

# Examining *in vivo* Tympanic Membrane Mobility Using Smart Phone Video-Otoscopy and Phase-based Eulerian Video Magnification

Mirek Janatka<sup>a</sup>, Krishan S. Ramdoo<sup>b</sup>, Taran Tatla<sup>b,c</sup>, Krittin Pachtrachai<sup>a</sup>, Daniel S. Elson<sup>c</sup>,  
Danail Stoyanov<sup>a</sup>

<sup>a</sup>Centre for Medical Image Computing (CMIC) and Dept. Computer Science,  
University College London, UK

<sup>b</sup>Dept. of Ear, Nose and Throat, Northwich Park Hospital, London, UK

<sup>c</sup>The Hamlyn Centre, Dept. Surgery and Cancer, Imperial College London, UK

## ABSTRACT

The tympanic membrane (TM) is the bridging element between the pressure waves of sound in air and the ossicular chain. It allows for sound to be conducted into the inner ear, achieving the human sense of hearing. Otitis media with effusion (OME, commonly referred to as ‘glue ear’) is a typical condition in infants that prevents the vibration of the TM and causes conductive hearing loss, this can lead to stunting early stage development if undiagnosed. Furthermore, OME is hard to identify in this age group; as they cannot respond to typical audiometry tests. Tympanometry allows for the mobility of the TM to be examined without patient response, but requires expensive apparatus and specialist training. By combining a smartphone equipped with a 240 frames per second video recording capability with an otoscopic clip-on accessory, this paper presents a novel application of Eulerian Video Magnification (EVM) to video-otology, that could provide assistance in diagnosing OME. We present preliminary results showing a spatio-temporal slice taken from an exaggerated video visualization of the TM being excited *in vivo* on a healthy ear. Our preliminary results demonstrate the potential for using such an approach for diagnosing OME under visual inspection as alternative to tympanometry, which could be used remotely and hence help diagnosis in a wider population pool.

**Keywords:** Otoscopy, Eulerian Video Magnification, Otitis media with effusion

## 1. INTRODUCTION

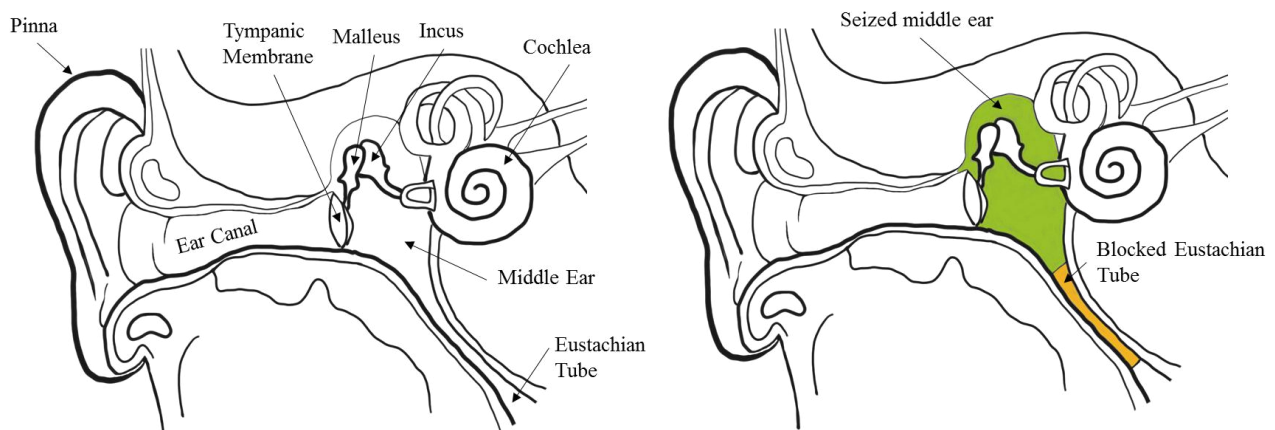
The tympanic membrane (TM) is the principal element for conducting sound waves through ossicular chain in the middle ear. Pressure waves that propagate through air and into the ear canal are received by the TM, which then transmits the sound via conduction across the middle ear to the nerve center located in the cochlea, enabling the human sense of hearing. Therefore, the ability for the TM to vibrate is paramount to human hearing. Physiologically, this is maintained by ensuring that the pressure differential between the outer ear and the middle ear is neutral. The Eustachian Tube (ET) connects the middle ear chamber to both the nasal passage and throat and opens during mastication and yawning, ensuring both sides of the TM are at atmospheric pressure. This anatomy can be seen in Figure 1 on the left.

However, if the ET ceases to function, a vacuum is created in the inner ear chamber, that draws surrounding fluid into the tympanic cavity, causing a condition called otitis media with effusion (OME), or more commonly known as ‘glue ear’. With no exhaust, this fluid can solidify within the chamber and dampen the mechanical transmission of sound from the TM across the ossicular chain, due to an immobilized TM, as demonstrated in Figure 1, right. OME is prevalent in newborns and young infants and can be undetectable due to lack of symptoms<sup>1</sup>, leading to lack of sensory interaction with the child; which can in turn cause great issues with their early development due to loss of hearing. Furthermore, the condition is difficult to diagnose due to the inability of the child to signal when they can hear pure-tones used in audiometry examinations.

In such cases, tympanometry is used to supplement audiometric diagnosis of OME and other causes of conductive hearing loss<sup>2</sup>. This technique involves generating a pure-tone directly inside a sealed ear canal, whilst a connected pneumatic pump attached to the probe changes the pressure within the canal through a range of negative, neutral and positive levels. A microphone based inside the probe monitors the change in returned sound energy at each pressure increment, which reflects the change in volume of the ear canal, indicating the mobility of the TM in the form of a tympanogram.

Tympanometry side steps the requirement of having a patient respond to an audio signal for a diagnosis, yet are costly (typically between US\$2000~3500<sup>2</sup>) and are therefore not readily available for home use, where concerned parents could conveniently inspect their child.

Over the last decade, smart phone cameras have greatly improved in terms of sensor and display resolution, as well as optical quality. The large-scale demand has also made such optically enhanced devices abundant and affordable on the consumer market. This eruption of readily available high quality digital camera systems has opened up a market for comparatively low cost auxiliary components and software that can repurpose the multi-functional smart phone into a medical diagnostic instrument<sup>3</sup>. The Cellscope Otoscope (Cellscope, Inc. San Francisco, California, USA) is an optical add-on to the iPhone (Apple, Inc. Cupertino, California, USA) that allows the device to be used as a digital video-otoscope. It consists of a positive lens for optical magnification of the TM and ear canal and a fibreoptic relay, which allows the LED light source on the back of the phone to be redirected to the distal end of the scope, providing illumination inside the ear. The Cellscope is compatible with Welch Allyn specular (Welch Allyn, Inc. Skaneateles Falls, New York, USA), which twist on as it would on a conventional otoscope; permitting for reusability and cleanliness of the product with minimum cost of consumables. The principle of the system is that a visualization of the TM and ear canal can be achieved of the child by the parent at home, who then can send the video of the ear to a specialist over the Internet, who can then advise on further treatment if needed. This allows for continuous monitoring of the child's ear by an expert, without having to attend a physical appointment.



**Figure 1:** Anatomy of the ear (left). Ear with Otitis Media with Effusion

This procedure is a proven pipeline of remote diagnostics for ear infections using video, however it is unable to verify if the TM has mobility in a similar vein to tympanometry and therefore is not ideal for diagnosing OME. This is due to the inability to distinguish visually if the TM can be excited and thus transmit sound to the inner ear. Motion magnification would enable such a video-otoscopic approach to identify such oscillations, yet is dependent on the high-speed video acquisition. Currently, the CellScope Otoscope acquires video at 30 fps at 1080p resolution. By utilizing the addition of high-speed video capture capabilities of the recent iPhone models and the recent advent of Eulerian Video Magnification (EVM), such a low-cost system could determine if the TM is functioning and mobile, as an alternative substitute to tympanometry and allowing for a home diagnosis of OME in infants, without needing pneumatics.

## 2. METHODS

### 2.1 Data Acquisition

A healthy volunteer was recruited with no known hearing complications and seated in a quiet room. A pure tone generator was connected to a loud speaker (SoundDock Series 2, Bose Corp, Massachusetts, USA) and a continuous 60 Hz sine wave was transmitted at a comfortably high volume. The smart phone (iPhone 6S, Apple, California, USA) was set to 240 fps slow motion video mode, which records at 720p resolution. The light source on the phone was turned on with the Cellscope Otoscopic device attached. With the specula placed into the volunteer's left ear canal, as demonstrated in Figure 2, and a clear visual of the TM established with the use of the camera's fixed auto-focus system, the diagnostic high-speed video was captured.



**Figure 2:** Left) The CellScope Otoscope attached to the iPhone, placed inside an ear phantom (Kyoto Kagaku Co. Ltd, Kyoto, Japan). Middle) Close-up of insertion. Right) Typical visualization of TM on the iPhone screen.

## 2.2 Motion Video Magnification

Spatial-temporal motion magnification was first introduced by Liu *et. al.*<sup>4</sup>, which tracked the motion of user selected regions from an initial frame through the video via segmentation based on similarities in colour, proximity and correlated movement. This would allow for the motion of the selected regions to be amplified. Borrowing terminology from fluid dynamics, Liu's approach uses an Lagrangian formulation, as individual particle trajectories (or in this case, the trajectories of points on an object within a video) are being traced through time.

Another perspective on motion magnification is an Eulerian perspective<sup>5</sup>, where in fluid dynamics a specific volume within a fluid would be observed for changes, such as pressure, velocity or temperature in its state over time. Pixels within a stationary video can be similarly observed in an Eulerian manner, such as examining variation of intensity<sup>6</sup> and phase from a 2D Fourier spectra of the frames<sup>7</sup>. These subtle changes can be amplified and allows the accentuation of minute changes that would not normally be seen by the naked eye, such as vibrations of cranes<sup>7</sup> and subtle changes in colour of people to detect blood pulsation rate<sup>6</sup>. Sound retrieval from delicate structures (plant leaves, crisp packets) have also been demonstrated, which show the versatility and potential of the technique<sup>8</sup>. Potential surgical applications of EVM include vessel identification in laparoscopic surgery and neurosurgery<sup>9</sup>.

However, EVM has two caveats to output a meaningful result. It requires both a high-speed camera and a relatively static scene. To decipher sound vibrations visually on the TM, the sample rate of the video has to be twice that of the frequency of the pure tone used to excite the membrane.

As an iPhone 6s is capable of recording at 240 frames per second, it was decided to use an excitation tone of 60Hz. Due to the spatial-temporal nature of EVM, large movements that are not of interest, such as camera movement, can cause artifacts on the output video, which would occlude the desired observation. This confined, relatively stationary nature of the ear canal makes the otoscopic application of EVM suitable.

For the purpose of this paper, phase-based motion magnification was used in the form of a MatLab toolkit provided by Wadhwa *et. al.*<sup>7</sup>, as it is capable of amplifying subtle motions in the presences of larger movements, compared to prior intensity-based developments. In essence, their method uses complex steerable pyramids<sup>10</sup> to decompose the video and then separating the amplitude of local wavelets from their phase. Temporal filtering is then performed at each location, orientation and scale. The temporally-bandpassed phases can then be amplified by a factor and the pyramid collapsed to create the motion-magnified video. In this study, the band filter was set to +/-1 the excitation frequency of the audible pure-tone (60Hz) and to several amplification factors, x5, x15 and x30. The EVM processing toolkit was ran on an Intel i7 laptop with 16 GB of RAM. The dataset video was trimmed to 6 seconds and the computational time required typically 365 seconds. Octave bandwidth pyramids were used for faster processing time.

## 2.3 Data Processing

A spatial-temporal video probe tool program was written to observe the magnified video through time across a user defined single spatial dimension from the scene, in what this paper refers to as a spatial-temporal slice (STS). This utilised the Bresenham's line algorithm to allow the user to select a sample graphically any line between two point on the video.

To extract a qualitative reading from the STSs, additional processing on these temporally composed images had to be performed. Assuming that a transverse drum vibration is being observed by the video otoscope, a STS would give a one dimensional sample of this per frame. Such vibration would cause an intensity difference between each step, therefore a vibration signal can be extrapolated by observing the change in aggregate intensity per time step:

$$f(t) = \sum_{I \in L_t} I. \quad (1)$$

Where  $f(t)$  is the signal to be extracted,  $I$  is the intensity value of a pixel and  $L_t$  is the sampled line on the image at time step  $t$ . With the assumption that an STS from an unamplified video should appear to be static, any signal extracted can be regarded as noise from artifacts or subtle motions from the surrounds (patient or inspector).

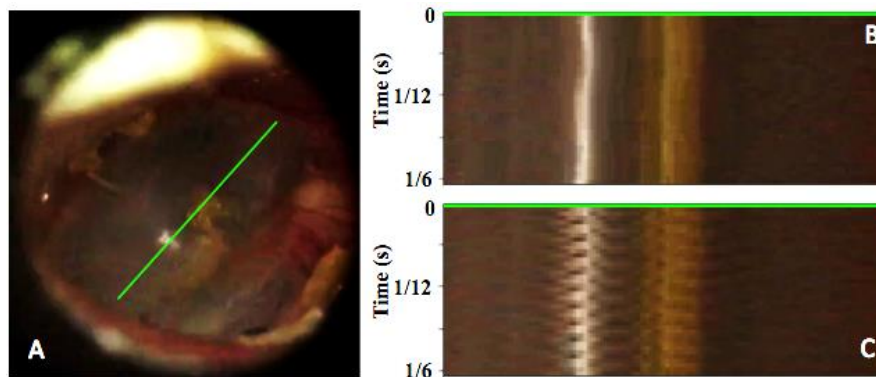
$$\sum_{I \in L_t} I = \sum_{I \in L_{t+1}} I. \quad (2)$$

$$f_{result}(t) = f_{amplified}(t) - f_{unamplified}(t). \quad (3)$$

This signal from the unamplified STS can be subtracted from the amplified STS to give a cleaner signal. This outlines how the results are extrapolated from the images in the next section.

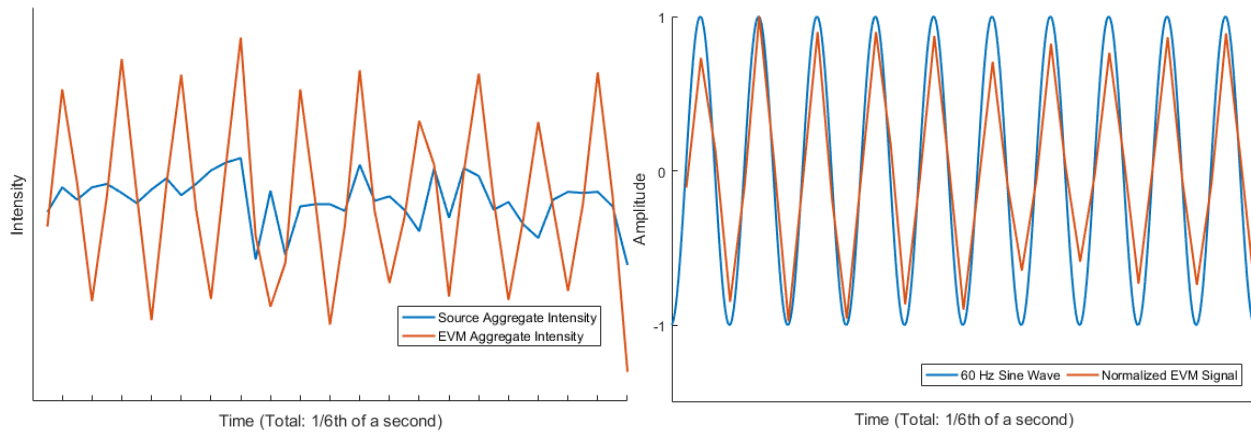
### 3. RESULTS

A clear illuminated visual of the volunteer's TM was achieved in the experiment, as shown in Figure 4a. The STS in Figure 4b shows that there was some slight motion present, despite best efforts to maintain a static scene, however such motions are to be expected *in vivo* due to the respiratory system, movement of the patient's head and stability of the inspector. Figure 4c shows the STS of the EVM processed video amplified by x30, where a rippling pattern is immediately identifiable in parts of the image. This visual confirms that the TM is oscillating at 60Hz, as 10 ripples can be counted within the sixth of a second sampled.

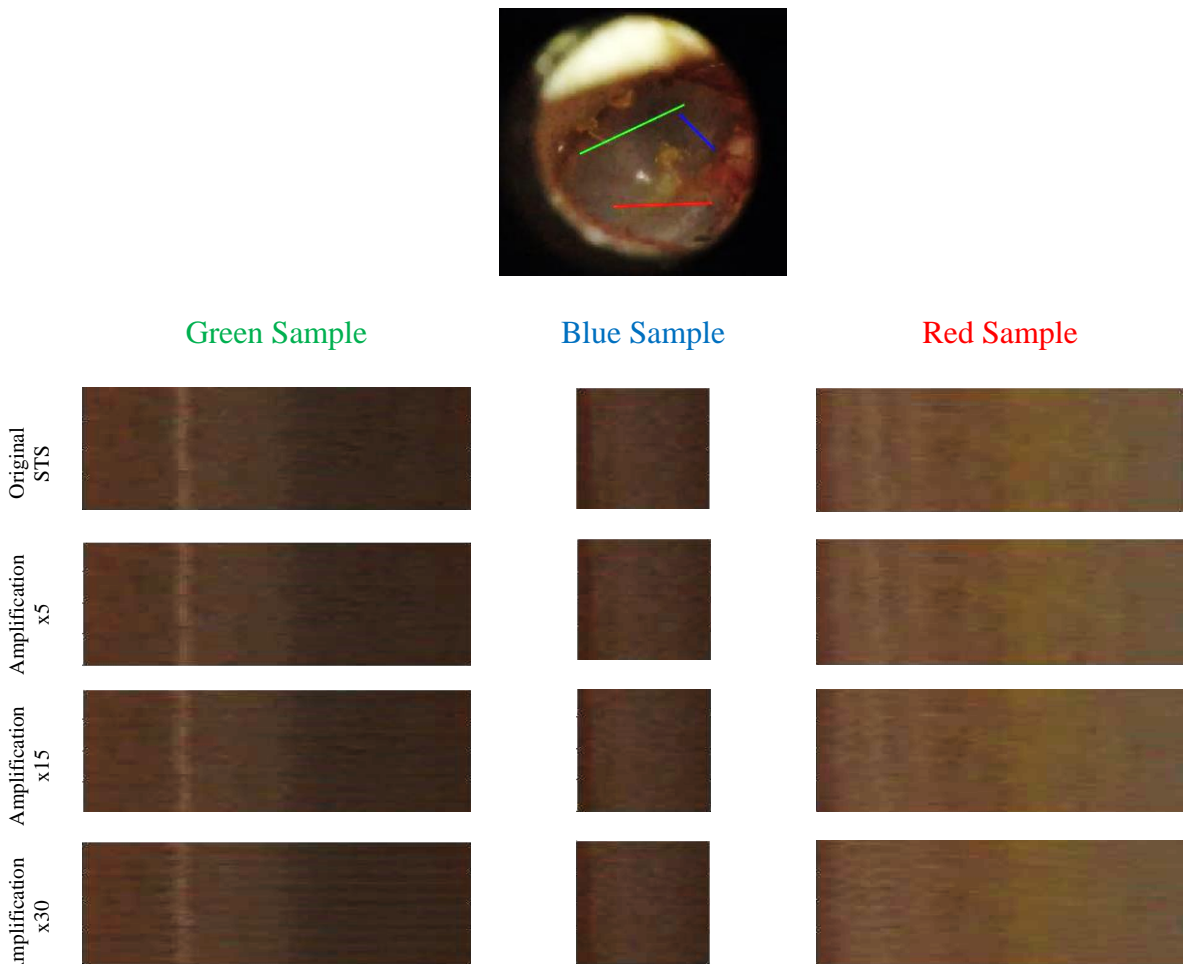


**Figure 4:** Video-otoscopic visualization of the TM. The green line indicates the spatial location in the frame of where the STS was sampled A) Visualization of the ear from the otoscope with placement of STS. B) STS of original unamplified video. C) STS of EVM processed video. The ten ripples observed are attributed to the 60Hz audible pure-tone excitation from the loud speaker within a sixth of a second.

Figure 5 (left) compares the extracted signal from both STS. This graph shows comparatively how EVM allows for the mobility of the TM to be identified compared to the insignificant raw source video. Figure 5 (right) shows the correspondence to the audible pure tone input signal has to the visually derived signal.



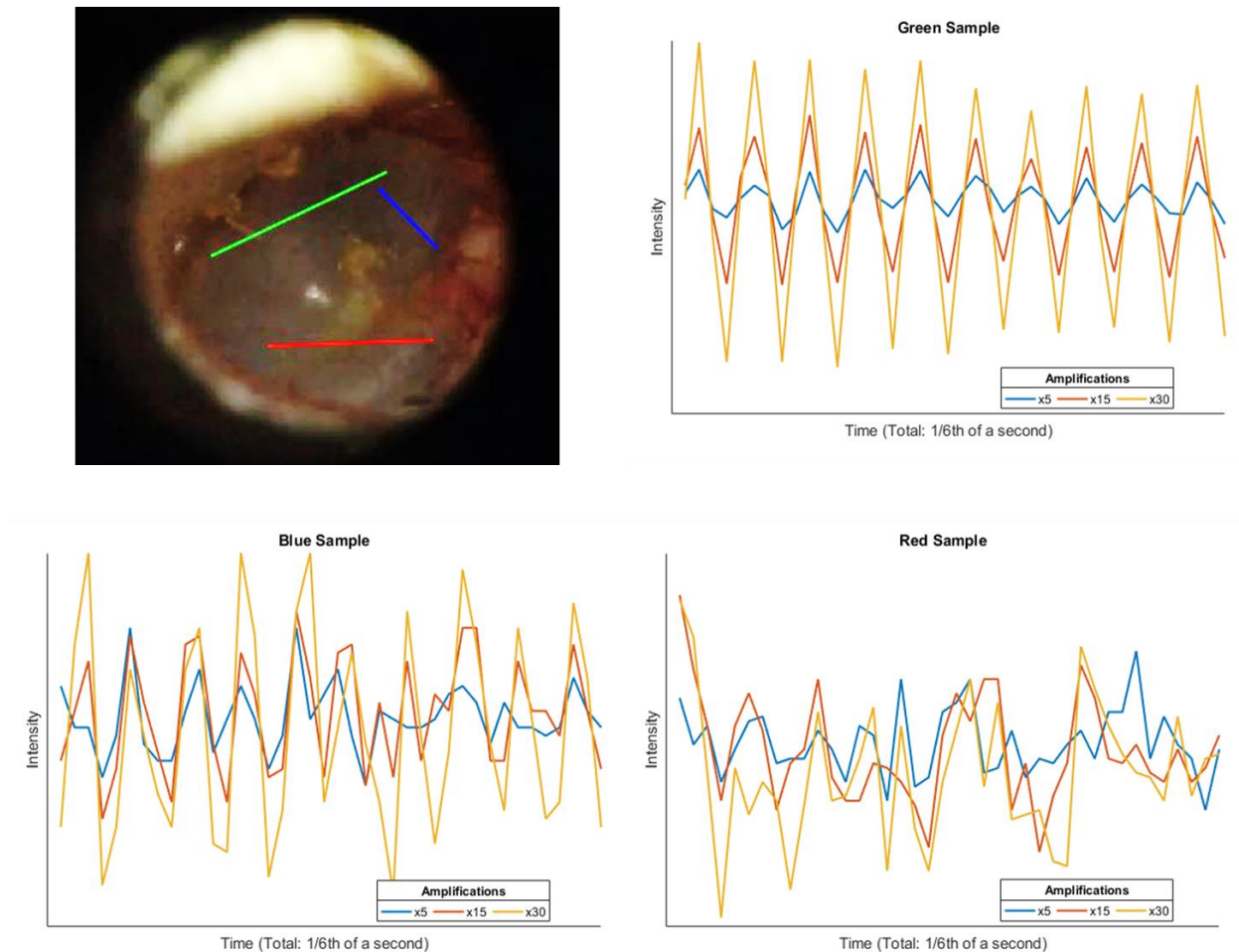
**Figure 5:** Left) Extrapolated data from both source (blue,  $f_{unamplified}(t)$ ) and processed (orange,  $f_{amplified}(t)$ ) STS. Summation of intensities from both STS are shown for each time sample, the 10 peaks in blue represents the visualization of the sound vibrating the TM. Bottom) Signal retrieval and comparison. The signal was retrieved by subtracting the source STS from the EVM processed STS ( $f_{result}(t)$ ), which partially removes the background noise. The final output is normalized and shown alongside the phase aligned input audible pure tone that was used to excite the TM.



**Figure 6:** Comparison of different sample sites at different amplifications



A variety of sample locations on the TM and several amplification factors were trialed to confirm consistent results. Figure 6 shows a visual comparison how vibrations from the sound wave can be found in several locations on the TM. The signals extrapolated ( $f_{result}(t)$ ) from these samples are compared in Figure 7.



**Figure 7:** Comparative extrapolated signals from different sample sites and amplifications

Figure 4 shows the capability to detect vibration of the TM within the ear canal, however a salient features present in the video were selected and can easily be seen in the STS. This is shown in Figure 5, which shows how the signal can be proceed via subtraction of the unprocessed footage and how similar the retrieved signal is to that of the excitation sine sound wave. To prove that the method would also work on other sample sites, Figure 6 shows a comparison of various sample sites that are lacking salient features, as sampled in Figure 4. Visual inspection of the resultant STSs does not show the vibrations as apparently as Figure 4, however it is still clear to extract a clear frequency response from the green and blue samples taken in Figure 7, however, this is not possible to see in the red sample. Potentially, this could be due to either the orientation of the sample on the membrane or that the sample taken in an area that behaves different to the center of the TM due to attachments to other parts of the ear. It can be seen from all samples in Figure 7 that the greater the amplification the clearer the frequency response can be deciphered.

## 4. CONCLUSION

This study demonstrates the novel use of existing low cost consumer products combined with algorithms for examining the mobility of the TM. Our preliminary results show that the approach could be an effective alternative to expensive pneumatic driven tympanoscopes, when the combination of EVM and high-speed video capture are used. Such methods could allow for an increase in the diagnosis of OME for both concerned new parents and possibly in developing countries. Because the approach is low cost and can be applied remotely there could be impactful applications in healthcare systems where the medical resources and staff concentrated in urban centers and are scarce amongst the rural population. Ultimately such capabilities could lead to improvements in pediatric care and assist global child development.

The preliminary results that we show demonstrate a successful retrieval of an audio pure-tone from a healthy volunteer's TM proving its mobility using visual video-otoscopy. While the result is encouraging additional experiments are necessary to clearly demonstrate the approach within a clinical context with more data samples and a combination of healthy and impeded TMs. In addition to the technique being performed on patients with OME, comparisons to tympanometry for validation and a combined tone generation and signal acquisition are needed. For a practical system, the EVM software needs to be deployed for on board processing on a smart phone and complemented by recording software and algorithms for removing residual motion which is not clinically relevant but present in the signal due to hand shake or movement of the patient.

## 5. ACKNOWLEDGEMENTS

The authors would like to thank Neal Wadhwa and Micheal Rubenstein at MIT CSAIL for sharing their MatLab toolkit used in the data processing<sup>7</sup>. The work was funded by the EPSRC (EP/N013220/1, EP/N022750/1, EP/N027078/1, NS/A000027/1).

## REFERENCES

- [1] American Academy of Family Physicians and others, "Otitis media with effusion.", *Pediatrics* 113(5), 1412 (2004).
- [2] Onusko, E. "Tympanometry," *Am. Fam. Physician*, 70(9), 1713–1720 (2004).
- [3] Rappaport, K. M. McCracken, C. C. Beniflah, J. Little W. K., Fletcher, D. A. Lam, W. A. and Shane, A. L. "Assessment of a Smartphone Oscope Device for the Diagnosis and Management of Otitis Media", *Clinical pediatrics.*, 1, 1–11 (2015).
- [5] Rubinstein, M. Wadhwa, N. Durand F. and Freeman, W.T. "Revealing Invisible Changes in the World," *Science*, 339(6119), 519 (2013).
- [6] Wu, H. Y. Rubinstein, M. Shih, E. Guttag, J. Durand, F. and Freeman, W. T. "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, 31(4), 1–8 (2012).
- [7] Wadhwa, N. Rubinstein, M. Durand, F. and Freeman, W. T. "Phase-based video motion processing," *ACM Trans. Graph.* 32(4), 1 (2013).
- [8] Davis, A. Rubinstein, M. Wadhwa, N. Mysore, G. J. Durand, F. and Freeman, W. T. "The Visual Microphone : Passive Recovery of Sound from Video," *SIGGRAPH 2014*, 1–10 (2014).
- [9] McLeod, J. Baxter, J. S. H. de Ribaupierre, S. Peters, T. M. "Motion magnification for endoscopic surgery," *SPIE Med. Img.*, 9036(1), 1–8 (2014).
- [10] Wadhwa, N. Rubinstein, M. Durand, F. and Freeman, W. T. "Riesz pyramids for fast phase-based video magnification," *ICCP*, (2014).