



FIV2018-218

ZONE BASED ACTIVE NOISE CONTROL VIA HEAD TRACKING FOR AN AIRCRAFT SEAT

Wintta Ghebreiyesus
Ryerson University
Toronto, Ontario, Canada

Fengfeng (Jeff) Xi
Ryerson University
Toronto, Ontario, Canada

ABSTRACT

Aircraft passengers are in constant need for sufficient cabin noise reduction for travel comfort. In the past, conventional active noise control (ANC) systems have been used inside aircraft cabins to provide noise reduction. In recent years, the concept of placing compact ANC systems on either side of the headrest of a passenger seat is being applied to produce a zone of quiet (ZoQ) around the ears. This paper aims at improving the zone based active noise control for an aircraft passenger seat using virtual sensing and head tracking techniques.

A virtual sensing method will be introduced in order to represent the noise that a passenger would expect to hear at the ear. Virtual sensing allows for this representation even though the sensing microphones are installed in the seat's headrest. The locations of the virtual sensors will be determined using a Matlab-based simulation with the goal of creating an ideal zone of quiet (ZoQ) with respect to the physical sensors. The second part of the paper will outline the heading tracking process of localising the ZoQ to around a moving mannequin head. The head tracking method will be developed to track the coordinates of the mannequin's ears for the virtual sensing model when a passenger's head is moving while seated. The software program Python is used with MS Kinect to use HSV parameters and face cascading for head tracking of the mannequin head.

The methods presented in this paper can be developed to accurately predict zones of quiet for various multiple input multiple output (MIMO) local ANC configurations. This work will provide a basis for the shape and size of the ZoQ with respect to the sound sources and desired location of cancellation. The virtual sensing and head tracking results will be presented in the final paper.

INTRODUCTION

There is a need for business jet cabin noise reduction research. The current methods for cabin noise reduction are not sufficient as passengers continue to be dissatisfied with loud flying environments. Generally, aircraft noise control can be classified as being environmental (outside of the aircraft) or cabin-related (inside of the aircraft). Cabin-related aircraft noise control can further be classified as being overall or zone-based. Most methods of business jet cabin noise cancellation occur outside the cabin and within the cabin trim or inner lining. Unfortunately, these methods do not target vibration-borne noise from the cabin itself or noise that propagates from inside of the cabin. In addition to this, the dominant sources of noise in an aircraft, which are the engine harmonic tones, still remain undamped by these noise reduction techniques. Noise reduction inside the cabin of a business jet is necessary for providing passengers with optimal sound comfort.

According to Guigou et al., turbofans, propellers and turbulent boundary layers are some causes of noise fields which propagate sound waves through the fuselage and create extremely high sound levels inside the cabin [1]. Major noise sources originating from the airframe of an aircraft include horizontal stabilizers, spoiler, flaps, leading and trailing edge devices, the wings, landing gears, nacelles, and the fuselage. Moreover, the concept of head-tracking for localized active noise control creates a quiet zone in close vicinity to a single passenger. The head-tracking system has a fast face detector. The applications of a fast face detector are diverse and include user interfaces, image databases, and teleconferencing [2].

Review of zone-based ANC

A Zone of Quiet (ZoQ) is created when a user is isolated from the outside acoustical environment. One of the major goals of the

ZoQ is to focus the noise control strictly within the local zone of quiet. From a practical standpoint, the ZoQ is limited by the spatial distribution of sound sources and the geometry of the desired location of cancellation. A common local ANC application of a ZoQ is a noise reducing headrest in a passenger seat, attenuating noise around the passenger’s ears. The physical microphone (located near the aircraft headrest) used to generate the error signal in the ANC scheme cannot be located inside a human ear. Active noise control permits the generation of a zone-of-quiet based on destructive interference. The size and shape of this zone-of-quiet depend on the type of the sound sources and the frequency components of the signal to be canceled [3].

Figure 1 provides an illustration of a person enclosed in a zone-of-quiet around his head and headrest area. Generally, a mean squared error (MSE) criterion is applied to one point of interest away from the physical microphone. Then a ZoQ-based approach can be used to control the noise at that point by deriving a filter that can minimize this MSE. A zone-of-quiet is defined by a sinc function, with the primary sound pressure level reduced by 10 dB over a sphere of diameter one tenth of the excitation wavelength, $\lambda / 10$. The main benefit of the zone-of-quiet ANC technique versus conventional ANC methods for controlling the sound field in rooms is that the far field sound pressure level remains mostly unaffected, a typical increase of less than one dB [4].



Fig. 1 Concept of Zone of Quiet.

METHODOLOGY FOR NOISE REDUCTION

Active noise control has been used for several decades to provide attenuation for various applications. ANC systems provide noise reduction by introducing a controllable secondary noise source at the originating source of noise. In the case of an aircraft passenger seat, the reduction is near the seat which is a new definition of near-field cancellation – close to the desired location of cancellation. For local active noise control systems, virtual microphones are meant to be used in diffuse sound fields. This will explained in the proceeding sections.

A diffuse sound field is one where sound pressure is uniform throughout. A perfect diffuse field is almost unachievable,

however, it is the model commonly used for reverberant sound field analysis. In these cases, the sound field is assumed to be sufficiently diffuse. This means that the field is diffuse above the Schroeder frequency. This is the frequency above which there exists at least three room acoustic modes within the 3 dB bandwidth of any one mode. For most enclosures, the Schroeder frequency is between 100 and 200 Hz.

Localisation of noise reduction

There are a lot of factors which affect the spatial extent of a zone of quiet. Theoretically-speaking, in order to determine the size of a zone of quiet, there needs to exist a perfectly diffuse sound field. The equations presented in this section have been modified from previous work by Rafaely [4]. The pressure in a perfect diffuse field can be expressed as a function of space and time in spherical coordinates, $\mathbf{x} = (x, \theta, \phi)$ as

$$e(\mathbf{x}, t) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N e_n(\mathbf{x}, t) \quad (1)$$

where $e(\mathbf{x}, t)$ is the total pressure at position \mathbf{x} and time t , N is the number of plane waves which approaches infinity, and e_n is the pressure associated with the n th plane wave. The autocorrelation in the pure tone diffuse field can be expressed using a sinc function as

$$\rho(\Delta\mathbf{x}, t) = \frac{\text{sinc}(k \Delta\mathbf{x})}{k \Delta\mathbf{x}} \quad (2)$$

where k is the wave number and ρ is the correlation coefficient. The acoustic wave number, k , can be determined by dividing $2\pi f$ by c , where f is the natural frequency (in Hz) and c is the speed of sound (in m/s). The correlation coefficient, ρ , can be defined in space and time (cross-correlation) assuming the sound field is stationary over both as

$$\rho(\Delta\mathbf{x}, \Delta t) = \frac{E[e(\mathbf{x}_1, t_1)e(\mathbf{x}_0, t_0)]}{E[e^2]} \quad (3)$$

where $\Delta\mathbf{x}$ represents the distance between the two points, Δt represents the time delay, $E[\cdot]$ is the expectation operation which is determined by averaging several samples of diffuse sound fields, and $E[e^2]$ is the variance of the pressure which is independent of \mathbf{x} and t because of the assumption that the sound field is stationary with respect to space and time. The introduction of a time delay makes Equation 3 applicable to broadband diffuse sound fields. Now that the diffuse sound field has been explained, the following section will show how to determine the autocorrelation for zones of quiet in broadband diffuse sound fields.

Local active noise control in a diffuse sound field can be achieved by using a secondary source, which is oftentimes a loudspeaker, and canceling the total pressure in the near field of the source of noise. A theoretical model which addresses this approach models a monopole secondary source in a primary

diffuse sound field. A derivation of the spatial extension of the zone of quiet in a broadband diffuse sound field for local active noise control will be presented. The primary diffuse sound field is denoted by $d(\mathbf{x}, t)$. When a secondary monopole source is placed at the origin of the same spherical coordinate system $\mathbf{x} = (x, \theta, \phi)$, the resulting pressure is expressed as $y(\mathbf{x}, t)$. The total pressure is found by superimposing the primary and secondary pressures as follows

$$e(\mathbf{x}, t) = d(\mathbf{x}, t) + y(\mathbf{x}, t) \quad (4)$$

The pressure cancellation is assumed to be at a point $\mathbf{x}_0 = (x_0, \theta_0, \phi_0)$ such that

$$d(\mathbf{x}_0, t) + y(\mathbf{x}_0, t) = 0 \quad (5)$$

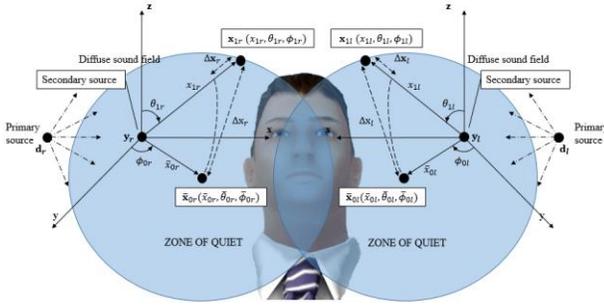


Fig. 1 Diagram of spatially fixed cancellation points $\bar{\mathbf{x}}_{0r}$, $\bar{\mathbf{x}}_{0l}$ and positions \mathbf{x}_{1r} , \mathbf{x}_{1l} near them, relative to the secondary sources for the right and left ears respectively.

This shows that the primary and secondary fields are equal with opposite phase at the cancellation point \mathbf{x}_0 . Figure 3 below is a visual representation of this concept. The secondary field is what allows for noise cancellation and is often referred to as *anti-noise*.

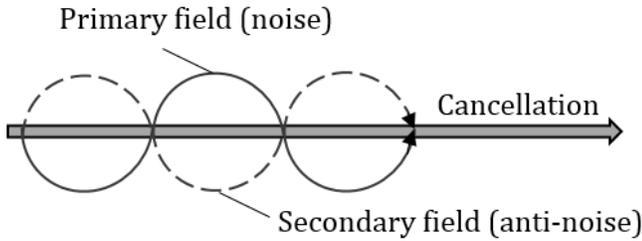


Fig. 3 Visual representation of primary and secondary fields at the cancellation point \mathbf{x}_0 .

The position \mathbf{x}_0 can now be assumed to be in the near field of the secondary source, such that the indirect secondary sound field caused by reflections is negligible. The reverberation distance is defined as the distance from the source at which the direct field dominates. The reverberation distance depends on the room

volume and reverberation time (length of time required for sound to decay 60 dB from its original level). The size of the zone of quiet is directly related to the effectiveness of the primary pressure attenuation around the cancellation point. The averaged squared total pressure at position $\mathbf{x}_1 = (x_1, \theta_1, \phi_1)$ near the cancellation point using the expectation operation $E[\cdot]$ over several samples of diffuse sound fields. The variance of the total pressure at point \mathbf{x}_1 can now be determined using

$$E[e^2(\mathbf{x}_1, t)] = E[d^2(\mathbf{x}_1, t)^2] + E[y^2(\mathbf{x}_1, t)] + 2E[d(\mathbf{x}_1, t)y(\mathbf{x}_1, t)] \quad (6)$$

The variance of the total pressure at \mathbf{x}_1 depends on the variance of the primary and secondary fields at the same point as well as on the correlation between the primary and secondary fields at \mathbf{x}_1 . If the diffuse primary field is assumed to be stationary such that the variance of the pressure is the same for all \mathbf{x} and t , the following equality holds

$$E[d^2(\mathbf{x}_1, t)] = E[d^2(\mathbf{x}, t)] = E[d^2] \quad (7)$$

The secondary source has been modelled as a monopole point source. Under some assumptions, the pressure produced by a monopole is similar to that produced by a piston in a baffle. Under one assumption, the source radius is significantly smaller than a wavelength λ and the source can be considered omni-directional. Under another assumption, only pressure further away than one source radius is considered for cancellation. These assumptions are sufficient for practical local active noise control systems such as active headrests, so the monopole model can be used successfully to obtain the behaviour of such systems. The secondary sound field produced by a monopole point source in the near field is assumed to generate spherical waves which propagate away from the source. The spherical waves also decay in amplitude. The secondary sound field can be expressed as

$$y(\mathbf{x}, t) = \frac{\rho_0}{4\pi r} \dot{q} \left(t - \frac{x}{c} \right) \quad (8)$$

which is now dependent only on the distance from the source \mathbf{x} , with q representing the source strength (volume velocity per unit volume) and \dot{q} its derivative with respect to time. The source strength q can be determined from the surface area and the amplitude of vibration of the spherical monopole point source. The secondary pressure at \mathbf{x}_1 can now be expressed in terms of the secondary pressure at \mathbf{x}_0 using Equation 8 as

$$y(\mathbf{x}_1, t) = \frac{x_0}{x_1} y \left(\mathbf{x}_0, t - \frac{\Delta x}{c} \right) \quad (9)$$

where Δx is the difference in the distances of the two points \mathbf{x}_1 and \mathbf{x}_0 to the source. The averaged squared secondary pressure at \mathbf{x}_1 can now be written using Equations 9, 7 and 5 as

$$E[e^2(\mathbf{x}_1, t)] = \left(\frac{x_0}{x_1} \right)^2 E[d^2] \quad (10)$$

The last term in Equation 6 can be expressed using Equations 9, 5, and 3 as

$$E[d(\mathbf{x}, t)y(\mathbf{x}_1, t)] = -\frac{x_0}{x_1} \rho \left(\Delta \mathbf{x}, \frac{\Delta x}{c} \right) E[d^2] \quad (11)$$

The variance of the total pressure at position \mathbf{x}_1 in Equation 6 can now be expressed in terms of the variance of the primary pressure by substituting Equations 11, 10 and 7 into Equation 6 yielding

$$E[e^2(\mathbf{x}_1, t)] = E[d^2] + \left(\frac{x_0}{x_1}\right)^2 E[d^2] - 2\frac{x_0}{x_1}\rho\left(\Delta\mathbf{x}, \frac{\Delta x}{c}\right)E[d^2] \quad (12)$$

The final expression for the sound attenuation ϵ at \mathbf{x}_1 (measurement point) can now be obtained by dividing Equation 12 by the variance of the primary pressure. Assuming cancellation at \mathbf{x}_0 , this will yield

$$\epsilon(\mathbf{x}_1, \mathbf{x}_0) = \frac{E[e^2(\mathbf{x}_1, t)]}{E[d^2]} = 1 + \left(\frac{x_0}{x_1}\right)^2 - 2\frac{x_0}{x_1}\rho\left(\Delta\mathbf{x}, \frac{\Delta x}{c}\right) \quad (13)$$

where the sound attenuation in dB can be found by $10 \log_{10} \epsilon$. It should be noted that in Equation 13, $\Delta\mathbf{x}$ is the distance from \mathbf{x}_1 to the cancellation point \mathbf{x}_0 , whereas Δx is the difference between the distances of the two points \mathbf{x}_1 and \mathbf{x}_0 to the secondary source (refer to Figure 2). The cross-correlation function in a diffuse field derived in the previous section, along with Equation 13 can be used to determine the zone of quiet with respect to the speaker or secondary source location. Generally, the size of the zone of quiet decreases at higher frequencies and at closer cancellation points, as can be seen in Figure 2. As the distance from the speaker to the cancellation point, \mathbf{x}_0 is increased, Δx , the difference between the distances of the two points \mathbf{x}_1 and \mathbf{x}_0 to the secondary source is also increased. This is a performance metric that can be used to determine the optimal zone of quiet.

Adaptive LMS moving virtual microphone technique

The virtual microphone technique was the first ANC virtual sensing method. This technique used the assumption of equal primary sound pressure at the physical and virtual microphone locations. This method can be carried out using equal numbers of physical and virtual microphones ($M_0 = M$). The microphones are located in M_0 pairs, with one physical microphone and one virtual microphone in each pair. The primary sound pressure is assumed to be equal at the physical and virtual microphones in each pair such that $\mathbf{d}_0(n) = \mathbf{d}(n)$.

The adaptive LMS moving virtual microphone technique developed by Moreau et al. [5] employs the adaptive LMS virtual microphone technique to estimate the virtual error signals at the moving virtual locations. The adaptive LMS moving virtual microphone technique determines estimates of the virtual error signals, $\hat{\mathbf{e}}_0(n)$, at the moving virtual locations, $\mathbf{x}_0(n)$. In this moving virtual sensing algorithm, the adaptive LMS virtual microphone technique is first used to obtain estimates of the virtual error signals, $\hat{\mathbf{e}}_0(n)$ at the spatially fixed virtual locations $\bar{\mathbf{x}}_0$. The primary component of the physical error signals is first calculated using the matrix of physical secondary transfer functions, $\hat{\mathbf{T}}_{ce}$, and is given as

$$\hat{\mathbf{d}}(n) = \mathbf{e}(n) - \hat{\mathbf{y}}(n) = \mathbf{e}(n) - \hat{\mathbf{T}}_{ce}\mathbf{u}(n) \quad (14)$$

Matrices of the primary and secondary weights, $\bar{\mathbf{w}}$ and $\bar{\mathbf{w}}_e$, of size $M \times \bar{M}_0$, at the \bar{M}_0 spatially fixed virtual locations, $\bar{\mathbf{x}}_0$, are then estimated separately. Estimates, $\hat{\mathbf{e}}_0(n)$, of the total virtual error signals at the spatially fixed virtual locations, $\bar{\mathbf{x}}_0$, can then be calculated using the following equation

$$\hat{\mathbf{e}}_0(n) = \hat{\mathbf{d}}_0(n) + \hat{\mathbf{y}}_0(n) = \bar{\mathbf{w}}^T \hat{\mathbf{d}}(n) + \bar{\mathbf{w}}_e^T \hat{\mathbf{y}}(n) \quad (15)$$

Estimates, $\hat{\mathbf{e}}_0(n)$, of the virtual error signals at the moving virtual locations, $\mathbf{x}_0(n)$, are now obtained by spatially interpolating the virtual error signals, $\hat{\mathbf{e}}_0(n)$, at the spatially fixed virtual locations, $\bar{\mathbf{x}}_0$.

Detection of cancellation point, \mathbf{x}_0 , via head tracking model

The global coordinate system in Figure 2 can be used to obtain the locations of the moving virtual microphones \mathbf{x}_{0r} and \mathbf{x}_{0l} , for the right and left ears respectively. As the head is rotated or translated, the distance from the speaker to the cancellation point, \mathbf{x}_0 will change. The difference between the distances of the two points \mathbf{x}_1 and \mathbf{x}_0 to the secondary source, Δx can be maintained in order to sustain the same size of zone of quiet. Estimates, $\hat{\mathbf{e}}_0(n)$, of the virtual error signals at the moving virtual locations, $\mathbf{x}_0(n)$, are now obtained by spatially interpolating the virtual error signals, $\hat{\mathbf{e}}_0(n)$, at the spatially fixed virtual locations, $\bar{\mathbf{x}}_0$.

The head tracking model will produce a better estimate of the matrix of physical secondary transfer functions, $\hat{\mathbf{T}}_{ce}$ (which represents the physical plant response between speaker and microphone) because the physical microphones will be close to the tracked head versus the instability which would be present in an untracked head system. This in turn will produce more accurate estimates, $\hat{\mathbf{e}}_0(n)$, of the total virtual error signals. Figure 4 demonstrates the use of a head tracker to find the head position for a local ANC system. The updated head positions can then be fed to the virtual plant to determine the moving virtual error signals at the listener's ears via the aforementioned moving virtual microphone method.

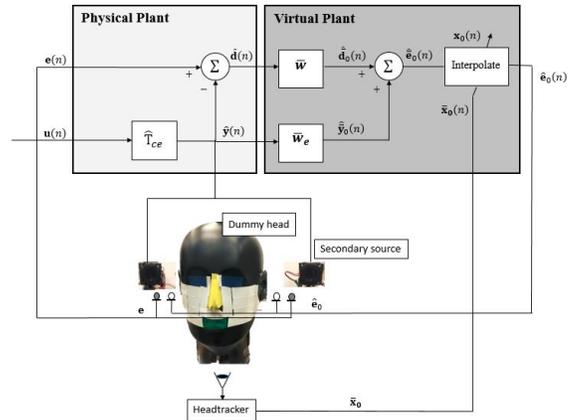


Fig. 4 Head tracking ANC setup and block diagram.

METHODOLOGY FOR HEAD-TRACKING

HSV scale and masking process

The hue, saturation and value (HSV) scale is representative of the purpose of RGB, however it organizes the geometry in such a way so that it becomes more meaningful in comparison to its Cartesian or cube counterpart. The HSV scale is a cylindrical system where the hue is represented by the angle around the central vertical axis, which is the azimuth or theta coordinate. The saturation is represented by the distance from this vertical axis which is the radial coordinate, and the value is represented by the distance along the axis which is the height coordinate. The value parameter is sometimes referred to as lightness or brightness. Table 1 below demonstrates the use of the HSV parameters and their corresponding lower and upper bound pixel values for each facial feature.

Table 1 Pixel Ranges for HSV Parameters.

HSV Bounds	HSV Parameters ([Hue, Saturation, Value])			
	Eyes	Nose	Mouth	Ears
Lower	[103,	[33, 53,	[78, 152,	[82, 23,
Limit	168, 54]	131]	57]	200]
Upper	[114,	[46, 253,	[96, 255,	[112, 69,
Limit	255, 153]	235]	147]	240]

The first step of the masking process is to represent the selected feature as a binary one which shows up on the display as white. On the other hand, all of the negatives or non-selected regions that do not fall into the HSV classification or range are binary zeroes and show up on the display as black. The secondary step in the masking process is to combine the ones and zeroes from each mask (one mask per facial feature) into a matrix format. This information is then displayed on a computer monitor or screen in the form of a white and black image. The final step is to combine the original red, green, and blue (RGB) image with the masked image to create a display of the detected and coloured facial features. A sample of this combined RGB and masked image can be seen in Figure 5 below.

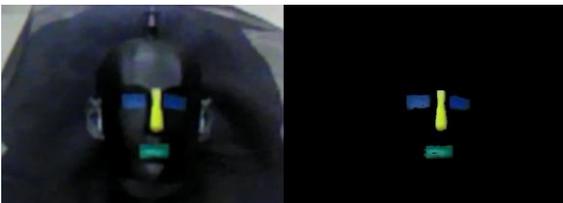


Fig. 5 Separated colour facial feature detection.

Integrated method for head tracking via MS Kinect

The first step in creating a rotation and translation system for head tracking via MS Kinect, is to capture the frame and then to obtain a window near the headrest. The size of the window for positive images (images of the face) is 25 x 35 pixels. Then the window is converted to a grayscale image. Then the Python program makes a decision to obtain the face from the window if it is detected, or to select another cascade to search for the face until it is found. A cascade is a special classifier that is trained using sets of positive and negative images.

After the training process, the cascade becomes an object detector and in this case a face detector. Once the face has been selected, the translation and rotation schemes can be applied. The most computationally-challenging part is detecting the face. In order to determine the translated distance, the position of the face's centroid is calculated and then fed into a predetermined translation equation to finally output the translated distance in centimetres. The head rotation angle determination process begins with storing the selected face in HSV format, which uses cylindrical coordinates to express colour in terms of hue, saturation, and value. Then the centroids of the nose and ears are stored separately. Similar to the translation scheme, the various centroid values are fed into predetermined rotation equations, and averaged. This whole process is then repeated for the next frame of the live video feed.

EXPERIMENTAL RESULTS AND DISCUSSION

A head tracking scheme, which utilizes HSV parameters and face cascading methods for an acoustic dummy head, was developed in order to improve the near-field zone of quiet for broadband noise near an aircraft passenger seat. The head tracking measurements proved that the translation of the head could be tracked within +/- 0.2 cm, whereas the head rotation angle could be matched to the closest degree. Sufficient prediction intervals of 0.5 cm for translation and 5 degrees for rotation were selected. Figure 6 demonstrates the relationships between translation and pixel position as well as rotation and pixel position for the dummy head experiment. At 17 degrees, there is a pixel range gap about 10 pixels wide, as can be seen by the red circled region on the right graph in Figure 6. The two cascades used for determining the yaw angle converge to the 17 degree output mark, so if the pixel value falls into this region on the graph, the angle is assumed to be 17. These results are sufficient, however, when 6 degree-of-motion tracking is implemented, it will be much more difficult to reproduce similar outcomes along each axis of rotation and translation.

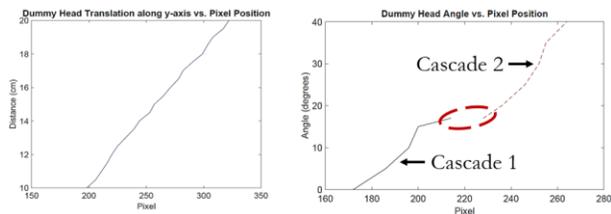


Fig. 6 Left: Dummy head translation vs. pixel position; Right: Dummy head horizontal angle vs. pixel position

Enhancement in Hearing Aids. *IEEE Transactions on Audio, Speech, and Language Processing*, pp. 1685 – 1697.

- [4] Rafaely, B. (2001). Zone of quiet in a broadband diffuse sound field. *Journal of the Acoustical Society of America* **110** (1), pp. 296-302.
- [5] Moreau, D., Cazzolato, B., Zander, A., & Petersen, C. (2008). A review of virtual sensing algorithms for active noise control. *Algorithms* **1**, pp. 69 – 99.

CONCLUSION

This research has aimed to improve zone-based local active noise control for an aircraft passenger seat primarily using two methods: (1) adaptive LMS moving virtual microphone technique, and (2) head tracking via MS Kinect. The main contributions of this work include the development of a zone-based integrated ANC system for an aircraft cabin seat, the creation of and the development of a head tracking tool which interfaces with Python programming language and Microsoft’s Kinect RGB camera and depth sensor. Local ANC in a diffuse sound field was reviewed using the introduction of a secondary source. The near field or localised zone of quiet was improved by developing a new method which combines conventional virtual sensing methods with head tracking using a unique coordinate system. A detailed simulation of the adaptive LMS moving virtual microphone method for an aircraft passenger is currently being generated.

Moreover, a head tracking scheme, which utilizes HSV parameters and face cascading methods for an acoustic dummy head, was developed in order to improve the near-field zone of quiet for broadband noise near an aircraft passenger seat. The head tracking measurements proved that the translation of the head could be tracked within ± 0.2 cm, whereas the head rotation angle could be matched to the closest degree which is acceptable. The theory and methods developed in this work will aid in the accurate prediction of zones of quiet for various multiple-input multiple-output (MIMO) local ANC configurations, which is suggested for future work.

REFERENCES

- [1] Guigou, C., & Fuller, C. R. (1999). Control of aircraft broadband noise with foam-PVDF smart skin. *Sound and Vibration*, pp. 541 – 557.
- [2] Viola, P., & Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *Accepted Conference on Computer Vision and Pattern Recognition*.
- [3] Serizel, R., Moonen, M., Wouters, J., & Jensen, S. H. (2012). A Zone-of-Quiet Based Approach to Integrated Active Noise Control and Noise Reduction for Speech