The effects of visual training on multisensory temporal processing

Ryan A. Stevenson[*,1,3], **Magdalena M. Wilson**[*,4], **Albert R. Powers**[5], and **Mark T. Wallace**[1,3,6,7]

[1]Vanderbilt University Medical Center, Dept. Hearing and Speech Sciences

[2]Vanderbilt Brain Institute

[3]Vanderbilt University Kennedy Center

[4]Vanderbilt University, Undergraduate Program in Neuroscience

[5]Vanderbilt University Medical Center, Medical Scientist Training Program

[6]Vanderbilt University Dept. of Psychology

[7]Vanderbilt University Medical Center, Dept. of Psychiatry

Abstract

The importance of multisensory integration for human behavior and perception is well documented, as is the impact that temporal synchrony has on driving such integration. Thus, the more temporally coincident two sensory inputs from different modalities are, the more likely they will be perceptually bound. This temporal integration process is captured by the construct of the temporal binding window - the range of temporal offsets within which an individual is able to perceptually bind inputs across sensory modalities. Recent work has shown that this window is malleable, and can be narrowed via a multisensory perceptual feedback training process. In the current study, we seek to extend this by examining the malleability of the multisensory temporal binding window through changes in *unisensory* experience. Specifically, we measured the ability of visual perceptual feedback training to induce changes in the multisensory temporal binding window. Visual perceptual training with feedback successfully improved temporal visual processing and more importantly, this visual training increased the temporal precision across modalities, which manifested as a narrowing of the multisensory temporal binding window. These results are the first to establish the ability of unisensory temporal training to modulate multisensory temporal processes, findings that can provide mechanistic insights into multisensory integration and which may have a host of practical applications.

Keywords

Multisensory integration; Perceptual training; Audiovisual; Vision; Temporal processing; Plasticity; Synchrony

Introduction

The integration of information from the different senses is critical to our ability to interact with the environment. While individual sensory systems transduce distinct forms of stimulus energy, the information extracted is often bound into a single unified percept of the external event, also known as a "unity assumption," where the observer assumes that two sensory inputs originate from a single event as opposed to two separate events (for review, see Spence 2007). Such multisensory integration results in improvements in detection (Stein and Wallace 1996; Lovelace et al. 2003) and localization (Wilkinson et al. 1996; Nelson et al. 1998) as well as response and reaction times (Hershenson 1962; Diederich and Colonius 2004). A number of factors impact the integration of stimuli across the different senses. These include the spatial relationships of the paired stimuli (Meredith and Stein 1986a; Macaluso et al. 2004), their relative efficacy (Meredith and Stein 1986b; Stevenson and James 2009; James et al. 2012), and most germane to the current study, their temporal relationship (Meredith et al. 1987; Miller and D'Esposito 2005; Vatakis and Spence 2007). For this final circumstance, the more temporally coincident the stimuli, the larger the multisensory interactions generated by their pairing. This finding is robust and has been replicated across a number of experimental levels of analysis, including in single neurons (Meredith et al. 1987; Stein and Meredith 1993; Wallace et al. 2004; Royal et al. 2009), in neuronal ensembles revealed by electroencephalography (EEG; Senkowski et al. 2007; Schall et al. 2009; Talsma et al. 2009) and functional magnetic resonance imaging (Macaluso et al. 2004; Miller and D'Esposito 2005; Stevenson et al. 2010; Stevenson et al. 2011), and in psychophysical studies of behavior and perception (Dixon and Spitz 1980; Conrey and Pisoni 2004; 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; Foss-Feig et al. 2010; Hillock et al. 2011; Stevenson et al. In Press).

As alluded to above, while temporal synchrony generally results in a high degree of multisensory integration, paired stimuli need not be precisely synchronous in order to generate significant neural, behavioral and perceptual changes. Multisensory stimuli presented in close temporal proximity are often still integrated into a single, unified percept (McGurk and MacDonald 1976; Stein and Meredith 1993; Shams et al. 2000; Andersen et al. 2004; Vatakis and Spence 2007). Not surprisingly, such perceptual binding only occurs over a limited temporal interval, a process that is best captured in the construct of a multisensory temporal binding window (TBW; Colonius and Diederich 2004; Hairston et al. 2005). The TBW is a probabilistic concept, reflecting the likelihood that two stimuli from different modalities will be perceptually bound across a range of stimulus asynchronies. Although numerous paradigms have been used to study the TBW, one of the most established of these is a simultaneity judgment (SJ) task in which participants report on the perceived simultaneity (i.e., synchrony) of paired stimuli (Seitz et al. 2006b; Vatakis and Spence 2006; Vatakis et al. 2008; Roach et al. 2011). Using this method, a response distribution is created that plots perceived simultaneity as a function of the stimulus onset asynchrony (SOA) between the paired stimuli, a distribution that has previously been used to define the multisensory TBW (E.g. Hairston et al. 2005; Powers et al. 2009; Stevenson et al. 2012). It should be noted here that the perception of synchrony does not necessarily mean that the two inputs are bound. However, previous studies using identical stimulus sets have shown that the perception of synchrony and perceptual binding are closely related (Stevenson et al. 2010; Stevenson et al. 2011; Stevenson and Wallace Under Review). Using these and related measures, a number of investigators have shown the window to span several hundred milliseconds (for review see van Eijk et al. 2008).

Recently, work has focused not only on defining the temporal window within which multisensory stimuli are bound, but also in exploring the malleability of multisensory

temporal processes. Several of these studies have shown that the response distribution is plastic to temporal recalibration, a process by which the peak of the TBW can be shifted towards an auditory- or visual-leading presentation through intensive exposure to temporally-asynchronous pairs (Fujisaki et al. 2004; Vroomen et al. 2004; Vroomen and Baart 2009) and through online recalibration (Navarra et al. 2005; Vatakis et al. 2007), thus changing the point of subjective simultaneity. More germane to the current study is the work showing that the response distribution which serves to create the proxy measure for the TBW can be substantially and significantly narrowed via perceptual feedback training (Powers et al. 2009). While this work illustrates the plasticity inherent in the TBW, it remains unclear whether changes in multisensory experience are a necessary component for the changes that can be induced, or whether these effects can be similarly driven through changes in the temporal processing of the individual unisensory components.

In the current study, we attempted to address this issue by exploring the malleability of the multisensory TBW following unisensory (i.e., visual) perceptual feedback training. The multisensory TBW was measured before and after visual-only training with the goal of increasing individuals' ability to detect asynchronies between two visual stimuli. We hypothesized that this increased precision in visual temporal processing would translate into upstream changes in audiovisual temporal processing, specifically in a narrowing of the TBW. These findings are discussed in terms of the novel insights for our understanding of multisensory integration that can be gleaned from the transference of unisensory training into multisensory perceptual change, as well as the possible practical and clinical applications of such training methods in altering multisensory performance.

Materials and methods

Participants

Data was collected and analyzed from 22 Vanderbilt undergraduate students and university employees (mean age 22, 11 female). All participants had self-reported normal hearing and vision and no personal or close family history of neurological or psychiatric disorders. All recruitment and experimental procedures were approved by the Vanderbilt University Institutional Review Board.

Experiment protocol

The paradigm lasted 5 days in its entirety. On the initial day, participants completed a baseline audiovisual SJ task, followed by a training session. Training sessions included an initial audiovisual SJ block, followed by three blocks of visual temporal order judgment (TOJ) training, and ending with a final audiovisual SJ block. On days two through five, participants also completed a training session, with an additional post-participation survey on day five. The visual modality was chosen for training primarily due to a number of previous studies that have demonstrated that marked perceptual plasticity can be induced within the visual system (Adini et al. 2002; Seitz et al. 2006a).

Apparatus and Materials

The experiment took place in a dimly lit, sound-attenuated testing booth (Model SE 2000; Whisper Room Inc, Morristown, TN) with participants facing a computer monitor that served as the experimental display. Visual stimuli were presented using a 19-inch monitor (NEC MultiSync FE992, 160 Hz refresh rate) placed approximately 60 cm from the participant at eye level. Auditory stimuli were presented using Philips SBC HN-110 headphones. Stimulus presentation and timing were controlled via Matlab 7.11 (MathWorks, Natick, MA) psychophysics toolbox (Brainard 1997; Pelli 1997). All auditory stimuli were calibrated with a Larson-Davis Sound Track LxT2 sound level meter, and timing of all

stimuli and SOAs were verified using a Hameg 507 oscilloscope with a photovoltaic cell and microphone. External SOA measurement was used to calibrate internal estimates of SOA, which were recorded and monitored for all trials.

Visual temporal order judgment training task

The visual stimuli used in this task consisted of two identical white circles (1.5 cm diameter) presented on a black screen. These two circles were located 2.5 cm above and below a fixation X (1×1 cm). No auditory stimulus was present.

In this TOJ task, trials began with a black screen with a white fixation X, followed by a visual presentation in which the two circles appeared with a randomly chosen SOA. SOAs ranged from −37.5 ms (bottom stimulus leading) to 37.5 ms (top stimulus leading) at 6.25 ms intervals, including a 0 ms SOA condition in which the circles were presented simultaneously, for a total of 13 conditions. For a total of 14 individuals referred to as the "experimental" group, following a response, both circles disappeared and subjects were given feedback as to their response. If subjects responded correctly, a yellow smiling face appeared with the word "correct." If subjects responded incorrectly, a blue frowning face appeared with the word "incorrect" (see Figure 1A). Feedback was presented for 500 ms. For an additional 8 individuals, referred to as the "control" group, no feedback on task performance was given. Trial orders were randomized, and 20 trials of each of the 13 SOAs were presented for a total of 260 trials per block. Participants were instructed to indicate which circle appeared first, the top or bottom. Participants were instructed to respond as quickly as possible, and responded with a button press on a keyboard. Each experiment day included 3 blocks of visual TOJ training. See Figure 1A for a graphical representation of the visual training.

Two-interval forced choice audiovisual simultaneity judgment task

The visual stimulus in the SJ tasks consisted of a white ring on a black background (diameter = 9.4 cm) presented for 6.25 ms (i.e., one refresh cycle), with this duration verified via oscilloscope. Auditory stimuli consisted of 6ms, 1800 Hz tone bursts presented binaurally with no interaural-time or level differences at 70 dB SPL.

In this 2-IFC task, trials began with a black screen with a fixation cross, followed by two randomly ordered audiovisual pairs, one pair presented simultaneously and one pair presented non-simultaneously. The two pairs were separated by 1 s, during which participants were presented with a fixation cross alone. The non-simultaneous pair had a SOA ranging from −300 ms (auditory stimulus leading) to 300 ms (visual stimulus leading) at 50 ms intervals including a 0 ms SOA condition in which both presentations were simultaneous for a total of 13 conditions (Figure 1B). After presentation of both audiovisual pairs, the monitor returned to the black screen and white fixation cross until a response was entered. Once a response was entered, there was a 500 ms delay before the next trial began. Trial orders were randomized, and 20 trials of each of the 13 SOAs were presented for a total of 260 trials per block. Participants were instructed to indicate which audiovisual pair contained the simultaneous flash and beep, the first or the second. Participants were instructed to respond as quickly as possible, and responded with a button press on a keyboard. See Figure 1B for a graphical representation of the audiovisual task.

Data Analysis

All data were collected and analyzed in Matlab 7.11 (MathWorks, Natick MA). Individual subjects' raw data were used to calculate the mean accuracy of simultaneity judgments at each SOA for each day's audiovisual SJ tasks and to calculate the mean accuracy of the visual task at each SOA for each visual TOJ block.

With the visual TOJ tasks, overall group accuracies were calculated by averaging each individual's means at each SOA (Figure 2). Because the negative SOAs (bottom first) and positive SOAs (top first) have no ecological significance, they were combined by absolute SOA (e.g. −6.25 ms and 6.25 ms SOAs are grouped when calculating mean accuracy at each SOA). Statistical analysis included performing a two-factor (day, SOA) repeated-measures ANOVA and *post hoc*, bi-directional t-tests to determine which days showed significant differences in performance when compared to performance on day one.

With the audiovisual SJ tasks, overall group accuracies were calculated by averaging individuals' means at each SOA (Figure 3A). Statistical analysis involved performing a two-factor (SOA, assessment number) repeated-measures ANOVA and *post hoc*, bi-directional *t*-tests to determine which SOAs showed statistically-significant differences from baseline to training days two through five. Data from pre-training SJ tasks were used to measure the TBW as these data were more reliable than data from the post-training SJ task, presumably due to fatigue effects.

In addition to the simple mean accuracies, individual mean data from the audiovisual SJ tasks were fit with two sigmoid curves generated using the Matlab *glmfit* function, splitting the data into auditory-leading (negative SOA) and visual-leading (positive SOA) sides and fitting them separately, and method that has been used successfully in previous studies of the TBW (Powers et al. 2009; Hillock et al. 2011; Powers et al. 2012; Stevenson et al. 2012; Stevenson and Wallace Under Review). The criterion for measuring the individual temporal window size was established as half the distance between the individuals' lowest accuracy point at baseline assessment (i.e. 50%) and 100%, resulting in a threshold of approximately 75% perceived simultaneous. This TBW size criterion was then used to evaluate the width of each side of the distribution, and then the two sides were combined to estimate the total distribution width. This distribution width was considered the size of each individual's window (see Figure 3B for an example). Group-level analysis of window size changes across assessments was completed using a repeated-measures ANOVA (within-subject factor, assessment number) and *post hoc*, bi-directional *t*-tests to determine differences from the baseline at specific SOAs.

Results

Visual training improves visual temporal order judgments

To measure the impact that training had on the visual TOJ task in the experimental group, a repeated-measures ANOVA with within-subject factors SOA and experiment day was performed (Figure 4). Not surprisingly, SOA length had a significant main effect on accuracy ($F_{(5,420)} = 38.69$, $p = 2.14 \times 10^{-37}$, $\eta^2 = 0.29$, with subjects having higher response accuracies in correctly identifying stimulus order as SOA increased. In addition, experimental day had a significant main effect on accuracy ($F_{(4,420)} = 2.77$, $p = 0.027$, $\eta^2 = 0.02$), with accuracies improving with training. There was no significant interaction between SOA length and experimental day on accuracy of visual TOJs ($F_{(24,420)} = 0.17$). For specific SOAs, the experimental group significantly improved at the 12.5 ms (increased from 58% to 64%, $t_{(13)} = 2.80$, p = 0.01, $d = 0.72$), 18.75 ms (increased from 59% to 68%, $t_{(13)} = 3.78$, p = 0.002, $d = 0.78$), 25 ms (increased from 64% to 72%, $t_{(13)} = 3.77$, p = 0.002, $d = 0.61$), and 31.25 ms (increased from 69% to 76%, $t_{(13)} = 2.46$, p = 0.03, $d = 0.47$) SOAs. The experimental group showed a marginally significant improvement at the 37.5 ms (increased from 75% to 80%, $t_{(13)} = 1.95$, p = 0.07, $d = 0.29$). No significant gain was seen at the 6.25 SOA (no increased at 54%, $t_{(13)} = 0.80$) SOAs.

In the control group, a repeated-measures ANOVA with within-subject factors SOA and experiment day was also performed. SOA length had a significant main effect on accuracy

($F_{(5,140)} = 32.24$, $p = 1.56 \times 10^{-12}$, $\eta^2 = 0.52$), with subjects having higher response accuracies in correctly identifying stimulus order as SOA increased. Experimental day also had a significant main effect on accuracy ($F_{(4,140)} = 4.58$, $p = 0.01$, $\eta^2 = 0.02$), and there was no significant interaction between SOA length and experimental day on accuracy of visual TOJs ($F_{(20,140)} = 1.34$). For specific SOAs, the control group significantly improved at the 18.75 ms (increased from 72% to 82%, $t_{(7)} = 3.25$, p = 0.01, $d = 0.51$), 25 ms (increased from 74% to 81%, $t_{(7)} = 2.65$, p = 0.04, $d = 0.63$), and 31.25 ms (increased from 80% to 87%, $t_{(7)} = 2.95$, p = 0.02, $d = 0.62$) SOAs. The control group showed marginal improvement at the 12.5 ms (increased from 61% to 66%, $t_{(7)} = 2.15$, p = 0.07, $d = 0.85$) and 37.5 ms (increased from 84% to 89%, $t_{(7)} = 1.92$, p = 0.10, $d = 0.35$) SOAs. No significant gain was seen at the 6.25 SOA (increased from 56% to 59%, $t_{(7)} = 1.29$) SOAs.

Visual training improves audiovisual simultaneity judgments

In addition to examining the effects of visual training within modality on visual TOJ performance, we also examined the cross-modal effects of training on multisensory performance on an audiovisual simultaneity judgment task in the experimental group. In this analysis, group accuracy means were calculated at each SOA for each of the five days of the experiment (for an example comparing days 1 and 2, see Figure 3A). A repeated-measures ANOVA using within-subject factors of SOA and experimental day produced two significant main effects. First, there was a main effect of SOA on accuracy of simultaneity judgments ($F_{(13,720)} = 116.81$, $p = 6.61 \times 10^{-152}$, $\eta^2 = 0.62$). Also, there was a main effect of experimental day, with performance accuracy improving across days of training ($F_{(4,720)} = 6.06$, $p = 8.44 \times 10^{-5}$, $\eta^2 = 0.01$). There was no significant interaction between experimental day and SOAs on accuracy of simultaneity judgments ($F_{(44, 720)} = 0.77$). *Post hoc* t-tests were run to determine if there were significant effects at individual SOAs (Figure 3A), as previous research has suggested that training-related effects occur predominately at intermediate SOAs. With visual-leading SOAs, significant or marginally significant increases in accuracy were seen as early as day 2 relative to baseline at SOAs of 100 ms (increased from 60% to 68%, $t_{(13)} = 1.86$, p = 0.086, $d = 0.65$), 150 ms (increased from 66% to 78%, $t_{(13)} = 4.00$, p = 0.002, $d = 0.81$), 200 ms (increased from 78% to 86%, $t_{(13)} = 1.94$, p = 0.074, $d = 0.61$), and 250 ms (increased from 81% to 91%, $t_{(13)} = 4.71$, p = 0.0003, $d = 1.02$). The most extreme visual-leading SOA, at 300ms, did not significantly change with training (90% to 93%, $t_{(13)} = 1.26$). For auditory-leading SOAs, only marginally significant increases in accuracy were seen at SOAs of 150 ms (increased from 81% to 87%, $t = 1.84$, p = 0.089, $d = 0.70$) and 200 ms (increased from 89% to 94%, $t = 2.13$, p = 0.053, $d = 0.91$). The most extreme auditory-leading SOAs, at –250 and 300ms, did not significantly change with training (96% to 95%, $t_{(13)} = 0.17$ and 94% to 94%, $t_{(13)} = 0.16$, respectively). Finally, the shortest SOAs, –100 ms (64% to 71%, $t_{(13)} = 1.52$), –50 ms (59% to 59%, $t_{(13)} = 0.14$), and 50 ms (52% to 52%, $t_{(13)} = 0.00$), similarly showed no significant difference with training.

Data from the audiovisual simultaneity judgment task was also analyzed for the control group. A repeated-measures ANOVA was run using within-subject factors of SOA and experimental day produced only one significant main effect. There was a main effect of SOA on accuracy of simultaneity judgments ($F_{(13,308)} = 28.48$, $p < 1.00 \times 10^{-5}$, $\eta^2 = 0.38$). However, without feedback there was no main effect of experimental day ($F_{(4,308)} = 0.89$, p = 0.54), and there was no significant interaction between experimental day and SOAs on accuracy of simultaneity judgments ($F_{(44, 308)} = 0.91$, $p = 0.64$). *Post hoc* t-tests were run to determine if there were significant effects at individual SOAs. With visual-leading SOAs, a data from a single SOA showed a significant increase in accuracy at day 2 relative to baseline, the 300 ms SOA (87% to 95%, $t_{(7)} = 3.52$, p = 0.01, $d = 0.54$). No other visual-leading SOAs showed any significant changes including the 50 ms (52% to 55%, $t_{(7)} = $

0.51), 100 ms (71% to 66%, $t_{(7)} = 0.73$), 150 ms (86% to 86%, $t_{(7)} = 0.24$), 200 ms (81% to 89%, $t_{(7)} = 1.72$), and 250 ms (89% to 86%, $t_{(7)} = 1.67$). No auditory-leading SOAs showed any significant changes including the 50 ms (64% to 61%, $t_{(7)} = 0.77$), 100 ms (68% to 71%, $t_{(7)} = 0.68$), 150 ms (82% to 84%, $t_{(7)} = 0.22$), 200 ms (84% to 80%, $t_{(7)} = 1.08$), 250 ms (88% to 88%, $t_{(7)} = 0.31$), and 300 ms (86% to 89%, $t_{(7)} = 0.86$).

In an effort to extend these results to the construct of the TBW, binding windows were calculated for each individual in the experimental group based on the audiovisual SJ task at the beginning of each training session by fitting accuracy data with two sigmoid functions (see Figure 3B for an example subject). To ensure that these functions fit the data well, an omnibus ANOVA was run comparing each measured data point and the data point estimated by the function across SOAs (12), sessions (11), and participants (22). Data extracted from the sigmoid functions did not significantly differ from the measured data ($F_{(1,5807)} = 0.31$, mean deviation = 0.7% accuracy). These data illustrate that training on the purely visual task led to a substantial and rapid shift in the size of the audiovisual temporal binding window (Figure 4a). In measurements of the full TBW (including both the auditory- and visual-leading portions of the distribution), windows were narrower than baseline (300 ms) on day 2 (209 ms, $t_{(13)} = 3.49$, p = 0.004, $d = 0.92$), day 3 (209 ms, $t_{(13)} = 3.37$, p = 0.008, $d = 1.01$), and day 4 (216 ms, $t_{(13)} = 2.16$, p = 0.021, $d = 0.82$), with a marginally significant effect on day 5 (232 ms, $t_{(13)} = 1.71$, p = 0.11, $d = 0.67$). These changes were also analyzed independently for the auditory-leading and visual-leading side of the window. The largest changes were seen in conditions where visual preceded auditory presentations. There were significant decreases from baseline (204 ms) to day 2 (133 ms, $t_{(13)} = 2.86$, $p = 0.013$, $d = 0.80$), day 3 (135 ms, $t_{(13)} = 2.51$, $p = 0.026$, $d = 0.79$), and day 4 (139 ms, $t_{(13)} = 2.17$, p = 0.05, $d = 0.75$), and a marginally significant effect at day 5 (149 ms, $t_{(13)} = 1.63$, p = 0.13, $d = 0.60$). Where audition preceded vision, there were smaller yet still significant differences relative to baseline (96 ms) on days 2 (75 ms, $t_{(13)} = 2.19$, $p = 0.047$, $d = 0.66$) and 3 (73 ms, $t_{(13)} = 4.00$, $p = 0.002$, $d = 0.81$), a marginal difference at day 4 (77 ms, $t_{(13)} = 1.61$, $p = 0.13$, $d = 0.59$), and no significant effect at day 5 (83 ms, $t_{(13)} = 1.18$, $p = 0.26$).

TBWs were also measured for the control group. These calculations were performed for the entire window, as well as for visual-leading and auditory-leading TBWs. No significant narrowing was seen in the TBW in any portion of the window ($\alpha = 0.05$; see Figure 4b). For the whole TBW, relative to the baseline measurement (250 ms), there were no changes seen on day 2 (297 ms, $t_{(7)} = 1.04$), day 3 (298 ms, $t_{(7)} = 0.87$), day 4 (257 ms, $t_{(7)} = 0.39$), or day 5 (282 ms, $t_{(7)} = 0.90$). Likewise, for the visual-leading TBW, relative to the baseline measurement (133 ms), there were no changes seen on day 2 (148 ms, $t_{(7)} = 1.14$), day 3 (174 ms, $t_{(7)} = 0.93$), day 4 (129 ms, $t_{(7)} = 0.07$), or day 5 (166 ms, $t_{(7)} = 1.11$). Finally, for the auditory-leading TBW, relative to the baseline measurement (116 ms), there were no changes seen on day 2 (149 ms, $t_{(7)} = 0.86$), day 3 (124 ms, $t_{(7)} = 0.38$), day 4 (128 ms, $t_{(7)} = 0.85$), or day 5 (116 ms, $t_{(7)} = 0.03$).

To directly measure if there was a difference in training effect between experimental and control groups, a group × training day ANOVA was run. Importantly, for whole TBWs there was a significant group × training day interaction ($F_{(4,109)} = 2.78$, $p = 0.04$, $\eta^2 = 0.04$). Additionally, such an interaction was tested for with visual-leading and auditory-leading TBWs as well. Following the patter in which larger changes were measured in the visual-leading windows, a marginally significant interaction was seen in the visual-leading windows ($F_{(4,109)} = 2.16$, $p = 0.09$, $\eta^2 = 0.05$), where no difference was seen in the auditory-leading windows ($F_{(4,109)} = 1.59$).

Initial window size is a strong predictor of training effects

Previous research has shown that there is a correlation between initial multisensory TBW size and the degree to which the window could be narrowed as a result of audiovisual training. More specifically, the wider the initial multisensory TBW the greater the gain that could be observed with training (Powers et al. 2009). To test this hypothesis, a correlation analysis was run comparing initial TBW as measured with the audiovisual SJ tasks and the total change in individuals' windows following the visual training in the experimental group (no such analysis was completed in the control group given the lack of training effect). Indeed, this analysis showed a striking correlation, in that a wider initial TBW was associated with greater visual training-induced narrowing of the TBW ($r^2 = 0.73$, $p = 0.004$).

Discussion

The importance of multisensory integration on human perception is well documented, as is the impact that temporal synchrony has on driving such integration. Previous research has shown a remarkable malleability in adult multisensory processing, with substantial changes being driven by short-term manipulations of multisensory relations (Fujisaki et al. 2004; Vroomen et al. 2004; Powers et al. 2009). Here we extend these findings, specifically showing that the temporal aspects of multisensory integration can be changed via alterations in unisensory experience as well. Several key findings derive from the current experiments. First, perceptual learning via visual temporal order training can successfully improve temporal visual processing, replicating prior studies of visual perceptual plasticity (Adini et al. 2002; Seitz et al. 2006a). Second, we show that this same visual training when paired with feedback increases the temporal precision on a multisensory (i.e., visual-auditory) task; specifically visual temporal order training with feedback narrows the multisensory TBW as defined using a simultaneity judgment task. Third, this narrowing is driven in large part by changes in the visual-leading portion of the TBW. Fourth, this study suggests that this narrowing takes place rapidly, being seen after a single one hour training session. Finally, the degree of plasticity appears to be determined by the initial size of the multisensory TBW. Thus, individuals with large multisensory windows prior to training are those that see the greatest effects from training.

While previous studies (Kim et al. 2008; Alais and Cass 2010) have shown that audiovisual training can affect visual learning at the perceptual level, this is the first to show a reciprocal relationship in which visual training with feedback affects multisensory learning. The findings generated here provide some insight into the transfer of training across sensory domains, illustrating that visual training structured to improve temporal performance on visual tasks when paired with feedback extends to multisensory (i.e., audiovisual) temporal tasks as well. One possible explanation for this is that the training that includes explicit feedback impacts either supramodal or amodal temporal processes, altering the tolerance of an internal referential clock that is used to guide judgments on temporally based tasks such as the temporal order and simultaneity tasks used here (Gibbon 1977; Gibbon et al. 1984). With changes to such a supramodal clock, one would expect that the performance improvements would extend beyond the trained modalities (e.g., to auditory-tactile temporal judgments).

Alternatively, the results may be specific to visual processing through either modulating a central internal clock in a modality-specific manner or altering a modality-specific clock (Wearden et al. 1998). In this case, visual training when paired with feedback would alter multisensory temporal processing simply because the visual input is temporally more precise, which may explain the asymmetric impact of training where visual-leading conditions were more susceptible to improvement. Changes in an independent modality-specific clock would be limited to judgments made on multisensory stimuli that are within

the trained modality (i.e., visual-auditory and visual-tactile as opposed to auditory-tactile), and would be the result of changes to the neural architecture supporting visual temporal processes. Previous research investigating unisensory and multisensory temporal processing suggests that there are indeed sensory-specific as well as amodal internal reference clocks, but that they are in all likelihood highly interactive, and as such, our results may be driven by changes at both levels (Wearden et al. 1998). Further research, possibly testing for such a change in auditory-tactile temporal processing with visual feedback training, is needed to distinguish between these possibilities. Importantly, visual training without feedback also resulted in an increase in significant gains in visual accuracy. Without feedback though, there was no transfer in temporal precision to the audiovisual realm, strongly suggesting that the effects are in fact due to changes to a supramodal or amodal clock.

The effects of visual feedback training in narrowing the multisensory TBW were rapid, occurring after only a single hour of training. Additional training resulted in no appreciable change in the width of the TBW. Such a timeline for change closely mirrors that seen with multisensory feedback training (Powers et al. 2009). Also similar to multisensory training, the changes engendered following visual training were specifically dependent upon feedback. The lack of change in multisensory temporal processing with visual TOJ exposure in our control group showed that passive exposure was not sufficient to drive such a change in multisensory processing (Powers et al. 2009). Here we expand these results to show that visual-only training with feedback transfers to multisensory (i.e., audiovisual) temporal processing, as well as to a different task (SJ) that has previously been shown to be uncorrelated with the specific task use in training (van Eijk et al. 2008). Additionally, it appears that supervised training (training mechanisms driven by feedback) in particular is sufficient to engender this change, whereas exposure to the task and stimuli, and associated practice effects, do not appear to induce transfer of perceptual learning from unisensory visual to multisensory performance. However, whether unsupervised training (based on exposure to controlled environmental statistics) has the ability to induce such a transfer remains unknown.

While the narrowing of multisensory temporal processing was seen for both visual-leading and auditory-leading stimuli, the largest effect was seen on the visual-leading portion of the TBW. Greater plasticity here is consistent with the ethological relevance of visual-leading stimuli, which reflects the statistical relationship of visual and auditory cues (in which visual energy always impinges on the retina before auditory energy impinges on the cochlea; Pöppel et al. 1990). Also, it is possible that since the visual-leading portion of the TBW was initially larger than the auditory-leading portion, the asymmetrical effect of visual temporal training was the result of a correlation between training effect size and initial window size.

Another salient feature of the results is the significant training-related effect seen at intermediate SOAs, again a finding consistent with the results seen after multisensory training (Powers et al. 2009). The lack of change at the longest SOAs is likely a result of a ceiling effect: participants were already responding with extremely high accuracy and as such did not have room for improvement. The lack of change at the shortest SOAs may simply reflect task difficulty at these intervals and inherent neural limitations in distinguishing two stimuli that are temporally proximate. Consistent with these interpretations, and which further reinforce the plasticity at intermediate SOAs, is the result of a followup neuroimaging study to the multisensory perceptual feedback training work (Powers et al. 2012). In this fMRI study, which identified the posterior superior temporal sulcus (pSTS) as a major node in the multisensory training-induced changes in temporal perception, the largest hemodynamic changes following training were seen for intermediate SOAs.

The ability to manipulate the TBW through both multisensory and visual training, in addition to providing insights into multisensory integration on a basic level, has potential clinical implications. A number of clinical populations, including those with autism (Foss-Feig et al. 2009; Kwakye et al. 2011), dyslexia (Hairston et al. 2005), and schizophrenia (Foucher et al. 2007) show atypically wide TBWs. Training paradigms aimed at narrowing the TBW provide a possible remediation strategy in which individuals within these populations may improve sensory processing abilities. This possibility is further supported by neurological evidence that regions such as superior temporal sulcus showing anatomical and functional abnormalities in autism (Levitt et al. 2003; Boddaert et al. 2004; Gervais et al. 2004; Pelphrey and Carter 2008a; Pelphrey and Carter 2008b; Bastien-Toniazzo et al. 2009), dyslexia (Pekkola et al. 2006; Richards et al. 2008), and schizophrenia (Shenton et al. 2001) also underlie multisensory temporal processes (Miller and D'Esposito 2005; Stevenson et al. 2010; Stevenson et al. 2011).

The ability to impact multisensory temporal processing through visual-only, perceptual-feedback learning also leads to a number of related possibilities. Future work should attempt to examine the effects of auditory training on multisensory function. Our expectation is that these changes will be less dramatic than the changes following visual training because auditory temporal discrimination is more precise than visual temporal discrimination (Alais and Cass 2010). Furthermore, as previously mentioned, these data suggest that this effect is caused by modifications to a higher-level internal timing mechanism as opposed to visual processing specifically, however, this is not directly assessed through this paradigm. Measuring possible transfer of these effects to an untrained modality pair, such as an auditory-tactile task, would provide further insight into this question, and should be pursued.

The results reported here provide evidence that visual-only perceptual learning via feedback training is capable of modulating related multisensory integration. Specifically, increasing the temporal precision of visual perception can narrow the multisensory temporal binding window. This narrowing is rapid, occurring after a single day of training, and is driven by changes in intermediate SOAs in the visual-leading TBW. The ability to modulate multisensory temporal processing has possible clinical applications that warrant further investigation, including remediation strategies in clinical populations with atypical temporal processing such as autism, dyslexia, and schizophrenia.

Acknowledgments

References

Adini Y, Sagi D, Tsodyks M. Context-enabled learning in the human visual system. Nature. 2002; 415:790–793. [PubMed: 11845209]

Alais D, Cass J. Multisensory perceptual learning of temporal order: audiovisual learning transfers to vision but not audition. PLoS One. 2010; 5:e11283. [PubMed: 20585664]

Andersen TS, Tiippana K, Sams M. Factors influencing audiovisual fission and fusion illusions. Brain Res Cogn Brain Res. 2004; 21:301–308. [PubMed: 15511646]

Bastien-Toniazzo M, Stroumza A, Cavé C. Audio-visual perception and integration in developmental dyslexia: An exploratory study using the McGurk effect. Current Psychology Letters. 2009; 25

Boddaert N, Chabane N, Gervais H, Good CD, Bourgeois M, Plumet MH, Barthelemy C, Mouren MC, Artiges E, Samson Y, Brunelle F, Frackowiak RS, Zilbovicius M. Superior temporal sulcus

anatomical abnormalities in childhood autism: a voxel-based morphometry MRI study. Neuroimage. 2004; 23:364–369. [PubMed: 15325384]

Brainard DH. The Psychophysics Toolbox. Spat Vis. 1997; 10:433–436. [PubMed: 9176952]

Colonius H, Diederich A. Multisensory interaction in saccadic reaction time: a time-window-of-integration model. J Cogn Neurosci. 2004; 16:1000–1009. [PubMed: 15298787]

Conrey B, Pisoni DB. Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. J Acoust Soc Am. 2006; 119:4065–4073. [PubMed: 16838548]

Conrey, BL.; Pisoni, DB. Detection of Auditory-Visual Asynchrony in Speech and Nonspeech Signals. In: Pisoni, DB., editor. Research on Spoken Language Processing. Vol. vol 26. Indiana University; Bloomington: 2004. p. 71-94.

Diederich A, Colonius H. Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. Percept Psychophys. 2004; 66:1388–1404. [PubMed: 15813202]

Dixon NF, Spitz L. The detection of auditory visual desynchrony. Perception. 1980; 9:719–721. [PubMed: 7220244]

Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT. An extended multisensory temporal binding window in autism spectrum disorders. Exp Brain Res. 2009

Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT. An extended multisensory temporal binding window in autism spectrum disorders. Exp Brain Res. 2010; 203:381–389. [PubMed: 20390256]

Foucher JR, Lacambre M, Pham BT, Giersch A, Elliott MA. Low time resolution in schizophrenia Lengthened windows of simultaneity for visual, auditory and bimodal stimuli. Schizophr Res. 2007; 97:118–127. [PubMed: 17884350]

Fujisaki W, Shimojo S, Kashino M, Nishida S. Recalibration of audiovisual simultaneity. Nat Neurosci. 2004; 7:773–778. [PubMed: 15195098]

Gervais H, Belin P, Boddaert N, Leboyer M, Coez A, Sfaello I, Barthelemy C, Brunelle F, Samson Y, Zilbovicius M. Abnormal cortical voice processing in autism. Nat Neurosci. 2004; 7:801–802. [PubMed: 15258587]

Gibbon J. Scalar expectancy theory and Weber's law in animal timing. Psychological Review. 1977; 84:179–325.

Gibbon J, Church RM, Meck WH. Scalar timing in memory. Ann N Y Acad Sci. 1984; 423:52–77. [PubMed: 6588812]

Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT. Altered temporal profile of visual-auditory multisensory interactions in dyslexia. Exp Brain Res. 2005; 166:474–480. [PubMed: 16028030]

Hershenson M. Reaction time as a measure of intersensory facilitation. J Exp Psychol. 1962; 63:289–293. [PubMed: 13906889]

Hillock AR, Powers AR, Wallace MT. Binding of sights and sounds: age-related changes in multisensory temporal processing. Neuropsychologia. 2011; 49:461–467. [PubMed: 21134385]

James, TW.; Stevenson, RA.; Kim, S. Inverse effectiveness in multisensory processing. In: Stein, BE., editor. The New Handbook of Multisensory Processes. MIT Press; Cambridge, MA: 2012.

Keetels M, Vroomen J. The role of spatial disparity and hemifields in audio-visual temporal order judgments. Exp Brain Res. 2005; 167:635–640. [PubMed: 16175363]

Kim RS, Seitz AR, Shams L. Benefits of stimulus congruency for multisensory facilitation of visual learning. PLoS One. 2008; 3:e1532. [PubMed: 18231612]

Kwakye LD, Foss-Feig JH, Cascio CJ, Stone WL, Wallace MT. Altered auditory and multisensory temporal processing in autism spectrum disorders. Front Integr Neurosci. 2011; 4:129. [PubMed: 21258617]

Levitt JG, Blanton RE, Smalley S, Thompson PM, Guthrie D, McCracken JT, Sadoun T, Heinichen L, Toga AW. Cortical sulcal maps in autism. Cereb Cortex. 2003; 13:728–735. [PubMed: 12816888]

Lovelace CT, Stein BE, Wallace MT. An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. Brain Res Cogn Brain Res. 2003; 17:447–453. [PubMed: 12880914]

Macaluso E, George N, Dolan R, Spence C, Driver J. Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage. 2004; 21:725–732. [PubMed: 14980575]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264:746–748. [PubMed: 1012311]

Meredith MA, Nemitz JW, Stein BE. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci. 1987; 7:3215–3229. [PubMed: 3668625]

Meredith MA, Stein BE. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. Brain Res. 1986a; 365:350–354. [PubMed: 3947999]

Meredith MA, Stein BE. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J Neurophysiol. 1986b; 56:640–662. [PubMed: 3537225]

Miller LM, D'Esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci. 2005; 25:5884–5893. [PubMed: 15976077]

Navarra J, Vatakis A, Zampini M, Soto-Faraco S, Humphreys W, Spence C. Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. Brain Res Cogn Brain Res. 2005; 25:499–507. [PubMed: 16137867]

Nelson WT, Hettinger LJ, Cunningham JA, Brickman BJ, Haas MW, McKinley RL. Effects of localized auditory information on visual target detection performance using a helmet-mounted display. Hum Factors. 1998; 40:452–460. [PubMed: 9849103]

Pekkola J, Laasonen M, Ojanen V, Autti T, Jaaskelainen IP, Kujala T, Sams M. Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: an fMRI study at 3 T. Neuroimage. 2006; 29:797–807. [PubMed: 16359873]

Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis. 1997; 10:437–442. [PubMed: 9176953]

Pelphrey KA, Carter EJ. Brain mechanisms for social perception: lessons from autism and typical development. Ann N Y Acad Sci. 2008a; 1145:283–299. [PubMed: 19076404]

Pelphrey KA, Carter EJ. Charting the typical and atypical development of the social brain. Dev Psychopathol. 2008b; 20:1081–1102. [PubMed: 18838032]

Pöppel E, Schill K, von Steinbüchel N. Sensory integration within temporally neutral systems states: A hypothesis. Naturwissenschaften. 1990; 77:89–91. [PubMed: 2314478]

Powers AR 3rd, Hevey MA, Wallace MT. Neural correlates of multisensory perceptual learning. J Neurosci. 2012; 32:6263–6274. [PubMed: 22553032]

Powers AR 3rd, Hillock AR, Wallace MT. Perceptual training narrows the temporal window of multisensory binding. J Neurosci. 2009; 29:12265–12274. [PubMed: 19793985]

Richards T, Stevenson J, Crouch J, Johnson LC, Maravilla K, Stock P, Abbott R, Berninger V. Tract-based spatial statistics of diffusion tensor imaging in adults with dyslexia. AJNR Am J Neuroradiol. 2008; 29:1134–1139. [PubMed: 18467520]

Roach NW, Heron J, Whitaker D, McGraw PV. Asynchrony adaption reveals neural population code for audio-visual timing. Proceedings of the Royal Society. 2011; 278:9.

Royal DW, Carriere BN, Wallace MT. Spatiotemporal architecture of cortical receptive fields and its impact on multisensory interactions. Exp Brain Res. 2009; 198:127–136. [PubMed: 19308362]

Schall S, Quigley C, Onat S, Konig P. Visual stimulus locking of EEG is modulated by temporal congruency of auditory stimuli. Exp Brain Res. 2009; 198:137–151. [PubMed: 19526359]

Seitz AR, Nanez JE Sr. Holloway SR, Watanabe T. Perceptual learning of motion leads to faster flicker perception. PLoS One. 2006a; 1:e28. [PubMed: 17183655]

Seitz AR Sr. JEN, Holloway SR, Watanabe T. Perceptual Learning of Motion Leads to Faster Flicker Perception. PLoS ONE. 2006b; 1

Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG. Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia. 2007; 45:561–571. [PubMed: 16542688]

Shams L, Kamitani Y, Shimojo S. Illusions. What you see is what you hear. Nature. 2000; 408:788.

Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. Schizophr Res. 2001; 49:1–52. [PubMed: 11343862]

Spence C. Audiovisual multisensory integration. Accoustics, Science, & Technology. 2007; 28:61–70.

Stein, B.; Meredith, MA. The Merging of the Senses. MIT Press; Boston, MA: 1993.

Stein BE, Wallace MT. Comparisons of cross-modality integration in midbrain and cortex. Prog Brain Res. 1996; 112:289–299. [PubMed: 8979836]

Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW. Neural processing of asynchronous audiovisual speech perception. Neuroimage. 2010; 49:3308–3318. [PubMed: 20004723]

Stevenson RA, James TW. Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. Neuroimage. 2009; 44:1210–1223. [PubMed: 18973818]

Stevenson RA, VanDerKlok RM, Pisoni DB, James TW. Discrete neural substrates underlie complementary audiovisual speech integration processes. Neuroimage. 2011; 55:1339–1345. [PubMed: 21195198]

Stevenson RA, Wallace MW. Multisensory temporal integration: Task and stimulus dependencies. Experimental Brain Research. (Under Review).

Stevenson RA, Zemtsov RK, Wallace MT. Individual Differences in the Multisensory Temporal Binding Window Predict Susceptibility to Audiovisual Illusions. J Exp Psychol Hum Percept Perform. 2012

Stevenson RA, Zemtsov RK, Wallace MT. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. Journal of Experimental Pschology: Human Perception and Performance. (In Press).

Talsma D, Senkowski D, Woldorff MG. Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli. Exp Brain Res. 2009; 198:313–328. [PubMed: 19495733]

van Atteveldt NM, Formisano E, Blomert L, Goebel R. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb Cortex. 2007; 17:962–974. [PubMed: 16751298]

van Eijk RL, Kohlrausch A, Juola JF, van de Par S. Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type. Percept Psychophys. 2008; 70:955–968. [PubMed: 18717383]

van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. Neuropsychologia. 2007; 45:598–607. [PubMed: 16530232]

Vatakis A, Ghazanfar AA, Spence C. Facilitation of multisensory integration by the "unity effect" reveals that speech is special. J Vis. 2008; 8:14–11. [PubMed: 18831650]

Vatakis A, Navarra J, Soto-Faraco S, Spence C. Temporal recalibration during asynchronous audiovisual speech perception. Exp Brain Res. 2007; 181:173–181. [PubMed: 17431598]

Vatakis A, Spence C. Audiovisual synchrony perception for music, speech, and object actions. Brain Res. 2006; 1111:134–142. [PubMed: 16876772]

Vatakis A, Spence C. Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. Percept Psychophys. 2007; 69:744–756. [PubMed: 17929697]

Vroomen J, Baart M. Phonetic recalibration only occurs in speech mode. Cognition. 2009; 110:254–259. [PubMed: 19059584]

Vroomen J, Keetels M, de Gelder B, Bertelson P. Recalibration of temporal order perception by exposure to audio-visual asynchrony. Brain Res Cogn Brain Res. 2004; 22:32–35. [PubMed: 15561498]

Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA. Unifying multisensory signals across time and space. Exp Brain Res. 2004; 158:252–258. [PubMed: 15112119]

Wearden JH, Edwards H, Fakhri M, Percival A. Why "sounds are judged longer than lights": application of a model of the internal clock in humans. Q J Exp Psychol B. 1998; 51:97–120. [PubMed: 9621837]

Wilkinson LK, Meredith MA, Stein BE. The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. Exp Brain Res. 1996; 112:1–10. [PubMed: 8951401]

Zampini M, Guest S, Shore DI, Spence C. Audio-visual simultaneity judgments. Percept Psychophys. 2005; 67:531–544. [PubMed: 16119399]
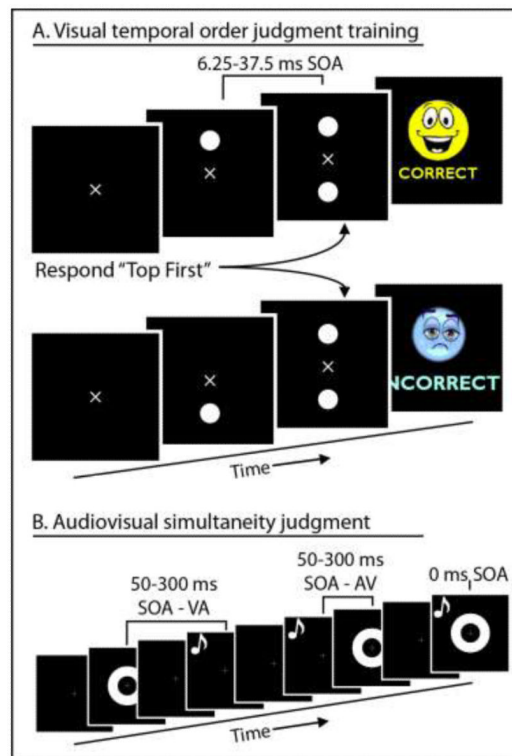
Figure 1. Experimental methods
Unisensory visual training consisted of a temporal order judgment task with feedback (Panel A). Control groups underwent an identical procedure without feedback presented. The multisensory temporal binding window was assessed with a 2-IFC simultaneity judgment task without feedback (Panel B).
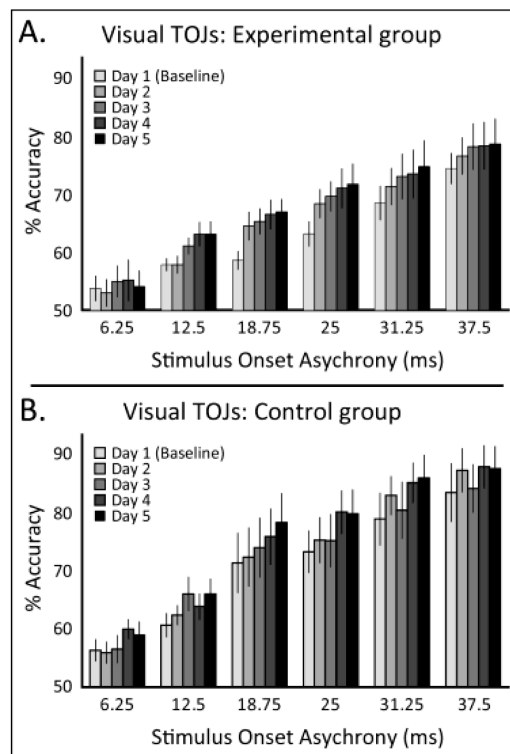
Figure 2. Visual temporal order judgments
Individuals were more accurate with larger stimulus onset asynchronies, and visual feedback training (A) and visual exposure (B) increased accuracy levels. Data points depicted here are group averages and standard errors from the experimental (A) and control (B) groups.
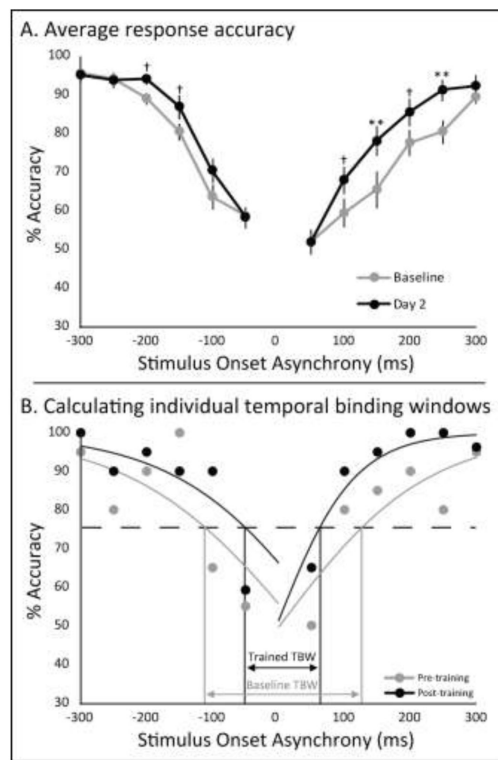
Figure 3. Audiovisual simultaneity judgments
Individuals showed significant improvements in temporal precision as early as the second day of training (Panel A) primarily at intermediate stimulus onset asynchronies, and with stronger effects seen when visual stimuli preceded auditory stimuli. Significance levels are denoted as $*p < 0.05$, $**p < 0.01$, and $^{\dagger}p < 0.10$. Each individual's 2IFC assessment data were fit with sigmoid curves from which a multisensory temporal binding window was derived at a 75% criterion (for an example subjects data points and sigmoid curve, see Panel B).
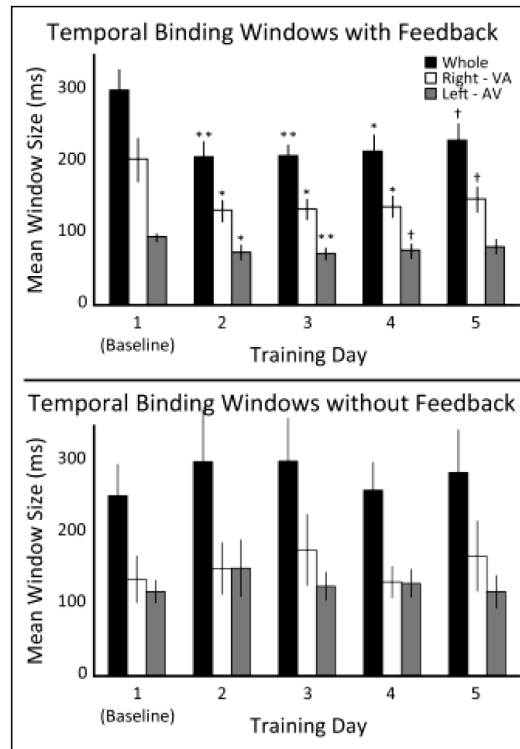
Figure 4. Narrowing of the temporal binding window following visual training
Average window sizes (total, right, left) prior to each of the 5 days of training are depicted.
Window size decreased significantly after the first day of training, then remained stable.
Error bars indicate 1SEM, all p-values representing change from baseline are denoted as **
$p < 0.01$, * $p < 0.05$, and † $p < 0.15$.