LISTENING EFFORT BETWEEN AUDITORY-ONLY, VISUAL-ONLY AND AUDITORY-VISUAL SPEECH PERCEPTION IN NORMAL HEARING LISTENERS AS MEASURED USING PUPILLOMETRY

Saiful Adli Jamaluddin[a], Nur Asyiqin Khairol Azmi[a], Mohd. Zulfaezal Bin Che Azemin[b], Nurlin Ali Hanafiah[a], Greg O'Beirne[c]
[a] Department of Audiology and Speech-Language Pathology, Kulliyyah of Allied Health Sciences, [b] Department of Optometry and Visual Sciences, International Islamic University Malaysia, Kuantan, Pahang, Malaysia.
[c] Department of Speech and Hearing, School of Psychology, University of Canterbury, New Zealand.

INTRODUCTION

Listening effort is the purposeful allocation of the mental resources used to overcome difficulties when listening to speech in a noisy environment. Increased cognitive load may result in decreased speech perception performance and mental fatigue. Pupillometry can be applied to investigate listening effort where changes in pupil diameter can be measured as a function of cognitive load. By measuring peak pupil dilation and latency, we can gauge listening effort by changing listening factors such as sentence intelligibility, type of interrupting noise to speech perception, and speech perception modalities. This study aims to establish the baseline of listening effort in normal hearing young listeners in three modalities of speech perception which are auditory-only, visual-only, and auditory-visual.

RESEARCH OBJECTIVES

1. Describe average peak pupil dilation (PPD) and peak latency in auditory-only, visual-only and auditory-visual modes by pupillometry in three different signal–to-noise ratios using the Malay Matrix Sentence Test (MMST).
2. Compare peak pupil dilation (PPD) and peak latency in auditory-only (A-only) in three different signal–to-noise ratios.
3. Compare peak pupil dilation (PPD) and peak latency in visual-only (V-only) vs. auditory-visual (AV) speech perception in three different signal–to-noise ratios.

PUPILLOMETRIC DATA

Recording unit	Pupil Pro Headset (Pupil Labs, Berlin Germany)
Cameras (two units)	1. Scene Camera (eye-to-screen calibration) 2. Infrared hi-speed (pupil recording)
Pupil recording	1. 2000 frames per second 2. Accuracy (0.60º) 3. 2D & 3D eye model 4. Pupil diameter relative size in eye camera pixels in mm
Recording resolution	800 x 600 pixels at 30 Hz
Recording setup	

SPEECH PERCEPTION STIMULI

Speech test	Auditory-Visual Malay Matrix Sentence Test - everyday Malay sentences in varying levels of signal-to-noise ratio presented in A-only, V-only and AV test modalities
Output	Sound Blaster X 5.1 external soundcard (Creative Technology Ltd., Singapore) Sennheiser HD 280 Pro headphones (Sennheiser GmbH, Germany)
Recording task	Closed-set task (participant select words perceived based on collection of words seen on computer screen)
Sample of auditory-visual stimulation presented to participants	

METHODS

Participants' head were positioned at approximately 50 cm away from the centre of a 22-inch LCD computer monitor (forehead to screen). They sat on a adjustable chair and strapped on a chin rest to ensure a constant distance between the head and the screen. Visual field calibration was achieved by looking at five points in the visual field (centre, top left and right, and bottom left and right). Participants were required to complete a training list for the MMST speech perception task (20 sentences per task). This is subsequently followed by randomly assigned MMST tests in three modalities (A, V, and AV) & three signal–to-noise ratio: -7, -5 and -2 dB SNR (lower SNR indicates more demanding listening). Pupil dilation were recorded for each set of trials. Adequate breaks were given to avoid fatigue.

STUDY SAMPLE

Twenty Malay native speakers (15 females and 5 male) were recruited, however ten out of the 20 participants were excluded due to hearing loss (n = 2), history of difficulty in listening in noise (n = 1) and unreliable pupil data (n = 7). Three male and seven female participants form the final number of experiment group.

RESULTS & ANALYSIS

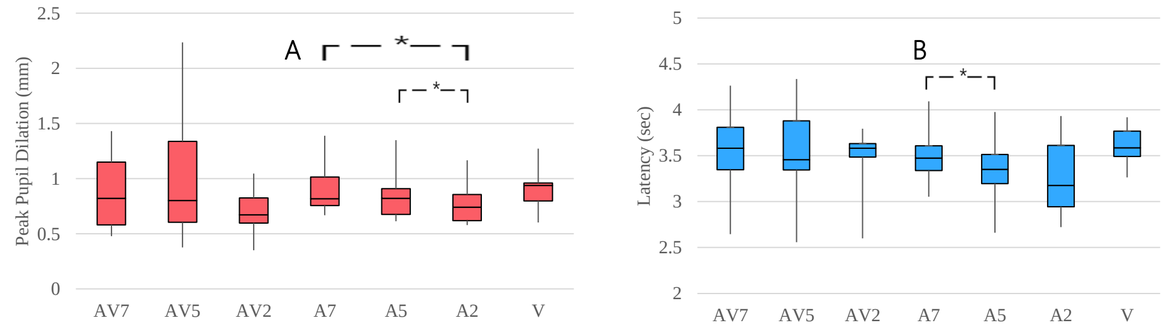

Figure 1. Peak Pupil Dilation (A) and latency (B) in multimodal speech perception trials where, A7 = A-only at -7dB SNR; A5 = A-only at -5dB SNR; A2 = A-only at -2dB SNR; AV7 = AV at -7dB SNR; AV5 = AV at -5dB SNR; AV2 = AV at -2dB SNR; & V = V-only)

Shapiro-Wilk test revealed the data to be not normally distributed ($p > .05$) for both dilation and latency. The Friedman test result showed statistically significant differences observed in A-only, V-only, and AV depending on SNRs where, $X2(5) = 43.543$, $p = .000$, $X2(5) = 40.645$, $p = .000$ $X2(7) = 56.267$, $p = .000$ respectively. Subsequently, the PPD and peak latencies were compared using Wilcoxon Signed Rank Test within each three modalities (A-only, V-only and AV) and SNRs (-7, -,5 and -2 dB SNR). Result revealed significant main effect of SNRs in A-only towards PPD between A2 and A7, and A2 and A5 where, $Z = -2.092$ and -2.091, $p < .05$ and latency between A5 and A7 where, ($Z = -2.807$, $p = .005$).

DISCUSSION & CONCLUSION

The finding suggests that the PPD increased significantly with decreasing SNR in A-only modality between A7 vs. A2, and A5 vs. A2. The peak PPD latency across SNRs in AV were inconsistent with previous studies. The finding suggests that the PPD increased significantly with decreasing SNR in A-only modality between A7 vs. A2, and A5 vs. A2. Meanwhile PPD latency varies significantly between A5 vs. A7 only. Current study demonstrates pupillometry as an objective measure of listening effort in speech perception in noise. The finding of the study showed PPD and peak latency systematically increases as SNR decreases in A-only modality, revealing more effort is exerted at acoustically demanding conditions. More data is necessary to make further conclusions.