Computer Aided Hearing Assessment: Detection of Eye Gesture Reactions as a Response to the Sound

A. Fernández(✉), Marcos Ortega, and Manuel G. Penedo

Departamento de Computación, Uiversidade da Coruña, A Coruña, Spain
{alba.fernandez,mortega,mgpenedo}@udc.es

Abstract. A methodology for the detection of eye gestural reactions as a response to auditory stimuli is presented in this work. A precise hearing evaluation is important to improve the quality of life of those who suffer from hearing loss. In the case of patients with cognitive decline or other communication disorders this evaluation becomes much more complicated. The audiologist needs to focus his attention on spontaneous gestural reactions that might indicate some sort of perception. The detection of this gestural reactions is sometimes imprecise and it requires a broad experience from the audiologist. To facilitate this task, we present a fully automated method that analyzes video sequences recorded during the audiometric evaluation and identifies these unconscious gestural reactions. The presented methodology achieves an accuracy of the 94.21 % in the detection of these reactions to the auditory stimuli, which makes of it an interesting tool to assist the audiologists in the hearing assessment of this specific group of patients.

Keywords: Hearing assessment · Gesture information · Eye movement analysis

1 Introduction

Hearing plays a key role on everyday living for every one of us. Among older adults, hearing loss is one of the most common self-reported conditions [1], which is also one of the most widely under-treated. Different studies [2] have demonstrated the considerable negative effects that untreated hearing loss may have on the physical, social, psychological and cognitive well-being of a person. In fact, those who suffer from hearing loss can experience an incomplete communication that impacts negatively to their social lives, at times leading to isolation, withdrawal and lack of independence.

Hearing loss can occur at all ages, but it is more common among the elderly adults. Approximately one in three people between ages of 65 and 74, and nearly half of those older than 75, suffer hearing loss. Age-related hearing loss, or presbyacusis, is the cumulative effect of aging on hearing, the consequence is the slow

A. Campilho and M. Kamel (Eds.): ICIAR 2014, Part II, LNCS 8815, pp. 39–47, 2014.
DOI: 10.1007/978-3-319-11755-3_5

loss of hearing that occurs as people gets older. For all these, regular hearing tests are totally necessary for elderly adults and highly recommended in case of any doubt about the ability of hearing at any age. Pure Tone Audiometry (PTA) is a behavioral test for the evaluation of the hearing sensitivity. This exam is one of the "gold standard" tests for the measuring the hearing capacity.

Since the PTA is a behavioral test, it involves some operational limitations, specially among patients with special needs or disabilities. Co-pathology is a major complication for the diagnosis of hearing problems. Almost all elderly adult will develop some degree in cognitive capacity as time progresses. This slow decline oftentimes progress into more serious conditions such as dementia or Alzheimer's disease. In these cases, the standard protocol of a PTA becomes unenforceable since the interaction between the audiologist and the patient is almost impossible. Since aging is highly related to both hearing loss and age related cognitive decline, the coexistence these two conditions is substantially likely, and it represents a challenge for the audiologists. With these specific group of patients, the audiologist needs to focus his attention on unconscious eye gestural movements that indicate some kind of perception. The proper detection of these spontaneous reactions requires broad experience since each patient may show different gestures as a reaction All the subjectivity involved make of this evaluation an imprecise problem, difficult to reproduce and very prone to errors.

In [3] an initial approach was proposed to provide an automatic solution for the detection and identification of these eye gestural reactions. This first approximation was developed in order to confirm the viability of the proposed methodology but, as pointed out in the conclusions, more studies need to be conducted in order to obtain a fully automated computer-driven system capable of solving this specific problem. The experimental results of this initial and novel solution showed the possibility of reliably distinguishing between the different categories of eye movement established by the experts, however, these categories have later been extended which makes necessary the development of new experiments.

It is important to note that, due to the nature of these patients and the features of the eye gestural reactions that they show, a typical method for the detection and classification of gestural reactions is not applicable. In this case, the reactions manifested by the patients are fully opened and can not be stereotyped into typical gestures associated to the classical emotions.

The scope of this paper is to extend and corroborate the results previously obtained and to define the complete methodology that receives as input a video sequence recorded during the performance of an audiometric evaluation and obtains as result the identification of the precise moments when an spontaneous eye gestural reaction may have happened. The initial proposal classified the detected eye moments into one of the established categories, but in order to obtain a fully automated solution it is necessary to give meaning to these movements and identify those that really correspond with a reaction to the sound.

The remainder of this paper is organized as follows: Section 2 is devoted to explain the clinical protocol.Section 3 explains the methodology for the automatic detection of the eye gestural reactions. Section 4 shows the experimental results. And finally, conclusions and future work lines are presented in Section 5.

2 Clinical Protocol for Pure Tone Audiometry

As mentioned before, the Pure Tone Audiometry (PTA) is the standard test for the evaluation of the hearing capacity. It allows the audiologist to determine the hearing threshold levels of the patient, and thus, to determine if hearing loss exists. The setup is this test is quite consistent (see Fig. 1): the patient is seated in front of the audiologist wearing earphones connected to the audiometer. Through this device, the audiologist delivers pure tone sounds at different frequencies and intensities. In the case of patients without any communication disorder, they are asked to raise their hands when they perceive the sound.

Fig. 1. Typical setup of the video sequences

There is a need of communication between the audiologist and the patient. Firstly, the audiologist needs to explain to the patient the protocol for the audiometric test, and it is highly important that the patient understands the instructions given. Also, during the performance of the test, the patient needs to indicate when he perceives the auditory stimuli, so there is a question-answer communication during all the procedure. This need of communication is what makes unenforceable the assessment of patients with a high degree of cognitive impairment. The evaluation of this patients is much more complex, but it is still possible if the audiologist is experienced enough and he focuses his attention of unconscious and spontaneous eye gesture reactions that represent some kind of perception for these patients, and they can be considered as a response to assess the hearing. The proper assessment of these patients is important, since the detection of hearing loss may allow its correction by the use of hearing aids, thus, achieving a decrease in the feeling of isolation of these patients.

3 Analysis of Eye-Based Gestural Reactions

The development of an automated solution for the identification of eye gestural movements as a reaction to the auditory stimuli may be of great relevance between the audiologist community. In order to analyze the assessments, the audiometries are recorded using a video camera located behind the audiologist. This way, the recorded video sequence has a viewpoint similar to the one observed by the expert during the test. This location ensures recording the audiometer (which is going to be necessary in order to correlate the stimuli and the reactions), and also that the face of the patient is recorded in frontal position.

Using computer vision techniques over the video sequences we propose a methodology for the detection and interpretation of the eye gestural reactions. The main scheme of this methodology is detailed in Fig. 2. The more complex stage is the *Eye Movement Classification*, which is presented more in detail in Fig. 4. Each stage is going to be explained next.

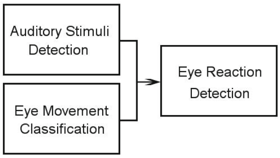


Fig. 2. General representation of the methodology

3.1 Auditory Stimuli Detection

In order to correlate the reactions with the delivery of the auditory stimuli it is necessary to know when an stimulus is being sent. In our particular case, experts are working with analogical audiometers, so this information can not be automatically extracted. It is necessary to develop a solution that analyzes the video sequence and provides this information. There are two devices employed by the experts, one of them (Fig. 3(a)) has a single light that turns on when the auditory stimulus is being sent; and the other one (Fig. 3(b)) has two lights, one light represents the left channel, and the other one the right channel.

Template matching is used to locate this light indicators, after that, in the HSV color space H and S components are thresholded in order to determine whether the lights are on or off, and consequently, if an auditory stimulus is being delivered or not.

(a) (b)

Fig. 3. Analogical audiometers: (a) Beltone Electronics and (b) Madsen Otometrics

3.2 Eye Movement Classification

The main stage of the methodology is the *Eye movement classification*. It analyzes the video sequences in order to detect eye movements and classify them into one of the categories established by the audiologists. The main steps of this stage are detailed in Fig. 4.

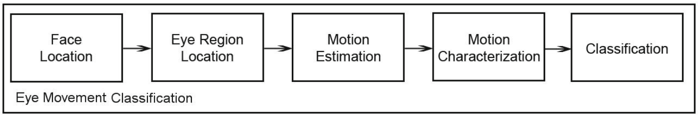
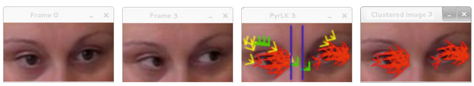


Fig. 4. Main steps of the *Eye Movement Classification*

Face Location. An initial face location facilitates the subsequently step. Due to the stability of the domain, it is possible to ensure that faces will always be in frontal position. Thanks to this particularity, the Viola-Jones [4] face detector is highly appropriate, since it is low computational cost and very accurate.

Eye Region Location. In order to obtain the location of the eye region, we specifically trained a Viola-Jones cascade. This new detector was built using more that 1000 images of the eye area.

Motion Estimation. Global movement analysis is now applied in order to estimate the movements produced within the detected eye region. The use of a global viewpoint justified in previous works. The motion is estimated by applying the iterative Lucas-Kanade [5] optical flow method with pyramids [6]. Since our frame rate is 25 FPS, the optical flow is computed between frame i and frame $i+3$. In Figure 5 a sample of the optical flow results is showed. Fig. 5(a) and Fig. 5(b) are the images to be compared and Fig. 5(d) represents the final vectors to be considered in the next step of the methodology.



(a) (b) (c) (d)

Fig. 5. Sample optical flow images. Optical flow is calculated between (a) and (b). Optical flow results in (c): green vectors for the softer movements, yellow for intermediate and red for strongest movements. Stronger vectors in (d) after filtering the rest.

Motion Characterization. Once the significant movement are detected, it is necessary to characterize them in order to lately classify them. When no significant movement occurs, the classification does not take place, but in other case it is necessary to have some descriptors that allow the movement categorization. To that end, a set of descriptors is going to be extracted.

The movement features considered for the descriptors are: orientation, magnitude and dispersion. Orientation provides information about changes in the gaze direction. Magnitude provides information about the intensity of the movement. Dispersion allows to discriminate between localized an global movements. According to this, the descriptor is comprised of a vector of 24 values: 8 for orientation, 8 for the average length for each orientation and, 8 for the dispersion. These descriptors are described with detail in previous works.

Classification. After new meetings with the experts, six typical movements are considered: no significant movement (Class NM), eye opening (Class EO), eye closure (Class EC), gaze shift to the left (class GL), gaze shift to the right (Class GR) and global movement (Class GM). It must be noted that a new category was introduced since the previous version of this work: Class NM (no significant movement), in order to classify vectors that represent a slight movement which should not be considered as a relevant reaction. More samples are included in this new training and also a new classifier has been tested obtaining better results. The results of this new training are detailed in Section 4.

3.3 Eye Reaction Detection

Finally, after detecting the auditory stimuli delivery and having all the movements classified, it is necessary to identify those movements that correspond with a reaction to the stimuli. To that end, both information are correlated and we consider as a positive reaction any significant movement produced after the beginning of an auditory stimuli. To consider a significant movement as a positive reaction it has to last at least two frames, since it is considered that with a frame rate of 25 FPS a one frame reaction would be too fast to correspond with a real positive reaction. In Fig. 6 it can be observed a visual example where, on the one hand the auditory stimulus has been detected, and on the other hand, an eye movement has been detected and classified as relevant moment (Class GR). The result of the conjunction of these two situations allows the system to determine that a positive gestural reaction has occurred. This correlation also allow to measure the reaction times, which is another important information in the assessment of the hearing.

4 Experimental Results

In order to conduct this experiment a total number of 820 movement descriptors were considered. A new classifier was included in this experiment, the Support Vector Machine (SVM). A summary of the results obtained from this new experiment is detailed in Table 1. Ten different trainings were conducted in order to generalize the results, all of them were quite consistent, thus, in order to be brief, only the results of one of them is detailed here. From these experiments it can

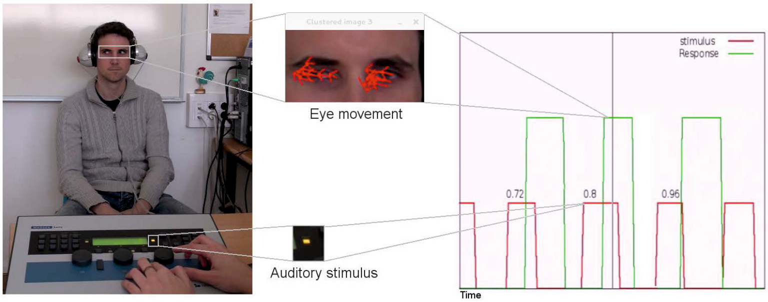


Fig. 6. Correlation between the stimulus and the reaction. Red signal for the stimuli (up when delivered) and green for reaction (up when reaction occurs). The response time is measured from the beginning of the stimulus until the beginning of the reaction.

Table 1. Accuracy of the classifiers by classes for the training dataset number 10

	Naive Bayes	Random Tree	Logistic	LMT	Perceptron	Random Forest	Random Committee	SVM
Class NM	47.1%	67.6%	55.9%	55.9%	61.8%	64.7%	58.8%	**69.1%**
Class EO	35.5%	67.1%	65.8%	67.1%	61.8%	**73.7%**	**73.7%**	73.6%
Class EC	73.7%	57.9%	76.3%	76.3%	77.6%	69.7%	75.0%	**92.5%**
Class GL	**90.6%**	73.4%	76.6%	76.6%	79.7%	78.1%	82.8%	76.9%
Class GR	64.7%	55.9%	38.2%	38.2%	50.0%	58.8%	61.8%	**85.0%**
Class GM	44.7%	80.3%	78.9%	81.6%	89.5%	86.8%	85.5%	**89.9%**
Average	59.2%	68.1%	69.2%	70.0%	73.1%	74.2%	75.6%	**81.4%**

be concluded that the new classifier, SVM, is the one which offers the higher accuracy, so it will be the one considered into our final methodology.

The final evaluation of the methodology is complicated due to the difficulties in obtaining video sequences of this particular group of patients. Most of them are entered in senior centers, and special permissions are required in order to record them. For this reason, and in order of being able to evaluate the accuracy of the proposed methodology two video sequences were recorded with two volunteers from our research group. These volunteers were instructed to reproduce the eye gestural reactions that the target patients show spontaneously.

In this experiment, the eye movements considered as a positive reaction are: gaze shift to the left and gaze shift to the right. This is because the audiologists have established that changes in the gaze direction are the most common reaction. In most of the cases, these patients are in a static attitude, and when they perceive an auditory stimulus through one if their ears, they slightly change the direction of their gaze to the side of where they have perceived the sound. This change on the gaze direction is totally spontaneous and unconscious, but our volunteers are going to reproduce it so we can test the methodology.

The correlation of the auditory stimuli delivery and the eye movement classification allow the identification of the eye movements considered as positive reactions to the auditory stimuli. The results of the evaluation of the method are depicted in Table 2. In the first experiment 31 positive reactions occur, and the method is able of reliably detecting 29 of them. In the second case, a total number of 39 positive reactions occur, and 37 of them are correctly identified. These results offer a total accuracy of the 94.21%.

It is important to note, that the previous results besides of being objectively good (almost a 95% of accuracy), are greatly important since the ground truth was obtained by labeling the images frame by frame, which makes easier the detection of the eye moments. However, in the traditional protocol, the detection of the positive reactions is conducted in real time while the audiologist handles the audiometer, what makes highly likely that the he misses any of these reactions.

Table 2. Accuracy of the methodology in the detection of eye gestural reactions

	Experiment 1	Experiment 2
Detected reactions	29	37
Lost reactions	2	2
Accuracy	93.55%	94.87%
Combined accuracy	**94.21%**	

5 Conclusions

This work evaluates the suitability of a novel approach for the detection of eye gestural reactions as a response to the auditory stimuli. It facilitates the hearing assessment of patients with cognitive decline or communication disorders that do not interact with the audiologist in the standard way. More samples, a new movement category and a new classifier were tested with regard to previous works. Due to the difficulties in the obtaining of video sequences of this particular group of patients, the evaluation was conducted with volunteers. The obtained results show an accuracy of the 94.21% for the detection of the eye gestural reactions, which is a promising result. This study demonstrates the possible contribution that this work would have for the audiologists, since nowadays not all the audiologist are trained for the evaluation of these patients and it is a very subjective task. This methodology would provide objectivity and reproducibility to the process. As future works it would be important to obtain the required permissions for recording real patients and test the methodology with them.

References

1. Cruickshanks, K.J., Wiley, T.L., Tweed, T., et al.: Prevalence of hearing loss in older adults in beaver dam, wisconsin: The epidemiology of hearing loss study. American Journal of Epidemiology **148**(9), 879–886 (1998)
2. Davis, A.: The prevalence of hearing impairment and reported hearing disability among adults in great britain. Int. J. Epidemiol. **18**, 911–917 (1989)
3. Fernandez, A., Ortega, M., Penedo, M.G., Cancela, B., Gigirey, L.M.: Automatic eye gesture recognition in audiometries for patients with cognitive decline. In: Kamel, M., Campilho, A. (eds.) ICIAR 2013. LNCS, vol. 7950, pp. 27–34. Springer, Heidelberg (2013)

4. Viola, P., Jones, M.: Robust real-time object detection. Int. J. Comput. Vision **57**, 137–154 (2004)
5. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence, IJCAI 1981, vol. 2, pp. 674–679 (1981)
6. Bouguet, J.Y.: Pyramidal implementation of the Lucas-Kanade feature tracker: Description of the algorithm. Intel Corporation, Microprocessor Research Labs (2000)