Temporal alignment of pupillary response with stimulus events via deconvolution [a)]

Daniel R. McCloy, Eric D. Larson, Bonnie Lau, and Adrian KC Lee
Institute for Learning and Brain Sciences, University of Washington, 1715 NE Columbia Rd., Box 357988, Seattle, WA, 98195-7988
drmccloy@uw.edu, larsoner@uw.edu, blau@uw.edu, akclee@uw.edu

Running title: Deconvolution of pupil size

[a)]Portions of the research described here were previously presented at the 37th Annual MidWinter Meeting of the Association for Research in Otolaryngology.

ABSTRACT

Analysis of pupil dilation has been used as an index of attentional effort in the auditory domain. Previous work has modeled the pupillary response to attentional effort as a linear time-invariant system with a characteristic impulse response, and used deconvolution to estimate the attentional effort that gives rise to changes in pupil size. Here we argue that one parameter of the impulse response ($t_{\max}$) has been mis-estimated in the literature; we present our own estimate and show how deconvolution with our value of $t_{\max}$ yields more intuitively plausible and informative results.

Keywords: listening effort, pupillometry, deconvolution

I. INTRODUCTION

Pupillometry, the tracking of pupil diameter, has been used to measure attentional effort,[1,2] including in the auditory domain.[3--5] The pupillary response to attentional effort has been modeled as a linear time-invariant system comprising a train of theoretical "attentional pulses" and a characteristic impulse response approximated by an Erlang gamma function:

$$h = t^n e^{\frac{-nt}{t_{\max}}} \tag{1}$$

The impulse response h has empirically-determined parameters for the latency of response maximum $t_{\max}$ and the shape parameter of the Erlang distribution n; the latter is proposed to be analogous to the number of steps in the neural signalling pathway transmitting the attentional pulse to the pupil.[6] This model allows estimation of the timing and magnitude of the attentional signal by deconvolving the measured pupillary response using the estimated impulse response function as a deconvolution kernel,[7] in a method similar to that used in fMRI analysis of the BOLD response.

Hoeks and Levelt have empirically estimated the kernel parameters n=10.1 and $t_{\max}$=0.91 s using both auditory and visual stimuli, but a crucial shortcoming was the inclusion of button-press responses in all trials used for parameter estimation (non-button-press trials were included in their experimental design, but they report pupillary responses to these trials were "too small and noisy for further data analysis").[6] This is problematic in light of recent findings showing that up to 70% of pupil response can be attributed to preparatory and motor commands in tasks with button-presses, with effects beginning as early as 400 ms prior to the button press event.[8] In consequence, Hoeks and Levelt's estimate of the latency of response maximum ($t_{\max}$) may be inappropriate for processing pupillary responses to stimuli absent of motor responses. For this reason, we re-estimated $t_{\max}$ for both target (with button press) and non-target (no button press) auditory stimuli (Experiment 1), and show how our estimate of $t_{\max}$ yields better temporal alignment of stimulus and deconvolved pupil response in an auditory attention

switching task (Experiment 2), when compared to deconvolution using previous estimates. We expect the improvement in temporal alignment between stimulus and pupil response to be useful in addressing questions related to cognition, listening effort, and auditory attention.

II. GENERAL METHODS

All procedures were performed in a sound-treated booth illuminated only by the LCD monitor on which visual stimuli were presented. Auditory stimuli were delivered over Etymotic ER-2 insert earphones via a TDT RP2 real-time processor (Tucker Davis Technologies, Alachula, FL) at a level of 65 dB SPL. Pupil size was measured continuously at a 1000 Hz sampling frequency using an EyeLink1000 infra-red eye tracker (SR Research, Kanata, ON). Participants were seated 50 cm away from the EyeLink camera with their heads stabilized by a chin rest and forehead bar. All participants had normal audiometric thresholds (20 dB HL or better at octave frequencies from 250 Hz to 8 kHz), were compensated at an hourly rate, and gave informed consent to participate as overseen by the University of Washington Institutional Review Board.

III. EXPERIMENT 1

Experiment 1 tested the pupillary response to a simple auditory target detection task. The aim was to compare pupillary response to non-target tones versus response to target tones (with button press response to the target tones) and estimate the latency of maximum pupil response ($t_{\max}$). Ten adults (five female) aged 21 to 35 years (mean 26.6) participated in Experiment 1.

A. Pupil dynamic range

To maximize our ability to detect changes in pupil size, we assessed the dynamic range of each participant's pupil, then selected a background grayscale value for the visual display that yielded a resting dilation where the pupil's response was steepest. We began by presenting a

10-second rest period comprising a black screen with a centered, dark gray fixation dot (value 0.2 on 0–1 scale; 1 = maximum luminance). Next, a series of monochromatic screens with central fixation dots were presented for 3 s each, with background values ranging from 0 (black) to 0.5 (mid-gray) in eight exponential (base-2) steps; on each step the luminance value of the fixation dot was 0.2 higher than the background. After reaching the brightest level, the rest period and series of increasing luminance steps was repeated. To choose the best background value, we calculated median pupil size between 1.25 and 3.0 seconds after each change of screen luminance, averaged those median values across the two repetitions of the calibration sequence, and selected the background value exhibiting the greatest change in pupil size compared to the (darker) level preceding it.

B. Pupil response to auditory stimuli

To determine pupil response to auditory stimuli, participants were asked to respond by button press to tones with frequency modulation (FM) and ignore constant frequency tones. Steady tones were 1000 Hz with a 10 ms Hann window taper at both ends and a total duration of 100 ms. Target tones had a frequency centered at 1000 Hz that varied sinusoidally with a range of 200 Hz and a period matching the duration of the stimulus, and were otherwise identical to the steady tones. Tones were presented in four blocks of 75 stimulus presentations with breaks between blocks; each block began with a 10 second rest period to allow pupil size to stabilize. One-fourth of all tones were target tones, randomly distributed through the task. Inter-stimulus interval was randomly and evenly distributed between 3 and 5 s. Examples of both tone types were played for the listener prior to the task. Three participants repeated the task with standard and target tones swapped, to confirm that pupil responses were insensitive to the small differences between the tone types; these data are not presented.

Pupil size measurements were time-aligned to the onset of each tone and epoched from −0.5 s to 3.0 s. Pupil size was then baseline-corrected relative to the period from −0.5 s to 0.0 s and z-score normalized within each epoch, consistent with Wierda and colleagues' procedure.[7] The

first epoch of each block was excluded, as were epochs with an incorrect behavioral response, and epochs beginning less than 2.5 s after a button press.

C. Results & Discussion

Plots of pupil response to standard and target tones are shown in Figure 1. Response to standard tones shows a peak around 0.5 s after stimulus onset, whereas response to target tones shows an early peak around 0.75 s and a larger, later peak around 1.4 s. Differences in both magnitude and peak latency are attributable to the behavioral response (button press) in the target trials; the differences are consistent with previous work showing that when button press responses occur up to 70% of the pupillary response is attributable to them.[8]

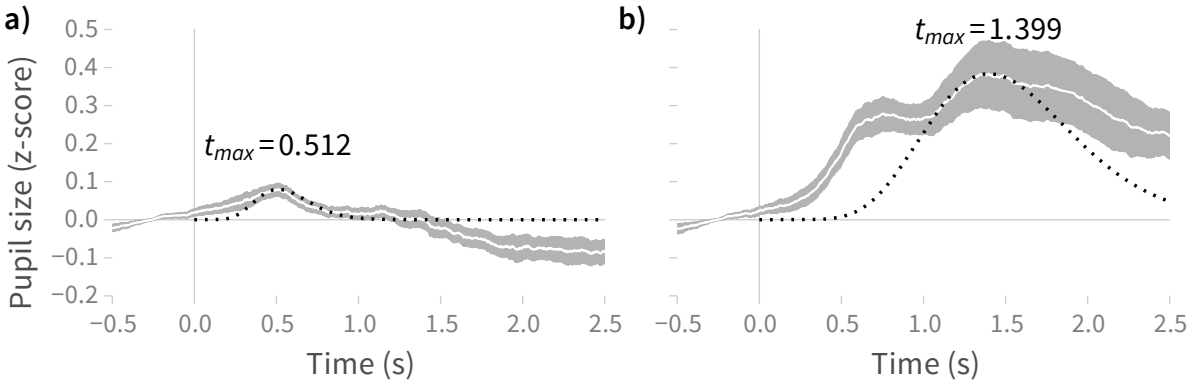

Figure 1: Mean (± 1 standard error) pupil size across subjects in response to (a) steady tones and (b) FM tones, with latency of maximum response ($t_{\max}$) labeled. The late peak for FM tones is attributable to the behavioral response (button press) in those trials. Dark dotted lines show deconvolution kernels calculated from the different $t_{\max}$ values.

Given the simplicity of the stimulus design in this experiment, we can suppose that $t_{\max}$ in the non-target condition (512 ms; Figure 1a) is close to the minimum possible latency for a pupillary change resulting from an auditory stimulus. In contrast, the larger value of $t_{\max}$ (930 ms) derived by Hoeks and Levelt[6] and subsequently used by Wierda and colleagues in their deconvolution analysis[7] likely reflects contributions to pupil dilation from a combination of stimulus, motor planning, and motor command activities (as does our estimate of $t_{\max}$ to target

tones; Figure 1b). As such, our estimate of $t_{\max}$ for non-target tones should yield a more appropriate deconvolution kernel for analysis of pupil responses to auditory stimuli absent a rapid motor response, as well as pupil responses to continuous auditory stimuli. This does not preclude using our estimate of $t_{\max}$ when analyzing auditory tasks that include rapid motor responses: as long as button presses are balanced across experimental conditions, it should still be possible to analyze the difference in (deconvolved) pupil size across conditions by treating the pupillary response to motor planning and execution as noise.

IV. EXPERIMENT 2

To illustrate the effect of appropriate parameterization of the deconvolution kernel in pupillometric analysis, we applied the deconvolution technique of Wierda and colleagues[7] to measurements of pupil size from an auditory attention switching experiment, using estimates of $t_{\max}$ from Experiment 1 and from Hoeks and Levelt.[6] Sixteen adults (eight female) aged 19 to 35 years (mean 25.5) were recruited for Experiment 2. The experiment included two stimulus manipulations (number of noise-vocoder bands; mid-trial gap duration) and one cued behavioral manipulation (maintain attention to one talker throughout, or switch attention between talkers); methods for all three manipulations are described, but for brevity the deconvolution analysis will only be shown for the behavioral manipulation.

A. Stimuli

Stimuli comprised spectrally degraded spoken alphabet letters ADEGOPUV from the ISOLET v1.3 corpus[9] from one female and one male talker. The mean fundamental frequencies of the unprocessed recordings were 103 Hz for the male talker and 193 Hz for the female talker. Letter durations ranged from 351 to 478 ms, and were silence-padded to a uniform duration of 500 ms, RMS normalized, and windowed at the edges with a 5 ms cosine-squared envelope. Two streams of four letters each were generated for each trial, with a gap of either 200 or 600 ms

between the second and third letters of each stream.

Spectral degradation of the letters followed conventional noise vocoding strategy, maintaining temporal and amplitude cues and removing fine structure.[10] The stimuli were fourth-order bandpass filtered into 10 or 20 spectral bands of equal equivalent rectangular bandwidths,[11] with lower and upper bounds of 200 and 8000 Hz. The amplitude envelope of each band was extracted with half-wave rectification and a 160 Hz low-pass fourth order Butterworth filter. The resulting envelopes were used to modulate white noise that had been bandpass filtered at the same frequencies as the extracted bands, and the resulting modulated noise bands were summed and presented diotically at 65 dB SPL. A white-noise masker with π-interaural-phase was played continuously during experimental blocks, to provide additional masking of environmental sounds (e.g., friction between earphone tubes and subject clothing) and to provide parity with follow-up MEG neuroimaging experiments. The masking noise was presented at a level of 45 dB SPL, yielding a stimulus-to-noise ratio of 20 dB.

B. Procedure

Participants were instructed to maintain their gaze on a white fixation dot centered on a black screen throughout test blocks. Each trial began with a 1 s auditory cue (spoken letters "AA" or "AU") indicating (by the sex of the talker) whether to attend first to the male or female voice, and additionally indicating whether to maintain attention to that talker throughout the trial ("AA" cue) or to switch attention to the other talker at the mid-trial gap ("AU" cue). The cue was followed by 0.5 s of silence, followed by the main portion of the trial: two concurrent, dichotic 4-letter streams (one male voice, one female voice), with a variable-duration gap between the second and third letters. The task was to respond by button press to the letter "O" spoken by the target talker (Figure 2). To allow unambiguous attribution of button presses, the letter "O" was always separated from another "O" (in either stream) by at least 1 s, and its position in the letter sequence was balanced across trials and conditions.

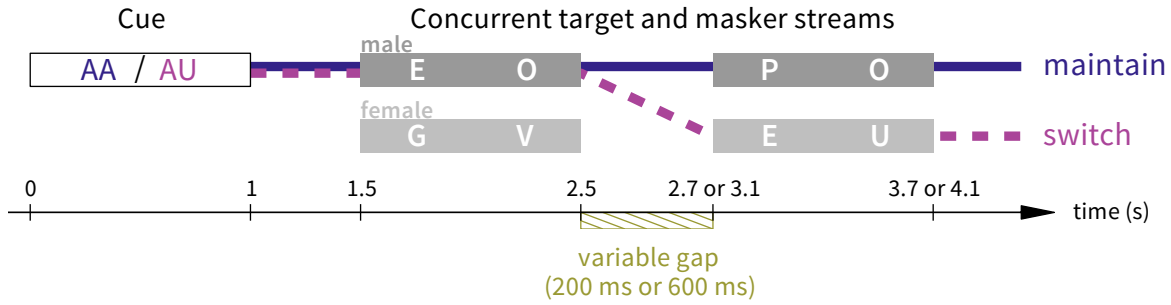

Figure 2: (Color online) Illustration of trial types in Experiment 2. In the depicted "switch" trial (heavy dashed line), listeners would hear cue "AU" in a male voice, attend to the male voice ("EO") for the first half of the trial and the female voice ("EU") for the second half of the trial, and respond once (to the "O" occurring at 2–2.5 s). In the depicted "maintain" trial (heavy solid line), listeners would hear cue "AA" in a male voice, attend to the male voice ("EOPO") throughout the trial, and respond twice (once for each "O").

C. Analysis

Deconvolution kernels were calculated as in Equation 1, with $n = 10.1$ (following Hoeks and Levelt) and values of $t_{\max}$ from both Hoeks and Levelt (930 ms) and from Experiment 1 (512 ms). Fourier transforms of the deconvolution kernels and the subject-level mean pupil size time series indicated no appreciable energy at frequencies above 10 Hz, so for efficiency of computation (and following the procedure of Wierda and colleagues) deconvolved signals were generated as a best-fit linear sum of kernels spaced at 100 ms intervals, as implemented in pyeparse.[12] Statistical comparison of pupil dilation time series was performed using a non-parametric cluster-level one-sample T-test on the within-subject differences between experimental conditions (clustering across time only),[13] as implemented in mne-python.[14]

D. Results & Discussion

Deconvolved pupil size for the behavioral contrast "maintain" versus "switch" is presented in Figure 3; the effects of gap duration and number of vocoder bands are not discussed. Mean deconvolved pupil size was statistically significantly larger in trials requiring mid-trial switches of attention than in trials where subjects maintained attention to the same talker throughout

the trial. Z-score normalized pupil size exhibits the same pattern of statistically significant difference between "maintain" and "switch" trials (i.e., a single cluster from point of divergence to end of trial; not shown).

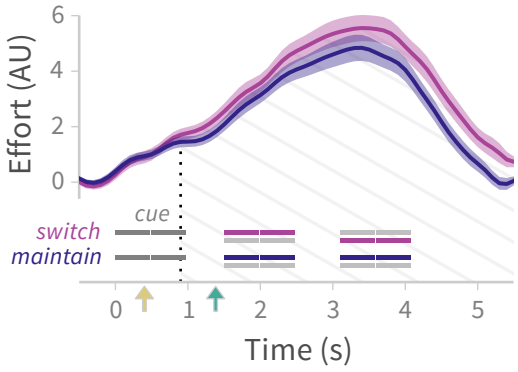

Figure 3: (Color online) Mean (± 1 standard error) deconvolved pupil size across subjects for "maintain" versus "switch" trials, with trial schematic showing to-be-attended streams (dark bars). Overall curve shapes and statistically significant differences between curves (hatched region) are similar to plots of z-score normalized pupil size (not shown). However, the divergence of the deconvolved signals aligns temporally to the end of the cue (dotted line); the right (green) arrow indicates divergence of the curves for normalized pupil size measurements, and the left (yellow) arrow indicates divergence for deconvolved signals using kernel parameters from Hoeks and Levelt.[6] AU = arbitrary units (deconvolution procedure yields "kernel weights" at each time point).

However, the divergence of the z-score normalized pupil size time series occurs around 1.3 s, indicated by the right (green) arrow in Figure 3, whereas the divergence of the deconvolved signals is temporally aligned with the offset of the "AA"/"AU" cue (cf. dotted vertical line in Figure 3). The left (yellow) arrow in Figure 3 indicates time of significant divergence if deconvolved using a kernel computed with the estimate of $t_{\max}$ from Hoeks and Levelt;[6] such early divergence indicates acausal behavior (different effort associated with different trial types occurs *before* listeners have heard the portion of the cue that differentiates "maintain" trials from "switch" trials). The temporal alignment of the trial type cue and the divergence of the pupil size time series using our estimate of $t_{\max}$ is consistent with the view that pupil dilation reflects cognitive load or attentional effort, and that effort/load increases *as soon as listeners know they are hearing a (more difficult) "switch" trial.*

V. CONCLUSION

Deconvolution of pupil size measurements allows insight into the unfolding of attentional effort over the course of an experimental trial, by temporally aligning the measured response with the stimulus events that induced it. However, pupil size is also affected by non-stimulus events; motor planning and execution associated with rapid button press responses are a particularly likely source of noise in the pupillometric signal in experimental settings. Nonetheless, careful attention to experimental design — combined with appropriate parameterization of the deconvolution kernel — preserves the ability to make inferences from the temporal relationship between stimulus events and (deconvolved) pupillary response.

ACKNOWLEDGMENTS

This research was supported by NIH grant R01-DC013260 to Adrian KC Lee. The authors are grateful to Zach Smith for spectral degradation code used in Experiment 2, and to Matt Winn for helpful suggestions on an earlier draft of this paper.

REFERENCES

[1] Hess EH & Polt JM (1964). "Pupil size in relation to mental activity during simple problem-solving." *Science* **143**(3611), 1190–1192. doi:10.1126/science.143.3611.1190.

[2] Kahneman D & Beatty J (1966). "Pupil diameter and load on memory." *Science* **154**(3756), 1583–1585. doi:10.1126/science.154.3756.1583.

[3] Kuchinsky SE, Ahlstrom JB, Vaden KI, Cute SL, Humes LE, Dubno JR, & Eckert MA (2013). "Pupil size varies with word listening and response selection difficulty in older adults with hearing loss." *Psychophysiology* **50**(1), 23–34. doi:10.1111/j.1469-8986.2012.01477.x.

[4] Koelewijn T, Shinn-Cunningham BG, Zekveld AA, & Kramer SE (2014). "The pupil response is sensitive to divided attention during speech processing." *Hearing Res.* **312**, 114–120. doi:10.1016/j.heares.2014.03.010.

[5] Winn MB, Edwards JR, & Litovsky RY (2015). "The impact of auditory spectral resolution on listening effort revealed by pupil dilation." *Ear Hear.* **36**(4), e153–e165. doi:10.1097/AUD.0000000000000145.

[6] Hoeks B & Levelt WJM (1993). "Pupillary dilation as a measure of attention: A quantitative system analysis." *Beh. Res. Meth. Ins. C.* **25**(1), 16–26. doi:10.3758/BF03204445.

[7] Wierda SM, van Rijn H, Taatgen NA, & Martens S (2012). "Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution." *P. Natl. Acad. Sci. USA* **109**(22), 8456–8460. doi:10.1073/pnas.1201858109.

[8] Hupé JM, Lamirel C, & Lorenceau J (2009). "Pupil dynamics during bistable motion perception." *J. Vision* **9**(7), paper 10. doi:10.1167/9.7.10.

[9] Cole RA, Muthusamy Y, & Fanty M (1990). "The ISOLET spoken letter database." Technical Report 90-004, Oregon Graduate Institute, Hillsboro, OR. Paper 205.

[10] Shannon RV, Zeng FG, Kamath V, Wygonski J, & Ekelid M (1995). "Speech recognition with primarily temporal cues." *Science* **270**(5234), 303–304. doi:10.1126/science.270.5234.303.

[11] Moore BCJ & Glasberg BR (1987). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns." *Hearing Res.* **28**(2-3), 209–225. doi:10.1016/0378-5955(87)90050-5.

[12] Larson ED & Engemann DA (2015). "pyeparse: 0.1.0." doi:10.5281/zenodo.14566.

[13] Maris E & Oostenveld R (2007). "Nonparametric statistical testing of EEG- and MEG-data." *J. Neurosci. Meth.* **164**(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024.

[14] Gramfort A, Luessi M, Larson ED, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, & Hämäläinen MS (2014). "MNE software for processing MEG and EEG data." *NeuroImage* **86**, 446–460. doi:10.1016/j.neuroimage.2013.10.027.

LIST OF FIGURES