Speech modifications to increase the intelligibility of vocal messages broadcast by driving assistance systems intended for hearing-impaired drivers

Jérémy Rouch[1)], Étienne Parizet[2)]
Univ Lyon, INSA-Lyon, Laboratoire Vibration Acoustique,
F-69621, Villeurbanne, France.
[1)]jeremyrouch@aol.com, [2)]etienne.parizet@insa-lyon.fr

Summary

The purpose of the AIDA (Automobile Intelligible pour Déficients Auditifs - Intelligible automobile for hearing impaired) project is to develop speech modifications tools to increase the intelligibility of voice messages broadcast by driving assistance systems embedded in personal vehicles. The target audience for this project is people suffering from mild and moderate hearing loss. This article presents an approach to increase the intelligibility of natural speech signals by signal processing methods based on filtering and dynamic compression. The magnitude of the filter was defined from a global SII (Speech Intelligibility Index) optimisation method in a driving situation incorporating hearing loss via the use of a hearing loss simulator. The effect of these speech modifications on intelligibility level was assessed using a word test on ten hearing-impaired subjects. The results show a significant increase in the speech reception threshold (SRT) in two different in-car noises. The SRT was shifted by about -9 dB to -11 dB depending on the in-car noise compared to unmodified words with identical sound pressure levels.

1 Introduction

Advanced driver assistance systems (ADAS) embedded in personal vehicles communicate part of the information intended for the driver by voice messages via the vehicle sound system. Correct processing of the information communicated by these systems largely depends on how intelligible the message is for the driver [1, 2]. This intelligibility is a function of three main factors: the masking noise inside the passenger compartment, the quality of the voice signal and the driver's hearing ability. Reduction of the effect of the masking noise is achieved by reducing the level of this noise using passive or active methods. Passive methods involve the use of materials designed to insulate the passenger compartment or damp the vibrations producing the noise. Active methods involve reducing the noise directly in the passenger compartment by the addition of a sound signal in phase opposition to the noise signal. The use of passive materials is already widespread and an extra addition would mean extra cost and work on the design of the vehicle's structure itself. Active methods require the use of microphones to be placed in the passenger compartment and possibly the addition of loudspeakers. In addition, active control systems for the noise inside the vehicle, even the most effective, are really effective at frequencies lower than 500 Hz[3, 4] whereas the useful frequencies of the voice are higher[5]. For the designers of ADAS systems, the most practical and cheapest means of increasing the intelligibility of voice messages broadcast by these devices is to change the speech signal directly. Increasing the intelligibility of a partially masked voice signal without changes to the background noise is discussed in the literature as a problem of near-end listening enhancement (NELE)[6]. The extension of this problem to cases of hearing impaired has so far only been accorded very little attention even though the difficulties of listening to speech in noise are particularly significant for hearing impaired[7]. The problem of understanding noisy speech for hearing impaired has above all been dealt with by methods for the adjustment of hearing aids[8, 9, 10]. However, in this particular case, the speech and the masking noise cannot be processed in totally separate ways because these aids work on the global signal presented at the entrance of the device's microphone. This signal is made up of a mixture of speech and noise.

The algorithms for natural speech modifications presented in the literature to deal with the NELE problem[12, 13] go from a simple filtering followed by a wide dynamic range compression[14] to more complex signal processing techniques using empirical mode decomposition (EMD) methods[15] or reproducing components of the Lombard effect[16, 17]. Recently, three groups of authors published results of the evaluations of different speech modifications algorithms on hearing-impaired subjects in different masking noise conditions. Schepker *et al.*[18] suggested an algorithm based on an evaluation of the SII of the noisy signal by time windows. The value of the SII for each window controls the parameters for frequency rebalancing and dynamic compression. For high SII values, the speech signal is only slightly modified because the intelligibility in the time window is already high whereas for low SII values the signal undergoes a greater modification. The evaluation of these modifications at a constant SPL level compared with the original speech signals on hearing-impaired subjects in the case of a masking noise recorded in a cafeteria show a reduction in SRT of 1.5 dB on average[19]. Nathwani *et al.*[20] used a speech modifications algorithm based on the reproduction of three Lombard effect components to increase the intelligibility of a speech signal in an in-car noise. This algorithm is based on detection and then separate processing of voiced and unvoiced segments of the speech signal. A time dilation is applied to the entire signal but in a more significant manner for the unvoiced parts. A detection of the frequency positions of the formants in the unvoiced segments is carried out[21]. These formants are then shifted to higher frequencies. The energy ratio of the unvoiced parts to that of the voiced parts is increased. The results of the performance assessment for these speech modifications on hearing-impaired subjects in an in-car noise show a reduction in SRT of about 1.8 dB compared with the unmodified situation at the same SPL level. Zorila *et al.*[22] tested the efficacy of four NELE algorithms on hearing-impaired subjects. The most efficient of these algorithms works on the speech signal by time windows[23]. A voicing indicator is calculated on each window. This controls the speech modifications levels. These modifications involve a shift in the energy of the signal below 500 Hz to higher frequencies and the application of a dynamic compression. The effect produced by these modifications on hearing-impaired subjects resulted in an intelligibility increase of about 35% to 40% when the speech signal was broadcast at an SNR (Signal to Noise Ratio) corresponding to the SRT of the unmodified speech.

The present article proposes a speech modifications method for ADAS systems intended for hearing impaired. It incorporates the effect of hearing loss into its design. This method is based on frequency filtering of the speech signal followed by a dynamic compression. The definition of the filter frequency response is performed by a metaheuristic global optimisation process of the SII in an environment reproducing that of the passenger compartment of a car. The calculation of the SII[11] is performed on a sound signal passing through a hearing loss simulator. Due to the use of this hearing loss simulator, the standard methods for calculating the SII proposed in the standard cannot be directly applied so an alternative but exact method, which leads to the same results as the standardised method, is therefore proposed. The first part of this article presents this speech modifications method and the way to obtain its parameters. The second section is concerned with the preliminary results: on the one hand on the stability of the equalisation curves for the speech signal produced by the optimisation process, and on the other hand on the effect of the speech modifications on intelligibility for normal-hearing subjects with a simulated hearing loss. This gain was expressed by a shift in the SRT of broadcast words when these words are modified compared with the same, but unmodified, words broadcast at identical SPL levels. The effects of these modifications on hearing-impaired subjects are presented in a third section, including the possibility of using a general equalisation curve independent of the hearing loss profile of the subject. Finally, a discussion of the results and a general conclusion for this work are presented.

2 Speech modifications

The speech modifications method proposed took place in two successive steps. These steps are close to those developed by Schepker *et al.* [18] and Zorila *et al.* [23] in the sense that they involve a change in the frequency balance of the speech signal followed by a dynamic compression. The originality of this method lies mainly in the fact that the search of the optimal frequency balance took into account a hearing loss profile in the maximisation process of the SII.

2.1 Filtering of the vocal signal by maximisation of the SII incorporating hearing loss

Two successive filters were used. The first one was a 12th order Butterworth band-pass filter for which the cut-off frequencies were determined from a general consideration of the frequency bands that are useful in the calculation of the SII. In one-third octave bands, the calculation of the SII was performed between 141 Hz and 8913 Hz. The cut-off frequencies of the first filter were chosen to observe these limits. They were located between 125 Hz and 10 kHz. This first filter essentially suppressed the background noise, because speech only has very little energy below 125 Hz or above 10 kHz.

The second filter was specific to a hearing loss profile and to the masking noise. Its magnitude was determined from a search for an optimum equalisation on the useful octave bands for the SII (160 Hz to 8 kHz). This SII optimisation process was carried out by a genetic algorithm. The calculation of the SII for each equalisation possibility proposed during the execution of the genetic algorithm was performed from a speech-shaped noise added with an in-car noise played in a listening booth. The speech noise was equalised by one-third octave band and then added

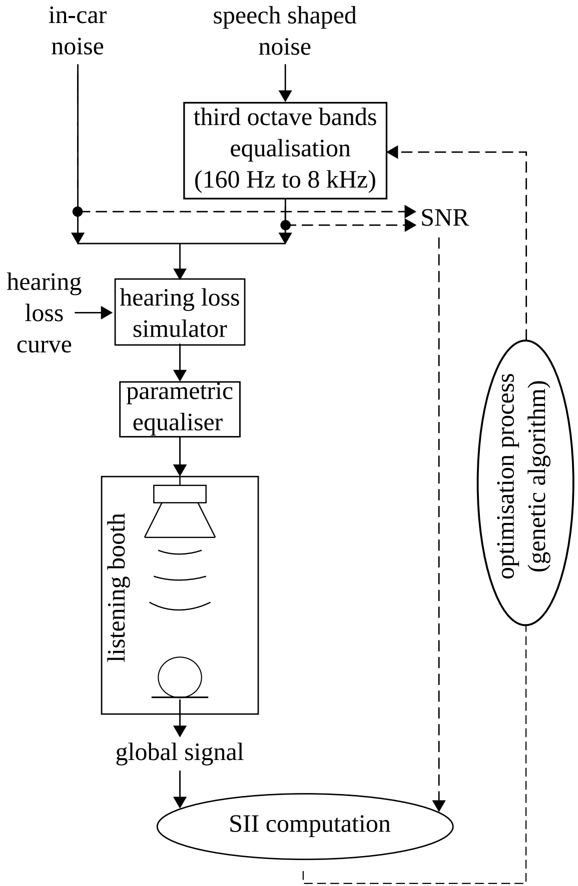

Figure 1: Search procedure for speech signal equalisation in an in-car noise by optimisation of the SII with hearing losses taken into account.

to a masking noise recorded in the passenger compartment of a moving car. The resulting sound signal passed through a hearing loss simulator configured in accordance with a given hearing loss profile. Before being broadcast in the listening booth, the signal passed through a parametric equaliser to minimise the effect of the frequency responses of the listening booth and the loudspeaker. When the signal was broadcast in the listening booth, it was recorded by a microphone. The resulting signal at the position of the microphone was used for the calculation of the SII. This calculation also involved the signal-to-noise ratio measured upstream of the hearing loss simulator. The value 1-SII was used as the cost function for the genetic algorithm. Therefore the genetic algorithm sought the equalisation values which maximised the SII. The genetic algorithm used here was supplied by the Matlab™ Global Optimisation Toolbox. The entire process for the determination of this optimum equalisation search is illustrated in Figure 1.

2.1.1 Sound signals

The original speech noise was obtained by a 5 s white noise equalised by one-third octave bands according

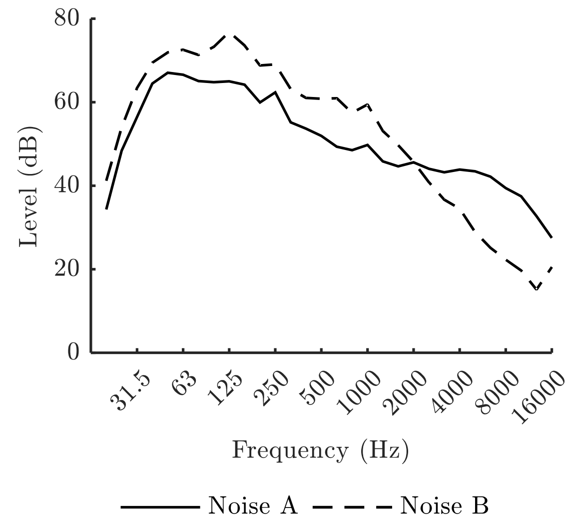

Figure 2: Noise spectra A (50 km/h on a smooth road during a rain shower) and B (130 km/h on a smooth road) recorded in motion at the listener's position.

to the long-term spectrum of the mixed speech with normal vocal effort as defined in the standard relating to the calculation of the SII [11]. The equalisation of this speech noise signal designed to determine the equalisation values optimising the SII was performed by one-third octave bands from 160 Hz to 8 kHz. An amplification factor defined between 0 dB and 20 dB in 1 dB steps was applied independently to these 18 frequency bands. The global signal level was then adjusted so that the signal-to-noise ratio at the measurement microphone was equal to -25 dB.
The masking noise signal came from a 5 s recording in the passenger compartment of a real vehicle travelling at 130 km/h on a smooth road. This recording was performed on an acoustic head placed at the front passenger position. Its broadcast level in the listening booth was adjusted to correspond to the actual level measured during its recording, *i.e.* 69.1 dB(A). Its spectrum is shown in Figure 2 (noise B).

2.1.2 The hearing loss simulator

The hearing loss simulation used[24, 25] was based on the proposal by Irino *et al.*[26, 27]. It reproduces the increase of the absolute threshold of hearing and the loss of compression due to the deterioration of the outer hair cells. It uses a process of inverse compression which is based, for each auditory filter estimated by a gammachip filter, on the combination of passive pass-band filter and active high-pass filter, the characteristics of which depend on the real level of the input signal and the hearing loss level. This process is applied to short time windows. The final signal is reconstructed by an overlap-and-add method. The simulator's input parameters are the auditory threshold levels as measured by a tonal audiometry.

2.1.3 Listening booth

The listening booth used for the broadcasting and recording of signals was a floor-insulated booth with double walls. The interior walls of the listening booth were treated with an absorbent coating. The background noise level in the listening booth was 18 dB(A) and the reverberation times RT30 were 0.24 s for the 125 Hz octave band, 0.12 s for the 250 Hz octave band and less than 0.1 s for the higher octave bands.
The sound signal playback was performed by a Tapco S.8 loudspeaker. The sound signal reception was performed by a Presonus PRM1 omnidirectional microphone. The loudspeaker was placed at a height of 1.2 m and a distance of 1.3 m from the microphone. The microphone was placed at a height of 1.2 m, above a chair in the position of the centre of the head of a virtual listener.

2.1.4 SII calculation method

The SII is a tool for evaluating the intelligibility of a speech signal in a stationary noise. It is calculated on a set of frequency bands. An audibility value A is calculated for each band taking into account the energy ratio between the speech signal and the noise signal, the global signal level and the effect of inter-band masking. These audibility values are weighted by importance values I depending on the vocal material used. The final value of the SII is obtained by adding the result for all the n frequency bands.

$$\mathrm{SII} = \sum_{i=1}^{n} I_i A_i \tag{1}$$

In the following sections of this study concerning the evaluations of the proposed speech modifications method, the intelligibility tests were performed on short words in French. In their structure, these words were very close to those established by Peckels and Rossi[28] to adapt the Voiers' diagnostic rhyme test[29] to the French language. Thus the band importance function used here was the function corresponding to the values defined by Voiers for short words.
According to the ANSI standard, the calculation of the SII may be carried out from the speech signal and the noise signal analysed separately or from one of these signals and the global signal. This latter calculation method assumes that the global signal is the sum of the speech signal and the noise signal. In the method proposed here, the calculation of the SII from the signal recorded in the listening booth could not be made directly because the inverse compression produced by the hearing loss simulator is a non-linear procedure. The gain applied by the simulator is dependent on its input signal level. However, by considering sufficiently narrow frequency bands it may be considered that the hearing loss simulator does not modify the signal-to-noise ratio. In a given frequency band, the inverse compression processed by the hearing loss simulator changes the level of the global sound but does not change the ratio between each of its components (noise and speech). Therefore, using the energy for each frequency band of the speech and noise signals entering the simulator and the energy of the global signal recorded in the listening booth downstream of the simulator, it is possible to determine the energy of the speech and noise contributions in this global signal and therefore to calculate a SII value.
The signal-to-noise ratio at a given frequency value $\mathrm{SNR}(f)$ before and after the simulator is given by:

$$\mathrm{SNR}(f) = \frac{{p_s}^2(f)}{{p_n}^2(f)} = \frac{{p_s}'^2(f)}{{p_n}'^2(f)} \tag{2}$$

where ${p_s}^2(f)$ and ${p_n}^2(f)$ are the components of the speech noise and the masking noise respectively, at the frequency f upstream of the simulator, and ${p_s}'^2(f)$ and ${p_n}'^2(f)$ are the components of the speech noise and the masking noise respectively, at the frequency f downstream of the simulator. ${p_n}'^2(f)$ is obtained from $\mathrm{SNR}(f)$ and the global signal level ${p_g}'^2(f)$ at the frequency f downstream of the simulator by:

$${p_n}'^2(f) = \frac{{p_g}'^2(f)}{1+\mathrm{SNR}(f)} \tag{3}$$

${p_s}'^2(f)$ is obtained by:

$${p_s}'^2(f) = \mathrm{SNR}(f)\frac{{p_g}'^2(f)}{1+\mathrm{SNR}(f)} \tag{4}$$

2.1.5 Wide dynamic range compression

Fast compressions of the sound signal performed by hearing aids are often listed under the generic name of syllabic compression [30]. Globally, this type of compression reduces the energy ratio between consonants and vowels by acting faster than the average duration of a syllable. This has the effect of increasing the audibility of consonants which have a lower energy time density than vowels. During the evaluation of the four NELE algorithms performed by Zorila *et al.*, these authors found that the two most efficient algorithms were those using a wide dynamic range compression. In the speech modifications method presented here, a very similar compression process was used.
The compression algorithm used in this study was proposed by Giannoulis *et al.*[31]. The compression attack and release times were set respectively at 30 ms and 100 ms as defined in the standard, ANSI S3.22-2003[32]. A compression ratio of 3 was applied in the range of levels defined between -15 dB and +15 dB around the median level of a word. The samples of words for which the level was below -20 dB compared with the median level were not compressed to avoid

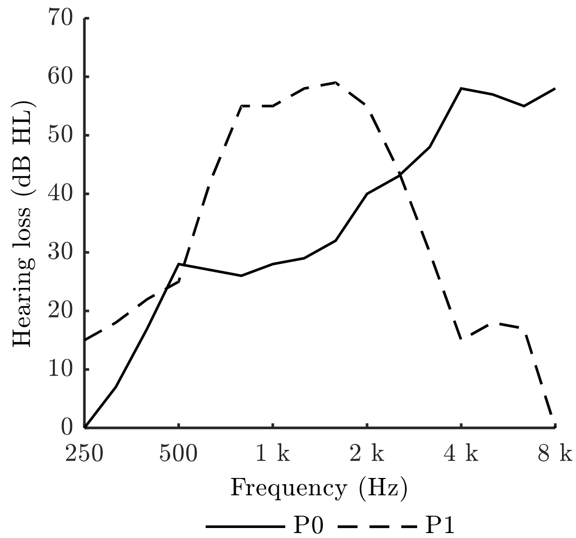

Figure 3: Two hearing loss profiles used to evaluate the stability of the optimisation process.

amplifying the background noise. The samples for which the level was higher than + 15 dB compared with the median level were limited to this level.

3 Pre-experiments

3.1 Stability of the equalisation values obtained by the optimisation process

Due to its heuristic nature, the genetic algorithm introduces uncertainty to the equalisation optimising the SII. This heuristic nature is a result of the random aspects of the initial population draw and the evolution mechanisms[33]. The final solution obtained from this type of algorithm is not the optimum solution but an optimised solution. So, depending on the problem, the initial population and the evolution mechanism, the solution provided by the use of a genetic algorithm may vary. In order to check the stability of the optimised equalisation curves given by this method, ten repetitions of the genetic algorithm were performed on two different hearing loss profiles. The first one (denoted as P0 in Figure 3) was defined from the average of the 12 hearing loss profiles observed for 12 hearing-impaired subjects presenting mild hearing loss (taken from [25]). This profile is due to presbycusis, with an impairment increasing with frequency. The second profile (P1 in Figure 3) was measured at the best ear of a participant suffering from a pronounced impairment in the median frequency range (between 400 Hz and 3.2 kHz).

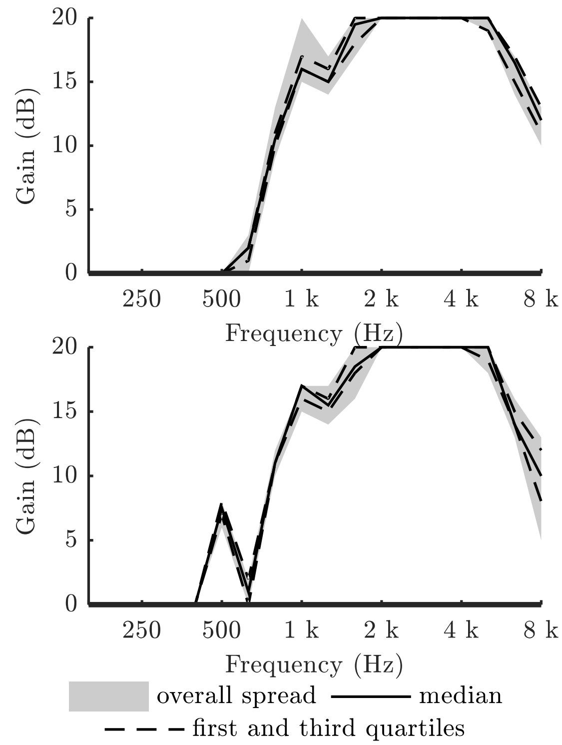

Figure 4: Results of ten repetitions of the genetic algorithm for the search of optimal equalisation values of a speech noise in the maximisation of the SII taking into account a hearing loss level. Top figure: with use of the hearing loss profile P0; bottom figure: with use of the hearing loss profile P1.

3.1.1 Results

The equalisation values maximising the SII obtained are shown in Figure 4. For both hearing loss profiles, the dispersion of the equalisation values optimising the SII was most significant for frequencies between 800 Hz and 1.6 kHz and above 6.3 kHz. The maximum deviation between the first and third quartile was 2 dB at 1.6 kHz for values obtained from the profile P0 and 4 dB at 8 kHz for those obtained from profile P1. For the recommended values between 2 kHz and 5 kHz, no dispersion was observed. At these frequencies, all of the ten repetitions of the genetic algorithm set the maximum gain value of 20 dB. Generally and for the two hearing loss profiles a concentration of the values obtained for the ten repetitions around the median value was observed, showing the stability of the solutions provided by the use of the genetic algorithm[1].

[1] The relative similarity between the two median equalisation curves can be noticed, this point is discussed in Section 4.2.

3.2 Normal-hearing subjects with hearing loss simulation

A pre-experiment to validate the efficacy of the speech modifications proposed was performed on 12 normal-hearing subjects with simulation of a mild hearing loss. The hearing loss simulator was the same as that described in Section 2.1.2. The hearing loss profile entered as the parameter for this simulator was the profile P0 shown in Figure 3.

3.2.1 The word test

The effect of increasing intelligibility produced by the speech modifications proposed was investigated by comparison of the SRT measurement with and without speech modifications in a masking noise. The SRT measurements were performed from the psychometric curve plottings obtained by a FAAF (Four Alternative Auditory Feature) test[34] adapted for the French language[35]. The FAAF test involves the subject listening to a monosyllabic word in a masking noise and then identifying it in a list of four words with close phonetic characteristics. Several words are presented to the subject for a given SNR. This procedure is repeated for several SNR values. For each SNR a correct identification score is measured. The scores for each SNR values are used to plot a regression psychometric curve. It is assumed that this psychometric curve is a sigmoid function:

$$\text{Score} = \frac{0.75}{1 + \exp(-\alpha(\text{SNR} - \beta))} + 0.25 \qquad (5)$$

where α and β are two parameters to be adjusted. The value 0.25 represents the expected performance in a case of zero intelligibility because the subject must choose one word from four. The SRT is defined as the SNR leading to a performance equal to 0.625 corresponding to the abscissa of the sigmoid inflexion point. The lower the SRT value the better the intelligibility of the words in the noise. Therefore during the comparison of the two situations a reduction in SRT, or a negative gain on the SRT, indicates an increase in intelligibility.

3.2.2 Stimuli

The words used for the FAAF test performed in this study were French words with a CVC (Consonant-Vowel-Consonant) phonetic structure. In a single presentation list the four words differed by the final consonant, e.g. “GALE, GAVE, GAZ, GAGE”. 36 words were used for the score measurement of a FAAF test at a given SNR. The original words used were spoken by a female voice and came from recordings made in a sound-proofed and acoustically treated booth. The modified words were the same as those previously described with the use of the voice processing for which the equalisation optimising the SII was performed using the hearing loss profile P0. The average SPL levels for each original and modified word were adjusted independently depending on the desired SNR at the microphone placed in the listening booth.
Two masking noises were used. Their spectra are shown in Figure 2. The first noise (noise A) was recorded in the passenger compartment of a moving car using an acoustic head placed on the front passenger seat. This noise was recorded at 50 km/h on a smooth road during a rain shower. The noise of the rain hitting the windscreen and the bodywork was audible. Its recording level was 61.1 dB(A). Its playback level in the listening booth was identical. The second noise (noise B) was that used during the search for the optimum equalisation maximising the SII (130 km/h on a smooth road) played at the same level as during the search for equalisation (see Section 2.1.1).

3.2.3 Test conditions

The FAAF tests took place in the same listening booth as that described in Section 2.1.3. The playback equipment was also identical. The subjects sat on the chair and a touch screen was placed in front of them for the written presentation of the list of four words and the task of selecting one of them. Before the test was performed a training phase was carried out to familiarise the subject with the FAAF test protocol and the list of the 36 words used. This phase involved performing the test for detecting words without modifications for the different SNR values in a rolling noise which differed from those used for the effective phase of the test but recorded in the same conditions (rolling noise on a rough road at 90 km/h). The order in which the situations were presented (with or without modification, with noise A or noise B) was random.

3.2.4 Participants

12 normal-hearing subjects took part in the tests, 6 men and 6 women aged between 21 and 31 with an average age of 26.1. All had French as their mother tongue. The normal-hearing characteristic was verified with an audiogram and defined as a hearing loss level less than 20 dB HL for each frequency between 125 Hz and 8 kHz on each ear.

3.2.5 Results

The effect of the speech modifications on the results of the FAAF test in noises A and B for normal-hearing subjects with hearing loss simulation are presented in Figure 5. On average, SRT decreased by 9.7 dB in noise A and 12.6 dB in noise B. The reduction was homogeneous over all the subjects: the scattering of individual gain values was 6.4 dB in noise A and 3.4 dB in noise B. The SRT reduction was observed

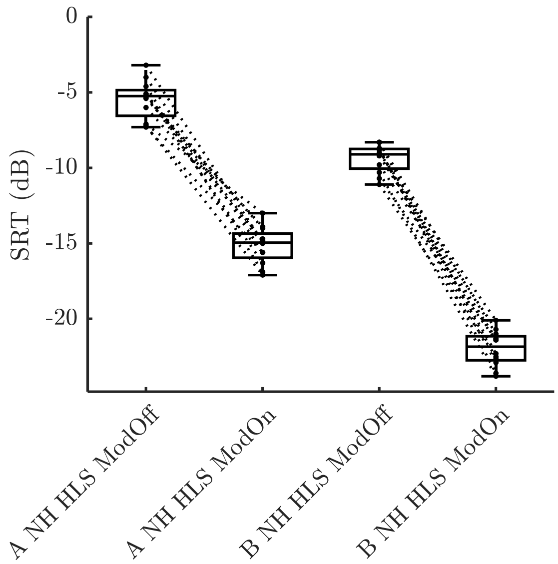

Figure 5: Box and whisker plots and paired comparisons of the effect of voice modifications on the SRT of 12 normal-hearing subjects with hearing loss simulation (hearing loss profile P0). A NH HLS ModOff (ModOn): test in noise A of normal-hearing subjects with hearing loss simulation without (with) use of speech modifications. B NH HLS ModOff (ModOn): similar results in noise B.

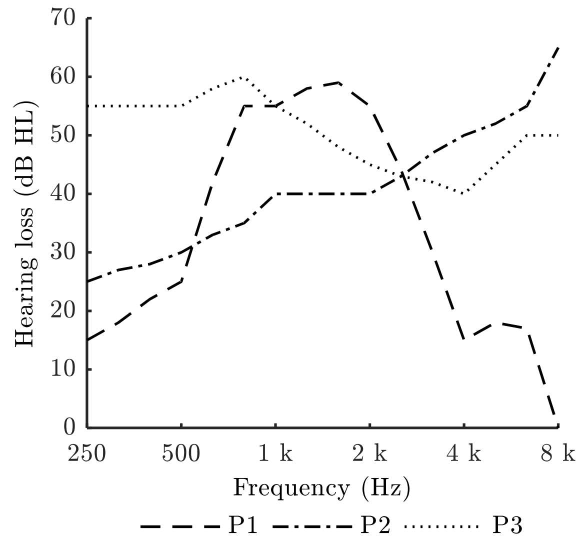

Figure 6: Typical hearing loss profiles for the hearing-impaired subjects. P1, P2 and P3 are hearing loss profiles of the best ear observed in 3 hearing-impaired subjects (speech-frequency PTA: 38 dB HL, 49 dB HL, 40 dB HL).

for all subjects. The hypothesis tested was only the reduction of the SRT, *i.e.* an increase of intelligibility due to the proposed modifications. Thus, a one-sided hypothesis test was used. A one-sided Wilcoxon signed rank test on paired data gave p-values lower than 5.10^{-4} for each of these two listening conditions.

4 Experiment: hearing-impaired subjects

The FAAF test performed on the hearing-impaired subjects was identical to that performed on the normal-hearing subjects as described in Section 3.2.1. The experimental conditions were also the same. The speech modifications used for each subject in this experiment used either the equalisation curve optimising the SII obtained from the hearing loss profile of the subject tested or an equalisation curve common to all the subjects. This equalisation curve was obtained from the average values of the different equalisation curves dedicated to the hearing loss profiles of the ten hearing-impaired subjects and the hearing loss profile P0.

4.1 Participants

10 hearing-impaired subjects took part in the tests, 4 men and 6 women. Their ages were between 30 and 81 with an average age of 59.3. All had French as their mother tongue. These subjects had low to moderate levels of hearing loss characterised by a speech-frequency PTA index between 25 dB HL and 54 dB HL for their best ear (PTA stands for Pure Tone Average *i.e.* average level of hearing loss at 500 Hz, 1 kHz, 2 kHz and 4 kHz). These subjects were selected so as to have relatively homogeneous hearing losses in their two ears. The speech-frequency PTA difference between the two ears on each subject was equal to or lower than 10 dB. Three typical hearing loss profiles could be seen among participants. An example of each of them is presented in Figure 6. P1 was detected in a participant suffering from a pronounced impairment in the medium frequency range (this profile was also used in Section 3.1). P2 is typical of presbycusis and two subjects had rather constant losses between 250 Hz and 8000 Hz (P3).

4.2 Speech modifications with average equalisation

In Figure 4 it is possible to note the similarity of the equalisation curves obtained when maximising the SII for two very different hearing loss profiles (P0 and P1). The major difference between these equalisation curves was a gain of 6 dB for the P1 profile on the one third octave band centred on 500 Hz. This difference could not be explained because the optimisation process takes different factors into account in a complex way. A statistical analysis by two-sided Wilcoxon-Mann-Whitney test comparing the samples of the equalisation values obtained from the two hear-

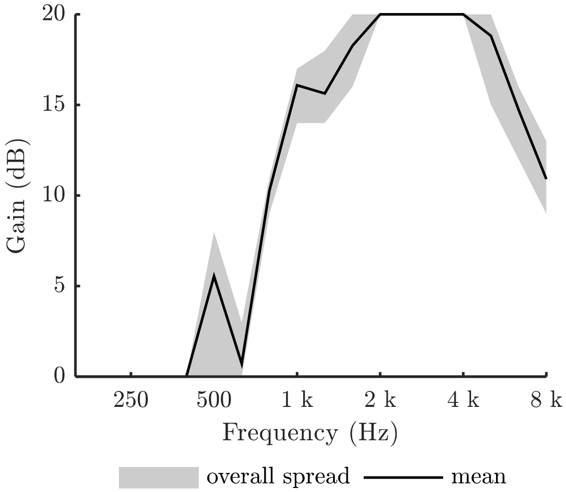

Figure 7: Dispersion and average of the optimal equalisation values of a speech-shaped noise signal by optimisation of the SII obtained from the 11 hearing loss profiles.

ing loss profiles indicated a significant difference (p-value<0.05) only in the one third octave bands centred at 500 Hz, 6.3 kHz and 8 kHz. The dispersion and the average of all the equalisation curves obtained from one execution of the genetic algorithm for each of the 11 hearing loss profiles used in this study (profile P0 and the 10 profiles of each hearing-impaired subject) are presented in Figure 7. On average, amplification was made maximum between 1.6 kHz and 4 kHz. No amplification was applied below 500 Hz. Between 630 Hz and 8 kHz, the dispersion of curves is low (less than 5 dB). This dispersion is higher in the 500 Hz frequency band. Nevertheless, as this band is of less importance in the SII computation, it was possible to consider the use of an average equalisation curve instead of the individually-adjusted ones.

4.3 Results

The results of the FAAF tests performed on the hearing-impaired subjects when using an equalisation dedicated to each individual's hearing loss profiles are presented in Figure 8. SRT values (with or without speech modifications) show a strong inter-individual variability whatever the masking noise (between 16.1 dB and 26.8 dB). This result was not surprising due to the scattering of hearing losses (speech-frequency PTA between 25 and 54 dB). The decreases of SRT due to speech modifications were more homogeneous (gain dispersion was 7.5 dB in noise A and 10.9 dB in noise B). On average the SRT reduction was 9.0 dB in noise A and 11.0 dB in noise B. A one-sided Wilcoxon signed rank test on paired data gave p-values lower than 5.10^{-4} for each of the two listening conditions. When the speech was modified using the mean equalisation curve, the average SRT reduc-

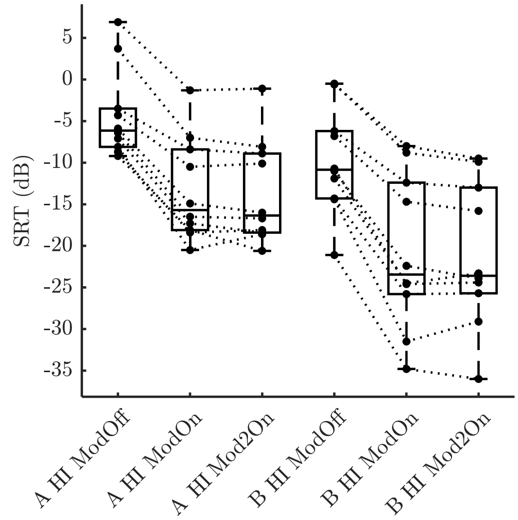

Figure 8: Box and whisker plots and paired comparisons of the effect of speech modifications on the SRT of 10 hearing-impaired subjects. A HI ModOff: test in noise A without speech modifications. A HI ModOn: speech modifications determined from individual hearing loss profiles. A HI ModOn2: speech modifications determined from the averaged hearing loss profile. B HI ModOff, B HI ModOn, B HI ModOn2: same results for noise B.

tion was 9.4 dB in noise A and 11.3 in noise B. A one-sided Wilcoxon signed rank test on paired data gave p-values lower than 5.10^{-4} for each of these two listening conditions. A better performance for the modifications using an average equalisation was not observed in all subjects: some had higher performances when using the dedicated equalisation and others with the modifications using the average equalisation. On average, the effects produced by both kinds of modifications using an average equalisation and the modifications using a dedicated equalisation were not significantly different (p-value of 0.33 for noise A and 0.45 for noise B). The calculation of the Type II risk gave a value for the power of the test of 0.54 for the listening condition in noise A and 0.44 for the listening condition in noise B.

5 Discussion

For hearing-impaired subjects, the absence of a significant difference between the effects produced by dedicated or averaged modifications has not been shown because the statistical power of the test remains low (less than 80%). However, the data comparison indicates that even if this effect exists, it is weak because the difference in medians in the two conditions is also small from a physical point of view, 0.4 dB in noise A and 0.3 dB in noise B.

This potential absence of effect or the weakness of this effect should be compared with the low dispersion of the equalisation curves obtained from the different hearing loss profiles. This low dispersion indicates that the equalisation process depends very little on the hearing loss profile. This can be due to the fact that the optimisation process is based on the SII calculation, which mainly takes into account SNR values in frequency bands. Yet the hearing loss simulation does not modify these SNR values to a great extent. It can also be expected that, without any hearing loss simulation, the optimised equalisation curve would be quite similar to the ones presented in Figure 7.

It has not been possible to show the effect of the shape of the background noise spectrum on the equalisation curve because the optimisations of the equalisation values were only performed from a single background noise. However, the results indicated that the filter optimised for noise B was also efficient for noise A (see Figure 8), even if the SRT reduction was slightly smaller (9 dB instead of 11 dB). Part of the explanation for this result is found in the relative similarity in the spectra of the two noises. Both noises were mainly made up of low frequencies characteristics of in-car noises, even if the high-frequency levels were higher for noise A. However if another type of noise was used, with a very different power spectral density function, the equalisation curves obtained could have been relatively different in shape.

Generally, in the in-car context in which this study took place, it seems that for the different hearing loss profiles the optimisation process results in equalisation curves that, in their overall shapes, are close to each other and close to the importance curves of the speech frequency bands used in the calculation of the SII. Thus (a) as a priority this optimisation process would have a tendency to concentrate the energy in the useful speech frequency bands including a possible background noise effect, in other words it seems that for mild to moderate hearing-impaired subjects, the speech intelligibility in the in-car masking noise context is mainly governed by the general speech important frequencies (established from normal hearings); (b) an average equalisation curve can be used in the word modification process regardless of the hearing loss profile. To a certain extent, this can be considered as an extension the results from Zorila *et al.* [22]. Indeed, these authors applied an equalisation curve which did not depend on the hearing ability of the participants. This improved the speech intelligibility for normal-hearing or hearing-impaired listeners. In the study from Zorila *et al.*, hearing-impaired participants had rather similar profiles while, in the present study, a larger set of different hearing loss profiles was used.

The increase of intelligibility described in this paper is higher than what can be seen in the existing literature when the comparison is possible. The decrease of SRT observed by Rennies *et al.* [19] (1.5 dB) is not directly comparable with our results (between 7.5 dB and 11.3 dB on average) because the listening conditions were not identical. Rennies *et al.* used a masking noise recorded in a cafeteria. This type of background noise is naturally more detrimental for the intelligibility of isolated speech because it is made up of a set of speech signals and its spectrum is similar to that of isolated speech. In the study presented here, the energy of the background noise spectrum was concentrated in the low frequencies outside the useful frequencies for speech. In addition, unlike the method proposed by Rennies *et al.* the optimum equalisation proposed here was done off-line, which left the time for the genetic algorithm to converge to its optimum value. For the same masking noise, Nathwani *et al.* [20] observed a gain of -1.8 dB by the use of voice modifications obtained from an algorithm reproducing Lombard effect components. Moreover, as shown in [22], simulating the Lombard effect is not as efficient as using compression and filtering. Finally, it is not possible to compare the results from the present study to the one from Zorila *et al.* [22] since these authors relate intelligibility improvement at a constant SNR while, in the current study, a shift in SRT was measured.

In the present study, the modified words were assessed in the case of a relatively detrimental listening condition (*i.e.* for SNR corresponding to a SRT of 50%). In better listening conditions, a decrease of voice quality due to the speech modifications proposed here can be expected. For instance, the fast dynamic compression can create audible artefacts. However, the intended use of this speech modification process is to enable the user to adjust the gain in order to reach a satisfactory compromise between speech intelligibility and quality.

6 Conclusion

A method of improving speech intelligibility in cars for hearing-impaired listeners has been proposed. It is based on a wide dynamic range compression and equalisation optimising the SII in an in-car noise by including the effect of hearing loss. The efficiency of the modifications was evaluated on a panel of ten participants presenting mild to moderate hearing loss. A significant increase in intelligibility could be noted, as the SRT was reduced by 9 to 11 dB. This efficacy was also noted when the masking noise (recorded inside a driving car) was different from the one which was used to optimise the speech filtering. The difference in intelligibility gain between the modifications using a dedicated equalisation or an average equalisation was either insignificant or very small. Thus regardless of the hearing loss profile, an average equalisation curve can be used without any performance loss.

A future development of the work conducted here would be the study of the effect of the modifications under a constraint of equal loudness rather than equal SPL level. This work was addressed by Zorila *et al.*. However, these authors used a loudness indicator which is not completely dedicated to the case of masked speech used for hearing impaired. Such an indicator should, therefore, be developed first. A more ecologically valid study could also be performed by working directly on hearing-impaired subjects placed in the passenger compartment of a car, being asked to adjust the speech level so as to reach a low-enough listening effort.

Acknowledgement

This work was supported by the French National Research Agency (ANR) as a part of the AIDA project and by the LabEX CeLyA (Centre Lyonnais d'Acoustique) of Université de Lyon (ANR-10-LABX-60/ ANR-11-IDEX-007).

References

[1] M. A. Nees and B. N. Walker: Auditory displays for in-vehicle technologies. Rev Hum Factors Ergon **7** (2011) 58–99.

[2] F. Chen, M. Jonsson, J. Villing, and S. Larsson: Application of speech technology in vehicles. Speech Technology, Springer, 2010.

[3] P. N. Samarasinghe, W. Zhang, and T. D. Abhayapala: Recent advances in active noise control inside automobile cabins: Toward quieter cars. IEEE Signal Process Mag **33** (2016) 61–73.

[4] R. Schirmacher: Current status and future developments of ANC systems, SAE Int J Passeng Cars-Mech Syst **8** (2025) 892–896.

[5] T. S. Bell, D. D. Dirks, and T. D. Trine: Frequency-importance functions for words in high-and low-context sentences. J Speech Lang Hear Res **35** (1992) 950–959.

[6] B. Sauert and P. Vary: Near end listening enhancement: speech intelligibility improvement in noisy environments. ICASSP 2006 Proceedings (2006).

[7] B. R. Glasberg and B. C. Moore: Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. Scand Audiol Suppl **32** (1988) 1–25.

[8] D. Byrne, H. Dillon, T. Ching, R. Katsch, and G. Keidser: NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures. J Am Acad Audiol **12** (2001) 37–51.

[9] B. C. Moore, B. R. Glasberg, and M. A. Stone: Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF. Int J Audiol **49** (2010) 216–227.

[10] H. Dillon: Hearing aids, Second edition. Thieme, Boomerang press Sydney, New York, 2012.

[11] ANSI S3.5–1997: Methods for calculation of the speech intelligibility index. Am Natl Stand Inst, 1997.

[12] M. Cooke, C. Mayo, C. Valentini-Botinhao, Y. Stylianou, B. Sauert, and Y. Tang: Evaluating the intelligibility benefit of speech modifications in known noise conditions. Speech Commun **55** (2013) 572–585.

[13] M. Cooke, C. Mayo, and C. Valentini-Botinhao: Intelligibility-enhancing speech modifications: the hurricane challenge. Interspeech, (2013) 3552–3556.

[14] R. Niederjohn and J. Grotelueschen: The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression. IEEE Trans Acoust Speech Signal Process **24** (1976) 277–282.

[15] M. E. Hamid, S. Das, K. Hirose, and M. K. I. Molla: Speech enhancement using EMD based adaptive soft-thresholding (EMD-ADT). Int J Signal Process Image Process Pattern Recognit **5** (2012) 1–16.

[16] M. D. Skowronski and J. G. Harris: Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments. Speech Commun **48** (2006) 549–558.

[17] J. G. Harris and M. D. Skowronski: Energy redistribution speech intelligibility enhancement, vocalic and transitional cues. J Acoust Soc Am **112** (2002) 2305–2305.

[18] H. Schepker, J. Rennies, and S. Doclo: Speech-in-noise enhancement using amplification and dynamic range compression controlled by the speech intelligibility index. J Acoust Soc Am **138** (2015) 2692–2706.

[19] J. Rennies, J. Drefs, D. Hülsmeier, H. Schepker, and S. Doclo: Extension and evaluation of a near-end listening enhancement algorithm for listeners with normal and impaired hearing. J Acoust Soc Am **141** (2017) 2526–2537.

[20] K. Nathwani, G. Richard, B. David, P. Prablanc, and V. Roussarie: Speech intelligibility improvement in car noise environment by voice transformation. Speech Commun **91** (2017) 17–27.

[21] K. Nathwani, M. Daniel, G. Richard, B. David, and V. Roussarie: Formant shifting for speech intelligibility improvement in car noise environment. ICASSP 2016 Proceedings (2016).

[22] T. C. Zorilă, Y. Stylianou, S. Flanagan, and B. C. J. Moore: Evaluation of near-end speech enhancement under equal-loudness constraint for listeners with normal-hearing and mild-to-moderate hearing loss. J Acoust Soc Am **141** (2017) 189–196.

[23] T. C. Zorilă, V. Kandia, and Y. Stylianou: Speech-in-noise intelligibility improvement based on power recovery and dynamic range compression. 2012 Proceedings of the 20th European Signal Processing Conference, (2012) 2075–2079.

[24] N. Grimault, E. Parizet, A. Corneyllie, L. Brocolini, R. Weyn, and S. Garcia: Real-time simulation of hearing impairment: Application to speech-in noise intelligibility. J Acoust Soc Am **140** (2016) 3438–3438.

[25] E. Parizet, S. Garcia, A. Corneyllie, and L. Brocolini: Hearing loss simulator for sound quality applications.

INTER-NOISE and NOISE-CON Congress and Conference Proceedings (2016) **253** 535–539.

[26] T. Irino, T. Fukawatase, M. Sakaguchi, R. Nisimura, H. Kawahara, and R. D. Patterson: Accurate estimation of compression in simultaneous masking enables the simulation of hearing impairment for normal-hearing listeners, in Basic Aspects of Hearing, Springer, 2013.

[27] M. Nagae, T. Irino, R. Nisimura, H. Kawahara, and R. D. Patterson: Hearing impairment simulator based on compressive gammachirp filter. Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA) (2014).

[28] J. P. Peckels and M. Rossi: Le test de diagnostic par paires minimales. Adaptation au francais du Diagnostic Rhyme Test de WD Voiers. Rev Acoust **27** (1973) 245–262.

[29] W. D. Voiers: Evaluating processed speech using the diagnostic rhyme test. Speech Technol **1** (1983) 30–39.

[30] R. E. Sandlin: Textbook of hearing aid amplification. Singular Publishing group, 2000.

[31] D. Giannoulis, M. Massberg, and J. D. Reiss: Digital dynamic range compressor design: A tutorial and analysis, J Audio Eng Soc **60** (2012) 399–408.

[32] ANSI S3.22–2003: Specification of hearing aid characteristics. Am Natl Stand Inst, 2003.

[33] S. N. Sivanandam and S. N. Deepa: Introduction to genetic algorithms. Springer Science & Business Media, 2007.

[34] J. R. Foster and M. P. Haggard: The Four Alternative Auditory Feature test (FAAF): linguistic and psychometric properties of the material with normative data in noise. Br J Audiol **21** (1987) 165–174.

[35] S. Favier: Adaptation à la langue française d'un test anglo-saxon de perception de la parole le FAAF test, Institut des sciences et techniques de réadaptation, Lyon, 2009.