TESTING INTELLIGIBILITY OF RE-ADDED NOISY SIGNALS AFTER NOISE SUPPRESSION PROCESSING

Juan Johansson, Nikos Menoudakis, Marcell Balogh, Sergi Escanes

Aalborg University, CREATE, Rendsburggade 14, 9000 Aalborg, Denmark

ABSTRACT

Due to byproducts of noise suppression in audio, unwanted auditory artifacts are introduced during processing which affect quality and intelligibility of the processed signal. A proposition expressed in a paper [1] on intelligibility suggests that adding a certain proportion of the original unaltered signal to its processed counterpart could increase intelligibility, giving us inspiration to carry out an experiment testing this proposal. An array of recorded segments of speech with added noise were processed with noise suppression, modified and tested with 20 participants for intelligibility using both objective and subjective tests, in order to assess the aforementioned proposal. We found that using a proportion of 15 percent original unprocessed speech signal to 85 percent processed speech signal resulted in highest results for intelligibility.

***Index Terms*—** Signal processing, digital hearing aids, noise reduction

1. INTRODUCTION

One of the most crucial aspects to efficient communication in day-to-day life is the ability to understand one another. Whether it is visual or auditory, coherence in communication is essential to being able to get your message across adequately. Although it seems trivial, it is taken for granted that many of us can do this naturally, the most common example being the physical act of speaking to one another, which will be referred to as vocal communication. Unfortunately, due to either pre-existing conditions or injury, there is an estimated 350 million people world-wide that suffer from disabling hearing loss [2], and many of them therefore struggle with vocal communication due to loss of the capacity to properly capture important frequency ranges that the human ear uses to take in and process auditory information, especially in the case of human speech. The human ear is known to have a hearing range of $20 - 20000$Hz, of which $200 - 8000$Hz is the range where human speech is heard, so when hearing impairment affects the ability to adequately capture vital frequencies within this speech range, the ability to understand what is said, what we consider as intelligibility, could be hindered.

Fortunately, technology has evolved to a point where this can be helped with hearing aids, devices created to improve hearing for those with hearing impairments. These devices use a variety of techniques in order to help improve auditory information relevant to the type of hearing impairment it is used for. Although giant steps have been made to improve hearing with these devices, due to the complexity of the human auditory mechanism along with technological constraints, we are still at a point in time where hearing aids cannot completely compensate for the sound quality of unimpaired, natural hearing, and even in certain conditions fall short in being able to adequately capture and transmit auditory information.

One example is the presence of different types of unwanted auditory noise that exist in everyday situations, such as babble at a crowded restaurant or the sound of inside an aircraft cabin during flight, causing in mild cases minor annoyance and in more serious cases masking of other more important auditory information taking place at the time. Unfortunately, for hearing aids the problem of prevalent unwanted noise is also present, and in certain situations can negatively affect the mechanics of the hearing aid and the processes it undergoes to enhance certain aspects of hearing. One of these problems is the loss of intelligibility due to unwanted artefacts introduced during the process of noise suppression that hearing aids use to reduce unwanted noise in the very situations stated above.

Digital noise suppression is achieved by an algorithm which identifies noisy regions of the incoming audio signal and suppresses their volume level until it comes across audio which it considers not to be noise, at which suppression stops. Under adequate conditions, this successfully suppresses unwanted noise and overall improves the audio signal being processed. It is quite common, however, that there are situations in which the undesired noise is at a high enough volume level where it is hard for the noise suppression algorithm to differentiate between what is noise and what is not noise, possibly causing it to suppress auditory information it was not meant to suppress. The consequence to this is spurious volume modification of the audio signal that can hinder audio quality and intelligibility in the case where speech is the particular type of audio information attempted to be transmitted in these types of noisy environments. Whilst there are different approaches to help correct this problem, one technique proposed by Mads G. Christensen et al.[1] suggests that the addition of some of the original noisy signal to the noise-suppressed signal (which will be referred to as processed signal) could help improve the overall speech intelligibility. The scope of the project presented in this paper will be the investigation of this particular technique, information gathered and used to help apply and test this technique by conducting a formal experiment, and analysis of data gathered to conclude whether this technique can improve intelligibility in noisy speech signals that have had noise suppression processing applied to them.

For testing this technique, several types of speech were chosen as well as some stationary background noise. First of all, these signals were cleaned of noise using the Wiener filter. Afterwards, a proportion of the original noisy signal was added to then test whether intelligibility grows or not according to our initial hypothesis. Once this was done, we carried out a test for checking that as well as checking the optimal amount of unprocessed signal that has to be added back to the enhanced signal for highest intelligibility. The results will be presented and discussed in the sections below.

2. METHODS

We chose 10 different speech phrases with car background noise for testing. The initial SNR before processing was -15dB. The Ephraim-Malah algorithm (Wiener filtering) was then used via a MATLAB implementation [3] and was applied for suppressing noise. The processed signals were then mixed with their original noisy ones in three different ratios: 15, 30, 45 percent. That is, for instance, that in the last case 45 percent of the signal tested was noisy and the rest was processed. These ratios were chosen taking into account two criteria: to keep SNRs low and to be sufficiently spaced among them to so as to tell them apart. The mixing was done as follows:

$$Y' = (1 - \alpha)Y + \alpha X \tag{1}$$

where α is the percentage, Y' is the signal to be tested, Y is the enhanced (processed) signal and X is the unaltered noisy signal. A total of 20 participants were asked to do the test.

The participants filled in a quick form with the following questions: name (optional), nationality (optional), hearing problems (yes or no; mandatory), English level (intermediate, advanced, native; mandatory). These questions were intended to identify the relevance of factors such as English level in the test.

The actual test procedure was the following. Participants were asked to take a seat in the sound lab and to put a set of headphones (Urbanears Plattan) on. They were granted a training session for volume adjustment and to get themselves familiar with the steps to follow in the test. The sequence of steps after the training session were as follows: participants listened to a sentence and were asked to repeat back as much they could. Moreover, they were asked to rate how intelligible they thought the message was according to a certain scale [3]. This was done to exclude correct answers due to guessing.

Parallel to the experiment, we performed a Short-Time Objective Intelligibility (STOI). This tool allowed us to test intelligibility before performing the actual test and helped us decide upon what percentages should be tested. The results will be presented in the following section.

3. RESULTS

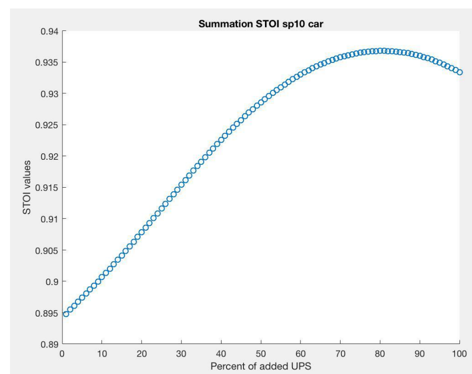

Fig. 1. STOI scores for one of the sentences used in the test varying the amount of noisy unprocessed signal (UPS) from 0 to 100 percent in the tested signal

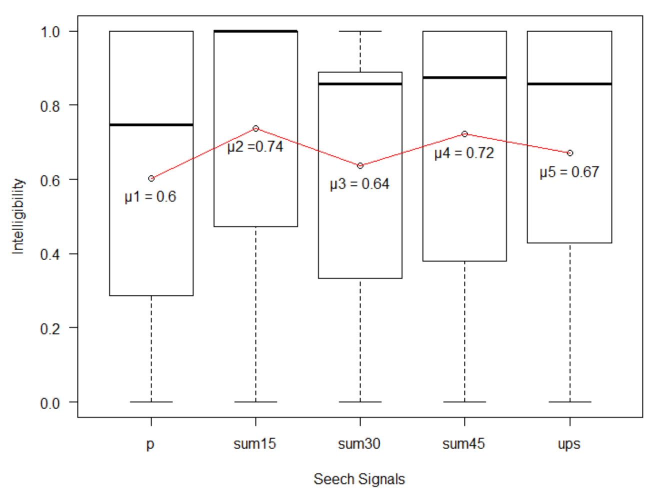

Fig. 2. Box plot of scores for the audio intelligibility test.

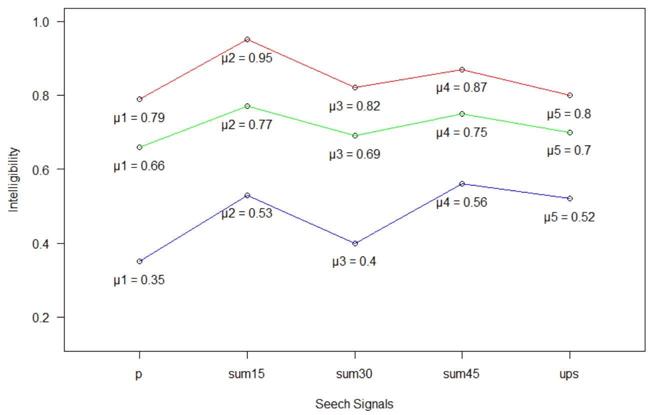

Fig. 3. The influence of English levels on intelligibility scores. From the top to bottom: Native (red); Advanced (green); Intermediate (blue).

4. DISCUSSION

In this section, a brief discussion on the results gathered will be given. Figure 1 shows STOI score as a function of percentage (from 0 to 100) of noisy signal added given a sentence. The initial intention was to use STOI's results to help us decide upon the optimal percentages to be tested in the experiment. As we can see in the graph, its maximum is around 70 percent for this sentence in particular. Similar results were obtained with the rest of the sentences. Since we wanted to work with reasonable SNRs, that would not have been a good choice, hence our choice was that which was mentioned in section 2. Therefore, the results of the STOI were used to check the tendency of adding noisy signal to the enhanced one in relation to intelligibility. We saw that there was a positive correlation of these two variables (at least until it reaches the maximum), as we wanted to check according to our initial hypothesis.

The last two figures present the results of the experiment. Figure 2 shows that the optimal percentage to be chosen amongst the ones we tested is 15 percent (sum15).

It was found that participants with higher English skills performed better than the others. According to this, one could assume that the language played an important role in the experiment biasing our results. However, according to figure 3 the outcome of the separated levels is very similar compared to the analysis done when they are not isolated. Consequently, the hypothesis stating that language proficiency influenced the results can be rejected.

5. CONCLUSION

Based on the results gathered, we can conclude that our hypothesis has been proven true. Additionally, our results are in good agreement with the ones obtained by Mads G.C. et al.[1], showing highest intelligibility when 15 percent of noisy signal is added to its processed counterpart. Although results showed that language barriers provided no significant influence to results, further studies which could be done with more participants, more combinations of sentences used and different types of noise could shed light upon the mechanics underlying speech perception with regards to noise suppression.

6. REFERENCES

[1] M. G. Christensen J.B. Boldt M. S. Kavelekalam J. K. Nielsen. Model Based Speech Enhancement for Intelligibility Improvement in Binaural Hearing Aids. 2017.

[2] World Health Organization. Deafness and hearing loss. http://www.who.int/mediacentre/factsheets/fs300/en/ . 2017.

[3] Esfandiar Zavarehi. Wiener Filter - Ephraim-Malah algortihm. https://se.mathworks.com/matlabcentral/fileexchange/7673-wiener-filter?focused= 5061187tab=function.

[4] C Bowen. Table 1: Intelligibility. http://www.speech-language-therapy.com/ . 2011.