Emotion, voices and musical instruments: Repeated exposure to angry vocal sounds makes instrumental sounds angrier

Casady Bowman, Takashi Yamauchi and Kunchen Xiao
Department of Psychology
Texas A&M University
College Station, U.S.A.
casadyb@tamu.edu

Abstract—The perception of emotion is critical for social interactions. Nonlinguistic signals such as those in the human voice and musical instruments are used for communicating emotion. Using an adaptation paradigm, this study examines the extent to which common mental mechanisms are applied for emotion processing of instrumental and vocal sounds. In two experiments we show that prolonged exposure to affective non-linguistic vocalizations elicits auditory after effects when participants are tested on instrumental morphs (Experiment 1a), yet no aftereffects are apparent when participants are exposed to affective instrumental sounds and tested on non-linguistic voices (Experiment 1b). Specifically, results indicate that exposure to angry vocal sounds made participants perceive instrumental sounds as angrier and less fearful, but not vice versa. These findings suggest that there is a directionality for emotion perception in vocal and instrumental sounds. Significantly, this unidirectional relationship reveals that mechanisms used for emotion processing is likely to be shared from vocal sounds to instrumental sounds, but not vice versa.

Keywords—emotion perception; adaptation; music; speech

I. INTRODUCTION

Speech and music are two of the most effective means used to express emotion through sound; they provide the basis for everyday social interactions [1]. Research in vocal acoustic [2], infant-directed speech [3][4] and music [5] suggest shared mechanisms for emotion perception in music and speech. For example in [6], seven psychoacoustic features—loudness, tempo/speech rate, melody/prosody contour, spectral centroid, spectral flux, sharpness, and roughness—can commonly explain affect ratings given to both music and speech. A review by [7] further highlights the interdependence of the two domains; musical experience and linguistic experience mutually influence each other such that experience in one domain improves the other. These findings point to the idea that emotions in music and speech are processed by shared general mechanisms.

Despite these compelling findings, emotion processing underlying speech and music remains elusive due to three limitations. First, the majority of the studies investigating emotional processing of speech and music are correlational, relying mainly on regression analysis [1][2][9]. While regression analyses can determine what features of sound predict emotion ratings, it only indicates an indirect associative relationship. Second, the majority of speech and music research has been conducted separately. Only in the past ten years have topics of interest in research expanded to include the perception of emotion in music and speech. Third, past literature does not make clear the effect of motivational aspects of emotion (e.g., approach vs. avoidance).

Due to these limitations, a full understanding of emotion processing in music and speech is still unclear [1]. It is unknown whether the perception of emotion in speech and music is merely associative or structural. Correlational studies are unsuitable to uncover the functional specificity underlying the two domains (e.g., whether the same or different mental mechanisms mediate emotion processing in speech and music) (see [10][12] for exceptions). In addition, it is unknown exactly how speech and music are related beyond two dimensions of emotion. Research on emotion processing between music and speech has focused almost exclusively on dimensional aspects of emotion (arousal and valence), where other important characteristics such as adaptive properties of emotion (e.g., approach and avoidance) have not been investigated. By applying the adaptation paradigm developed by [1], the current study investigated these questions.

II. ADAPTATION IN EMOTION PROCESSING

Adaptation is a process during which continued exposure to a stimulus results in a biased perception toward opposite features of the adapting stimulus [11]. Adaptation is often used in face perception where participants are shown a face during a short adaptation period, and then they are shown ambiguous test images created by morphing between two faces. Reference [12] showed that extended exposure to distorted faces caused non-manipulated faces to appear distorted in the opposite direction of the adapting stimulus. This aftereffect is attributed to a reduction in neural responses evoked by the adapting face.
which leads to a bias of perception towards the unadapted stimuli [13]. These results suggest that adaptation studies are a useful and important means of uncovering the nature of the neural representations.

Fig. 1 demonstrates visual adaptation. In this figure the viewer is adapted to an American flag that is yellow, green and black, and directed to look at this flag for approximately one minute. Following this adaptation phase, the viewer looks at a blank white space and will see the same image in the opposite colors, red, white and blue. The prolonged exposure to the colors yellow, green and black, wears out the neurons in the brain that respond to those colors; when the viewer looks at a blank white space the neurons respond in the opposite direction, so that they see the American flag with the correct colors. Because of these contrastive aftereffects, adaptation paradigms are utilized to probe functional specificity of neural populations [14]. This type of adaptation also takes place for sound stimuli; for example, [10] demonstrated that auditory adaptation to angry vocalizations causes voices at test to be perceived as more fearful, and vice versa.

Adaptation research shows that neurons respond to specific stimulus attributes and are active at early stages of information processing, particularly for high-level properties such as facial identity [10, 14]. Researchers interpret these aftereffects to mean that a recalibration of neural processes takes place in response to continuously updated stimulation [10, 14], such that neurons are “worn out” from responding to an angry stimulus adaptor and then recalibrate so that an ambiguous sound at test is perceived as less angry.

Prolonged exposure to stimuli can also result in the opposite effect—sensitization. Sensitization results when an observer is repeatedly exposed, for instance, to an angry face and rates a subsequent neutral face as angrier [15]. The exact interpretation of what causes sensitization is still unclear. However, recent behavioral and neuroimaging research (functional magnetic resonance imaging—fMRI) points to the idea that sensitization is mediated by similar processes as adaptation and that sensitization occurs when stimuli serve a salient adaptive purpose [16]. Reference [16], for example, demonstrated that angry vocalizations evoked changes in the brain such as an increased alertness, which caused sensitivity to emotional information that is important for adaptive behavior. Participants listened to four speech-like, non-word stimuli and rated prosody discrimination of voices (e.g., if the voice was neutral or angry) while being recorded on fMRI. Results show sensitization where the bilateral superficial (SF) complex and the right laterobasal (LB) complex of the amygdala were sensitive to emotional cues from speech prosody that were similar to a melody in music. This offers evidence that anger, which has negative valence but approach motivation, is processed separately from fear, which has negative valence and avoidance motivation.

While the adaptation paradigm has been used to explore neural mechanisms underlying face perception, it is not yet clear if these aftereffects exist for processing other types of nonlinguistic auditory information, such as vocal sounds and instrumental sounds. To empirically investigate the relationship between music and speech, we focus on the link between instrumental sounds and vocalizations as substitutes for music and speech. Instrumental sounds and vocalizations are used as an initial starting point for studying music and speech because these sounds are simple and lack some of the complex variables such as rhythm or prosody. By using an adaptation paradigm, we investigate the structural relationships between vocal sounds and instrumental sounds in emotion in terms of their motivational salience—approach and avoidance. Approach is associated with positive feelings, and avoidance with negative feelings [17][18][19][20], however, the emotion anger serves as a confound—anger is associated with approach but coupled with negative feelings [21][22][23].

In Exp. 1a, participants heard either an angry or fearful vocalization from the Montreal Affective Voices [24] four times to elicit adaptation. Following this exposure phase, participants heard a test sound from a morphed continuum of instrumental sounds ranging from anger to fear and judge whether the sound is angry or fearful, see Fig. 2.

---

Fig. 1. When participants view a yellow green and black American flag for 60 seconds and then view a blank white space, they will see an American flag with the correct colors. This is due to neurons responding to the colors of the first flag “wearing out” and responding in the opposite direction after viewing a blank white space.

Fig. 2. Adaptation paradigm procedure for Exp. 1a (left) and Exp. 1b (right). Participants were adapted to vocal sounds (left—Exp. 1a) and later judged if the test sounds (instrumental sounds) were angry or fearful. This order was reversed in Exp. 1b (right panel).
In Exp 1b. Participants are first adapted to an instrumental sound and tested on a voice sound. If emotion processing in the domains of music and speech make use of shared neural mechanisms, and if emotion processing in the two domains is related in terms of their motivational characteristic [16], prolonged exposure to vocal sounds (e.g., angry voice) should result in after effects (either adaptation or sensitization) in the processing of instrumental sounds and vice-versa.

III. METHOD

A. Participants

Thirty-six undergraduate students took part in Experiment 1a (adapt to voice, test on instrument) (19 female, mean age = 18.7, standard deviation (SD) = 0.82), 17 male, mean age = 19.7, SD = 2.02). All participants reported normal hearing and received course credit. Fifty-two undergraduate students took part in Experiment 1b (adapt to instrument, test on voice) (24 female, mean age = 18.96, standard deviation (SD) = 0.91, 28 male, mean age = 19.32, SD = 1.09). All participants reported normal hearing and received course credit.

B. Materials

For the instrumental sounds used in the baseline and experimental test phases, stimuli were created from instrumental recordings taken from two classes of musical instruments, brass and woodwind. Selected instruments were the French horn, baritone, saxophone, and flute, recorded at 440Hz. Instrumentalists from which the sounds were recorded were directed to play both an angry and a fearful sound for each instrument. From these recordings angry to fearful continua were created from each instrument in seven steps that corresponded to 5/95%, 20/80%, 35/65%, 50/50%, 65/35%, 80/20% and 95/5% anger/fear. For the prolonged exposure sounds used in the experimental condition, two female and two male voices were taken from the Montreal Affective Voices [MAV, 24]. The MAV were designed as an auditory equivalent of the affective faces by Ekman and Friesen (1986); these are nonverbal affect bursts that correspond to anger, disgust, fear, pain, sadness, surprise happiness and pleasure. Analyses of the MAV show a mean rating of 68% for valence and arousal, which indicates high recognition accuracy. These stimuli have been used by Bestelmeyer [10, 14]. To create the MAV’s actors were instructed to produce emotional interjections used the vowel /a/. For prolonged exposure sounds, voices from four identities were chosen, two male, and two female; each expressing anger and fear. Stimuli were normalized in energy and presented in stereo via JVC Flats stereo headphones. The program STRAIGHT [24] was used to create the anger/fear morphs, this is based on MatlabR2007b (Mathworks, Inc).

C. Procedure

The experiment consisted of two phases—a baseline phase without prior prolonged exposure sounds and an experimental phase with prior prolonged exposure sounds. In the baseline phase, subjects received 84 trials, with 2 blocks of trials, one for each instrument class (2 brass and 2 woodwind) and was always given prior to the experimental phase. The sound for each instrument at each of the seven morph steps was repeated six times, leading to 84 trials per instrument class block, with a total of 168 trials. Within each block, sounds were presented randomly with an inter-stimulus interval of 2-3s. Following the baseline phase participants took part in the experimental phase where the trial structure consisted of one voice played four times and followed by an ambiguous morph after a silent gap of 1 second. There were four adaptation blocks (2 emotion x 2 gender) and each of the seven test stimuli per identity was repeated six times leading to 84 trials per block with a total of 336 trials. Table 1 demonstrates the structure of the baseline and test phases of Exp. 1a and Exp. 1b.

D. Design

For all data analysis, data were averaged as a function of the seven morph steps, where each participant had an average emotion rating score for each sound at each step, subsequently one-way repeated measures ANOVA was applied. For both the baseline and experimental tasks, participants were presented with first a sound and then a rating decision.

IV. RESULTS

A. Experiment 1a (Voice-Instrument)

A one-way repeated measures ANOVA on behavioral responses revealed a significant main effect for affective vocal sounds when participants were tested on instrumental sounds, Fig. 1a, (F (2, 70) = 21.71, MSE = .070, p < .001, η²p = .383). Prolonged exposure to an angry voice in Experiment 1a showed that participant’s consistently rated instrumental sounds as angrier, demonstrating a sensitization effect rather than adaptation. To examine the direction of this effect, paired t-tests were run and indicate that there is a significant difference for the baseline and anger conditions, t (35) = 4.61, p < .001, t = .91, 95% CI[.41, 1.40], where listeners rated sounds as close to neutral at baseline (M = .55, SD = .12) and only slightly more fearful (M = .59, SD = .17) after exposure to fear.
Fig. 3. Behavioral results for prolonged exposure to vocalizations, when tested on instrumental sounds (a). The grand average of all participants is displayed. Psychophysical function for the grand average of the three experimental conditions: baseline (solid), anger (light dashed) and fear (dark dashed). The point of subjective equality (PSE) values are denoted with a star (b).

To further investigate the relationship between vocal and instrumental sounds data were averaged as a function of the seven morph steps and a psychophysical curve (the hyperbolic tangent function) was fitted to the mean data for each adaptor type (baseline, anger and fear). Good fits were obtained for all three conditions; baseline ($R^2 = .76$), anger ($R^2 = .74$), and fear ($R^2 = .77$). The point of inflection of the function (point of subjective equality—PSE) was computed for all curves (baseline, anger and fear) as illustrated with an asterisk in Fig. 3b. The point of inflection refers to the point on the test continuum where the instrument at test was equally likely to be labelled as angry or fearful and can range from $– 0$ (100% anger / 0% fear) to $1$ (0% anger / 100% fear). Higher values indicate that the center of symmetry was closer to fear and lower values indicate that it was closer anger.

A one-way repeated measures ANOVA on inflection values revealed that there was a significant main effect of adaptation to affective voices ($F(2, 68) = 15.03$, $MSE = 1.79$, $p < .001$, $\eta^2_p = .30$). Exploring the main effects with t-tests show that the PSE as a result of adaptation to anger was significantly larger ($M = 4.39$, $SD = 2.13$) than the baseline condition ($M = 3.31$, $SD = 1.41$), ($t(35) = 3.11$, $p < .05$), suggesting that adaptation to an angry vocalization causes sensitization. In addition, anger was also rated significantly higher ($M = 4.39$, $SD = 2.13$) than fear ($M = 2.69$, $SD = 2.10$), ($t(35) = 6.41$, $p < .05$).

B. Experiment 1b (instrument-voice)

In contrast to the sensitization effect found in Experiment 1a, we found no sign of either adaptation or sensitization when participants were first adapted to instrumental sounds and tested with vocal sounds, $F(2, 102) = 1.53$, $MSE = .065$, $p = .221$, $\eta^2_p = .029$ (Fig. 4).
Overall results show that an angry voice has an effect on subsequent emotion processing of instrumental sounds, where the sound will be rated as more angry, but instrumental sounds do not have this effect on vocalizations. This indicates that exposure to an angry instrument does not cause adaptation, but sensitization.

V. DISCUSSION

The aim of this study was to examine the extent to which emotion processing for instrumental sounds and vocalizations make use of common mental mechanisms. Employing an adaptation framework modeled after [10][14] participants in Experiment 1a were exposed multiple times to an angry or fearful vocal sound and judged whether an instrumental test sound (on a morphed anger-fear continuum) was angry or fearful. Results demonstrated that exposure to angry voices made instrumental stimuli sound angrier and less fearful, while exposure to fearful voices did not show sensitization. In Experiment 1b participants were exposed multiple times to an angry or fearful instrumental sound and judged whether a vocal test sound (on a morphed anger-fear continuum) was angry or fearful. Results showed that when participants were exposed to an angry or fearful instrumental sound and tested on affective voices, sensitization did not occur. Overall, when exposed to angry vocal sounds, listener’s showed a marked increase in anger responses. This indicates that affective vocal sounds have an effect on the emotion perception of affective instrumental sounds. This result was not present for exposure to fearful vocalizations or for repeated exposure to affective instrumental sounds.

The finding of increased ratings of anger to angry vocalizations suggests two ideas. First, there is a unidirectional mechanism responsible for emotion processing in vocal and instrumental sounds, and second, that this unidirectional mechanisms are likely to be specific to a particular emotion. Below we discuss these ideas in detail.

A. Directionality and Emotion Perception

There is a growing support for the relationship between music and language processing such that they share overlapping cognitive resources [26][27]. However, less is known about the directionality of emotion perception in vocal and instrumental sounds. Results from this study indicate that there is a specific directionality for emotion perception in vocal and instrumental sounds. Significantly, this unidirectional relationship reveals that mechanisms used for emotion processing may be shared from vocal sounds to instrumental sounds, but not vice versa. This is in line with other studies that show a unidirectional auditory mechanism for speech and music perception.

For example, [28] compared responses to voices and musical instruments using event related potentials (ERPs), finding evidence for a directional mechanism of emotion perception. Their results show a voice-specific response to sung voices and tones of musical instruments where mechanisms were more activated by voice-stimuli compared to non-vocal stimuli [28][29]. This ‘voice-specific’ response is related to the salience of voice stimuli, reflecting the way attention is allocated, and suggests that emotion perception is mediated by vocalizations.

Further, the second idea resulting from this study suggests that in addition to a unidirectional mechanism, there are some sub-mechanisms used for processing different emotions. Because there was an effect for the emotion anger and not fear, where responses were increased when participants were repeatedly exposed to angry vocalizations, this indicates a critical difference in emotion processing. This difference is likely due to the adaptive value of the emotion anger compared to fear [30].

For example, [16] showed that angry vocalizations evoked changes in the brain such as increased alertness, which caused a sensitivity to information that is important for adaptive behavior [16]. Participants listened to four speech-like, non-word stimuli and rated prosody discrimination of voices (e.g., if the voice was neutral or angry) while being recorded on fMRI. Results show sensitization where the bilateral superficial (SF) complex and the right laterobasal (LB) complex of the amygdala were sensitive to emotional cues from speech prosody, similar to a melody in music. This offers evidence that there are different sub regions of the amygdala that are sensitive to emotional cues from angry voices and that more than one mechanism is used to process emotion in vocal sounds.

In addition, [31] addressed whether brain regions associated with processing the adaptive value of affective expressions are also employed by affective music. Using an event-related fMRI, responses to basic emotions (fear, sadness and happiness, as well as neutral) expressed through faces, nonlinguistic vocalizations and short, novel musical excerpts were compared. Results showed that responses in the amygdala to fearful music and vocalizations were correlated, revealing that the mechanisms used for emotion processing in music might be shared with mechanisms that evolved for vocalizations [30], though this does not address whether emotion processing in music is mediated by emotion processing of voices.

Within the music, speech and emotion literature, effects have been found to indicate the possibility of a directional mechanism for emotion perception and a sub-mechanism that processes categorical differences in emotion. Evidence from ERP studies show voice-specific responses where brain mechanisms are more activated by vocal stimuli, and studies using fMRI have shown specific regions in the auditory cortex that elicit a greater response for vocal sounds, demonstrating the likelihood of a unidirectional mechanism that mediates emotion perception of vocal and instrumental sounds. Additionally, rating studies have shown that angry sounds are perceived as threatening, thereby increasing adaptive behaviors. Behavioral and fMRI studies of face perception demonstrate a sensitization to angry faces in the amygdala that reflects a categorical difference for processing expressed emotion. Taken together, these studies support a unidirectional mechanism of emotion processing in vocal and instrumental
sounds as well as a sub-mechanism that is used for processing categorical differences in sound.

As with previous adaptation studies, a number of limitations apply. Because typical aftereffects to prolonged exposure were not found, one suggested limitation of this study includes the selection of the emotions used in the 2-forced choice tasks. While there is a clear category boundary between anger-fear or anger-sadness continua for face and voice adaptation [10], the difference between these emotions may not be large enough to encompass emotions shared between music and speech. In addition, the forced choice task paradigm may not allow for enough variety in responses to account for musical emotions in that there is not enough variability as compared to arousal and valence ratings of emotion. A wider variety of stimuli, perhaps expanding the types of emotion and instrument classes could be addressed in future work either by utilizing more diverse stimuli, including musical excerpts, or by including other emotions that have a larger categorical boundary.

VI. CONCLUSIONS

Previous research within the music and speech domains in addition to the current findings provide a strong foundation to support the existence of a general and unidirectional mechanism of emotion perception for vocal and instrumental sounds. Previous correlational research has revealed acoustic features that are shared for speech and music such that emotion ratings can be predicted from one domain to the other. More recently, adaptation research has shown that exposure to angry stimuli causes a reduction in the perception of anger, rather than an increase typically seen with the adaptation paradigm. This reduction in anger reflects the underlying motivational salience (approach versus avoidance) that may be driving the perception of emotion, indicating that there is a sub-mechanism used for processing different types of emotions. Implications for work regarding affective-cognitive learning and affective speech indicate that perhaps using only valence and arousal or basic emotions is not enough to encompass emotion in speech and other sounds. Moore thought should be given to the motivational salience of emotions, which will help lead researchers to a better understanding of the degree of overlap between affective qualities in music and speech.

Future studies can explore the overlap of emotion processing for different types of sound stimuli such as the voice, music and environmental sounds. This research will encourage the building of a model for emotion perception in voice and music and further specify mental mechanisms used for emotion processing.

REFERENCES


