Routledge
Taylor & Francis Group

Auditory Perception & Cognition

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rpac20

Triggering Misophonia: The Importance of Spectral Information, Temporal Information, and Action Identification

Savvas Kazazis, Iza Korsmit & Stephen McAdams

To cite this article: Savvas Kazazis, Iza Korsmit & Stephen McAdams (08 Jan 2024): Triggering Misophonia: The Importance of Spectral Information, Temporal Information, and Action Identification, Auditory Perception & Cognition, DOI: 10.1080/25742442.2024.2302295

To link to this article: https://doi.org/10.1080/25742442.2024.2302295

Published online: 08 Jan 2024.

AUDITORY PERCEPTION & COGNITION
https://doi.org/10.1080/25742442.2024.2302295

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Triggering Misophonia: The Importance of Spectral Information, Temporal Information, and Action Identification

Savvas Kazazis[a,b,c], Iza Korsmit[c] and Stephen McAdams[c]

[a]Faculty of Fine Arts, School of Music Studies, Aristotle University of Thessaloniki, Thessaloniki, Greece; [b]Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece; [c]Schulich School of Music, McGill University, Montreal, Canada

ABSTRACT

Misophonia is characterized by severe negative emotional responses to specific environmental sounds. In this experiment, we investigated the importance of spectral and temporal acoustic information, as well as the role of action identification (e.g., chewing) in triggering misophonia. Eighteen participants with severe misophonia completed the experiment. In total, three stimulus sets were used: the first set consisted of recorded sounds that were either common misophonic triggers or neutral sounds that are not expected to trigger misophonia; the other two sets were each generated by applying temporal and spectral modifications to the unmodified (recorded) stimulus set. Participants rated how triggered they were by each sound (i.e., aversiveness) and were asked to identify the action category of each sound. The unmodified trigger sounds were rated to be more aversive than neutral sounds ($p < 0.0001$). The main effects of modification type and identification on aversiveness of trigger sounds were significant ($p = 0.0001$ and $p = 0.006$, respectively), but their interaction was marginally significant ($p = 0.053$). Although the unmodified and temporally modified sounds were not significantly different from each other ($p = 0.4$), spectrally modified sounds were rated significantly less aversive than both the temporally modified ($p = 0.003$) and unmodified sounds ($p = 0.001$). Regarding identification, the sounds that were incorrectly identified were on average rated as less aversive than the correctly identified sounds ($p = 0.006$). Furthermore, the interaction shows that the identification effect was largest for the spectrally modified sounds. This shows that both identification and spectral information play an important role in triggering misophonia.

ARTICLE HISTORY
Received 29 January 2023
Accepted 22 December 2023

KEYWORDS
Misophonia; action identification; spectral information; temporal information; psychological disorders

Introduction

Misophonia is a psychological disorder that is characterized by severe aversive responses to specific environmental sounds (i.e., triggers). Although there are many idiosyncrasies in what may trigger a person with misophonia, the most common triggers are created by

CONTACT Savvas Kazazis savvas.kazazis@mail.mcgill.ca Faculty of Fine Arts, School of Music Studies, Aristotle University of Thessaloniki, University Campus, Thessaloniki 54124, Greece

other humans, such as the sound of someone chewing, clearing their throat, tapping their foot, or typing on a keyboard. Given the high prevalence of such sounds in everyday life, having misophonia can have large negative effects on one's functioning in personal, academic, and work environments. Wang et al. (2022) found that anxiety and depression symptoms, negative beliefs about emotions, and disgust sensitivity were related either to the functional impact of misophonia or to misophonic outbursts (Vitoratou et al., 2021).

The disorder is not yet recognized by the Diagnostic and Statistical Manual – 5th version (DSM-5; American Psychiatric Association, 2013), but there has been an increasing amount of research on the characterization and treatment of misophonia (Vitoratou et al., 2021; see also Brout et al., 2018, for a review). Recently, Rosenthal et al. (2022) investigated the relationship between misophonia and psychiatric disorders in the DSM-5. They found that among the disorders that were correlated with misophonia, the most significant predictors of misophonia severity were borderline personality disorder, obsessive compulsive disorder, and panic disorder.

Research has shown that misophonia manifests itself on a physiological, neurophysiological, and neurobiological basis. Both the subjective judgment of aversiveness and the physiological measure of skin conductance response (SCR) increase when people with misophonia are presented with triggers (Edelstein et al., 2013). Furthermore, an EEG study showed that the N1-component, which is related to auditory attention and sound detection, was diminished for misophonics in response to unexpected (oddball) sounds compared to control participants (Schröder et al., 2014). Furthermore, an fMRI study found that people with misophonia show increased response in the anterior insular cortex (AIC) in response to misophonic sounds, compared to control participants and other unpleasant or neutral sounds (Kumar et al., 2017). They also show increased connectivity between the AIC and the default mode network (DMN): the AIC is related to the salience that people attribute to environmental stimuli (Seeley et al., 2007) as well as visceral emotions (Craig, 2009), such as the processing of disgusting stimuli (see, for instance, Kober et al., 2008); the DMN is related to memory recall and internal thoughts (Raichle et al., 2001). Consequently, the increased AIC reactivity may be causing intense visceral emotions, and the AIC-DMN connectivity may be interpreted as an increased salience attributed to environmental sounds due to associative learning and memory. Both AIC and DMN function may potentially be linked to the identification of the source of the sound, an aspect we address in our study.

Although it has been suggested that misophonia is mainly an auditory disorder that affects auditory processing in the early stages (Schröder et al., 2014), misophonics can also be triggered by visual cues (Brout et al., 2018). Samermit et al. (2022) examined whether misophonic responses can be modulated by visual cues. Using a sound-swapped video database, they found that when trigger sounds were paired with "positive attributable video sources," they were rated as significantly less unpleasant than when they were paired with the original video sources. During interviews that Edelstein et al. (2013) conducted with misophonics, some people reported that they react less severely or not at all when their trigger sounds are produced by animals or infants. Moreover, they report especially aversive responses to trigger sounds produced by family and friends, but less aversive responses when the sounds are produced by strangers or even themselves, further suggesting a role of sound source identification. Indeed, a later study by Edelstein et al. (2020) found that incorrect identification or incorrect text descriptions accompanying trigger sounds changed the subjective aversiveness of those sounds. This indicates that misophonia is not a purely auditory processing disorder but is also influenced by a top-down process of source identification. More recently, Heller and Smith (2022) found that the identification of spectrally degraded sounds had an effect on participants' pleasantness ratings of everyday sounds. This study did not include exclusively misophonic participants, although a few common misophonic triggers were used. A similar effect of spectral degradation could potentially be found on the experienced (un) pleasantness of trigger sounds in a strictly misophonic population sample as was investigated in the current study.

We consider both the effects of top-down identification, as well as bottom-up auditory attention in terms of spectral and temporal acoustic information. We present an experiment that exclusively addressed misophonic participants. First, we investigate the importance of spectral and temporal acoustic information in triggering misophonia. We test whether and the extent to which spectral and temporal modifications of trigger sounds add to or attenuate the feelings of aversiveness. We do not have a hypothesis as to whether the spectral or the temporal information of sound events is more important in triggering misophonia. We also test whether action identification (e.g., chewing, slurping, typing, etc.) of original recorded sounds and spectral or temporal modifications of those sounds affects the severity of misophonic response. Our main hypothesis concerns only misophonic trigger sounds: we hypothesize that incorrect identification will attenuate the severity of aversive responses to these sounds.

Methods

Participants

Participants were recruited online through misophonia support groups on Facebook. They filled out a pre-screening questionnaire (Misophonia Assessment Questionnaire; MAQ, see materials) and were selected to complete the main experiment if their score classified them to have at least moderate misophonia. The pre-screening questionnaire was completed by 33 participants, of which three were rejected because of low MAQ scores. From the 30 invited participants, only 18 (16 female) with an average age of 36.1 years ($SD = 14.1$; range: 19–61) were willing to proceed and completed the main experiment. Their average MAQ score was 44.83 out of 63 ($SD = 10.5$), which would be classified as severe misophonia. Participants were also instructed to complete the Amsterdam Misophonia Scale questionnaire (A-MISO-S; see **Materials**) at the end of the experiment. Their average A-MISO-S score was 15.44 out of 24 ($SD = 2.66$), which again classifies them as severe misophonics.

Eight participants resided in the United States; two each were from Canada, the United Kingdom, the Netherlands, and Belgium; followed by one participant from Germany and one from South Africa. Three participants reported auditory disorders of tinnitus and hyperacusis. Seven participants reported comorbid psychological disorders of anxiety, obsessive compulsive disorder, and depression. It is also worth noting that 50% of the participants reported that they have experienced Autonomous Sensory

Meridian Response (ASMR), as has been reported by previous studies as well (Rouw & Erfanian, 2018). Participants were compensated for their time through PayPal.

Materials

Questionnaires

Participants filled out the Misophonia Assessment Questionnaire (MAQ; Dozier, 2015) as a pre-screening test. It contains 21 items about the negative impact of misophonia on one's activities, thoughts, and feelings, which are rated on a Likert-scale from 0 (not at all) to 3 (almost all the time). Scores can range from 0 to 63. Johnson (2014, as cited by Dozier, 2015) classified a score of 0–21 as mild, 22–42 as moderate, and 43–63 as severe misophonia. Consequently, all participants that scored 22 or higher on the MAQ were invited to participate in the main experiment. Additionally, in the pre-screening experiment we asked participants the types of sounds they are usually triggered by. They could tick the boxes of "Eating or drinking sounds (e.g., eating an apple)," "Respiratory sounds (e.g., sniffling)," "Tapping sounds (e.g., typing)," "Material sounds (e.g., plastic)," or "Other" with the option to fill in other triggers. This was to ensure that the participants were triggered by the types of sounds that we included in the experiment, and not merely by a very specific sound that was not included.

After the listening task of the main experiment, participants filled out demographic questions concerning their age, country of residence, gender identity, biological sex, hearing issues, and comorbid psychological disorders. We asked them whether the action sounds included in the experiment were habitually triggering for them and also triggered them in the experiment to further confirm that our stimuli caused a misophonic response. Additionally, we asked the participants from what age their misophonia started and whether any of their family members suffered from misophonia. Finally, we described the phenomenon of ASMR, asked participants if they had ever experienced the phenomenon, and to describe which, if any, sounds elicited the response.

In addition to the pre-screening MAQ, in the post-experiment survey participants filled out the Amsterdam Misophonia Scale (A-MISO-S; Schröder et al., 2013) and the Misophonia Coping Responses questionnaire (MCR; Dozier, 2015). We wanted to include several misophonia scales, as there are no validated misophonia assessment questionnaires, and this allowed us to corroborate our findings on the severity of misophonia in our participant pool. The A-MISO-S measures the severity of misophonia symptoms with six items that are rated on a multiple-choice scale that scores zero to four. Total scores can range from zero to 24, with a score between 0–4 considered as subclinical misophonia, 5–9 as mild, 10–14 moderate, 15–19 severe, and 20–24 extreme severity of misophonia symptoms. The MCR measures the coping mechanisms of people with misophonia when they are confronted with trigger sounds. It contains 22 items that are rated on the same scales as the MAQ. There are no classifications available for the MCR, but scores can range from zero to 66.

Stimuli

In total, three stimulus sets were used. The first set consisted of recorded sounds that were either common misophonic triggers or neutral sounds that are not expected to trigger misophonia. Table 1 lists all the stimuli (organized in categories of trigger and

Table 1. List of stimuli and their respective categories and action labels.

Stimulus Name	Category	Action Label
Almonds	Misophonic	Chewing
Apple	Misophonic	Chewing
Banana	Misophonic	Chewing
Chips	Misophonic	Chewing
Noodles	Misophonic	Chewing
Soup	Misophonic	Slurping
Straw	Misophonic	Slurping
Nose Sniffling	Misophonic	Respiration
Snoring	Misophonic	Respiration
Sighing	Misophonic	Respiration
Plastic Crinkling	Misophonic	Crinkling
Swallowing Drink	Misophonic	Drinking
Typing Keyboard	Misophonic	Typing
Apple Pealing	Neutral	Pealing
Brushing Hair	Neutral	Brushing Hair
Crashing Plastic	Neutral	Crushing Plastic
Door Creaking	Neutral	Creaking Door
Flipping Pages	Neutral	Handling Paper
Ripping Paper	Neutral	Handling Paper
Pouring Lentils	Neutral	Handling Dried Beans
Stirring Lentils	Neutral	Handling Dried Beans
Shaking Water Bottle	Neutral	Shaking Liquid
Twisting Bottle Cap	Neutral	Opening Bottle
Writing Marker	Neutral	Writing on Paper
Writing Pen	Neutral	Writing on Paper
Zipper	Neutral	Zipping

neutral sounds) along with their respective action labels that describe the physical action that produced the sound. The sounds were recorded in the acoustically treated sound-isolation labs of the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT). Two microphones were placed in close and distant positions (approximately 12 in and 30 in, respectively) from the sound sources leading to two different recording sets. After listening to comparisons between the two recording sets, we decided to use only the distant recordings, because they sounded more natural with respect to everyday listening conditions. For instance, in some cases the close-miked recordings sounded as if someone was chewing right inside one's ears. The recording equipment consisted of two Neumann U87 condenser microphones, which were interfaced through an RME Audio Fireface UFX. The outputs of each microphone were monitored by two people through two Sennheiser HD280 headphones.

The second stimulus set was generated by applying temporal modifications on the unmodified (i.e., as recorded) stimulus set. Temporal modifications were achieved by segmenting the audio signal into short overlapping frames of 25 ms and shuffling them over a radius of 250 ms. The amount of overlap was set to 50% and the frames were Hann-windowed. The shuffling was performed by a random permutation of the frames (within the given radius) without replacement. The permutation was conditioned by a Gaussian probability distribution which controls the difference between the original and final frame positions (Ellis, 2010). This minimizes possible audible artifacts that would result from large displacements from the original frame positions (e.g., if a uniform probability distribution was used instead). The standard deviation of the distribution was empirically set to 16 (frame indices) after informal listening tests

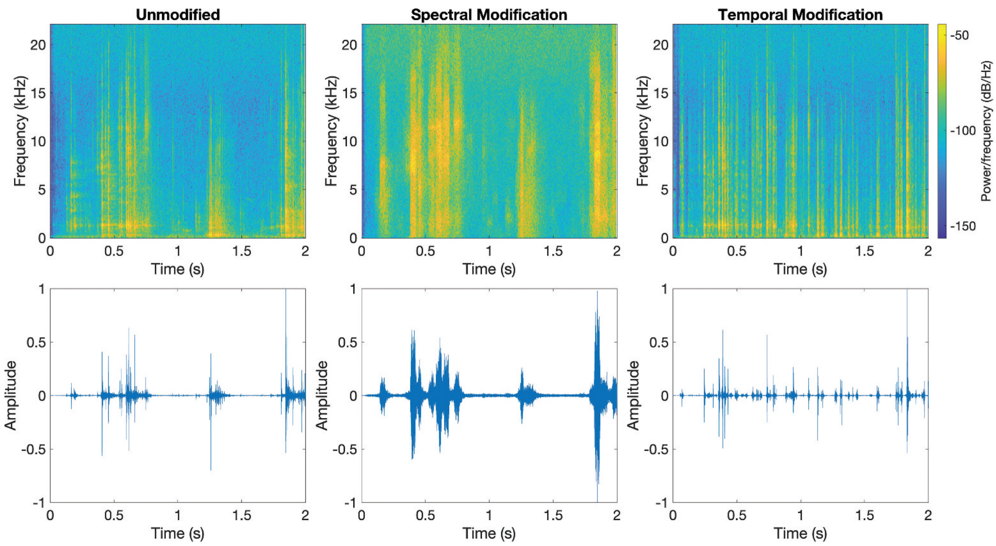

Figure 1. Spectrograms (top) and respective waveforms (bottom) of the first two seconds of the stimulus set "almonds".

conducted among the authors. The resynthesis was performed by overlap-add. Overall, this process preserved the short-term average spectrum (over the 250-ms radius) but distorted the global amplitude envelope (Figure 1).

The third stimulus set was generated by applying spectral modifications on the unmodified stimulus set. Spectral modifications were achieved through spectral whitening while preserving the global amplitude envelope. This was done by first estimating the spectral envelope of the input signal on 20-ms frames through a 13^{th}-order Linear Predictive Coding (LPC) model. The signal was resynthesized by feeding white noise (instead of the original excitation signal) into the filter's structure and the amplitude envelope was controlled by the gain of the filter. The code for analyzing and resynthesizing the signal was adopted from Slaney's Auditory Toolbox (Slaney, 1998). Overall, this process had the effect of removing the spectral fine structure of the input signal through whitening while preserving its global amplitude envelope (within 20-ms frames) through the time-varying gain of the filter (Figure 1).

All stimuli (including the unmodified stimulus set) had a duration of 10 s and were high-pass filtered with an 80-Hz cutoff frequency to reduce audible artifacts caused by the modification processes. The last processing stage was loudness normalization, which was implemented according to the algorithm of Moore et al. (1997). All stimuli were processed and synthesized in MATLAB (The MathWorks, Inc., Natick, MA).

Procedure

The pre-screening and main experiment were executed online, hosted on secure servers at McGill University. Before the experiment, participants gave informed consent. As mentioned above, only participants with a MAQ score of at least 22 were selected to proceed to the main experiment. The experimental instructions required participants to have normal hearing thresholds; use a laptop or desktop computer; use headphones or earphones, be inside a quiet room, and not currently be taking medical or recreational drugs that affect wakefulness and vigilance. Before the main trials, participants were asked to adjust the volume of their headphones/earphones to a comfortable listening level according to a reference sound sample. They were further instructed to not change this level throughout the experiment.

In each trial, participants were first presented with a stimulus and then were asked to rate with a slider "How effectively does this sound trigger misophonia for you?" The rating scale was a continuous 9-point analogical-categorical scale (Weber, 1991) and the two extremes were labeled as "Not effectively" and "Very effectively". Afterwards, they were asked to "Identify the action that produced the sound" through a drop-down menu that displayed all the action labels listed in Table 1. Identification was deemed *correct* whenever the participant's answer matched the "Action Label" of the respective "Stimulus Name," as shown in Table 1. No feedback was provided on the correctness of participants' responses. Participants were allowed to play back each stimulus as many times as they wished prior to making their judgments. The two tasks were presented in consecutive screens and participants were not able to go back from the identifying task to the rating task (i.e., they could not adjust their rating after identifying the action).

The presentation order of each stimulus was randomized across all stimulus categories and for each participant. At the end of the experimental session, participants were asked to fill in the post-experimental survey as described in the **Materials** section. The total duration of the experiment was approximately 1 hour.

Data Analysis

All statistical analyses were conducted in R version 4.2.1 (www.r-project.org). For the Shapiro-Wilk test we used the *shapiro_test* function from the *rstatix* library (Kassambara, 2023) on the aversiveness ratings grouped by stimulus category, modification, and identification. For the non-parametric Wilcoxon signed-rank test, we used the inbuilt *wilcox.test* function, and the *wilcox_effectsize* from the *rstatix* library to test the *r* effect size. The linear mixed model (LMM) was implemented using the *lmer* function with the *bobyqa* optimizer from the *lmerTest* package (Kuznetsova et al., 2020). Eta-squared is not a reliable statistic for evaluating a local effect size (i.e., a particular main effect) in mixed-effects models because of the presence of random effects (and therefore shared/partitioned variance). Post-hoc analyses were done with the *emmeans* function to calculate the estimated means, and the *eff_size* function to calculate the Cohen's *d*, both from the *emmeans* package (Lenth et al., 2022). In the absence of appropriate effect size measures for LMMs, Cohen's *d* in the post-hoc contrasts should give a sufficient idea of effect sizes. For the independent factorial ANOVA, we used the inbuilt *aov* function, and the *partial_eta_squared* function from the *rstatix* library to calculate the η_p^2 effect size. Planned contrasts were defined with the inbuilt *contrasts* function. With respect to the sound modification effect, the first contrast compared the unmodified sounds to both the spectrally and temporally modified sounds. Then, a second contrast compared the spectrally and temporally modified sounds to each other. Effect sizes are reported only for statistically significant results ($p \leq 0.05$).

Results

Due to the relatively small number of participants, we investigated the normality distribution of the variable combinations in order to determine the appropriate statistical methods. Shapiro-Wilk tests showed that the rating distributions for all combinations of stimulus category (i.e., misophonic or neutral stimulus), modification (i.e., unmodified, spectral, or temporal), and identification (i.e., correct or incorrect) showed evidence of non-normality (all $p < .01$). For this reason, and due to the unbalanced design of the experiment (i.e., number of correct versus incorrect identification responses), we used linear mixed modeling (LMM) to predict the aversiveness ratings of misophonic triggers, and to test for the main and interaction effects of modification type and identification.

Since our main hypothesis concerns only (misophonic) trigger sounds, the effectiveness of trigger manipulation on the aversiveness ratings of the unmodified stimulus set was studied first (Figure 2). The non-parametric Wilcoxon signed-rank test showed that the group of sounds which were categorized *a priori* as trigger sounds (Table 1) were indeed rated significantly higher (i.e., as more aversive; $M = 5.04$, $SD = 3.30$) than the neutral sounds ($M = 2.20$, $SD = 2.40$; $Z = -18.77$, $p < .0001$, $r = .43$). Therefore, the following analysis concerns only the groups of trigger sounds within each stimulus set.

After experimenting with different LMM effect structures, we selected the model that performed the best in terms of the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and log likelihood.[1] The employed model had the following structure: the aversiveness ratings of the misophonic stimuli were set as the dependent variable; the fixed effects consisted of modification type and identification correctness (the latter coded as a binary variable per participant and per stimulus), along with their

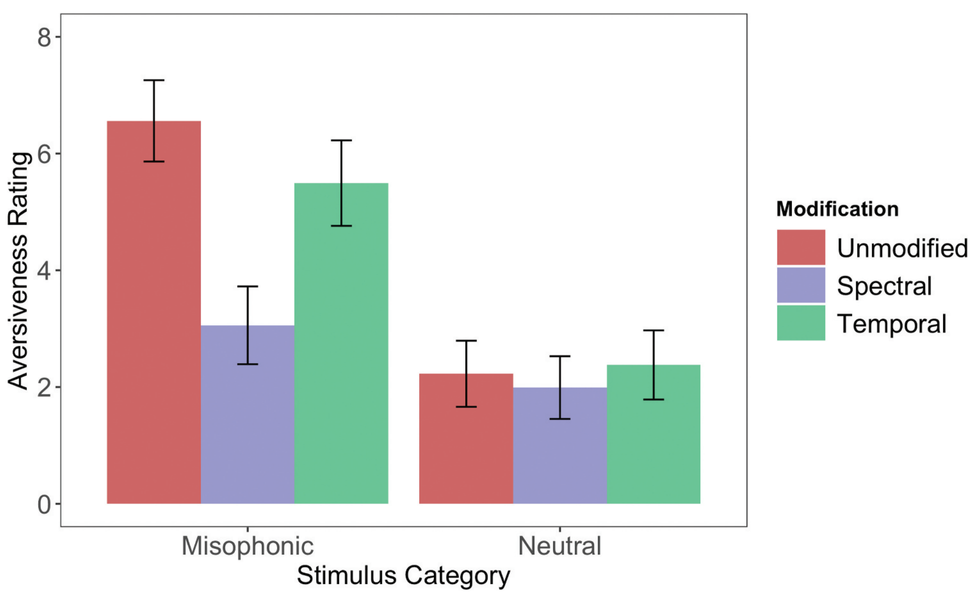

Figure 2. Mean and standard error of aversiveness ratings separated by stimulus category and modification type.

interaction; random intercepts were used per participant, and per stimulus; random slopes were used for both modification type and identification per participant.

We conducted Type III Wald F-tests to analyze the main and interaction effects of the LMM. The main effects of modification type, $F(2, 51.85) = 11.04$, $p = .0001$, and identification, $F(1, 19.87) = 9.69$, $p = < .006$, were found to be significant. The interaction effect between these two variables was marginally significant, $F(2, 641.88) = 2.96$, $p = .053$. With respect to modification type, Bonferroni-corrected post-hoc comparisons showed that although there was no significant difference between the aversiveness ratings of the unmodified stimuli (estimated $M = 6.01$, $SE = 0.60$) and temporally modified stimuli (estimated $M = 5.32$, $SE = 0.60$), the spectrally modified stimuli (estimated $M = 3.40$, $SE = 0.53$) were rated as significantly less aversive than both the unmodified, $t(48.5) = 4.61$, $p = .0001$, Cohen's $d = 1.62$ (hereafter reported as d), and temporally modified stimuli, $t(39.9) = -3.61$, $p = .003$, $d = -1.19$. With respect to the identification effect, the correctly identified stimuli (estimated $M = 5.64$, $SE = 0.53$) were overall rated as significantly more aversive, $t(20.1) = 3.11$, $p = .006$, $d = 0.96$, than the incorrectly identified stimuli (estimated $M = 4.18$, $SE = 0.56$).

Regarding the interaction between modification type and identification (correct vs. incorrect), the identification effect (i.e., the differences in aversiveness ratings between correctly and incorrectly identified stimuli) per modification type was significant for both the spectral and temporal modifications, $t(28.6) = 3.96$, $p = .0004$, $d = 1.26$; and $t(26.4) = 2.18$, $p = .004$, $d = 0.68$, respectively. However, for the unmodified stimuli, the identification effect was marginally significant, $t(62) = 1.97$, $p = .05$, $d = 0.77$. As can also be inferred from Figure 3, the identification effect was largest for the spectrally modified stimuli, followed by the unmodified and temporally modified stimuli. However, only the difference in identification effect between the spectral and temporal modifications was found to be marginally significant ($t(638) = 2.38$, $p = .05$, $d = 1.87$), whereas the rest were insignificant ($p > .3$).

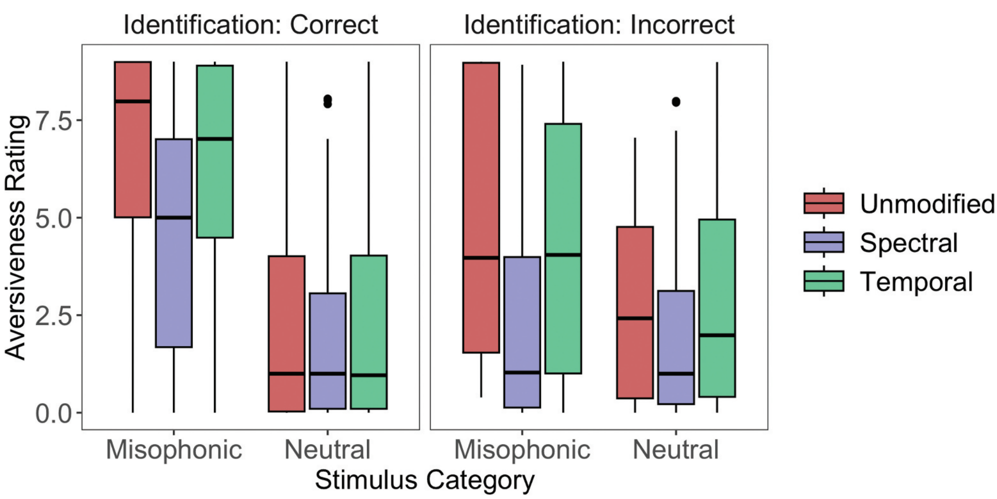

Figure 3. Boxplot of aversiveness ratings separated by stimulus category, modification type, and identification. Plots show the median (box midline), first and third quartile (box outlines), minimum (Q1–1.5*IQR) and maximum (Q3 + 1.5*IQR) scores (whiskers), and outliers (dots).

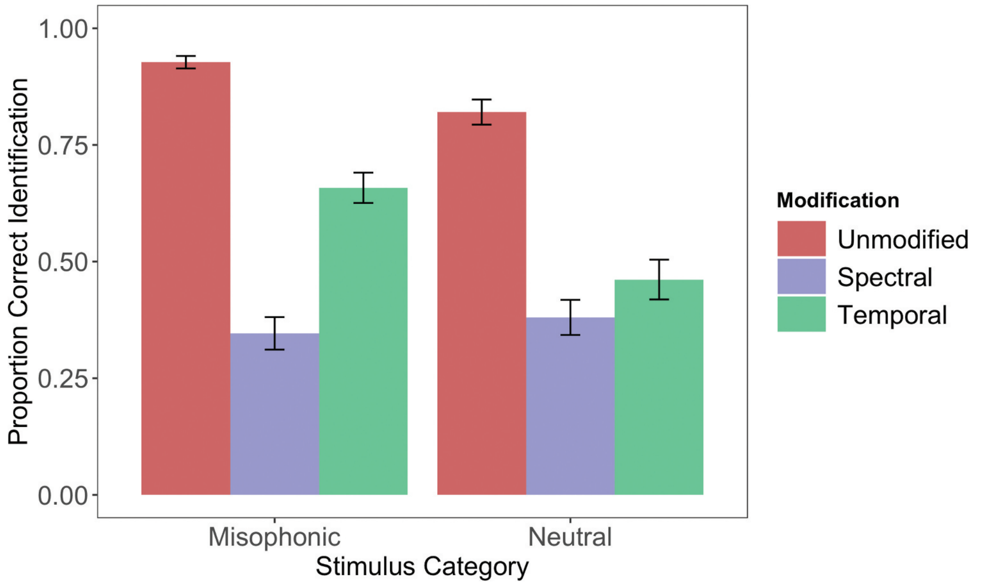

Figure 4. Mean and standard error of proportion correct identification separated by stimulus category and modification type.

As a more general remark, we investigated the stimuli's proportion of correct identification, i.e., the proportion of participants that correctly identified each stimulus. An independent factorial ANOVA with the proportion correct identification as dependent variable and stimulus category (misophonic/neutral) and modification as independent variables showed that the difference in proportion correct identification between the misophonic and neutral stimulus sets (Figure 4) was not significant, $F(1, 72) = 1.89$, $p = .17$. However, there was an effect of modification type, $F(2, 72) = 20.71$, $p < .0001$, $\eta_p^2 = .37$, where the planned contrasts showed that proportion correct identification was higher for the unmodified sounds compared to both the spectral and temporal modifications, $t(72) = -5.95$, $p < .0001$, $d = -1.43$. The proportion correct identification was also slightly higher for the temporal than the spectral modifications, $t(72) = -2.46$, $p = .02$, $d = -0.68$. There was no significant interaction between the stimulus category and modification, $F(2, 72) = 1.06$, $p = .35$.

Discussion

The main purpose of this experiment was to test whether action identification plays a role in evoking misophonic responses and the extent to which these relate to spectral and temporal stimulus information. To this end, we modified recorded stimuli that are generally considered to be triggers while aiming to preserve either their spectral or temporal similarities to the recorded sounds.

Post-hoc tests related to modification type show that temporal modifications of 25-ms frames within a 250-ms radius did not reduce aversiveness responses compared to the unmodified stimuli. It can therefore be concluded that spectral information averaged over 250 ms is important in triggering misophonia, and that the random temporal micro-fluctuations over 25-ms frames (or longer) did not statistically affect the misophonic responses compared to the unmodified stimuli. In addition, given that the identification effect on temporally modified stimuli was significant (although weak) and that there was a relatively high proportion of correct identification of temporally modified trigger sounds (65.8%; Figure 4), it can be concluded that both spectral information and action identification contributed to high aversiveness ratings.

This is further supported by the results obtained from spectral modifications. Preserving the temporal information, and therefore any repetitive patterns of the unmodified stimuli over 20-ms frames while whitening the signal, significantly reduced the aversiveness ratings to a level similar to the neutral stimulus sets (Figure 2). This leads again to the conclusion that spectral information is more important than the temporal organization of amplitude micro-fluctuations in triggering misophonia. The identification effect of this modification type was the largest, although percent correct action identification was relatively poor (34.6%; Figure 4). Nonetheless, correct identification of these stimuli significantly increased the aversiveness ratings (Figure 3). In general, these results are also similar to those obtained by Heller and Smith (2022) according to which spectral degradations made the sounds more neutral across their "Positive" and "Negative" valence groups of sounds. The authors attributed this effect to the high rate of misidentification and uncertainty about the sounds' causal properties (e.g., with respect to actions, such as impact sounds, or to the geometric/material properties of the object).

Although we acknowledge that the interaction effects between modification type and identification were marginally significant ($p = .05$), we believe that the reported results offer additional insights and complete the picture of the present analysis. The marginally significant p-value of the interaction effect is probably due to the relatively small number of participants and the high variance of ratings, especially with respect to the incorrect identifications of the unmodified stimulus set (Figure 3). The unmodified stimulus set can be thought of as a control set in which action identification should not play a role at all. In fact, almost all participants were able to correctly identify the action labels of the unmodified stimuli (92.7% correct identification of trigger sounds, 82.1% correct identification of neutral sounds; Figure 4). We attribute the high variance of the aversiveness ratings of the unmodified stimulus set (see right panel of Figure 3) to systematic misidentification: for instance, a participant may have consistently misidentified a stimulus as "Chewing" whereas the experimenters had labeled it as "Slurping," as in the "Noodles" stimulus label (Table 1). Nonetheless, the weak p-value of .05 and the small number of participants due to the difficulty of recruiting moderate to severe misophonic participants notwithstanding, the overall conclusion of the present analysis is that both action identification and spectral information play an important role in triggering misophonia.

The results are consistent with our initial hypothesis regarding action (or source) identification and support the findings reported by Edelstein et al. (2020) according to which incorrect identification of trigger sounds changed the subjective aversiveness of those sounds. The spectrally modified sounds reduced the misophonic responses, which suggests that this type of modification acted as a "neutralizer" for the misophonic stimulus set (Figure 2). Similar results were reported by Heller and Smith (2022) with respect to pleasantness ratings of everyday sounds. On a final note, we do not claim that temporal information is unrelated to action identification. The present analysis concerns only the aforementioned type of temporal modification on misophonic sounds. Examining the importance of individual amplitude and frequency modulation rates is currently under investigation.

Note

1. The results of this report are slightly different from the results presented in the extended abstract of the 21st Annual Auditory Perception, Cognition & Action Meeting (APCAM) due to differences between the employed LMM effect structures.

Acknowledgments

We would like to thank Xin Wang and Valentijn Nieman for their help during the recording sessions and Bennett K. Smith for setting up the online experiment and for programming the user interface.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This project was partially funded by CIRMMT Student Awards to SK and IK and a Canadian Natural Sciences and Engineering Discovery Grant [RGPIN-2020-04022] and a Canada Research Chair [950-231872] to SMc.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Brout, J. J., Edelstein, M., Erfanian, M., Mannino, M., Miller, L. J., Rouw, R., Kumar, S., & Rosenthal, M. Z. (2018). Investigating misophonia: A review of the empirical literature, clinical implications, and a research agenda. *Frontiers in Neuroscience*, *12*, 36. https://doi.org/10.3389/fnins.2018.00036

Craig, A. D. (2009). Emotional moments across time: A possible neural basis for time perception in the anterior insula. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1933–1942. https://doi.org/10.1098/rstb.2009.0008

Dozier, T. H. (2015). Counterconditioning treatment for misophonia. *Clinical Case Studies*, *14*(5), 374–387. https://doi.org/10.1177/1534650114566924

Edelstein, M., Brang, D., Rouw, R., & Ramachandran, V. S. (2013). Misophonia: Physiological investigations and case descriptions. *Frontiers in Human Neuroscience*, *7*, 296. https://doi.org/10.3389/fnhum.2013.00296

Edelstein, M., Monk, B., Ramachandran, V. S., & Rouw, R. (2020). Context influences how individuals with misophonia respond to sounds. *bioRxiv*, Preprint. https://doi.org/10.1101/2020.09.12.292391

Ellis, D. P. W. (2010). *Time-domain scrambling of audio signals in Matlab*. Retrieved January 16, 2023, from http://www.ee.columbia.edu/~dpwe/resources/matlab/scramble/

Heller, L. M., & Smith, J. M. (2022). Identification of everyday sounds affects their pleasantness. *Frontiers in Psychology*, *13*, 894034. https://doi.org/10.3389/fpsyg.2022.894034

Johnson, M. (2014). 50 cases of misophonia using the MMP. Paper presented at the misophonia conference of the Tinnitus Practitioners Association, Atlanta, GA.

Kassambara, A. (2023). *Rstatix: Pipe-friendly framework for basic statistical tests (0.7.2)*. https://cran.r-project.org/web/packages/rstatix/index.html

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage*, *42*(2), 998–1031. https://doi.org/10.1016/j.neuroimage.2008.03.059

Kumar, S., Tansley-Hancoc, O., Sedley, W., Winston, J. S., Callaghan, M. F., Allen, M., Cope, T. E., Gander, P. E., Bamiou, D. E., & Griffiths, T. D. (2017). The brain basis for misophonia. *Current Biology*, *27*(4), 527–533. https://doi.org/10.1016/j.cub.2016.12.048

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). *lmerTest: Tests in linear mixed effects models* (3.1-3). https://CRAN.R-project.org/package=lmerTest

Lenth, R. V., Buerkner, P., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2022). *Emmeans: Estimated marginal means, aka Least-Squares means (1.8.0)*. https://CRAN.R-project.org/package=emmeans

Moore, B. C. J., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, *45*(4), 224–240.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682. https://doi.org/10.1073/pnas.98.2.676

Rosenthal, M. Z., McMahon, K., Greenleaf, A. S., Cassiello-Robbins, C., Guetta, R., Trumbull, J., Anand, D., Frazer-Abel, E. S., & Kelley, L. (2022). Phenotyping misophonia: Psychiatric disorders and medical health correlates. *Frontiers in Psychology*, *13*, 941898. https://doi.org/10.3389/fpsyg.2022.941898

Rouw, R., & Erfanian, M. (2018). A large-scale study of misophonia. *Journal of Clinical Psychology*, *74*(3), 453–479. https://doi.org/10.1002/jclp.22500

Samermit, P., Young, M., Allen, A. K., Trillo, H., Shankar, S., Klein, A., Kay, C., Mahzouni, G., Reddy, V., Hamilton, V., & Davidenko, N. (2022). Development and evaluation of a sound-swapped video database for misophonia. *Frontiers in Psychology*, *13*, 890829. https://doi.org/10.3389/fpsyg.2022.890829

Schröder, A., van Diepen, R., Mazaheri, A., Petropoulos-Petalas, D., Soto de Amesti, V., Vulink, N., & Denys, D. (2014). Diminished N1 auditory evoked potentials to oddball stimuli in misophonia patients. *Frontiers in Behavioral Neuroscience*, *8*, 123. https://doi.org/10.3389/fnbeh.2014.00123

Schröder, A., Vulink, N., Denys, D., & Fontenelle, L. (2013). Misophonia: Diagnostic criteria for a new psychiatric disorder. *PLoS One*, *8*(1), e54706. https://doi.org/10.1371/journal.pone.0054706

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Slaney, M. (1998). *Auditory toolbox*. Interval Research Corporation, Tech. Rep., 10. Retrieved August 6, 2023, from https://engineering.purdue.edu/~malcolm/interval/1998-010/

Vitoratou, S., Uglik-Marucha, C., Hayes, E., & Gregory, J. (2021). Listening to people with misophonia: Exploring the multiple dimensions of sound intolerance using a new psychometric tool, the S-Five, in a large sample of individuals identifying with the condition. *Psych*, *3*(4), 639–662. https://doi.org/10.3390/psych3040041

Wang, Q., Vitoratou, S., Uglik-Marucha, N., & Gregory, J. (2022). Emotion processes predicting outbursts and functional impact in misophonia. *Frontiers in Psychology*, *13*, 903142. https://doi.org/10.3389/fpsyg.2022.903142

Weber, R. (1991). The continuous loudness judgement of temporally variable sounds with an 'analog' category procedure. In A. Schick, J. Hellbrück, & R. Weber (Eds.), *Fifth Oldenburg Symposium on Psychological Acoustics* (pp. 267–294). BIS.