Audiovisual integration deficits in schizotypal personality and implications for populations diagnosed with schizophrenia

by

Dawson Overton

Primary Advisor: Susanne Ferber

Subsidiary Advisor: Nicholas Rule

A thesis submitted in conformity with the requirements for the degree of Master of Arts

Graduate Department of Psychology

University of Toronto

© Copyright by Dawson Overton, 2015
Audiovisual integration deficits in schizotypal personality and implications for populations diagnosed with schizophrenia

Dawson Overton

Master of Arts

Graduate Department of Psychology

University of Toronto

2015

Abstract

Some evidence exists for audiovisual speech integration deficits in schizophrenia, but the generality of these deficits is unclear. We sought to characterize these deficits more fully by testing a general student population with varying degrees of schizotypal personality (SP) traits. Audiovisual integration of speech was measured with a McGurk task and non-speech audiovisual integration with a Sound-Induced Flash Illusion task. Visual integration capabilities were measured to determine if deficits were uniquely multisensory. Participants also completed the Schizotypal Personality Questionnaire (SPQ), which is predictive of SP traits. We found a negative correlation between SPQ score and performance on both audiovisual integration tasks, and no correlation between SPQ score and visual integration tasks. These findings support the claim that populations diagnosed with schizophrenia and populations with SP traits have a common deficit in a perceptual process and suggest that audiovisual integration deficits in these populations are not specific to speech integration.
Acknowledgements

I would like to thank my supervisor, Dr. Susanne Ferber, for useful and insightful feedback at all stages of this project. Dr. Ryan Stevenson’s suggestions and support were also invaluable, as he directed me to relevant background research and aided in interpretation of data. Discussions with Sol Sun, Rannie Xu, and Justin Ruppel were helpful as well, as were comments on my thesis proposal by Dr. Nicholas Rule.

I would like to additionally thank Sol Sun for the use of the Composite Face and Navon Letter Matlab scripts upon which two of my tasks were based (in slightly modified form). I would like to thank Dr. Ladan Shams for the use of her Sound-Induced Flash Illusion task script, and Dr. Ryan Stevenson for the use of the McGurk task script.
# Table of Contents

**List of Figures**

Introduction .............................................................................................................................. 1

Multisensory integration background .................................................................................. 1

Neuroanatomy of audiovisual binding .................................................................................. 2

Measuring audiovisual binding capability ............................................................................. 3

Audiovisual binding deficits in schizophrenia ..................................................................... 4

Measuring audiovisual integration in a population with schizotypal personality traits ....... 6

Hypotheses ............................................................................................................................... 7

Methods .................................................................................................................................. 8

Participants .............................................................................................................................. 8

Stimuli and apparatus ............................................................................................................. 9

McGurk task procedure ......................................................................................................... 9

Sound-Induced Flash Illusion task procedure ...................................................................... 11

Composite Face task procedure ............................................................................................ 12

Navon Letter task procedure ................................................................................................. 15

Questionnaires ....................................................................................................................... 17

Results ..................................................................................................................................... 18

McGurk task ............................................................................................................................ 19

Sound-Induced Flash Illusion task ....................................................................................... 20

Composite Face task ............................................................................................................. 21
List of Figures

Figure 1. Example of McGurk stimulus........................................................................................................10
Figure 2. Stimuli used in the Sound-Induced Flash Illusion task .................................................................12
Figure 3. Progression of Composite Face task trial .......................................................................................13
Figure 4. Example of Composite Face trial ....................................................................................................14
Figure 5. Example of Navon Letter stimuli..................................................................................................17
Figure 6. McGurk effect perception rate versus SPQ score .........................................................................19
Figure 7. Sound-Induced Flash Illusion perception rate versus SPQ score .................................................21
Figure 8. Composite Face Integration Measure versus SPQ score ...............................................................22
Figure 9. Composite Face task reaction time in seconds versus SPQ score ...............................................23
Figure 10. Navon Letter Integration Measure versus SPQ score .................................................................24
Figure 11. Navon Letter task reaction time in seconds versus SPQ score ...................................................25
Introduction

Multisensory integration background

Multisensory integration is an omnipresent phenomenon that is largely taken for granted in daily life, but its precise mechanisms and neural correlates still remain relatively unclear. Research in perception has historically restricted itself to a single modality at a time, and the study of how each modality influences the other and how modalities are combined into a coherent percept is relatively young. It was classically believed that sensory modalities were processed separately, and did not converge until they reached brain regions associated with much higher levels of cognition. A study by Meredith and Stein (1986) provided early evidence that this was not the case, by identifying neurons in the superior colliculus of the cat brain that appeared to take cues from afferent neurons coming from multiple sensory pathways (for example, the auditory and visual streams). A later book by the same investigators (1993) and a review paper (2014) expanded on these results and demonstrated that the auditory and visual sensory pathways (as well as other sensory pathways) appear to influence each other long before reaching the cortex.

Although integration appears to be fairly ubiquitous throughout the human perceptual systems, the most commonly investigated interaction is between auditory and visual stimuli, and this study concerns interaction between these two modalities alone. In addition, recent research in multisensory integration considers crossmodal effects on perception from the very earliest levels of perceptual processing (Senkowski, 2008), to feature extraction and integration (Taylor, 2006), to high-level cortical processing and object integration (Winters, 2010). For the purposes of this study, I am interested in integration at the “perceptual fusion” level, due to its role in
many higher-level cognitive abilities (such as the recognition of speech). I will be defining successful audiovisual integration as a subjective report of a percept containing a fusion of auditory and visual input, and will consider such a percept successfully “bound.”

How the brain segregates continuous sensory information from auditory and visual modalities into information belonging to a single object, and how this information from separate modalities is combined to represent that object as a coherent blended percept is not yet known in detail, though steps have been made toward an answer. Meredith and Stein have identified qualities of the stimuli that are predictive of whether or not a percept will successfully bind, and tasks in this study make use of two of them: temporal and spatial congruency (the closer in time or space two stimuli are perceived, the more likely they are to be bound).

Deficits in perceptual fusion of auditory and visual information are particularly interesting because of their ability to plausibly contribute to the symptoms of some developmental and psychotic disorders. For instance, an inability to integrate auditory and visual information in this way could underlie source recognition deficits that are a hallmark of schizophrenia. This possibility is the motivation of the present study, which seeks to more fully characterize previously observed audiovisual integration deficits in populations diagnosed with schizophrenia (SCZ) by first testing a student population with varying degrees of schizotypal personality traits.

**Neuroanatomy of audiovisual binding**

Many brain regions have been implicated in various stages of the audiovisual integration process. In macaques, premotor cortex was implicated in audiovisual integration (Smiley, 2009), close to or overlapping with regions involved in this integration in humans (e.g., Heschl’s gyrus).
Visual input to the primary auditory cortex was found in the superior temporal polysensory area (STP) of the superior temporal sulcus (STS), which is consistent with findings in humans (Jiang, 2001). Specific brain regions have also been implicated in perceptual fusion, or an actual combination of input across modalities (Miller and D’Esposito, 2005). These regions include Heschl’s gyrus and STS (like the previously mentioned studies), in addition to middle intraparietal sulcus and inferior frontal gyrus.

Most relevant to the present study, human STS activation has been found in audiovisual integration of speech. An fMRI study by Stevenson (2009) implicated this region in the perception of the McGurk effect (which describes the phenomenon of integrating incongruent auditory and visual speech stimuli into a blended percept; McGurk and MacDonald, 1976) in healthy controls and children with autism spectrum disorder (see Methods for a more detailed explanation of the McGurk task). Even more compellingly, a later study used fMRI-guided TMS to disrupt STS processing while participants participated in a McGurk task (Beauchamp, 2010). The application of TMS significantly reduced the perception of the McGurk effect in contrast to completing the task without TMS, lending more evidence to the region’s crucial role in audiovisual integration of speech. The role of STS in audiovisual integration has not been tested in SCZ populations explicitly, though some evidence exists for abnormalities in STS in these populations (Ross, 2007).

**Measuring audiovisual binding capability**

Several paradigms exist to investigate audiovisual binding capability. Both auditory and visual input have the ability to influence the opposite modality in a consciously noticeable way, though in general, these modalities do not have an equal influence on the final percept. In
particular, the visual modality tends to be stronger in most people (Shams, 2010), and will often “overpower” stronger, incongruent information in the auditory modality. A well-known example of this is the aforementioned McGurk effect (McGurk and MacDonald, 1976). If an observer sees a person enunciating a vowel (the visual stimulus) and hears a person speaking a different vowel simultaneously (the auditory stimulus), in most cases they will perceive a “blending” of these two vowels. For example, if the syllable /ba/ is the visual stimulus and the syllable /ga/ is the auditory stimulus, the syllable /da/ is usually perceived. If the same observer closes their eyes, they will perceive the “true” auditory stimulus. This effect is independent of knowledge of the illusion, appears to be involuntary, and illustrates the complex nature of perceptual blending (as an apparent continuum between auditory and visual information).

It is not always the case that visual information is the dominant determinant of a percept, however; auditory stimuli can sometimes “drive” the perception of the stimulus, particularly if the visual stimulus is very weak. In what is now known as the “Sound-Induced Flash Illusion” (Shams, 2000), the perception of a weak visual stimulus (a very short flash) can be influenced by auditory stimuli in close temporal proximity. More specifically, a single flash presented temporally close to two short auditory “beeps” will often cause the perception of two visual flashes. This somewhat counterintuitive example, in combination with the example of the McGurk effect, illustrates that the interaction between auditory and visual input is bidirectional (though the weighting each modality is given when determining a percept is almost certainly different).

**Audiovisual binding deficits in schizophrenia**
Audiovisual binding deficits have been observed in SCZ populations. A study by Pearl (2009) found that the rate of perception of the McGurk effect was significantly lower in SCZ adolescents than healthy controls (HC) adolescents, but did not find a large difference between SCZ adults and HC adults. Interestingly, SCZ populations appeared to have an auditory bias in this and other tasks, whereas a visual bias is common in HC populations. Another study by Ross and colleagues (2007) found that SCZ groups performed more poorly than controls on a speech in noise task, also indicating that audiovisual speech perception is impaired. The authors noted that unimodal (auditory) speech processing appeared to be intact in these individuals, pointing specifically to an integration deficit.

De Gelder and colleagues (2002) also found that SCZ groups performed more poorly than controls on an audiovisual speech integration task, but found no difference between the SCZ and HC groups on an audiovisual spatial localization task. Though inconclusive, these results suggest the possibility that audiovisual integration deficits found in later studies may be restricted to speech perception. Another possibility is that the spatial localization nature of the task made it possible for SCZ groups to perform well, whereas a speech integration task (which heavily relies on temporal simultaneity) is more challenging due to a decreased sensitivity to simultaneity in SCZ groups. This makes tasks such as the Sound-Induced Flash Illusion task ideal for future studies with schizophrenic populations, as it is a non-speech audiovisual binding task that makes use of temporal congruency for the perception of the illusion. Performance on a task of this nature would help to determine if audiovisual integration deficits are indeed limited to speech, or apply to audiovisual integration more generally in cases where temporal simultaneity is the key determinant in binding. To summarize, the Pearl, Ross, and De Gelder
studies agree that an audiovisual speech integration deficit exists in schizophrenic populations, but are ambiguous with respect to non-speech integration.

Limited evidence also exists for unimodal visual integration deficits in SCZ groups. Silverstein and colleagues (2009) found that SCZ groups performed worse than controls on a contour integration task. They also found reduced activations in visual areas V2-V4, but no change in activations in V1, suggesting that the most basic visual processes are likely intact in SCZ groups but processes for representing global configurations visually may not be. The role that visual integration has on observed audiovisual integration deficits is not known (though, as mentioned, unimodal auditory integration deficits are not observed).

Though it can be difficult to measure these deficits in populations diagnosed with schizophrenia directly, one way to measure them indirectly is in a general population with varying degrees of schizotypal personality traits. High levels of these traits often results in the diagnosis of schizotypal personality disorder (SPD), which is thought to be a “schizophrenia spectrum” disorder, though it is somewhat unclear if it represents a precursor to schizophrenia (see Kendler et al., 1981 for a discussion). SPD does, however, share several symptoms with schizophrenia, such as difficulty in interpreting social cues and speech mannerisms, and there is fairly strong evidence for a genetic link between schizophrenia and SPD (Kendler et al., 1981). Because of this relatively established link, and the complications involved with testing a clinical population, a student population with varying degrees of schizotypal personality traits was tested in this study, with the intention of using the results to make predictions about schizophrenic populations.

Measuring audiovisual integration in a population with schizotypal personality traits
The first aim of this study was to replicate the audiovisual speech integration deficits observed in schizophrenic populations (as measured by the McGurk effect) in schizotypal personality (SP) populations. The second aim was to investigate non-speech audiovisual integration in this population. To address this, general audiovisual integration was measured using the Sound-Induced Flash Illusion task, in order to better characterize the possible etiology of this audiovisual integration deficit. If the rate of perception of the illusion on the Sound-Induced Flash Illusion task correlated with SP trait measures, this would indicate an audiovisual integration deficit not only for speech, but also for more basic stimuli. In addition, unimodal visual integration was measured, in order to determine if audiovisual integration deficits might be partially explained by integration deficits in the visual modality.

In order to investigate multisensory integration in SP, a healthy student population took questionnaires measuring schizotypal personality traits. The McGurk task was used as a metric of audiovisual speech integration, and the Sound-Induced Flash Illusion task as a metric of audiovisual (non-speech) integration. Both social and non-social visual stimuli were used to test visual integration ability. This was done with a Composite Face task (Young, Hellawell, & Hay, 1987) and Navon Letter task (Navon, 1977) respectively (these tasks are explained more thoroughly in Methods).

Hypotheses

It was hypothesized that a participant’s performance on the audiovisual speech integration (McGurk) task would be negatively correlated with their score on the Schizotypal Personality Questionnaire (SPQ). As mentioned, previous work has shown that patients diagnosed with schizophrenia may exhibit lower performance on the McGurk task (Pearl, 2009).
Though this has not been shown with populations with schizotypal personality traits, it appeared reasonable to make a similar hypothesis in this case due to the similarity between the symptomology of these two disorders and the evidence of a genetic link between them (Kendler, 1981).

Performance on the Sound-Induced Flash Illusion task was of particular interest. If one makes the assumption that the aforementioned audiovisual integration deficits are due to the task reliance on temporal sensitivity (rather than being strictly speech-related), there was good reason to expect a negative correlation to be found with this task as well. It is, however, possible that the observed deficits in schizophrenics are due to a completely different mechanism (e.g., one which affects only social stimuli), in which case performance on this task should be no different from HC for the SP population. Predicted performance on the Navon Letter task and the Composite Face task was also less clear, though as mentioned earlier in the “Audiovisual binding deficits in schizophrenia” subsection, there were previously observed deficits in visual integration in SCZ groups (Silverstein, 2009).

Methods

Participants

Participants were recruited from a student population at the University of Toronto for course credit (using the internal Psynup and Psynup-Brain recruitment pools) or cash compensation. Participants had normal or corrected-to-normal vision and were native speakers of
English. In total, 41 participants (34 female) completed the battery of tasks (mean age 20.68 years, range 18-24 years). All participants provided written informed consent. Procedures were approved by the University of Toronto Research Ethics Board.

**Stimuli and apparatus**

Stimuli were presented on a 36 x 27 cm CRT monitor at a resolution of 1280x1024 pixels and a refresh rate of 85 Hz. Participants performed the tasks in a dimly lit room with the monitor placed approximately 40 cm in front of a chinrest, which the participant used throughout the study. Audio stimuli were presented binaurally through a pair of Cyber Acoustics ACM-500 noise-cancelling headphones. All tasks were presented using Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) for MATLAB (R2014a; MathWorks, Natick, MA). Each participant completed 4 tasks in the same order, and took breaks between tasks (but not between trials). The 4 tasks included a McGurk task, a Sound-Induced Flash Illusion task, a Navon Letter task (Navon, 1977), and a Composite Face task (Young, Hellawell, & Hay, 1987). Directly following the 4 tasks, participants completed the SPQ, the Community Assessment of Psychotic Experiences (CAPE), the 16 Item Prodromal Questionnaire (PQ-16), and the Revised Launay-Slade Hallucination Scale (LSHS-R).

**McGurk task procedure**

In the McGurk task, participants were presented with auditory and visual stimuli as illustrated in Figure 1. The visual stimulus consisted of a video of a woman pronouncing a syllable, and the auditory stimulus consisted of the same woman speaking a (possibly incongruent) syllable. The auditory and visual syllables presented varied by trial, but were
always one of 3 conditions: auditory “ba”, visual “ba” (congruent); auditory “ga”, visual “ga” (congruent); and auditory “ba”, visual “ga” (incongruent). A trial consisted of a fixation cross presented for 504-1500 ms (randomly jittered in increments of approximately 12 ms), a synchronous presentation of the auditory and visual stimuli, an additional presentation of the fixation cross for 250 ms, and a centered presentation of the message “What did she say: b, d, g, or th?” until the participant entered a response (each letter or pair of letters corresponding to the beginning of the syllable the participant perceived). This sequence was repeated for each trial, and the participant responded by pressing the corresponding “b”, “d”, “g”, or “t” key on a keyboard. If the participant indicated perceiving the blended syllable (“da” or “tha”) during an incongruent trial, they were considered to have perceived the McGurk effect for that trial, and if they indicated having perceived the auditory (“ba”) syllable or the visual syllable (“ga”), they were considered to not have perceived the McGurk effect. The task consisted of 60 trials, with 20 trials for each condition in randomized order.
Figure 1. Example of stimulus used in the McGurk task. (A) Shows 5 frames of the visual stimulus (in this case, “ga”; the actual stimulus is a video). (B) and (C) show the waveform and spectrogram, respectively, of the auditory stimulus (in this case, also “ga”). Visual and auditory stimuli are temporally aligned such that it appears the woman is uttering the (possibly incongruent) syllable. Image adapted from Stevenson (2014).

Sound-Induced Flash Illusion task procedure

In the Sound-Induced Flash Illusion task, participants were presented with visual “flash” stimuli (a white ring around the central fixation cross presented for 12 ms) and auditory “beep” stimuli (3,500 Hz pure tone with a duration of 7 ms). A total of 4 conditions with 20 trials per condition were presented, in randomized order. The 4 conditions were: two controls (one flash and one beep; two flashes and two beeps), and the illusory conditions (one flash and two beeps; two flashes and one beep). Flashes and beeps were both offset by the same small amount of time (43 ms) in conditions that included two flashes and/or two beeps (see Figure 2). A trial consisted of a fixation cross presented for 504-1500 ms (randomly jittered in increments of approximately 12 ms), followed by the presentation of the stimuli, another presentation of the fixation cross for 250 ms, and finally a centered presentation of the message “How many flashes did you see: 1 or 2?” until the participant responded. This sequence was repeated for each trial, and the participant responded with either the ‘1’ or ‘2’ key on a keyboard. Participants were explicitly instructed to report only the number of flashes and to ignore the beeps. If they responded with the number of beeps presented during an illusory trial, the illusion was considered perceived, while if they responded with the number of flashes presented, it was considered not perceived.
Figure 2. Stimuli used in the Sound-Induced Flash Illusion task. (A) and (B) show the two control conditions (one flash and one beep, and two flashes and two beeps, respectively). (C) and (D) show the illusory conditions (one flash and two beeps, and two flashes and one beep, respectively). Note that each case is preceded by a fixation cross presented for 504-1500 ms (jittered randomly in increments of approximately 12 ms) and succeeded by a fixation cross for 250 ms. A circle around the fixation cross indicates a flash, a speaker symbol indicates a beep, and all symbols indicate what was actually presented (rather than perceived).

Composite-face task procedure

In the Composite-face task, participants were shown a picture of a chimeric face (a picture of a face with top and bottom halves taken from different people) followed by another chimeric face with either the same or different bottom half and either the same or different top half (compared to the first face). Faces were either aligned or misaligned (offset by 50% of their width; see Figure 3). A 2x2x2 design with a total 8 conditions was used: faces may be the same (the top half of the first face is the same as the top half of the second face) or different, congruent (if the top half of the first face is the same as the top half of the second face, the bottom half of the first face is the same as the bottom half of the second face; if the two top halves are different,
the two bottom halves are also different) or incongruent (if the top halves are the same, the bottom halves are different; if the top halves are different, the bottom halves are the same), and misaligned or aligned. The task consists of 96 trials, with 12 trials per condition in randomized order. At the start of each trial, the message “Press SPACE” was displayed until the participant pressed the ‘space’ bar on a keyboard, followed by presentation of a fixation cross for 500 ms, presentation of the first face for 700 ms, presentation of a fixation cross for 200 ms, presentation of a mask (consisting of multiples of the letter ‘X’ in a rectangular pattern, in the same position on the screen as the first face) for 1000 ms, presentation of a fixation cross for 200 ms, presentation of the second face for 5000 ms (or until participant response), and finally presentation of a fixation cross for another 500 ms. Participants were instructed to press ‘s’ if the two top halves were the same or ‘d’ if they were different, and to answer as quickly as possible without sacrificing accuracy.

Figure 3. Schematic of a Composite Face task trial. Timing is shown beginning when the participant presses the space bar, and each fixation cross, stimulus, or mask is shown for the length of time given above the frame. Fixation cross, stimuli, and mask are not to scale.

If the participant correctly identified the top halves as being same or different, the trial was recorded as correct (otherwise it was recorded as incorrect). If the participant did not
respond within 5000 ms, the trial was excluded. In a participant with normal visual integration, judging if the top half has changed should be easier in trials where the face stimulus is misaligned. More specifically, performance should be worse for participants with normal visual integration on the aligned, incongruent trials compared with the aligned, congruent trials, and should show little to no difference between misaligned, incongruent trials and misaligned, congruent trials. No difference between the two aligned condition trials (in addition to the expected lack of difference between misaligned condition trials) may indicate a visual integration deficit. Reaction time was also recorded (measured from the time of presentation of the second face to the time of the participant’s press of the ‘s’ or ‘d’ key), with a slower reaction time on a correct trial taken to be indicative of the task being more difficult for the participant (due to, for example, a need to ignore an integrated face in the aligned condition). The top and bottom half-faces were randomly selected from a pool of 42 male and 42 female faces for each trial (though in congruent trials, the top face determined the bottom face). Face stimuli were presented with most of the hair, neck and background masked with solid white color. Given the position of the headrest, face stimuli subtended a visual angle of 6.8° vertically and 5.8° horizontally.
Figure 4. Example of Composite Face stimuli for the offset condition. In this example, the top halves of the chimeric faces have not changed, and the participant should respond “s”. In the actual trial, the second face is shown after a mask and a delay, not side-by-side.

An “integration measure” (IM) was calculated as follows (e.g., see Young et al., 1987):

$$IM = (Acc_{A,C} - Acc_{A,I}) - (Acc_{M,C} - Acc_{M,I})$$

Where $Acc_{A,C}$ is the accuracy on the aligned, congruent trials, $Acc_{A,I}$ is the accuracy on the aligned, incongruent trials, $Acc_{M,C}$ is the accuracy on the misaligned, congruent trials, and $Acc_{M,I}$ is the accuracy on the misaligned, incongruent trials. A positive value of IM indicated a higher difference in performance on the aligned faces when they were congruent relative to incongruent, and a smaller difference in performance between congruent and incongruent cases when the faces were not aligned. This is thought to be a measure of holistic face processing ability because individuals with high holistic face processing ability will integrate the top and bottom halves more strongly when the halves are aligned, making the aligned, incongruent case significantly more difficult than the aligned, congruent case. Reaction times were calculated using the same formula, using reaction times for the corresponding trials instead of accuracies.

Navon-letter task procedure

In the Navon-letter task, participants were shown a visual stimulus of a letter composed of smaller letters (see Figure 4) and instructed to either report the “small” letter (the local letter) or the “large” letter (the global letter). At the position of the headrest, stimuli subtended a visual angle of 4.0° vertically and 3.1° horizontally. Only the letters ‘s’ and ‘h’ were used for both global and local letters. The task consisted of a total of 192 trials. At the beginning of the first
trial, the participant was randomly instructed to report either the local or global letter for each subsequent trial, and at the beginning of the 97th trial (halfway), the alternate instruction was given for each subsequent trial (i.e., if the local instruction was given for trials 1-96, the global instruction was given for trials 97-192, and vice versa). Additionally, prior to the first and 97th trial, participants were given 8 “practice trials” (with the same global or local instructions as that set of 96 trials) that displayed whether their response was correct or incorrect (this feedback allowed the experimenter to determine if the participant understood the instructions, and was not given during actual trials). Each block of 96 trials consisted of 24 trials of each of the 4 possible stimuli types (local ‘s’, global ‘s’; local ‘s’, global ‘h’; local ‘h’, global ‘s’; local ‘h’, global ‘h’). The conditions in which the global and local letter was the same (both “s” or both “h”) were defined “congruent”, and the conditions in which they were not the same were defined “incongruent.” This resulted in a total of 4 conditions (congruent or incongruent stimuli, and global or local instructions). At the beginning of each trial, the message “Press SPACE” was displayed until the user pressed the “space” key on a keyboard. After the user response, a fixation cross was shown for 500 ms, the Navon letter was shown for 5000 ms (or until the user pressed the ‘s’ or ‘h’ key), and the fixation cross was shown for another 500 ms. If the user did not press the ‘s’ or ‘h’ key while the stimulus was present, the trial was excluded. If the user pressed the key corresponding to the current instruction (i.e., if they pressed the key corresponding to the local letter during a local instruction trial, or the global letter during a global instruction trial), the trial was recorded as correct, while the opposite response was recorded as incorrect. Reaction time was also recorded (measured from the beginning of stimulus presentation to time of the press of the ‘s’ or ‘h’ key), with a slower reaction time on a correct
trial taken to be indicative of the task being more difficult for the participant (due to, for example, a need to ignore an incongruent letter).

Figure 5. Example of two possible Navon Letter stimuli (on left, global ‘h’ and local ‘s’; on right, global ‘s’ and local ‘h’). Only one global letter was shown in each actual trial.

Similar to the Composite Face task, an “Integration Measure” was calculated as follows:

\[ IM = (Acc_{L,C} - Acc_{L,I}) - (Acc_{G,C} - Acc_{G,I}) \]

Where \( Acc_{L,C} \) is the accuracy on the local instruction, congruent trials, \( Acc_{L,I} \) is the accuracy on the local instruction, incongruent trials, \( Acc_{G,C} \) is the accuracy on the global instruction, congruent trials, and \( Acc_{G,I} \) is the accuracy on the global instruction, incongruent trials. A positive value of IM indicated a higher difference in performance between congruent and incongruent on the local cases, but no such difference on the global cases, thus suggesting a higher integration ability. Reaction times were calculated using the same formula, using reaction times for the corresponding trials instead of accuracies.

**Questionnaires**

Participants were given 4 questionnaires, meant to measure likelihood of a SPD diagnosis and levels of schizotypal personality traits. In all questionnaires, higher scores represent higher likelihoods of schizotypal personality traits. These questionnaires were: the Schizotypal
Personality Questionnaire (SPQ; Raine, 1991), the 16 Item Prodromal Questionnaire (PQ-16; Ising, 2012), the Community Assessment of Psychotic Experiences (CAPE; Konings, 2006), and the Revised Launay-Slade Hallucination Scale (LSHS-R; Launay, 1981). SPQ consists of 74 questions (divided into 9 symptom categories), each given a score of either 0 or 1 (yes or no), for a maximum score of 74. CAPE consists of 42 questions, each given a score between 0-3 (a Likert scale), for a maximum score of 126. PQ-16 consists of 16 questions, each given a score between 0-3 (a Likert scale), for a maximum score of 48. LSHS-R consists of 12 questions, each given a score between 0-4 (a Likert scale), for a maximum score of 48. As an example, one question in the SPQ for the symptom category “excessive social anxiety” is “Do you ever get nervous when someone is walking beyond you?”, and a question for the symptom category “unusual perceptual experiences” is “Have you often mistaken objects or shadows for people, or noises for voices?” (Raine, 1991).

Raine (1991) found that the SPQ has high internal validity, and the majority of the top 10 percent of scorers has a clinical diagnosis of schizotypal personality disorder. The SPQ directly measures likelihood of a SPD diagnosis, whereas CAPE, PQ-16 and LSHS-R measure risk of psychosis more generally (which may include SPD). Among these questionnaires, it was expected that the SPQ would be most useful for the population of interest, due to having a higher score range (increasing statistical power) and mostly regarding less severe symptoms of psychosis. The PQ-16 and LSHS-R asked questions regarding more severe symptoms, and were unlikely to have high scores for a general student population (but were still collected for completeness). Correspondingly, the following analyses focus on the SPQ score.

Results
McGurk task

The mean rate of perception of the McGurk effect for the 41 participants was 0.5634 ($SD = 0.3221$). The mean SPQ score was 30.56 ($SD = 11.19$).

A Pearson correlation was calculated between SPQ score and rate of perception of the McGurk effect. Rate of perception of the McGurk effect was found by considering only illusory trials (different auditory and visual stimuli) and calculating the percentage of trials in which the participant reported a bound percept.
Figure 6. McGurk effect perception rate plotted against SPQ score with 95% confidence band. The perception rate is calculated by dividing the number of trials in which the illusion is perceived by the total number of illusory trials.

A significant negative correlation between McGurk perception rate and SPQ was found, $r(39) = -.3264$, $p = .0372$, 95% CI [-.5762, -.0208]. This indicates that, in line with our hypothesis, individuals with higher SPQ scores were less likely to perceive the McGurk effect during illusory trials. Power was calculated to be 0.56.

**Sound-Induced Flash Illusion task**

The mean rate of perception of the Sound-Induced Flash Illusion for the 41 participants was 0.6878 ($SD = 0.1046$).

A Pearson correlation was calculated between SPQ score and rate of perception of the Sound-Induced Flash Illusion (SIFI). Again, the rate of perception of the SIFI was found by considering only illusory trials and calculating the percentage of trials in which the participant reported the number of auditory stimuli rather than the number of visual stimuli.
A significant negative correlation between McGurk perception rate and SPQ was found, $r(39) = -.3261$, $p = .0374$, 95% CI [-.5760, -.0206]. This indicates that, in line with our hypothesis, individuals with higher SPQ scores were less likely to perceive the Sound-Induced Flash Illusion during illusory trials. Power was also calculated to be 0.56.

**Composite Face task**
The mean Composite Face Integration Measure (see Methods) for the 41 participants was 0.0438 ($SD = 0.1112$). The mean Integration Measure reaction time for the task was 0.816 s ($SD = 0.171$ s).

In order to assess the relationship between schizotypal personality traits and holistic face processing, we calculated a Pearson correlation between the Composite Face Integration Measure and SPQ score.

*Figure 8.* Composite Face Integration Measure plotted against SPQ score with 95% confidence band.
The integration measure was not found to be correlated with the SPQ score, $r(39) = .0182, p = .9111, 95\% \text{ CI } [-.2949, .3279]$.

![Graph showing the correlation between SPQ score and integration measure](image)

*Figure 9.* Integration Measure reaction time in seconds on Composite Face task plotted against SPQ score with 95% confidence band.

In addition, we did not find a significant correlation between reaction time on the task and SPQ score, $r(39) = .1648, p = .3096, 95\% \text{ CI } [-.1547, .4530]$.

**Navon Letter task**
The mean Navon Letter Integration Measure (see Methods) for the 41 participants was 0.00051 ($SD = 0.0540$). The mean Integration Measure reaction time for the task was 0.509 s ($SD = 0.089$ s).

To assess the relationship between schizotypal personality traits and the global precedence effect, we calculated a Pearson correlation between the Navon Letter Integration Measure and SPQ score.

*Figure 10.* Navon Letter Integration Measure plotted against SPQ score with 95% confidence band.
The integration measure was not found to be correlated with the SPQ score, $r(39) = -0.0091, p = 0.955$, 95% CI [-0.3159, 0.2994].

**Figure 11.** Integration Measure reaction time in seconds on Navon Letter task plotted against SPQ score with 95% confidence band.

In addition, the reaction time on the task was not found to be correlated with the SPQ score, $r(39) = 0.1369, p = 0.3934$, 95% CI [-0.1783, 0.4266].

Partial correlations between McGurk, SIFI, and visual integration tasks
In order to determine the extent to which performance on the two visual integration tasks might explain performance on the audiovisual integration tasks, partial correlations between SPQ score and rate of perception on both audiovisual tasks (accounting for either Navon or Composite Face integration measure) were calculated.

A significant partial correlation between SPQ score and rate of McGurk perception, accounting for the integration measure of the Composite Face task, was found, \( r(39) = -0.347, p = 0.0241 \). A significant partial correlation between SPQ and rate of SIFI perception (also accounting for the integration measure of the Composite Face task) was found, \( r(39) = -0.315, p = 0.0434 \). No changes in the strength of correlation were observed, suggesting variance in audiovisual integration cannot be explained by visual integration as measured by the Composite Face task.

A significant partial correlation between SPQ score and rate of McGurk perception, accounting for the integration measure of the Navon Letter task, was found, \( r(39) = -0.326, p = 0.0334 \). A significant partial correlation between SPQ and rate of SIFI perception (also accounting for the integration measure of the Navon Letter task) was found, \( r(39) = -0.326, p = 0.0332 \). No changes in the strength of correlation were observed for this task either, suggesting variance in audiovisual integration cannot be explained by visual integration as measured by the Navon Letter task.

Finally, a partial correlation between SPQ score and rate of SIFI perception, accounting for rate of McGurk perception, was calculated, \( r(39) = -0.300, p = .0517 \). A partial correlation between SPQ score and rate of McGurk perception, accounting for rate of SIFI perception, was also calculated, \( r(39) = -0.301, p = .0514 \). The fact that these correlations remain relatively strong
suggests that performance on each task accounts for a substantially independent portion of the variance in SPQ score.

**Discussion**

In this study, we sought to determine if audiovisual speech integration deficits could be predicted by schizotypal personality severity in a general student population (replicating previous findings in populations diagnosed with schizophrenia), and if so, if this only applied to speech-related deficits. Consistent with the original hypotheses, a significant negative correlation was found between the rate of perception of the McGurk effect and SPQ score, as well as between the rate of perception of the Sound-Induced Flash Illusion and SPQ score. To our knowledge, these findings are novel and these relationships have not been investigated previously.

The results for the McGurk task extend previous findings that McGurk effect perception is disrupted in populations diagnosed with schizophrenia (Pearl, 2009), and suggest that the deficits observed in that study may not be limited to those with severe schizophrenic symptoms, but may also exist in a general population, varying according to degree of severity of schizotypal personality traits. A similar deficit with a non-social stimulus (the Sound-Induced Flash Illusion) has, to our knowledge, not been previously investigated in SCZ groups, but was found to correlate with schizotypal personality trait severity, suggesting a possible analogous deficit in SCZ groups.
In the Composite Face and Navon Letter tasks, no significant correlation was found with SPQ score. Reaction times for the Composite Face and Navon Letter tasks also did not vary significantly with SPQ score. The possibility that performance on the visual integration tasks could explain audiovisual integration performance was also tested (despite seeming extremely unlikely given the lack of correlation between visual tasks and SPQ), and it was confirmed that such an explanation was not plausible.

The McGurk task results confirm a symptomatic link between SCZ and SPD groups. As mentioned previously, there is some debate regarding the similarity between these classifications, ranging from the belief that SPD is simply a “milder” form of SCZ with a similar genetic basis (Kendler, 1981; Vollema, 2002), to the claim that they are different disorders which share a limited number of symptoms but differ in others ways. Although these results do not necessarily strongly support one of these claims over the other, they do support the claim that SPD and SCZ share a perceptual deficit. It was not clear at the beginning of the study if the SIFI perception rate would also correlate with schizotypal personality trait severity, and the fact that it does helps to qualify the SPD deficit. The results from the SIFI task support a general model of audiovisual integration deficit in SPD and not merely a deficit in the integration of “social” stimuli, or even more specifically, speech recognition.

Developmental reasons for a general audiovisual integration deficit have been suggested in other clinical populations (Altieri, 2013). For the tasks used in this study, the lack of temporal offset appears to be the key quality that causes the stimuli to be bound across modalities. This had led some investigators to posit that individuals may have a specific sensitivity to judging simultaneity, or a “temporal binding window” (TBW), representing the length of time that the
visual stimulus can be offset from the auditory stimulus and still be likely to be perceptually bound (Wallace, 2014). As mentioned previously, temporal congruency is not the only factor that determines if a percept will be bound, but this measure provides a useful way to measure the relative integration capability between individuals.

The same authors propose that a narrow TBW is a product of typical brain development in a stimulus-rich environment. As we observe the environment throughout our lives, we often experience stimuli from different modalities that seem to reliably originate from the same spatial source and at the same time (e.g., a cat opening its mouth and a meowing sound). After enough of these observations across a wide enough variety of objects occur, it is thought that this generalized “rule” is represented neurally at an early stage in the auditory and visual streams (i.e., temporospatial congruence implies the same source and thus should be perceptually bound). Since the temporal offset between auditory and visual stimuli originating from the same source is very small in most natural cases (excluding cases where the stimuli are very far away from the observer and the much slower speed of sound compared to light becomes important), it is plausible that typical development would bias towards a narrow TBW.

Developmental studies appear to be in agreement with this view; children tend to have a much larger TBW (as measured in the manner explained by Wallace, 2014), which narrows with age. Importantly, they also have a reduced ability to bind audiovisual stimuli (Hillock, 2011), suggesting a link between a narrow TBW and audiovisual binding capability (Powers, 2009). Conversely, a wide TBW appears to be associated with a lack of audiovisual binding capability. This makes intuitive sense; if an individual has not “learned” that natural stimuli tend to have a very small temporal offset across modalities, they may have difficulty recognizing that such stimuli originated from the same source. These deficits in development of audiovisual binding
 AUDIOVISUAL INTEGRATION DEFICITS IN SCHIZOPHRENIA

capabilities have recently been proposed as an etiology for certain symptoms of developmental disorders such as in autism spectrum disorder (ASD). Stevenson has argued (2014) that previously described deficits in binding associated with a wide TBW may “cascade” up to much higher-level perceptual phenomena such as speech recognition.

Given this discussion on audiovisual integration deficits in other clinical populations and in typically developing adults, the fact that the observed SCZ deficit appears to be a general audiovisual deficit could guide further studies investigating its etiology. Since this developmental reason for a general audiovisual integration deficit exists and has found support in other clinical populations, it is possible that similar developmental support could be found for audiovisual integration deficits in SCZ groups. In any case, the present results would suggest investigating a mechanism that is relatively early in the auditory and visual processing streams, and not necessarily specific to language-specific regions of the auditory cortex.

There are some possible alternative interpretations of the findings. For example, given the nature of the tasks, it is possible that the SPQ scores could be correlated with a general performance deficit (such that high-scoring individuals would be less likely to pay attention to the stimuli properly and thus hear only the auditory stimulus in the McGurk task, for example), or a visual integration performance deficit. The partial correlation analyses, and the lack of correlation between the SPQ and visual tasks, however, render these explanations unlikely.

The mean values of the Integration Measures in the visual integration tasks were somewhat surprising (being very close to 0). In a general student population, the average of this measure was expected to be slightly positive, as there was no reason to expect a consistent visual integration deficit in this population. Since visual integration ability is also subject to individual
differences and negative values of IM are possible, it is likely that these variations balanced each other out (since several participants did have reasonably positive IM values). In any case, replicating the baseline effects for these tasks was less important for this study than determining if they offered an explanation for audiovisual integration performance (which was, again, shown to be unlikely).

In addition, a limitation of this study is the gender ratio (34 out of 41 are female). This could result in difficulties generalizing to a population consisting of both sexes (in fact, the prevalence of both SCZ and SPD is slightly higher in males than females). Previous work in patients with SCZ participants that included more males than females, however, produced results in line with the current study’s results, mitigating this factor. Finally (as seen in Methods), statistical power, while reasonable, is still slightly low; in order to raise it to a value of 0.8 for the observed correlations, a total of 52 participants would be needed; this can be addressed by increasing the sample size by 11 participants.

Conclusions and Future Work

The present study supports a previously uninvestigated, perceptual link between SPD and SCZ, and replicated previous findings in the schizophrenia literature. In addition, it has begun to characterize the nature of this deficit in SPD, and if this link between SPD and SCZ is indeed valid, it could guide future investigations of deficits in SCZ (i.e., to verify that they are also of a general audiovisual nature). Performing the same tasks outlined in this study with a population diagnosed with schizophrenia would be important for ensuring that the schizotypal personality
findings do extend to schizophrenic populations. Little work has been done investigating the role of temporal insensitivity to audiovisual speech integration deficits in SCZ, so this would be a worthwhile follow-up study (and would help to further cement similarities in perceptual deficits in SPD and SCZ).

In addition to this, a neuroimaging study investigating level of activation of STS would be useful in beginning to elucidate mechanistic reasons for the observed deficits in both SPD and SCZ, especially since some evidence of abnormalities in STS for SCZ groups already exists (Ross, 2007). Similar to previous studies in other clinical populations (e.g., Beauchamp, 2010), TMS could be used to investigate the causal role of STS in the deficits. This would be particularly interesting because it could either help connect SPD and SCZ deficits to the existing body of evidence in other clinical populations, or suggest that the deficits have a unique mechanism in SCZ (despite similarities at the perceptual level).
References


Pearl, D., Yodashkin-Porat, D., Katz, N., Valevski, A., Aizenberg, D., Sigler, M., … Kikinzon, L. (2009). Differences in audiovisual integration, as measured by McGurk phenomenon,
among adult and adolescent patients with schizophrenia and age-matched healthy control groups. Comprehensive Psychiatry, 50(2), 186–92. doi:10.1016/j.comppsych.2008.06.004


Vollema, M. G., Sitskoorn, M. M., Appels, M. C. M., & Kahn, R. S. (2002). Does the Schizotypal Personality Questionnaire reflect the biological-genetic vulnerability to

