Published in final edited form as:

Perception. 2015 ; 44(7): 764–778.

Direction of auditory pitch-change influences visual search for slope from graphs

Stacey Parrott¹, Emmanuel Guzman-Martinez¹, Laura Ortega¹, Marcia Grabowecky^{1,2}, Mark D. Huntington³, and Satoru Suzuki^{1,2}

¹Department of Psychology, Northwestern University, Evanston, IL

²Interdepartmental Neuroscience Program, Northwestern University, Evanston, IL

³Department of Materials Science and Engineering, Northwestern University, Evanston, IL

Abstract

Linear trend (slope) is important information conveyed by graphs. We investigated how sounds influenced slope detection in a visual search paradigm. Four bar graphs or scatter plots were presented on each trial. Participants looked for a positive-slope or a negative-slope target (in blocked trials), and responded to targets in a go/no-go fashion. For example, in a positive-slope-target block, the target graph displayed a positive slope while other graphs displayed negative slopes (a go trial), or all graphs displayed negative slopes (a no-go trial). When an ascending or descending sound was presented concurrently, ascending sounds slowed detection of negative-slope targets whereas descending sounds slowed detection of positive-slope targets. The sounds had no effect when they immediately preceded the visual search displays, suggesting that the results were due to crossmodal interaction rather than priming. The sounds also had no effect when targets were words describing slopes, such as “positive,” “negative,” “increasing,” or “decreasing,” suggesting that the results were unlikely due to semantic-level interactions. Manipulations of spatiotemporal similarity between sounds and graphs had little effect. These results suggest that ascending and descending sounds influence visual search for slope based on a general association between the direction of auditory pitch-change and visual linear trend.

Introduction

The ability to extract linear trend (or slope) from scatterplots and bar graphs is essential for the interpretation of quantitative information. Because slope extraction is a skill that is difficult for students to master (Hadjidemetriou and Williams, 2001), it is helpful to understand factors that influence slope perception. Graphs utilize spatial relationships to convey quantitative concepts (e.g., Pinker, 1990; Tversky, 2005), so that spatial attention is a general factor that influences graph comprehension (e.g., Grant & Spivey, 2003).

Here we investigated the possibility that multisensory correspondences may also influence slope perception. In particular, we examined auditory influences on slope detection because

a graph conveys quantitative information as a pattern of visual elevation that may be perceptually associated with auditory pitch. A higher pitch is typically associated with the perception of and attention to visual elements at a higher elevation (e.g., Evans & Treisman, 2010; Pratt, 1930; Roffler & Butler, 1967; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006), and this crossmodal association may derive from natural perceptual experience (Parise, Knorre & Ernst, 2014). It has also been shown that ascending and descending pitch draw visual spatial attention and bias motion perception in upward and downward directions (e.g., Maeda, Kanai & Shimojo, 2004; Mossbridge, Grabowecky, & Suzuki, 2011). Furthermore, Walker (1987) demonstrated that people tend to subjectively associate a left-to-right pattern of visual elevation (e.g., low, high, high, high) with a corresponding temporal pattern of auditory pitch (e.g., low, high, high, high). Consistent with this finding, people can extract linear trends (e.g., the direction and strength of linear correlation) similarly whether the information is visually presented as a scatter plot or auditorily presented as a pitch progression over time (with pitch as the Y axis and time as the X axis) (Flowers, Buhman & Turnage, 1997). These results suggest that hearing an ascending or descending pitch may automatically influence visual processing of slopes in bar graphs and scatter plots.

To simulate search for a data pattern in a situation such as reading a report containing multiple graphs, we simultaneously presented four bar graphs (or scatter plots), and asked participants to respond as quickly and as accurately as possible as to whether an upward (or downward) trend was present in any of the bar graphs (or scatter plots). If a positive visual slope is crossmodally associated with an ascending auditory pitch, hearing an ascending sound should facilitate detection of a positive slope and/or interfere with detection of a negative slope. Similarly, if a negative visual slope is crossmodally associated with a descending auditory pitch, hearing a descending sound should facilitate detection of a negative slope and/or interfere with detection of a positive slope.

To gain insight into the representation in which the crossmodal interaction may occur, we manipulated spatiotemporal similarity between the auditory and visual stimuli. Specifically, the trend conveyed in each graph was defined as a pattern of elevations of ten discrete bars or points from left to right. Thus, each graph conveyed an ascending or descending trend with ten discrete elements and a spatial progression from left to right. In three experimental conditions, we varied the degree to which the auditory stimuli captured these spatiotemporal properties. In the *continuous-sound* condition the pitch continuously ascended or descended, capturing the increasing or decreasing aspects of a visual trend without capturing visual discreteness or left-to-right spatial directionality. In the *discrete-sound* condition the pitch ascended or descended in ten discrete steps, additionally capturing visual discreteness with auditory articulation. In the *moving-discrete-sound* condition the pitch ascended or descended in ten discrete steps while it also spatially moved from left to right, additionally capturing left-to-right directionality. If auditory-visual interactions are enhanced by greater spatiotemporal similarity, ascending/descending sounds should produce progressively stronger effects on the detection of positive/negative graph slopes in the discrete-sound and moving-discrete-sound conditions relative to the continuous-sound condition. Alternatively, if the relevant auditory-visual interactions depend primarily on an abstract relationship

between auditory-temporal processing of pitch direction and visual-spatial processing of slope, the spatiotemporal similarity manipulations should have little effect.

Experiment 1: Slope search with concurrent auditory and visual information

Method

Participants—Thirty-six Northwestern university undergraduate students gave informed consent to participate for partial course credit. All had normal or corrected-to-normal vision and normal hearing, and were tested individually in a normally lit room.

Stimuli—Visual stimuli were presented on a 19" CRT monitor at a 75 Hz refresh rate and a 1024 by 768 pixel resolution, controlled by a Dell desktop computer using MATLAB software (version 2010a) running Psychtoolbox (Brainard, 1997; Pelli, 1997). The viewing distance was 70 cm, so that each pixel subtended approximately 0.029° visual angle. We generated a large number of bar graphs and scatter plots, each containing ten data points (i.e., 10 bars or 10 points). The spacing between the centers of the adjacent bars or points was 0.90° and the center of the first bar or point was 0.44° to the right of the left end of the X-axis. The length of the horizontal line depicting the X-axis was 8.90° . The height of each bar or the elevation of each point was randomly assigned a value between 2.03° and 4.93° from the X-axis in 0.29° increments. For each generated bar graph or scatterplot, we computed Pearson's correlation (heights of the bars or points against their left-to-right serial positions), and chose positively sloped graphs with their r-squared values lying between 0.75 and 0.85. For this range of r-squared values, the slopes were accurately judged (as to whether they were negative or positive), but they were not trivially easy to judge (based on pilot results). Negatively sloped bar graphs and scatter plots were generated by mirror reversing the positively sloped ones. On each trial, four graphs of appropriate type (bar graphs or scatterplots with positive or negative slopes) were randomly selected from this larger set, and no graph was presented more than once.

Each bar graph or scatterplot was placed in one of the quadrants of the display, upper-left, upper-right, lower-left, or lower-right. The center of each graph was randomly jittered by $\pm 1.45^\circ$ in the vertical direction with respect to the center of the quadrant. This random jitter prevented horizontal alignment of adjacent graphs (which would have provided a common horizontal reference to those graphs), ensuring that participants needed to evaluate the heights of the bars and points separately within each graph to determine the slopes.

Each array of four graphs was accompanied by either (1) a *continuous* sound consisting of complex tones (59.7 dB SPL_A)¹ continuously ascending or descending in frequency, (2) *discrete* sounds, identical to the continuous sound except that the frequency ascended or descended in 10 successive steps, or (3) *moving-discrete* sounds, identical to the discrete sounds except that the sound source was perceived to move from left to right. Each tone

¹We allowed each participant to select a lower or higher volume setting upon request. For those who chose the lower volume setting, the loudness of the non-moving sound was 56.3 dB SPL_A , whereas for those who chose the higher volume setting, the loudness of the non-moving sound was 62.3 dB SPL_A . The loudness profile of the moving sound was proportionally adjusted.

consisted of the fundamental and two upper harmonics (the harmonics helped to make the tones sound more natural and pleasant than do pure tones). The fundamental frequency was increased from 675 Hz to 1575 Hz in an ascending sound, and decreased from 1575 Hz to 675 Hz in a descending sound. For the continuous sounds, the ascending and descending frequency changes were linear and continuous. For the discrete and moving-discrete sounds, the frequency changes were made in 10 steps (including the first tone) of 100 Hz². For the moving-discrete sounds, the intensities in the left and right ears were modulated as shown in Figure 1. Note that the sound intensity in each ear started at zero so as to avoid uncomfortably abrupt sound onsets. Based on our pilot observation, this modulation generated a compelling impression of a sound moving smoothly from left to right. Each sound stimulus had a total length of 1000 ms, and was presented via headphones³ (Sennheiser HD280 pro).

A group of twelve participants heard the continuous sounds, another group of twelve participants heard the discrete sounds, and the remaining group of twelve participants heard the moving-discrete sounds. Each group also heard a flat (control) sound (with a fundamental frequency of 1031 Hz, the geometric mean of 675 Hz and 1575 Hz for the first group; the fundamental frequency of the flat sound was slightly higher [1087 Hz] for the last two groups due to a technical error).

Procedure—On each trial, participants viewed a visual array consisting of four bar graphs or four scatter plots (see Figure 2a and 2b for an illustration of a trial sequence). Participants were instructed to press the space bar when any of the four graphs contained the target slope (positive or negative), and to refrain from pressing the space bar on target absent trials—a go/no-go procedure. The graph type (bar graph or scatter plot) was blocked. Within each graph type there were four sub-blocks of 60 trials, alternating positive-slope and negative-slope targets. The block order was counterbalanced across participants according to whether the bar graph block or the scatter plot block was presented first and whether each sub-block sequence began with the positive-slope-target block or the negative-slope-target block.

Each trial began with a 500 ms blank screen, followed by an array of four graphs. The graphs were presented concurrently with an ascending, descending, or flat sound (randomly intermixed across trials and equally probable). The search display remained until participants made a response or the trial timed-out after 3000 ms. Participants were told that the sounds were uninformative and that they should be ignored.

Analysis—All response-time analyses reported were based on non-transformed data. Across the four experiments (1, 2, 3A, and 3B), our response-time analyses included 60 means. We conducted a Shapiro-Wilk test of normality for the response-time distribution for each of these means, and 9 cases significantly deviated from normality at $p < 0.05$, which

²Although human pitch perception is known to be logarithmic, for the range of fundamental frequencies used in our ascending and descending sounds, the linear intervals we used minimally deviated from logarithmic ones (Pearson's $r = 0.99$ for linear correlation), and the intervals were perceived to be constant.

³We used headphones because our goal was to examine non-spatial crossmodal interactions between auditory pitch dynamics and visual pattern. Headphones also provided the advantage of shielding ambient sounds. Had we presented sounds through loud speakers placed at the same distance as the visual display (thereby placing auditory and visual stimuli at the same spatial depth), we may have obtained larger crossmodal effects.

was group-wise significant (relative to $3 = 60 \times 0.05$ deviations expected by chance), $\chi^2(1) = 12.63$, $p < 0.001$. To ensure that our results were unaffected by deviations from normality, we also conducted all analyses using log-transformed response time data. Of the 60 transformed distributions, only two significantly deviated from normality at $p < 0.05$, which was group-wise below the chance level of 3 (see above). The results of all analyses remained unchanged after log transformation except for one comparison as noted below. All accuracy analyses were based on arcsine-transformed proportions.

Results and discussion

Response time data (excluding incorrect trials) were analyzed in a $2 \times 2 \times 3 \times 3$ Mixed ANOVA with Graph Type (bar graph or scatter plot), Target Slope (positive or negative) and Crossmodal Congruence (congruent sound, flat sound, or incongruent sound) as within-participant factors, and Sound Condition (continuous sound, discrete sound, or moving discrete sound) as a between-participants factor.⁴ There was no significant main effect of Sound Condition, $F(2, 66) = 0.442$, $p = 0.65$., nor did it interact with any other factors ($F_s < 1.5$), suggesting that whether the sound's pitch was continuously or discretely changed or whether the sound source was static or moved from left to right did not make a difference. There was also no main effect of Graph Type, $F(1, 33) = 0.76$, $p = 0.39$., nor did Graph Type interact with any other factors ($F_s < 1.5$, except in one case described below), suggesting that the sounds generally produced equivalent effects on the speed of target slope detection for both bar graphs and scatter plots.

There was a significant main effect of Target Slope, $F(1, 33) = 13.23$, $p < 0.001$, driven by faster response times for positive-slope targets ($M = 899$ ms) than for negative-slope targets ($M = 923$ ms). This main effect was qualified by a significant Graph-Type-by-Target-Slope interaction, $F(1, 33) = 31.46$, $p < 0.001$, indicating that positive slopes were detected faster than negative slopes for scatter plots ($M = 895$ ms for Positive slopes and $M = 948$ ms for Negative slopes; $t[35] = 7.51$, $p < 0.001$), but not for bar graphs ($M = 905$ ms for Positive slopes and $M = 900$ ms for Negative slopes; $t[35] = 0.60$, $p = 0.55$).

Importantly, there was a significant main effect of Crossmodal Congruence, $F(2, 66) = 15.64$, $p < 0.001$ (Figure 3). This effect was driven by slower response times with the incongruent sounds (i.e., a positive slope paired with a descending sound or a negative slope paired with an ascending sound) compared with both the congruent sounds (i.e., a positive slope paired with an ascending sound or a negative slope paired with a descending sound) ($t[35] = 5.02$, $p < 0.001$) and the flat sound ($t[35] = 4.47$, $p < 0.001$), with the latter two not significantly different from each other ($t[35] = 0.06$, $p = 0.95$).

Overall response accuracy was high (98.7%), though there was a main effect of Crossmodal Congruence, $F(2, 66) = 6.65$, $p < 0.003$; accuracy was slightly but significantly lower with the flat sound (98.3%) compared with both the congruent sounds (98.9%, $t[35] = 3.50$, $p < 0.002$) and the incongruent sounds (98.8%, $t[35] = 2.70$, $p < 0.02$), with the latter two not significantly different from each other ($t[35] = 0.63$, $p = 0.53$). This small congruency effect

⁴The effect of target location, which did not interact with Crossmodal Congruence, is reported in Parrott et al., 2014.

on accuracy (with the flat sound yielding slightly reduced accuracy) cannot explain the response time difference between the congruent-sound and incongruent-sound conditions as the result of a speed-accuracy trade-off. There was also a significant Graph-Type-by-Target-Slope-by-Crossmodal-Congruence interaction, $F(2, 66) = 8.17, p < 0.001$; this 3-way interaction was driven by the greater accuracy for the negative targets (99.2%) than for the positive targets (98.7%) in the congruent-sound condition and the greater accuracy for the positive targets (99.0%) than for the negative targets (98.7%) in the incongruent-sound condition for the bar graphs, with the opposite pattern obtained for the scatter plots, yielding greater accuracy for the positive targets (99.2%) than for the negative targets (98.8%) in the congruent-sound condition and greater accuracy for the negative targets (99.4%) than for the positive targets (98.3%) in the incongruent-sound condition. These accuracy effects are small and do not provide any evidence of a speed-accuracy trade-off for the significant main effect of crossmodal incongruence obtained for response time (Figure 3). No other effects were significant for response time or accuracy (F 's < 2.66).

In addition to producing congruency effects, the ascending (or descending) sound might have directed attention upward (or downward) relative to the flat sound. However, Mossbridge and colleagues (Mossbridge et al., 2011) have shown that, although an ascending (or descending) sound directs attention upward (or downward) along the vertical meridian, it does not direct attention to upper (or lower) locations away from the vertical meridian. Consistent with this prior result, we found no significant interaction between Sound Type (ascending, descending, or flat) and Target Location (upper left, upper right, lower left, or lower right), $F(6, 198) = 1.4, p = 0.22$.

Overall, the results demonstrate an effect of hearing an ascending or descending pitch on the detection of a positive or negative slope in bar graphs and scatter plots. The fact that the manipulation of auditory-visual spatiotemporal similarity had little effect suggests that the effect of crossmodal congruency is driven by a general association between the direction of auditory pitch-change and the direction of visual linear trend.

Experiment 2: Slope search with sequential auditory and visual information

We conducted a follow-up experiment to examine the possibility that the observed effect could be due to target confusion or priming rather than crossmodal interaction. For example, when participants heard a descending sound while looking for a positive slope, they might have been momentarily confused as to which slope they were looking for or might have been primed to look for a negative slope. Although these possibilities are unlikely because participants were always looking for either a positive or a negative slope within a block of trials, we conducted Experiment 2 to evaluate them. We repeated the continuous-sound condition of Experiment 1, except that each sound ended 200 ms before the onset of the visual search display. We reasoned that either a target-confusion effect or a priming effect would persist across a brief gap of 200 ms. In contrast, crossmodal perceptual interactions typically require that auditory and visual stimuli be presented concurrently to be integrated into a single perceptual event (e.g., Holmes & Spence, 2005). If the crossmodal effects obtained in Experiment 1 were due to target confusion or priming, an incongruent sound should still interfere with the visual target response even when the sound immediately

preceded the search display. Alternatively, if the results of Experiment 1 reflect crossmodal interactions, asynchronous presentations of the sounds and visual search displays should abolish the crossmodal effects.

Method

Participants—Twenty new Northwestern undergraduate students gave informed consent to participate for partial course credit. All had normal or corrected-to-normal vision and normal hearing, and were tested individually in a normally lit room.

Stimuli and Procedure—All stimuli and procedures were identical to those used in Experiment 1 in the continuous-sound condition except that each sound ended 200 ms before the visual display; thus, the initial blank screen was lengthened from 500 ms to 1700 ms (i.e., 500 ms [silent] + 1000 ms [sound] + 200 ms [silent]; see Figure 2)).

Results and discussion—Response time data (excluding incorrect trials) were analyzed in a $2 \times 2 \times 3$ Repeated-measures ANOVA with Graph Type (bar graph or scatter plot), Target Slope (positive or negative) and Crossmodal Congruence (congruent sound, flat sound or incongruent sound) as within-participant factors. As in Experiment 1, response times were faster for positive-slope targets than for negative-slope targets, $F(1, 19) = 16.9, p < 0.001$, but unlike in Experiment 1, Target Slope did not significantly interact with Graph Type, $F(1,19) = 0.004, p = 0.95$.

Importantly, there was no significant main effect of Crossmodal Congruence, $F(2, 38) = 0.12, p = 0.89$, suggesting that asynchronously presenting the sounds and visual displays abolished the crossmodal congruence effects (Figure 4). To demonstrate that the congruency effect was uniquely obtained in the auditory-visual synchronous condition (Experiment 1) relative to the auditory-visual asynchronous condition (this experiment), we found a significant Experiment-by-Crossmodal-Congruence interaction, $F(2, 108) = 5.93, p < 0.004$, indicating that an incongruent sound interfered with slope search only when the sound was presented synchronously with the visual display. The overall response time was slower in Experiment 2 than in Experiment 1, $F(1, 54) = 4.09, p < 0.05$, potentially because the simultaneous onset of the sound and the visual display in Experiment 1 may have speeded response times by increasing arousal and/or by providing a temporal cue for the onset of the visual display.⁵ No other main effects or interactions were significant for Experiment 2 (F 's < 3). Response accuracy was overall high (98.4%), and there were no significant effects on accuracy (F 's < 2.03).

The fact that an incongruent sound slowed search for a target slope when presented concurrently with the visual display (Experiment 1), but had no effect when presented immediately before the visual display (Experiment 2) suggests that the incongruent sounds crossmodally interfered with the detection of visual slope rather than producing target confusion or priming. Alternatively, an incongruent sound may have interfered semantically at the response stage. For example, participants might have categorized a positive-slope

⁵This difference was not significant with the log transformation, $F(1,54) = 1.30, n.s.$

target as conceptually positive in an abstract semantic representation before making a target-present response; if so, a descending sound, which may be categorized as conceptually negative, might have interfered at the response stage. To test this possibility, we replaced positive and negative slopes with words “positive” and “negative” (Figure 5). We chose to use these words because the experimenter described the target slopes as positive or negative while giving task instructions in Experiment 1. If the crossmodal interference effect obtained in Experiment 1 was due to semantic incongruence at a response stage, a concurrent ascending sound should slow responses to the target word “negative,” whereas a concurrent descending sound should slow responses to the target word “positive.” In contrast, if an incongruent sound interfered with the detection of visual slope in Experiment 1, the sounds should have no effect on the word search task.

Experiment 3: Word search—semantic control

Experiment 3A: “Positive” or “Negative” as targets

Method

Participants: Twelve new Northwestern undergraduate students gave informed consent to participate for partial course credit. All had normal or corrected-to-normal vision and normal hearing, and were tested individually in a normally lit room.

Stimuli and Procedure: All stimuli and procedures (Figure 5) were identical to those used in Experiment 1 in the continuous-sound condition except for the following. Visual stimuli were presented on a smaller 15” LCD monitor at a 60 Hz refresh rate and 1440 by 900 pixel resolution driven by an Apple MacBook Pro. The viewing distance was approximately 62 cm, so that each pixel subtended approximately 0.021 degrees of visual angle (27% smaller than in Experiment 1). The words, “Positive” (3.28° [horizontal] $\times$ 0.99° [vertical]) and “Negative” ($3.72^\circ \times 0.99^\circ$), were presented in Times New Roman font, size 50. These words replaced the positively and negatively sloped graphs used in Experiment 1. As in Experiment 1, the words were centered in the four quadrants with random jitter in the vertical direction.

Results and discussion: Response time data (excluding incorrect trials) were analyzed in a 2×3 Repeated-measures ANOVA with Target Word (“Positive” or “Negative”) and Crossmodal Congruence (congruent sound, flat sound or incongruent sound) as within-participant factors. There was a marginal main effect of Target Word, $F(1, 11) = 4.80$, $p = 0.051$, reflecting slightly faster responses to “Negative” targets ($M = 738$ ms) than to “Positive” targets ($M = 765$ ms). Importantly, there was no significant main effect of Crossmodal Congruence, $F(2, 22) = 0.94$, $p = 0.41$ (Figure 6), which also did not interact with Target Word, $F(2, 22) = 0.33$, $p = 0.72$.

Response accuracy was overall high (98%). Although there was a significant main effect of Crossmodal Congruence, $F(2, 22) = 5.36$, $p < 0.013$, it reflected slightly greater accuracy with flat sounds (98.4%) than with congruent sounds (97.4%). No other effects on accuracy were significant (F 's < 1).

These results do not support the possibility that the crossmodal incongruence effect obtained in Experiment 1 stemmed from semantic interactions occurring at a response stage. Nevertheless, although the experimenter described the graph slopes as positive or negative, participants might have conceptualized them differently. To determine the likely conceptualization chosen by our participants, in a pilot study we asked fifteen Northwestern students how they would describe the slopes of the graphs used in our study. The majority (9/15) described the slopes as “increasing” or “decreasing,” some (3/15) as “positive” or “negative,” two (2/15) as “up” or “down,” and one (1/15) as sinusoidal. Thus, in the following experiment, we used “Increasing” and “Decreasing” as the stimuli instead of “Positive” and “Negative.”

Experiment 3B: “Increasing” or “Decreasing” as targets

Method

Participants: Fourteen new Northwestern undergraduate students gave informed consent to participate for partial course credit. All had normal or corrected-to-normal vision and normal hearing, and were tested individually in a normally lit room.

Stimuli and Procedure: The stimuli and procedure were the same as those in Experiment 3A except that the words “Increasing” (4.35° [horizontal] $\times$ 0.99° [vertical]) and “Decreasing” ($4.66^\circ \times 0.99$ pixels) were used instead of “Positive” and “Negative.”

Results and discussion: Response time data (excluding incorrect trials) were analyzed in a 2×3 Repeated-measures ANOVA with Target Word (“Increasing” or “Decreasing”) and Crossmodal Congruence (congruent sound, flat sound or incongruent sound) as within-participant factors. There was a main effect of Target Word, $F(1, 13) = 21.80$, $p < 0.001$, reflecting faster responses to “Decreasing” targets ($M = 673$ ms) than those to “Increasing” targets ($M = 729$ ms). Note that responses to “Negative” targets were also faster than those to “Positive” targets in Experiment 3A. These effects likely reflect the fact that the words “Decreasing” and “Negative” were longer than “Increasing” and “Positive,” respectively, in our stimuli. It is known that a longer target is easier to detect among shorter distractors than vice versa (e.g., Treisman & Gormican, 1988).

Importantly, there was no significant main effect of Crossmodal Congruence, $F(2, 26) = 1.16$, $p = 0.33$ (Figure 7), which also did not interact with Target Word, $F(2, 26) = 0.95$, $p = 0.40$. Response accuracy was overall high (98.4%). There was a significant Target-Word-by-Crossmodal-Congruency interaction, $F(2, 26) = 3.40$, $p < 0.05$, in that the accuracies in the congruent-sound (99.0%) and incongruent-sound (98.9%) conditions were both greater than that in the flat-sound condition (98.5%) for “Decreasing” targets, whereas the accuracies in the congruent-sound (98.3%) and incongruent-sound (97.0%) conditions were both less than that in the flat-sound condition (98.7%) for “Increasing” targets. Nevertheless, these accuracy effects are small, and there were no other significant effects on accuracy (F 's < 3.28). Importantly, consistent with the response time result, there was no significant main effect of Crossmodal Congruence on accuracy, $F(2, 26) = 0.39$, $p = 0.68$.

The results from Experiments 3A and 3B taken together have shown that the ascending and descending sounds did not influence a word search whether we used the words “Positive” and “Negative” that the experimenter used to describe the slopes in Experiment 1 (Experiment 3A) or the words “Increasing” and “Decreasing” that people tend to spontaneously use to describe the slopes (Experiment 3B). These results suggest that crossmodal interference effect obtained in Experiment 1 is unlikely to be due to semantic incongruence at a response stage.

General discussion

The present study has demonstrated the existence of a crossmodal association between auditory pitch-change and visual trend such that an ascending sound interferes with the search for negative slopes and a descending sound interferes with the search for positive slopes for both bar graphs and scatter plots. Importantly, these crossmodal effects occurred when sounds were presented concurrently with visual displays (Experiment 1), but did not occur when sounds immediately preceded visual displays (Experiment 2), consistent with a crossmodal interaction rather than a temporally non-specific priming effect. Furthermore, the crossmodal effects did not occur when we presented written words conveying semantic information about graph slopes such as “Positive” and “Negative” (the labels given to participants in Experiment 1) or “Increasing” and “Decreasing” (the terms people tend to spontaneously use to describe graph slopes), providing evidence against the possibility that the crossmodal effects were mediated by amodal semantic representations.

These results are generally consistent with prior results showing that auditory pitch and visual elevation are associated (e.g., Evans & Treisman, 2010; Rusconi et al., 2006), that pitch change and vertical attention shifts are associated (e.g., Mossbridge et al., 2011), and that people can extract the direction and strengths of correlation from an auditory pitch progression over time as well as from a visual scatter plot (Flowers et al., 1997). The current results extend these prior results by demonstrating that ascending and descending pitches automatically interact with visual detection of increasing and decreasing trends even when the sound is task irrelevant. Further, although the increasing and decreasing visual trends were conveyed by discrete elements in our graphs (as in conventional graphs), the use of continuous versus discrete auditory pitch-changes made no difference. Although the visual trends conveyed in our graphs (as in conventional graphs) were to be interpreted as progressing from left to right (as in conventional graphs), there was no difference between the pitch-changing sound moving from left to right and not moving. These results suggest that auditory temporal and visual spatial processes share a common representation of ascending and descending trends that is insensitive to auditory-visual matches in spatiotemporal properties such as discreteness and spatial direction.

It may seem puzzling that an incongruent sound slowed visual slope search but a congruent sound did not speed it. However, such an asymmetry is relatively common in behavioral facilitation and interference effects. For example, a similar asymmetry has been reported for the Stroop effect where an interference effect from an incongruent word (e.g., naming the red ink depicting the word “blue” vs. depicting a neutral symbol) can be much larger than a facilitation effect from a congruent word (e.g., naming the red ink depicting the word “red”

vs. depicting a neutral symbol). Although the origin of this asymmetry is still debated (e.g., Brown, 2011), asymmetric behavioral responses do not necessarily indicate that the underlying inhibitory mechanisms are stronger than the underlying excitatory mechanisms. In fact, symmetric inhibitory and excitatory processes could generate asymmetric behavioral responses due to a sigmoidal function that links the processing of task-relevant information to effector activation (e.g., Cohen, Dunbar, & McClelland, 1990). Specifically, the task-relevant information in the neutral condition may be high because the primary stimulus feature (ink color in a Stroop task and visual slope in our task) may provide sufficient task-relevant information (e.g., Lindsay & Jacoby, 1994) and/or because the neutral condition (neutral symbols for a Stroop task and the flat sound for our task) may provide a near-optimal task context. If so, additional task-congruent information would not speed responses because the effector activation would already be near the floor in the neutral condition, whereas task-interfering information would slow response by shifting the effector activation away from the floor. Thus, although we observed only an interference effect from an incongruent sound, this does not necessarily indicate that the underlying crossmodal interaction is predominantly inhibitory. As with the Stroop effect, additional research will be necessary to determine whether the behavioral asymmetry we observed reflects a predominantly inhibitory effect from an incongruent sound, or whether congruent and incongruent sounds produce equivalent excitatory and inhibitory effects on visual slope processing while the asymmetry arose due to factors such as nonlinearity linking stimulus processing to response time.

In summary, we have demonstrated that ascending and descending sounds influence visual search for positive and negative slopes in both bar graphs and scatter plots. These crossmodal interactions require concurrent presentations of visual and auditory stimuli, but are insensitive to auditory-visual matches in spatiotemporal properties and are unlikely to be driven by semantic-level interactions. These results suggest that the auditory-temporal processing of pitch direction and the left-to-right visual-spatial processing of linear trend interact at a common representation of ascending and descending trends.

Acknowledgements

This study was supported by NIH grant R01 EY021184.

References

- Brainard DH. The Psychophysics Toolbox. *Spatial Vision*. 1997; 10:433–436. [PubMed: 9176952]
- Brown TL. The relationship between Stroop interference and facilitation effects: statistical artifacts, baselines, and a reassessment. *Journal of Experimental Psychology: Human Perception and Performance*. 2011; 37(1):85–99. [PubMed: 20718567]
- Cohen JD, Dunbar K, McClelland JL. On the control of automatic processed: A parallel distributed processing account of the Stroop effect. *Psychological Review*. 1990; 97:332–361. [PubMed: 2200075]
- Connell L, Cai ZG, Holler J. Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and cognition*. 2013; 81(1):124–130. [PubMed: 23195703]
- Evans KK, Treisman A. Natural cross-modal mappings between visual and auditory features. *Journal of Vision*. 2010; 10(1):6. [PubMed: 20143899]

- Flowers JH, Buhman DC, Turnage KD. Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. *Human Factors*. 1997; 39(3):341–351. [PubMed: 9394628]
- Grant ER, Spivey MJ. Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*. 2003; 14:462–466. [PubMed: 12930477]
- Hadjidemetriou C, Williams JS. Children's graphical conceptions: Assessment of learning for teaching. *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education*. 2001; 3:89–96.
- Holmes NP, Spence C. Multisensory integration: space, time and superadditivity. *Current Biology*. 2005; 15:762–764.
- Lindsay DS, Jacoby LL. Stroop process dissociations: the relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*. 1994; 20(2):219–234. [PubMed: 8189189]
- Maeda F, Kanai R, Shimojo S. Changing pitch induced visual motion illusion. *Current Biology*. 2004; 14(23):R990–R991. [PubMed: 15589145]
- Mossbridge JA, Grabowecky M, Suzuki S. Changes in auditory frequency guide visual-spatial attention. *Cognition*. 2011; 121(1):133–139. [PubMed: 21741633]
- Parise CV, Knorre K, Ernst MO. Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*. 2014; 111(16):6104–6108.
- Parrott S, Guzman-Martinez E, Ortega L, Grabowecky M, Huntington MD, Suzuki S. Spatial position influences perception of slope from graphs. *Perception*. 2014; 43:647–653. [PubMed: 25223108]
- Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*. 1997; 10:437–442. [PubMed: 9176953]
- Pinker, S. A theory of graph comprehension. In: Freedle, R., editor. *Artificial Intelligence and the Future of Testing*. Hillsdale, NJ: Erlbaum; 1990. p. 73-126.
- Pratt CC. The spatial character of high and low tones. *Journal of Experimental Psychology*. 1930; 13(3):278.
- Roffler SK, Butler RA. Factors that influence the localization of sound in the vertical plane. *The Journal of the Acoustical Society of America*. 1968; 43(6):1255–1259. [PubMed: 5659493]
- Rusconi E, Kwan B, Giordano BL, Umiltà C, Butterworth B. Spatial representation of pitch height: the SMARC effect. *Cognition*. 2006; 99(2):113–129. [PubMed: 15925355]
- Treisman A, Gormican S. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*. 1988; 95:15–48. [PubMed: 3353475]
- Tversky, B. Visuospatial reasoning. In: Holyoak, K.; Morrison, R., editors. *The Cambridge Handbook of Thinking and Reasoning*. Cambridge: Cambridge University Press; 2005. p. 209-241.
- Walker R. The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Perception & Psychophysics*. 1987; 42(5):491–502. [PubMed: 2447557]

Figure 1. Modulations of sound intensity in the left ear (dashed curve) and right ear (solid curve) used for the moving-discrete sound, generating a compelling impression of the sound moving smoothly from left to right.

Figure 2.

Trial sequence for the slope search task with (a) bar graphs or (b) scatter plots (graph type was blocked). Target slope (positive or negative) was also blocked. In a block of positive-slope target trials (shown here), participants pressed the space bar (on "go" trials) when they detected a positive slope, but otherwise refrained from making a response (on "no-go" trials). In a block of negative-slope target trials (not shown), participants pressed the space bar when they detected a negative slope, but otherwise refrained from making a response. The sound was presented either concurrently with the visual search display (Experiment 1: AV concurrent) or immediately before the visual display (Experiment 2: AV sequential).

Figure 3.

Results from Experiment 1: Mean response times for the slope search task when the sound was presented concurrently with the visual search display. The data have been averaged across target slope (positive or negative), graph type (bar graph or scatter plot) and sound condition (continuous, discrete, or moving-discrete). The error bars represent ± 1 standard error of the mean adjusted for within-participant comparisons. *** indicates, $p < 0.001$.

Figure 4.

Results from Experiment 2: Mean response times for the slope search task when the sound was presented immediately before the visual search display. The data have been averaged across target slope (positive or negative) and graph type (bar graph or scatter plot). The error bars represent ± 1 standard error of the mean adjusted for within-participant comparisons.

Figure 5.

Trial sequence for the word search task (Experiment 3). Target word, “Positive” or “Negative,” was blocked. In a block of positive-word target trials (shown here), participants pressed the space bar (on “go” trials) when they detected a word “Positive,” but otherwise refrained from making a response (on “no-go” trials). In a block of negative-word target trials (not shown), participants pressed the space bar when they detected a word “Negative,” but otherwise refrained from making a response. The sound was presented concurrently with the visual search display.

Figure 6.

Results from Experiment 3A: Mean response times for the word search task when the sound was presented concurrently with the visual search display. The data have been averaged across target words (“Positive” or “Negative”). The error bars represent ± 1 standard error of the mean adjusted for within-participant comparisons.

Figure 7.

Results from Experiment 3B: Mean response times for the word search task when the sound was presented concurrently with the visual search display. The data have been averaged across target words (“Increasing” or “Decreasing”). The error bars represent ± 1 standard error of the mean adjusted for within-participant comparisons.