Journal of Engineering and Applied Sciences 13 (11): 4038-4045, 2018
ISSN: 1816-949X

Development of Malay Language Based Spatial Audio Simulator for Auditory Training Software

[1]Abrar Hussain, [1]Kalaivani Chellappan and [2]Siti Zamratol Mai-Sarah Mukari
[1]Department of Electrical, Electronic and Systems Engineering,
Faculty of Engineering and Build Environment, National University of Malaysia,
43600 Bangi, Malaysia
[2]Institute of Ear, Hearing and Speech, National University of Malaysia,
50300 Kuala Lumpur, Malaysia

Abstract: Auditory training software operating with embedded technology of virtual auditory space is a new concept in Malay language application. This concept stems from the necessity of maintaining cultural appropriateness. Virtual auditory space is created by employing spatial audio simulator. This simulator can convey sound to the listeners through headphone and loudspeaker from distinct directions, so that, listeners can perceive the sounds coming from various spatial positions in space by using their spatial hearing mechanism. In this study, three dimensional virtual auditory space is created by pre-synthesizing Malay speech with time-domain Head-Related Impulse-Response (HRIR) data that acquired from KEMAR database. For this study, simulations are done under the platform of Intel® Core ™ i5-2450M CPU@2.5 GHz. Moreover, IDT high definition audio codec is used as a built-in sound controller. The current study uses MATLAB tool for the simulation of virtual auditory space. Results of spatialization are analyzed through the spectrogram analysis of the spatialized signals and show that spatial rendering of speech sound is possible.

Key words: Virtual auditory space, spatial hearing, binaural technique, head-related transfer function, auditory training, platform

INTRODUCTION

One of the most sophisticated biological abilities of human being is spatial hearing. Spatial hearing develops extensively through learning and adaptation to gain added accuracy and better performance in complex audio environments. Spatial hearing is often studied using arrangements enabling the signals entering the two ears to be controlled separately (Pulkki and Karjalainen, 2015). This control eases the binaural hearing which is collateral to the concept of spatial hearing. The underlying process, can be credited to binaural hearing is auditory spatial processing (Vause and Grantham, 1998). Binaural hearing (Hawley *et al*., 2004) that uses spatial separation of speech sources further helps in effective speech understanding in noise as they provide basic binaural cues of ITD and ILD in the closed acoustic environment of headphone. These cues are first detected in the peripheral ear and interpreted in the central auditory pathways. Older adults are less served with these cues (Warren *et al*., 1978) and hence, left-right sound distinction. This difficulty is attributed to spatial processing inability in general and Spatial Processing Disorder (SPD) in particular. SPD mainly occurs due to inability of use of binaural cues to selectively attend to sounds from one direction while simultaneously suppressing sounds arriving from another. SPD problem cannot largely be solved by earlier formal auditory training software as they did not include spatial separation of speech and noise (Tyler *et al*., 2010). Recently, children who have been diagnosed with spatial processing disorder (Cameron and Dillon, 2011) are remediated with auditory training software.

Development of spatial audio simulator depends on the creation of virtual auditory space. It is a complex yet efficient signal processing method for binaural hearing. It has many applications in clinical diagnosis. In the event of clinical diagnosis (Bronkhorst and Plomp, 1988; Peissig and Kollmeier, 1997), speech intelligibility measurement in virtual free field is a dominated area of auditory spatial processing (Hawley *et al*., 2004, 1999; Bronkhorst and Plomp, 1988; Peissig and Kollmeier, 1997; Ozimek *et al*., 2013; Drullman and Bronkhorst, 2000; Nelson *et al*., 1998, 1999; Yost *et al*., 1996; Crispien and

Corresponding Author: Abrar Hussain, Department of Electrical, Electronic and Systems Engineering,
Faculty of Engineering and Build Environment, National University of Malaysia, 43600 Bangi, Malaysia

Ehrenberg, 1995; Ericson and McKinley, 1997, 2001; Gilkey and Anderson, 2014; Yang and Hodgson, 2007; MacDonald *et al.*, 2002; McKinley *et al.*, 1994; Brown *et al.*, 2010). Moreover, authentic reproduction of sound is imperative for different training tasks (Pulkki and Karjalainen, 2015). This technique helps in spatial hearing in particular while listeners use binaural headphones. Two channel headphone listening is an important part in binaural hearing which is inspired by binaural technique of recording and reproducing sounds. The advantage with binaural recordings and reproduction is the capability to record, process, store and reproduce 3-dimensional sound fields with high realism with only two channels. Auditory scenarios, generated by usual recordings and headphone presentation do not have the natural ear-specific auditory cues produced with a free-field presentation. However, headphone speech test results are often taken as an exact depiction of real-world performance. Spatial audio solutions using headphones have been around for a long time in real-time applications but yielding spatial cues that more closely simulate real life accuracy has been a computational issue and has often been solved by hardware solutions. This has long been a restriction but now with more powerful computers this is becoming a lesser and lesser concern and software solutions are now applicable. Additionally, auditory displays of virtual reality have been advocated as a means of facilitating multichannel listening process (Begault and Wenzel, 1993; Burdea *et al.*, 1996; Doll and Hanna, 1995; Doll *et al.*, 1992; King and Oldfield, 1997; Noro *et al.*, 1996; Ricard and Meirs, 1994). This kind of system can greatly enhance auditory interfaces to computers to improve the sense of presence for virtual reality simulations and enhancement of acoustics.

Historically, methods to synthesize real life sound directions through stereo have been developed. A general method to synthesize spatial sound is through the exploitation of Head-Related Transfer Functions (HRTFs), a Digital Signal Processing (DSP) method. Previous studies have shown that the use of HRTFs for virtual environments is suitable when trying to simulate convincing binaural audio for localization of sound sources in three dimensions (Bronkhorst, 1995; Moller *et al.*, 1995, 1996; Murphy and Neff, 2010). In fact, digital signal processing can employ filters to reproduce the same spectral response that an individual would receive at his or her two eardrums. These techniques can achieve a virtualization of sound under headphones and simulate realistic virtual spatial locations and listening environments (Koehnke and Besing, 1996; Wenzel *et al.*, 1990; Wightman and Kistler 1989a, b). When filtered appropriately, sounds are heard externalized in headphone presentation. In other words, listening through traditional headphones gives the listener the illusion that stimuli emanate from external locations as in a natural listening environment. Several studies in recent years have attempted to simulation under headphones of the actual acoustic signals that would arrive at the eardrums of the listener if the sound source were in a specific location relative to the listener (Koehnke and Besing, 1996; Wenzel *et al.*, 1990; Wightman and Kistler, 1989a, b).

In this study, the overview on HRTF function is described in the very next section followed by other sections which describe materials and methods, results and discussion conclusion.

Head-related transfer function: Head-related Transfer Function (HRTF) is a technique used in processing signals that travels through space towards the ear canal of body. HRTF is a very simple function in virtual auditory model compared to free field model. HRTF is important for front-back, up-down discrimination of sound maintaining sources directly ahead of the listener. A pair of HRTFs can be used to synthesize a binaural sound that appears to attain from a particular point in space for a pair of ears. The monaural cues originate from the interaction between the sound source and the human anatomy. The original source sound is modified before it infiltrates the ear canal for processing by the auditory system. These modifications encode the source location and to be captured via. an impulse response which relates the source location and the ear location. This impulse response is termed as Head-Related Impulse Response (HRIR). HRIRs have been used to produce virtual surround sound (Begault, 1994; So *et al.*, 2006). Convolution of an arbitrary source sound with the HRIR converts the sound will be heard by the listener if it had been played at the right source location with the listener's ear at the right receiver location; if it had been played at the left source location with the listener's ear at the left receiver location. A monophonic sound signal can be virtually positioned in any direction in headphone listening if the HRTF's for both ears are available for the desired virtual source direction (Moller *et al.*, 1995; Xie, 2013). A signal x_m, meant to be perceived to be arriving from a assigned direction, expected to convolve with the HRIR pair $\{H_r, H_l\}$ measured with the source in the same direction and the convolved signals are defined mathematically as:

$$y_l = H_l \otimes x_m \quad (1)$$

$$y_r = H_r \otimes x_m \quad (2)$$

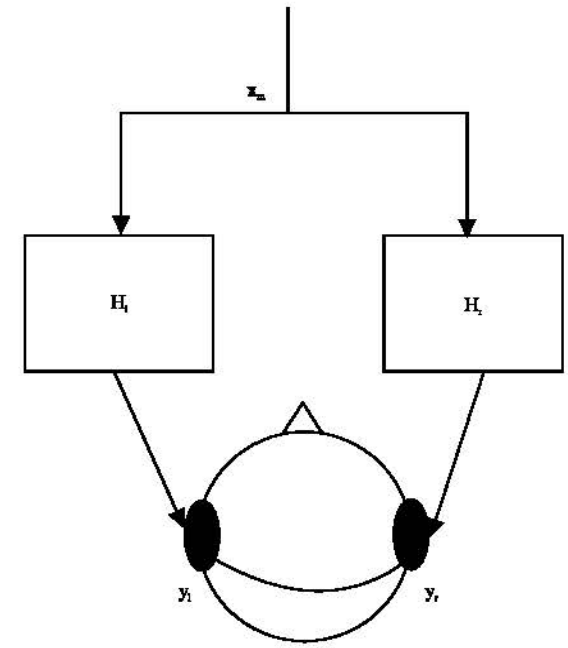

Fig. 1: Creating a virtual source with HRTF processing (Wightman and Kistler, 1989a, b)

From Eq. 1 and 2, it is seen that H_l and H_r are the HRIR signals from left and right ear, respectively are applied to the headphone as shown in Fig. 1. As the decay time of the HRIR is consistently less than a few milli seconds, 256-512 taps in the filters are sufficient for sampling rate of 44.1 kHz. The method ideally reproduces the ear canal signals that would have been produced if the sound source existed in the desired direction.

HRIR data is measured from the sound that arrived into their ear and a dummy head is used to complete this objective. Dummy heads are more accustomed and do not suffer from geometrical variations of head in the front-back, up-down for each individual listener. The impulse responses measured using dummy head can be used for the virtualization of the sound. The widely chosen dataset for HRIR was recorded from Massachusetts Institute of Technology (MIT) and published by Bill Gardner and Keith Martin (George and Martin, 1994).

MATERIALS AND METHODS

Recorded digital stereo speech signals from the laptop are converted into mono by using Sound Forge11 Software to get compatible with the HRIR signals which are also mono in nature. Subsequently mono speech signals go through the convolution process with HRIR signal. Convolved signals are converted to stereo using

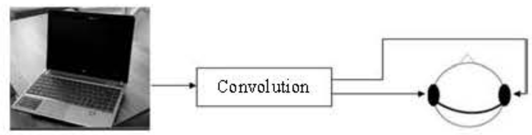

Fig. 2: Flow diagram for convolution and playback

MATLAB coding. Finally, these convolved signals are passed to headphone for playback. The total procedure can be depicted in a flow diagram in Fig. 2.

Data preparation

HRIR data preparation: The specific MATLAB function only accesses the raw HRIR data from KEMAR database that was attained previously by KEMAR measurements. KEMAR database is useful for broadband input, i.e., speech for differentiating the sound from front to back. The provided folder 'compacttar' contains the recorded data at elevations ranging in 10° intervals from 40 through to 90°, each sorted into respective folders. Within these folders featured data for 0, 45, 90 and 180° on the horizontal plane or the Azimuth are obtained. These horizontal angles are conscripted to conform with the fact that listening is most difficult at these angles for loosing accuracy. Moreover, many of the intelligibility measurement are done using these angles (Hawley *et al*., 2004, 1999). The HRTF data that are used in this study are compact as to produce pure speech synthesis. On the top of that are effective for 3D audio, since, they are smaller (128 samples). Instead of using a response recorded from both the right ear and the left ears, each HRTF is based on only left ear measurements and utilizes the idea of symmetry. For example, a sound source at 90° and 0 elevation or directly to the left of the left ear would use the left 90° HRTF for the left ear and left 270° for the right ear. KEMAR has two different pinnae. The left pinna is normal, the right pinna is the large red model and consequently HRIR's are not identical.

Preparation of audio content: The speech data, competing speech data and the multi-talker babble speech data are acquired from the compiled database prepared by Audiology and Speech Science Clinic, Universiti Kebangsaan Malaysia, in Malay language. The speech data are from female speakers. The competing speech data are multi-talker babble and two male talker speech who speak simultaneously. Speech data are linguistically correct and semantically meaningful.

Development of simulator

Convolution: The target speech is convolved with HRIR data that is accessed for 0° Azimuth and presented to the

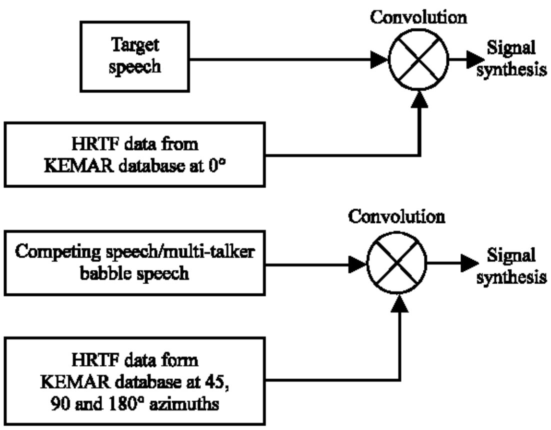

Fig. 3: Convolution of the sound sources

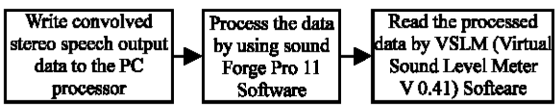

Fig. 4: Block diagram for the evaluation of spatial simulator

listener binaurally through headphones. The apparent source position can be changed by selecting the appropriate pair of HRIRs. Therefore, the two-talker competing speech and multi-talker babble speech are convolved with HRIR data which is accessed for 45, 90 and 180° Azimuth. Figure 3 illustrates this task. The convolution process is done for both the left and right channels of the final signal and summed into a final output stereo array.

Evaluation of simulator: Figure 4 presents the block diagram for evaluating the simulator. The convolved speech data are produced as stereo output signal. By using MATLAB tool, these data are written as a '.wav' format file to the system PC. Further the data are again processed by the Sound Forge Pro 11 Software. The processing consists of the 'batch conversion' for the incoming stereo output signal. In this batch conversion, stereo signal is stored as mono signal. This mono output is required for the VSLM Software to be processed for spectrogram analysis.

Playback: The output from the stereo array is played back through the inbuilt sound function from MATLAB tool.

RESULTS AND DISCUSSION

MATLAB GUI interface for spatial auditory simulator: Figure 5 shows the spatial auditory simulator using

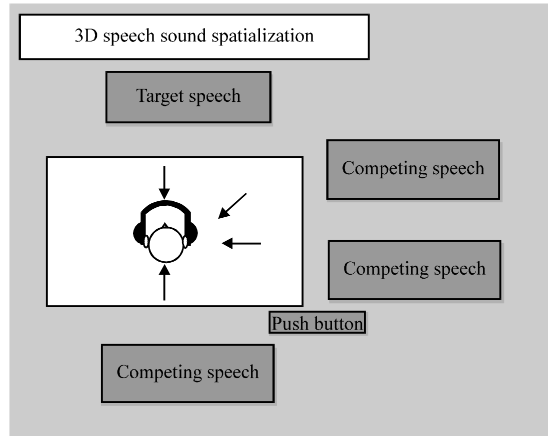

Fig. 5: Spatial auditory simulator

MATLAB Software. The spatialization system virtually places target speech at 0° position, so that, listener with the Sennheiser HD215 headphone can perceive the speech emanating from 0° location. By the same token, listeners perceive the competing speeches from 45, 90 and 180° location. The subsequent subsections evaluate the spatialization of speech sound by using spectrogram analysis.

Spectrogram analysis: Spectrograms use the periodogram power spectrum which is a non-parametric estimation method and are extensively used by speech and audio engineers. The human hearing mechanism is based on the time-frequency analysis of ear canal signals (Pulkki and Karjalainen, 2015). In this research, narrowband spectrogram analysis is conducted by the Virtual Sound Level Meter (VSLM) Software. FFT size of 4096 is chosen to visualize the rapid change of frequency in measured time slots. Therefore, frequency resolution is favorable. The frequency resolution is 10.7 Hz. Hanning window is used as it touches zero at both ends, removing any discontinuity. It has wide peak but accurate low side lobes.

Figure 6-15 visually represent how the input speech signal, competing speech and multitalker babble speech signals are filtered by the impulse responses. The frequency dependent peaks and troughs are seen in the filtered signals. It shows how the input speech audio signal is manipulated by head-related-impulse response signals to produce the listeners percieveness for directivity change. The spectrograms are shown as 3D plot where amplitude is encoded as a height as well as a colour.

From Fig. 6, it is evident that target speech is denser at high frequency region due to female voice. For target

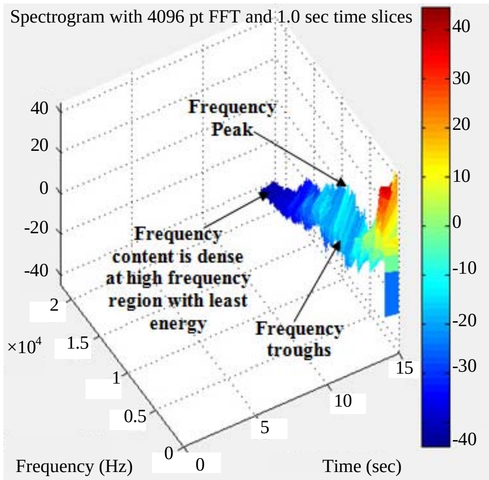

Fig. 6: Spectrogram of recorded mono output speech signal

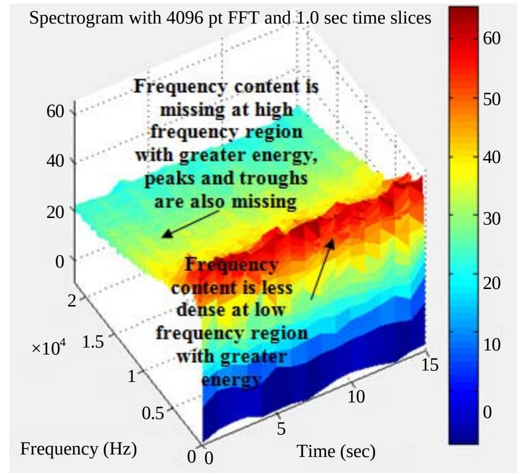

Fig. 9: Spectrogram of mono output two talker competing speech signal for 45° spatial position

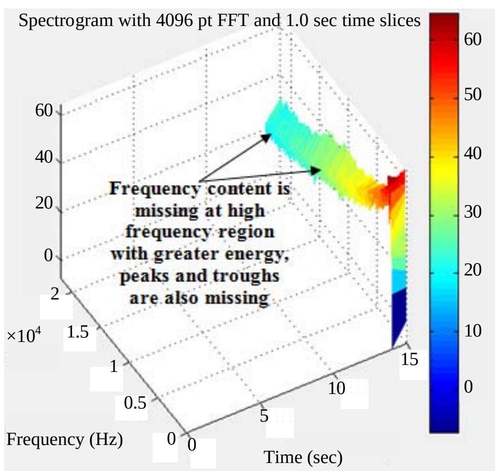

Fig. 7: Spectrogram of mono output speech signal for 0° spatial position

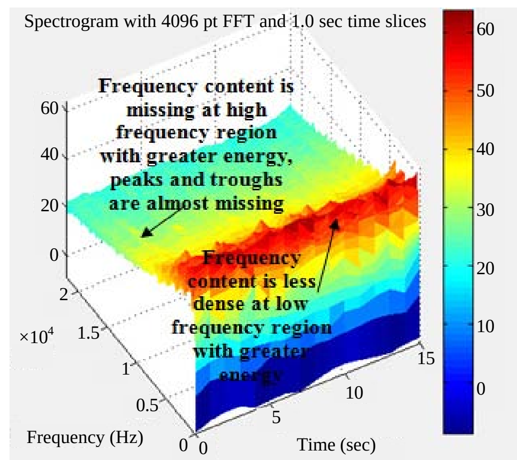

Fig. 10: Spectrogram of mono output two talker competing speech signal for 90° spatial position

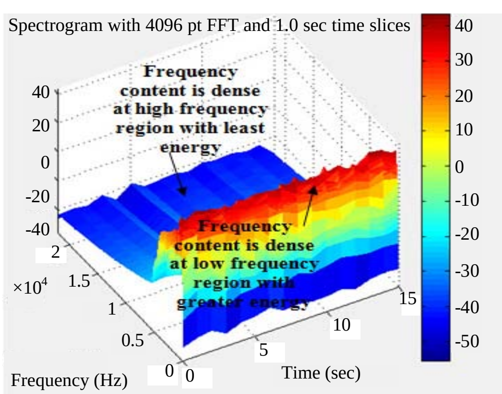

Fig. 8: Spectrogram of recorded mono output two talker competing speech signal

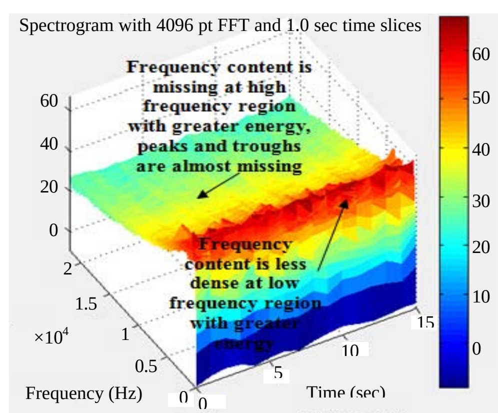

Fig. 11: Spectrogram of mono output two talker competing speech signal for 180° spatial position

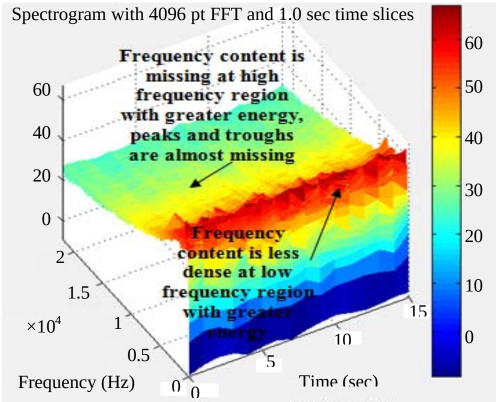

Fig. 12: Spectrogram of recorded mono output multi-talker babble noise signal

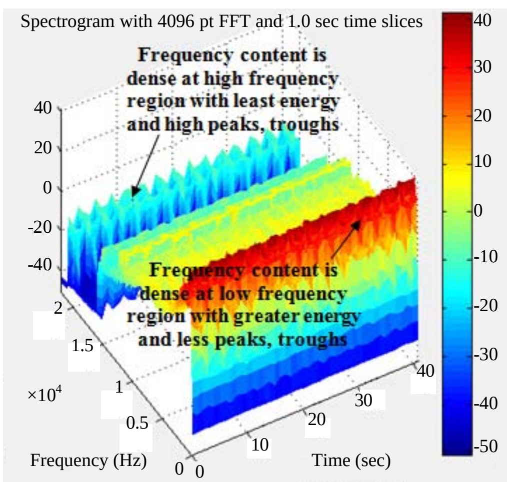

Fig. 13: Spectrogram of mono output multi-talker babble noise for 45° spatial position

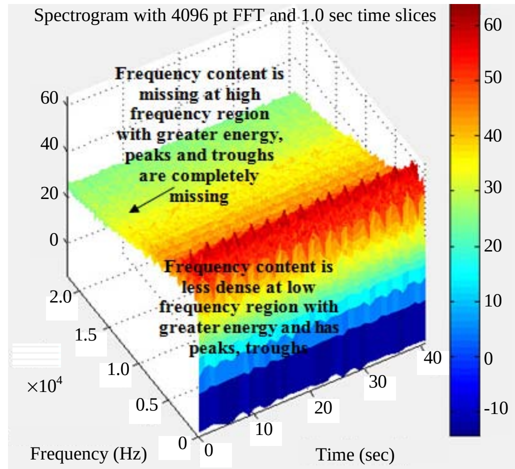

Fig. 14: Spectrogram of mono output multi-talker babble noise for 90° spatial position

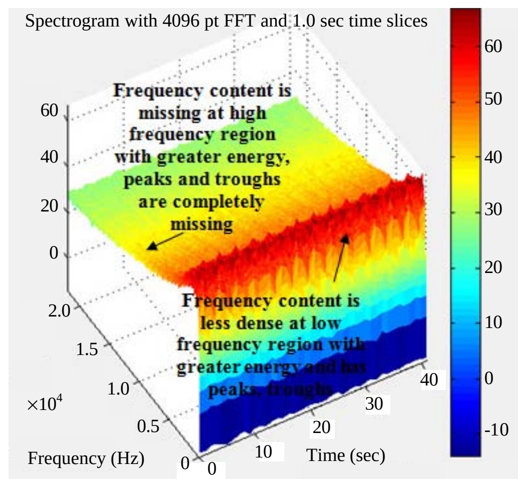

Fig. 15: Spectrogram of mono output multi-talker babble noise for 180° spatial position

speech placed at 0° spatial position, frequency content dramatically decreases below 300 Hz which is shown in Fig. 7. An informal listening test confirms that the output signal is recreated with poor quality when played back via. headphone. Moreover, energy of target speech is intensified at lower frequency region. However, informal listening test ensures that directivity is well maintained.

Figure 8 illustrates the energy distribution across frequency for the competing speech which consists of two male talkers. It is seen that energy distribution is fluctuating in the whole frequency region. Overall, low frequency region is denser due to male voice. It means that more energy mounts up at low frequency area.

Specifically, frequency content suddenly gets low above 300 Hz and as a consequence energy is the lowest above that frequency. For 45, 90 and 180° spatial positions (Fig. 9-11), high energy is accumulated in the higher frequency region compared to competing speech while no spatialization was occurred. In other way, energy in high frequency increases compared to competing speech while spatialization was absent. For all the spatial positions, frequency content is rapidly declined above 300 Hz. Therefore, output signal is highly altered for 45, 90 and 180° spatial positions as high frequency contents are missing. However, directivity is highly preserved for 45 and 90° spatial situations compared to 180° spatial position as perceived by the listeners wearing headphone. This is due to front-back confusion of spatial system.

Figure 12 demonstrates the time-frequency representation of multitalker babble speech. From the spectrogram it is observed that both the lower frequency region and high frequency region contain dense energy. Nevertheless, frequency content gets low in both the high frequency and low frequency part. However, this fall down of frequency is more visible in the high frequency area above 300 Hz.

Figure 13-15 show the spectrograms for spatialized condition of 45, 90 and 180° position respectively. For all the spatial positions, frequency content is quickly turns down above 300 Hz. Therefore, output signal experiences a lesser amount of alternation for 45, 90 and 180° spatial positions and it is not restored with good quality when played back through headphone for these virtual locations. Nonetheless, directivity is again highly preserved for 45 and 90° spatial situations compared to 180° spatial location. This is also due to front-back confusion of spatial system.

From the spatialization system it is evident that for both type of competing speech signals, directivity is not changed for the 45 and 90° position, although, significant high frequency loss is prevalent in the spatialized signals for multitalker babble speech.

CONCLUSION

This study is about the development of a spatial auditory simulator which will be used will be used as one of the embedded modules in the auditory training software. The proposed Malay language based spatial audio simulator is first of its kind. This system is expected to provide an alternative to the existing clinical based training to be conducted at patient's home with supervision by audiology through the system. It is targeted to ageing population in maintaining the health of their auditory system and also others with hearing impairment.

ACKNOWLEDGEMENT

The researchers would like to express their sincere appreciation to Nurul Huda bt. Bani for her preparation of audio content.

REFERENCES

Begault, D.R. and E.M. Wenzel, 1993. Headphonelocalization of speech. Hum. Factors, 35: 361-376.

Begault, D.R., 1994. 3-D Sound for Virtual Reality and Multimedia. AP Professional, Buffalo, New York, USA., ISBN:9780120847358, Pages: 293.

Bronkhorst, A.W. and R. Plomp, 1988. The effect of head-induced interaural time and level differences on speech intelligibility in noise. J. Acoust. Soc. Am., 83: 1508-1516.

Bronkhorst, A.W., 1995. Localization of real and virtual sound sources. J. Acoust. Soc. Am., 98: 2542-2553.

Brown, D.K., S. Cameron, J.S. Martin, C. Watson and H. Dillon, 2010. The North American Listening in Spatialized Noise-Sentences Test (NA LiSN-S): Normative data and test-retest reliability studies for adolescents and young adults. J. Am. Acad. Audiology, 21: 629-641.

Burdea, G., P. Richard and P. Coiffet, 1996. Multimodal virtual reality: Input-output devices, system integration and human factors. Intl. J. Hum. Comput. Interact., 8: 5-24.

Cameron, S. and H. Dillon, 2011. Development and evaluation of the LiSN and Learn auditory training software for deficit-specific remediation of binaural processing deficits in children: Preliminary findings. J. Am. Acad. Audiology, 22: 678-696.

Crispien, K. and T. Ehrenberg, 1995. Evaluation of the-cocktail-party effect-for multiple speech stimuli within a spatial auditory display. J. Audio Eng. Soc., 43: 932-941.

Doll, T.J. and T.E. Hanna, 1995. Spatial and spectral release from masking in three-dimensional auditory displays. Hum. Factors, 37: 341-355.

Doll, T.J., T.E. Hanna and J.S. Russotti, 1992. Masking in three-dimensional auditory displays. Hum. Factors, 34: 255-265.

Drullman, R. and A.W. Bronkhorst, 2000. Multichannel speech intelligibility and talker recognition using monaural, binaural and three-dimensional auditory presentation. J. Acoust. Soc. Am., 107: 2224-2235.

Ericson, M.A. and R.L. McKinley, 1997. The Intelligibility of Multiple Talkers Separated Spatially in Noise. In: Binaural and Spatial Hearing in Real and Virtual Environments, Gilkey, R.H. and T.R. Anderson (Eds.). Lawrence Erlbaum Associates, New Jersey, USA., pp: 701-724.

Ericson, M.A. and R.L. McKinley, 2001. The intelligibility of multiple talkers separated spatially in noise. AFRL-HE-WP-SR-2001-0009, Air Force Research Laboratory, Ohio, USA.

George B. and K. Martin, 1994. HRTF measurements of a KEMAR dummy-head microphone. MIT Media Lab, Cambridge, Massachusetts, USA. http://sound. media. mit.edu/resources/KEMAR.html.

Gilkey, R.H. and T.R. Anderson, 2014. Binaural and Spatial Hearing in Real and Virtual Environments. Psychology Press, Hove, England, UK., ISBN-13:978-0-805-81654-9, Pages: 795.

Hawley, M.L., R.Y. Litovsky and H.S. Colburn, 1999. Speech intelligibility and localization in a multi-source environment. J. Acoust. Soc. Am., 105: 3436-3448.

Hawley, M.L., R.Y. Litovsky and J.F. Culling, 2004. The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. J. Acoust. Soc. Am., 115: 833-843.

King, R.B. and S.R. Oldfield, 1997. The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays. Hum. Factors, 39: 287-295.

Koehnke, J. and J.M. Besing, 1996. A procedure for testing speech intelligibility in a virtual listening environment. Ear Hearing, 17: 211-217.

MacDonald, J.A., J.D. Balakrishnan, M.D. Orosz and W.J. Karplus, 2002. Intelligibility of speech in a virtual 3-D environment. Hum. Factors, 44: 272-286.

McKinley, R.L., M.A. Erickson and W.R. D'Angelo, 1994. 3-Dimensional auditory displays: Development, applications and performance. Aviat. Space Environ. Med., 65: A31-A38.

Moller, H., M.F. Sorensen, C.B. Jensen and D. Hammershoi, 1996. Binaural technique: Do we need individual recordings?. J. Audio Eng. Soc., 44: 451-469.

Moller, H., M.F. Sorensen, D. Hammershoi and C.B. Jensen, 1995. Head-related transfer functions of human subjects. J. Audio Eng. Soc., 43: 300-321.

Murphy, D. and F. Neff, 2010. Spatial Sound for Computer Games and Virtual Reality. In: Game Sound Technology and Player Interaction: Concepts and Developments, Grimshaw, M. (Eds.). Information Science Reference, Hershey, Pennsylvania, USA., ISBN:978-1-61692-830-8, pp: 287-312.

Nelson, W.T., R.S. Bolia, M.A. Ericson and R.L. McKinley, 1998. Monitoring the simultaneous presentation of spatialized speech signals in a virtual acoustic environment. Proceedings of the 1998 Conference on IMAGE, June 22, 1998, Image Society Inc., Chandler, Arizona, pp: 159-166.

Nelson, W.T., R.S. Bolia, M.A. Ericson and R.L. McKinley, 1999. Spatial audio displays for speech communications: A comparison of free field and virtual acoustic environments. Proc. Hum. Factors Ergon. Soc. Ann. Meeting, 43: 1202-1205.

Noro, K., T. Kawai and H. Takao, 1996. The development of a dummy head for 3-D audiovisual recording for transmitting telepresence. Ergon., 39: 1381-1389.

Ozimek, E., J. Kocinski, D. Kutzner, A. Sek and A. Wicher, 2013. Speech intelligibility for different spatial configurations of target speech and competing noise source in a horizontal and median plane. Speech Commun., 55: 1021-1032.

Peissig, J. and B. Kollmeier, 1997. Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. J. Acoust. Soc. Am., 101: 1660-1670.

Pulkki, V. and M. Karjalainen, 2015. Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics. John Wiley & Sons, Hoboken, New Jersey, USA., ISBN:978-1-118-86654-2, Pages: 98.

Ricard, G.L. and S.L. Meirs, 1994. Intelligibility and localization of speech from virtual directions. Hum. Factors, 36: 120-128.

So, R.H.Y., N.M. Leung, J. Braasch and K.L. Leung, 2006. A low cost, non-individualized surround sound system based upon head related transfer functions: An ergonomics study and prototype development. Appl. Ergon., 37: 695-707.

Tyler, R.S., S.A. Witt, C.C. Dunn and W. Wang, 2010. Initial development of a spatially separated speech-in-noise and localization training program. J. Am. Acad. Audiology, 21: 390-403.

Vause, N.L. and D.W. Grantham, 1998. Speech intelligibility in adverse conditions in recorded virtual auditory environments. Proceedings of the 1998 International Conference on Auditory Display (ICAD'98), November 01-04, 1998, ACM, Swindon, England, UK., pp: 28-28.

Warren, L.R., J.W. Wagener and G.E. Herman, 1978. Binaural analysis in the aging auditory system. J. Gerontology, 33: 731-736.

Wenzel, E.M., S.S. Fisher, P.K. Stone and S.H. Foster, 1990. A system for three-dimensional acoustic visualization in a virtual environment workstation. Proceedings of the 1st Conference on Visualization, October 23-26, 1990, IEEE, San Francisco, California, USA., ISBN:0-8186-2083-8, pp: 329-337.

Wightman, F.L. and D.J. Kistler, 1989a. Headphone simulation of free-field listening II: Psychophysical validation. J. Acoust. Soc. Am., 85: 868-878.

Wightman, F.L. and D.J. Kistler, 1989b. Headphone simulation of free-field listening. I: stimulus synthesis. J. Acoust. Soc. Am., 85: 858-867.

Xie, B., 2013. Head-Related Transfer Function and Virtual Auditory Display. 2nd Edn., J. Ross Publishing, Fort Lauderdale,Florida,USA.,ISBN-13:978-1-60427-070-9, Pages: 486.

Yang, W. and M. Hodgson, 2007. Validation of the auralization technique: Comparative speech-intelligibility tests in real and virtual classrooms. Acta Acustica United, 93: 991-999.

Yost, W.A., R.H. Dye and S. Sheft, 1996. A simulated cocktail party with up to three sound sources. Percept. Psychophysics, 58: 1026-1036.