ARCHIVOS DE NEUROCIENCIAS | Original article

received 2 November 2022 | accepted 9 February 2023 | published 13 February 2023

Doi: 10.31157/an.v28i2.436

EVALUATION OF A BATTERY OF AUDITORY EMOTIONAL RECOGNITION STIMULI

Kreither Johanna[1,2] | Contreras Renata[1] | Cueto José[1]

1. Associative Research Program (PIA) in Cognitive Sciences, Cognitive Sciences Research Center (CICC), Faculty of Psychology, Universidad de Talca, Chile.
2. Center for Applied Psychology (CEPA), Faculty of Psychology, University of Talca, Chile.

Correspondence

Dra. Johanna Kreither
Faculty of Psychology, Universidad de Talca, Campus Lircay. Avenida Lircay s/n, Talca, Chile. CP 3460000.

jkreither@utalca.cl

Summary

Introduction: auditory emotional recognition (AER) is the ability to recognize emotional states in others, emphasizing prosodic characteristics such as pitch, intensity, and frequency. The literature on this area is scarce and there are currently no validated AER stimulus batteries in Latin America.
Objective: to evaluate a battery of OER stimuli in a sample of Chilean adult population.
Material and methods: 140 adults between 18 and 50 years of age answered an online form containing a battery of auditory stimuli and associated questions. The stimuli were elaborated through a synthetic modulator according to the parameters suggested in the literature. Participants were asked to associate the auditory stimuli with emotional states.
Results: the characteristic F0M is key for the recognition of the emotions Joy and Anger, F0SD is important for the emotion Sadness and HF500 for the emotion Anger. Significant differences were found between the response means between participants who reported a psychiatric disorder at the time of the study and those who did not.
Conclusions: several of the stimuli were evaluated by the participants as representative of some emotional state in a high proportion. Similarities were found between the results of this study and the literature regarding the importance of the presence/absence of HF500 in the identification of the emotions Joy, Sadness and Anger, as well as F0M in the case of Joy and F0SD in the case of Joy and Sadness.
Keywords: *emotion; auditory emotional recognition; prosody.*

Introduction

Emotional recognition (ER) plays a fundamental role in social cognition, as it allows us to predict the behavior of others in our environment and, with this, to adapt to the demands of the social context.[1]

Auditory emotional recognition (AER) is the ability to interpret perceptual components of speech[2] and identify emotional states through these prosodic features,[3] which is of great relevance since it provides more information for a thorough understanding of the social and communicative environment.[4] The study of this type of recognition is limited compared to visual recognition;5 most of the research has focused on a combination of stimuli of different nature.[6,7,3]

Emotion is conveyed by variations in the pitch of the voice, which leads to the process of receptive vocal emotional recognition (also called auditory emotional recognition); the interpretation of this characteristic (pitch) allows individuals to infer the speaker's true emotional state, even if the verbal content of the speech is neutral.8 The voice, due to its pitch and amplitude characteristics, allows expressing more complex emotions than facial expression, since the speaker has the possibility of adding intentions more effectively.[4]

Relatively distinctive patterns of acoustic signals have been identified according to specific emotions: 1) pitch, which has the components of mean fundamental frequency (F0M), fundamental frequency variability (F0SD), and fundamental frequency contour shape (F0contour); 2) intensity, which is made up of mean voice intensity (VoiceintM), voice intensity variability (VoiceintSD), and amplitude rate of rise (ATTACK); 3) quality, composed of the relative proportion of acoustic energy above and below 500Hz (HF500) and first formant bandwidth (F1BW), and 4) temporality, established by the speed of speech (speech rate) and number of pauses in speech (pause ratio).[2,8] Currently, in Latin America and Chile there are no batteries or validation studies of these on the AER. The literature suggests that the vocal characteristics that express these emotional states can be approximated by means of non-vocal and pure sound wave audios, of the same nature as those used in this research. Thus, it becomes necessary to know the exact parameters of these physical characteristics and how changes in them affect the recognition of certain emotions in the Chilean population. Therefore, we sought to evaluate a battery of synthetically generated auditory emotional recognition (AER) stimuli in a population of Chilean adults, establishing the response rate and frequency of AER stimuli according to the physical characteristics associated with each type of emotion.

Material and methods

Sample

We worked with 140 adults of Chilean nationality (104 women), aged between 18 and 50 years. Of the total number of participants, 31 reported having been diagnosed with a psychiatric disorder (this was indicated in the responses to the demographic questionnaire included in the online survey, which did not consider clinical verification, only the report of the participants). Participants read and validated an informed consent form, in which they accepted voluntary participation in the task and the use of the data in the research.

Design and procedure

This project was reviewed and approved by the Ethics Committee of the Faculty of Psychology of the Universidad de Talca.

An online questionnaire was developed through the SurveyMonkey platform, which included the informed consent, with its corresponding acceptance or refusal, a demographic questionnaire and the sequence of auditory stimuli. The questionnaire continued only for those participants who agreed to participate in the study at the informed consent stage.

The instructions for the auditory stimulus sequence required the use of headphones and setting the volume to 50%, a test stimulus, and finally, the 42 auditory stimuli were presented, one stimulus at a time, randomized in their presentation among the participants. Each participant had to play every sound and to identify the emotion linked to sound selecting one of 5 responses (joy, sadness, anger, fear, or no emotion).

Instrument and Variables

a. Demographic questionnaire. The demographic questionnaire included 14 items covering self-reported information related to age, sex, educational level, psychiatric diagnoses, among others.

b. Battery of auditory stimuli. The battery included 42 stimuli representing 4 basic emotions (fear, anger, joy and sadness) and an expression without emotion or neutral. The approximate application time was 15 minutes.

As for the AER stimuli, they were audio files, 500 ms long, created through a synthetic modulator and of non-vocal character, since they correspond to a modulated sound. They were created by Dr. Johanna Kreither, as part of the Fondecyt 11180961 project, using Matlab software. The stimuli followed the characteristics that Juslin and Laukka[9] and Kantrowitz et al.[8] propose for the sound to represent a particular emotion. Additionally, the operational characteristics of the auditory stimuli are detailed below: 1) F0M: middle fundamental frequency, its increase varies the perception of emotions, its levels being 125=low, 225=medium and 378=high. This modulated tone is typically perceived as "joy" (label "Joy 378/125"); 2) F0SD: variability of the fundamental frequency, its fluctuation in pitch varies the perception, its levels being "low" 20, 40 and 60, "medium" 80 and 125 and "high" 150 and 175. This modulated tone is typically perceived as "sadness" (label "Sadness 125/20"); 3) HF500: high energy noise that overlaps the standard tone acting as interference without affecting the fundamental frequency or pitch variability. This sound is typically perceived as "anger" or "fear" (label "Anger/Fear HF500"). This parameter was either present (Yes) or absent (No) in the audios, which divided the stimuli into two groups. These characteristics are summarized in Table 1.

Finally, a section was made after each stimulus, referring to the emotion represented by the item according to the participant (Appendix 3).

Table 1. Simplified summary of sound signal patterns specific to certain basic emotions

Features	Anger	Fear	Happiness	Sadness
Tone F0 M	High	Medium	High	Under
F0 SD	High	Under	High	Under
F0 cont	Medium	High	High	Under
Intensity				
Int M	High	Under	High	Under
Int SD	Big	Big	Big	Small
Attack	High	Under	Medium	Medium
Quality				
F1 BW	Under	High	High	High
HF 500	High	Under	High	Under
Temporary Signals				
Speed of sound	Medium	High	Medium	Medium
Pause ratio	Under	Medium	Medium	High

Nota. Adaptación del resumen aparecido en "Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion", de Juslin y Laukka, 2001.[9]

Analysis Plan

Statistical analysis of the data was performed using SPSS version 26 and Microsoft 365 Excel software.

a. Analysis plan for frequencies per stimulus. In order to show how many responses were obtained for each auditory stimulus and for each emotion, the stimuli were divided into two groups: from 100 to 120, which included auditory stimuli without background noise (HF500), and from 200 to 220, which did present it. After this, the frequency percentages of the emotions associated with each stimulus were calculated.

b. Analysis plan to determine differences according to the report of psychiatric disorder. To identify possible intervening variables in the AER process, the frequency means for each emotion were compared according to the presence or absence of a psychiatric disorder at the time the questionnaire was completed. For this purpose, Student's t-test for independent samples was performed.

Results

Analysis results for frequencies per stimulus

The percentage of responses for stimuli from 100 to 120 are presented in Table 2. Stimuli with high frequencies (over 35%) were found to be associated with Joy, Sadness and No Emotion.

Stimuli 200 to 220 showed high response percentages (> 35%) for the emotions Anger and Joy (Table 3).

Results of analysis to determine differences according to reported psychiatric disorder

Significant differences were obtained in the emotion Fear ($p < 0.01$) between those who did report a psychiatric disorder (M=11.42; SD=5.94) and those who did not (M=8.52; SD=5.15), as well as in the emotion Sadness ($p < 0.05$), between participants who reported a psychiatric disorder (M=10; SD=5.61) and those who did not (M=7.93; SD=4.44). Something similar occurred in the case of No Emotion: the group of those who reported a psychiatric disorder (M=5.26; SD=5.26) and the group who did not report a psychiatric disorder (M=10.37; SD=11.12) presented significant differences ($p < 0.01$).

Table 2. Summary of percentage of responses per stimulus from 100 to 120

Stimulus	Features		Percentage of responses (%)				
	F0M/F0SD (Hz)	HF500	Joy	Anger	Fear	No Emotion	Sadness
100	125/20	No	2,9	10,7	20,7	26,4	**39,3**
101	125/40	No	5,0	15,0	20,0	33,6	26,4
102	125/60	No	5,7	17,9	22,9	30,0	23,6
103	125/80	No	7,1	17,1	17,1	34,3	24,3
104	125/125	No	7,1	11,4	29,3	28,6	23,6
105	125/150	No	13,6	9,3	20,7	**38,6**	17,9
106	125/175	No	16,4	16,4	19,3	30,0	17,9
107	225/20	No	5,0	1,4	20,7	15,7	**57,1**
108	225/40	No	13,6	5,0	14,3	21,4	**45,7**
109	225/60	No	12,9	5,7	20,7	21,4	**39,3**
110	225/80	No	15,0	7,1	14,3	34,3	29,3
111	225/125	No	22,1	10,0	16,4	27,9	23,6
112	225/150	No	30,7	10,7	13,6	25,7	19,3
113	225/175	No	30,7	12,1	14,3	30,0	12,9
114	378/20	No	17,9	0,7	14,3	12,1	**55,0**
115	378/40	No	22,9	5,7	17,9	13,6	**40,0**
116	378/60	No	33,6	5,0	14,3	18,6	28,6
117	378/80	No	**39,3**	5,7	12,9	17,1	25,0
118	378/125	No	**45,0**	5,7	17,1	18,6	13,6
119	378/150	No	**56,4**	7,9	7,9	18,6	9,3
120	378/175	No	**62,1**	7,9	10,0	11,4	8,6

Note: Values > 35% are highlighted in the percentages of responses per emotion. The operational and modulated characteristics are F0M: fundamental frequency, F0SD: frequency height and HF500: high energy noise.

Discussion

This research used synthetic stimuli with variations in tone (F0M/F0SD) and quality (HF500), which allowed the effects of the variability of these three physical components on emotional recognition to be assessed. This potentially allows, as pointed out by Kantrowitz et al.,[8] the use of the same stimulus battery in cross-cultural populations without the barriers of language, adaptation and translation. Therefore, it was possible to assess how the information from the literature presented and the results of previous studies were replicated in the Chilean population.

As for the physical characteristics of the stimuli used, tone is composed of fundamental frequency (F0M) and fundamental frequency variability (F0SD). Regarding the results of this

Table 3. Summary of percentage of responses per stimulus from 200 to 220.

Stimulus	Features		Percentage of responses (%)				
	F0M/F0SD (Hz)	HF500	Joy	Anger	Fear	No Emotion	Sadness
200	125/20	Sí	2,9	**38,6**	32,9	17,9	**7,9**
201	125/40	Sí	0,7	**45,0**	27,1	21,4	5,7
202	125/60	Sí	2,9	**40,7**	31,4	15,0	10,0
203	125/80	Sí	3,6	**40,7**	30,0	20,0	5,7
204	125/125	Sí	3,6	**38,6**	30,0	17,9	10,0
205	125/150	Sí	7,1	32,1	22,9	**25,7**	12,1
206	125/175	Sí	2,1	**35,7**	30,7	25,7	5,7
207	225/20	Sí	3,6	18,6	32,9	13,6	**31,4**
208	225/40	Sí	4,3	25,7	27,9	20,0	**22,1**
209	225/60	Sí	7,1	34,3	23,6	22,9	**12,1**
210	225/80	Sí	13,6	31,4	22,1	23,6	9,3
211	225/125	Sí	10,7	27,1	25,7	27,1	9,3
212	225/150	Sí	13,6	27,1	25,7	25,0	8,6
213	225/175	Sí	12,1	**35,0**	23,6	25,7	3,6
214	378/20	Sí	8,6	9,3	33,6	15,0	**33,6**
215	378/40	Sí	17,1	13,6	28,6	15,7	**25,0**
216	378/60	Sí	20,7	22,1	21,4	16,4	19,3
217	378/80	Sí	**19,3**	15,7	27,9	23,6	13,6
218	378/125	Sí	**35,0**	24,3	19,3	15,7	5,7
219	378/150	Sí	**37,1**	22,9	23,6	13,6	2,9
220	378/175	Sí	**35,7**	27,9	17,1	14,3	5,0

Note: Values > 35% are highlighted in the percentages of responses per emotion. The operational and modulated characteristics are F0M: fundamental frequency, F0SD: frequency height and HF500: high energy noise.

research, it was evidenced that the level of the fundamental frequency (low=125: medium=225; high=378) plays a key role in the percentage of identification of emotions such as Joy and Anger. In stimuli with high levels of F0M, Joy is recognized to a greater extent, while in stimuli with low levels of F0M, Anger is identified to a greater extent. In relation to the variability of the fundamental frequency, it was found that this characteristic determines the percentage of accuracy for the identification of Joy, that is, there is a higher percentage of identification of this emotion when the variability increases, which can be seen in the increase in stimuli 114 to 120 (See Table 2). In the case of Sadness, this variability mainly plays a significant role in its recognition, although it also plays a role in accuracy, which is evident in the best identified stimuli, which had the characteristics 225/20, 378/20 and 225/40, demonstrating that F0SD is the parameter that allows the recognition of this emotion, according to the results of stimuli 107, 114 and 108. Regarding Joy, this feature must be greater than 60 Hz (medium=80 - 125; high=150 - 175) and its increase is directly proportional to the percentage of recognition. For example, the stimulus with parameters 378/80 had a lower percentage than the stimulus with features 378/175. In contrast, for Sadness recognition, F0SD must be found at low levels (20 - 40 - 60), even when the F0M feature is at high or medium level, which extends the findings of the studies by Juslin and Laukka[9] and Kantrowitz et al.,[8] who propose that the parameters for recognizing Sadness should be low F0M and low F0SD. It is worth mentioning that, although the "stereotypical" Sadness stimulus (125/20) is identified in this study as such an emotion, the response rate is among the lowest, implying that its recognition is more linked to the F0SD feature than to the F0M level.

The (HF500) presence in the stimuli, according to the literature, is linked to an increase in the perception of Anger and a decrease with respect to Sadness.[8] This is consistent with the results of this study: the first set of stimuli, without HF500 noise, did not show significant percentages of the perception of Anger. By contrast, the second set, did and included an increasing of such emotion, while significant percentages in the perception of Sadness were eliminated. The emotion Joy was identified transversally in both sets of stimuli, but its recognition was reduced in the group containing the HF500 noise, contrary Juslin and Laukka findins.[9]

No significant perceptions (>35%) for Fear were expected. The study by Kantrowitz et al.,[8] on which this research was based, did include fear. Thus, its specific effects and its recognition are unknown. On the other hand, Juslin and Laukka[9] claim possible characteristics for a stimulus to be correctly identified as Fear. However, such result is not verified in the present research. This may be due to the fact that the stimuli in the present study were modified according to 3 prosodic aspects and Fear may require the F0SD modulation at low levels along with other aspects, such as the shape of the fundamental frequency contour (F0cont), which would have to be at high level.[2] This pitch characteristic varies in this emotion more than in others, for example, Sadness, with which it shares several parameters.

In the studies that worked with auditory stimuli for OER, together with samples of clinical population (schizophrenia) and controls,[2,8] significant differences were found in the performance of the groups according to emotional recognition, without emotional discrimination. The present investigation yielded significant differences in the percentage of identification responses between participants who reported a psychiatric disorder and those who did not with respect to the emotions Fear and Sadness and the expression No Emotion, which is partially consistent with the literature, since the emotions Joy and Anger showed no differences, in contrast to previous studies. This partial difference could be due to the fact that the stimulus parameters generate different responses in the Chilean population and in the studies by Leitman et al.[2] and Kantrowitz et al.[8] .

It is worth mentioning that, to the authors' knowledge, this is the first evaluation of a battery of stimuli in the Chilean population that aimed at establishing the parameters of emotion recognition by means of auditory stimuli in adults, and to evaluate them in an exploratory manner. For this same reason, it was decided not to discard the results of the participants who reported a psychiatric disorder, since the percentage of prevalence of these is the one that would be found in the Chilean population.

In conclusion, this study in the Chilean population achieved results like those of previous research by Juslin and Laukka[9] and Kantrowitz et al.[8] based on an English-speaking population. On the one hand, it was confirmed that the F0M parameter at high levels determines the identification of Joy, and the F0SD parameter at low levels that of Sadness, with the exception that, unlike previous studies, high F0M and low F0SD still allow the recognition of Sadness. The former shows that in this population F0M does not play a determining role in the identification of this emotion. On the other hand, the presence or absence of the high energy sound HF500 is related to the appearance of Anger and the disappearance of Sadness, although it is amplified for the emotion of Joy. Previous research claims that high energy sound should not recognized emotion clearly. In this study, by contrast, it only decreased its percentage of recognition. It never dropped below 35%, especially in stimuli 218 to 220.

This study is relevant given the limited number of studies on auditory emotional recognition and the lack of research in Spanish on the subject. In addition, it was found that, contrary to what Kantrowitz et al.[8] postulate regarding the use of batteries of synthesized nonverbal stimuli and the recognition of the emotions they evoke, according to different cultural contexts, there are smaller differences in the identification of certain emotions. Probably, this is due to the fact that these emotions are determined by cultural factors, both in their expression and interpretation. Such issue opens new lines of research on the evaluation of specific parameters of the stimuli and their physical characteristics (F0M, F0SD and HF500) adapted to the participant population.

Among the recommendations, for future studies in Latin American populations, we recommend having a larger number of participants to ensure larger number of participants the validity and usefulness of the data, as well as control participants and patients with a clinically verified diagnosis, as part of the inclusion criteria. In addition, a greater production of this type of studies with different prosodic parameters of auditory emotional stimuli is needed.

Bibliography

1.Celdrán J. The importance of emotional recognition (ER) in childhood : study, proposal and validation of an ER assessment test aimed at children [Doctoral dissertation]. Murcia: University of Murcia; 2018. Available at: http://hdl.handle.net/10201/55679

2.Leitman D, Laukka P, Juslin P, Saccente E, Butler P, Javitt D. Getting the cue: Sensory contributions to auditory emotion recognition impairments in schizophrenia. Schizophr Bull. 2010;36(3):545-56. doi: 10.1093/schbul/sbn115.

3.Tabernero M, Politis D. Facial recognition of basic emotions and emotional prosody in behavioral variant Frontotemporal Dementia. Latin American Neuropsychology. 2012;4(4):36-41. Available at: https://www.neuropsicolatina.org/index.php/Neuropsicologia_Latinoamericana/article/view/127

4.Morningstar M, Nelson E, Dirks M. Maturation of vocal emotion recognition: Insights from the developmental and neuroimaging literature. Neurosci Biobehav Rev. 2018;90:221-30. doi: 10.1016/j.neubiorev.2018.04.019.

5.Chaturvedi R, Kraus M, Keefe RS. A new measure of authentic auditory emotion recognition: Application to patients with schizophrenia. Schizophr Res. 2020;222:450-4. doi: 10.1016/j.schres.2019.11.043.

6.Pascual S, Solé B, Castillón J, Jesús A, Tejedor C. Affective prosody and facial and verbal recognition of emotion in schizophrenia. Rev Psychiatry Fac Med Barna. 2005;32(4):179-83. Available at: https://sid-inico.usal.es/idocs/F8/ART13456/prosodia_afectiva.pdf

7.Philip RC, Whalley HC, Stanfield AC, Sprengelmeyer R, Santos IM, Young AW, et al. Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. Psychol Med. 2010;40(11):1919-29. doi: 10.1017/s0033291709992364.

8.Kantrowitz JT, Leitman DI, Lehrfeld JM, Laukka P, Juslin PN, Butler PD, et al. Reduction in tonal discriminations predicts receptive emotion processing deficits in schizophrenia and schizoaffective disorder. Schizophr Bull. 2013;39(1):86-93. doi: 10.1093/schbul/sbr060.

9.Juslin PN, Laukka P. Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. Emotion. 2001;1(4):381-412. doi: 10.1037/1528-3542.1.4.381.