RESEARCH ARTICLE

Multisensory temporal integration: task and stimulus dependencies

Ryan A. Stevenson · Mark T. Wallace

Abstract The ability of human sensory systems to integrate information across the different modalities provides a wide range of behavioral and perceptual benefits. This integration process is dependent upon the temporal relationship of the different sensory signals, with stimuli occurring close together in time typically resulting in the largest behavior changes. The range of temporal intervals over which such benefits are seen is typically referred to as the temporal binding window (TBW). Given the importance of temporal factors in multisensory integration under both normal and atypical circumstances such as autism and dyslexia, the TBW has been measured with a variety of experimental protocols that differ according to criterion, task, and stimulus type, making comparisons across experiments difficult. In the current study, we attempt to elucidate the role that these various factors play in the measurement of this important construct. The results show a strong effect of stimulus type, with the TBW assessed with speech stimuli being both larger and more symmetrical than that seen using simple and complex non-speech stimuli. These effects are robust across task and statistical criteria and are highly consistent within individuals, suggesting substantial overlap in the neural and cognitive operations that govern multisensory temporal processes.

Keywords Audiovisual · Multisensory integration · Temporal binding window · Psychophysics · Temporal perception · Speech perception · Synchrony · Multimodal

R. A. Stevenson (✉) · M. T. Wallace
Department of Hearing and Speech Sciences, Vanderbilt
University Medical Center, 7110 MRB III BioSci Bldg 465,
21st Ave South, Nashville, TN 37232, USA
e-mail: ryan.andrew.stevenson@gmail.com

R. A. Stevenson · M. T. Wallace
Vanderbilt Kennedy Center, Nashville, TN, USA

R. A. Stevenson · M. T. Wallace
Vanderbilt Brain Institute, Nashville, TN, USA

M. T. Wallace
Department of Psychology, Vanderbilt University,
Nashville, TN, USA

M. T. Wallace
Department of Psychiatry, Vanderbilt University,
Nashville, TN, USA

Introduction

A primary role of the nervous system is to convert sensory inputs into directed and adaptive actions. This is achieved through the different sensory systems working in concert, integrating information across the senses and improving behavioral outcomes as seen in higher detection rates (Stein and Wallace 1996; Lovelace et al. 2003), more accurate localization (Wilkinson et al. 1996; Nelson et al. 1998), and faster responses (Hershenson 1962; Diederich and Colonius 2004). The ability to integrate information to generate neural, behavioral, and perceptual gains is known to be highly dependent upon the spatiotemporal relationship of the different sensory inputs (Wallace et al. 2004). A great deal of work has focused on the importance of temporal factors in dictating multisensory interactions and has shown robust effects of stimulus timing at the level of single neurons (Meredith et al. 1987, 1992; Royal et al. 2009), neural populations measured by EEG (Senkowski et al. 2007; Schall et al. 2009) and fMRI (Macaluso et al. 2004; Miller and D'Esposito 2005; Stevenson et al. 2010, 2011), and in behavioral and perceptual paradigms

(Dixon and Spitz 1980; Wallace et al. 2004; Keetels and Vroomen 2005; Zampini et al. 2005; Conrey and Pisoni 2006; van Atteveldt et al. 2007; van Wassenhove et al. 2007; Powers et al. 2009; Hillock et al. 2011).

Given the importance of temporal factors for multisensory integration and emerging evidence for deficits in multisensory temporal processing in clinical populations such as autism and dyslexia (Hairston et al. 2005; Foss-Feig et al. 2009; Kwakye et al. 2011; de Boer-Schellekens et al. 2013; Woynaroski et al. 2013), a number of experimental paradigms have been developed to examine these effects. Many of these studies have gravitated toward the construct of the *temporal binding window* (TBW)—the range of stimulus onset asynchronies (SOAs) in which multiple sensory inputs have a high probability of altering responses. As such, the TBW can serve as a proxy measure for when multisensory integration occurs. Not surprisingly, the TBW has been measured using a variety of different tasks, stimulus types, and statistical criteria (Dixon and Spitz 1980; Miller and D'Esposito 2005; Zampini et al. 2005; Conrey and Pisoni 2006; Vatakis and Spence 2006; van Atteveldt et al. 2007; van Wassenhove et al. 2007; Stevenson et al. 2010). These differences make comparisons across studies very difficult, thus limiting our ability to map this construct onto the relevant neural and cognitive operations, previously identified in regions of the insula, interior parietal sulcus, inferior frontal gyrus, and superior temporal sulcus (Miller & D'Esposito 2005; Stevenson et al. 2011).

To tackle this issue, the current study provides the first parametric analysis of how task, stimulus type, and statistical criterion impact the TBW. The TBW is measured using stimuli ranging from simple flashes and beeps to complex audiovisual speech using four distinct tasks and employing two different criteria. These data provide a novel view into the task and stimulus dependencies of the TBW, showing robust effects of both task and stimulus type. Perhaps most importantly, these effects were found to be highly correlated within individuals, hinting at a common set of neural and cognitive operations underlying multisensory temporal processes and that serve to bind multisensory stimuli into unitary events.

Methods

Participants

Thirty-nine participants (15 male, mean age = 20.9 years, SD = 3.8 years, range = 18–35 years) were run, and all procedures were approved by the Vanderbilt Institutional Review Board.
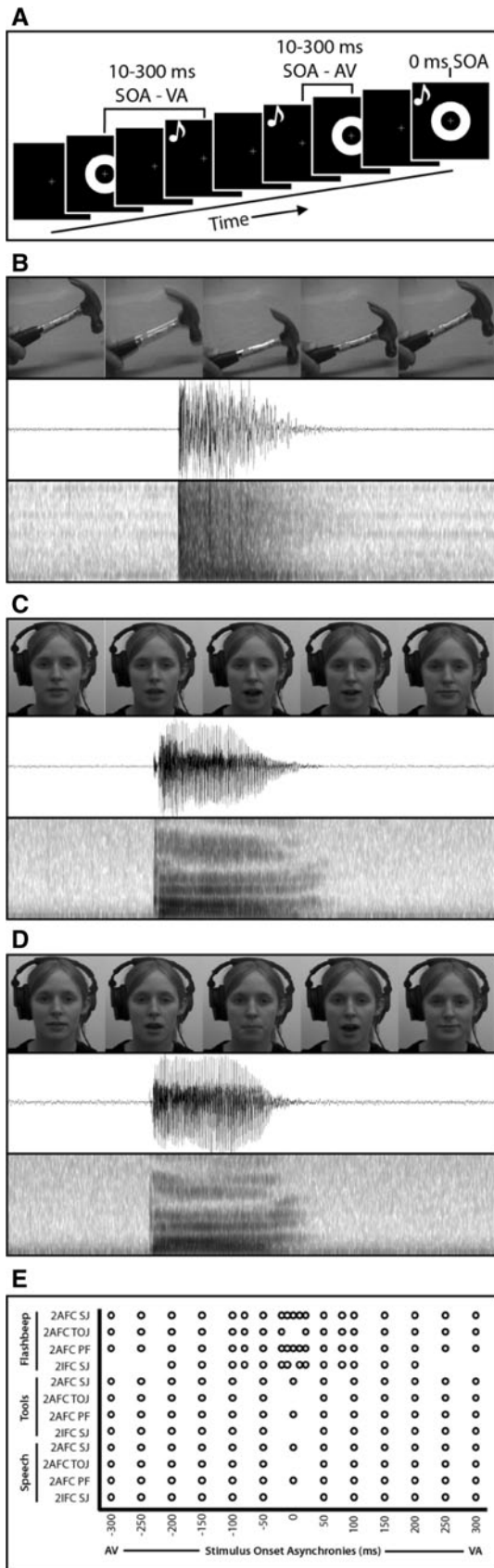
Stimuli

All stimuli were presented using MATLAB (MATH-WORKS Inc.) through the Psychophysics Toolbox extensions (Brainard 1997; Pelli 1997). Visual stimuli were presented at a distance of approximately 60 cm from the participants, and auditory stimuli were presented binaurally via noise-canceling headphones. Experimental timing was confirmed via oscilloscope.

Three distinct categories of audiovisual stimuli were presented: simple flashbeeps, dynamic handheld tools, and single syllable utterances. These stimulus types were chosen for a number of reasons. First is their use in previous studies. Speech stimuli have been used in a plethora of studies, and speech integration is by itself a widely studied perceptual process. Similarly, in terms of basic studies of sensory perception, it is quite common to use extremely limited stimuli, such as pure tones, noise bursts, flashes, and checkerboards. Dynamic handheld tool stimuli were also included as a complex, non-speech stimulus class. These tool stimuli, like the speech stimuli, are relatively complex, include biological motion, and presentation of a body part, and as such can be used as a contrast to assess speech-specific differences in processing.

The visual component of the flashbeep stimuli consisted of a white ring circumscribing the visual fixation cross on a black background and was 18.25 cm in diameter or 17.3° of visual angle. Visual stimulus duration was 10 ms. Auditory beep stimuli consisted of a 3,500 Hz pure tone with a duration of 13 ms. See Fig. 1a for a graphical representation.

Dynamic handheld tool stimuli were selected from a stimulus set that has been previously used successfully in studies of multisensory integration (Stevenson et al. 2007, 2009; James et al. 2009, 2011, 2012; Stevenson and James 2009; Wallace and Murray 2011). Stimuli consisted of two audiovisual clips showing a hammer in action. Hands were visible in the recordings. Recordings were made with a DCR-HC85 MiniDV Digital Handycam camcorder. Separate video and audio files were extracted from the raw recordings, such that they could be manipulated and presented separately as visual and auditory stimuli. Video was acquired at the camera's native resolution of 1,024 × 720. Audio was acquired with 16 bits at a sampling rate of 32 kHz with the camcorder's onboard microphone. Visual stimuli were cropped to square, down-sampled to a resolution of 400 × 400 pixels spanning 18.25 cm per side or 17.3° of visual angle, and converted from color to grayscale. Audio was converted from stereo to mono. Presentations were shortened to 1 s, and each presentation included a full cycle of motion for the given tool. See Fig. 1b for a graphical representation.

Single syllable utterances were selected from a stimulus set that has been previously used successfully in studies of

multisensory integration (Schall and Hanes 1993; Quinto et al. 2010). Stimuli consisted of two audiovisual clips of a female speaker uttering single instances of the syllables/ba/and/ga/. Visual stimuli were cropped to square, downsampled to a resolution of 400 × 400 pixels spanning 18.25 cm per side or 17.3° of visual angle, and converted from color to grayscale. Presentations were shortened to 2 s, and each presentation included the entire articulation of the syllable, including pre-articulatory gestures. See Fig. 1c, d for graphical representations.

Procedure

Each participant completed 12 experiments, including 4 cognitive tasks with each of the 3 stimulus types. With each task, stimulus onset asynchronies (SOA) were parametrically varied. See Fig. 1e for SOAs with each stimulus type and task. Tasks included a two-alternative, forced-choice simultaneity-judgment task (SJa: "Were the auditory and visual stimuli presented at the same or different times"), a temporal-order judgment task (TOJ: "Was the auditory or visual stimulus presented first"), a perceptual-fusion task (PF: "Did you perceive the presentation as a single event"), and a two-interval, forced-choice task (SJi: two presentations, one synchronous and one asynchronous, "Was the first or second presentation synchronous"). Experiment orders were randomized across participants. SOAs were tailored for each task and stimulus type based on previous research, with notable differences being that TOJ and SJi tasks did not include a 0 ms SOA (as there would then be no correct response) and that simple flashbeep stimuli included more dense SOAs between 0 and 150 ms than the more complex stimuli (Vatakis and Spence 2006; Vatakis et al. 2008). In detail, the SOAs for each task and stimulus type were as follows (note that all nonzero values included an auditory-leading and a visual-leading condition with that respective SOA). Flashbeep SJa and flashbeep PF tasks used SOAs of 0, 10, 30, 50, 80, and 100–300 ms in 50-ms increments. The flashbeep TOJ task used SOAs of 30, 50, 80, and 100–300 ms in 50-ms increments. The flashbeep SJi task used SOAs of 10, 20, 50, 100, and 100–200 ms in 50-ms increments. The tool SJa, tool PF, speech SJa, and speech PF tasks used SOAs of 0–300 ms in 50-ms increments. The tool SJi, tool TOJ, speech SJi, and speech TOJ tasks used SOAs of 50–300 ms in 50-ms increments. See Fig. 1d for a graphical representation.

Participants sat inside a light- and sound-attenuating WhisperRoom™ (Model SE 2000; Whisper Room Inc). Participants were asked to fixate toward a fixation cross at all times and were monitored by close circuit infrared cameras throughout the experiment to ensure fixation. Each task began with an instruction screen, after which the participant was asked whether he or she understood the instructions. Each trial included a 501–1,500-ms fixation, stimulus presentation, a 250-ms fixation, and a response screen (in the case of the SJi task, two stimulus presentations occurred separated by 501–1,500 ms). Following a response via button press, the subsequent trial began with the 501–1,500-ms fixation. For all conditions, participants were instructed to respond as accurately as possible. Trial orders were randomly varied using the method of constant stimuli, with 20 trials per condition (in the case of the tool and speech stimuli, which both used two audiovisual clips, this included 10 presentations of each stimulus for each SOA).

Analysis

For both the SJa and the PF tasks, analyses were identical. Responses from the SJa and PF tasks were used to calculate a TBW for each subject. The first step in this process was to calculate a rate of perceived synchrony/fusion with each SOA, which was simply the percentage of trials in a given condition in which the individual reported that the presentation was synchronous/fused. As is conventional for measures of the TBW (Powers et al. 2009; Hillock et al. 2011; Hillock-Dunn and Wallace 2012; Stevenson et al. 2013), two psychometric sigmoid functions were then fit to the rates of perceived synchrony/fusion across SOAs one to the audio-first presentations and a second to the visual first presentations which will be referred to as the left and right windows, respectively. Data used to create both first-pass, best-fit sigmoids functions included the synchronous, 0 ms SOA condition. These best-fit functions were calculated using the *glmfit* function in MATLAB. Following this first-pass fit, the intersection of the left and right best-fit curve was used to estimate the point of subjective simultaneity (PSS) or the point at which the participant maximally fused the two stimuli. To account for nonzero PSS measurements, refitting of the data was then performed in an iterative fashion using the measured PSS to split the mean data points until these fittings converged. For example, if a measured PSS value was 45-ms visual leading, all data points with an SOA less than 45-ms visual leading would be used to re-calculate the left sigmoid, and all values greater than 45-ms visual leading would be used to calculate the right sigmoid. Each participant's individual left and right TBWs were then estimated at two levels relative to the individual's PSS, at the SOA at which the best-fit sigmoids *y* value equaled a 50 and 70 % rate of perceived synchrony (see Fig. 2). Each participant's

left and right TBWs were then summed to produce a third whole TBW. Group TBWs were then calculated by taking the arithmetic mean of the left, right, and whole TBWs from each participant.

For both the TOJ and the SJi, analysis was identical. Responses from the TOJ and SJi tasks were used to calculate a TBW for each subject. The first step in this process was to calculate a rate of responses with each SOA, which was simply the percentage of trials in a given condition in which the individual gave a specific response (in the case of the TOJ, that the visual was first, and in the SJi, that the second presentation was synchronous). A single psychometric sigmoid function was then fit to the response rates across all SOAs. These best-fit functions were calculated using the *glmfit* function in MATLAB. Each participant's PSS was then extracted as the SOA at which the individual's sigmoid curve was equal to 50 %. Individual left and right TBWs were then estimated at the same two levels, 50 and 70 %. The 50 % TBW was measured as the difference in ms between the SOA at which the best-fit sigmoids *y* value equaled a 25 % and the individual's PSS for the left window, and between the 75 % and the individual's PSS for the right window. The 70 % TBW was measured as the difference in ms between the SOA at which the best-fit sigmoids *y* value equaled a 35 % and the individual's PSS for the left window, and between the 65 % and the individual's PSS for the right window (see Fig. 2). Each participant's left and right TBWs were then summed to produce a third, whole TBW. Group TBWs were then calculated by taking the arithmetic mean of the left, right, and whole TBWs from each participant.

Results

Each of the 12 experimental conditions, consisting of 4 multisensory temporal tasks using three different stimulus types (i.e., simple flashbeep stimuli, complex-tool stimuli, and complex speech stimuli) were analyzed in a similar manner. Individuals' response rates were fit to a psychometric function (see methods) and averaged into a group-mean curve for each condition (Fig. 2). To assess how well these functions fit the measured data, predicted data points corresponding to the SOAs presented to participants were extracted from these psychometric functions for each individual, task, and stimulus type. An omnibus ANOVA was run comparing fitted data points and actual measured data points ($F_{(1,8997)} = 0.19$, $p = 0.67$, mean deviation = 0.003), showing a very good fit. Using these distributions, the TBW was then calculated over its entire width, and for left and right sides. These TBW were calculated at two statistical criterion levels, 50 and 70 %, to determine how strongly these criteria shape the final conclusions about the characteristics of the TBW.
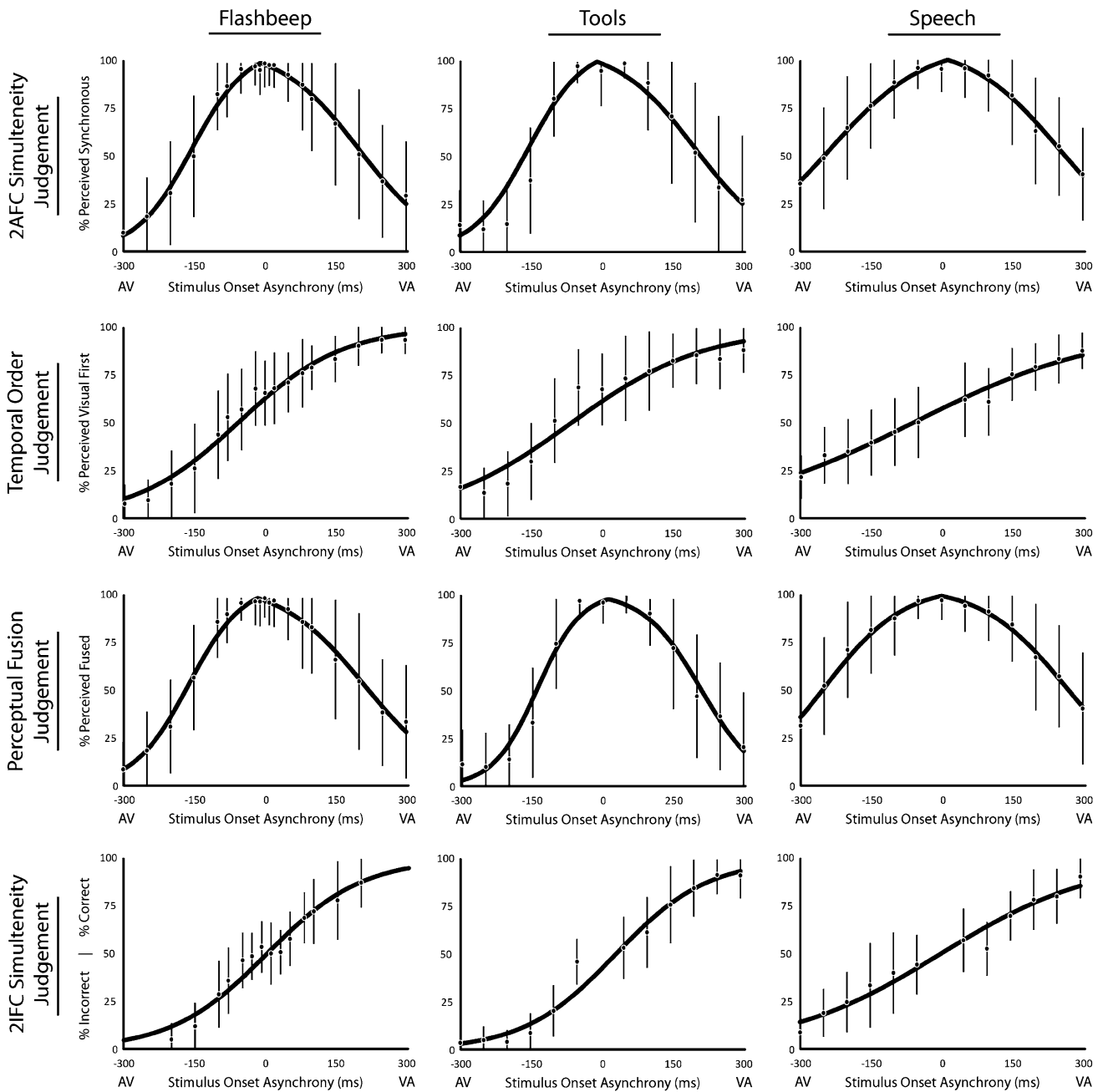
Fig. 2 Best-fit curves used to define the temporal binding window. *Sigmoid* curves of averaged data are depicted by task (*row*) and stimulus type (*column*). Data points represent the means of individuals' responses, with *error bars* representing the standard deviation between subjects

Measurements of the full TBW

The width of the full TBW was strongly dependent upon both task and stimulus type (Fig. 3). A two-way, repeated-measures ANOVA across stimulus type and task at the 50 % criterion revealed main effects of both task and stimulus type, and a significant interaction. This pattern was confirmed at the 70 % criterion. Given significant main effects across both factors, protected, follow-up *t* tests between stimulus types at the 50 % criterion showed the full TBW to be larger for speech (461 ms) compared to non-speech audiovisual stimuli, but to not differ between flashbeep (322 ms) and tool (317 ms) stimuli. Statistical tests for collapsed and pairwise comparisons of stimulus type within tasks are reported in Table 1. The full TBW was also found to be dependent upon task. Collapsed across stimulus type, follow-up *t* tests at the 50 % criterion revealed TBWs measured with the 2AFC simultaneity-judgment (SJa; 420 ms)
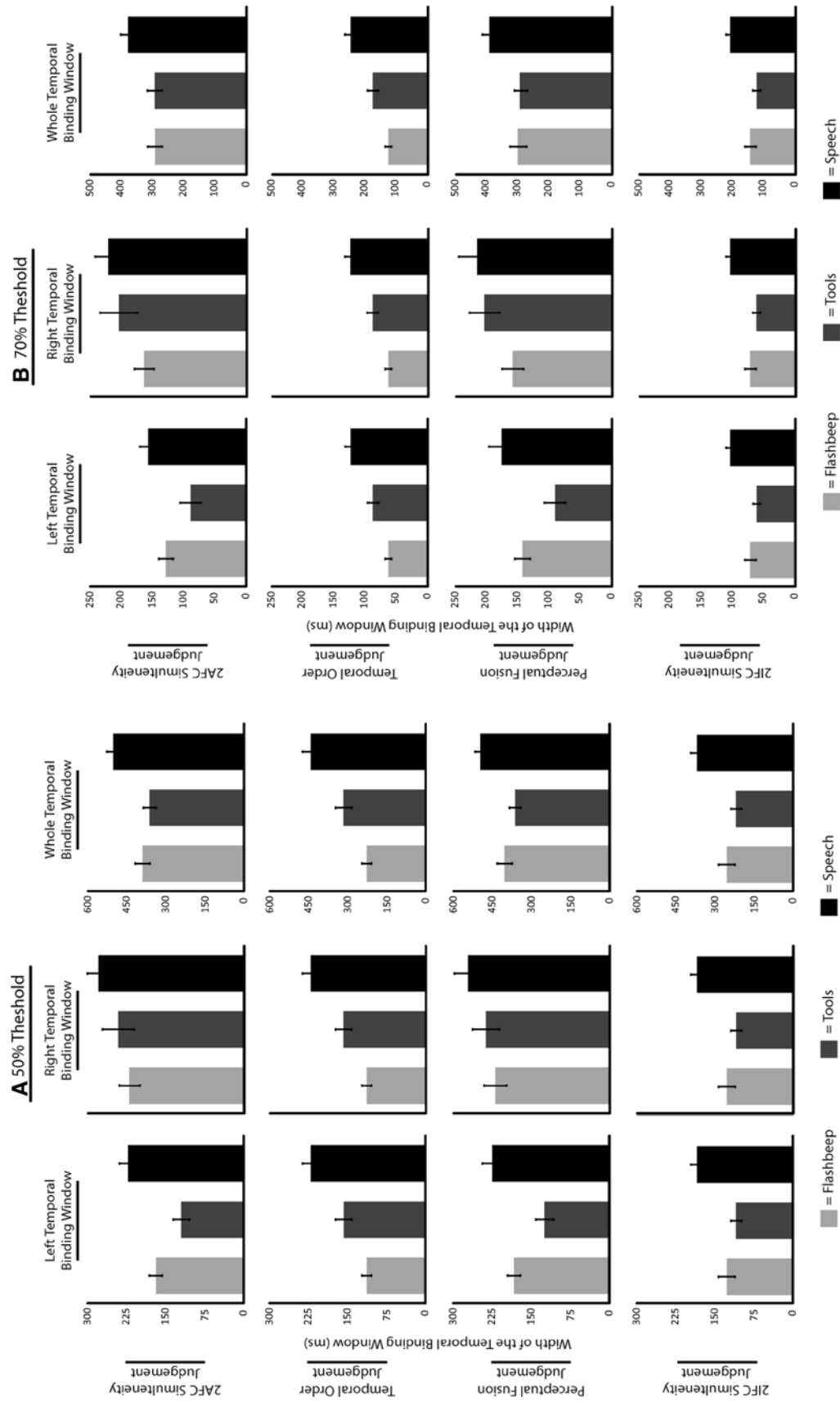
Fig. 3 Temporal binding windows. Average individual TBWs are depicted by task (*row*) and stimulus type (*column*) at the 50 % (*left*) and 70 % (*right*) criteria. *Error bars* represent standard error

Table 1 Effect of stimulus type on the temporal binding window

2-way ANOVA						
Main effect of stimulus type	Left TBW		Right TBW		Whole TBW	
	$p<$	F	$p<$	F	$p<$	F
50 % criterion	$1.00e^{-15}$	53.74	$1.00e^{-13}$	40.72	$1.00e^{-15}$	53.68
70 % criterion	$1.00e^{-13}$	39.14	0.02	4.22	$1.00e^{-8}$	31.72

Follow-up t tests							
Task comparison	Stimulus	Left TBW		Right TBW		Whole TBW	
		$p<$	T	$p<$	t	$p<$	T
Flashbeep versus tools	Collapsed	n.s.	1.89	n.s.	1.91	n.s.	0.13
	SJa	n.s.	1.65	n.s.	1.40	n.s.	0.05
	TOJ	n.s.	1.98	n.s.	2.00	n.s.	1.99
	PF	0.03	2.31	n.s.	1.18	n.s.	0.75
	SJi	n.s.	1.91	n.s.	1.91	n.s.	1.91
Flashbeep versus speech	Collapsed	$5.71e^{-12}$	7.74	$1.29e^{-3}$	3.31	$4.92e^{-17}$	9.99
	SJa	$3.39e^{-3}$	3.01	n.s.	0.34	$9.07e^{-7}$	5.26
	TOJ	$1.30e^{-3}$	3.59	$1.33e^{-3}$	3.58	$1.32e^{-3}$	3.58
	PF	$1.38e^{-4}$	4.54	n.s.	0.34	$7.37e^{-5}$	4.70
	SJi	$4.56e^{-7}$	7.04	$4.47e^{-7}$	7.05	$4.52e^{-7}$	7.05
Tools versus speech	Collapsed	$7.35e^{-9}$	6.28	$2.50e^{-6}$	4.97	$1.46e^{-13}$	8.46
	SJa	0.02	2.44	$9.91e^{-3}$	2.63	$2.01e^{-5}$	4.49
	TOJ	$3.38e^{-5}$	4.96	$3.35e^{-5}$	4.97	$3.36e^{-5}$	4.96
	PF	0.04	2.23	n.s.	0.80	$6.06e^{-3}$	2.99
	SJi	$1.23e^{-4}$	4.65	$1.21e^{-4}$	4.66	$1.22e^{-4}$	4.66

and perceptual-fusion (PF; 425 ms) tasks to be wider than with both temporal-order judgment (TOJ; 313 ms) and 2IFC simultaneity-judgment (SJi; 280 ms) tasks, but no difference was seen between the SJa and PF tasks or the TOJ and SJi tasks. Statistical tests for collapsed and pairwise comparisons of task within stimulus type are reported in Table 2.

Measurements of the left and right TBW

In addition to looking at full window width, analyses focused on both the left (auditory leading) and right (visual leading) sides of the TBW. The rationale for this division was the fact that in the natural environment, the conduction speed of light and sound energy inherently bias the order in which visual and auditory cues arrive at their respective sense organs, such that light energy arriving at the retina always precedes sound energy arriving at the cochlea. Left and right TBW analysis began with a three-way, repeated-measures ANOVA across side (left/right), task, and stimulus type at the 50 % criterion. This analysis revealed main effects of side, task, and stimulus type, as well as a side-by-task interaction. Main effects and interactions were confirmed using the 70 % criterion, with a single difference of no main effect of side (see Fig. 3). See Table 3 for associated statistics.

Given the main effects of window side, subsequent two-way, repeated-measures ANOVAs were run on the left and right TBW data across stimulus type and task. Focusing first on the left TBW, a main effect of stimulus type was seen at both the 50 and 70 % criteria. Given this significant main effect protected, follow-up t tests at the 50 % criterion revealed that the left TBW as measured using speech stimuli (215 ms) was wider than that measured using flashbeep (150 ms) and tool (129 ms) stimuli. In contrast, no difference between flashbeep and tool stimuli was seen. Statistical tests for collapsed and pairwise comparisons of stimulus type within tasks are reported in Table 1. In addition, a main effect of task was seen at both the 50 and 70 % criteria. Similar protected, follow-up t tests across task and collapsed across stimulus types revealed that the left TBW as determined using the SJi task (140 ms) was narrower than that for the SJa (173 ms) and PF (181 ms) tasks, and the TBW with the TOJ task (156 ms) was narrower than with the PF task. Statistical tests for collapsed and pairwise comparisons of task within stimulus type are reported in Table 1. No stimulus-type-by-task interaction was observed.

For the right TBW at the 50 % criterion, a main effect of stimulus type was seen. Follow-up t tests revealed that the right TBW as measured using speech stimuli was wider (246 ms) than for flashbeep (173 ms) and tool (188 ms)

Table 2 Effect of task on the temporal binding window

2-way ANOVA main effect of task	Left TBW		Right TBW		Whole TBW	
	$p<$	F	$p<$	F	$p<$	F
50 % criterion	$1.00e^{-5}$	14.05	$1.00e^{-15}$	50.62	$1.00e^{-13}$	34.66
70 % criterion	$1.00e^{-5}$	18.79	$1.00e^{-13}$	44.03	$1.00e^{-8}$	32.76

Follow-up *t* tests

Task comparison	Stimulus	Left TBW		Right TBW		Whole TBW	
		$p<$	T	$p<$	t	$p<$	T
SJa versus TOJ	Collapsed	n.s.	0.28	$5.29e^{-6}$	4.85	$1.93e^{-4}$	3.89
	Flashbeep	0.03	2.23	$1.80e^{-4}$	3.91	$5.39e^{-4}$	3.59
	Tools	n.s.	1.12	$2.59e^{-3}$	3.32	0.04	2.29
	Speech	n.s.	0.06	n.s.	0.96	n.s.	0.46
SJa versus PF	Collapsed	n.s.	0.67	n.s.	0.80	n.s.	0.22
	Flashbeep	n.s.	0.87	n.s.	0.42	n.s.	0.83
	Tools	n.s.	1.10	n.s.	0.35	n.s.	0.29
	Speech	n.s.	0.01	n.s.	1.09	n.s.	1.62
SJa versus SJi	Collapsed	$5.57e^{-3}$	2.84	$5.12e^{-7}$	5.41	$1.45e^{-9}$	6.75
	Flashbeep	$8.89e^{-3}$	2.67	$1.89e^{-4}$	3.90	$2.50e^{-4}$	3.81
	Tools	n.s.	0.34	$8.21e^{-5}$	4.63	$3.05e^{-5}$	5.86
	Speech	n.s.	1.77	n.s.	1.46	0.04	2.20
TOJ versus PF	Collapsed	0.03	2.32	$1.08e^{-4}$	4.05	$2.86e^{-5}$	4.42
	Flashbeep	$1.89e^{-4}$	3.90	$1.25e^{-3}$	3.34	$9.12e^{-5}$	4.10
	Tools	n.s.	0.38	$7.12e^{-4}$	3.82	$9.32e^{-4}$	3.72
	Speech	n.s.	0.55	n.s.	0.01	n.s.	0.29
TOJ versus SJi	Collapsed	n.s.	1.94	n.s.	1.97	n.s.	1.95
	Flashbeep	n.s.	0.88	n.s.	0.87	n.s.	0.88
	Tools	0.05	2.10	0.05	2.14	0.05	2.12
	Speech	n.s.	1.52	n.s.	1.54	n.s.	1.53
PF versus SJi	Collapsed	$4.76e^{-5}$	2.84	$2.65e^{-6}$	5.42	$8.38e^{-10}$	6.75
	Flashbeep	$2.86e^{-3}$	2.67	0.02	3.90	$2.75e^{-3}$	3.82
	Tools	n.s.	0.34	$1.79e^{-4}$	4.63	$9.15e^{-6}$	5.86
	Speech	$4.42e^{-3}$	1.77	n.s.	1.46	$1.30e^{-3}$	2.20

stimuli, with no differences being found between flashbeep and tool stimuli. Statistical tests for collapsed and pairwise comparisons of stimulus type within tasks are reported in Table 1. A main effect of task was also seen at the 50 % criterion. The right TBW with the SJa (248 ms) and PF (244 ms) tasks were wider than both the TOJ (156 ms) and SJi (140 ms) tasks, with no significant difference between either the SJa and PF or the TOJ and SJi tasks. These effects of both stimulus type and task were comparable at the 70 % criterion. Statistical tests for collapsed and pairwise comparison of stimulus type within task are reported in Table 2.

To explore the significant interaction between side and stimulus type revealed by the three-way ANOVA, protected, follow-up *t* tests were run comparing the left and right TBWs across stimulus type collapsed across task at the 50 % criterion. Differences were found between the left and right TBW for the flashbeep, tool, and speech stimuli

(mean difference = 23, 59, and 32 ms, respectively). This pattern was similar for the 70 % criterion. Thus, the side-by-stimulus type interaction was driven by a difference between speech and non-speech stimuli, where non-speech stimuli are associated with a more asymmetrical TBW when compared with speech stimuli. Statistical tests for collapsed and pairwise comparisons are reported in Table 3.

Within-subject correlations

In order to determine how tightly coupled these various indices of multisensory temporal binding are within individuals, the consistency of the TBWs as measured using the different stimuli and tasks was assessed through pairwise correlations between each task-stimulus condition. Full, left, and right TBWs were correlated within individuals at both the 50 and 70 % thresholds (Fig. 4). At the 70 % criterion, the

Table 3 Differences between the left and right TBW

3-Way ANOVA	Effect	$p<$	F
50 % criterion	Side	n.s.	0.97
	Stimulus type	$1.00e^{-15}$	40.72
	Task	$1.00e^{-15}$	50.62
	Side × Stim. type	0.003	6.49
	Side × task	$1.00e^{-5}$	23.97
70 % criterion	Side	$1.00e^{-20}$	90.12
	Stimulus type	$2.78e^{-8}$	58.96
	Task	$2.14e^{-11}$	29.84
	Side × stim. type	0.11	2.24
	Side × task	$1.00e^{-15}$	63.70

Follow-up *t* tests			
Stimulus type	Task	$p<$	T
Flashbeep	Collapsed	$1.31e^{-3}$	3.30
	SJa	$4.60e^{-3}$	2.90
	TOJ	n.s.	0.57
	PF	n.s.	1.81
	SJi	n.s.	0.83
Tools	Collapsed	$1.77e^{-4}$	3.88
	SJa	$7.84e^{-3}$	2.72
	TOJ	0.03	2.30
	PF	$5.10e^{-3}$	3.06
	SJi	n.s.	0.77
Speech	Collapsed	0.05	2.03
	SJa	n.s.	1.93
	TOJ	n.s.	0.62
	PF	n.s.	1.11
	SJi	n.s.	0.33

strongest correlations were seen for the full TBW (average $R = 0.58$) and the right TBW (average $R = 0.23$). Also of note is that the correlations were consistently stronger for the right TBW when compared with the left TBW (average $R = 0.18$). A very similar pattern was seen at the 50 % criterion, with average R values of 0.39, 0.28, and 0.24 for the full, right, and left TBWs, respectively. A review of the patterns revealed by these correlational analyses revealed that the SJ tasks were the most correlated with other tasks, whereas the TOJ task was the least correlated with other tasks (van Eijk et al. 2008). For statistical results with each individual correlation, see Fig. 4.

Discussion

In studies of multisensory interactions, the temporal relationship between paired sensory inputs has proven to be one of the most important features in determining if these inputs will be integrated, what form the interaction will take (i.e., enhancements vs. depressions) and the magnitude or strength of the resultant interaction (Meredith et al. 1987; Stevenson et al. 2010; Vroomen and Keetels 2010). It has been well established that sensory inputs do not have to be simultaneously presented in order to be integrated, but rather can occur within a limited range of temporal offsets, a construct known as the multisensory temporal binding window (TBW). Given the great impact temporal processing has on multisensory integration and the emerging evidence for disrupted multisensory temporal processing in clinical conditions (Hairston et al. 2005; Foss-Feig et al. 2009; Kwakye et al. 2011), substantial efforts have been put into characterizing the TBW. However, these studies have utilized a variety of tasks, stimulus types, and statistical criteria, making comparisons across experiments difficult. The current study represents the first to our knowledge designed to parametrically assess the impact of these factors on the TBW within the same set of experimental subjects. Collectively, these data show that both stimulus type and task strongly affect the TBW, with the window as assessed with speech stimuli being both larger and more symmetrical than that defined using simple and complex non-speech stimuli. In addition, these effects were found to be robust across statistical criteria and to be highly consistent within individuals, suggesting substantial overlap in the neural and cognitive operations that govern multisensory temporal processes.

One of the most salient features of the current data was the asymmetry between the left and right TBW when assessed using non-speech stimuli, a finding concordant with prior work (Dixon and Spitz 1980; Miller and D'Esposito 2005; Zampini et al. 2005; Conrey and Pisoni 2006). The asymmetry was driven by the fact that individuals are much more likely to perceive trials in which the visual stimulus precedes the auditory stimulus as synchronous when compared with the converse. This finding has a great deal of ecological validity, since, in the natural environment, visual stimulus energy will always reach the retina prior to auditory energy reaching the cochlea (Pöppel et al. 1990). Thus, if the TBW is shaped by the normal temporal statistics of stimuli within the environment, one would expect that the shape and symmetry of the TBW would reflect the greater likelihood of there being meaningful stimulus pairings in which the visual stimulus precedes the auditory stimulus. Interestingly, and novel to this study, is the finding that the criterion which is used to define the TBW may impact this measure of asymmetry.

Another of the strongest features of the current data was how the TBW was impacted by stimulus type, an effect driven largely by the differences between the speech and non-speech audiovisual stimuli. The finding that the TBW is wider for speech-related stimuli fits well with two previous lines of work studying multisensory temporal processing.
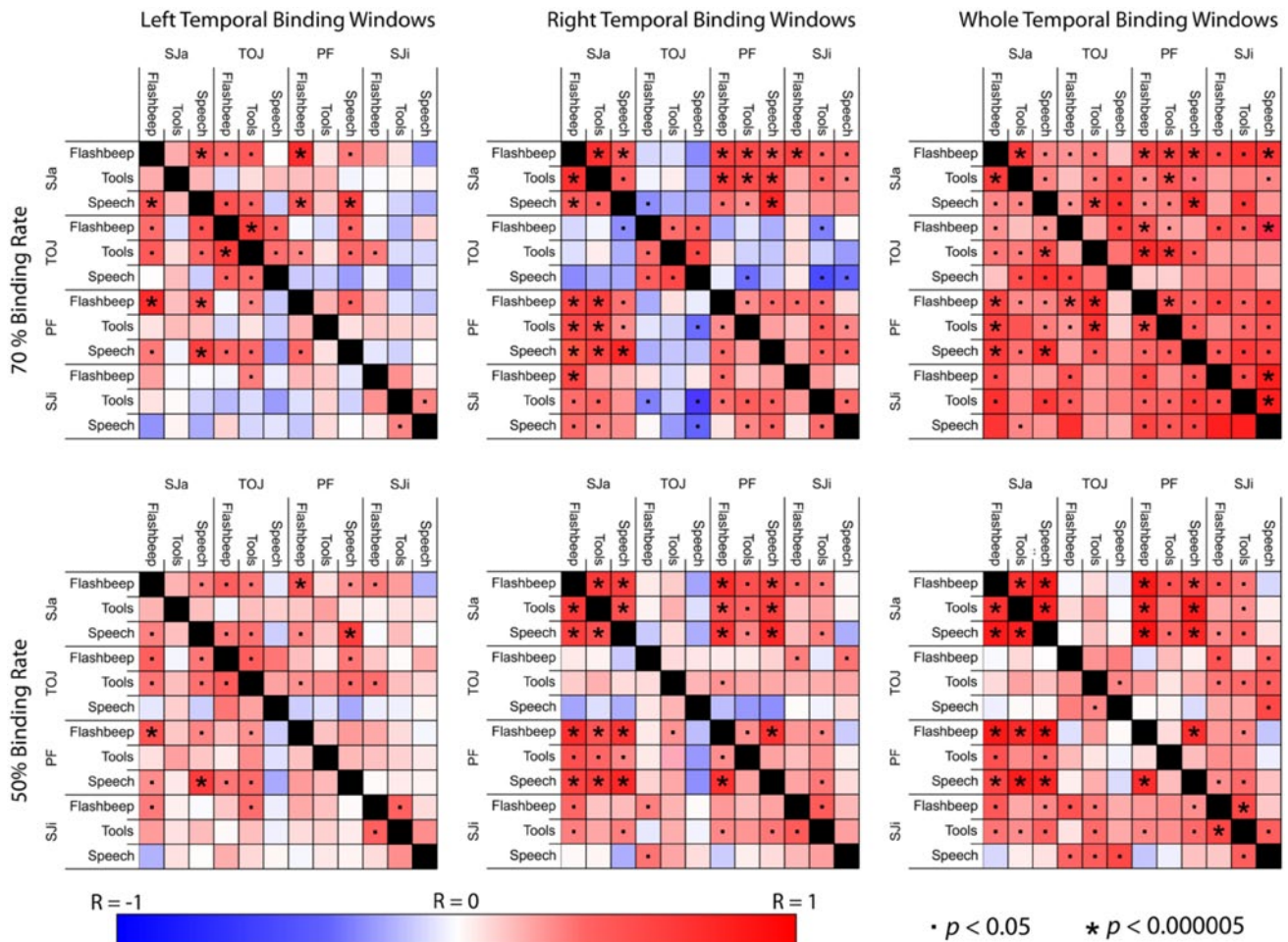
Fig. 4 Within-subject correlation matrices. Correlations between each stimulus/task pairing is depicted for the 50 % (*top*) and 70 % (*bottom*) thresholds. Correlations are shown for the *left* (VA), *right* (AV), and whole TBWs

First, while the TBW has been shown to narrow substantially with development (Hillock et al. 2011), there appears to be evidence for differential narrowing dependent upon the nature of the stimuli used (Hillock-Dunn and Wallace 2012). Second, in its mature state, the TBW has previously been shown to be wider with complex stimuli than with simple stimuli regardless of stimulus length (Vatakis and Spence 2006). A number of plausible explanations could explain the differential width of the TBW based on stimulus type. The most compelling of these arguments claims that certain stimuli, particularly speech stimuli, may require additional processing *within* modality prior to, or concurrent with, integration *across* modalities, and thus, the multisensory process may be optimized if there is more tolerance to temporal offsets.

It should also be noted here that each stimulus type varied in total length; however, previous work has shown that stimulus length itself may not significantly impact the perception of synchrony (Vatakis and Spence 2006). Differences in

the length of the individual auditory and visual components have been shown to have an effect on the PSS such that when these components differ in duration (Kuling et al. 2012). However, duration effects were not seen when the durations of the auditory and visual components were matched, as they were in the current study. However, the Kuling et al. study replicated a previous study showing that the inter-subject variability of the PSS decreased with increased absolute duration (Boenke et al. 2009). Such a change in variability of the measure of interest here, the TBW, may also be thus impacted by stimulus duration, a possibility that should be explored in future research. While these previous studies found no net impact of absolute stimulus duration on the mean PSS, it should be noted that in these two studies, maximum durations were 500 ms (Boenke et al. 2009) and 294 ms (Kuling et al. 2012), whereas the stimuli in the current study ranged from 10 to 2,000 ms. As such, it cannot be said with certainty whether such a duration effect may only be seen with stimuli exceeding 1 s.

Also difference between the simple flashbeep stimuli and both the tool and speech stimuli is the fact that the temporal offsets of the simplest stimuli used here are such that the offset of the first stimulus component precedes the onset of the second stimulus component, creating a gap between the auditory and visual stimuli, while the auditory and visual components of the tool and speech stimuli overlapped even in the conditions with the greatest SOAs. With that said, differences between stimulus types due to this difference would appear between the flashbeep and both the tools and speech stimuli; however, this was not the case. In fact, the lack of difference between the TBWs measured with the flashbeep and tool stimuli, coupled with the significant differences between the TBW measured with the speech stimuli and both non-speech stimulus types, suggests that the presence of a stimulus gap in the flashbeep stimuli did not drive the effects observed here.

In addition to the effects of stimulus type, the task used to measure the TBW was also found to be an important factor in the size of the window. Thus, the TOJ and the SJi tasks produced reliably narrower windows than the SJa and PF tasks. Such a result implies that the perceptual and/or cognitive operations in the latter pair of tasks take a greater time to reach completion, suggesting additional processing steps following the lower-level analyses of temporal relationships. In the case of the perceptual-fusion task, this makes a great deal of sense, since subjects are asked to report on a judgment not of the relative timing of the paired stimuli, but rather of their likelihood of being derived from the same event, a judgment likely to invoke comparison with higher order perceptual representations.

One important consideration in all studies of the TBW is the threshold or criterion chosen to define the window. Unfortunately, no convention exists across studies. The construct of the TBW and its usefulness in describing the cognitive phenomenon of perceptual binding in the absence of objectively simultaneous sensory inputs is dependent upon the reliability of the measure across these somewhat arbitrary analytical choices. To gain a better view into how changes in these criteria affect the TBW, the current study chose two specific statistical criterion levels, 50 and 70 %, that are similar to the most commonly used values in prior work. In large measure, the results across these two different thresholds were surprisingly concordant, with the one exception being that there was a differential effect of window asymmetry across criteria. Nonetheless, the caution remains in regard to comparisons across studies in which thresholds differ, strongly limiting any statements that can be made regarding the absolute size of the TBW.

Perhaps of greatest interest here are the strong within-subject correlations. This finding suggests that while each task undoubtedly taps into distinct cognitive operations (Hirsh and Sherrick 1961; van Eijk et al. 2008; Vatakis et al.

2008), they are each likely to share a common underlying process responsible for the ascription of temporal identity at a stimulus level. All tasks, with the exception of the TOJ task, showed strong correlations in the ecologically relevant right side of the TBWs, where vision precedes audition. All tasks including the TOJ task showed strong correlations in the whole TBW. This difference between the right and whole TBW with the TOJ task was the product of a shifted PSS, a shift previously discussed by van Eijk et al. (2008). These results provide evidence that while the PSS and TBW as measurements may be differentially impacted by task and stimulus type as previously suggested (Keetels and Vroomen 2005), all four tasks show a striking correlation in their elicitation of the TBW (see Fig. 4). Emerging evidence points to the site for this multisensory temporal calculation to be cortical associative areas within the superior temporal sulcus (STS). Neuroimaging studies have shown the STS to be differentially responsive based on the temporal alignment of audiovisual stimuli across a wide range of tasks and stimulus types (Macaluso et al. 2004; Miller and D'Esposito 2005; Stevenson and James 2009; Stevenson et al. 2010, 2011), necessary for audiovisual perceptual fusion (Beauchamp et al. 2010; Nath and Beauchamp 2011; Stevenson et al. 2011), and dysfunctional in clinical populations showing impaired multisensory temporal integration (Zilbovicius et al. 2006; Pelphrey and Carter 2008).

In conclusion, the data reported here provide the first systematic analysis of the multisensory temporal binding window across stimulus types, cognitive task, and threshold level. These findings provide evidence for both stimulus- and task-related effects on the TBW, as well as significant asymmetries between the right and left sides of the window. Perhaps, most importantly, these data show that the construct of the TBW shows a remarkable reliability within subjects, suggesting a common underlying neural architecture for the temporal binding of multisensory stimuli highly likely to have originated from the same event.

References

Beauchamp MS, Nath AR, Pasalar S (2010) fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. J Neurosci 30:2414–2417. doi:10.1523/JNEUROSCI.4865-09.2010

Boenke LT, Deliano M, Ohl FW (2009) Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. Exp Brain Res 198:233–244. doi:10.1007/s00221-009-1917-z

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436

Conrey B, Pisoni DB (2006) Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. J Acoust Soc Am 119:4065–4073

de Boer-Schellekens L, Eussen M, Vroomen J (2013) Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. Front Integr Neurosci 7:8. doi:10.3389/fnint.2013.00008

Diederich A, Colonius H (2004) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. Percept Psychophys 66:1388–1404

Dixon NF, Spitz L (1980) The detection of auditory visual desynchrony. Perception 9:719–721

Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT (2009) An extended multisensory temporal binding window in autism spectrum disorders. Exp Brain Res. doi:10.1007/s00221-010-2240-4

Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT (2005) Altered temporal profile of visual-auditory multisensory interactions in dyslexia. Exp Brain Res 166:474–480. doi:10.1007/s00221-005-2387-6

Hershenson M (1962) Reaction time as a measure of intersensory facilitation. J Exp Psychol 63:289–293

Hillock AR, Powers AR, Wallace MT (2011) Binding of sights and sounds: age-related changes in multisensory temporal processing. Neuropsychologia 49:461–467. doi:10.1016/j.neuropsychologia.2010.11.041

Hillock-Dunn A, Wallace MT (2012) Developmental changes in the multisensory temporal binding window persist into adolescence. Dev Sci 15:688–696. doi:10.1111/j.1467-7687.2012.01171.x

Hirsh IJ, Sherrick CE Jr (1961) Perceived order in different sense modalities. J Exp Psychol 62:423–432

James TW, Stevenson RA, Kim S (2009) Assessing multisensory integration with additive factors and functional MRI. The International Society for Psychophysics, Dublin

James TW, VanDerKlok RM, Stevenson RA, James KH (2011) Multisensory perception of action in posterior temporal and parietal cortices. Neuropsychologia 49:108–114. doi:10.1016/j.neuropsychologia.2010.10.030

James TW, Stevenson RA, Kim S (2012) Inverse effectiveness in multisensory processing. In: Stein BE (ed) The new handbook of multisensory processes. MIT Press, Cambridge, MA

Keetels M, Vroomen J (2005) The role of spatial disparity and hemifields in audio-visual temporal order judgments. Exp Brain Res 167:635–640. doi:10.1007/s00221-005-0067-1

Kuling IA, van Eijk RL, Juola JF, Kohlrausch A (2012) Effects of stimulus duration on audio-visual synchrony perception. Exp Brain Res 221:403–412. doi:10.1007/s00221-012-3182-9

Kwakye LD, Foss-Feig JH, Cascio CJ, Stone WL, Wallace MT (2011) Altered auditory and multisensory temporal processing in autism spectrum disorders. Front Integr Neurosci 4:129. doi:10.3389/fnint.2010.00129

Lovelace CT, Stein BE, Wallace MT (2003) An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. Brain Res Cogn Brain Res 17:447–453. doi:10.1016/S0926-6410(03)00160-5

Macaluso E, George N, Dolan R, Spence C, Driver J (2004) Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage 21:725–732

Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci 7:3215–3229

Meredith MA, Wallace MT, Stein BE (1992) Visual, auditory and somatosensory convergence in output neurons of the cat superior colliculus: multisensory properties of the tecto-reticulo-spinal projection. Exp Brain Res 88:181–186

Miller LM, D'Esposito M (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci 25:5884–5893

Nath AR, Beauchamp MS (2011) A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage. doi:10.1016/j.neuroimage.2011.07.024

Nelson WT, Hettinger LJ, Cunningham JA, Brickman BJ, Haas MW, McKinley RL (1998) Effects of localized auditory information on visual target detection performance using a helmet-mounted display. Hum Factors 40:452–460

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442

Pelphrey KA, Carter EJ (2008) Charting the typical and atypical development of the social brain. Dev Psychopathol 20:1081–1102. doi:10.1017/S0954579408000515

Pöppel E, Schill K, von Steinbüchel N (1990) Sensory integration within temporally neutral systems states: a hypothesis. Naturwissenschaften 77:89–91

Powers AR 3rd, Hillock AR, Wallace MT (2009) Perceptual training narrows the temporal window of multisensory binding. J Neurosci 29:12265–12274. doi:10.1523/JNEUROSCI.3501-09.2009

Quinto L, Thompson WF, Russo FA, Trehub SE (2010) A comparison of the McGurk effect for spoken and sung syllables. Atten Percept Psychophys 72:1450–1454. doi:10.3758/APP.72.6.1450

Royal DW, Carriere BN, Wallace MT (2009) Spatiotemporal architecture of cortical receptive fields and its impact on multisensory interactions. Exp Brain Res 198:127–136. doi:10.1007/s00221-009-1772-y

Schall JD, Hanes DP (1993) Neural basis of saccade target selection in frontal eye field during visual search. Nature 366:467–469. doi:10.1038/366467a0

Schall S, Quigley C, Onat S, Konig P (2009) Visual stimulus locking of EEG is modulated by temporal congruency of auditory stimuli. Exp Brain Res 198:137–151. doi:10.1007/s00221-009-1867-5

Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG (2007) Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia 45:561–571

Stein BE, Wallace MT (1996) Comparisons of cross-modality integration in midbrain and cortex. Prog Brain Res 112:289–299

Stevenson RA, James TW (2009) Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. Neuroimage 44:1210–1223. doi:10.1016/j.neuroimage.2008.09.034

Stevenson RA, Geoghegan ML, James TW (2007) Superadditive BOLD activation in superior temporal sulcus with threshold nonspeech objects. Exp Brain Res 179:85–95

Stevenson RA, Kim S, James TW (2009) An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. Exp Brain Res 198:183–194. doi:10.1007/s00221-009-1783-8

Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW (2010) Neural processing of asynchronous audiovisual speech perception. Neuroimage 49:3308–3318. doi:10.1016/j.neuroimage.2009.12.001

Stevenson RA, VanDerKlok RM, Pisoni DB, James TW (2011) Discrete neural substrates underlie complementary audiovisual speech integration processes. Neuroimage 55:1339–1345. doi:10.1016/j.neuroimage.2010.12.063

Stevenson RA, Wilson MM, Powers AR, Wallace MT (2013) The effects of visual training on multisensory temporal processing. Exp Brain Res. doi:10.1007/s00221-012-3387-y

van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb Cortex 17:962–974. doi:10.1093/cercor/bhl007

van Eijk RL, Kohlrausch A, Juola JF, van de Par S (2008) Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type. Percept Psychophys 70:955–968

van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in auditory-visual speech perception. Neuropsychologia 45:598–607. doi:10.1016/j.neuropsychologia.2006.01.001

Vatakis A, Spence C (2006) Audiovisual synchrony perception for music, speech, and object actions. Brain Res 1111:134–142. doi:10.1016/j.brainres.2006.05.078

Vatakis A, Navarra J, Soto-Faraco S, Spence C (2008) Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. Exp Brain Res 185:521–529. doi:10.1007/s00221-007-1168-9

Vroomen J, Keetels M (2010) Perception of intersensory synchrony: a tutorial review. Atten Percept Psychophys 72:871–884. doi:10.3758/APP.72.4.871

Wallace MH, Murray MM (eds) (2011) Frontiers in the neural basis of multisensory processes. Taylor & Francis, London

Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA (2004) Unifying multisensory signals across time and space. Exp Brain Res 158:252–258. doi:10.1007/s00221-004-1899-9

Wilkinson LK, Meredith MA, Stein BE (1996) The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. Exp Brain Res 112:1–10

Woynaroski TG, Kwakye LD, Foss-Feig JH, Stevenson RA, Stone WL, Wallace MT (2013) Multisensory speech perception in high-functioning children with autism spectrum disorders. J Autism Dev Disord. doi:JADD-D-12-00456R1

Zampini M, Guest S, Shore DI, Spence C (2005) Audio-visual simultaneity judgments. Percept Psychophys 67:531–544

Zilbovicius M, Meresse I, Chabane N, Brunelle F, Samson Y, Boddaert N (2006) Autism, the superior temporal sulcus and social perception. Trends Neurosci 29:359–366. doi:10.1016/j.tins.2006.06.004