"HENRI COANDA"
AIR FORCE ACADEMY
ROMANIA

"GENERAL M.R. STEFANIK"
ARMED FORCES ACADEMY
SLOVAK REPUBLIC

INTERNATIONAL CONFERENCE of SCIENTIFIC PAPER
AFASES 2012
Brasov, 24-26 May 2012

METHODS FOR THE EVALUATION OF SPEECH ENHANCEMENT ALGORITHMS FROM SPEECH INTELLIGIBILITY POINT OF VIEW

Jaroslav, **Hovorka**

Mesit pristroje spol. s r.o., Uherske Hradiste, Czech republic

Abstract: *Cockpits of the aircrafts, especially those propeller-driven and the interior of a broad range of ground combat platforms, especially tracked ones, are characterized by high amplitude low frequency noise. This noise is coming from the engine and is interfering with communication. Since command and control in critical combat mission is a key to mission success, high speech intelligibility is the basic requirement in practice. A broad range of digital techniques and algorithms exists today which can be used and implemented either in personal headsets or in vehicular communications harness to increase the speech quality. However, these digital techniques and algorithms can have influence on the intelligibility of the processed speech. For this reason, digital algortihms must be assessed from intelligibility point of view, since high intelligibility is a crucial and basic requirement. This article is focused on the methods which are suitable for the evaluation of intelligibility of the processed speech and for the assessment of speech enhancement algorithms. Various methods which can be used for the assessment of speech enhancement algorithms from this point of view are described here.*

Keywords: *Intelligibility, nonsense syllable test, word tests, sentence tests, SPIN, HINT, speech reception threshold.*

1. INTRODUCTION

Noise levels on the board of a broad range of combat platforms can reach extremely high values. As indicated in [1], noise levels can reach up to 117 dB on the board of tracked combat vehicles. Very high noise is entering communications systems which are installed on the board of these platforms, examples of these systems can be radios or vehicle intercoms. The noise is entering communications channel through the microphone of personal headsets used by the crew members [2]. Since the noise is in the frequency range of the speech, it cannot be removed by frequency filtering. Some of the speech enhancement algorithms must be implemented to increase the quality or intelligibility of the processed speech [3]. Improvement of speech quality, however can result in degradation of speech intelligibility [3]. Speech quality and speech intelligibility are totally different attributes of the speech.

Speech intelligibility is very important especially in the field of military communications. The main role of all communications systems, either onboard or man-wearable is to provide reliable and intelligible speech under high noise of the battlefield.

Today, a broad range of speech enhancement algorithms can be used to process speech in these systems [3]. These algorithms and procedures can influence the intelligibility of the speech which is processed by particular digital algorithm. It is the main

reason why methods and tests must exist so that designers of the algorithms and also users of the systems can make assessment of the intelligibility of the processed speech and find out which algorithms and procedures are the best from this point of view.

A wide range of tests is used to assess intelligibility of the processed speech and ability of the particular algorithm to provide intelligible speech at its output. Developing a good intelligibility test is very difficult task. It is important that all major speech phonemes are well represented in the test, all the lists in the tests should be of equal difficulty and contextual information should be under full control.

2. INTELLIGIBILITY TESTS

2.1 The main tests of intelligibility. Intelligibility tests can be divided into three main groups. These groups [3] are nonsense syllable tests, word tests and sentence tests. All these methods have their own advantages and disadvantages.

The first group, nonsense syllable tests [4], is based on using a random list composed of the defined number of nonsense monosyllables. These monosyllables must have the format Consonant-Vowel-Consonant. Each initial and final consonant and each vowel are used only once. These syllables are being presented to a group of listeners and their role is to identify these syllables. This way the number of syllables identified correctly is obtained. This test is called articulation test. Articulation test was modified and modifications of this test exist today. The common disadvantage of using nonsense syllable test is difficulty and inability to prepare test lists of syllables in which all items are equally difficult to recognize [3].

The second group of methods, for the assessment of intelligibility, are the word tests. These tests are based on single meaningful words, which differ in the leding or trailing consonant. Here are the two main ways to make word tests – phonetically balanced word tests and rhyming word tests.

Phonetically balanced word tests [5] are based on presentation of 20 lists of 50 common monosyllables. All the lists must be designed so that the requirement of average difficuilty and equal range of difficuilty was met. It is also necessary words in the list have the very same phonemic distribution as expected in normal speech.

The other method of word tests is represented by the rhyming word test. These tests are fully based on rhyming words [6]. All items of the test are monosyllables. These monosyllables must have the form Consonant-Vowel-Consonant. The trailing vowel-consonant is given and known and the role of the listener is to identify the leading consonant. Test words must be chosen very carefully. This method is also modified to method which is called Modified Rhyme Test (MRT) [7]. Here, in this method, listener response is restricted to a finite set of rhyming words. It means that for aech of the words presented in the test, reponse contains the group of possible rhyming words. These words are words which the listener can choose from.

Other modification of the general word test is Diagnostic Rhyme Test (DRT) [3]. This algorithm is widely used for the evaluation of intelligibility of speech coders. In DRT, rhyming words do not differ only in the leading consonant, but they also differ in one distinctive feature of the leading phoneme. The role of the listener is to choose from two options, of course one of the options is the stimulus word. Distinctive features which are being used in these tests are voicing, nasality, sustention, sibilation, graveness and compactness. DRT test has a big advantage over the other tests, since DRT tests except the overall intelligibility score for the speech enhancement system gives the diagnostic score for each of the distinctive features. DRT tests were recognized to be very reliable tests in practice.

Disadvantage of the word tests is the fact, they do not reflect real-world situations, since humans are using sentences in real situations, not only words. It is the main reason why sentence tests were introduced in practice and are widely used in assessment of speech intelligibility. It is maximally important in sentence tests, phonetic content is balanced and distribution of the sounds in the language

is properly reflected. Two main types of the speech tests exist today – Speech Perception in Noise (SPIN) and Hearing in Noise Test (HINT) [3]. SPIN test is composed of 8 lists, each list contains 50 sentences. Each sentence in the list is composed of 5 to 8 words. It is maximally important to reach equal difficuilty among the lists. It is the main reason why half of the sentences in each list contain words of high predictability and half of the sentences contain words of low predictability. Listeners are asked to give single word response. HINT test contains 25 phonetically balanced lists of the sentences. Each list contains 10 sentences. All the sentences are equalized for naturelness, length and intelligibility. The difference to SPIN test is that in HINT test the sentences are scored on word by word basis. HINT test is very popular and widely used.

2.2 Mesuring speech intelligibility. In the test described above, intelligibility is defined as percentage of the words or syllables which are identified correctly by the listener. Percentage intelligibility is often measured at speech to noise levels which are fixed.

Alternative to percentage correct scores are Speech Reception Threshold (SRT) [3]. SRT can generally be measured under quiet conditions or under a noise. The meaning of SRT is different and is dependent on conditions under which measurement is done, if measured in quiet conditions or under the noise. In quiet conditions, SRT means intensity level at which listeners identify words with accuracy 50 %. It is obtained by presenting speech material at different intensity levels. Performance intensity plot is obtained. It is easy to determine 50 % point, which corresponds to SRT.

Totally different situation is if the measurement is done under the noise. SRT under the noise represents signal-to-noise ration at which listeners identify words with accuracy 50 %. This test is done by presenting speech to the listeners at different signal-to-noise ratios. Performace plot can be made which is a plot of percentage correct scores as a function of signal-to-noise ratio. 50 % point representing the SRT in this plot can be obtained.

4.CONCLUSIONS

A widely used methods suitable for the intelligibility assessment of speech and digital speech processing algorithms were outlined in this article, these methods cover nonsense syllable tests, word tests and sentence tests. Speech reception Threshold under quiet conditions and under a noise were explained.

BIBLIOGRAPHY

1. Hovorka, J., Marko, M., (2009). Predikcia hlukovej expozície osádky bojových vozidiel, SES Liptovský Mikuláš.
2. Hovorka, J.,(2008), Moderní přístupy pro zabezpečení srozumitelnosti komunikace uvnitř bojových vojenských vozidel, *Měření Diagnostika a Spolehlivost palubních sousvat letadel, 8. mezinárodní vědecká konference*, Brno, 2008.
3. Loizou, P., (2007) *Speech Enhancement Theory and Practice*. CRC Press
4. Fletcher, H. and Steinberg, J. (1929), Articulation testing methods, *Bell Syst. Tech. J.,8.*
5. Egan, J. (1948), Articulation testing methods, *Laryngoscope, 58(9).*
6. Fairbanks, G. (1958), Test of phonemic differentiation: the rhyme test, *J. Acoust. Soc. Am.,30(7).*
7. House, A., Williams, C., Hecker, M. and Kryter, K. (1965), Articulation-testing

methods: consonantal differentiation with a closed-response set, *J. Acoust. Soc. Am., 37(1).*