, all stimuli presented to the participants are speech mixed with noise at a predetermined speech-to-noise ratio.

Speech was processed through the digital radios in the same way as the modified STIPA signals, as described above. Speech mixed with noise was played back through the artificial mouth of a Head Acoustics HMSII.3 headform, transmitted through the digital radio, and recorded for use in the experiment. All speech processing was done in the form of long concatenated speech files, which were afterwards split into individual test sentences.

In addition to the speech processed through the radios, conditions were also recorded for which no radio channel was present. Those conditions are simply speech mixed with noise, at different signal-to-noise ratios. These were used as reference conditions.

4 RESULTS

4.1 Subjective test results

Using speech material prepared at different signal-to-noise ratios, the matrix intelligibility score was determined. Figure 2 shows the average matrix scores for clear speech at signal-to-noise ratios between -13 and -1 dB. Using these scores, a logistic function was fitted with a minimum score chance of 10 %. This results in an SRT of -9.6 dB with a 12.9 % slope per dB.

Figure 3 shows similar test results but now with telephone bandwidth reduction (300 – 3600 Hz). Bandwidth reduction will result in an SRT of -8.9 dB and a less steeper slope of approximately 10 % per dB. Figures 4 and 5 shows the Matrix results for a Tetra and and an DMR radio. The Tetra radio has an SNR of -4.8 dB with a slope of 7.6 % while the DMR radio has an SRT of 0.2 dB. These results shows that for both digital radios, an increase in SNR of respectively 5 and 10 dB is needed to obtain similar objective speech intelligibility. Clearly a difference in SRT between Tetra and DMR of 5 dB is measured.

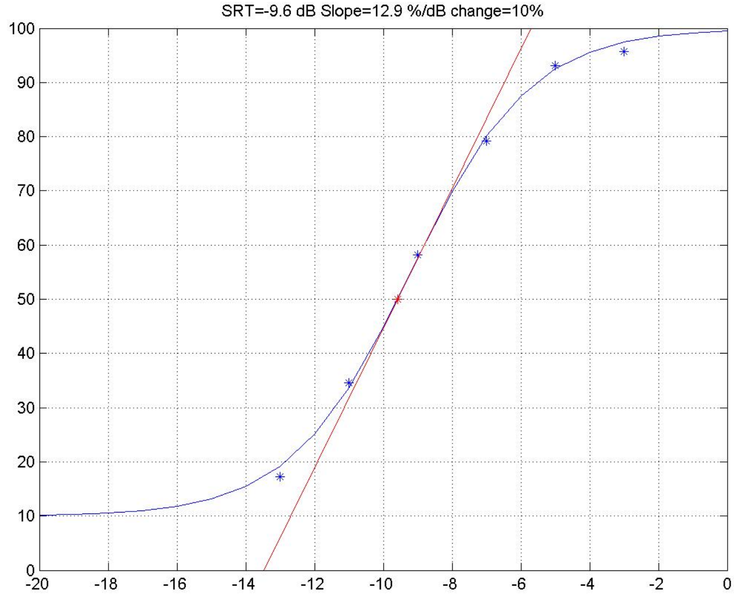

Figure 2. Matrix test results of speech in combination with speech noise at different signal-to-noise ratios

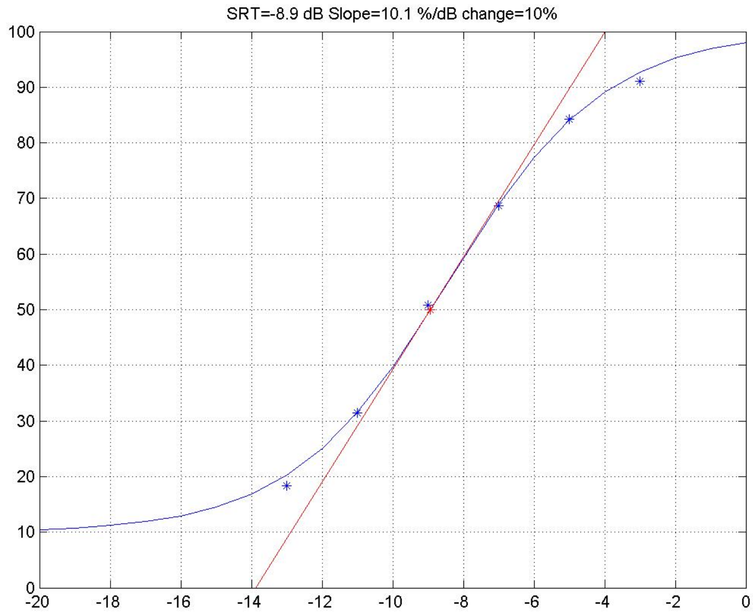

Figure 3. Matrix test results of bandwidth limited speech in combination with speech noise at different signal-to-noise ratios

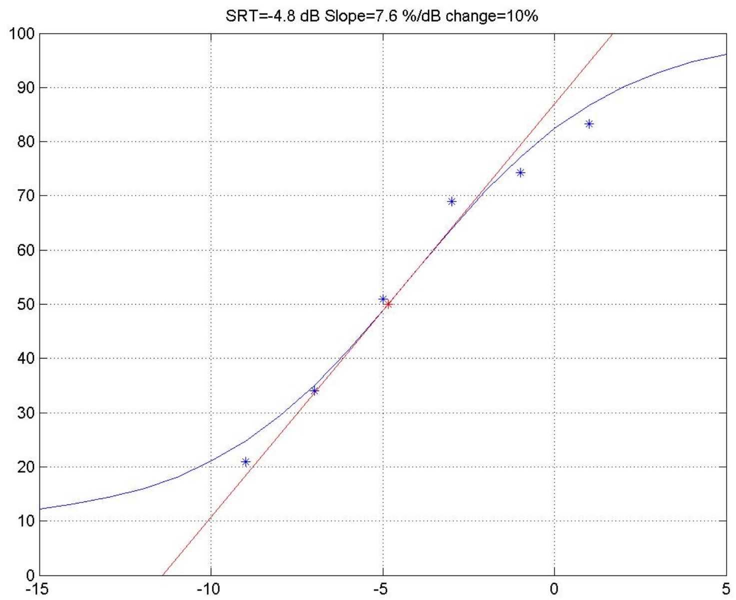

Figure 4. Matrix test results of speech in combination with speech noise at different signal-to-noise ratios for Tetra radio.

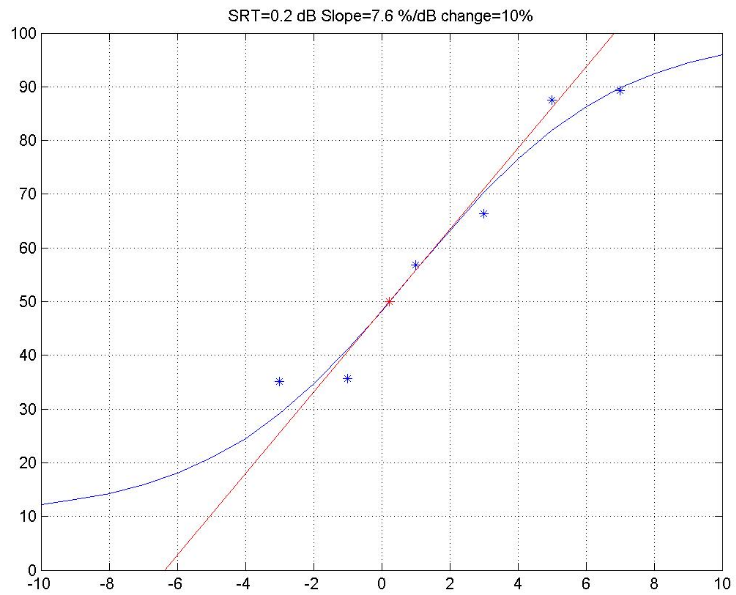

Figure 5. Matrix test results of speech in combination with speech noise at different signal-to-noise ratios for DMR radio.

4.2 STI results with the STIPA-VC test signal

Figure 6 show the STI results of the new STIPA signal measured for the range between -9 and +15 dB SNR. Measurements results lower than -9 dB SNR were omitted since they were not transmitted by either of the radios. Overall, the STI results do show the expected behavior as a function of the signal to noise ratio. Also, a difference between the STI data is found for the two radios, which corresponds with the subjective findings.

Tetra radios show better result than DMR. Especially for the high STI range (0.7 and higher), the difference in SNR is comparable with the subjective results.

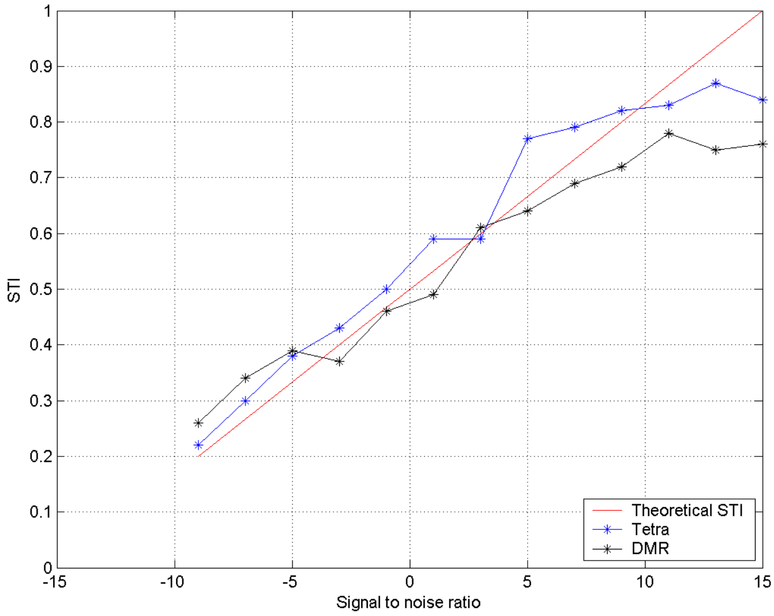

Figure 6. STI results using the new STI carrier signal at different signal-to-noise ratios for Tetra and DMR radio.

5 CONCLUSIONS

The results described in this paper show that replacing the standard noise carrier from the STIPA signal with a deterministic carrier signal is a successful strategy to enable STIPA measurements on systems featuring narrow-band voice coders. A benefit of this approach is that standard STIPA test equipment can be used.

We advise to limit the use of the modified STIPA signal to those situations for which noise carriers are unacceptable.

6 REFERENCES

1. Thomas Brand and Birger Kollmeier, Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests, Acoustical Society of America, Vol. 111 No. 6, June 2002
2. Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America*, *66*(5), 1333-1342.
3. Rolph Houben, Jan Koopman, Heleen Luts, Kirsten C. Wagener, Astrid van Wieringen, Hans Verschuure and Wouter A. Dreschler, Development of a Dutch matrix sentence test to assess speech intelligibility in noise, International Journal of Audiology, april 2014.