A Methodology for the Analysis of Spontaneous Reactions in Automated Hearing Assessment

Alba Fernández, Marcos Ortega, Manuel González Penedo, Covadonga Vázquez, and Luz M. Gigirey

***Abstract*—Audiology is the science of hearing and auditory processes study. The evaluation of hearing capacity is commonly performed by an audiologist using an audiometer, where the patient is asked to show some kind of sign when he or she recognizes the stimulus. This evaluation becomes much more complicated when the patient suffers some type of cognitive decline that hinders the emission of visible signs of recognition. With this group of patients, a typical question–answer interaction is not applicable, so the audiologist must focus his attention on the patient's spontaneous gestural reactions. This manual evaluation entails a number of problems: it is highly subjective, difficult to determine in real time (since the expert must pay attention simultaneously to the audiological process and the patient's reactions), etc. Considering this, in this paper, we present an automatic methodology for processing video sequences recorded during the performance of the hearing test in order to assist the audiologist in the detection of these spontaneous reactions. This screening method analyzes the movements that occur within the eye area, which has been pointed out by the audiologists as the most representative for these patients. By the analysis of these movements, the system helps the audiologist to determine when a positive gestural reaction has taken place increasing the objectivity and reproducibility.**

***Index Terms*—Auditory responses, computer vision, gestural reactions, hearing screening, movement detection, screening methods.**

I. Introduction

POPULATION aging is a demographic revolution affecting the entire world. Moreover, this longevity involves a parallel increase of the years lived with incapacity and invalidity. It has been established that one of the most common disabilities in elder people is the decrease of hearing [1], which is also one of the most widely undertreated. Hearing impairment commonly implies problems in understanding speech and communicating, which results in a feeling of progressive confinement. In turn, with age increases the possibility of emergence of neurodegenerative disorders and communication problems. The conjunction of neurodegenerative disorders and hearing loss has a negative impact on the emotional state and also on the physical and social well being of our elders [2].

Manuscript received April 2, 2014; revised September 18, 2014; accepted September 19, 2014. Date of publication October 3, 2014; date of current version December 31, 2015. This work was supported in part by the Secretaría de Estado de Investigación of the Spanish Government through the research project TIN2011-25476.

A. Fernández, M. Ortega, and M. González Penedo are with the Departamento de Computación, Universidade da Coruña, A Coruña 15071, Spain (e-mail: alba.fernandez@udc.es; mortega@udc.es; mgpenedo@udc.es).

C. Vázquez and L. M. Gigirey are with the Dual Sensory Loss Unit, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain (e-mail: mariacovadonga.vazquez@rai.usc.es; luz.gigirey@usc.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2014.2360061

Lack of hearing is one of the most frequent sensory deficits among elder population. It certainly extends to all age ranges (a number of 700 million people with hearing loss has been estimated by Davis [2] for 2015), but it is between elder people where it has a higher incidence. According to the National Institute of Deafness and Other Communication Disorders [3] almost the 50% of North Americans over age 75 years have affected their hearing. Meanwhile, the Australian Hearing Annual Report [4] states that more than half of individuals aged between 60–70 years have some hearing deficit, increasing to 70% for people with 70 years old or more. In general terms, population aging is a global reality [1]. Its onset is usually insidious but it gradually worsens.

Hearing loss is the disability more closely related to aging [2], [5]. Furthermore, recent investigations show that hearing loss is a potential risk factor for cognitive impairment [6]. This study also estimates that for 2050, 100 million people may suffer different problems related with cognitive decline. In turn, there is scientific evidence of a possible association between decreased hearing and an increase in Alzheimer's disease [6]. Degraded hearing may lead the individuals to social isolation, and this social isolation has long-term consequences to healthy brain functioning. Besides, hearing loss may also force the brain to devote too much energy on processing sound, reducing the energy spent on memory or thinking.

The use of hearing aids and hearing rehabilitation process is closely related to the development of social, emotional, psychological, and physical well being of people with hearing problems [3]. All these considerations make important the conduction of regular hearing checks.

Pure-Tone Audiometry (PTA) is the key test used to measure the hearing sensitivity. It determines the faintest tones a person can hear at selected frequencies. This test allows the audiologists to evaluate the hearing capacity and also to determine the prevalence of hearing problems. It is a subjective behavioral measurement of hearing threshold, as it relies on patient response to pure tone stimuli. Therefore, it is needed the cooperation of the patient during the test procedure, which may involve certain limitations which will be discussed shortly.

During the audiometric test, pure tones are delivered to the patient via earphones and the patient must indicate when he or she perceives the stimulus (typically, by raising his or her hand). The performance of this test is typically a completely manual method, which entails certain problems. As mentioned before, it is expected that the patient raises his or her hand to indicate the perception of the stimulus. There is a previous work [7], where we have developed a method capable of detecting the hand raising of the patients without cognitive decline (those who interact with the audiologist in the conventional way); this method provides an objective and automatic measure of the

time that it takes between the delivery of the auditory stimulus and the reaction of the patient. This measure is relevant for the identification of patients with abnormally slow response times, which can be a symptom of any medical condition.

However, in the case of patients with cognitive decline, this standard protocol becomes unenforceable since no active interaction audiologist-patient is possible. This specific group of patients has limitations when it comes to maintaining a normal interaction, limitations that are aggravated as the cognitive decline worsens.

Although the evaluation of these patients becomes much more complex, it is still possible if the audiologist is experienced enough. Whereas "normal patients" react raising the hand or with voice, patients with cognitive decline typically react unconsciously with subtle facial reactions. These facial reactions occur mainly on the eye region, so the audiologist needs to focus his attention within this region in order to detect changes in the gaze direction, eyes opening or closing, or another specific expression change that could indicate some kind of perception by the patient. It is important to emphasize that these gestural reactions are particular for each patient, even the same patient may react in different ways during the same session. This variability requires that the audiologist must have broad experience so he or she can properly detect and interpret the gestural reactions. The subjectivity involved in the gesture interpretation makes this task an imprecise problem, prone to errors, and it greatly limits the reproducibility and robustness of the measurements performed in different sessions or by different experts, leading to inaccuracies in the assessment.

All these considerations make clear the improvements that an automated solution could offer, helping the audiologists in the detection and interpretation of these gestural reactions. In this paper, we propose a novel method for the analysis of the eye movement specifically designed for this field. This proposal makes use of computer vision techniques in order to analyze video sequences recorded during the performance of the audiometry. The methodology needs to detect the patient, locate the eye region, and be able to detect movements produced within the eye region.

In recent years, research efforts seeking methods based on computer vision for the analysis of human motion or the recognition of human gestures have gained growing interest. These solutions can be applied in a wide range of applications such as video surveillance, human–computer interaction, human performance analysis, virtual reality, clinical studies, ambient intelligent systems, and so forth.

Regarding to the audiometric domain, there is no computer vision solution for supporting the audiologists in the evaluation of patients with cognitive decline given the particular and specific characteristics of these gestures as discussed before, to the best of the authors' knowledge.

Although we are interested on detecting facial reactions, a typical method for the detection and classification of gestural reactions is not applicable on this domain. This is because this kind of methods are based on gestural reactions typically associated with known expressions or human emotions (such as happiness, anger, surprise, disgust, etc.) proposed by Ekman [8]. In our case, the gestural reactions manifested by the patients do not correspond to any of these emotions, they are fully opened and they cannot be stereotyped into particular emotions. They are more subtle and can be highly variable depending on the patient, they may even be variable for a same patient during the test. This is why, a typical solution for facial expression recognition such as [9] or [10] is not applicable on this domain.

Fig. 1. Sample of the different eye movements target of detection.

The main challenge of this study is the identification of gestures associated with reactions to the sound, which are totally dependent of the patient. In most cases, reactions are associated with changes on the gaze direction, some images samples can be observed on Fig. 1. From these images, it can be inferred that reactions can be more subtle or marked depending on the patient, in addition, sometimes, the presence of wrinkles or the absence of eyebrow modify the appearance and the features of the area, and also changes in the illumination or other lighting conditions may affect the process. To solve this problem, a novel approach based on the optical flow method is proposed; orientation, magnitude, and dispersion are the main features of the optical flow considered to characterize the movement. In this proposal, eight video sequences are considered for training and five for validation. Although there are more than 150 video sequences, the number is patients with communication disorders between this population is quite low, which limits our video dataset. Anyway, the available sequences cover different degrees of cognitive decline, which implies a high representation of the real conditions. To the best of the authors' knowledge, this problem has never been attempted to address through a computational solution, which may be very helpful for the audiologists.

The remainder of this paper is organized as follows. Section II introduces the traditional protocol. Section III is devoted to explain the methodology. In Section IV, we show and discuss the experimental results obtained for the evaluation of the methodology. Finally, Section V provides some discussion and conclusions.

II. Audiometric Process

As mentioned in the Introduction, the conduction of regular hearing checks is specially important for patients with more than 60 years old. The communication difficulties resulting from hearing loss can cause isolation and frustration. Research indicates that diagnosed patients who wear hearing aids are less affected by depression and have improved health. It is also recommended to undergo a hearing test if any problems in hearing or communication are detected. Next, we detail the clinical procedure to perform a PTA.

A. Clinical Protocol for PTA

PTA is the standard test to identify the hearing threshold levels of an individual. It allows the audiologist to diagnose the

presence or absence of hearing loss by determining the softest sound that can be perceived in a controlled environment. PTA is a behavioral test, where the patient wears earphones connected to an audiometer, so auditory stimuli can be delivered to each ear. These auditory stimuli are pure-tone sounds at different frequencies and intensities. Cooperation is needed during the test procedure, since the patient taking the test is typically asked to raise his hand when he or she perceives the stimulus. This is why, the evaluation of patients with cognitive decline becomes much more complex as mentioned in Section I The results of hearing sensitivity are plotted on an audiogram, which is a graph displaying intensity as a function of frequency. The frequencies commonly tested are 100, 250, and 500 Hz, and 1, 2, 4, and 8 KHz, and the intensities commonly plotted range from −10 to 110 dB, in multiples of 10. The range between 100 and 8 KHz represents the most important levels for clear understanding of speech. The results of the audiometric test determine the subject's hearing levels. Normal conversation speech is about 45 dB. Normal hearing is expected to be between −10 and 20 dB. According to the results, hearing can be classified in: normal hearing, mild hearing loss, moderate hearing loss, moderately severe hearing loss, severe hearing loss, and profound hearing loss.

Prior to any exploration, an otoscopic examination is performed to check the absence of any obstruction (e.g., earwax) in the outer ear canal, because obstructions can interfere to the hearing capacity and must be removed. In the case of excessive cerumen in the ear canal, the patient must be referred to the appropriate specialist and the audiometry must be rescheduled. This otoscopic examination also allows to determine if the eardrum presents any damage that can reduce hearing, such as perforations in the eardrum or congenital malformations.

Next, the audiologist explains to the patient, the protocol for the audiometric test. Since it is a behavioral test, it is important that the patient understands the instructions given. This need of understanding in the communication is what makes difficult the assessment of patients with cognitive decline or patients with a profound hearing loss without hearing aids. For patients without impairments, the audiologist indicates to them that they are going to receive different types of auditory stimuli via earphones and they must respond to them affirmatively, usually by raising their hands, when they perceive them. Since each ear is evaluated separately, patient must respond consistently, by raising his or her right hand, when he or she perceives the auditory stimuli on the right ear and equivalently for the left ear.

The performance of the PTA allows the audiologist to evaluate air and bone conduction. For the air conduction audiometry, the patient wears earphones and the results establish the extent of sound transmission through the bones of the middle ear. For bone conduction, patient wears a vibrating earpiece behind the ear next to the mastoid bone. This bone vibrator uses the skull to transfer vibrations cochlear and bypasses the outer and middle ear. Results of bone conduction determine the extent to which there is neurosensory hearing loss.

The auditory stimuli are sent through an audiometer, where the audiologist sets the different frequencies and intensities that he or she wants to test. Once the expert has selected the frequency and the intensity, the audiologist sends the auditory stimuli to the patient and waits for his or her reply. If the patient is able to perceive the stimulus he or she typically will raise his or her hand; in the event that the patient does not respond, the audiologist can try to send the same stimulus again or to increase the intensity and send a new stimulus. In the case of patients with cognitive decline, the communication process will be more complex, but the handling of the audiometer and the delivery of the auditory stimuli remain the same.

An initial framework for automatizing audiometries has been introduced in [7]. This framework automatically detects stimuli and hand gestures on patients to obtain precise measurements of reaction times. Using the same audiometric process, the present study will deal with spontaneous eye-based gestures to extend the system for patients with cognition or motion conditions unable to raise their hand accordingly. In this new approach, specifically aimed patients with severe cognitive decline, the hand raising does not occur or it is not relevant (as referred by the audiologists), so all the attention is focused only on the eye movements.

The proposed method for the analysis of the eye-based gestural reactions, addressed in the next section, fully respects this clinical protocol, without involving changes in the behavior of the patient or the audiologist. This is an advantage since the only need is to record the audiometries with a video camera, but the rest of the audiometric protocol is not influenced.

III. Analysis of Eye-Based Gestural Reactions

As mentioned in Section I, an automated solution for the detection and interpretation of the gestural reactions could be of great relevance for improving the objectivity and repeatability in the evaluation of these patients.

By the application of computer vision techniques on the video sequences recorded during the performance of the audiometries, we have developed a method aimed to support the audiologists in the detection of eye-based gestural reactions as a response to the auditory stimuli. A schematic representation of the main steps of this method can be seen in Fig. 2. Next sections discuss each one of the stages showed in the schema.

A. Face Location

Proper face location will allow us to narrow the search area of the eye region. This first location reduces the computational cost of the next step and makes it less error prone.

Although the location of a face is a natural process for a human being, it becomes a challenging task in computer vision. In addition to the inherent complexity of defining a face for a computer, the variations in scale, orientation, pose, facial expression, lighting conditions, and background, increase the complexity of the problem.

Different methods have been developed in order to detect faces in a scene. First, easiest proposals worked with monocolor backgrounds or a predefined static background, in these cases, the face was obtained by removing the background. Other approaches use color information as base, seeking for skin color regions [11]. The use of color as the main base of the method may involve limitations with some skin colors or with varying lighting conditions. If we work with video sequences instead of

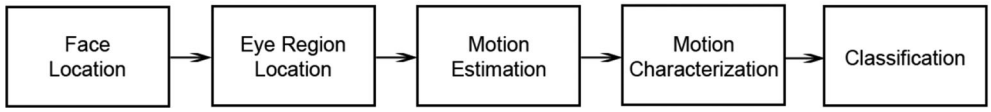


Fig. 2. Schematic representation of the methodology.

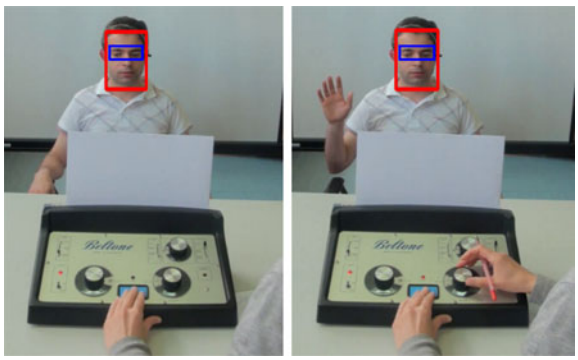


Fig. 3. Face detection at different times during the test.

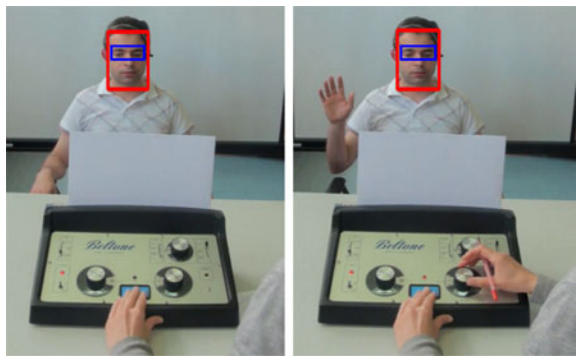


Fig. 4. Eye detection at different times during the test.

static images, we can use motion information to find the face [12], considering that face is almost always moving. The problem with motion information arrives when we have nonstatic objects in the background. There also exist many other different alternatives: using feature analysis, active shape models, neural networks, etc.

The domain is very stable in terms of location; the audiologist is always seated in front of the patient and the video camera is located behind the audiologist to ensure that the patient's face will always be recorded in frontal position. This particular setup can be observed in Fig. 3. The certainty of having a frontal position of the patient's face allows us to apply the Viola and Jones approach [13]. The Viola–Jones detector is a general object detection framework, which provides competitive object detection rates in real time. It can be trained to detect a variety of different objects; however, its initial motivation was to provide a solution to the face detection task. As consequence of this, an optimized classifier for the face detection was obtained. Particularly, a classifier for the detection of frontal faces is available in the OpenCV library. This classifier is not as flexible as other approaches, but it is low computational and very robust for the detection of frontal faces, so it is a good solution for this domain.

It is important to note here that, in this stage of the methodology, the video sequences are going to be processed as individual images, frame by frame. Since in our case, we are working with high-resolution video sequences, the corresponding frames are going to be scaled for subsequent processing. This way, the resolution of the image is reduced to reach a compromise size, which allows us a fast processing without loss of information. Over this reduced image, we apply the Viola–Jones face detector provided by the OpenCV library. Samples of face detections can be seen in Fig. 3. With this face location, the method retrieves the face with the original resolution and uses it as entry for the next step.

B. Eye Region Location

Once the search area has been limited to the face region, the next step is the location of the eye region. We could have considered the detection of the gestural reactions all over the face, but the extensive experience of our audiologists with this type of patients allow them to claim that the gestural reactions that really correspond to responses to the auditory stimuli most prominently occur within the eyes region. This statement allows us to limit the movement analysis to this particular area and work without considering the other movements that may occur in the rest of the face which could lead to confusion.

Eye detection can be broadly divided into three types: template-based, feature-based, and appearance-based methods. For example, in [14] deformable templates are used to extract the eye boundaries. A sample of the second group of methods can be found in [15], where blinks are detected based on differences between successive images. The appearance-based ones can be integrated with machine-learning techniques and have been widely developed by the research community during the recent years. Some representative algorithms can be found in [16], where a neural-network-based approach is proposed, and in [17], where a SVM is applied.

For the location of the eye region, a cascade was specifically trained for this study using more than 1000 images of the eye area. Each one of this 1000 images was manually selected, since there was no training image database for this specific region. The training images where cropped from different face images from different face databases. An accuracy of the 98% was obtained for this eye detector. It is capable of detecting the eye region regardless of the expression and even when the eyes are closed, which is a relevant feature given the unconstrained and unpredictable gestures and expressions of target patients. Samples of eye region detections can be observed in Fig. 4.

With the aim of facilitate subsequent steps, it was established that the eye regions captured during an audiometric evaluation must have the same size. Since the Viola and Jones object detector does not fulfill this condition, a later correction is required. To that end, a fixed size is established based on the measurements of the first location. The subsequent locations are going to be scaled to this fixed size.

(a) (b)

Fig. 5. Eye region detection. (a) Without considering cross correlation. (b) Applying cross correlation.

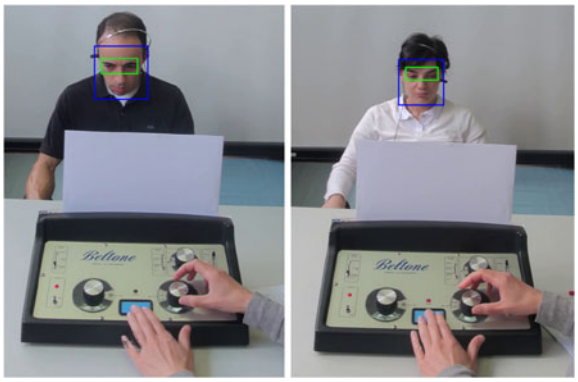


Fig. 6. Eye region detection samples.

Although the Viola–Jones locations are fairly stable, there might be a small displacement of a few pixels between locations of consecutive frames. Even though this displacement is almost nonsignificant to the human eye, since the aim of this methodology is the analysis of the movements within these regions, it may introduce noise to the results. To solve this, cross correlation between images is calculated.

The cross correlation between images calculates the greater similarity R of a template T inside an image I, according to the classical (1). In this case, the template corresponds with the eye region located in the previous frame, and the image used for the correlation is based on the current location of the eye region slightly enlarged

$$R(x,y) = \frac{\sum_{x',y'}(T(x',y')I(x+x',y+y')}{\sqrt{\sum_{x',y'} T(x',y')^2 \sum_{x',y'} I(x+x',y+y')^2}}. \quad (1)$$

The difference between applying this optimization and not can be observed in Fig. 5. In Fig. 5(a), the cross correlation was not applied, so by the overlap of the images it can be observed that there exists a displacement of several pixels; however, when the cross correlation is applied, the displacement is almost nonexistent as it can be seen in Fig. 5(b). A few more samples of eye detection are shown in Fig. 6.

C. Motion Estimation

After the eye region location, this step is aimed to start the detection and analysis of the movements or expression changes that occur within this particular area. Due to the nature of the problem, movements are analyzed in a global sense, so a classic point to point feature registration (such as in [18]) is less effective, when the expected set of movements cannot be initially expressed as a function of particular point in the ROI. Besides,

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

Fig. 7. Sequence of consecutive eye region locations. For $t = 3$ the optical flow would be performed between (a) and (d).

(a) (b) (c)

Fig. 8. Motion estimation with optical flow. (a) Detected reference points in frame i. (b) Optical flow results with the location of the reference points over the frame $i + 3$. (c) Motion vectors (from blue to red).

each individual may show different gestures as a reaction and even the same patient can acts erratically showing different movements along the audiometric test. Therefore, a template analysis (e.g., [19], [20]) is not possible either, since the gestures of these patients are erratic and they do not correspond to any typical gestural expression.

In order to address this problem, a novel approach specifically aimed to this domain and based on global movement (GM) analysis for description was proposed. By the evaluation of the domain and the features of the images to be treated, it was decided to analyze the optical flow between eye region images. The motion is estimated by applying the iterative Lucas and Kanade [21] optical flow method with pyramids [22]. Optical flow has shown optimal results in the identification of general and unconstrained movements produced by expression changes, as in [23], where an interactive photoframe analyzing activity and facial expressions was developed.

It must be noted at this point that the video sequences considered for this research have a frame rate of 25 frames per second (FPS). This is a fairly high rate, so comparisons between a frame and the next one may not show changes notable enough, because expression changes cannot occur so quickly. With the purpose of allowing expression changes significant enough, we consider a time window (t) between frames, i.e., optical flow is computed between frame i and frame $i + t$ (see Fig. 7). If t is too high, small movements will not be detected. So t must be chosen as a tradeoff between ignoring irrelevant movements and not losing relevant movements. Empirically it was calculated that most of relevant movements for this domain can be detected with $t = 3$.

The optical flow operation is based on the detection of interest points. Traditionally, the interest operator associated with the optical flow is Good Features to Track [24]. Usually, this interest operator is applied over the first reference image and the obtained interest points are used for all the images in the video sequence. In this case, this behavior was modified, so that the interest operator is applied over the first frame of each comparison. This is done because changes in the eye expression (e.g., open eyes versus eyes closed) may highly affect the amount of detected interest points. A sample of the application of the interest operator and the optical flow can be observed in Fig. 8.

Fig. 9. Sample of movement vectors represented as arrows for different eye movements. (a) Gaze shift and (b) EO.

Fig. 10. Movement vectors ranked by magnitude. (a) Reference frames, (b) frame to be compared, and (c) shows the ranked movement vectors: green for v_{short}, yellow for v_{interm}, and red for v_{long}.

Fig. 8(a) represents the interest points detected by the interest operator (i.e., Good Features to Track) over the reference frame *i*. Fig. 8(b) shows the correspondence of the interest points located by the optical flow over the second image $i + 3$. Finally, Fig. 8(c) shows the motion vectors with origin at the interest points in frame *i* (represented in) and end at their corresponding points in frame $i + 3$ (represented in red).

Since vectors in Fig. 8(c) represent the direction and the amount of movement, this representation can be modified by showing arrows instead of vectors, where the arrow for a particular point represents its movement from the initial time considered to the final one. Fig. 9 shows a couple of samples with this type of representation. In Fig. 9(a), the direction of the gaze changes, the optical flow is able to detect this movement and gives as a results vectors pointing to the side. In the case of Fig. 9(b), the eye opening (EO) increases, so vectors are pointing up, properly representing the movement produced.

In order to adapt the obtained results to this specific domain and considering only the significant vectors for the subsequent steps, some optimizations are conducted.

1) Nonsignificant Vectors Removal: Since every movement is detected regardless of its strength, it can be considered that small movements should not be considered. This approach removes the nonsignificant vectors in order to only consider the vectors that really correspond with a significant movement and, thus, facilitate the movement classification.

Movement vectors are ranked according to their magnitude into three different classes. This clustering was done empirically after evaluating the movement vectors for this domain. Since the eye region size has been fixed at the beginning of the procedure, the thresholds can be normalized according to these proportions. Equation (2) shows the established thresholds for an eye region size of 62×115 pixel, and Fig. 10 shows a sample of this classification. Vectors labeled as v_{short} are considered too small to be significant and will be removed, in Fig. 11 it can be observed this situation. The second class v_{interm} contains those vectors with an intermediate length that does not always correspond

Fig. 11. Movement vectors ranked by magnitude. (a) Reference frame, (b) frame to be compared, and (c) shows the ranked movement vectors.

Fig. 12. Correction for too long vectors. (a) and (b) are the frames to be compared. (c) Shows the movement vectors. Vectors in gray are removed due to their excessive size.

with relevant movements, therefore, in principle, they are not considered. Vectors in v_{long} have a length significant enough to always correspond to significant movements, so they are the vectors considered for the next stages of the methodology. It can be observed that there is a limit for vectors in v_{long}; as occurs with too small vectors, very long vectors must be removed. These vectors are usually related to inaccurate associations or interest points that do not appear in the second image (these behavior can be observer in Fig. 12)

$$\text{vector classification} \begin{cases} 0px \leq v_{\mathrm{short}} \leq 1.5px \\ 1.5px \leq v_{\mathrm{interm}} \leq 2.5px \\ 2.5px \leq v_{\mathrm{long}} \leq 13px. \end{cases} \tag{2}$$

2) Discarding the Displacement Component: It can occur sometimes that the detected motion is due to global motion between the two images instead of motion within the region. This global displacement implies a significant number of vectors with the same strength and direction. In order to correct the displacement component of the image, the number of equal vectors (both in angle and magnitude) was considered. This value is defined as

$$C_{\theta,m} = \{v \in C \mid \theta_v \simeq \theta \ \wedge \ |v| \simeq m\} \tag{3}$$

where $C_{\theta,m}$ will contain the set of vectors with a similar angle (θ) and magnitude (m). v represents the vector, C represents the entire set of vectors, θ_v is the angle of the vector, and $|v|$ is the magnitude of the vector.

Between all the $C_{\theta,m}$, the one with a higher number of elements is chosen [following (4)], since if there is a global displacement this occurs only in one direction

$$C_{\mathrm{mode}} = C_{\theta,m} \mid \forall \theta', m', \ \theta' \neq \theta \vee m' \neq m, \ |C_{\theta,m}| > |C_{\theta',m'}| \tag{4}$$

where θ' and θ are the angles and m' and m are the magnitudes.

To consider a global displacement, a high number of vectors with the same angle and magnitude is required. This is established with

$$C_{\mathrm{mode}} \geq |C| \cdot \lambda \tag{5}$$

Fig. 13. Correction for displacement vectors. (a) and (b) are the frames to be compared, (c) shows the movement vectors, which point up when there is no real movement, and (d) shows the vectors after the correction (vectors in gray are the discarded vectors).

where λ is a parameter that sets the limit for discarding vectors depending on the number of vectors in C_{mode}.

When a global displacement is detected, the removal of the displacement vectors is not enough as it is also necessary to correct the remaining vectors. To that end, a subtraction of the vectors is computed, where the displacement component is subtracted to the remaining vectors. The consequences of this optimization can be observed in Fig. 13, where it can be noted that after the correction of the displacement vectors no significant movement is detected.

D. Motion Characterization

The information provided by the optical flow serves as basis to characterize the produced movements. Movement detection allows the identification of those instants during the process, where the patient shows a sign of perception. Since each patient is going to react differently to the auditory stimuli, a classical solution for the global characterization of the facial expression changes is not applicable. Considering this, the proposal relies on the detection of basic gestures within the detected ROI. So that the aim is to process the complete video sequence by the analysis of the optical flow, in order to detect where the significant movements occur.

In order of being able of reliably distinguishing the patient's movements it is needed to characterize the movement when it occurs using as base a group of properties associated with movement. It is necessary to find a set of features that describe the movement that shows the patient as a sign of perception, in such a way that all these movements are equally described. This is an important contribution of this study, since it enables the possibility of modeling any spontaneous movement from the patients in a compact and homogeneous feature space, which allow the subsequent analysis of these in a formal and repeatable way. When no significant movement occurs, the classification is not applicable, so it is not needed to characterize the optical flow in these time intervals. Instead, when a significant movement occurs, it needs to be characterized. With the aim of capturing all those features that are relevant for the motion characterization, we propose a descriptor based on some features that are going to be detailed next.

These properties try to group all vectors obtained after the previous step. The considered properties are: orientation, magnitude, and dispersion. A general idea of this feature extraction can be observed from Algorithm 1.

First, each eye is considered separately, so that each detected movement generates two movement descriptors, one for the right eye and the other one for the left eye. When generating

Algorithm 1: Optical flow feature extraction

Data: $iSamples$ which contains the optical flow vectors
Result: $histOrient$, $histMagnit$ and $histDist$
for $i = 0$ *to* $iSamples.size()$ **do**
  $p \leftarrow iSamples[i].p$; ▷ p is the origin
  $q \leftarrow iSamples[i].q$; ▷ q is the end
  $angle \leftarrow iSamples[i].angle$; ▷ angle in degrees
  $histOrient[ceil(angle/45) - 1] + +$;
  $histMagnit[ceil(angle/45) - 1] + = euclideanDistance(p, q)$;
  $histPos[ceil(angle/45) - 1].pushback(q)$;
for $i = 0$ *to* $histOrient.size()$ **do**
  if $histOrient[i] \neq 0$ **then**
    $histMagnit[i] = histMagnit[i]/histOrient[i]$;
for $i = 0$ *to* $histPost.size()$ **do**
  $centroides[i] \leftarrow$ compute centroid as the average;
for $i = 0$ *to* $histPost.size()$ **do**
  for $j = 0$ *to* $histPost[i].size()$ **do**
    $distances[i] \leftarrow distances[i] + euclideanDistance(centroid[i], histPos[i][j])$;
  $histDist[i] = distances[i]/histPost[i].size()$;

a movement descriptor, one of the relevant features is the orientation. The vector orientation provides information about the direction of the movement produced in the eye region, this orientation is different for a change in the gaze direction than for a movement of eye closure (EC). For the definition of these descriptors, vectors are divided in eight different ranges according to their angle. This classification can be represented mathematically as

$$R_i^* = \left\{ v \in C_f^* \mid \theta_v \in [45 \cdot i, 45 \cdot (i+1)] \right\} \tag{6}$$

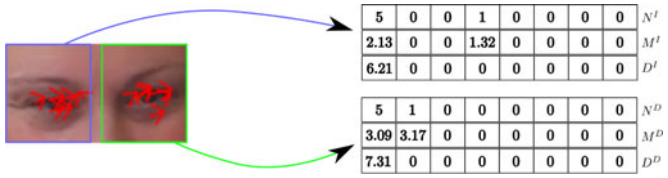
where $* \in \{L, R\}$, indicating the differentiation between the left and the right eye, and i takes values from 0 to 7. This way, vectors are grouped according to their angle, and the eight first values of the descriptor correspond to the number of vectors in each range as

$$n_i^* = |R_i^*|. \tag{7}$$

It is also important to know the vector's magnitude, because that feature provides information about the intensity of the movement. Considering this, the next eight values of the descriptor are associated with the vector's magnitude. With the vectors grouped by ranges, the average of the module of the vectors is calculated according to (8). This feature provides information about the intensity of the movement, allowing to distinguish between strong and soft movements

$$m_i^* = \frac{1}{n_i^*} \cdot \sum_{v \in R_i^*} |v|. \tag{8}$$

Finally, the dispersion of the optical flow vectors contributes with other eight values to the descriptor. The dispersion of the optical flow allows us to discriminate between localized and GLs. The computation is considered by range, it means, accord-



5	0	0	1	0	0	0	0	N^I
2.13	0	0	1.32	0	0	0	0	M^I
6.21	0	0	0	0	0	0	0	D^I

5	1	0	0	0	0	0	0	N^D
3.09	3.17	0	0	0	0	0	0	M^D
7.31	0	0	0	0	0	0	0	D^D

Fig. 14. Descriptors sample. Left image shows the movement vectors for each eye and right tables the corresponding descriptors for each eye. First row represents orientation, the second one magnitude, and the last one dispersion.

Fig. 15. Beltone electronics audiometer used in our audiometry sequences.

ing to the angle of the vectors. From each one of the vectors ($v = \overrightarrow{AB}$), the destination point is taken $B = (B_x, B_y)$, and the center of them is calculated according to (9) locating the centroid

$$c_i^* = \left(\frac{1}{n_i^*} \cdot \sum_{v=\overrightarrow{AB}, v \in R_i^*} B_x, \frac{1}{n_i^*} \cdot \sum_{v=\overrightarrow{AB}, v \in R_i^*} B_y \right). \quad (9)$$

Once the centroid is calculated, the dispersion is computed through the calculation of the average distance to that center, according to

$$d_i^* = \frac{1}{n_i^*} \cdot \sum_{v=\overrightarrow{AB}, v \in R_i^*} d(B, c_i^*) \quad (10)$$

where $d(p, q)$ is the Euclidean distance between p and q.

With all this, the descriptor is comprised of a vector of 24 values, where the eight first correspond to the orientation, N^*, the eight next are related to the magnitude, M^*, and the last eight provide the information about dispersion, D^*, as

$$N^* = \{n_i^* | i \in \{0...7\}\} \quad (11)$$

$$M^* = \{m_i^* | i \in \{0...7\}\} \quad (12)$$

$$D^* = \{d_i^* | i \in \{0...7\}\}. \quad (13)$$

A sample of these descriptors can be observed in Fig. 14.

Once the vector descriptors are computed for each movement, the next step is the classification of these, according to the movement classes determined for this domain.

E. Classification

The last step of the methodology is the classification of the descriptors. It must be noted that this classification is not strictly required, different alternatives could be attempted. This methodology allows to cover many more problematic, but considering that the experts tend to classify the type of movements that the patients show as a reaction, in this first approach, we are going to handle the problem as similar as possible to how the experts do it.

At this point, the different movement descriptors are associated with the different movement categories established by the audiologist for this domain. After reaching an initial consensus with the experts, five typical movements where identified as the most relevant: EO, EC, gaze shift to the right (ER), gaze shift to the left (GL) and GM. Also, an extra category was included to categorize those descriptors corresponding to small or insignificant movements.

In this stage, classification is conducted independently for each eye, meaning that the descriptor of one of the eyes can be classified as "EC," whereas the descriptor of the other eye can be classified as "ER." It must be also noted that if the obtained movement vector is totally composed by 0 values, the classification is not conducted since no movement (NM) exists.

A previous training of the classifier is needed to accomplish this step. To that end, a supervised training is conducted for several classifiers. The obtained results are shown in Section IV, obtaining as result the most suited classifier.

The different classifiers considered for this evaluation are: Naive Bayes, Random Forest, Random Committee, Logistic Model Tree (LMT), Random Tree, Logistic, Multilayer Perceptron, and Support Vector Machines (SVMs). The results for these classifiers are shown in the next Section.

IV. Experimental Results

To perform these experiments different video sequences were analyzed and the eye movements produced during them were manually labeled.

The audiologist were equipped with an audiometer Madsen Xeta from Otometrics (see Fig. 15). The auditory stimuli used were pure tone and the frequency range was between 125–8000 Hz for air conduction, and 250–8000 Hz for bone conduction. The stimulus levels can be set from -10 to 120 dB (HL) with 5dB (HL) steps for air conduction, and from -10 to 70 dB (HL) also with 5 dB (HL) steps for bone conduction.

The video sequences considered for the experiments had full HD resolution (1080 $\times$ 1920 pixels) and 25 FPS. The device used is a conventional video camera with full HD resolution and no particular hardware requiriments. The only requirement is to try to maintain favorable and constant lighting conditions in order to improve the recorded images and avoid shadows or occlusions. As mentioned before, the video sequences are focused on the patient who is seated in front of the camera. The image shows the patient waist up and also the surrounding scenario: the audiometer, the hand of the audiologist handling the audiometer, the background, etc. There is a reason for such a general scene, and it is due to the need of recording the audiometer in order to detect the moments, when the auditory stimuli are being sent, which allows to correlate this information with the eye movements.

TABLE I
CONTINGENCY TABLE FOR DETECTION.

		Labels		
		NM	Movement	
Classifications	NM	1176	12	
	Movement	42	270	
	Total:	1218	282	1500
	Sensitivity:	95.74%		
	Specifity:	96.55%		
	Accuracy:	96.4%		
	F-score:	90.9%		

Despite of the high resolution of the video sequences, the obtained eye regions do not have the same quality. This is motivated by the need of a general scene including elements other than the patient and also by changes in the lighting conditions during the audiometric evaluation. These considerations increase motion detection difficulties.

Although we have recorded more than 150 audiometric evaluations, only eight of them are suited for this initial evaluation. As mentioned in Section I these gestural reactions are very specific and they only occur when the patient suffers from cognitive decline or in the case of patients very expressive facially, otherwise they will raise their hands or respond vocally. This justifies the low number of video sequences available for this evaluation, since it is a ratio that corresponds to the percentage of people with these characteristics within a normal population. It is also important to note the difficulties for recording this specific group of patients; most of people with severe cognitive decline are entered in special centers and special permits and authorizations are needed to record them.

There are eight considered video sequences, all of them adults, male and female, with ages ranging from 45 to 85. Each of these video sequences takes between 4 and 8 min, so considering a frame rate of 25 FPS, it involves the evaluation of between 6000 and 12 000 different frames per video sequence.

The experiments detailed next show: a preliminary analysis about the quality of the movement detection; subsequently, over the classification of the detected movement a high number of different classifiers are evaluated in order to select the one that offers better results using the proposed motion descriptor; after that, applying that classifier, a more detailed analysis is conducted over specific video sequences of patients with cognitive decline. And finally, the movements are associated with the auditory stimuli in order to establish validity of characterized reactions respect to the stimuli, setting the suitability of our methodology in the target domain and purpose.

A. Quality of the Movement Detection

In order to evaluate the quality of the movement detection, a number of 1500 labeled frames were analyzed. Each one of them were processed to check if any significant movement took place. The results of the classification are detailed in Table I using a contingency table. It can be noted from Table I that most of frames are classified as nonsignificant movements; this is an expected consequence of the fact that, generally, the patients are static.

An important measurement in these cases is the sensitivity, i.e., the ability to detect significant movements when these occur. In our case, the sensitivity rate is 95.74%. Combining sensitivity with specificity (ability to detect nonsignificant movements), we obtain an accuracy of 96.4%. The F-score of our method is 90.9%.

B. Classifier Assessment in the Domain

Several classifiers are trained and evaluated for all the available video sequences in this Section. Only those frames where the optical flow detected a significant movement are considered. It is important to note that most of the frames do not show a significant movement, since, by default, the patient does not show any reaction. When a significant movement was detected by the optical flow, it was manually classified into one of the possible categories (i.e., class EO, class EC, class GL, class GR, class GM and no movement class NM)]. No other category was needed in this dataset by the experts. A total number of 820 descriptors were detected as significant movements and they were classified into one of the possible categories obtaining the distribution showed in Table II. As it can be observed in this table, the number of samples is not well balanced because changes in the gaze direction are not very common. In order to improve the training dataset, it is necessary to balance the number of samples of the different classes. To that end, for those classes with a high number of samples (class EO, class EC, and class GM), a limit number of 75 samples was established. These 75 samples were randomly selected for ten different trainings. Thus, the final dataset will be composed by 357 frames.

As previously mentioned, since to balance the training datasets 75 samples of three of the six classes (classes EO, EC, and GM) are randomly selected, it is necessary to conduct several trainings in order to obtain reliable results. Ten different experiments were conducted, where each one of them corresponds to a tenfold cross validation. Furthermore, each one of the ten training datasets was trained for each one of the eight different classifiers considered for this experiment: Naive Bayes, Random Tree, Logistic, LMT, Perceptron, Random Forest, Random Committee and SVMs. Results of this experiment are summarized in Table III. In this Table, each column corresponds with one of the eight considered classifiers, and each row corresponds with one of the ten experiments. Each cell shows the accuracy for the combination of training dataset and classifier. Finally, the last two rows show the average and the variance of the ten experiments for each classifier.

Although for reasons of space and simplicity, only a summary of the results is shown here, all the experiments were discussed in detail. In Table III, only the global accuracy of the experiment is shown; however, the accuracy was also analyzed considering the different classes in order to discuss the behavior of the classifier related to each class. It was observed that, whereas, almost all the classifiers offer balanced accuracy for all the classes, the Naive Bayes classifier provides high accuracy in the

TABLE II
DISTRIBUTION OF THE SIGNIFICANT MOVEMENTS BETWEEN THE CONSIDERED CATEGORIES

	Eye opening (EO)	Eye closure (EC)	Gaze left (GL)	Gaze right (GR)	Global mov. (GM)	No mov. (NM)
Number of samples	241	339	64	34	108	34

TABLE III
ACCURACY OF THE CLASSIFIERS FOR TEN DIFFERENT TRAININGS

	Naive Bayes	Random Tree	Logistic	LMT	Perceptron	Random Forest	Random Committee	SVM
Test 1	55.6%	73.3%	73.3%	73.6%	71.4%	77.5%	76.1%	75.8%
Test 2	55.0%	72.8%	69.7%	72.5%	72.8%	76.7%	77.2%	77.2%
Test 3	51.9%	68.9%	71.7%	73.3%	69.2%	74.2%	73.1%	75.6%
Test 4	58.6%	72.2%	73.6%	74.7%	72.5%	76.4%	76.9%	75.8%
Test 5	53.6%	70.6%	67.8%	70.6%	70.3%	74.2%	75.6%	75.8%
Test 6	55.3%	71.4%	69.7%	73.6%	72.2%	75.8%	78.6%	77.5%
Test 7	56.6%	73.6%	71.4%	72.8%	76.9%	78.1%	79.2%	79.2%
Test 8	55.3%	73.6%	67.5%	71.7%	71.7%	77.2%	76.9%	77.5%
Test 9	55.2%	71.3%	69.1%	69.6%	74.1%	77.2%	75.2%	76.3%
Test 10	59.2%	68.1%	69.2%	70.0%	73.1%	74.2%	75.6%	81.4%
Average	55.6%	71.6%	70.3%	72.2%	72.4%	76.1%	76.4%	77.2%
Variance	4.55	3.78	4.54	2.94	4.52	2.20	3.10	3.40

Last two rows show the average and the variance for each classifier.

TABLE IV
ACCURACY BY CLASSES OF NAIVE BAYES FOR DIFFERENT EXPERIMENTS

	Test 3	Test 5	Test 7	Test 9
Class NM	0.529	0.529	0.529	0.471
Class EO	0.303	0.329	0.408	0.276
Class EC	0.632	0.697	0.776	0.789
Class GL	0.859	0.859	0.859	0.891
Class GR	0.647	0.647	0.618	0.618
Class GM	0.276	0.263	0.263	0.307
Global	51.9444%	53.6111%	56.6667%	55.1532%

classification of class GL, but very low for classes EO or GM. This behavior can be observed in Table IV, where the accuracy by classes is detailed for several experiments. From the global results in Table III, it can be concluded that Naive Bayes is the worst classifier in terms of accuracy, but even if this did not happen, it would not be a valid classifier due to the imbalance of the different classes. The problem of imbalance does not happen to the other classifiers, so they can be considered.

Analyzing the global results from Table III, it can be observed that the best results are obtained with the Random Committee and the SVM classifiers. The average accuracy is better for the SVM classifier, whereas the variance is better for Random Committee. Although the variance for SVM is not the minimum, it is one of the lowest values and, thus, it is acceptable. These results only show the global accuracy of the classification, so in order to evaluate the classification by classes and check for imbalance in Table V the accuracy by classes is detailed for the experiment number 10.

As it can be observed from Table V, no major imbalances occur neither for Random Committee nor for SVM. So going back to the main table (see Table III) and analyzing the obtained accuracies, it can be observed that the best accuracy is obtained for the combination of experiment number 10 and the SVM classifier (accuracy of 81.4%). So this combination is selected, as the most suited for the classification task and it will be the one applied for the following experiments.

C. Classifier Evaluation

In order to evaluate the performance of the trained classifier, it was applied to five different sequences from three different patients, who reacted with some kind of eye movement. These three patients were elderly, but they did not have any cognitive impairment, this is why their spontaneous reactions expressed like eye movements were few and far between. A total number of 1950 frames were analyzed in this experiment, within these 1950 frames, a total number of 545 were classified as significant movements, for the remaining it is considered that no significant movements occur. The results of this classification are detailed on Table VI. The columns of this table correspond to the number of frames evaluated, the number of frames where a significant movement was detected, the number of frames correctly classified, and finally, the accuracy of the classification in terms of percentage.

By the evaluation of the classification results, it was observed that a couple of optimizations could be applied. First, having into consideration the domain knowledge, it can be established that it must exist continuity along the movement., i.e., if a movement of EC is detected for three consecutive frames in one eye, and in the other eye, two frames are classified as EC, whereas an intermediate frame is classified as GM, it is very likely that a miss classification has occurred and that particular frame should be classified as EC too. By the application of a voting system,

TABLE V
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%

Last row shows the average for each classifier.

TABLE VI
INITIAL CLASSIFICATION RESULTS

	Number of frames	Significant frames	Correctly classified	Accuracy
Video 1, seq. 1	350	124	87	70.16129%
Video 1, seq. 2	400	134	80	59.70149%
Video 2, seq. 1	400	90	68	75.55556%
Video 2, seq. 2	400	122	80	65.57377%
Video 3, seq. 1	400	75	59	78.66667%
Global	1950	545	374	68.62385%

TABLE VII
CLASSIFICATION RESULTS AFTER THE OPTIMIZATIONS

	Number of frames	Significant frames	Correctly classified	Accuracy
Video 1, seq. 1	350	124	93	75.0%
Video 1, seq. 2	400	134	82	61.19403%
Video 2, seq. 1	400	90	73	81.11111%
Video 2, seq. 2	400	122	86	70.49180%
Video 3, seq. 1	400	75	69	92.0%
Global	1950	545	403	73.94495%

TABLE VIII
ACCURACY BEFORE AND AFTER THE OPTIMIZATION

	Accuracy before optimization	Accuracy after optimization
Video 1, seq. 1	70.16129%	75.0%
Video 1, seq. 2	59.70149%	61.19403%
Video 2, seq. 1	75.55556%	81.11111%
Video 2, seq. 2	65.57377%	70.49180%
Video 3, seq. 1	78.66667%	92.0%
Global	68.62385%	73.94495%

based on the requirement of this continuity, some miss classifications may be corrected and, thus, the accuracy may be improved. Furthermore, considering the domain and according to the experts opinion, isolated movements of only frame of length are discarded, because a movement without continuity does not represent a significant movement. Moreover, this system attempts to automate the expert behavior, and the expert does not consider movements of one frame of length since he or she is not able to detect them and, thus, they are irrelevant for the characterization of patients as they can induce error.

Taking into account these two considerations, the results were optimized and the new obtained results are detailed in Table VII. For an easier comparison, the accuracies before and after the optimizations are compared in Table VIII. This last table shows the improvement in the accuracy due to the optimizations applied. It can be noted that for Video 3, seq. 1, the accuracy suffers a greater increase. By the analysis of this sequence, it was observed that the lighting conditions were significantly better; thus, under optimal recording conditions, the influence of the proposed optimizations is even greater. Nevertheless, under normal recording conditions, the optimizations also provide an improvement.

D. Association of Movements and Stimuli Reactions

Finally, the most relevant results are related to the correct detection of eye gestural reactions to the stimuli. Previously, it has been seen that the movements are detected and they are correctly classified. Now it must be demonstrated that by the correlation of the detected movements and the auditory stimuli, the system is able to detect the reactions to the stimuli.

For this last analysis, the amount of data is not high, but even so, it is interesting to conduct a preliminary analysis to determine if the reactions can be correctly associated with the detected movements. With five different video sequences from three different patients, the aim is to corroborate if the detected reactions correspond to the ones labeled by the experts.

For this experiment, it is considered that a gestural reaction exists, when a significant movement occurs during two frames or more. For this particular group of patients, besides all the rest of eye movements, they expressed their unconscious reactions by gaze movements, so, in this case, only movements of classes GL and GR are interpreted as positive reactions to the stimuli. Classification results are processed and it is established that a eye gestural reaction exists when two or more consecutive frames are classified as GL or GR. According to this, the number of detected reactions for each video sequence is detailed in Table IX. As it can be observed, all the existing eye gestural reactions are correctly detected with this methodology, achieving a 100% of accuracy in the detection of these reactions, which is the main goal of this proposal.

Finally, associated with the auditory stimuli, we have shown here the reactions labeled by the experts against those obtained

TABLE IX
NUMBER OF EXISTING AND DETECTED EYE GESTURAL REACTIONS FOR EACH VIDEO SEQUENCE

	Classification accuracy	Number of reactions	Detected reactions
Video 1, seq. 1	75.0%	2	2
Video 1, seq. 2	61.19403%	1	1
Video 2, seq. 1	81.11111%	3	3
Video 2, seq. 2	70.49180%	1	1
Video 3, seq. 1	92.0%	1	1
Global	73.94495%	8	8

The correlation between natural reactions and our methodology is maximum.

by the system. Although the data are insufficient for a definitive validation, it has been shown that the methodology here explained offers useful characteristics for this domain.

V. CONCLUSION AND FUTURE RESEARCH

This methodology is capable of characterizing the eye movement of patients with conditions in the audiometric domain, something that had not been addressed so far. Besides, it is not only able to classify the movement with reasonable detection rates, but also these rates seem to indicate that the methodology is appropriate for the detection of eye gestural reactions to the stimuli. The methodology has shown encouraging and positive results, especially considering that it is the first fully automated approximation proposed for this domain.

For future works, there is a need of more video sequences of patients with this particular conditions so a more comprehensive analysis can be conducted. Also, deeper studies on applicability of the movement description to learn from the patient's responses will be done. The behaviour of this methodology suggests that it may be a useful tool in different domains, where eye gestural reactions could provide information, i.e., in [25], a survey about different physical and cognitive aspects that may influence cognitive decline is conducted. In a different survey [26], eye blink rate is used as a biological marker of cognitive decline. With these surveys in mind, it might be proposed the study and adaptation of our methodology to facilitate cognitive evaluation of elder people based on gestures and reactions.

ACKNOWLEDGMENT

The authors would like to thank the Fogar da Terceira Idade Porta do Camiño de Santiago de Compostela (Spain) and their residents for allowing them to evaluate and record them.

REFERENCES

[1] *Libro Blanco Del Envejecimiento Activo (in Spanish) IMSERSO*, Oct. 2010.
[2] A. Davis, "The prevalence of hearing impairment and reported hearing disability among adults in great britain," *Int. J. Epidemiol.*, vol. 18, pp. 911–17, Dec. 1989.
[3] N. I. of Deafness and O. C. Disorders. (2009, Mar.). Quick statistics. [Online]. Available: http://www.nided.uih.gov/health/statistics/hearing.asp
[4] A. H. A. Report. (2009). [Online]. Available: http://www.hearing.com.au/annual-reports
[5] C. Mulrow, C. Aguilar, J. Endicott, M. Tuley, R. Velez, W. Charlip, M. C. Rhodes, J. A. Hill, L. A. DeNino *et al.*, "Quality-of-life changes and hearing impairment. A randomized trial," *Ann. Internal Med.*, vol. 113, pp. 188–194, Aug. 1990.
[6] F. R. Lin, "Hearing loss and cognition among older adults in the united states," *J. Gerontol., Series A*, vol. 66, pp. 1131–1136, 2011.
[7] A. Fernandez, M. Ortega, B. Cancela, M. Penedo, C. Vazquez, and L. Gigirey. (2012). Automatic processing of audiometry sequences for objective screening of hearing loss. *Expert Syst. Appl.* [Online]. *39(16)*, pp. 12 683–12 696. Available: http://www.sciencedirect.com/science/article/pii/S0957417412007130
[8] P. Ekman, W. Friesen, and P. Ellsworth, *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*, (Pergamon General Psychology Series). Oxford, U.K.: Pergamon Press, 1972. [Online]. Available: http://books.google.es/books?id=uKe3MQEACAAJ
[9] I. Cohen, N. Sebe, L. Chen, A. Garg, and T. S. Huang, "Expression recognition from video sequences: Temporal and static modelling," *Comput. Vis. Image Understanding*, vol. 91, pp. 160–187, 2003.
[10] S. Bashyal and G. K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning vector quantization," *Eng. Appl. Artif. Intell.*, vol. 21, no. 7, pp. 1056–1064, 2008.
[11] K. Sandeep and A. N. Rajagopalan, "Human face detection in cluttered color images using skin color and edge information," in *Proc. Indian Conf. Comput. Vis., Graph. Image Process.,* 2002, pp. 230–235.
[12] H. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, and E. Petajan, "Multimodal system for locating heads and faces," in *Proc. 2nd Int. Conf. Autom. Face Gesture Recog.,* 1996, pp. 88–93.
[13] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
[14] "Locating and extracting the eye in human face images," *Pattern Recog.*, vol. 29, no. 5, pp. 771–779, 1996.
[15] S. Kawato and N. Tetsutani, "Detection and tracking of eyes for gaze-camera control," *Image Vis. Comput.*, vol. 22, no. 12, pp. 1031–1038, 2004.
[16] J. Huang and H. Wechsler, "Eye detection using optimal wavelet packets and radial basis functions (RBFs)," *Int. J. Pattern Recog. Artif. Intell.*, vol. 13, pp. 1009–1026, 1999.
[17] Z. Zhu and Q. Ji, "Robust real-time eye detection and tracking under variable lighting conditions and various face orientations," *Comput. Vis. Image Understanding*, vol. 98, pp. 124–154, 2005.
[18] A. Geetha, V. Ramalingam, S. Palanivel, and B. Palaniappan, "Facial expression recognition—A real time approach," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 303–308, 2009.
[19] S. Kumano, K. Otsuka, J. Yamato, E. Maeda, and Y. Sato, "Pose-Invariant facial expression recognition using variable-intensity templates," *Int. J. Comput. Vis.*, vol. 83, no. 2, pp. 178–194, Jun. 2009.
[20] H. C. Akakin and B. Sankur, " Robust classification of face and head gestures in video," *Image Vis. Comput.*, vol. 29, no. 7, pp. 470–483, Jun. 2011.
[21] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell.,* 1981, pp. 674–679.
[22] J.-Y. Bouguet, "Pyramidal implementation of the Lucas-Kanade feature tracker: Description of the algorithm," *Intel Corporation, Microprocessor Res. Labs*, Santa Clara, CA, USA, 2000.
[23] H. Dibeklioglu, M. Ortega, I. Kosunen, P. Zuzanek, A. Salah, and T. Gevers, "Design and implementation of an affect-responsive interactive photo frame," *J. Multimodal User Interfaces*, vol. 4, pp. 81–95, 2011.
[24] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.,* 1994, pp. 593–600.
[25] P. Bamidis, A. Vivas, C. Styliadis, C. Frantzidis, M. Klados, W. Schlee, A. Siountas, and S. Papageorgiou, "A review of physical and cognitive interventions in aging," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 206–220, 2014.
[26] A. Ladas, C. Frantzidis, P. Bamidis, and A. B. Vivas, "Eye blink rate as a biological marker of mild cognitive impairment," *Int. J. Psychophysiol.*, vol. 93, no. 1, pp. 12–16, 2014.

Authors' photographs and biographies not available at the time of publication.