Onset asynchrony influences audiovisual interaction in temporal modulation discrimination tasks

Master's Thesis

Presented to

The Faculty of Graduate School of Arts and Science
Brandeis University
Department of Psychology
Robert Sekuler, Advisor

In Partial Fulfillment
of the Requirements for the Degree

Master of Science
in
Psychology

by
Tianyou Zhou

February 2019

ABSTRACT

Onset asynchrony influences audiovisual interaction in temporal modulation discrimination tasks

A thesis presented to the Department of Psychology

Graduate School of Arts and Sciences
Brandeis University
Waltham, Massachusetts

By Tianyou Zhou

A target stimulus in one sensory modality influenced by an unattended stimulus in another sensory modality is called cross-modal interaction. Based on previous studies, temporal correlation and information from the unattended stimulus can cause cross-modal interaction. However, the attentional cue provided by the unattended stimuli can also be a way that unattended stimuli influence participants' performance in multisensory tasks. In this study audio-visual oscillated stimuli was used, and the congruency of the oscillations in audio-visual stimuli and the onset asynchrony of the two stimuli were manipulated. We used diffusion decision model (DDM) to analyze the result as a decision making process. I found no increase of accuracy in different onset asynchrony conditions, but the responses were faster when unattended stimuli have earlier onset than the target stimuli. The DDM analysis have shown that when the unattended stimuli present earlier than the target stimuli, the drift rate was faster but the boundary separation was also larger than other onset asynchrony conditions. The conclusion is that the unattended stimuli can act as an attentional cue and can provide information to influence our judgment on the target stimuli.

Content

1 Intruduction 1

2 Experiment 1 5

2.1 Participants 5

2.2 Stimuli 5

2.3 Procedures 6

2.4 Results 8

2.4.1 Descriptive statistics of accuracy and reaction times 9

2.4.1.1 The effect of congruency and onset asynchrony on accuracy 10

2.4.1.2 The effect of congruency and onset asynchrony on response time 12

2.4.1.3 Individual differences 13

2.4.1.4 Response bias 15

2.4.2 Diffusion decision model (DDM) analysis 16

2.4.2.1 Drift rate (v) 17

2.4.2.2 Boundary separation (a) 19

2.4.2.3 Non-decision time (t) 20

2.4.2.4 Response Bias (z) 21

3 Experiment 2 23

3.1 Participants 23

3.2 Stimuli and procedure 23

3.3 Result 24

4 Discussion 27

Tables and figures

Figures

Figure 1 6

Figure 2 9

Figure 3 11

Figure 4 14

Figure 5 15

Figure 6 16

Figure 7 18

Figure 8 25

Tables

Table 1 12

Table 2 12

Table 3 17

Table 4 19

Table 5 20

Table 6 21

Table 7 25

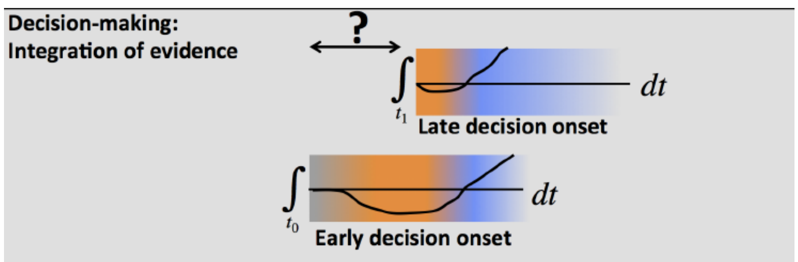
1. Introduction

Processing information from different sensory modalities can influence each other. Even when people intentionally direct their attention away from one of the sensory modalities, perception is inevitably influenced by information that should be ignored (Molholm, Martinez, Shpaner, & Foxe, 2007). In many previous studies this phenomenon occurred particularly when stimuli in different sensory modality have similar temporal properties (Spence, 2011). For example, a single visual flash of light present with multiple auditory beeps can be perceived as multiple flashes (Shams, Kamitani, & Shimojo, 2000), and the perception of a train of visual flashes can be altered by the auditory flutters presented at the same time (Shipley, 1964). This cross-modal influence can also affect the perception of continuous stimuli. When visual and auditory stimuli modulate continuously and are presented together, when the time of changes have high enough correlations, the stimuli in the unattended sensory modality will have influence on the perception of the target stimulus (Maddox, Atilgan, Bizley, & Lee, 2015; Sun, Hickey, Shinn-Cunningham, & Sekuler, 2017; Varghese et al., 2017).

From the above examples, we can see that one of the cause of cross-modal influence is the temporal correlation between stimuli. If the temporal distance between the stimuli is short enough(within the *temporal binding window*), the stimuli will be perceived as synchronized (Spence & Squire, 2003; Fujisaki & Nishida, 2005). The length of the *temporal binding window* can be adjusted depending on the task, which can be as short as 15 ms or as long as over 100 ms (Stone et al., 2001; Stevenson & Wallace, 2013). The stimuli presented in different sensory modalities, once they have been perceived as synchronized, can have cross-modal influence on each other, which not only change the perception on temporal properties, but also change the perception on stimuli properties that are unrelated to temporal properties (Maddox et al., 2015; Atilgan et al., 2018). It is also possible that without temporal correlation, the cross-modal influence can still happen. For example, in "sound-induced flash illusion" the perception of flash numbers can be altered by sound beep number, but the number of flashes can have little influence on the perception of sound beeps (Shams et al., 2000; Parker et al., 2017). This asymmetry in cross-modal influence have also been shown for continuous audiovisual stimuli (Varghese et al., 2017). The above phenomenon was explained with modality appropriateness hypothesis: temporal information is easier to process in auditory than in visual system (Welch & Warren, 1980; Wada, Kitagawa, & Noguchi, 2003), which made auditory stimuli easier to influence participants' temporal judgments. In Andersen, Tiippana, and Sams (2004) they found that four factors can contribute to the cross-modal influence in audio-visual tasks. Except for the modality appropriateness, people are also more likely to be influenced by stimuli as discrete rather than fused together, and they were more likely to be influenced by the more reliable stimuli. Following this logic, if the background stimuli have some properties that make them more relevant to the task, they will cause more cross-modal influence.

If we think about the process of doing a multisensory task as a decision making process, the influence from the unattended stimulus can be analyzed in a sequential sampling model frame (for a review, see (Ratcliff & Smith, 2004)). A widely used sequential sampling model is the diffusion decision model (Ratcliff, 1978; Ratcliff & McKoon, 2008; Ratcliff, Smith, Brown, & McKoon, 2016). In the diffusion decision model a simple choice decision process was modeled as an 'evidence accumulating' process. After a period of time for sensory encoding, the person will start to accumulate evidence and become more likely to make a certain choice. The person will stop the evidence accumulation and make a response at a certain point, which in the model is that the evidence accumulation reach a decision criterion that set beforehand.

For a multisensory task, such as the task with audio-visual stimuli like the one in Sun et al. (2017) and Varghese et al. (2017), the effect of cross-modal influence should mainly influence the evidence accumulation, because the perception on the incoming information was changed by the cross-modal influence from the unattended stimuli. However, in Varghese et al. (2017) the non-decision time was also changed when the unattended stimuli appear to be related to the task. When the stimuli in the decision task was simple enough, the non-decision time, which is modeled as sensory encoding and motor execution time, are usually fixed (Wagenmakers, Van Der Maas, & Grasman, 2007; Grasman, Wagenmakers, & Van Der Maas, 2009). However, when the stimuli was ambiguous, the participants might deliberately delay the onset of decision making process, thus add time on the non-decision time (Teichert, Grinband, & Ferrera, 2015). A diagram for the delay of decision making process was shown in Teichert, Ferrera, and Grinband (2014) figure 1, which I include it below:



In Teichert et al. (2014) this delay was assumed to be a top-down control of decision process. With longer time to prepare the decision making process, the participants can improve their accuracy in task performance. The diffusion decision model analysis shows that the delay of decision onset can increase the evidence accumulation speed in the decision process (Teichert et al., 2014). The effect of decision onset delay provide a confusion to the effect of cross-modal influence in the multisensory tasks: the mere attentional cue provided by the unattended stimuli might also play a role in multisensory tasks, and this effect can mix with the cross-modal influence on the perception of target stimuli.

In this study we used a simplified version of the audio-visual oscillation judgment task in Sun et al. (2017) and Varghese et al. (2017). In order to separate attentional cuing and the cross-modal perceptual influence in this task, I varied both the oscillation congruency (the audio-visual oscillations are in the same or different frequency and phase) between the audio-visual stimuli and the onset time of the two stimuli (the onset of the unattended stimulus is earlier or later than the target stimulus). When the unattended stimulus onset is earlier than the target stimulus, the task performer should have both the attentional cuing benefit and the extra information from the unattended stimulus. However, the extra information from the unattended stimulus can have different effect in different temporal correlation conditions: when the oscillation between the audio-visual stimuli are incongruent, the early onset of the unattended stimulus can provide the same attentional cuing but the extra information from it will be disrupting for the perception of the target. Thus, we can expect an interaction of task performance between onset and congruency conditions: in congruent condition the early onset of the unattended stimulus can provide enhancement in task performance with both attentional cuing and the oscillation information, but in the incongruent condition, while the early onset of the unattended stimulus still provide attentional cuing, the information it provide will be contrary to the target stimulus, which compromised the enhancement effect of the attentional cuing. The influence of onset asynchrony can be more clear when we put the result in a diffusion decision model: during the decision process the stimuli should be the same for different onset asynchrony, so only the attentional cuing effect of the early present unattended stimuli can influence the drift rate, and the influence of the information from early presented stimuli should have influence on response bias, which reflected the influence of the information the task performer receives before decision making onset.

2. Experiment 1

2.1 Participants

27 participants (five males) from Brandeis University took part in the experiment. The participants are undergraduate or graduate students from 19 to 32 years old (mean = 21.64, standard deviation = 2.83). All the participants are healthy. Before the experiment every participants have their visual acuity measured using ETDRS test (For a review of ETDRS visual acuity test, see(Rosser, Laidlaw, & Murdoch, 2001)) at 60 cm distance. The result shows that all the participants have normal or corrected to normal visual acuity (less than 0.2 on log MAR scale). The participants received ten dollars as compensation after the experiment.

2.2 Stimuli

The experiment had visual and auditory stimuli. The visual stimulus was a moving white disc (60.78 cd/m^2) on gray background (5.94 cd/m^2), presented on a 21.5 inch iMac screen. The screen resolution was 1920 × 1080 pixels, and viewing distance was 57cm. On each trial, the white disc appeared at one side (left or right) of the screen and moved to the other side of the screen in a constant speed ($72°/s$) until it disappeared at the other side two seconds after. While the disc was moving, its radius varied sinusoidally between 2° and 3° visual angle. At the onset of the disc's oscillation, its radius was always 2° visual angle. In each trial. the disc's radius' modulation rate was either 6 Hz or 7.5 Hz.

The auditory presentation was a two seconds' sound stream from a BOKA 81000 computer speaker. The components of the sound was a harmonic series from 220 Hz to 1320 Hz, and the energy of each component from 220 Hz each decreased by a factor of two. The sampling rate was 44.1 kHz. The average sound pressure level was 60 dB/SPL at the location of participants' head, which is approximately equal to the conversation sound level at one meter distance (*Noise Chart*, n.d.). An amplitude modulation with 100% modulation depth had been superimposed on the sound wave, which created 6 Hz and 7.5 Hz amplitude modulation, which are shown in 1. During the experiment the sound was modulated or not modulated (control condition), and the duration of the sound presentation vary depending on the conditions (see **Procedure** and Figure 1).

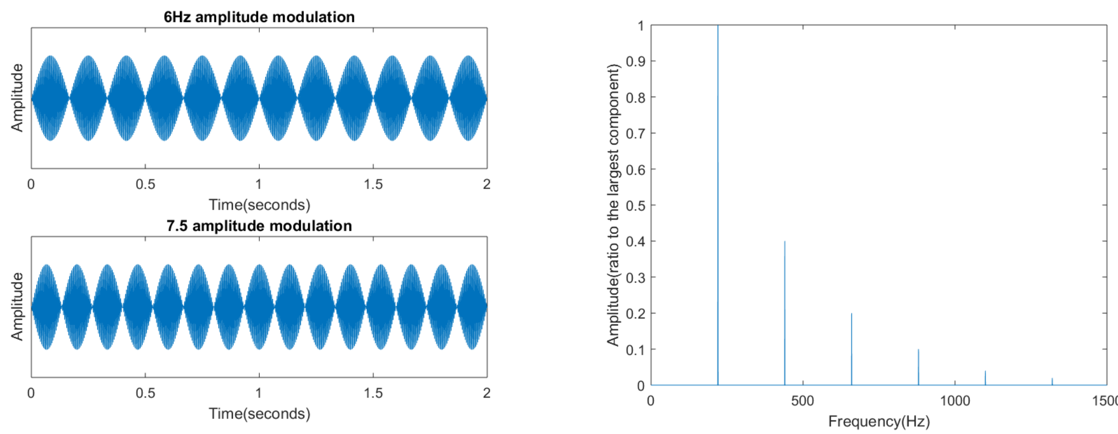


Figure 1: Left: two examples of modulated sound waves, which have 6 Hz (upper figure) and 7.5 Hz (lower figure) amplitude envelop with 100% modulation depth. Right: the frequency spectrum of the sound wave, which has a mix of 1st to 6th harmonics with of 220Hz fundamental.

2.3 Procedures

The experiment was programed in Matlab (version 2015a) and was presented using an 21.5 inches iMac under OSX Yosemite (version 10.10.5). The testing room was sound attenuated (The attenuation of sound pressure at the room's door is 17 dB/SPL), with a constant dim light on. The participants were asked to put their heads on a chin rest when they were performing the tasks. The chin rest was 57 cm away from the screen. For every trial during the experiment, a black fixation cross was presented at the center of the screen for 500 ms. After the fixation disappear, a blank waiting screen was presented for a period of time vary between 300 ms and 800 ms. Then the stimuli presented for 2000 ms (or +267/333 ms when auditory stimulus present before visual stimulus), while the participants perform a frequency judgment task on the visual stimuli. The judgments were made by key pressing. After that a 500 ms blank screen was presented as inter-trial waiting time, and a feedback beep sound was presented to the participants during that period of time (high pitch beep indicated correct responses and low pitch beep indicated error responses or no response within 2000 ms of stimuli presentation time).

For every participant, the experiment started with two practice blocks. The first block consisted of 40 trials of visual only tasks. In each trial participants only saw the white disc move on the gray background but did not hear the auditory stimuli. The participants' tasks were to judge whether the circle change size at a fast rate (7.5 Hz) or slow rate (6 Hz), by pressing 'Q' (for fast rate) or 'W' (for slow rate) on the keyboard. The second block of practice were 40 trials of auditory only judgment tasks. In this block only sound will be presented for two seconds, with either 6 Hz or 7.5 Hz modulation. During sound presentation only blank gray background was presented on the screen. The participants was asked to make response on the sound oscillation rate by pressing 'Q' (for fast) or 'W' (for slow) on the keyboard. During the practice the experimenter stayed in the room to ensure that the participant was able to perform the task as instructed.

After the practice, participants were tested on four blocks of experiment tasks, with 180 trials in each block (a total of 720 trials, last for an hour). Unlike the practice trials, the experiment trials presented both auditory and visual stimuli in every trials. However, the participants' tasks were only to judge the frequency of the visual presentations, and the response keys were the same as in practice trials ('Q' for fast size changing and 'W' for slow size changing). The visual stimuli, which was what the participants were required to make judgments on, were the same as the first practice block, which were a circle move from one side of the screen to the other side while change size sinusoidally at 6 Hz or 7.5 Hz frequency. The sound stimuli were vary on two variables in different trials. The first variable was the amplitude modulation frequency. Except for the 6Hz and 7.5Hz of amplitude modulation, the sound stimuli could also have no modulation in one third of the trials, which served as a control condition. These three kinds of sound were coupled with the same amount of fast and slow size changing disc, which created conditions of audio-visual congruent (modulations were in the same phases), audio-visual incongruent (modulations had different frequencies, thus were in different phases) and control conditions (discs' size change in sinusoidal way, but sound did not change in amplitude). The other variable is the onset asynchrony between auditory and visual presentation. In one-third of the trials the sound were presented two cycles earlier than the circle on the screen, and in another one third of the trials the sound presented two cycles later than the circle, and in the last one-third of the trials the sound and circle presented at the same time. When the two stimuli did now present simultaneously, the differences in presentation onset time was based on the frequency of the stimuli that presented first. Thus, if the sound was presented before the visual stimuli, the onset asynchrony would be two cycles of sound amplitude modulation, which were 333 ms when the sound was modulated at 6Hz and 267 ms when the sound was modulated at 7.5Hz; if the sound was presented after the visual stimuli, the asynchrony time would be the time of two cycles of the circle's size change. When the trial was in control condition and the sound did not modulate, the asynchrony time were two cycles of disc size change when the audio-visual stimuli onset was not simultaneous. The procedure of one experiment trial has been shown in Figure 2.

2.4 Results

The experiment data were collected in Matlab (2015a) and analyzed on RStudio (version 1.1.423) with R 3.4.3. I took out the trials with no response, and trials with response time shorter than 250 ms or longer than 2000 ms. After the filtering, 6.4% of the trials had been excluded.

The experiment was a 3×3 design, with three levels of audio-visual congruency conditions (congruent, incongruent and control) and three levels of onset asynchrony conditions (auditory stimuli before visual stimuli, simultaneous, auditory stimuli after visual stimuli). The dependent variables were response accuracy and response time for each trial. The median response time for correct trials and accuracy of response have been calculated in each one of the nine conditions for each participant, then fit into the linear models using 'lmer' function in R package 'lme4'. The effect of each factors on dependent variables were firstly analyzed

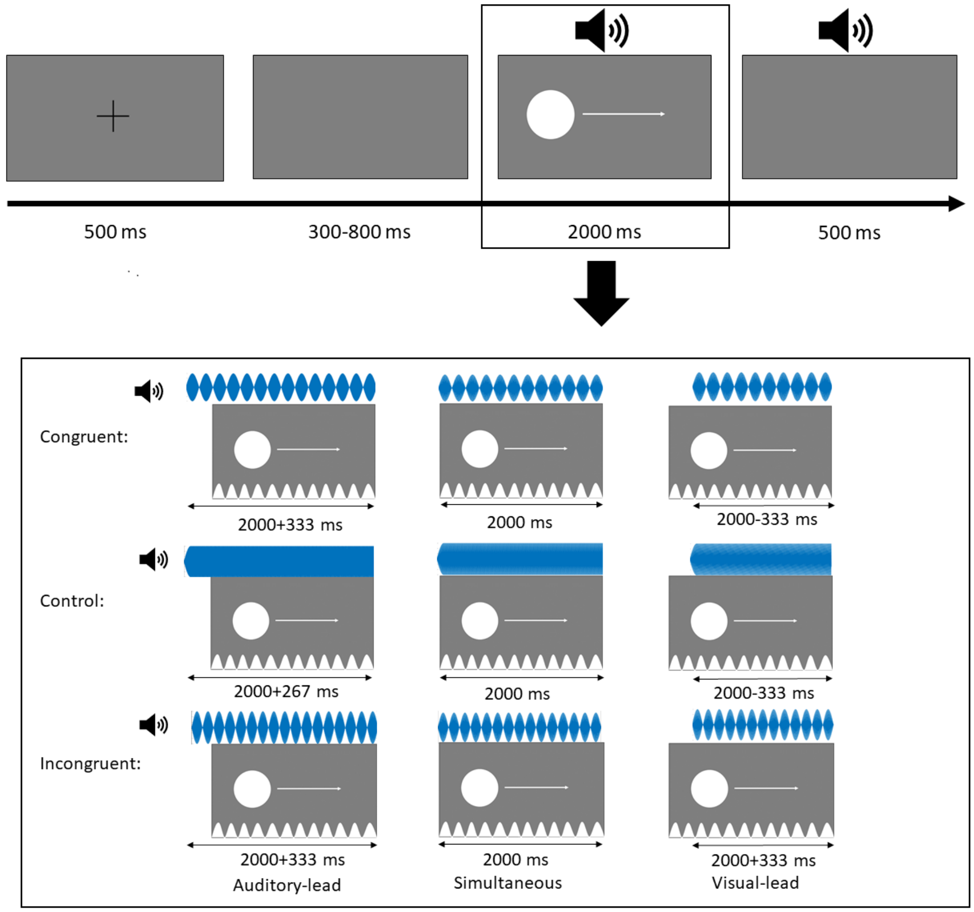


Figure 2: The figure shows an example of the experiment trials. The stimuli presented in each trial during the experiment was visual accompanying sound. The visual stimulus in each trial was a disc change size in sinusoidal way while moving horizontally through the middle of the screen. The examples we are showing here are only discs change size in 6Hz frequency. The sound amplitude modulation rate could be the same or different from the circle, or had no modulation in the control condition. The onset of two stimuli could also be simultaneous or different. The asynchrony time was always two cycles of modulation of the stimuli presented first. In the control condition, the asynchrony was always two cycles of the circle's size change. The disc size and sound amplitude are always smallest at the onset of presentation

by hierarchical regression analysis, then the model best explained the result will be used for ANOVA and pairwise comparison test.

2.4.1 Descriptive statistics of accuracy and reaction times

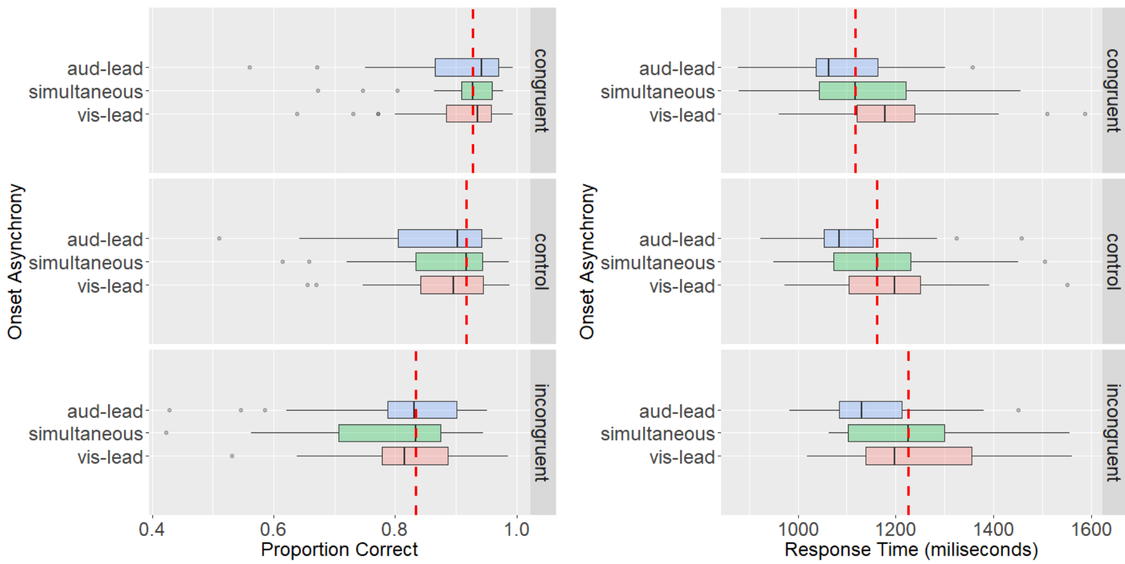
The accuracy was defined by proportion of correct trials. The overall accuracy for all the subjects was 0.86. When we separately calculate the percent of correct for each subject in each conditions, the proportion of correct was different across three levels of congruency conditions, but generally the same across three levels of stimuli onset asynchrony. When the auditory and visual stimuli had the same modulation frequency, thus was congruent in phase, the accuracy is the highest (M = 0.90, SD = 0.09), which is just slightly higher than the control condition (M = 0.87, SD = 0.10). When auditory and visual stimuli had different modulation frequency, the accuracy was the lowest (M = 0.80, SD = 0.12). The accuracy data are shown in Figure 3a.

We calculated the median of each participant's response time in correct trials in each condition. The median response time was different across the three levels of congruency conditions and three levels of onset asynchrony. When the auditory and visual stimuli had the same modulation rate, the response time was the shortest (M = 1143.07 ms, SD = 141.94 ms), and was highest when the modulation rate of auditory and visual stimuli were different (M = 1207.91 ms, SD = 138.73 ms). When auditory stimuli onset lead the visual stimuli by two cycles, the response time was shorter (M = 1116.05 ms, SD = 113.13 ms) compare to when the auditory and visual stimuli onsets are simultaneous (M = 1183.78 ms, SD = 144.29 ms) and when the auditory onset was two cycles behind the visual stimuli onset (M = 1209.98 ms, SD = 140.15 ms). The response time data are shown in Figure 3b.

2.4.1.1 The effect of congruency and onset asynchrony on accuracy

The first analysis on accuracy was a hierarchical regression analysis. The results of hierarchical regression analysis is shown in Table 1. The result have shown that phase congruency between auditory and visual stimuli had significant effect on accuracy, while the onset asynchrony did not have significant effect. However, adding the interaction between congruency and onset asynchrony make the model fit marginally better than the additive model ($p = .06$) suggest that the effect of onset asynchrony might have different effects on accuracy in different congruency levels. Thus, we chose the interaction model for further analysis.

The amount of variances in accuracy explained by each factors have been parsed out in an omnibus ANOVA on the interaction model. The analysis shows that the interaction between congruency and onset asynchrony is marginal, $F(4, 104) = 2.18$, $p = .07$, $\eta_p^2 = .04$. The main



(a) (b)

Figure 3: Figures for proportion of correct and median response time in correct trials for all the participants, which shows the effect of onset asynchrony in each congruency condition. The onset asynchrony had three levels: auditory-lead, refers to the condition when the auditory stimuli presented two cycles earlier than the visual stimuli. Simultaneous, refers to the condition when the auditory and visual stimuli onset at the same time. Visual-lead, refers to the condition when the visual stimuli presented two cycles earlier than the auditory stimuli. The congruency had three conditions: congruent, the visual and auditory stimuli had the same frequency and were in phase with each other. Control, the auditory stimuli did not oscillate while the visual stimuli oscillate. Incongruent, the auditory stimuli and the visual stimuli oscillate in different frequency. The red dashed lines shows the median percent correct/response time when the auditory and visual stimuli onsets are simultaneous. (a) Figure for percent of correct. (b) Figure for response time of correct trials.

effect of congruency is significant, $F(2,52) = 95.65$, $p < .001$, $\eta_p^2 = .79$, and the main effect of onset asynchrony is not significant, $F(2,52) = 1.79$, $p = .18$, $\eta_p^2 = .06$.

We did a pairwise comparison with the interaction model using the 'lsmeans' function in R (Lenth, 2016). To correct the multiple comparison error, the results were adjusted by Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). Different congruency conditions in all the onset asynchrony conditions are significantly different, and the difference in proportion of correct responses shows that congruent conditions always have higher proportion of correct responses than control conditions, and both congruent and control conditions are always higher in percent of correct than incongruent conditions. The differences among

Table 1: Model specification for proportion of correct

Model Specification	DF[a]	AIC[b]	χ^{2c}	ΔDF[d]	p value[e]
$Pc \sim (1 \mid Subjects)$	3	-549.65			
$Pc \sim Congruency + (1 \mid Subjects)$	5	-680.66	135.02	2	<.001
$Pc \sim Congruency + Onset + (1 \mid Subjects)$	7	-680.21	3.54	2	.17
$Pc \sim Congruency + Onset + Congruency : Onset + (1 \mid Subjects)$	11	-681.06	8.86	4	.06

[a] Degrees of freedom
[b] Akaike Information Criterion
[c] χ^2 value for model comparison between the current model and the model above it
[d] Change in degree of freedom from the above model to current model
[e] Probability of type I error for model comparison

levels of onset asynchrony were not significant except for one comparison under the incongruent condition, where the simultaneous condition had higher proportion of correct than when the visual stimuli appear earlier than auditory stimuli, $t(208) = 2.69$, $p = .02$.

2.4.1.2 The effect of congruency and onset asynchrony on response time

The effect of congruency, onset asynchrony and their interaction on response time also were examined by model comparison. For response time we used the median of correct response time for each subject in each conditions. The results of model comparison are presented in Table 2. The model comparison shows that both congruency and onset asynchrony improve the model significantly, but the interaction did not improve the model fit ($p = .11$). Thus, we are going to use the additive model (without interaction) for further analysis.

Table 2: Model specification for response time

Model Specification	DF[a]	AIC[b]	χ^{2c}	ΔDF[d]	p value[e]
$RT \sim (1 \mid Subjects)$	3	-513.21			
$RT \sim Congruency + (1 \mid Subjects)$	5	-552.00	42.79	2	<.001
$RT \sim Congruency + Onset + (1 \mid Subjects)$	7	-677.18	129.18	2	<.001
$RT \sim Congruency + Onset + Congruency : Onset + (1 \mid Subjects)$	11	-676.62	7.44	4	.11

[a] Degrees of freedom
[b] Akaike Information Criterion
[c] χ^2 value for model comparison between the current model and the model above it
[d] Change in degree of freedom from the above model to current model
[e] Probability of type I error for model comparison

The amount of variances in accuracy explained by each factors have been parsed out in an omnibus ANOVA on the additive model. The analysis shows that both the congruency and onset asynchrony have significant influence on response time. The effect of congruency yielded an $F(2, 52) = 42.24$, $p < .001$, $\eta_p^2 = 0.62$. The effect of onset asynchrony yielded an $F(2, 52) = 86.77$, $p < .001$, $\eta_p^2 = .77$.

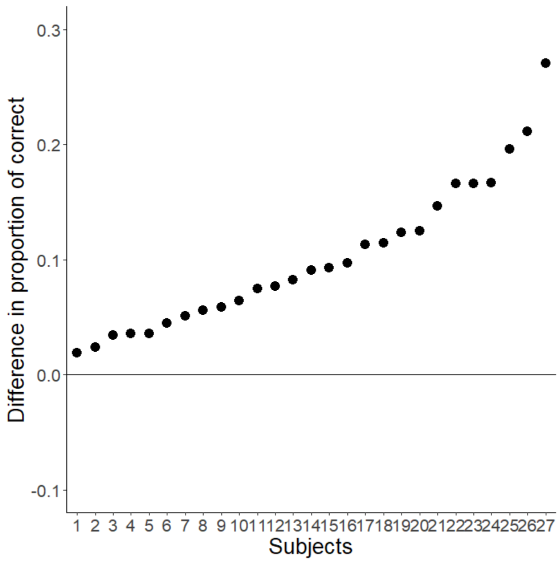
I did a pairwise comparison on the additive model for response time. The results shown that differences exist in every level of both of the factors. Participants made responses faster in congruent trials than in control trials ($t(208) = 2.14$, $p = .03$), and in incongruent trials the responses are slower than both control ($t(208) = 6.67$, $p < .001$) and congruent trials ($t(208) = 8.81$, $p < .001$). The responses were slower in the visual-lead condition than the simultaneous conditions ($t(208) = 3.56$, $p < .001$). The response in auditory-lead condition was faster than in both the simultaneous condition ($t(208) = 9.20$, $p < .001$) and the visual-lead condition ($t(208) = 12.76$, $p < .001$).

A brief conclusion from the accuracy and response time analysis is that the congruency between audio-visual stimuli have changed both participants' response speed and accuracy, but the onset asynchrony only affected participants' response speed and had no effect on accuracy. This result seems contradictory to the hypothesis that based on Teichert et al. (2014). A possibility for this result is that the overall accuracy was very high and the individual difference was large enough to cover the small difference between onset asynchrony conditions. Another explanation is that the participants might have chosen to use the advantage afforded by the auditory-lead condition to make faster responses, rather than make the responses more accurate. The next analysis will examine individual differences in this experiment.

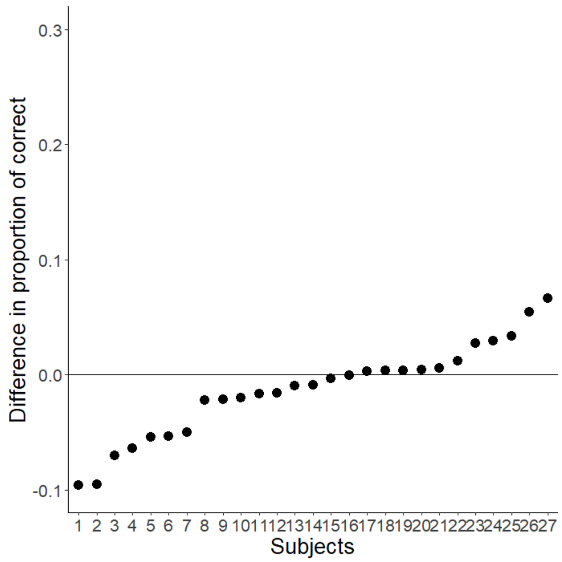
2.4.1.3 Individual differences

Firstly we want to address the question: did the conditions have similar influence on all the participants' performance? To answer this question, I calculated the difference of each participant's performance in different congruency conditions. After sorted by performance, across congruency conditions the difference in accuracy was different for different participants. It appears that some participants were not influenced by the congruency conditions and some of them have been strongly influenced in one direction, which is their performance is better in congruent than incongruent condition.

Individual difference in accuracy was also observed across conditions. However, the difference in each conditions was much smaller, and participants are influenced in different directions.



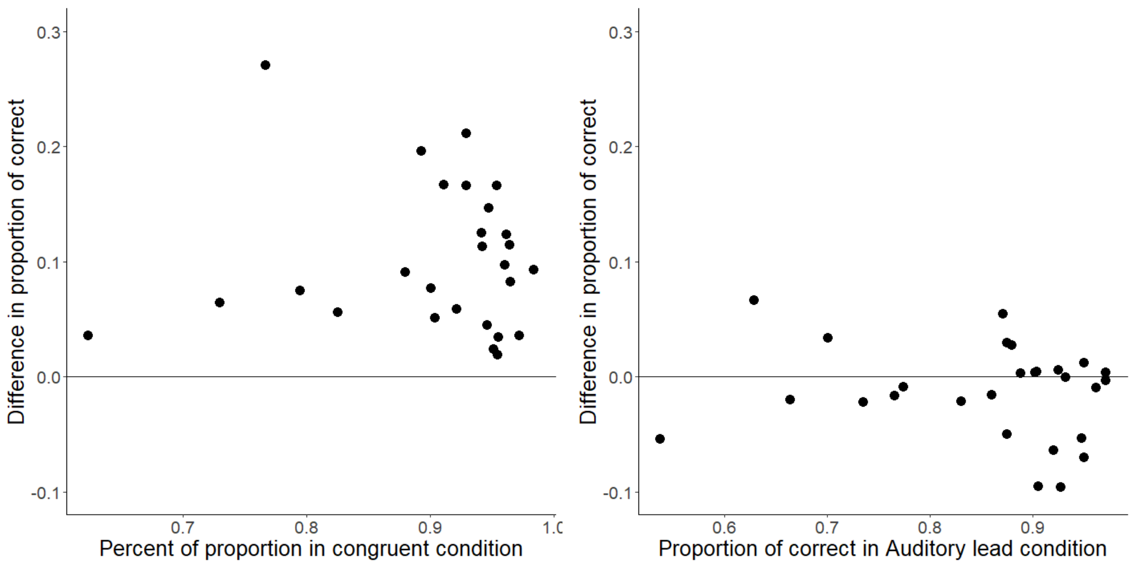
(a)



(b)

Figure 4: Individual differences in response accuracy. *Figure (a)*: the difference in proportion of correct between audio-visual congruent condition and incongruent condition for each participants. The x-axis scale is subjects sorted by their difference in proportion of correct trials between congruent and incongruent conditions. *Figure (b)*: the difference in proportion of correct between auditory-lead condition and visual-lead conditions for each participants. The x-axis scale is subjects sorted by their difference in proportion of correct trials between auditory-lead and visual-lead conditions.

The next analysis looked at if the difference of condition influence correlated with participants' performances in single conditions. It is possible that participants were less influenced by conditions if they were very good or very poor performers for the task, and the ones that perform fairly would be influenced most by the conditions. We did Pearson correlation analysis between participants' accuracy in different conditions and difference in accuracy between conditions. The result shown no correlation between performance in any of the conditions and the difference between conditions (Congruency conditions: congruent vs. difference: $r = -.0001$, incongruent vs. difference:$r = -.27$; onset asynchrony conditions: auditory-lead vs. difference: $r = -.16$, visual-lead vs. difference: $r = -.18$). It shows that the amount of influence of conditions on each individual was not related to individuals' ability in performing the tasks accurately.



(a) (b)

Figure 5: Figures as examples of no correlation between accuracy in each conditions and the difference of accuracy between conditions. Each point represent an individual participant. *Figure (a)*: the relationship between the accuracy in congruent condition and the accuracy difference between congruent and incongruent conditions. *Figure (b)*: the relationship between the accuracy in auditory-lead condition and difference between auditory-lead and visual-lead conditions.

2.4.1.4 Response bias

It is possible that participants have a bias towards making 'fast' or 'slow' response. We have used an equivalence test (Lakens, 2017) to look at the difference in error rate between fast and slow trials. If the difference in error rate was significantly different from zero, it indicates a bias in making a fast or slow response. We have separated the error rate in fast and slow trials (when the visual stimuli oscillated in fast or slow rate) for each participants by three congruency and three onset asynchrony conditions, and ran two one-sided test (TOST) for each conditions using R package TOSTer (version 0.3.4) (Lakens, 2017). The result shows that the equivalence was not significant for any conditions ($p > 0.9$), which means the participants are biased in their responses. The Figure 6 shows that participants are biased to make fast responses.

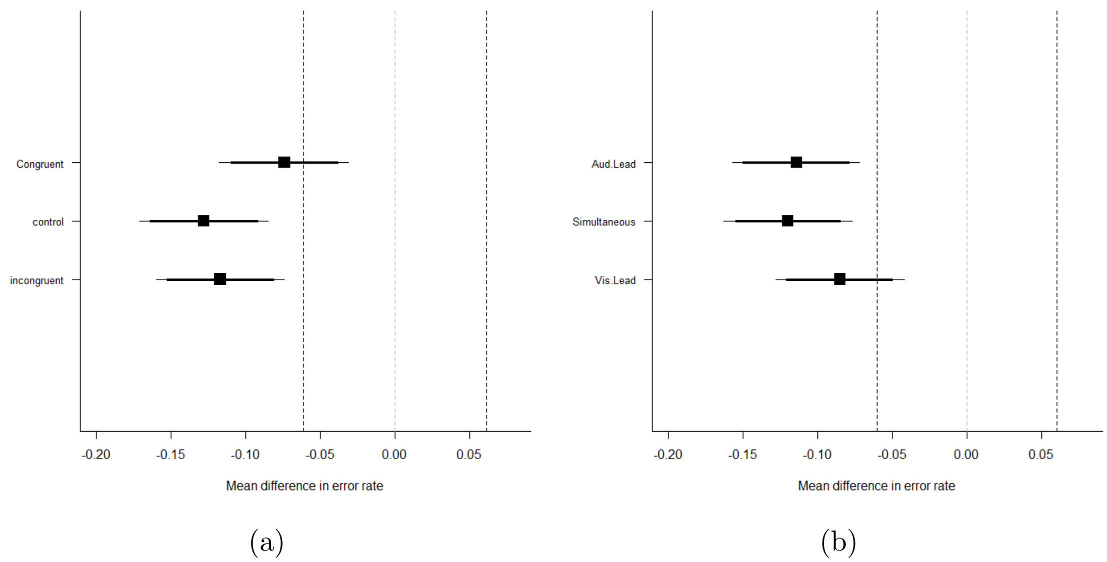


Figure 6: The results of equivalence test. The black vertical dashed lines defines 95% confidence interval for the zero point (within the confidence interval indicates equivalent). The horizontal segment for each point represent the 95% confidence interval for error rate difference between fast and slow trials in each conditions. From the figure we can see that in all the conditions the difference in error rate are deviated from the mean in the negative direction, indicating participants make less errors in fast trials, leading to the conclusion that the participants have a bias toward responding 'fast'. *Figure (a)*: the error rate differences in different congruency conditions. *Figure (b)*: the error rate differences in different onset asynchrony conditions.

2.4.2 Diffusion decision model (DDM) analysis

In order to better understand the effect of congruency and onset asynchrony on participants' performance, we need to take speed-accuracy trade-off into consideration, and for this purpose the next analysis is fitting the response time and accuracy into a diffusion decision model (DDM) (Ratcliff, 1978; Ratcliff & McKoon, 2008). DDM describes the change of the likelihood that a participant will choose one of the two choices in a two choices task as an information accumulation procedure. Thus, DDM evaluate several parameters to describe the decision making process. The main parameters are (1) drift rate, describes the influence of information accumulation on the speed of changing decision making likelihood; (2)

boundary separation, the distance between two 'decision boundaries' (i.e. the threshold of making a choice); (3) non-decision time, the time that not involved in decision making, such as motor control and preparation time; (4) starting bias, the start point of decision making, which is usually at the middle point between two decision boundaries but some times can be biased by experiment settings. Other parameters are within or between trial variances of the above parameters.

For this analysis we used the python package of HDDM (hierarchical drift diffusion model), version 0.6.1 (Wiecki, Sofer, & Frank, 2013). HDDM estimates parameters with hierarchical Bayesian model, and updates the Bayesian model using Markov Chain Monte Carlo (MCMC) sampling under PyMC (version 2.3.6). This analysis have chosen four parameters for model estimation, drift rate (v), boundary separation (a), non-decision time (t) and response bias (z). Model parameters were simulated via MCMC with 5000 iterations, and the first 800 iterations were thrown out for better sampling.

The parameters estimated in HDDM was separated by individual participants in different conditions. The boxplot for the four parameters are shown in Figure 7.

The four parameters in different congruency and onset asynchrony conditions are compared in order to find out the effect of conditions on the perceptual-decision process.

2.4.2.1 Drift rate (v)

The effect of congruency and onset asynchrony on drift rate was first analyzed with hierarchical regression analysis. The results are in Table 3.

Table 3: Model specification for drift rate

Model Specification	DF[a]	AIC[b]	χ^{2c}	ΔDF[d]	p value[e]
$Driftrate \sim (1 \mid Subjects)$	3	162.42			
$Driftrate \sim Congruency + (1 \mid Subjects)$	5	32.6	133.82	2	<.001
$Driftrate \sim Congruency + Onset + (1 \mid Subjects)$	7	-70.03	106.63	2	<.001
$Driftrate \sim Congruency + Onset + Congruency : Onset + (1 \mid Subjects)$	11	-117.64	55.61	4	<.001

[a] Degrees of freedom
[b] Akaike Information Criterion
[c] χ^2 value for model comparison between the current model and the model above it
[d] Change in degree of freedom from the above model to current model
[e] Probability of type I error for model comparison

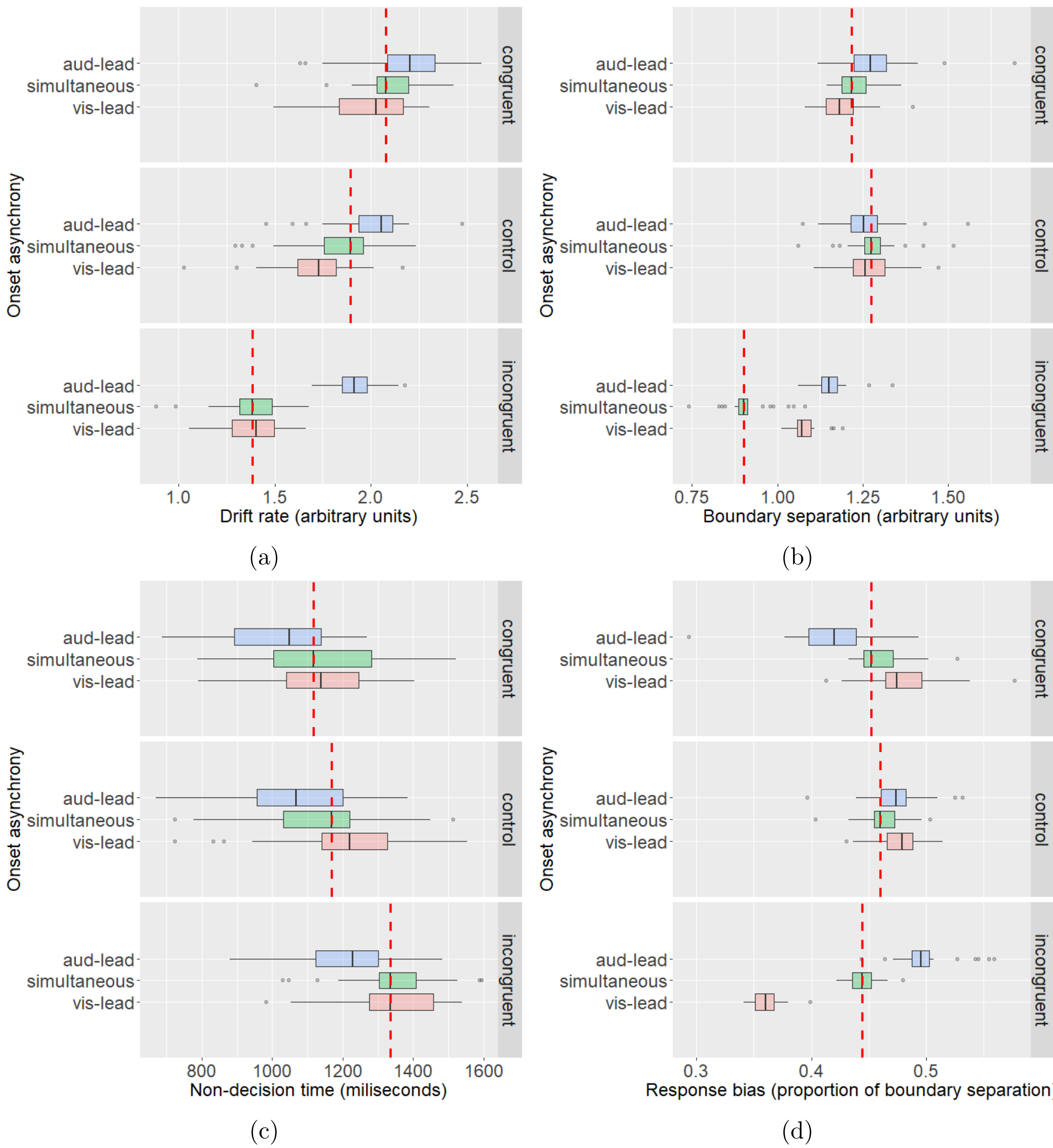


Figure 7: The estimated HDDM parameters in different conditions. The red dashed line are the median for the simultaneous condition in each level congruency condition. *Figure (a)* The drift rate. *Figure (b)* The boundary separation. *Figure (c)* The non-decision time. *Figure (d)* The response bias. The response bias was the proportion of boundary separation from 0 to 1. Boundaries was coded by response type, slow=0, fast=1.

The model comparison shows that both main effects and the interaction between congruency and onset asynchrony had significant effect on drift rate (v). The effect of interaction was significant, $F(4, 208) = 15.27$, $p < .001$, $\eta_p^2 = .23$. The effect of congruency was significant, $F(2, 52) = 189.13$, $p < .001$, $\eta_p^2 = .88$. The effect of onset asynchrony was significant, $F(2, 52) = 85.87$, $p < .001$, $\eta_p^2 = .77$.

The pairwise comparisons shows that in all the onset asynchrony levels, the effect of congruency were always significant, and in congruent condition the v was always faster than control condition ($p < .001$) and incongruent condition ($p < .001$), and the control condition always faster than incongruent condition ($p < .001$) except for in the auditory-lead condition ($p = .09$). For the effect of onset asynchrony, the auditory-lead condition had significant faster v than the visual-lead and simultaneous condition in the control and incongruent conditions ($p < .001$). In the congruent condition, v of auditory-lead condition was marginally faster than the simultaneous condition ($p = .07$) and faster than the incongruent condition ($p = .001$). The simultaneous condition had faster v than visual-lead condition in the control condition ($p = .02$), and the difference was marginal in the congruent condition ($p = .07$). When the auditory and visual oscillation were incongruent, the v in simultaneous and visual-lead conditions had no difference ($p = .77$).

2.4.2.2 Boundary separation (a)

The results of the hierarchical regression analysis on boundary separation is in Table 4.

Table 4: Model specification for boundary separation

Model Specification	DF[a]	AIC[b]	χ^{2c}	ΔDF[d]	p value[e]
$Boundary_sep \sim (1 \mid Subjects)$	3	-267.86			
$Boundary_sep \sim Congruency + (1 \mid Subjects)$	5	-429.86	166	2	<.001
$Boundary_sep \sim Congruency + Onset + (1 \mid Subjects)$	7	-469	43.14	2	<.001
$Boundary_sep \sim Congruency + Onset + Congruency : Onset + (1 \mid Subjects)$	11	-554.97	93.96	4	<.001

[a] Degrees of freedom
[b] Akaike Information Criterion
[c] χ^2 value for model comparison between the current model and the model above it
[d] Change in degree of freedom from the above model to current model
[e] Probability of type I error for model comparison

The model comparison have shown that both of the main effects and the interaction have had significant influence on boundary separation. The effect of interaction was $F(4, 208) =$

28.34, $p < .001$, $\eta_p^2 = .35$. The effect of congruency was $F(2, 52) = 213.39$, $p < .001$, $\eta_p^2 = .89$. The effect of onset asynchrony was $F(2, 52) = 35.37$, $p < .001$, $\eta_p^2 = .58$.

The pairwise comparison result indicates different effect of onset asynchrony on boundary separation (a) in different congruency conditions. In the congruent condition, the auditory-lead condition had significant larger a than simultaneous condition ($p < .01$), and simultaneous condition had marginally significant larger a than the visual-lead condition ($p = .07$). In the control condition, the three onset asynchrony conditions had similar a. In the incongruent condition, the auditory-lead condition still had larger boundary separation than both simultaneous ($p < .001$) and visual-lead conditions ($p < .001$), but the simultaneous condition have smaller a than the visual-lead condition ($p < .001$). The effect of congruency on a was generally consistent across different onset asynchrony conditions. The incongruent condition had smaller a than congruent and control conditions in all three onset asynchrony conditions ($p < .001$). The congruent and control condition had similar a in auditory-lead condition ($p = .11$), and in simultaneous and visual-lead conditions the control condition had larger a than the congruent condition ($p < .01$).

2.4.2.3 Non-decision time (t)

Non-decision time was assumed to include sensory encoding and motor execution time (Ratcliff & McKoon, 2008). In this experiment the variation in non-decision time should mainly come from differences in sensory encoding process in different conditions. The hierarchical analysis on non-decision time are shown in Table 5.

Table 5: Model specification for non-decision time

Model Specification	DF[a]	AIC[b]	χ^{2c}	ΔDF[d]	p value[e]
$Non_D T \sim (1 \mid Subjects)$	3	-122.43			
$Non_D T \sim Congruency + (1 \mid Subjects)$	5	-186.68	68.25	2	<.001
$Non_D T \sim Congruency + Onset + (1 \mid Subjects)$	7	-223.05	40.38	2	<.001
$Non_D T \sim Congruency + Onset + Congruency : Onset + (1 \mid Subjects)$	11	-218.19	3.14	4	.54

[a] Degrees of freedom
[b] Akaike Information Criterion
[c] χ^2 value for model comparison between the current model and the model above it
[d] Change in degree of freedom from the above model to current model
[e] Probability of type I error for model comparison

Model comparison shows that the main effects of congruency and onset asynchrony had significant effect on non-decision time (t), but the interaction did not have significant effect. The effect of congruency was $F(2, 52) = 47.49$, $p < .001$, $\eta_p^2 = .65$. The effect of onset asynchrony was $F(2, 52) = 21.79$, $p < .001$, $\eta_p^2 = .46$.

The pairwise comparison shows that the incongruent condition had significantly longer t than the congruent and control conditions ($p < .001$), and the congruent and control conditions had similar t. The auditory-lead condition had shorter t than the simultaneous and the visual lead conditions in congruent and incongruent conditions ($p < .01$). In the control condition the auditory-leading condition had shorter t than the visual-lead condition ($p = .001$) and marginally shorter t than the simultaneous condition ($p = .08$). The simultaneous had marginally shorter t than visual-lead condition in the control condition ($p = .08$). In other conditions the simultaneous and visual-lead conditions had similar t.

2.4.2.4 Response bias (z)

The effects of congruency and onset asynchrony on response bias were also analyzed by hierarchical regression analysis and the results are in Table 6.

Table 6: Model specification for response bias

Model Specification	DF[a]	AIC[b]	χ^{2c}	ΔDF[d]	p value[e]
$Bias \sim (1 \mid Subjects)$	3	-800.17			
$Bias \sim Congruency + (1 \mid Subjects)$	5	-822.07	25.90	2	<.001
$Bias \sim Congruency + Onset + (1 \mid Subjects)$	7	-831.82	13.75	2	.001
$Bias \sim Congruency + Onset + Congruency : Onset + (1 \mid Subjects)$	11	-1093.79	269.98	4	<.001

[a] Degrees of freedom
[b] Akaike Information Criterion
[c] χ^2 value for model comparison between the current model and the model above it
[d] Change in degree of freedom from the above model to current model
[e] Probability of type I error for model comparison

From model comparison we can see that all the main effects and the interaction had significant effect on response bias (z). The effect of interaction was $F(2, 208) = 120.65$, $p < .001$, $\eta_p^2 = .70$. The effect of congruency was $F(2, 52) = 42.83$, $p < .001$, $\eta_p^2 = .62$. The effect of onset asynchrony was $F(2, 52) = 20.95$, $p < .001$, $\eta_p^2 = .45$.

The pairwise comparison shows that the effect of congruency was different across different onset asynchrony conditions. In auditory-lead condition, the congruent condition had smaller z than the control condition ($p < .001$), and the control condition had smaller z than the incongruent condition ($p < .001$). In the simultaneous and the visual-lead conditions, the congruent and control conditions had similar z, and the incongruent condition had smaller z than both the congruent and control conditions ($p < .05$). The effect of onset asynchrony was also different across different congruency conditions. In the congruent condition, the auditory-lead condition had smaller z than the simultaneous condition($p < .001$), and the simultaneous condition had smaller z than the incongruent condition ($p = .005$). In the incongruent condition, the auditory-lead condition had larger z than the simultaneous condition ($p < .001$), and the simultaneous condition had larger z than the incongruent condition ($p < .001$). In the control condition three onset asynchrony conditions had similar z.

3. Experiment 2

The experiment left us with an unanswered question: how much influence would the auditory stimulus have if it were presented alone? In order to answer this question, we designed the experiment 2, which we make different length of auditory oscillating stimuli. The participants are asked to make judgment on the oscillation frequency of the stimuli with different number of circles of oscillation.

3.1 Participants

22 participants (four males) from Brandeis University undergraduate participants' pool took part in the experiment. The participants received course credits for participating the experiment.

3.2 Stimuli and procedure

In this experiment we did not present visual stimuli on the screen, and the auditory stimuli were using the same waves and sound pressure as experiment 1. The auditory stimuli are segmented by number of oscillation circles, which we chose two cycles, three cycles, four cycles and five cycles. As the length of the stimuli with the same oscillation number will differ in duration when they are in different frequency, we also created stimuli in the same duration for each cycle number. For example, two circles of 6 Hz oscillation have 333 ms, which is equal to the duration of 2.5 cycles of 7.5 Hz oscillation, and two cycles of 7.5 Hz oscillation is equal to the duration of 1.6 cycles of 6 Hz oscillation. Considering that the duration of stimuli might confuse with the number of cycles, we also included the "equal duration" stimuli in the experiment, such as 1.6 cycles in 6 Hz and 2.5 cycles in 7.5 Hz.

The experiment was separated into four blocks. In each block the stimuli are sound with certain number of oscillation and the "equal duration" stimuli. For example, in the block which the stimuli were two cycles' oscillation, there were 30 trials of 6 Hz two cycles oscillation, 30 trials of 7.5 Hz two cycles oscillations, 30 trials of 6 Hz 1.6 cycles oscillation, 30 trials of 7.5 Hz 2.5 cycles trial interleaved with each other. In this way we have blocks for two cycles, three cycles, four cycles and five cycles and the block order were randomized across participants. The total trial number for each experiment was 480 trials.

In each single trial, the participant saw a fixation on the screen indication the start a trial. The fixation presented for 500 ms. After the fixation disappear, the participants wait for a period of time range from 300 to 800 ms until the auditory stimulus presented, while the screen keeps in blank. At the end of the stimulus, an instruction appeared on the screen telling participants to make a key press response on whether the stimulus was oscillating in fast or slow frequency. After each response, the participants heard a sound feedback (high pitch beep for correct response and low pitch beep for error or no response). The participants started the experiment with 20 trials of practice with five circles stimuli, then they went to the experiment with 480 trials.

3.3 Result

The data was collected in Matlab (2015a) and analyzed on RStudio (version 1.1.423) with R 3.4.3. We mainly looked at the accuracy for different cycle numbers in this experiment. The main goal for the analysis was to look at in each cycle number conditions are the response accuracy different from chance level. The proportion of correct in each conditions are shown in Figure 8.

From Figure 8 we can see that participants' performance were approaching perfect when the number of cycles were more than three. The mean proportion of correct in two cycles condition was 0.78, in three cycles condition was 0.90, in four cycles condition was 0.92, in five cycles condition was 0.95. The lowest accuracy condition, which was the two cycles' condition, was significantly different from 50%, $t(21) = 11.94$, $p < .001$, indicating that even with two cycles of oscillation, participants can get enough amount of information from the stimuli to make judgment.

We also tested the effect of the duration of stimuli. In the two cycles and three cycles conditions, different time length have significantly different percent of correct. The analysis

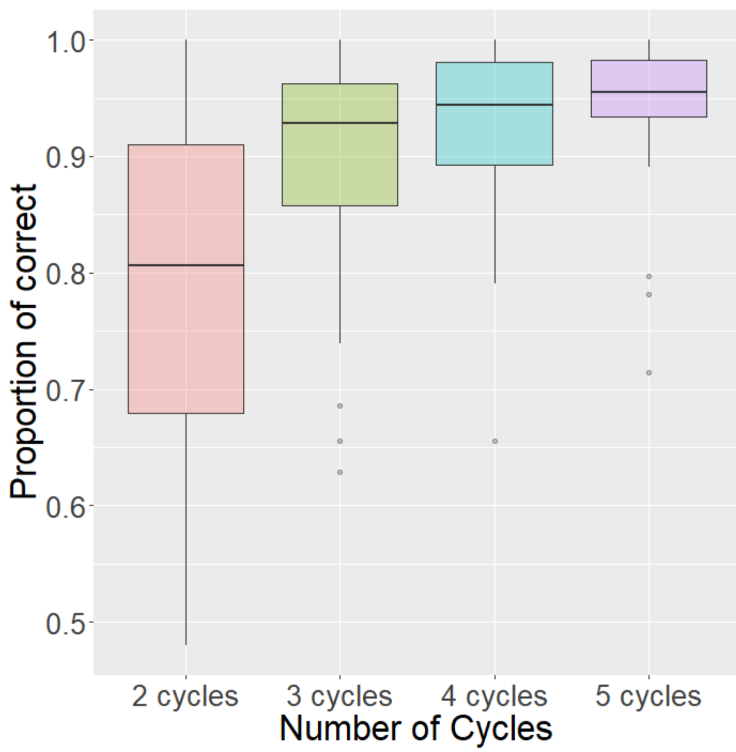


Figure 8: The figure for percent of correct in different cycle number conditions. The figure only included the trials with full cycles.

of time length effect are shown in Table 7.

Table 7: Percent of correct in each circle number and duration conditions

	Mean(short)[a]	Mean(long)[b]	t	df	p value
2 circles	0.70	0.87	5.53	83	<.001
3 circles	0.87	0.92	2.38	84	.02
4 circles	0.92	0.93	.48	84	.64
5 circles	0.95	0.95	.006	86	1

[a] Mean percent of correct for the stimuli length of 7.5 Hz oscillation in the cycle number of certain block
[b] Mean percent of correct for the stimuli length of 6 Hz oscillation in the cycle number of certain block

Although in two cycles' condition, the shorter duration lead to significantly lower performance than longer duration stimuli, the accuracy for the shorter duration stimuli was still better than chance level, $t(21) = 7.54$, $p < .001$. Thus, we can conclude that participants was able to receive extra temporal feature information from the two cycles' leading auditory stimuli if the unattended stimuli can cross the "attention barrier" to influence participants' perception.

4 Discussion

The result of accuracy and response time have shown the enhancement effect of cross-modal influence when audio-visual oscillation are congruent, which was consistent with the result of previous study (Sun et al., 2017). However, our study have also shown the disruption effect when audio-visual oscillation are incongruent, which was different from previous studies. It is possible that adding onset asynchrony in our design make the difference between congruency levels become easier to detect. Another possibility is that the experiment in Yile et al. (2017) have used more complicated stimuli than our study, which made their task more interesting while induced more noise in perception.

As the accuracy of response did not have significant difference across onset asynchrony conditions, only a difference in response time made it difficult to explain the effect of onset asynchrony. Considering speed-accuracy trade off, the faster response without better accuracy in auditory-lead condition can be explained as participants did receive extra information on frequency of oscillation from the early exposure of unattended stimuli, but they choose to use this advantage on making a faster response and did not make the response more accurate. It can also be argued that the early onset of the unattended stimuli only gave an alert to participants which made them respond earlier, and did not provide extra information that improve their task performance. We used diffusion decision model (DDM) to solve the confusion of this result.

We can find the effect of congruency and onset asynchrony conditions on drift rate. As drift rate reflected the effect of information accumulation on decision making, the significant differences in drift rate across congruency and onset asynchrony conditions indicated that the two factors did have influence on participants' perception. For congruency conditions, the higher drift rate in congruent condition have shown the beneficial of cross-modal influence on task performance when the audio-visual stimuli are temporally correlated, and the lower drift in the incongruent condition have shown the disruption effect of cross-modal influence when the audio-visual stimuli have different temporal feature. The same explanation cannot be used in explaining the drift rate difference between onset asynchrony conditions. If the drift rate difference between conditions were caused by the information from the unattended stimuli, the congruent and incongruent conditions should have opposite patterns of onset asynchrony influence: extra time of exposure to unattended stimuli should improve perception when the audio-visual stimuli were congruent, and the perception should be disrupted by extra exposure of unattended stimuli when the audio-visual stimuli were incongruent. However, in our result the auditory-lead condition had faster drift rate in all three congruency conditions. Thus, the increased drift rate cannot be explained by extra temporal feature information from the early presence of unattended stimuli.

We propose that the higher drift rate in auditory stimuli was mainly caused by our second hypothesis, which is the participants had been alerted by the early onset of auditory stimulus, and had extra time to prepare to judge the target stimulus. In Teichert et al. (2014) they found that delaying response onset can effectively improve response accuracy. Moreover, they found delaying the onset of decision making can significantly increase drift rate. Similar effect was found in an unpublished study in our lab (Sekuler, Sun, & Hickey, n.d.) which used the task in (Sun et al., 2017). This study shortened the inter-trial interval through 10 trials, and the average response accuracy dropped significantly when the inter-trial interval was shortened to a certain extent. Despite the auditory-lead condition, which the early presentation of unattended stimuli obviously increased participants' 'preparation time', the same reason can be used to explain the difference in drift rate between simultaneous and visual-lead condition. It is said that when the stimuli are ambiguous, participants need to integrate the information until they form a perception of the stimuli, then start the decision process (Teichert et al., 2014, 2015). This logic can be supported by the non-decision time result. The longer non-decision time in the incongruent condition could due to the process that the participants were trying to resolve the ambiguity of the stimuli and delayed the decision onset. The shorter non-decision time in auditory-lead condition could be caused by the fact that we record response time from the beginning of target presentation, so part of the non-decision time in auditory-lead condition was not recorded.

Both boundary separation and response bias had large interaction between congruency and onset asynchrony. The boundary separation, also termed as the decision threshold, only vary a lot in the incongruent condition, where the overall boundary separation was low and the simultaneous condition has very low boundary separation compare to other conditions. The low boundary separation should relate to the low accuracy in incongruent conditions, as short boundary separation make the decision process easier to reach the wrong side or response threshold (Ratcliff & McKoon, 2008; Ratcliff et al., 2016). The very low boundary separation in incongruent and simultaneous condition is possibly caused by a few subjects who performed at chance level only in the incongruent condition, and this individual effect have been exaggerated by the Markov Chain Monte Carlo process of HDDM. We also used the EZ-diffusion model (Wagenmakers et al., 2007) to estimate the parameters. The boundary separation was still significantly lower in incongruent condition and control condition ($p < .001$), but different onset asynchrony conditions in the congruent condition did not show large difference between each other.

The response bias in the congruent condition, the auditory-lead condition had a starting point closest to the 'slow' response and the visual-lead condition had a starting point closest to the 'fast' response. However, this pattern reversed in the incongruent condition. As the response bias is the bias at the decision onset, and the decision onset was the point of time after the 'preparation' of starting decision making process, it is possible that the bias reflect a correction of the general tendency of making a fast response. From the equivalent test we can see that participants have a general bias toward making a 'fast' response. As we have the same number of 'fast' and 'slow' trials in the experiment, the bias toward 'fast' response is detrimental on task performance. Thus, when participants have received more information from the early onset of unattended stimuli when the unattended and attended stimuli have the same oscillation frequency, the will tend to make better response and corrected from the tendency. On the other hand, when the unattended stimuli had different oscillation from the target stimuli, the early onset of unattended stimuli made the information in this trial become more confusing, and there would be less correction on the response bias. This explanation make the direct influence from the unattended stimuli become necessary.

As a conclusion, this study have found that cross-modal influence can be explained by mechanisms that different from temporal correlation. The information from the unattended stimuli can cross the 'barrier' of attention and influence the task performance. Moreover, we have shown a potential confusing factor that can happen in multisensory tasks. If the unattended stimuli can provide alert to the participants and make the participants prepare for the task earlier, it can improve the participants' perception of the target stimuli.

A caveat for the result of this study is that for an individual participant his or her performance can be very different from the overall performance. Some participants might have very different performance in different conditions while others have very similar performance across conditions. According to Ratcliff et al. (2016), individual difference have very small influence on the fitting of DDM model. It indicates that for the result of DDM model, we might need to consider that different participants may have different model parameters in an experiment.

References

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, *21*(3), 301–308.

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K., & Bizley, J. K. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron*, *97*(3), 640–655.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of The Royal Statistical Society. Series B (Methodological)*, 289–300. Retrieved from http://www.jstor.org/stable/2346101

Fujisaki, W., & Nishida, S. (2005). Temporal frequency characteristics of synchrony–asynchrony discrimination of audio-visual signals. *Experimental Brain Research*, *166*(3-4), 455–464. doi: 10.1007/s00221-005-2385-8

Grasman, R. P., Wagenmakers, E.-J., & Van Der Maas, H. L. (2009). On the mean and variance of response times under the diffusion model with an application to parameter estimation. *Journal of Mathematical Psychology*, *53*(2), 55–68.

Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. doi: 10.18637/jss.v069.i01

Maddox, R. K., Atilgan, H., Bizley, J. K., & Lee, A. K. (2015). Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. *Elife*, *4*, e04995. doi: 10.7554/eLife.04995

Molholm, S., Martinez, A., Shpaner, M., & Foxe, J. J. (2007). Object-based attention is multisensory: co-activation of an object's representations in ignored sensory modalities. *European Journal of Neuroscience*, *26*(2), 499–509.

Noise chart. (n.d.). Retrieved from https://www.noisehelp.com/noise-level-chart.html

Parker, J. L., Robinson, C. W., Gunzelmann, G., Howes, A., Tenbrink, T., & Davelaar, E. (2017). Auditory and visual contributions to multisensory integration. In *Proceedings of the 39th annual conference of the cognitive science society* (pp. 2858–2863).

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59-108.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: current issues and history. *Trends in cognitive sciences*, *20*(4), 260–281.

Rosser, D., Laidlaw, D., & Murdoch, I. (2001). The development of a reduced logmar visual acuity chart for use in routine clinical practice. *British Journal of Ophthalmology*, *85*(4), 432–436.

Sekuler, R., Sun, Y., & Hickey, T. (n.d.). *Audiovisual combination with temporal correlation and time pressure.* doi: 10.17605/OSF.IO/G5TMS

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, *408*(6814), 788. doi: 10.1038/35048669

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, *145*(3638), 1328–1330. doi: 10.1126/science.145.3638.1328

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. doi: 10.3758/s13414-010-0073-7

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology*, *13*(13), R519–R521. doi: 10.1016/S0960-9822(03)00445-7

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249–261. doi:

10.1007/s00221-013-3507-3

Stone, J., Hunkin, N., Porrill, J., Wood, R., Keeler, V., Beanland, M., ... Porter, N. (2001). When is now? perception of simultaneity. *Proceedings of the Royal Society of London B: Biological Sciences*, *268*(1462), 31–38. doi: 10.1098/rspb.2000.1326

Sun, Y., Hickey, T. J., Shinn-Cunningham, B., & Sekuler, R. (2017). Catching audiovisual interactions with a first-person fisherman video game. *Perception*, *46*(7), 793–814.

Teichert, T., Ferrera, V. P., & Grinband, J. (2014). Humans optimize decision-making by delaying decision onset. *PLoS One*, *9*(3), e89638.

Teichert, T., Grinband, J., & Ferrera, V. (2015). The importance of decision onset. *Journal of Neurophysiology*, *115*(2), 643–661.

Varghese, L., Mathias, S. R., Bensussen, S., Chou, K., Goldberg, H. R., Sun, Y., ... Shinn-Cunningham, B. G. (2017). Bi-directional audiovisual influences on temporal modulation discrimination. *The Journal of the Acoustical Society of America*, *141*(4), 2474–2488. doi: 10.1121/1.4979470

Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audio–visual integration in temporal perception. *International Journal of Psychophysiology*, *50*(1-2), 117–124.

Wagenmakers, E.-J., Van Der Maas, H. L., & Grasman, R. P. (2007). An ez-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). Hddm: hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, *7*. doi: 10.3389/fninf.2013.00014