Speech Intelligibility in Virtual Environments Simulating an Asymmetric Directional Microphone Configuration

Lorenzo Picinali[a], Silvano Prosser[b], Antonio Mancuso[c] and Giancarlo Vercellesi[c]

[a]De Montfort University, Music, Technology and Innovation Research Centre, LE1 9BH Leicester, UK
[b]Dipartimento di Audiologia, Università degli Studi di Ferrara, Corso Giovecca 203, 44100 Ferrara, Italy
[c]DICO-LIM, Università degli Studi di Milano, Via Comelico, 39, 20135 Milano, Italy
lorenzo@limsi.fr

In the hearing aids applications, the benefit of directional processing and bilateral listening in terms of speech intelligibility from frontal sound sources has been well documented in recent and past studies. Nevertheless, only a few of the situations in real life present a speaker located exactly in a frontal position, and this seems to constitute a limitation for the directional microphones mounted on the hearing aids. Although several attempts have been done to optimize the directional pattern of the hearing aid trough self-adapting or manually controlled settings, practical results tend to remain quite unsatisfactory.
The purpose of this study was to explore the advantage expected by a bilateral hearing aid with an asymmetric directional microphone configuration: responses in terms of speech intelligibility in noise were evaluated in normally hearing subjects for frontal and lateral sound sources.
Through a 3D Ambisonic virtual environment manipulation, the presence of two microphones (the two hearing aids) was simulated in a noisy environment with a speech sound source. The listeners were presented with the signals synthesized from the two simulated microphones calibrated with symmetrical and asymmetrical directional patterns, played through a pair of headphones. The speech intelligibility was measured for all the directional microphones' configurations and for reference speech sources located in frontal and lateral positions.

1 Introduction

In a normal listening situation, thanks to many mechanisms performed by the hearing system, a listener is able to analyze the sound scene and to make a selection between the sounds interpreted as noise and the sounds interpreted as speech, or as any other signal of interest. These mechanisms, often linked with the spatial attributes of the signals conveyed to the hearing system, allow a considerable improvement in the ability to detect and to understand speech signals.

It is well known that sensorineural hearing loss, due either to cochlear cells or neural fibers, damages consistently such mechanisms, introducing deterioration on the speech understanding most of all when the speech source is competing with other noises. These hearing losses can however take some advantage from hearing aid devices. Nevertheless, the use of BTE prostheses (Behind The Ear, for sure the most common typology of hearing aid present on the market [1]) can alter the perception of the cues linked with the space location of sound sources, aggravating the difficulties already brought by the hearing impairment in terms of speech intelligibility in noisy environments.

Among the various attempts made to improve the speech intelligibility in noise for hearing-impaired persons, a promising approach consists of equipping the hearing instruments with directional microphones. These may incorporate also adaptive switching systems between different symmetric setups (in terms of left and right directionality). However, these approaches do not seem to be completely satisfactory, and the speech intelligibility in different real world listening situations (for example both with frontal and lateral speech sources) is not properly restored.

2 Auditory scene analysis and spatial hearing

The auditory scene analysis (from now on, ASA) is the process by which the human auditory system organizes the sounds [2]. Three main mechanisms are involved in the ASA: segmentation, integration and segregation. In the following lines, the integration and segregation processes linked with the spatial attributes of the sound transferred to the hearing system will be taken in consideration and analyzed.

2.1 Integration and segregation

The sound reaches the ear and the eardrum as a whole, and is then analyzed by the hearing system: the ASA model proposes that sounds can be heard as "integrated" (heard as a whole, as a single sound event) or as "segregated" into individual components or, again, a combination of these two perceptual mechanisms can happen. As an example, it is possible to think about the sound of a piano and of a clarinet playing the same note at the same time: the individual components of each sound are segregated one from the other, than integrated into two different sound events that represent the individual perceptions of the two notes.

The processes of integration and segregation are often linked to the analysis of the four standard parameters of the sound signal, amplitude, frequency, duration and spectral content, and to mnemonic elements, as the familiarity with a specific sound and timbre. Nevertheless, other parameters play a really important role in the two ASA mechanisms: these parameters, known as the localization cues, can be grouped in interaural level differences, interaural time differences and direction dependent filtering.

2.2 The cocktail party effect

An example on how the localization cues can be used for the ASA, and in particular for the integration and the segregation of different sounds in a complex auditory scene, can be done citing one of the most known binaural effect: the cocktail party effect [3, 4].

The cocktail party effect arises from the fact that a given signal generated by a sound source in a certain position is less effectively masked by a noise generated by a source in a different position when the subject listens binaurally (with both the ears), The name "cocktail party" comes from the situation a listener have to face when he is reached by a number of different speech sources coming from different positions: even if, acoustically, the speech signals are masked one from the other, the properties of binaural perception helps the listener to select a desired speech from a desired source, and to isolate it from the others, with a consequent increase of the speech intelligibility for that specific source. This effect, also reported as spatial unmasking, is possible thanks to the mechanisms of integration and segregation based mainly on the three localization cues.

3 BTE hearing aids effects on the spatial hearing

How does the presence of a BTE hearing aid affect the ASA mechanisms when these are based on the localization cues? First of all, it should be noticed that the device is placed behind the pinna of the prosthesized subject, therefore the microphone that sends then the signals to the processing unit is located outside the outer ear system (in most of the cases, it is oriented towards the back of the subject itself). This causes, of course, alterations in terms of the filtering effects brought by the pinna to the signals that reach the outer ear, and is then translated in several problems during the sound sources localization process.

If sound sources cannot be properly localized, then some of the ASA mechanisms will not work properly, with a consequent reduction in terms of speech intelligibility in various real-life situations.

3.1 Hearing aids directionality

Given this situation, for which the application of BTE hearing aids risks to aggravate the situation already created by the hearing loss, the use of directional microphones on the prosthesis can be seen as an attempt to introduce an improvement in speech intelligibility for fontal sound sources. Most of the BTE hearing aids, in fact, are equipped with a two microphones system which, using simplified beam-forming techniques, allows the unit to have a variable polar pattern: cardioid, oriented towards the front, for a listening situation when the speaker is located in a frontal position, and omnidirectional for all the other listening situations.

3.2 Automatic or manual directional switching

Having a variable polar pattern hearing aid device introduces of course the problem of how to switch between directional and omnidirectional settings: a solution has been found allowing the subject to manually switch between the polar patterns through a button located on the device itself, or embedding within the hearing aid an automatic switching system which adapts the polar pattern characteristics to the listening situation where the subject is immersed.

Although the adaptive and manual switching systems have been positively judged in laboratory experiences [5, 6, 7], in the real use they do not seem to be well accepted: manual switching systems generate confusion in the prosthesized subjects, whilst the difficulties in the calibration of the automatic systems makes them unsuitable for many real-life situations.

For these reasons, the improvements brought to speech intelligibility for hearing-impaired persons are mainly given by noise reduction algorithms embedded within the hearing aids, while the loss of spatial information brought by BTE devices, with a consequent decrease of speech intelligibility for many real-life situations, does not seem to be regained through any of the previously discussed directional techniques.

3.3 The asymmetrical directionality

Within the past two years, a new approach has been attempted in order to overcome the problems outlined in the previous sections. The directional setups of hearing aids in a bilateral application have always been used symmetrically, calibrating the polar patterns of the left and right devices exactly in the same way: the new approach consists in using asymmetrical directional patterns, calibrating one of the two devices with a cardioid configuration, and the other with an omnidirectional one.

The benefits of this new configuration can be outlined mainly in the fact that the polar pattern of each device is fixed, and does not have to switch (or be switched) between cardioid and omnidirectional depending on the listening situation: if the speech source is located frontally in a surrounding noisy situation, then the signal to noise ratio would be better for the ear with the cardioid hearing aid, whilst if the speech source is lateral or rear, the benefit would be transferred to the ear with the omnidirectional device.

Of course this will generate interaural differences in terms of signal to noise ratios, and these might have effects on the speech intelligibility index in various listening situation: it is exactly the purpose of this research to establish whether the use of asymmetrical directional configurations can be considered suitable in order to overcome some of the problems brought by the use of BTE devices in terms of speech intelligibility.

4 An abstraction: SRT calculation with different directional setup

Other tests have already been carried out on the validity of asymmetrical directionality solutions [8]: nevertheless, further examinations seemed to be required.

The first step has been to create an abstraction of the problem, in order to be able to carry out a simple preliminary test for a first verification: the attempt was to synthesize really simple auditory situations, and to simulate the presence of two microphones, located exactly in the same position, with different polar pattern configurations. The signals of the two microphones could then be played back, one for each ear, through a pair of headphones.

The test has been carried out on normal-hearing subjects: testing hearing-impaired persons would have brought many other variables (hearing loss typology, adaptation to the new devices…), making the test more complex. Of course, as outlined in the conclusions, further testing with hearing-impaired subjects will be required.

4.1 Bformat soundfield synthesis and directional microphone simulation

The Ambisonic approach [9] has been used for the simulation of different auditory scenes: various sources, reproducing a filtered noise with a spectral profile similar to the one of a speech signal, have been virtually located all around the centre of the scene on a two-dimensional plane, and then coded in a 1st order B-Format Ambisonic signal. Within the same signal, a speech source has been coded in a frontal and in a lateral position, keeping always as a reference the centre of the scene.

The presence of two coincident microphones has then been simulated in the middle of the synthesized sound-field: the polar pattern configurations of the two microphones have been calibrated on four different directional situations:

- Symmetric directional setup with omnidirectional polar patterns
- Symmetric directional setup with cardioid polar patterns
- Asymmetric directional setup with omnidirectional polar pattern for the left ear
- Asymmetric directional setup with omnidirectional polar pattern for the right ear

The absence of an environmental simulation (no wall reflections and reverb) and of a simulation of the head between the two microphones made the left and right lateral position of the speech source absolutely equal, and a rear speech source simulation impossible. In fact, an ideal cardioid polar pattern oriented towards the front would generate a reduction of -∞ dBfs in the level of a rear source, making the test nearly useless. Further testing is required implementing environmental and head simulations.

4.2 The SRT measurement

4.2.1 The platform

The processed speech and noise files were loaded into a customized MaxMSP [10] platform for the mixing of the two signals with different weighting factors. The signal outputted from the computer was then routed into a professional audiometer (Aurical-ReSound): the use of an audiometer allowed a much higher precision in terms of level calibration and frequency response if compared to the one of a standard computer phone-out.

4.2.2 Subjects, stimuli and calibration

The subjects that performed the test were ten volunteers, aging between 22 and 25 years, with normal otoscopy, normal middle-ear air pressure and tympanic compliance and pure tone threshold within 10 dB between 0.125 and 8 kHz.

The speech reference signal consisted of 13 sentence lists, 20 sentences each [11]. All signals were calibrated with reference to a 1kHz calibration tone generated at the same RMS level values of the speech and the noise signals.

4.2.3 The measurement

Sentences embedded in noise were binaurally administered to the subjects, seating within a sound-treated room, using a pair of calibrated headphones (Aurical-ReSound). Background noise was kept at a constant level of 50 dB SPL(A). Speech reception threshold (SRT) corresponding to the 50% correct repetition rate [12] for the sentences was recorded through a simple up-down 2 dB step adaptive procedure going on up to 7-8 reversals. Usually, each SRT needed 14-18 sentences. Each subject furnished 8 SRT measurements, from the combination of 4 microphone conditions (bilateral omnidirectional, bilateral directional, right directional-left omnidirectional, left omnidirectional-right directional) for 2 speech signal directions (0° frontal, 90°-270° lateral). Acoustics conditions and sentence lists were administered in random order. For each subject, the whole time of data collection was within 30 minutes.

5 Results

5.1 SRT values in different directional configuration

In the figure number 1, it is possible to notice how the average between the different subjects of the calculated SRT varies for the different directional setups. From left to right, the diagram shows the SRT for frontal and lateral speech positions for symmetric omnidirectional setup (oo), symmetric cardiod setup (dd), asymmetric setup with omnidirectional for the left ear (lo) and asymmetric setup with omnidirectional for the right ear (ro).

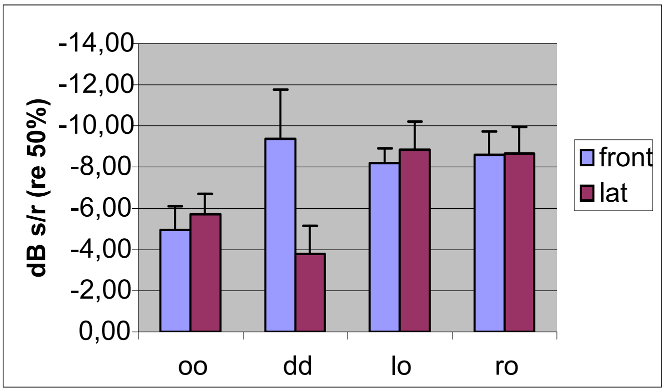

Fig. 1 Diagram of the SRT values for different directional configuration and for different position of the speech signal. The SRT data are reposted in dB s/n ratio

5.2 Brief analysis of the results

Considering only the symmetrical setups, the SRT data confirm the fact that, for a frontal speech source, the best result is given by the cardioid setup, whilst for a lateral speech source by the omnidirectional setup.

The interesting data comes from the two asymmetrical setups: even if the SRT for a frontal speech source is slightly worse than the one of the symmetrical cardioid setup (~1 dB s/r more for both the asymmetrical configurations), the SRT for a lateral speech source is far better (~5 dB s/r less for both the asymmetrical configurations). Focusing then on the SRT data for the symmetric omnidirectional configuration, it can be noticed that they are never better that the two asymmetrical setups, both for frontal and for lateral speech source positions.

Furthermore, the interaural differences in terms of signal to noise ratio brought by the asymmetric setups do not seem to result in a worsening in the speech intelligibility, but seem to improve it, at least comparing the SRT data between asymmetrical configurations and omnidirectional symmetrical ones

These data seem to make the two asymmetrical setups the best choice in terms of SRT average for all the speech source positions.

6 Conclusions and future work

In real-life auditory situations, when the position of a speech source in a noisy environment cannot easily be predicted, an asymmetric directional hearing aids setup seems to be the best choice in order to achieve the higher speech intelligibility both for frontal and lateral speech sources.

Nevertheless, further testing is required in order to confirm what outlined by this first research, trying to make the experimental tested situation more complex and closer to a real-life situation: an environmental simulation needs to be performed, in order to have wall reflections and reverb, simulating also the presence of the head between the two microphones. Furthermore, a test needs to be carried out in real auditory situations, using hearing aid devices and performing the test of hearing-impaired subjects.

Another interesting approach could be to start testing the perceived localization accuracy, also in terms of auditory spatial recreation, of the different directional and omnidirectional, symmetric and asymmetric setups, in order to investigate whether the use of asymmetrical directional configurations could help to overcome the problems of spatial perception linked with BTE hearing aids.

Acknowledgments

The group involved in this research project is supported by GNReSound Italia, with the collaboration of the Università degli Studi di Milano (Computer Science Department) and the Clinica Universitaria di Audiologia dell'Ospedale di Ferrara.

References

[1] J. J. DiGiovanni, *Hearing Aid Handbook*, Clifton Park, NY, US: Delmar, Cengage Learning (2008)

[2] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, Cambridge, MA, US: The MIT Press (1990)

[3] E. C. Cherry, "Some experiments on the recognition of speech with one and with two ears", *J. Acoust. Soc. Amer.* 25, 975-979 (1943)

[4] B. C. J. Moore, *An Introduction to the Psycholoy of Hearing,* London, UK: Academic Press (2003)

[5] T. C. Walden, B. E. Walden, "Predicting Success with Hearing Aids in Everyday Living", *J. Am. Acad. Audiol,* 15, 342-352 (2004)

[6] M. T. Cord, R. K. Surr, B. E. Walden, O. Drylund, "Relationship between Laboratory Measures of Directional Advantage and Everyday Success with Directional Microphone Hearing Aids", *J. Am. Acad. Audiol,* 15, 353-364 (2004)

[7] E. Mackenzie, M. E. Lutman, "Speech Recognition and Comfort Using Hearing Instruments with Adaptive Directional Characteristics in Asymmetric Listening Conditions" *Ear & Hearing* 26 (6), 669-679 (2005)

[8] B. W. Y. Hornsby, T. A. Ricketts, "Effects of Noise Source Configuration on Directional Benefit Using Symmetric and Asymmetric Directional Hearing Aid Fittings", *Ear & Hearing* 28 (2), 177-186 (2007)

[9] M. A. Gerzon, "Periphony: With-Height Sound Reproduction", presented at the *2nd Convention of the Central Europe Section of the Audio Engineering Society*, Munich, Germany (1972)

[10] www.cycling74.com

[11] F. Cutugno, S. Prosser, M. Turrini, *Audiometria Vocale*, GNReSound Italia (2000)

[12] H. Levitt, "Adaptive testing in audiology". *Scand. Audiol. Suppl.* 6, 241-91 (1978)