Cognition 208 (2021) 104529

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Training enhances the ability of listeners to exploit visual information for auditory scene analysis

Huriye Atilgan [1], Jennifer K. Bizley [*]

The Ear Institute, University College London, UK

ARTICLE INFO

Keywords:
Audiovisual integration
Selective attention
Auditory scene analysis
Temporal processing
Training

ABSTRACT

The ability to use temporal relationships between cross-modal cues facilitates perception and behavior. Previously we observed that temporally correlated changes in the size of a visual stimulus and the intensity in an auditory stimulus influenced the ability of listeners to perform an auditory selective attention task (Maddox, Atilgan, Bizley, & Lee, 2015). Participants detected timbral changes in a target sound while ignoring those in a simultaneously presented masker. When the visual stimulus was temporally coherent with the target sound, performance was significantly better than when the visual stimulus was temporally coherent with the masker, despite the visual stimulus conveying no task-relevant information. Here, we trained observers to detect audiovisual temporal coherence and asked whether this changed the way in which they were able to exploit visual information in the auditory selective attention task. We observed that after training, participants were able to benefit from temporal coherence between the visual stimulus and both the target and masker streams, relative to the condition in which the visual stimulus was coherent with neither sound. However, we did not observe such changes in a second group that were trained to discriminate modulation rate differences between temporally coherent audiovisual streams, although they did show an improvement in their overall performance. A control group did not change their performance between pretest and post-test and did not change how they exploited visual information. These results provide insights into how crossmodal experience may optimize multisensory integration.

1. Introduction

Integrating information across sensory modalities enables the brain to benefit from both redundant and complementary information. For example, being able to see a speaker's face provides both phonetic information (Grant, Walden, & Seitz, 1998; Summerfield, 1992) and grouping cues (Helfer & Freyman, 2005) that provide a benefit for speech comprehension in noisy environments. Benefiting from multisensory integration requires that the brain appropriately link signals across modalities (Lee, Maddox, & Bizley, 2019; Shams & Beierholm, 2010). Auditory and visual signals arrive and are processed with different latencies. Consequently, cross-modal signals can be perceived as synchronous across a range of onset asynchronies – known as the temporal binding window (Dixon & Spitz, 1980; Meredith, Nemitz, & Stein, 1987). Previous studies have demonstrated that there is short term plasticity in this window (Megevand, Molholm, Nayak, & Foxe, 2013; Navarra et al., 2005; Schormans & Allman, 2018; Vroomen, Keetels, DE Gelder, & Bertelson, 2004; Zmigrod & Zmigrod, 2015) and that experience and longer-term training can narrow this window such that listeners more accurately judge synchronous from asynchronous stimuli (Bidelman, 2016; Dixon & Spitz, 1980; Lee & Noppeney, 2011; Powers, Hillock, & Wallace, 2009).

Training listeners to optimize multisensory integration could provide rehabilitation to hearing impaired listeners and enable vision to augment auditory processing in noisy environments. However, training listeners to refine their temporal binding window has been observed to have varied consequences for multisensory integration. For example, training on an audiovisual temporal discrimination task led to a narrowing the temporal binding window (but did not narrow participants' spatial binding window). Training also led to a general decrease in the likelihood of cross-modal interactions across all temporo-spatial disparities, as indexed by spatial ventriloquism (Mcgovern, Roudaia, Newell, & Roach, 2016). In contrast, training that improved visual temporal discrimination abilities, did not influence the likelihood of

* Corresponding author.
E-mail address: j.bizley@ucl.ac.uk (J.K. Bizley).
[1] Current address: Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut, U.S.A.

https://doi.org/10.1016/j.cognition.2020.104529
Received 16 April 2020; Received in revised form 24 November 2020; Accepted 25 November 2020
Available online 26 December 2020

perceiving sound-induced flash illusions (Stevenson, Wilson, Powers, & Wallace, 2013). Finally, in another study in which listeners were trained to discriminate asynchronous from synchronous stimuli, listeners were subsequently shown to have stronger spatial ventriloquism effects when auditory-visual signals were temporally synchronous but spatially separated (Sürig, Bottari, & Röder, 2018).

A problem in interpreting these varied effects is that many lab-based tasks do not encompass the complexity that the brain faces in real-world situations. In most lab-based paradigms, observers often judge single audio and visual signals, presented in an otherwise quiet and dark environment. In contrast, in the world, the brain must match one of several competing sounds to a given visual object (or vice versa). Moreover, due to the variance in the timing of real-world signals, simply narrowing the window over which integration occurs may be a suboptimal strategy for effective information integration. Rather, what observers need to do is detect whether temporal coherence exists between signals in different modalities so that they may be appropriately grouped (Lee et al., 2019).

An additional consideration for training studies that focus on the temporal binding window is that it is unclear to what extent any adjustments in the temporal binding window extend to other multisensory processing tasks. In many cases, the task used to train observers is the same one used to measure the temporal binding window raising the question of how generalizable results are and whether they represent a genuine change in cross-modal binding, or whether listeners are simply shifting an internal criterion in order to improve their performance in this one task (Bizley, Maddox, & Lee, 2016; Lee et al., 2019; Powers et al., 2009; Setti et al., 2014).

In this study, our goal was to examine whether we could train listeners to improve their ability to detect audiovisual temporal coherence and, in doing so, whether this would improve their ability to use visual cues to appropriately group sound elements from one stream and separate them from those elements in a competing sound. To train listeners to detect audiovisual temporal coherence, we asked listeners to differentiate streams in which audio and visual stimuli were amplitude/radius modulated in a statistically independent manner from stimuli in which audio and visual elements maintained some degree of temporal coherence. We elected to train listeners to detect small amounts of correspondence (rather than incoherence) as we reasoned that detecting moments of genuine correspondence is more likely to be useful than detecting transient incoherence, both in this current task and in solving real-world binding problems.

In addition to measuring the ability of observers to assess temporal coherence before and after training (using similar stimuli to those used in training), we used the auditory selective attention task from Maddox et al. (2015) to assess how effectively listeners could utilize visual information during the performance of an auditory selective attention task. This task required participants to focus on one of two competing auditory streams and report brief timbre perturbations within the target stream. They also watched a visual stimulus whose radius could change in a manner that was temporally coherent with either the target, the masker, or neither auditory stream (but was never predictive about the timing of the timbre perturbations). In Maddox et al., we reported that the visual coherence condition significantly influenced performance in the auditory selective attention task such that performance was better when the visual stimulus was coherent with the target audio stream than when it was coherent with the masker stream.

In order to determine whether any training effects we observed were critically dependent on improved temporal coherence detection, as opposed to passive exposure to temporally coherent auditory-visual streams, we also trained another group of observers in an amplitude modulation rate discrimination task with the same temporally coherent audiovisual stimuli. A control group simply performed the pretest and post-test without any training. We hypothesized that an improved ability to detect temporal coherence might enable listeners to appropriately group temporally coherent audiovisual streams, which in turn would promote more effective auditory selective attention. Our results support this hypothesis and demonstrate that only listeners trained to detect audiovisual temporal coherence change the way in which they are able to use visual information to augment auditory scene analysis.

2. Methods

2.1. Participants

42 adults (age range 18–34 years; mean age 28 years; 11 males) with normal hearing and normal or corrected-to-normal vision, participated in the study. Six participants were excluded after the pretest due to poor performance (mean $d' < 0.8$, $n = 4$), or low visual hit rates (indicating inattention, <70%, $n = 2$). The remaining 36 participants were included for further analysis and were randomly allocated to 3 groups (12 listeners per group). The study was approved by the Ethics Committee of the University College London (ref: 5139) and all procedures performed were in accordance with the Declaration of Helsinki. All individuals were paid for their participation and signed an informed consent form before participation.

2.2. Testing procedure

We recruited participants and randomly assigned them to one of three groups, each performing a pretest and a post-test. Pretests and post-tests comprised the timbre variant of the selective attention task in Maddox et al. (2015) and an AV temporal coherence detection threshold test. In between the pretest and post-test, one group trained on an AV temporal coherence detection task (AV coherence training, $n = 12$), one group trained on an amplitude modulation rate discrimination task using temporally coherent AV stimuli (AV modulation group, $n = 12$), and a third group simply performed the pretest and post-test separated by a minimum of 5 days (control, n = 12; Fig. 1A). The pretest and post-test both took approximately 90 min. Participants in the two training groups performed 5 training sessions (each lasting not more than 40 min) on 5 separate days over not more than 2 weeks (Fig. 1A).

2.3. Stimuli and task design

2.3.1. Auditory selective attention task

The auditory selective attention task required that listeners attend to one of two competing auditory streams and report the presence of brief (200 ms) timbre perturbations in the target audio stream. They were additionally required to monitor a visual stimulus whose radius changed in time. The two audio streams were independently amplitude modulated and the visual radius was modulated with a time course that matched one or the other auditory stream or was independent of them both (Fig. 1B). Envelopes for the visual envelope and auditory amplitude were created using the same frequency domain synthesis. For each trial, an envelope was created by first setting all amplitudes of frequency bins above 0 Hz and below 7 Hz to unity and others to zero. At an audio sampling rate of 24,414 Hz, all non-zero bins were given a random phase from a uniform distribution between 0 and 2π, the corresponding frequency bins across Nyquist frequency were set to the complex conjugates to maintain Hermitian symmetry, and the inverse Fourier transform was computed yielding a time domain envelope. Second and third envelopes were created using the same method and orthogonalized using a Gram-Schmidt procedure. Each envelope was then normalized so that it spanned the interval [0,1] and then sine-transformed [$y = \sin^2(\pi x/2)$] so that the extremes were slightly accentuated. Visual envelopes were created by subsampling the auditory envelope at the monitor frame-rate of 60 Hz, starting with the first auditory sample so that auditory amplitude corresponded with the disc radius at the beginning of each frame.

Stimuli were presented in an unlit sound-attenuating room over headphones (HD 555, Sennheiser, Wedemark, Germany). Participants

were seated 60 cm from the screen with their heads held stationary by a chinrest. Auditory stimuli were created in MATLAB and presented using an RP2 signal processor (Tucker–Davis Technologies, Alachua, FL, USA). Each began and ended with a 10 ms cosine ramp. All stimuli were presented diotically. Visual stimuli were synthesized in MATLAB (The Mathworks, Natick, MA, USA) and presented using the Psychophysics Toolbox (Brainard, 1997). The visual stimuli were gray discs that subtended between 1° and 2.5° at the center of the computer screen. The white ring extended 0.125° beyond the gray disc.

The auditory stimuli were generated as described in the *timbre* variant of Maddox et al. (2015). On each trial two audio streams were presented (a target and a distractor), each of which had a distinct pitch (F0 = 175 or 195 Hz) and timbre (/u/ or /ɛ/). Across trials both vowels could take either pitch value, and pitch-timbre combination of target and distractor streams was fully counterbalanced. Each auditory stream was generated as a periodic impulse train and then filtered with synthetic vowels simulated as four-pole filters (formants F1–F4). The /u/ stream had formant peaks F1–F4 at 460, 1105, 2857, 4205 Hz and moved slightly towards /ɛ/ during timbre events, with formant peaks at 730, 2058, 2857, 4205 Hz. The /a/ stream had formant peaks F1–F4 at 936, 1551, 2975, 4263 Hz and moved slightly towards /i/ during timbre events, with formant peaks at 437, 2761, 2975, 4263 Hz. During timbre events, the formants moved linearly towards the deviant for 100 ms and then linearly back for 100 ms. Streams were calibrated to be 65 dB SPL (RMS normalized) using an artificial ear (Brüel & Kjær, Nærum, Denmark) and presented against a low level of background noise (54 dB SPL). Unlike Maddox et al., we did not assess individual timbre discrimination thresholds but instead used a fixed level of difficulty determined using the average individual thresholds measured previously. For [e] deviants in [u] stimuli, this corresponded to a shift of 42 Hz in F1 frequency and 143 Hz for F2, and for [i] deviants in [a] stream there was a shift of 75 Hz for F1, 196 Hz for F2.

Trials lasted 14 s. They began with only the target auditory stimulus and the visual stimulus, indicating the to-be-attended (target) auditory stream to the participant. The to-be-ignored auditory stream (masker) began 1 s later. As with the rest of the trial, the visual stimulus was only coherent with the auditory target during the first second if it was a match-target trial. All streams ended simultaneously. Events did not occur in the first two seconds (i.e. 1 s after the masker began) or the last 1 s of each trial, or within 1.2 s of any other events in either modality. A response made within 1 s following an event was attributed to that event. To ensure audibility and equivalent target to masker ratios without providing confounding information to the participants, an event in either auditory stream or the visual stream could only begin when both auditory envelopes were above 70% maximum. There were between 1 and 3 inclusive events (mean events = 2) in both the target and masker in each trial.

There were also between 0 and 2 inclusive visual flashes per trial (mean flashes = 1), in which the outer ring changed from white to cyan (0% red, 100% blue, 100% green) and back. Participants were also asked to report the colour change by pressing the button to make them watch the visual stimuli attentively while detecting the deviants in the auditory stream. Each participant completed 32 trials of each temporal coherence conditions (96 totals), leading to 64 potential hits and 64 potential false alarms for each condition (i.e., 128 responses considered for each d′ calculation) as well as 32 visual flashes per condition. When computing d′, auditory hit and false alarm rates were calculated by adding 0.5 to the numerator and 1 to the denominator so that d′ had finite limits. This task was used as a pretest and post-test for all three experimental groups (Fig. 1A).

2.3.2. Auditory-visual temporal coherence detection test

A two-interval forced-choice detection test was used to determine perceptual thresholds for detecting AV temporal coherence. In one stimulus interval, the sound was accompanied by a visual stimulus in which the radius changed over time independently of the auditory stimulus. In the other interval, the auditory and visual stimulus maintained some degree of temporal coherence (Fig. 1C). Auditory and visual stimuli were generated as described in the auditory selective attention task above, with a single auditory and visual stream presented on each occasion. The pitch and timbre of stimuli were varied across trials such that the auditory stimuli were either [u] or [e] (without any timbre deviants embedded), with F0 = 175 or 195 Hz, counterbalanced. The sounds in both stimulus intervals within the trial had identical pitch and timbre values and had a duration of 5 s. The method of constant stimuli was used to determine the threshold with participants performing 20 trials at each coherence level. AV stimuli were generated from 10% coherent in 10% steps to 100% coherent by multiplying the temporally coherent envelope with an independent envelope. Participants were required to select the interval (by pressing 1 or 2 on a button box) in which the temporally coherent pair was presented. Feedback was provided on every trial.

2.3.3. Auditory-visual temporal coherence detection training (AV coherence training)

The stimuli and procedure in the AV coherence training were identical to those used in the threshold test, but with an adaptive three-down one-up rule to determine the coherence level of the stimulus in the next trial. Previous work has demonstrated that task difficulty is an important aspect in driving multisensory learning (De Niear, Koo, & Wallace, 2016), so by using this approach, we required that participants worked near to their threshold for a large proportion of the training session. As in the threshold test, participants performed a two-interval forced choice task, and were asked to select the more coherent interval. The first training session started from the most distinguishable stimuli pairs; one interval had 100% temporally coherent audiovisual streams and the other interval was fully independent (i.e. 0% coherent). For the first 6 reversals, coherency was decreased in 10% steps followed by 5% steps for the following six reversals and by 2.5% steps for the remainder. The procedure was terminated at 18 reversals unless a maximum of 150 trials was reached first. For the 2nd-5th training session, the first "coherent" stimulus was generated with the average coherence level of the last ten reversals in the previous session. Each training session lasted less than 40 min. Feedback was provided on every trial.

2.3.4. Amplitude modulation rate discrimination training (AV modulation training)

For the AV modulation training, participants performed a two-interval forced-choice task in which one interval always had modulation envelope with a 7 Hz cut off rate, whereas the other was generated with a higher rate (maximum AM cut off rate = 11 Hz). Participants detected the "faster" interval. The audiovisual stimuli in both intervals were fully temporally coherent, each was 5 s long, with a constant pitch and timbre within a trial, counterbalanced across trials, Fig. 1D). In the sessions of AM rate training, an adaptive three-down one-up rule was used to determine the AM rate of the stimulus in the next trial. In the first session, the first stimulus was generated at the maximum AM rate and differed in AM rate by 1 Hz for the first six reversals and 0.5 Hz for the next six reversals and 0.25 Hz for the rest of the trials. The procedure was terminated at 18 reversals unless a maximum of 150 trials was reached first. In each consequent session, the first stimulus was generated with the average coherence level of the last ten reversals in the previous session. Participants pressed "1" or "2" on the press box to indicate the interval of the faster AV pair. Feedback was provided on each trial.

2.4. Statistical analysis

Statistics were performed using MATLAB (2011b, Mathworks, USA) and SPSS (IBM). The d′, hit rates, false alarm, and visual hit rates across AV coherence conditions and pretest versus post-test were calculated. Visual hit rates were calculated to ensure that participants were

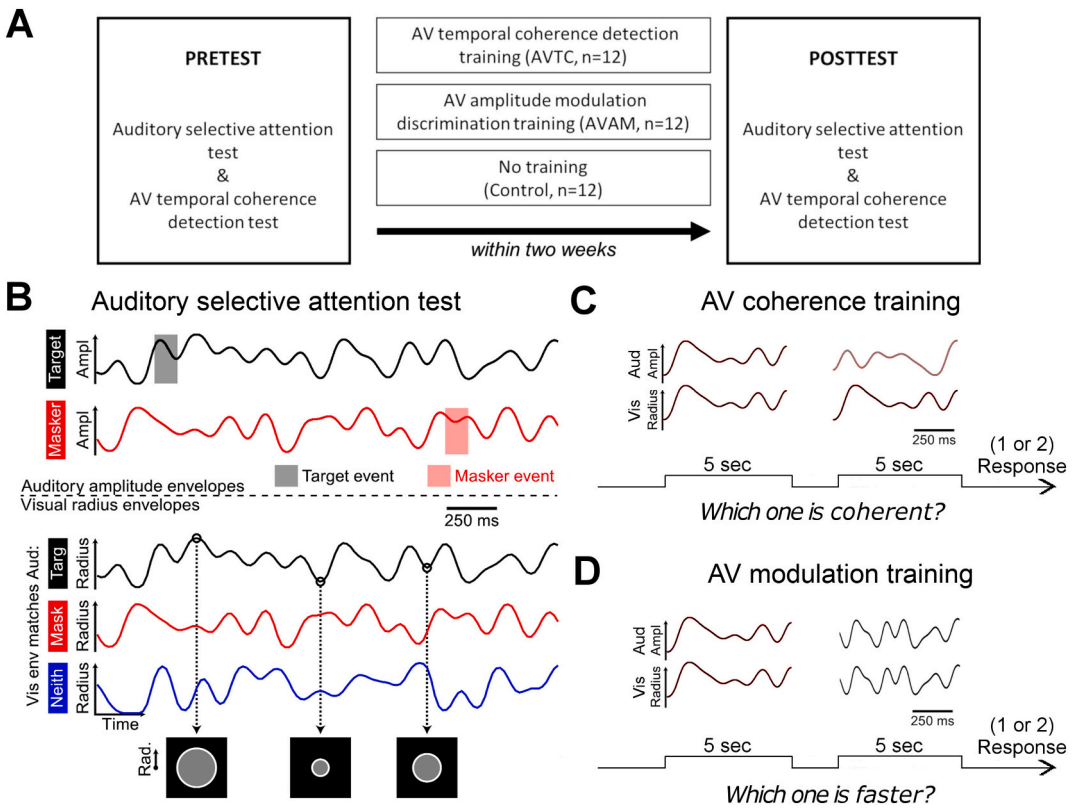

Fig. 1. **A** Experimental design. **B** Schematic representation of auditory and visual stimuli used in the auditory selective attention test (panel was taken from Maddox et al., 2015). Amplitude envelopes of target (black) auditory stream and masker (red) auditory stream and visual radius envelopes for three auditory visual (AV) coherence conditions; target coherent (black), masker coherent (red) and neither (blue). Examples frames of the visual stimuli at three radius level. **C** Schematic representation of AV temporal coherence detection test/AV coherence training. Two 5 seconds AV pairs were used. One maintained some degree of temporal coherence (left, here fully coherent) while the other was always fully independent (right) **D** Schematic representation of AV amplitude modulation rate discrimination (AV modulation) training. Two 5 second temporally coherent AV pairs were used. Each pair had a different modulation envelope (and rate) but stimuli were always temporally coherent across modalities.

attending the visual stimuli, and participants with a hit rate of <70% were excluded. Statistical significance across groups was assessed by two-tailed unpaired Student's *t*-tests, Mann-Whitney *U* tests, one-way analysis of variance (ANOVA), mix ANOVA or repeated measures of ANOVA where appropriate. Mauchly's test of sphericity was used for the assumption of homogeneity of variance for independent tests in the repeated measured analysis if not otherwise reported. Significant main effects or interactions were followed up with post-hoc testing using Bonferroni corrections where applicable. Significance was declared at $p < 0.05$, with a precise *p*-value stated in each case, and all tests were two-sided. Partial eta squared was reported to indicate the effect size. For individual differences in AV temporal coherence detection thresholds, 95% confidence intervals were calculated with linear regression.

2.5. Data availability

Data are available under CCBY 4 license (Atilgan & Bizley, 2020).

3. Results

3.1. Training was effective at improving performance

We first confirmed that both trained groups improved their ability on the trained stimulus feature. Fig. 2A and B show the thresholds derived from the last 5 reversals for the training session on day 1 and day 5 for the AV coherence training group and the AV modulation training group respectively. For both groups, thresholds were significantly lower for session 5, than for session 1 (pairwise *t*-test on S1 and S5, AV coherence training thresholds, $t_{11} = 2.961$, $p = 0.007$; AV modulation training thresholds: $t_{11} = 4.529$, $p < 0.001$).

3.2. Training affected performance in the auditory selective attention task

We calculated hit rates, false alarms and d′ values for all listeners in the auditory selective attention task across all visual coherence conditions and both sessions (Fig. 3A-C). To determine whether training led to

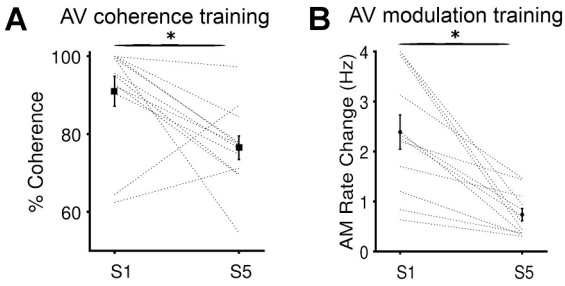

Fig. 2. Training improved performance in trained tasks and the Auditory Selective Attention task. **A** Training in audiovisual temporal coherence detection (AV coherence training) task was effective at driving an improvement in the coherence detection threshold between session (S1) and session 5 (S5). Black vertical lines show the mean ± SEM across participants. Gray lines are individual participants. **B** Training in audiovisual amplitude modulation (AV modulation training) was effective at driving an improvement in AM rate discrimination between S1 and S5.

changes in performance in the auditory selective attention task or in the ability of listeners to utilize visual information we ran a 3x3x2 mixed ANOVA for d', with a between-subjects factor of experimental groups (AV coherence training, AV modulation training and control) and within-subjects factors of session (pretest and post-test) and audiovisual coherence condition (target coherent, masker coherent, neither, data available as source data 1, for full statistical results, see Table 1).

The three factor ANOVA revealed a significant effect of session and visual coherence condition but not experimental group, and a significant three-way interaction between all three factors. Post-hoc tests examining the effect of coherence condition demonstrated that performance in the target-coherent condition was significantly better than both masker-coherent and independent conditions, replicating our previous findings (Maddox et al., 2015).

Since there was a significant three-way interaction between group, session and visual coherence condition, we separated the data according to the training group in order to perform further analysis (Fig. 3). To better understand the interaction between visual coherence condition, session and training group we also calculated the normalized d′ values (Fig. 3D-F, note: all statistical comparisons are performed only on the untransformed values). Normalized d′ distills the impact of the visual coherence condition by taking the difference between each condition and the across-condition mean. This effectively removes the effect of the absolute level of performance, which varies substantially across participants, as well as any difference in overall performance between pretest and post-test. Normalized d′prime allowed us to compare three distinct hypotheses. First, we hypothesized that if training did not change the way in which listeners used visual information, we would see the same across-coherence-condition pattern, simply shifted up to higher d′ values. In this situation the normalized d′ values would be unchanged by training as the increase in d′ was uniform across coherence condition. Second, if there was a change in the magnitude of the visual stimulus induced effects we predicted a similar pattern of across-coherence-condition performance, but a larger difference between target and masker coherent conditions. This would be reflected in normalized d′ measures being larger in magnitude, but equivalent in sign between pretest and post-test. Third, if training influenced the way in which listeners utilized visual information, we predicted that there would be a change in the pattern of across-coherence-condition d′ values, and a change in the sign of the normalized values.

Table 1
The results of 3x2x3 way ANOVA for within subject effects of coherence condition and session, and between subjects effects of experimental groups for d′ values with post-hoc pairwise multiple comparison corrected with Bonferroni ($p < 0.05$ in bold). Mauchly's test of sphericity was significant (<0.05); Greenhouse-Geisser corrected values are reported.

	Between/ Within Subject Effects			Pairwise multiple comparison	
	F	p	η^2		p
Session	35.411	**0.000**	0.518		
Exeriment Group	2.725	0.368	0.059		
Coherence Condition	11.052	**0.002**	0.251	Target vs Masker	**0.000**
				Target vs Independent	**0.002**
				Masker vs Independent	0.211
Session Experimental group int.	2.858	0.072	0.148		
Session Coherence condition int.	1.347	0.254	0.039		
Experimental group Coherence condition int.	1.927	0.162	0.105		
Experimental group Coherence condition Session int.	4.159	**0.024**	0.201		

For each group, we performed a two-way repeated measure within-subjects ANOVA with factors of coherence condition (target coherent, masker coherent and neither) and session (pretest and post-test; Table 2). In this framework, we expect a significant coherence condition effect in all groups and training effects in the AV coherence and AV modulation training group. Of particular interest is the interaction term, as this would indicate a training-induced change in the way in which visual information impacted performance.

3.3. AV coherence training improves performance in the auditory selective attention task and alters the way in which listeners use visual cues

In the AV coherence training group, participants trained on a task in which they were actively judging the temporal coherence of auditory and visual streams. Prior to training, participants showed the expected effect of coherence condition (i.e. target coherent > masker coherent). However, after training a different pattern was observed: both target and masker coherent conditions were superior to the neither condition in which the visual stimulus changed independently (Fig. 3A). Notably, the normalized d′ measures for the masker coherent condition changed from negative to positive in this group (Fig. 3D). Consistent with these observations a repeated ANOVA analysis of d′ scores revealed significant effect of session ($F(1,11) = 36.245$, $p < 0.001$), a borderline effect of coherence condition ($F(2,22) = 6.908$, $p = 0.05$) and a significant interaction (Fig. 3A, D, G; $F(2, 22) = 7.258$, $p = 0.002$; see also Table 2 which reports effect sizes). Post-hoc comparisons across AV coherence condition in the pretest data revealed that participants performed better when the visual stimulus was coherent with the target auditory stream versus the masker auditory stream (target coherent > masker coherent). In contrast, post-hoc comparisons of the post-test d′ scores revealed that, after training, performance was better when the visual stimulus was coherent with either the target or the masker stream than in the condition in which neither audio stream was coherent with the visual stimulus (target coherent > neither, masker coherent > neither, Bonferroni corrected post-hoc comparison $p < 0.05$).

3.4. AV modulation training enhances performance in the auditory selective attention task but does not change how listeners utilize visual information

Participants in the AV modulation training group were asked to detect the amplitude modulation rate of temporally coherent AV pairs. They were not actively detecting temporal coherence, but passively exposed to temporally coherent AV pairs. Although training improved their overall performance in the auditory selective attention task, the way in which they used visual information appeared unchanged after training. In both pretest and post-test performance was best in the target coherent condition (Fig. 3B) and the normalized d′ measures, which effectively factor out the overall d′ improvement, were overlapping (Fig. 3E) suggesting that there was no change in the way in which the coherence condition effected performance (Fig. 3H). Both session ($F(1, 11) = 23.134$, $p = 0.001$) and coherence condition ($F(2, 22) = 5.723$, $p = 0.010$) influenced d′, but – importantly – there was no interaction ($F(2, 22) = 0.09$, $p = 0.854$). Post-hoc comparisons ($p < 0.05$) across coherence conditions revealed that participants performed better when the visual stimulus was coherent with the target auditory stream compared to the masker coherent condition in both the pretest and post-test (target coherent > masker coherent). Therefore, this suggests an overall improvement in performance after AV modulation training, but no change in the way in which observers were able to exploit visual cues.

3.5. Control group performance was unchanged between pretest and post-test

Performance in the control group did not differ significantly between pretest and post-test (Fig. 3C, F, I): there was no effect of session (F

Table 2
The results of two-way repeated measures within-subjects ANOVA for each variable ($p < 0.05$ in bold) for d', hit rates, false alarm and bias for three experimental groups.

	Coherence conditions			Session (pre vs post)			Interaction between coh. condition & session		
	F	p	η^2	F	p	η^2	F	p	η^2
AV coherence training group									
d'	6.908	**0.050**	0.386	36.245	**<0.001**	0.767	7.258	**0.002**	**0.493**
Hit Rates	5.080	**0.015**	0.316	7.731	**0.018**	0.413	4.660	**0.021**	**0.298**
*False Alarm**	2.692	0.115	0.197	17.164	**0.002**	0.609	2.555	0.100	0.189
Bias	1.233	0.311	0.101	0.128	0.727	0.110	3.005	0.070	0.215
AV modulation training group									
d'	5.723	**0.010**	0.342	23.134	**0.001**	0.678	0.009	0.854	0.014
Hit Rates	2.952	**0**.073	0.212	24.148	**<0.001**	0.687	0.386	0.682	0.042
*False Alarm**	2.317	0.145	0.174	5.846	**0.034**	0.347	0.376	0.655	0.033
*Bias**	1.909	0.190	0.148	3.337	0.095	0.233	0.922	0.411	0.077
Control									
d'	4.653	**0.021**	0.297	1.431	0.257	0.115	0.039	0.962	0.004
Hit Rates	0.998	0.385	0.083	1.180	0.301	0.097	0.065	0.929	0.006
False Alarm	3.023	0.069	0.216	0.088	0.772	0.008	0.253	0.779	0.022
Bias	0.301	0.743	0.027	0.084	0.777	0.008	0.043	0.958	0.004

* Mauchly's test of sphericity was significant (<0.05), Greenhouse-Geisser corrected values are reported.

(1,11) = 1.431, $p = 0.257$), but there was a significant effect of coherence condition (F (2, 22) = 4.653, $p = 0.021$), with no interaction (F (2, 22) = 0.039, $p = 0.962$). Participants performed significantly better when the visual stimulus was coherent with the target auditory stream versus the masker auditory stream (target coherent > masker coherent) in pretest and post-test (Fig. 3F).

3.6. Changes in performance are driven by increased hit rates in the masker-coherent condition and decreased false alarm rates

To better understand the effect in the AV coherence training group we considered the hit rates and false alarm rates (which together define d′) to determine whether the changes were principally driven by an improved ability to detect the target timbre deviations in the masker coherent condition, or an improved ability to ignore deviants that

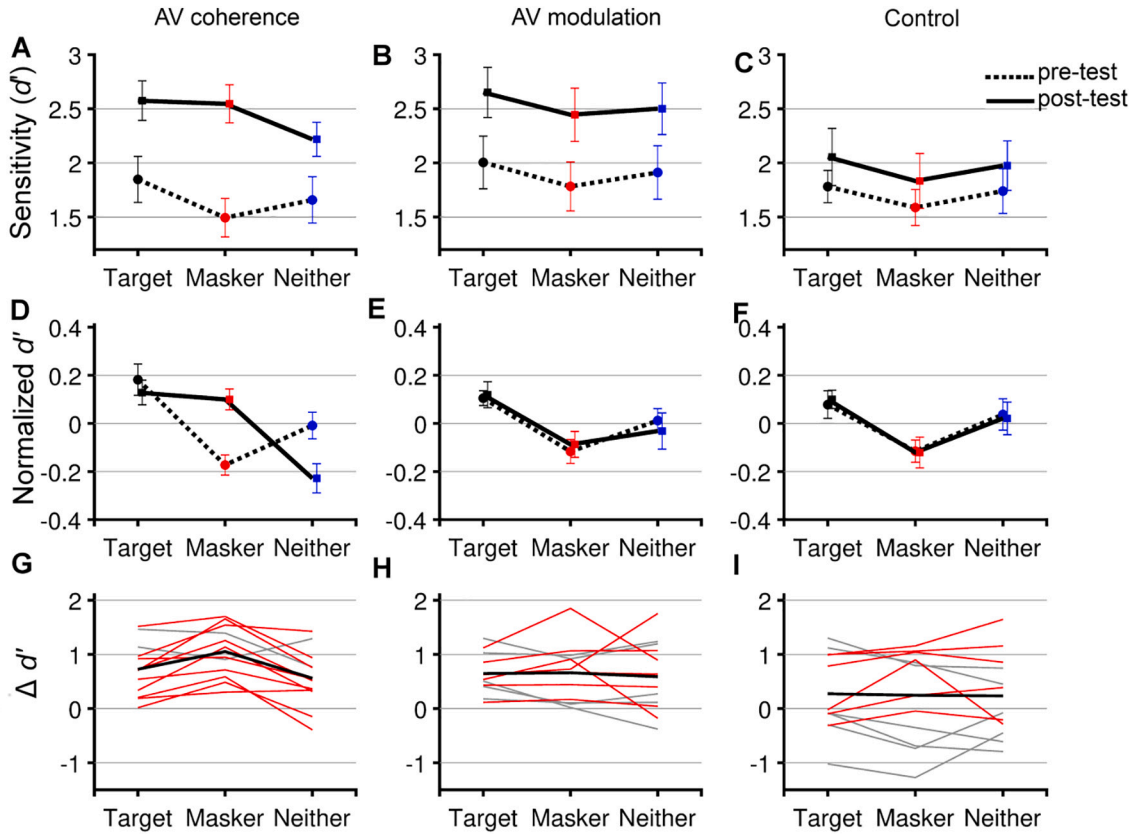

Fig. 3. Training to detect AV temporal coherence changed how listeners utilized visual information. **A-C** pretest (dashed line) and post-test (solid line) performance in the auditory selective attention task according to AV coherence conditions. A, D, G: AV coherence training group, B, E, H: AV modulation training group, C, F, G: control group. **D-F** Normalized mean ± SEM performance (calculated as within condition d' normalized to across condition performance for pretest and post-test separately). **G-I** the difference in d' between pre and post-test for each participant in either gray, or red. Participants who showed a greater increase in the masker coherent condition than the target coherent after training are plotted in red, the group mean is plotted in bold black.

occurred in the masker stream. Fig. 4 shows the changes in hit rates and false alarm rates between the pretest and post-test for the AV coherence training group and suggests that training drove an overall drop in false alarms across all coherence conditions, and a condition-specific increase in hit rates, with the largest increase occurring in the masker coherent condition. Two-way ANOVAs (Table 2) revealed that there was a significant increase in hit rates with training, with significant effects of session ($F(1, 11) = 7.731$, $p = 0.018$) and coherence condition ($F(2, 22) = 5.080$, $p = 0.015$), and a significant interaction ($F(2, 22) = 4.660$, $p = 0.021$). The decrease in false alarm rate between pretest and post-test was also statistically significant ($F(1,11) = 17.164$, $p = 0.002$) without a significant coherence condition effect ($F(2, 22) = 2.692$, $p = 0.115$).

3.7. AV coherence training improved temporal coherence detection

The pretest and post-test included a temporal coherence detection threshold test for all listeners. As anticipated, those listeners trained to detect audiovisual temporal coherence improved their thresholds between the pretest and post-test: (Fig. 5A, $t_{11} = 3.081$, $p = 0.005$). Listeners in the AV modulation group who were exposed to temporally coherent AV stimuli did not improve their thresholds ($t_{11} = 1.69$, $p = 0.104$) nor did the control group ($t_{11} = 0.234$, $p = 0.817$). Examination of these data also revealed there was considerable variability in how well observers could detect temporal coherence, and in the AV modulation training group, considerable individual variability in the way in which performance changed between tests. We therefore asked whether any of this individual variability predicted performance in the auditory selective attention task.

3.8. Sensitivity to temporal coherence is not predictive of performance in naïve listeners

We explored whether individual differences in temporal coherence detection ability accounted for the impact that the coherence condition had on performance in naïve listeners. Specifically, we tested the hypothesis that the ability of naïve listeners to detect AV temporal coherence would predict their ability to benefit from AV temporal coherence in the auditory selective attention task. We correlated each listener's AV temporal coherence threshold with the difference between the d' score in the target and masker coherent condition (Fig. 6A). Contrary to this hypothesis, there was no relationship between these values ($r = 0.1543$, $p = 0.3688$), nor was there any relationship between overall performance (across condition d′) and AV temporal coherence thresholds ($r = 0.2888$, $p = 0.0882$).

Having observed that the AV coherence training group improved their ability to utilize audiovisual temporal coherence in the masker coherent condition, we considered whether temporal coherence thresholds might be correlated with the magnitude of the benefit/impairment that the masker coherent condition had over the neither coherent condition. To assess this, we considered the difference in

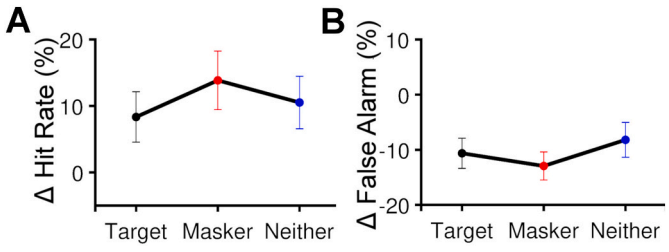

Fig. 4. Improved performance in the masker coherent condition in AV coherence training group was driven by a drop in false alarms and an increase in the masker-coherent condition hit rate. The changes in hit rates (**A**) and false alarm rate (**B**) between pretest and post-test; mean ± SEM.

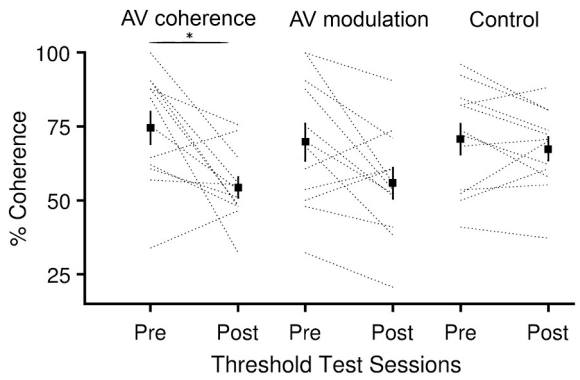

Fig. 5. Training decreased temporal coherence thresholds in the AV coherence training group. AV temporal coherence threshold values of pretest and post-test for three groups. * indicates significant paired t-test comparison ($p<0.05$).

d′ between the masker coherent condition and the neither condition. This comparison was weakly negatively correlated with AV temporal coherence thresholds for naïve listeners (Fig. 6B; $r = 0.339$, $p = 0.0438$) suggesting a trend where participants with better AV temporal coherence thresholds were more able to exploit the temporal coherence between masker stream and visual stimulus to yield a performance benefit relative to the neither condition. This finding mirrors the effect of training whereby improving AV coherence thresholds led to an improvement in the masker coherent condition.

Finally, since some participants in all groups showed improved temporal coherence detection thresholds between pre and post-test, we asked whether at an individual observer level whether the change in temporal coherence detection threshold correlated with the overall change in d′: Participants with a larger change in their AV coherence threshold showed larger improvements in overall performance (Fig. 4C; $r = 0.353$, $p = 0.0347$).

4. Discussion

Here we demonstrate that five short training sessions can improve a listener's ability to detect AV temporal correspondence and change the way in which they are able to exploit cross-modal temporal coherence. In naïve listeners, a visual stimulus that is temporally coherent with a target auditory stream enhances performance relative to when the visual stimulus is temporally coherent with the masker stream (with a condition in which the visual stimulus was coherent with neither yielding intermediary performance). After training, both target and masker visual coherence conditions yielded significantly better performance relative to the condition in which the visual stimulus was coherent with neither.

We had two control groups in this study. The first did not perform any training in between the pretest and the post-test. The second group was trained on an amplitude modulation rate discrimination task that utilized temporally coherent auditory visual stimuli. Therefore, like the AV coherence training group, they judged temporal features of the stimuli and were exposed to the stimulus streams that formed the target sounds in the auditory selective attention task. Unlike the AV coherence training group, the AV modulation training group did not require that observers make across-modal coherence discrimination and observers were free to base their decisions on auditory and/or visual features. Consistent with perceptual learning resulting from exposure to the sounds, both groups improved their performance of the auditory selective attention task. However, only observers that were required to explicitly judge cross-modal temporal coherence showed a change in the way in which visual information was used for auditory scene analysis.

We have previously argued (Maddox et al., 2015) that the visual stimulus impacts performance in the auditory selective attention task by

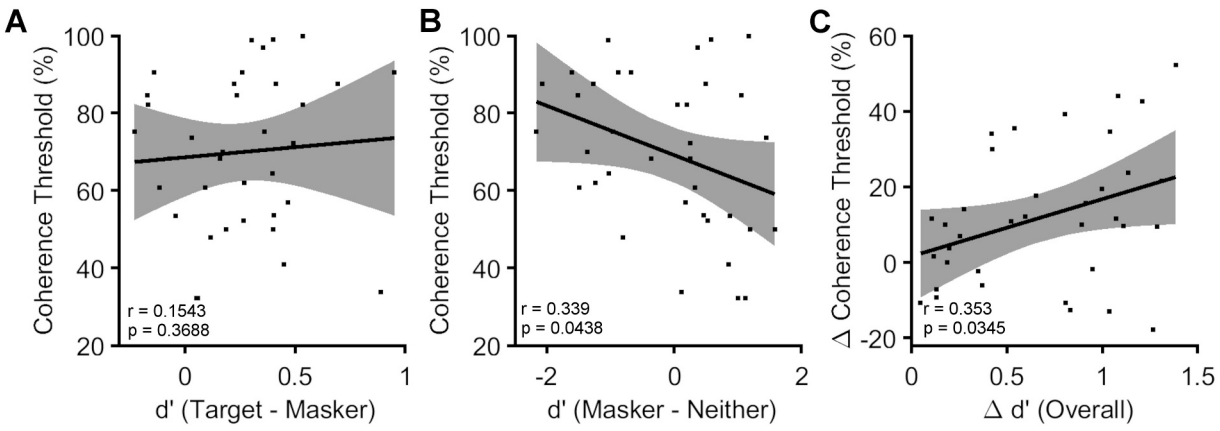

Fig. 6. Individual differences in AV temporal coherence detection **A** Target coherent -masker coherent d' difference (n=36 naïve listeners) versus AV temporal coherence threshold (low values indicate better thresholds) from all listeners pretest data. Error bars indicate 95% confidence intervals **B** Masker-neither d' difference versus AV coherence threshold for the pretest data. **C** The change in overall performance between the pretest and post-test versus change in AV coherence threshold.

altering how well listeners are able to separate the two competing streams and select the target. This is because the features that link the audio and visual streams (temporally coherent changes in auditory intensity and visual size) are independent of the changes in sound timbre that listeners are required to detect and the visual stimulus itself conveys no information about whether or when (or in which stream) the auditory timbre deviants occurred. Thus, improved performance in the auditory task demonstrated that auditory and visual streams have been bound into a single perceptual object (Bizley et al., 2016; Lee et al., 2019). Recordings in the auditory cortex of passively listening, naïve ferrets demonstrate that audiovisual temporal coherence causes an enhanced representation of the temporally coherent stream that extends to all of its features (Atilgan et al., 2018). As well as providing further evidence for audiovisual object formation, assuming that an audiovisual object has processing advantages, or captures selective attention more strongly, these data provide a bottom-up explanation for how the enhanced performance in the target coherent condition and the impaired performance in the masker coherent condition arises. What therefore might be the mechanism through which visual coherence with a to-be-ignored sound yields a processing advantage?

At a cellular level, successful stream segregation is thought to be a consequence of the activation of distinct neural populations in time, such that neurons representing the same stream are highly temporally coherent in their responses and those representing different streams share low coherence (Lu et al., 2017; Middlebrooks & Bremen, 2013). Under such a model temporal coherence with either the target *or* the masker stream should result in more distinct (and hence better segregated) neural responses that in turn offer a more effective substrate on which selective attention can operate. If AV temporal coherence allows the representation of each of two competing sounds to be more distinct within the sensory cortex then temporal coherence between target or masker stream should offer an advantage over an independently modulated visual stimulus. The data from the AV coherence training group suggest that after training listeners were able to benefit from AV temporal coherence when the visual stimulus was temporally coherent with either auditory stream. Possible mechanistic explanations for this would be that training has resulted in listeners being better able to use top-down control to actively suppress the masker stream in the masker coherent case, which in turn enables them to better detect the timbre deviants in the target stream. This suggestion is supported by the data in Fig. 4, which shows a condition-specific increase in hit-rates for the masker coherent condition after training. While we can only speculate about the mechanism underlying the effects observed here, a recent study that trained listeners to improve their audiovisual temporal perception reported enhanced beta band activity after training and suggested that enhanced top-down modulation was responsible for improved temporal processing (Theves, Chan, Naumer, & Kaiser, 2020). Further assessment of this hypothesis requires neurophysiological work to determine how selective attention and audiovisual object formation interact to shape the responses to target and masker streams in the auditory cortex.

We implemented an adaptive training procedure with feedback to force participants to work close to their perceptual threshold throughout the training periods. In keeping with other studies (Sürig et al., 2018), adaptive training was highly effective at rapidly driving learning in both of the temporal discrimination tasks. While many adaptive tasks show that the majority of learning occurs in the first session, repeated learning is thought to be critical for stabilizing learning (Shibata et al., 2017a; Shibata et al., 2017b). Follow up studies would be required to determine the optimal training strategy for maximizing long term perceptual gains.

Training listeners to narrow their temporal binding window often decreases their likelihood of integrating auditory and visual stimuli (Mcgovern et al., 2016; Setti et al., 2014), and the temporal binding window itself is task and stimulus dependent (De Niear, Gupta, Baum, & Wallace, 2018; Megevand et al., 2013; Stevenson & Wallace, 2013). Decreased integration can be successfully modelled within a Bayesian causal inference framework as resulting from both an increase in the precision of timing estimates and a decrease in a prior belief that signals originate from the same source (Mcgovern et al., 2016). Here, we report enhanced auditory visual integration after training listeners to make temporal coherence judgments. Unlike studies that train listeners to narrow their perceptual binding window, in our training paradigm participants were effectively trying to detect small amounts of temporal coherence and distinguish this from fully independent stimuli.

The width of the temporal binding window predicts susceptibility to sound induced flash illusions in naïve listeners (Stevenson, Fister, Barnett, Nidiffer, & Wallace, 2012). We did not find a relationship between the ability of naïve listeners to assess temporal coherence and their ability to exploit temporal coherence between the target and the visual stimulus. Nonetheless, we did observe a correlation between the ability of listeners to discriminate temporal coherence and the relative pattern of performance of the masker-coherent and independent condition with those people who were best able to assess temporal coherence showing an advantage for the masker coherent condition over the condition in which neither audio stream was coherent with the visual stimulus, and those people who were worse as assessing temporal coherence being relatively impaired on the masker coherent condition relative to the independent condition.

Temporal coherence across sensory modalities is a strong grouping cue. While this study focused on the impact that temporally coherent visual stimuli can have on listening, similar effects have been observed in the context of a visual discrimination (Lewis & Noppeney, 2010).

Here when task-irrelevant sounds were presented coincidentally with changes in the motion of the visual stimulus visual performance was improved relative to when such sounds were asynchronously presented. In this case temporally coherent audiovisual stimuli bidirectionally enhanced the connectivity between low level sensory cortical areas. This suggests that the effects we observe are not specific to auditory cortex, but represent a more general mechanism through which the brain links events in the world.

Previous studies have illustrated that visual cues can assist speech processing in noise (Grant et al., 1998; Helfer & Freyman, 2005; Schwartz, Berthommier, & Savariaux, 2004). While speech reading abilities are strongly predictive of audiovisual benefit for speech reception thresholds (Macleod & Summerfield, 1987), lip reading can influence auditory streaming (Devergie, Grimault, Gaudrain, Healy, & Berthommier, 2011), supporting the idea that, in addition to conveying phonetic information, lip reading benefits in noise potentially comprise of both bottom-up sensory effects that facilitate auditory scene analysis (Atilgan et al., 2018). Previous studies exploring the transfer of effects from training on temporal simultaneity judgments to other multisensory paradigms have had mixed results with transfer occurring to some tasks but not others (Mcgovern et al., 2016; Powers III, Hillock-Dunn, & Wallace, 2016; Setti et al., 2014; Sürig et al., 2018). An important question in interpreting the significance of our findings is whether the benefits in the auditory selective attention task transfer to other more real-world tasks such as utilizing speech reading in noisy listening conditions.

Context

We have previously (Maddox et al., 2015) demonstrated that a temporally coherent visual stimulus can enhance the ability of listeners to focus on one sound in a mixture. Using the same stimuli, we demonstrated a bottom up mechanism through which visual information could change the way in which sound mixtures were represented in auditory cortex, such that the neural representation of sounds that were temporally coherent with a visual stimulus were enhanced. One observation we made from our behavioural data was that listeners varied greatly in their ability to use visual to augment auditory scene analysis. In this study is therefore a first attempt to understand whether we could train listeners to use visual information more effectively. Our longer term goal is to relate these findings to other situations that require focusing on one sound in a mixture, such as listening to speech in noise, to understand whether listeners might benefit from training in order to exploit visual information more effectively.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104529.

Acknowledgments

This work was funded by a Wellcome Trust – Royal Society Sir Henry Dale Fellowship to JKB (ref: 098418/Z/12/Z) and an Action on Hearing Loss PhD studentship to HA. We are grateful to Suganya Mariyanesan for assistance in collecting the control data for this project, and to Ross Maddox and KC Lee for discussion of this work.

Appendix A. Supplementary data

Source data are freely available: Atilgan and Bizley (2020), "Training enhances the ability of listeners to exploit visual information for auditory scene analysis", Mendeley Data, V1, doi: 10.17632/dngrns68f8.1.

Supplementary videos are included to illustrate three trials; Supplementary video 1 shows a trial in which the target stream is coherent with the visual stimulus, Supplementary video 2 shows a trial in which the masker stream is coherent with the visual stimulus and Supplementary video 3 shows a trial in which the visual stimulus is coherent with neither stream. For demonstration purposes, the difficulty of the trial has been modified so that the timbre 'blips' are slightly more discriminable than in the experiment. Supplemental Table 1 details the timing of the blips in target, masker and visual streams for the three demo stimuli.

References

Atilgan, H., & Bizley, J. K. (2020). *Training enhances the ability of listeners to exploit visual information for auditory scene analysis. Mendeley Dataset, version (draft)*. https://doi.org/10.17632/dngrns68f8.1.

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., & Bizley, J. K. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron, 97*, 640-655.e4.

Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding: Experience-dependent effects in the double-flash illusion. *Experimental Brain Research, 234*, 3037–3047.

Bizley, J. K., Maddox, R. K., & Lee, A. K. (2016). Defining auditory-visual objects: Behavioral tests and physiological mechanisms. *Trends in Neurosciences, 39*, 74–85.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

De Niear, M. A., Gupta, P. B., Baum, S. H., & Wallace, M. T. (2018). Perceptual training enhances temporal acuity for multisensory speech. *Neurobiology of Learning and Memory, 147*, 9–17.

De Niear, M. A., Koo, B., & Wallace, M. T. (2016). Multisensory perceptual learning is dependent upon task difficulty. *Experimental Brain Research, 234*, 3269–3277.

Devergie, A., Grimault, N., Gaudrain, E., Healy, E. W., & Berthommier, F. (2011). The effect of lip-reading on primary stream segregation. *The Journal of the Acoustical Society of America, 130*, 283–291.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*, 719–721.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America, 103*, 2677–2690.

Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America, 117*, 842–849.

Lee, A. K., Maddox, R., & Bizley, J. K. (2019). An object-based interpretation of audiovisual processing. In A. K. Lee, A. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory processes*. Springer.

Lee, H., & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences of the United States of America, 108*, E1441–E1450.

Lewis, R., & Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *The Journal of Neuroscience, 30*, 12329–12339.

Lu, K., Xu, Y., Yin, P., Oxenham, A. J., Fritz, J. B., & Shamma, S. A. (2017). Temporal coherence structure rapidly shapes neuronal interactions. *Nature Communications, 8*, 13900.

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*, 131–141.

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

Mcgovern, D. P., Roudaia, E., Newell, F. N., & Roach, N. W. (2016). Perceptual learning shapes multisensory causal inference via two distinct mechanisms. *Scientific Reports, 6*, 24673.

Megevand, P., Molholm, S., Nayak, A., & Foxe, J. J. (2013). Recalibration of the multisensory temporal window of integration results from changing task demands. *PLoS One, 8*, Article e71608.

Meredith, M., Nemitz, J., & Stein, B. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience, 7*, 3215–3229.

Middlebrooks, J. C., & Bremen, P. (2013). Spatial stream segregation by auditory cortical neurons. *The Journal of Neuroscience, 33*, 10986–11001.

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Brain Research. Cognitive Brain Research, 25*, 499–507.

Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 29*, 12265–12274.

Powers, A. R., III, Hillock-Dunn, A., & Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Scientific Reports, 6*, 23374.

Schormans, A. L., & Allman, B. L. (2018). Behavioral plasticity of audiovisual perception: Rapid recalibration of temporal sensitivity but not perceptual binding following adult-onset hearing loss. *Frontiers in Behavioral Neuroscience, 12*, 256.

Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition, 93*, B69–B78.

Setti, A., Stapleton, J., Leahy, D., Walsh, C., Kenny, R. A., & Newell, F. N. (2014). Improving the efficiency of multisensory integration in older adults: Audio-visual temporal discrimination training reduces susceptibility to the sound-induced flash illusion. *Neuropsychologia, 61*, 259–268.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences, 14*, 425–432.

Shibata, K., Sasaki, Y., Bang, J. W., Walsh, E. G., Machizawa, M. G., Tamaki, M., … Watanabe, T. (2017a). Corrigendum: Overlearning hyperstabilizes a skill by rapidly making neurochemical processing inhibitory-dominant. *Nature Neuroscience, 20*, 1427.

Shibata, K., Sasaki, Y., Bang, J. W., Walsh, E. G., Machizawa, M. G., Tamaki, M., … Watanabe, T. (2017b). Overlearning hyperstabilizes a skill by rapidly making neurochemical processing inhibitory-dominant. *Nature Neuroscience, 20*, 470–475.

Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research, 219*, 121–137.

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

Stevenson, R. A., Wilson, M. M., Powers, A. R., & Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing. *Experimental Brain Research, 225*, 479–489.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 335*, 71–78.

Sürig, R., Bottari, D., & Röder, B. (2018). *Transfer of Audio-Visual Temporal Training to Temporal and Spatial Audio-Visual Tasks.. 31* p. 556).

Theves, S., Chan, J. S., Naumer, M. J., & Kaiser, J. (2020). Improving audio-visual temporal perception through training enhances beta-band activity. *Neuroimage, 206*, 116312.

Vroomen, J., Keetels, M., DE Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Research. Cognitive Brain Research, 22*, 32–35.

Zmigrod, S., & Zmigrod, L. (2015). Zapping the gap: Reducing the multisensory temporal binding window by means of transcranial direct current stimulation (tDCS). *Consciousness and Cognition, 35*, 143–149.