BRILL

Multisensory Research 32 (2019) 197–213

MULTISENSORY RESEARCH
brill.com/msr

Cross-Sensory Stimuli Modulate Reactions to Aversive Sounds

Patrawat Samermit *, **Jeremy Saal and Nicolas Davidenko**

Psychology Department, University of California, Santa Cruz, USA

Received 5 September 2018; accepted 21 February 2019

Abstract

We propose that cross-sensory stimuli presenting a positive attributable source of an aversive sound can modulate negative reactions to the sound. In Experiment 1, participants rated *original video sources* (OVS) of eight aversive sounds (e.g., nails scratching a chalkboard) as more aversive than eight *positive attributable video sources* (PAVS) of those same sounds (e.g., someone playing a flute) when these videos were presented *silently*. In Experiment 2, new participants were presented with those eight aversive sounds in three blocks. In Blocks 1 and 3, the sounds were presented alone; in Block 2, four of the sounds were randomly presented concurrently with their corresponding OVS videos, and the other four with their corresponding PAVS videos. Participants rated each sound, presented with or without video, on three scales: *discomfort*, *unpleasantness*, and *bodily sensations*. We found the concurrent presentation of videos robustly modulates participants' reactions to the sounds: compared to the sounds alone (Block 1), concurrent presentation of PAVS videos significantly reduced negative reactions to the sounds, and the concurrent presentation of OVS videos significantly increased negative reactions, across all three scales. These effects, however, did not linger into Block 3 when the sounds were presented alone again. Our results provide novel evidence that negative reactions to aversive sounds can be modulated through cross-sensory temporal syncing with a positive attributable video source. Although this research was conducted with a neurotypical population, we argue that our findings have implications for the treatment of misophonia.

Keywords

Cross-modal attenuation, multisensory integration, audition, vision, emotion

1. Introduction

Imagine a person scratching his fingernails on a blackboard in your lecture hall. Most people would respond with a multitude of negative sensations in-

---

[*] To whom correspondence should be addressed. E-mail: psamermi@ucsc.edu

DOI:10.1163/22134808-20191344

cluding physiological responses such as cringing, shivering, and screaming, and psychological responses such as generally feeling unpleasant and uncomfortable. Although there is a clinical condition known as misophonia, in which these reactions are marked, regular, and uncontrollable, often in reaction to specific sounds that originate from the mouth (Jastreboff and Jastreboff, 2002), these types of reactions are still prevalent in neurotypical populations, people who have not been diagnosed with misophonia. However, are these reactions the direct consequence of the auditory properties of these sounds, or could they also be mediated by higher-level knowledge of their physical source?

Certain sounds may elicit strong negative reactions among the general population. For example, the sound of fingernails scratching across a chalkboard, the sound of metal scraping glass, or the sound of someone chewing or sucking loudly may each produce negative emotional and physiological reactions in different observers (Kumar *et al.*, 2008; Zald and Pardo, 2002). Reuter *et al.* (2014) found that while aversive reactions to sounds often depended on their physical properties, certain reactions were based on deep emotional connections with the sound. Moreover, Thibodeau (2016) found that semantic features of certain words, like ‘moist’ and its association with disgusting bodily functions, was a more prominent source of people’s displeasure with the word than its phonological properties. Thus, semantic knowledge about the physical source of a sound may influence how we perceive the sound itself. For example, participants rated the psychoacoustics of chalkboard squeaking as worse when they knew the original source of the sound (chalkboard squeaks) as compared when they were told the sound was pulled from a modern musical composition (Reuter and Oehler, 2011).

Cox (2008) conducted a study to examine whether concurrently presenting an image that is thematically related to the sound could affect how negatively the sounds were perceived. Neurotypical participants heard *horrible* sounds (e.g., nails on a chalkboard, or a screaming baby) and *disgusting* sounds relating to human bodily functions (e.g., vomiting, eating, coughing, or spitting) that were paired with either a thematically associated still image (e.g., screaming baby sound paired with screaming baby picture), an unassociated still image (e.g., screaming baby sound paired with a picture of a lily flower and pad on a pond), or a green square as control. Results showed that the associated image made the participants perceive the horrible sounds as significantly more horrible than when the sound was presented with an unassociated or control image. However, the disgusting sounds were not perceived as significantly more or less disgusting based on the auditory–visual pairing. This finding suggests that some (but not all) negative reactions to sounds can be mediated by presenting concurrent, thematically related associated visual cues.

Critically, thematically *unrelated* images in Cox (2008), while more positive in nature, did not drive the horrible sounds to be perceived as any less horrible than the control. Therefore, the author proposes that the effect is driven by semantic congruence — the negative images are semantically associated with listeners' expectations for the sound's source, which removes ambiguity around what sound they are actually hearing, while the positive images were semantically unrelated and had no effect. Cox (2008) has shown that having horrible sounds paired with visual, semantically related images can increase neurotypicals' negative responses to horrible sounds themselves, but semantically unrelated, positive images do not decrease how horrible they find the sounds. This finding, which proposes a higher-level association between the sound and source, and not one simply related to the valence of a visual associate, lends itself to the question of association and attributability.

Based on this past body of research, we propose that neurotypicals' aversive responses to certain sounds are not driven purely by the auditory properties of the sounds, but rather by a multisensory simulation of the sounds' physical source. A large body of research supports the idea that sensory modalities are not modular but rather information is integrated across sensory modalities as it is processed (see Shimojo and Shams, 2001, for a review). For example, in the *ventriloquist effect*, the perceived spatial source of a sound is influenced by audio-visually synced movements, such that people perceive the sound as coming from a dummy's mouth rather than from the ventriloquist (Alais and Burr, 2004).

Studies on visual-auditory interactions have indicated that the *temporal synchronization* of visual and auditory information creates a stronger integration of the two sources. Sekuler *et al.* (1997) showed that ambiguous visual motion can be reinterpreted when synced with a correctly-timed sound. In the bounce–pass illusion, two identical disks move towards each other and cross, with the motion interpreted as *passing over* each other and continuing in their original directions. However, when a sound such as a click is synced at or near the point where these two disks cross each other, participants perceive the two disks *bouncing off* each other and moving in an opposite direction than they were moving before the moment they cross. Critically, the synchronization of the auditory click and the visual merging of the two disks significantly predicts the perception of a bounce, indicating that integrated visual cues may be understood as an attributable source of a sound and reinterpreted as such.

Similarly, in the McGurk effect, the same auditory phoneme, /ba/, is perceived as /da/ when participants view the lips synchronously creating the phoneme /ga/ (McGurk and MacDonald, 1976). Additionally, a follow-up study on the McGurk effect explored the role of temporal synchrony of the visual and auditory cues on the perceptions of phonemes (Munhall *et al.*, 1996). Results showed that an exact temporal synchrony between the visual and auditory cues is not necessary for the effect to arise, but when the two cues were in perfect temporal synchrony, the McGurk effect was strongest. Temporal synchronization, as in the McGurk effect, lends itself to stronger attribution of sound coming from the visual source.

However, the role of an attributable and synchronized video source on the perception of an aversive sound has yet to be investigated in a neurotypical population. With Cox's (2008) findings indicating a thematically related image can make a perceived sound worse for neurotypicals, it is worth investigating what a dynamic video of an attributable source of the sound can do, where the auditory and visual cues are temporally synced. In particular, does viewing a positive attributable source for an aversive sound help reduce neurotypicals' negative visceral and emotional reactions to the sound?

The present experiments examine whether presenting sounds synced with either the *original video source* (OVS) or a *positive attributable video source* (PAVS) differentially affects observers' responses to these sounds. Unlike Cox (2008), who only examined the *semantic* association between a thematically related visual image and sound, our experiments utilize videos that are time-synced to the sounds. The videos, then, function as an attributable source for the sound, allowing for more cohesive visual–auditory integration and a greater opportunity for change in the sound's perceived qualities. The dynamic aspect of the video, which is temporally synced to the sound, makes this attribution possible in a way that is similar to the McGurk effect, bounce–pass illusion, or other auditory–visual illusions. Our study goes beyond high-level semantic associations through a still image, and grounds the integration temporally through videos. We hypothesize that concurrent presentation of a PAVS video will attenuate negative responses to aversive sounds, whereas concurrent presentation of the OVS video will increase the negative responses, relative to hearing the sounds alone.

To test this, we conducted two experiments. In Experiment 1, participants watched 16 silent videos (the eight OVS videos that produced the eight aversive sounds, as well as eight PAVS videos that could plausibly produce the same sounds; see Supplementary Videos) and provided three behavioral ratings following each short ∼5-s silent video: how *uncomfortable* the video made them feel, how *unpleasant* it was, and the intensity of any felt *bodily sensations*. In Experiment 2, a different group of participants completed three blocks where they rated each of the eight aversive *sounds* on the same discomfort, unpleasantness, and bodily sensation scales, now asking how each *sound* made them feel. Participants were asked to rate each sound alone in the block 1 (pre-video) and block 3 (post-video). In block 2 (concurrent video), half of the sounds were presented with their associated OVS video, and the other half were presented with their associated PAVS video, showing an alternative, less aversive potential source of the sound.

We predicted that PAVS-paired aversive sounds would reduce ratings of discomfort, unpleasantness, and intensity of felt bodily sensations compared to sounds presented alone in block 1, whereas OVS-paired sounds would increase these ratings. We also predicted that there might be a *lingering* effect of the sound–video pairings, manifesting as lower ratings on the PAVS-paired sounds compared to OVS-paired sounds even when presented with no video in block 3.

2. Experiment 1: Reactions to Silent Videos

2.1. Participants

Twenty-three undergraduate students (16 female, 7 male; mean age = 20 years old) from the University of California, Santa Cruz psychology research pool completed an online experiment in exchange for course credit. Participants were required to sign a consent form that indicated they could stop the study at any time should they feel uncomfortable. None of the participants reported being diagnosed with misophonia or tinnitus.

2.2. Materials

Participants were presented with 16 silent videos (averaging 5 s each). Eight of the videos were demonstrations of actions that produced aversive sounds that induce negative emotional and visceral responses. The eight *original video sources* (OVS) were: (1) A knife grating on glass, (2) a chalk screeching on a chalkboard, (3) nails scratching a chalkboard, (4) a person rubbing a balloon to make it squeak, (5) a dry marker on paper, (6) someone loudly chewing/sucking on hard candy, (7) a fork scratching glass and (8) someone popping their fingers loudly. We filmed some of these events ourselves (e.g., rubbing a balloon, dry marker on paper, knife and fork scratching glass) and found other events on YouTube.

We generated the other eight *positive attributable video sources* (PAVS) that could provide a plausible alternative source for each of the aversive sounds. For example, for the sound of someone scratching their nails on a chalkboard, the PAVS video showed a person tearing a sheet of paper in sync with the nails being dragged on the board. To construct PAVS, we created or found videos of events that could plausibly produce the eight aversive sounds but involved less grating actions. These PAVS (respective to the eight OVS listed above) included: (1) a bird chirping, (2) a man playing a wooden flute, (3) paper being torn in half, (4) a cricket visibly chirping with its wings moving back and forth quickly, (5) a bunny licking another bunny, (6) someone handling rocks, (7) a child jumping on a bed, and (8) someone tapping a pencil on a table. The eight OVS videos, eight PAVS videos, and eight sounds used in these experiments are all available for download as Supplementary Videos.

2.3. Measures

Three 7-point scales were used to assess the participants' reaction to each silent video: a discomfort scale, an unpleasantness scale, and a bodily sensation scale. The *discomfort* scale measured participants' affective feeling toward the video, ranging from 1 indicating comfortable to 7 indicating uncomfortable. The *unpleasantness* scale measured participants' general perception of the video, where 1 indicated pleasant, and 7 indicated unpleasant. Finally, the bodily sensation measured the intensity of any bodily experiences when listening to the video, with 1 being 'None at all' and 7 being 'Very intense'. Participants were presented with these three scales, in the same order, after every video.

At the end of the study, participants were given a debriefing questionnaire asking about their general experiences with positive and negative sounds, autonomous sensory meridian response (ASMR), and whether they have experienced symptoms of tinnitus or misophonia (no participants did).

2.4. Procedure

Participants were asked to sign a consent form prior to participating in the study. They were then presented with instructions and watched the 16 silent videos in random order, and were asked to provide ratings on the three scales based on their reaction to each video. At the end of the study, the participants completed the debriefing questionnaire. The procedure took approximately ten minutes.

2.5. Results

To compare the differences between ratings for OVS and PAVS silent videos, we conducted paired-samples t-tests between each of the three types of ratings for OVS versus PAVS videos. Across all three scales, participants rated the OVS videos as significantly more negative than the paired PAVS videos (Fig. 1). Mean discomfort ratings were significantly higher for OVS videos (M: 4.45, SE: 0.17) than PAVS videos (M: 3.30, SE: 0.18; $t(22) = 5.99$, $p < 0.00001$); similarly, mean unpleasantness ratings were significantly higher for OVS videos (M: 4.59, SE: 0.15) than PAVS videos (M: 3.36, SE: 0.14; $t(22) = 8.20$, $p < 0.00001$); finally, mean bodily sensation ratings were significantly higher for OVS videos (M: 3.25, SE: 0.24) than PAVS videos (M: 2.57, SE: 0.18; $t(22) = 3.92$, $p = 0.0007$).

To account for variability in the use of the scales across participants, we also computed normalized responses across participants by subtracting the mean and dividing by the standard deviation of each participant's responses in each scale. Across all three normalized scales, participants rated OVS videos as significantly more negative than their respective PAVS videos. Mean normalized discomfort ratings were significantly higher for OVS videos than PAVS

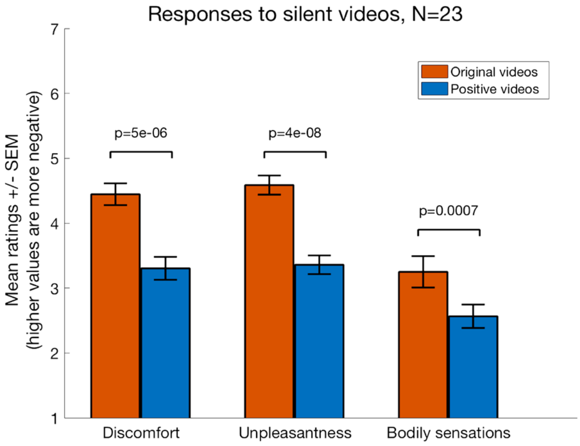

Figure 1. Mean responses across 23 participants in Experiment 1 to silent videos across three 7-point scales: discomfort, unpleasantness, and bodily sensations. Higher values indicate more negative responses. Original (OVS) videos are shown in red and positive (PAVS) videos are shown in blue. Error bars indicate standard error of the mean across participants.

videos ($t(22) = 6.67$, $p < 0.000001$); similarly for mean normalized unpleasantness ratings ($t(22) = 9.43$, $p < 0.00000001$), and for mean normalized bodily sensation ratings ($t(22) = 4.04$, $p = 0.0005$). Thus, the results based on the normalized scales are nearly identical to those based on the raw scales.

Finally, to determine whether the three scales are assessing independent information, we computed pairwise correlations among the three scales based on each participant's mean rating on each scale. We found that discomfort and unpleasantness were positively correlated across participants ($r = 0.84$, $p < 0.00001$), but bodily sensation ratings were not correlated with either discomfort ($r = 0.24$, $p = 0.26$) or unpleasantness ($r = 0.35$, $p = 0.1$). For clarity, we report results for all three measures separately in subsequent analyses, but note that the discomfort and unpleasantness scales appear to be interpreted similarly by participants.

2.6. Discussion

The results of Experiment 1 show that the silent OVS videos were perceived as more negative and elicited more intense bodily sensations than the silent PAVS videos. This was evident across all three scales. Overall, these data validate that our PAVS videos were less aversive than the OVS videos that produced each aversive sound. In Experiment 2, the eight aversive sounds were presented alone or paired with an OVS or PAVS video to test whether the concurrent presentation of an attributable source decreases (for PAVS videos) or increases (for OVS videos) the negative visceral and emotional response.

3. Experiment 2: Reactions to OVS- and PAVS-Paired Sounds

3.1. Participants

Forty undergraduate students (32 female, 8 male; mean age = 20 years old) from the University of California, Santa Cruz psychology research pool participated in exchange for course credit. Participants were required to sign a consent form indicating they were informed that in the experiment they would hear possibly unpleasant noises that may cause feelings of anxiety and that they could leave at any point.

3.2. Materials

We used the eight aversive sounds corresponding to the eight OVS videos described above: (1) A knife grating on glass, (2) chalk screeching on a chalkboard, (3) nails on a chalkboard, (4) a person rubbing a balloon so it squeaks, (5) a dry marker on paper, (6) someone loudly chewing/sucking on hard candy, (7) a fork scratching glass and (8) someone popping their fingers loudly. Figure 2 shows spectrograms for each of the eight sounds, a still image of the OVS video, and a still image of the PAVS video (see Supplementary Videos to view and hear the eight OVS and eight PAVS videos).

3.3. Measures

The same measures were collected in Experiment 2 as in Experiment 1, with one important difference: rather than rating the videos, participants were asked to rate the *sounds* that were presented with or without an accompanying video. At the end of the study, participants were given a debriefing questionnaire asking about their general experiences with positive and negative sounds, autonomous sensory meridian response (ASMR), and whether they have experienced symptoms of tinnitus or misophonia (no participants did).

3.4. Procedure

Participants were asked to sign a consent form before beginning the experiment. They were then read the instructions and presented with the three practice trials in order to familiarize them with the pleasantness, comfort, and bodily sensation scales. Participants were asked to rate each sound alone in block 1 (pre-video) and block 3 (post-video). In block 2 (concurrent video), half of the sounds were presented with their associated OVS video, and the other half were presented with their associated PAVS video, showing an alternative, less aversive potential source of the sound. The assignment of each sound to an OVS or PAVS video in the second block, as well as the presentation order of sounds within each block, was randomized across participants. Each participant was only exposed to one video per sound. Subjects were instructed to rate each sound they were presented with in each block, regardless

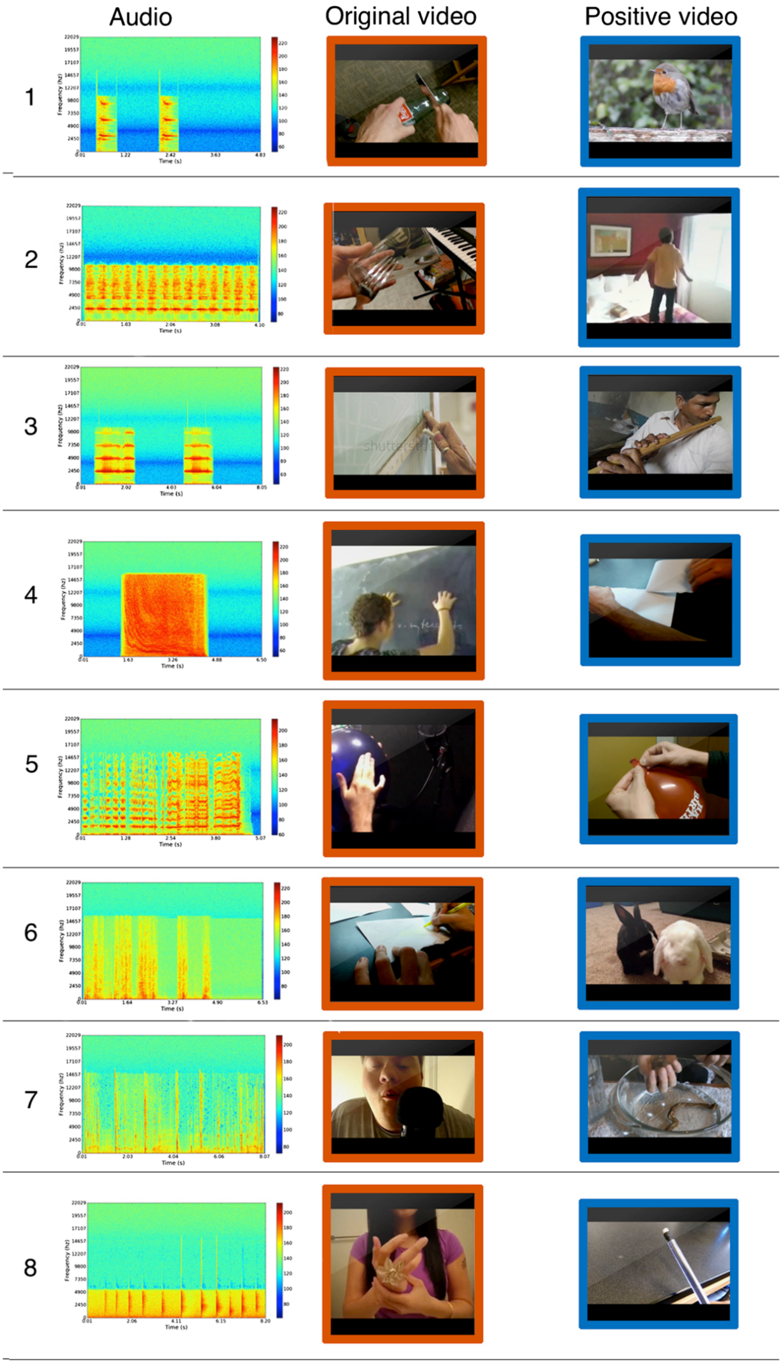

Figure 2. Audio spectrograms, original video source (OVS), and positive attributable video source (PAVS) videos for the eight sounds used in Experiment 1.

of whether or not a video was presented concurrently. Finally, participants completed a short questionnaire and were debriefed about the study.

3.5. Results

To compare the differences between responses to OVS-paired and PAVS-paired sounds, we computed pairwise t-tests between responses in block 1 (sound only), block 2 (sound concurrent with video), and block 3 (sound only), separately for the three scales. Because participants rated only four sounds in each condition on each scale, there was not enough within-participant variability in the responses to compute reliable normalized ratings, as in Experiment 1, so we report only results based on raw ratings on the 1–7 scale.

Across each of the three scales, we found that concurrent presentation of OVS videos significantly increased negative reactions to sounds (compared to pre- and post-video baselines), whereas concurrent presentation of PAVS videos significantly reduced negative reactions to the sounds (see Figs 3, 4, and 5). First, average *discomfort* ratings to OVS-paired sounds were 5.42 (SE: 0.11) when presented with the video, significantly higher than both the pre-video baseline (M: 4.97; SE: 0.12; $t(39) = 4.26$, $p = 0.0001$) and the post-video baseline (M: 5.05; SE: 0.12; $t(39) = 3.36$, $p = 0.002$; see Figure 3). Conversely, average discomfort ratings to PAVS-paired sounds were 4.77 (SE: 0.11) when presented with the video, significantly lower than the pre-video baseline (M: 5.09; SE: 0.09; $t(39) = 3.93$, $p = 0.0003$) and marginally lower than the post-video baseline (M: 4.93, SE: 0.11; $t(39) = 1.83$, $p = 0.08$).

A similar pattern of results was observed for ratings of unpleasantness (Fig. 4). Average unpleasantness ratings to OVS-paired sounds were 5.59 (SE:

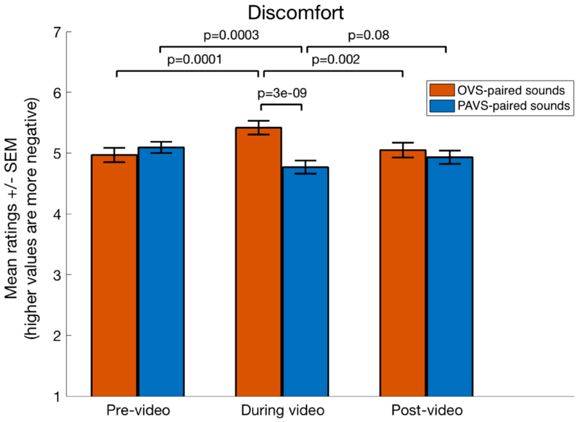

Figure 3. Discomfort ratings to aversive sound before (left), during (middle), and following (right) pairing with synced OVS (red) and PAVS (blue) videos. Each bar denotes the average rating on the 7-point discomfort scale across 40 participants' responses, and error bars indicate the standard error of the mean across participants.

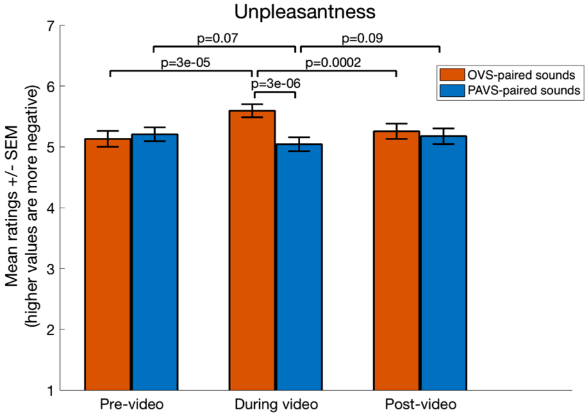

Figure 4. Unpleasantness ratings to aversive sounds before (left), during (middle), and following (right) pairing with synced OVS (red) and PAVS (blue) videos. Each bar denotes the average rating on the 7-point unpleasantness scale across 40 participants' responses, and error bars indicate the standard error of the mean across participants.

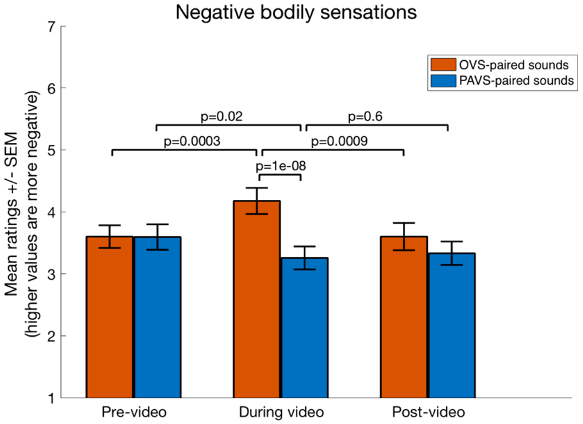

Figure 5. Ratings of bodily sensations to aversive sounds before (left), during (middle), and following (right) pairing with synced OVS (red) and PAVS (blue) videos. Each bar denotes the average rating on the 7-point bodily sensation scale across 40 participants' responses, and error bars indicate the standard error of the mean across participants.

0.11) when presented with the video, significantly higher than both the pre-video baseline (*M*: 5.13; SE: 0.13; $t(39) = 4.73$, $p = 0.00003$) and the post-video baseline (*M*: 5.26; SE: 0.13; $t(39) = 4.12$, $p = 0.0002$). Conversely, average unpleasantness ratings to PAVS-paired sounds were 5.04 (SE: 0.11) when presented with the video, marginally lower than both the pre-video baseline (*M*: 5.21; SE: 0.11; $t(39) = 1.85$, $p = 0.07$) and the post-video baseline (*M*: 5.18, SE: 0.13; $t(39) = 1.72$, $p = 0.09$).

Finally, the ratings of the intensity of bodily sensations to the sounds were also influenced by the concurrent presentation of OVS or PAVS videos (Fig. 5). Average bodily sensation ratings to OVS-paired sounds were 4.18 (SE: 0.21) when presented with the video, significantly higher than both the pre-video baseline (M: 3.60; SE: 0.18; $t(39) = 3.96$, $p = 0.0003$) and the post-video baseline (M: 3.60; SE: 0.22; $t(39) = 3.58$, $p = 0.0009$). Conversely, average bodily sensation ratings to PAVS-paired sounds were 3.26 (SE: 0.18) when presented with the video, significantly lower than the pre-video baseline (M: 3.59; SE: 0.21; $t(39) = 2.42$, $p = 0.02$). Interestingly, the post-video ratings of bodily sensations for PAVS-paired sounds remained low (M: 3.33, SE: 0.19) and were not significantly different from the during-video ratings ($t(39) = 0.57$, $p > 0.5$). In fact, the post-video ratings of PAVS-paired sounds were marginally lower than the pre-video ratings ($t(39) = 1.91$, $p = 0.06$), hinting at a potential lingering effect of pairing positive attributable video sources with aversive sounds on subsequent bodily sensations to the sounds alone.

3.6. Discussion

Overall, when aversive sounds were paired with their OVS videos, participants rated the sounds as producing more discomfort, being more unpleasant, and causing more intense bodily sensations than when they were presented alone. Conversely, when the same sounds were paired with their PAVS videos, participants rated them as producing less discomfort, being less unpleasant, and causing less intense bodily sensations than when they were presented alone. These findings support our first prediction that presenting synced attributable source videos with aversive sounds can modulate negative visceral and emotional responses. However, we did not find consistent evidence supporting our second prediction that these effects would linger into the post-video block. Only in one case (bodily sensation ratings on PAVS-paired sounds) did this lingering effect approach significance.

4. General Discussion

These experiments aimed to test whether an attributable visual source that was temporally synced with an aversive sound, such as the original video source for the sounds or a positive attributable video source (PAVS), could mediate neurotypicals' negative visceral and emotional reactions. Neurotypicals experience automatic negative emotional and visceral responses associated with some sounds such as nails on a chalkboard (Kumar *et al.*, 2008; Zald and Pardo, 2002). It has been shown that images that are visually associated with horrible sounds reliably make people perceive the sounds themselves as more horrible than visually unassociated images or a control (Cox, 2008). Additionally, cross-modal literature has illustrated that visual cues can be more powerfully integrated with auditory cues if they are temporally synced (Munhall *et al.*, 1996; Sekuler *et al.*, 1997).

Since cross-modal associations may play a role in the high-level perception of sound, we hypothesized that upon the presentation of a temporally synced positive attributable visual source, participants would have lower discomfort ratings, lower unpleasantness ratings, and less intense bodily sensations, whereas viewing the original video source would increase discomfort and unpleasantness ratings, and cause more intense bodily sensations compared to pre-test ratings of the sound. After validating the PAVS videos were less averise in Experiment 1, the findings of Experiment 2 support our hypothesis. These results indicate that neurotypicals' negative responses to everyday, aversive sounds may be momentarily attenuated by syncing positive attributable videos to the sounds. While the mechanism of these findings is still untested, our results suggest there is some type of visual–auditory integration occurring in the high-level perception of the sound. This may be due to a learned semantic association between the sound itself and its physical source, similar to people's perception of the word 'moist' (Thibodeau, 2016). This finding points towards a more complicated multimodal integration than may be predicted if visual and auditory cues were simply combined in an additive way.

In Experiment 2, we found that there was no statistically significant difference between the bodily sensation ratings during the video and post-video. Although the post-video and pre-video ratings were only marginally different, the data suggest that if participants were exposed repeatedly to pairings between the sounds and the PAVS videos, it could produce longer-term effects when sounds are heard alone.

We acknowledge an important limitation of our study is that we did not ask participants to rate the plausibility of the PAVS videos as being attributable sources for the sounds. It could be the case that PAVS videos were perceived as less attributable than OVS videos. If this is the case, the reduction in negative ratings to the PAVS-paired sounds may have been driven by another mechanism, such as participants being distracted by apparent lack of attributability, and this distraction reduced negative reactions. Assuming our PAVS videos attenuated negative emotional and visceral responses due to their attributability, a future control study could be run using positive videos that are not temporally synced or semantically related to the sound. If attributability does play a role, then temporally unsynced PAVS and semantically unrelated PAVS would not attenuate the negative responses compared to temporally synced and attributable PAVS from Experiment 2.

One possible mechanism may be the construction of a mental simulation of the sound's attributable source from the visual cue. Specifically, higher-level associations from the positive visual cues may down-modulate the perception of the original acoustic signals. Aversive sounds such as nails on a chalkboard, the sound of metal scraping glass, or someone chewing and sucking on objects cause higher activation of the amygdala and auditory cortex compared to similar but non-aversive sounds (Mirz *et al.*, 2000; Viinikainen, Kätsyri and Sams, 2012; Zald and Pardo, 2002). Kumar *et al.* (2012) found that the amygdala encodes not only the acoustic features of a sound stimulus, but also its valence, such as the sound's perceived unpleasantness. The perceived valence of a sound — its pleasantness or unpleasantness — provides feedback *back to* the auditory cortex, down-regulating the listener's perception of the sound itself. They conclude that the amygdala augments the representation of a sound's valence, making it so that a listener more readily and consciously registers the perception of an emotionally salient stimuli.

Additionally, Irwin *et al.* (2011) determined that the neural responses to natural sounds, such as a urban soundscapes, in the amygdala, posterior insula, and posterior auditory cortex are modulated by how pleasant or unpleasant the sound is perceived. Their results support a model of sound processing where two streams of information — one for acoustic signals, and one for processing emotional content — interact with each other to inform the holistic perception of a sound. The synced visual information may provide additional clues to the valence of the sound, creating a stronger 'positive' modulation of the perception of the sound itself and causing the sound to be perceived as less unpleasant overall.

While most of us experience some negative reactions to certain sounds (like nails scratching a chalkboard), those with clinically significant misophonia experience more extreme reactions that may make them feel disgust, anxiety, anger, and even a desire to harm those producing the sounds (Edelstein *et al.*, 2013). Sometimes these reactions are paired with a visceral 'fight-or-flight' response that manifests as pressure in the chest and head, tightened muscles, increased heart rate and body temperature, and even physical pain when individuals are exposed to trigger sounds (Edelstein *et al.*, 2013; Schröder *et al.*, 2013). Misophonic responses are not necessarily related to physical properties of a trigger sound, such as its loudness or timbre (Jastreboff and Jastreboff, 2014), but rather linked to higher-order features of the sound such as its meaning, social context, and interpretation (Bruxner, 2016; Schröder *et al.*, 2013).

It has been suggested that misophonia's negative visceral and emotional response to specific trigger sounds is similar but opposite to the autonomous sensory meridian response (ASMR). ASMR is a phenomenon in which individuals experience feelings of relaxation and well-being accompanied with a tingling sensation across the scalp, back of the neck, and other areas when they perceive specific audio-visual triggers (Barratt and Davis, 2015; Taylor, 2014). ASMR is not considered a clinical disorder, but can be experienced by the general population. Unlike people who experience misophonia, people who experience ASMR actively seek out audio-visual triggers such as videos of whispering into a microphone and the sound of squishing soft materials or giving massages (Barratt and Davis, 2015). However, McErlean and Banissy (2018) indicated that those who self-report experiencing ASMR also report having more misophonic symptoms as compared to a control group. The authors indicate that this finding is in agreement with Barratt and Davis (2015)'s suggestion that misophonia and ASMR exist across a valenced spectrum of sound sensitivity. Perhaps these similar, yet distinct, emotional and physiological experiences are both related to cross-modal associations that up- and down-regulate the perception of a sound's valence. However, to examine this relationship, future research needs to examine the mechanism for and lingering durability of PAVS's attenuation effect.

We do not know whether visual–auditory integration may mediate clinical misophonics' responses in the same way as in neurotypicals. Even if the attributability of sounds does play a role, new visual sources may still not be effective in misophonics if their associations to the aversive sounds are already overlearned across limbic and autonomic areas that control their perception of sounds (Jastreboff and Jastreboff, 2014). However, research has shown that misophonic responses have an onset in childhood and get worse over time (Kumar *et al.*, 2014; Rouw and Erfanian, 2018; Schröder *et al.*, 2013), which lends itself to the idea that cross-modal associations are susceptible to plasticity across the cortex (Shimojo and Shams, 2001).

Our findings, paired with future research on the mechanism behind the cross-modal attenuation, may serve as a basis for therapeutic relief from severe misophonic responses. Currently, there are several proposed therapies for treating misophonia, largely based on clinical experience and case reports. Tinnitus retraining therapy involves adding noise to the environment using a wearable sound generator, and many misophonics prefer to altogether leave or avoid a situation where a trigger sound may exist (Edelstein *et al.*, 2013; Jastreboff and Jastreboff, 2014). This form of therapy is seen as a sophisticated version of avoidance therapy, which can lead to reduced quality of life. Additionally, case studies have found that cognitive behavioral therapy (CBT) has positively affected misophonics' negative reactions (Bernstein *et al.*, 2013; Dozier, 2015). Schröder *et al.* (2017) conducted a mid-scale study finding CBT methods, such as attentionally shifting focus away from the sound, counterconditioning, stimulus manipulation, and relaxation exercises, effective in nearly half the patients over eight bi-weekly sessions. Part of the reframing process of cognitive behavioral therapy was stimulus manipulation, where a pleasant unconditioned stimulus (e.g., a positive video or image) would be repeatedly presented with a conditioned stimulus (e.g., a video of someone chewing). Unlike our design, these positive unconditioned stimuli are not attributable to the sound. The temporal syncing of the sound with positive attributability may help increase the likelihood of success in this therapeutic process.

Acknowledgements

The authors would like to acknowledge the contributions of John Collins in the preparation of these studies, and of Nick Antrilli in the preparation of this manuscript.

Supplementary Material

Supplementary material is available online at:
https://brill.figshare.com/s/154fe0137a60121c0bcf

References

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration, *Curr. Biol.* **14**, 257–262.

Barratt, E. L. and Davis, N. J. (2015). Autonomous sensory meridian response (ASMR): a flow-like mental state, *PeerJ* **3**, e851. DOI:10.7717/peerj.851.

Bernstein, R. E., Angell, K. L. and Dehle, C. M. (2013). A brief course of cognitive behavioural therapy for the treatment of misophonia: a case example, *Cogn. Behav. Therap.* **6**, e10. DOI:10.1017/S1754470X13000172.

Bruxner, G. (2016). 'Mastication rage': a review of misophonia — an under-recognised symptom of psychiatric relevance?, *Australas. Psychiat.* **24**, 195–197.

Cox, T. J. (2008). The effect of visual stimuli on the horribleness of awful sounds, *Appl. Acoust.* **69**, 691–703.

Dozier, T. H. (2015). Counterconditioning treatment for misophonia, *Clin. Case Stud.* **14**, 374–387.

Edelstein, M., Brang, D., Rouw, R. and Ramachandran, V. S. (2013). Misophonia: physiological investigations and case descriptions, *Front. Hum. Neurosci.* **7**, 296. DOI:10.3389/fnhum.2013.00296.

Irwin, A., Hall, D. A., Peters, A. and Plack, C. J. (2011). Listening to urban soundscapes: physiological validity of perceptual dimensions, *Psychophysiology* **48**, 258–268.

Jastreboff, M. M. and Jastreboff, P. J. (2002). Decreased sound tolerance and tinnitus retraining therapy (TRT), *Aust. N. Z. J. Audiol.* **24**, 74–84.

Jastreboff, P. J. and Jastreboff, M. M. (2014). Treatments for decreased sound tolerance (hyperacusis and misophonia), *Semin. Hear.* **35**, 105–120.

Kumar, S., Forster, H. M., Bailey, P. and Griffiths, T. D. (2008). Mapping unpleasantness of sounds to their auditory representation, *The Journal of the Acoustical Society of America* **124**, 3810–3817.

Kumar, S., von Kriegstein, K., Friston, K. and Griffiths, T. D. (2012). Features versus feelings: dissociable representations of the acoustic features and valence of aversive sounds, *J. Neurosci.* **32**, 14184–14192.

Kumar, S., Hancock, O., Cope, T., Sedley, W., Winston, J. and Griffiths, T. D. (2014). Misophonia: a disorder of emotion processing of sounds, *J. Neurol. Neurosurg. Psychiat.* **85**, P023. DOI:10.1136/jnnp-2014-308883.38.

McErlean, A. B. J. and Banissy, M. J. (2018). Increased misophonia in self-reported autonomous sensory meridian response, *PeerJ* **6**, e5351. DOI:10.7717/peerj.5351.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**, 746–748.

Mirz, F., Gjedde, A., Sdkilde-Jrgensen, H. and Pedersen, C. B. (2000). Functional brain imaging of tinnitus-like perception induced by aversive auditory stimuli, *Neuroreport* **11**, 633–637.

Munhall, K. G., Gribble, P., Sacco, L. and Ward, M. (1996). Temporal constraints on the McGurk effect, *Percept. Psychophys.* **58**, 351–362.

Reuter, C. and Oehler, M. (2011). Psychoacoustics of chalkboard squeaking, *J. Acoust. Soc. Am.* **130**, 2545.

Reuter, C., Oehler, M. and Mühlhans, J. (2014). Physiological and acoustical correlates of unpleasant sounds, in: *Proceedings of the Joint Conference ICMPC13-APSCOM5*. Yonsei University, Seoul, Korea, p. 97.

Rouw, R. and Erfanian, M. (2018). A large-scale study of misophonia, *J. Clin. Psychol.* **74**, 453–479.

Schröder, A., Vulink, N. and Denys, D. (2013). Misophonia: diagnostic criteria for a new psychiatric disorder, *PloS One* **8**(1), e54706. DOI:10.1371/journal.pone.0054706.

Schröder, A. E., Vulink, N. C., van Loon, A. J. and Denys, D. A. (2017). Cognitive behavioral therapy is effective in misophonia: an open trial, *J. Affect. Disord.* **217**, 289–294.

Sekuler, R., Sekuler, A. B. and Lau, R. (1997). Sound alters visual motion perception, *Nature* **384**, 308.

Shimojo, S. and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions, *Curr. Opin. Neurobiol.* **11**, 505–509.

Taylor, V. (2014). Youtube videos trigger tingling ‘brain orgasms’ in ASMR practitioners. Available from http://www.nydailynews.com/.

Thibodeau, P. H. (2016). A moist crevice for word aversion: in semantics not sounds, *PloS One* **11**, e0153686. DOI:10.1371/journal.pone.0153686.

Viinikainen, M., Kätsyri, J. and Sams, M. (2012). Representation of perceived sound valence in the human brain, *Hum. Brain Mapp.* **33**, 2295–2305.

Zald, D. H. and Pardo, J. V. (2002). The neural correlates of aversive auditory stimulation, *Neuroimage* **16**, 746–753.