



HAL
open science

Translating soundscape descriptors with facial emojis

Pierre Aumond, Marlène Gaillard, Lise Rouy, Enrique Suárez, Catherine Lavandier

► **To cite this version:**

Pierre Aumond, Marlène Gaillard, Lise Rouy, Enrique Suárez, Catherine Lavandier. Translating soundscape descriptors with facial emojis. *Applied Acoustics*, 2023, 208, pp.109342. 10.1016/j.apacoust.2023.109342 . hal-04071184

HAL Id: hal-04071184

<https://hal.science/hal-04071184>

Submitted on 17 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Translating soundscape descriptors with facial emojis

Pierre Aumond^a, Marlène Gaillard^b, Lise Rouy^a, Enrique Suárez^c,
Catherine Lavandier^b

^a*Univ Gustave Eiffel, CEREMA, UMRAE, Bouguenais, F-44344, France*

^b*CY Cergy Paris Univ, ENSEA, CNRS, ETIS lab, UMR
8051, Cergy-Pontoise, 95302, France*

^c*Institute of Acoustics, University Austral of Chile, General Lagos
2086, Valdivia, 5090000, Los Ríos, Chile*

Abstract

To describe the affects related to the perception of sound environments, the ISO 12913 standard proposes the use of 8 verbal scales (pleasant, chaotic, vibrant, uneventful, calm, annoying, eventful and monotonous) inspired by the theory of emotions proposed by Russell et al., and the Swedish Soundscape Quality Protocol. In the context of the international development of this standard, the issue of translating these terms into each language is currently being studied by about 20 research teams around the world in the framework of the "Soundscape Attribute Translation Project". As a counterpoint to this approach, for languages that will not be considered in the project or for illiterate populations (e.g. children), this study proposes to evaluate the relevance of using facial emojis to translate the verbal attributes. This work proposes a replication of the protocol set out in the "Soundscape Attribute Translation Project" by presenting facial emojis instead of the verbal attributes. A descriptive analysis of the results through a principal compo-

Email address: pierre.aumond@univ-eiffel.fr (Pierre Aumond)

ment analysis suggests that the variability in the description of emotions with non-verbal attributes can be decomposed on two principal axes (valence and arousal), using verbal attributes and that among 8 verbal attributes, 5 could be replaced by emojis in a direct way. Within the framework of ISO/TS 12913-3 and by transforming the attributes on the two main axes ISOPleasant and ISOEventful, a comparative analysis was also carried out with an experiment following the same protocol but using French verbal attributes. The latter shows a strong correlation between the positioning of the audio recordings on the ISOPleasant ($r=0.95$, $p<0.001$) and ISOEventful ($r=0.77$, $p<0.001$) axes. The results are encouraging and the use of facial emojis could potentially be proposed within the ISO standard, although a future extension of this study should be conducted by comparison with other languages/cultures and their acceptance by illiterate populations to confirm the relevance of using facial emojis to translate soundscape descriptors.

Keywords: soundscape, facial emojis, SATP

PACS: 0000, 1111

2000 MSC: 0000, 1111

1 **1. Introduction**

2 The definition of soundscape is standardized and defined since ISO 12913-
3 1:2014 as

4 ”the acoustic environment as perceived or experienced by and/or
5 understood by a person or people, in context”

6 . To best describe emotions related to experiences, Russell et al. presented
7 a model in 1980 in which affective verbal descriptors can be set in a circle

8 in the following order: pleasure, excitement, arousal, distress, displeasure,
9 depression, sleepiness, and relaxation called the the "circumplex model of
10 affect" (1). Then, an adaptation of these verbal attributes to the specific
11 perception of sound environments was first proposed by Axelsson et al. (2).
12 The popularity of this tool, called the Swedish Soundscape Quality Protocol
13 (SSQP), for describing the soundscape has led to its inclusion in the ISO/TS
14 12913-2:2018 through the following 8 adjectives: pleasant, vibrant, eventful,
15 chaotic, annoying, monotonous, uneventful, and calm.

16 However, the use of these verbal attributes (initially in Swedish, and
17 written in English in the ISO norm) immediately raises a linguistic issue.
18 The translation and multi-cultural interpretation of these terms is therefore
19 a matter of concern. For example, in a study conducted between Korea,
20 France and Sweden, while there were great similarities in the evaluation of
21 the pleasantness of soundscapes between the countries, differences were found
22 in the eventfulness of the soundscapes (attributed to different socio-cultural
23 effects) (3). Similarly, in a comparison of French and English soundscape
24 ratings in three studies conducted in Montreal, the authors show that the
25 French translation of the proposed SSQP terms yielded similar components
26 to the English translation, although slight differences were again found in
27 terms related to arousal descriptors (4). The studies conducted by Lam et
28 al. and Watcharasupat et al. have shed light on the challenges of translating
29 languages. Lam et al. focused on the translation difficulties of two neigh-
30 boring countries, Singapore and Malaysia, which share the same language -
31 Bahasa Melayu or standard Malay. Meanwhile, Watcharasupat et al. exam-
32 ined the intricacies of translating the Thai language. Both studies utilized

33 highly advanced translation methodologies, showcasing the complexities in-
34 volved in the translation process (5; 6).

35 To avoid any translation bias, Papadakis et al. proposed and applied a
36 methodology for the translation of soundscape attributes (7). This method is
37 very rigorous but requires significant resources: several bicultural and bilin-
38 gual translators, with knowledge of the tool to be evaluated, and more diverse
39 than only academy members. It then points out the difficulty of choosing
40 the most relevant attributes to describe the soundscape from one language
41 to another.

42 The Soundscape Attribute Translation Project (SATP) proposes to an-
43 swer these questions through a common protocol that will allow each of the
44 countries involved to stabilize a translation of the standard soundscape at-
45 tributes. An international collaboration has started in 2019 with now about
46 20 languages involved in this project. But this initiative can never be carried
47 out in all the languages used around the world. There is also the question
48 of specific audiences, illiterate as young children, who cannot be interviewed
49 through verbal attributes.

50 In psychology, there were many successful attempts to substitute words
51 with pictures. Since the nineties, the Self-Assessment Manikin (SAM) has
52 been probably one of the most popular initiatives, offering a non-verbal pic-
53 torial assessment tool to measure valence, arousal and dominance affective
54 dimensions of stimuli (8). It is a drawing of a man expressing different emo-
55 tional states through the expression of his face or the variation of his body
56 shape or the addition of abstract symbols (see Figure 1).

57 Although gender and age biases have sometimes been highlighted (9; 10),

Measuring Emotion

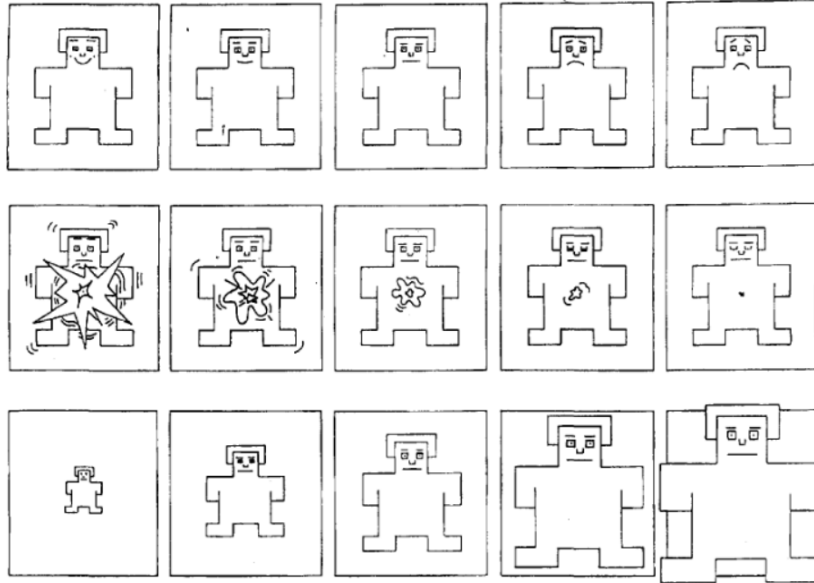


Figure 1: The Self-Assessment Manikin (SAM) used to rate the affective dimensions of valence (top panel), arousal (middle panel), and dominance (bottom panel) - figure from Bradley et al. (8)

58 SAM is regularly used in cross-cultural or cross-generational studies (11; 10)
59 as a reference tool. For example, after a review of the literature, Morris et
60 al. state that emotional responses to a "standardized global advertisement"
61 collected through the SAM, are very close between different cultures (e.g.
62 USA and Taiwan). More generally, the SAM is often used as a reference
63 when adapting to different cultural contexts stimulus databases for psychol-
64 ogy studies (audio - IADS (12), verbal - ANEW or pictorial - IAPS (13)). The
65 Affective Norms for English Words (ANEW) database is a set of normative
66 emotional ratings for a large number of words. When it comes to adapting
67 this database to other languages, the SAM is often used as a common lan-

68 guage or reference tool (e.g., in Italian, Spanish, or Portuguese (14; 15; 16)).
69 In a study led by Huang et al., the SAM is even used to highlight cultural dif-
70 ferences in image-related affect between Chinese and American young adults.
71 This reflects the team’s confidence that the interpretation of the tool is not
72 biased by the culture of the participants (13).

73 For soundscape assessment, this tool has already been investigated in
74 some studies reported by Stevens et al. (17; 18; 19). These authors state
75 that the SAM is an appropriate method for the assessment of soundscapes,
76 more intuitive and less time-consuming than the use of verbal scales. They
77 also show significant relationships between the SAM and soundscape catego-
78 rization or heartbeat measurement, as would be expected with verbal scales.
79 There do not seem to be any observed limitations to the use of non-verbal
80 tools such as the SAM for soundscape assessments.

81 Since the 2010s, there has been a very strong development of the pictorial
82 language linked to emojis. They now have their own Unicode and are stan-
83 dardized across platforms ¹. Also, it is commonly accepted that emotions in
84 facial expressions are perceived in a similar way in different cultures for a
85 similar context (20; 21; 22). It can therefore be assumed that cross-cultural
86 differences in the interpretation of facial emojis are less than those that exist
87 when using verbal scales (e.g. where differences may arise from irrelevant
88 translation). This is a shared hypothesis of many articles, particularly in
89 marketing, and food assessment researches (23; 24; 25; 26). A literature re-
90 view covering the period 2001-2021 and including 23 final studies, shows that

¹emojipedia - <https://emojipedia.org/>

91 a consensus has not yet been reached on the transcription of emotions from
92 smileys in psychological studies, but shows emerging concordances about
93 some pictorial symbols (27).

94 Among these studies, 1082 participants were asked to place 74 facial emo-
95 jis on a nine-point scale to assess their valence and arousal level. The authors
96 conclude that emojis could accurately display human emotional states (23).
97 Also, Jeager et al. state that facial emojis have the potential to be a strong
98 alternative to affective word questionnaires (24). Kaneko et al. examined
99 the cross-cultural use (Germany, Japan, the Netherlands, and the UK) of a
100 grid of (non-standard) emojis. They concluded that this tool appears to be
101 language-independent and ideal for cross-cultural research on food-related
102 emotions (28). The non-standardized emojis, called "EmojiGrid", were also
103 used by Toet and van Erp in 2021 (29) for audio and video clips. They
104 showed that facial emojis can be used as a tool for affective self-assessment
105 of emotions generated by sound and video in at least an equivalent manner to
106 the SAM. Finally, a study from 2020 presents a comparison of several inter-
107 faces for emotion collection including a verbal scale, the SAM, and the use of
108 emoticons (30). Their main conclusion is that all three tested user interfaces
109 are effective ways of assessing subjective feelings, each retaining its interests
110 and particularities. These recent researches enable us to hypothesize that the
111 attributes of facial emojis, compared to the SAM tool, are distinguished by
112 their greater modernity, more common use, and a focus on facial expressions
113 less likely to generate cultural or gender bias.

114 In this paper, we propose a replication of the SATP protocol by replacing
115 the verbal scales with the use of facial emojis as described in section 2. The

116 results in section 3.1 present the ability of emojis to translate soundscape
117 attributes through a principal component analysis of the collected data. The
118 section 3.2 offers a comparative analysis of the experiment conducted in
119 France with the French verbal attributes.

120 **2. Materials and Methods**

121 *2.1. Complementary works*

122 In addition to the main SATP experiment using facial emojis attributes,
123 conducted in November 2021 in the Ecole Centrale de Nantes laboratory and
124 presented in section 2.4, two unpublished works were important to design
125 the experiment and perform the analysis :

- 126 • a first one, conducted in August 2019 on the web with Chilean and
127 French participants, using facial emojis attributes but with a different
128 methodology from SATP, presented in section 2.1.1,
- 129 • a second one, conducted in April 2020 at the Cergy-Pontoise Univer-
130 sity laboratory, using French verbal attributes and following the SATP
131 protocol, presented in section 2.1.2.

132 *2.1.1. Pre-study using facial emojis*

133 In August 2019, the first study was carried out to determine whether the
134 use of facial emojis to describe the soundscape was relevant. This experi-
135 ment also aimed to see the sensitivity of the attributes (verbal and emojis)
136 to describe soundscape for two populations from different countries and with
137 different languages (Chilean and French). An online questionnaire, designed

138 at Austral University of Chile, was sent to French and Chilean participants.
 139 After giving their consent to participate in the study, the participants were
 140 first asked to recall in memory a familiar soundscape (near their home, near
 141 their workplace, etc.). Then 24 facial emojis and the 8 attributes were pre-
 142 sented to them one by one in a random order on their screen. They were
 143 then asked to indicate on a scale of 0 to 100 whether the verbal or emoji
 144 descriptor matched the recalled soundscape.

145 The table 1 shows the verbal attributes presented in Spanish and French
 146 and their aimed translation to English. The 24 facial emojis have been se-
 147 lected from the open source image database (see Section 2.3) and are shown
 148 in Figure 2.



Figure 2: Selected emojis for the online experiment using facial emojis with their corresponding unicode text

149 99 Chilean (63 males, 36 females, Mage = 39 years, SDage = 15 years)

Table 1: The soundscape attribute terms used in the French and Chilean versions of the experiment (and their equivalent in English).

English	Spanish	French
Pleasant	Agradable	Agréable
Annoying	Desagradable	Désagréable
Calm	Tranquilo	Tranquille
Chaotic (1)	Caótico	Chaotique
Chaotic (2)	Agitado	Agité
Vibrant (1)	Estimulante	Stimulant
Vibrant (2)	Interesante	Intéressant
Monotonous (1)	Monótono	Inintéressant
Monotonous (2)	Aburrido	Ennuyeux
Eventful (1)	Animado	Animé
Eventful(2)	Con muchos eventos	Mouvementé
Uneventful (1)	Sin eventos	Inerte
Uneventful (2)	Estático	Amorphe

150 and 46 French (34 males, 12 females, Mage = 36 years, SDage = 12 years)
151 have participated to the study. Figure 3 shows principal component analyses
152 of both data sets. The principal component analyses are computed only on
153 the verbal scales, the facial emojis are added as supplementary variables.

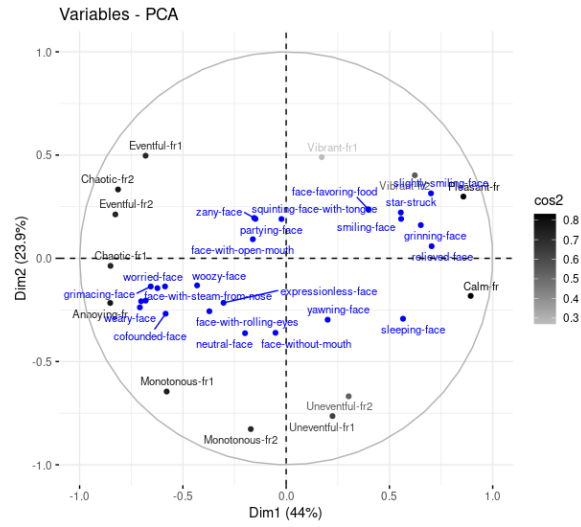
154 In view of these results, some initial conclusions can be drawn:

- 155 • It is possible to explore the circle of affects in a similar way, whether
156 with verbal attributes or with facial emojis.
- 157 • The position of facial emojis in relation to verbal attributes in the
158 French and Chilean groups is similar.

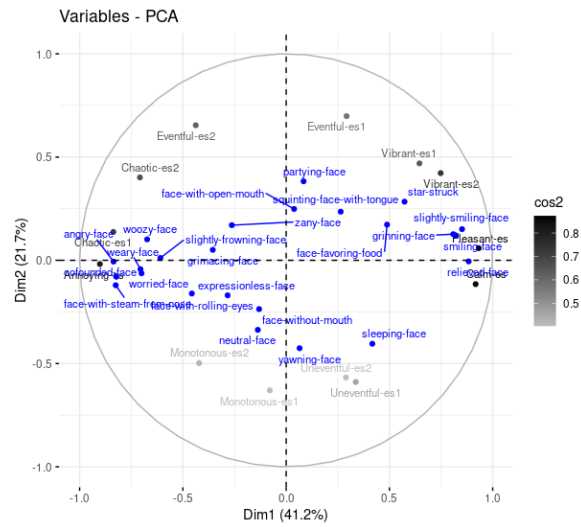
159 After a more detailed analysis, we considered that the results of this
160 pre-study are not strong enough to draw strong conclusions. The method of
161 imagining one's own soundscape raises more questions than it answers. While
162 a marked difference between the two populations could have allowed us to
163 reject the hypothesis of country agnosticity for either type of attribute (verbal
164 or pictural), this is not the case. Because of these concerns, it was necessary
165 to continue this work with the study presented in this paper. Despite this,
166 this exploration was necessary to motivate the present study and to properly
167 select the facial emojis presented to participants (11 of the 24 tested emojis
168 were chosen for the present study).

169 *2.1.2. French SATP*

170 To help to assess the relevance of translating soundscape descriptors with
171 facial emojis, a comparative analysis of the French results with the results
172 from this experiment is proposed in this study in section 3.2.



(a) French speaking participants



(b) Castilian speaking participants

Figure 3: Importance (squared cosinus) of the 24 facial emojis and 13 verbal scales in Component 1 and 2. The principal component analysis is computed only on the verbal scales, the facial emojis are added as supplementary variables.

173 The experiment protocol was approved by the Cergy-Pontoise ad hoc
174 committee (equivalent to the ethical committee). The perceptual test took
175 place between 25 October and 16 November 2021 in a sound-proof listening
176 room at the Ecole Centrale de Nantes, with a background sound level inferior
177 to 40 dB(A), using an external audio interface (MOTU Ultralite Mk-5) and
178 circumaural, acoustically open headphones (Sennheiser HD650). 38 native-
179 speaking or very good French-speaking students (level C2) who did not report
180 any hearing problems participated in the study (23 males, 15 females, Mage
181 = 21.7 years, SDage = 3 years). The participants all signed a contract
182 agreeing to share their results in this research, and as a thank you for their
183 participation, the participants received a €20 gift voucher.

184 The verbal scales presented in Table 2 are the French attributes that were
185 selected for the experiment.

186 The SATP Stage 2 protocol (see section 2.2) was followed from the cal-
187 ibration phase (open-circuit voltage calibration technique) to the data pro-
188 duction. The results of the study carried out in France were added to the
189 international database according to the protocol defined in the SATP.

190 *2.2. SATP stage 2 protocol*

191 In Stage 2 of the SATP protocol, standardized listening experiments are
192 conducted by each research partner with a sample of approximately 30 par-
193 ticipants (aged 18-30, balanced in terms of gender) who are native speakers
194 of the local language. These participants assess the soundscapes based on
195 the data collection instrument translated into the local language. The partic-
196 ipants had to listen to the stimuli with calibrated headphones. All research
197 groups use the same set of auditory stimuli, which have been provided by

198 University College London. This set consists of 27 binaural recordings (30 s)
199 recorded in London during the summer and autumn 2019. Following a pilot
200 listening experiment conducted at University College London, 27 desired au-
201 dio stimuli from a pool of 50+ excerpts have been selected based on having a
202 well-balanced sound sources types' composition and the potential to elicit a
203 broad range of responses for the attributes in use (31). The data and audio
204 of the SATP project can be found in a zenodo repository (32).

205 *2.3. Facial Emojis*

206 The facial emojis have been selected from Twitter's open-source image
207 database (33). Twitter is a very popular social network and many people
208 have already encountered these emojis. Twitter also has the advantage of
209 offering an open-source image database which, moreover, has been used in a
210 preliminary study on the use of facial emojis for the description of affects.(23).

211 Among the Unicode emoji characters and sequences, 107 represent human
212 faces and about 90 represent facial expressions related to emotions (34). From
213 this database, to keep the experimentation time reasonable, 12 smileys were
214 pre-selected. The emojis were selected with the aim of maximizing variance
215 on the circumplex circle of affect. This selection was partly based on the
216 pre-study presented in section 2.1.1 choosing the emojis that maximize the
217 valence/arousal variance. The 11 emojis were also checked to maximize the
218 variance in the Kutsuzawa et al. article (23). We added an additional emoji,
219 😊, even though it was not tested in the pre-study, but because its position in
220 the Kutsuzawa et al. paper (23) suggests it has a complementary placement
221 in terms of valence and arousal.

222 Figure 4 presents the 12 emojis selected for the study.



Figure 4: Selected emojis with their corresponding unicode text

223 *2.4. Experimental protocol*

224 The questionnaire and protocol were identical to that of the SATP Stage
 225 2, described in section 2.2. The experiment took place in April 2020, in a lis-
 226 tening room at ETIS laboratory in Cergy-Pontoise with a background sound
 227 level of less than 35 dB(A), using an external audio interface (RME fireface
 228 UC) and circumaural, acoustically open headphones (Sennheiser HD650).
 229 The following steps were followed for each participant. Before the partici-
 230 pant arrives, the sound levels are calibrated using the open-circuit voltage
 231 calibration procedure presented by Lam et al. (35). Then, the participant
 232 enters the room. After being briefly informed about the purpose of the
 233 experiment, the participant has to answer three quick questions: age, gen-
 234 der, and a self-evaluation of the French-speaking level (native, average, very
 235 good). They must also declare if they have known hearing problems. After
 236 an explanation of the test procedure, the participant is asked to listen to
 237 the 27 sound recordings of the SATP protocol on headphones, one after the
 238 other. The SATP dataset consists of 27 30-second binaural audio recordings
 239 made in urban public spaces in London and one 60-second stereo calibra-
 240 tion signal. For each of the sound recordings, the participant must answer

241 the following question: "To what extent would you say that these emoticons
242 represent the sound environment that you hear?" (in french : "*Dans quelle*
243 *mesure diriez-vous que ces émoticônes représentent l'environnement sonore*
244 *que vous entendez ?*"). To do this, the participant must move a slider initially
245 set at the center (50/100), to the left if the facial emoji presented does not
246 match the sound recording (towards 0/100) or inversely to the right (towards
247 100/100). For the rest of the analysis, results were normalized from 1 to 5,
248 to directly use the tools presented in the ISO/TS 12913-3:2019 standard and
249 presented in section 3.2.

250 At the end of the test, the participant is invited to leave the room after
251 signing a contract related to the acceptance of the disclosure of the results and
252 the good reception of an indemnity of 20€ in the form of a gift certificate. The
253 experiment protocol was approved by the Cergy-Pontoise ad hoc committee
254 (equivalent to the ethical committee).

255 44 individuals took part in the experiment (mostly students) and 38 of
256 whom spoke French as their native language were retained (23 males, 15
257 females, Mage = 27.1 years, SDage = 11 years).

258 *2.5. Statistical tools*

259 Analyses were performed with the R program (36). The analysis code is
260 available on the following github page: [https://github.com/pierromond/Translating-](https://github.com/pierromond/Translating-soundscape-descriptors-with-facial-emojis)
261 [soundscape-descriptors-with-facial-emojis](https://github.com/pierromond/Translating-soundscape-descriptors-with-facial-emojis). To reproduce the analysis in its
262 entirety, please contact the corresponding authors of the study to have access
263 to the input database.

264 **3. Results**

265 *3.1. Descriptive analysis*

266 All the presented results in the descriptive analysis are aggregated by
267 sound recordings using the median value of the attributes before computing
268 the proposed analysis. The aggregated dataset contains 27 sound objects
269 characterized by 12 variables.

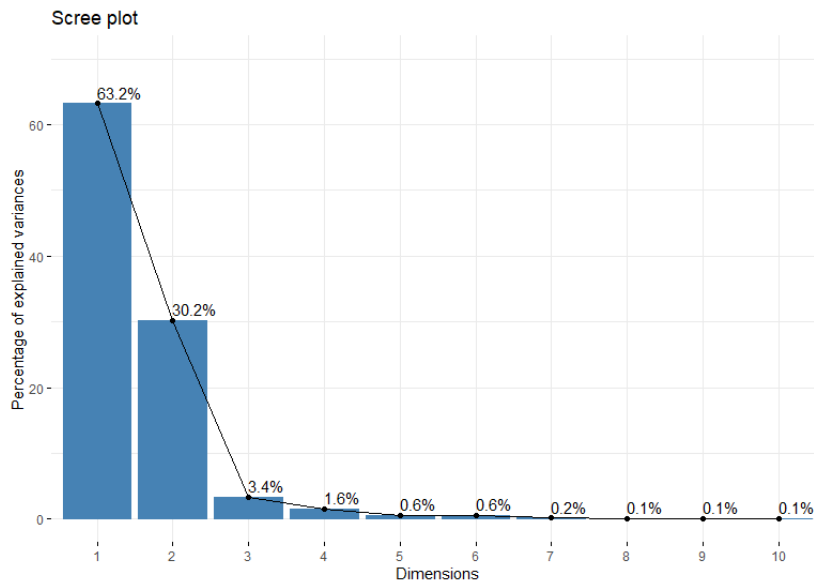


Figure 5: Decomposition of the total inertia

270 Figure 5 shows that the first two dimensions of analyze express 93.4%
271 of the total dataset inertia (63.2% for the first dimension and 30.2% for the
272 second). The two first dimensions are particularly salient for the description
273 of the variance of the data set.

274 Figure 6 shows the projection of the variables "facial emojis" on the two
275 main planes following the principal component analysis. The observation of

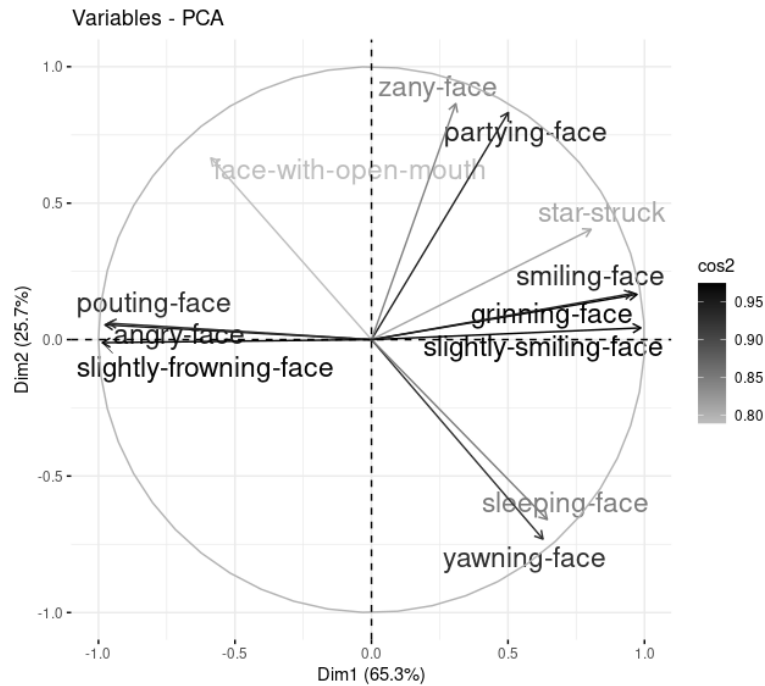


Figure 6: Importance (squared cosinus) of the 12 attribute scales in Component 1 and 2

276 emojis suggests that the first dimension corresponds to valence (e.g. 😊, 😞,
 277 😡) and the second dimension to arousal/eventfulness (e.g. 😄).

278 Figure 7 presents the contribution of each facial emoji on the 5 main
 279 dimensions. It can be seen that the emojis 😊, 😞, 😊, 😞, 😊 and 😡 are
 280 strongly and significantly correlated with the first dimension (valence). The
 281 second dimension (arousal) is mainly related to the emojis 😄, 🤩, 🤩 and 😴.

282 Figure 6, 7 but also Figure 8 show that emojis are well positioned to
 283 advantageously replace the verbal attributes *pleasant*: 😊 or 😊, *annoying*:
 284 😞 or 😞, *calm*: 🤩 or 😴, *vibrant*: 🤩 and *eventful*: 😄. The difficulty comes
 285 when the adjective is linked to the absence of the eventful character. For
 286 example, *uneventful* and *boring* have no easy equivalent.

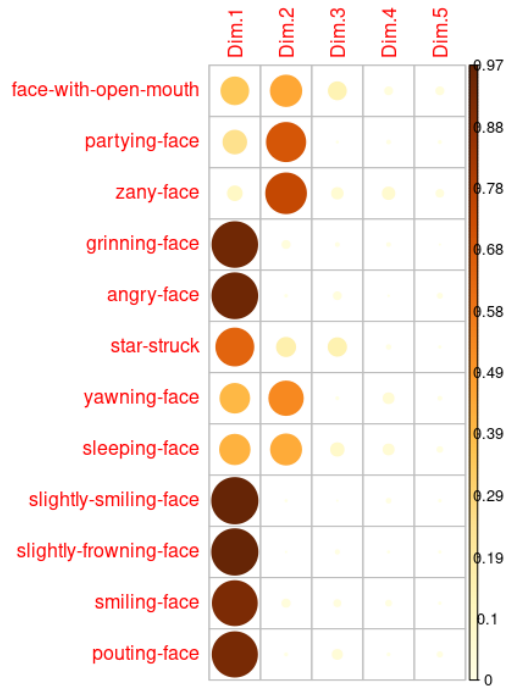


Figure 7: Importance (squared cosinus) of the 12 attribute scales in the first 5 components

287 Finally, the position of the 🙄 facial emoji is a special case. Initially, it was
 288 selected to represent part of the variance related to the arousal dimension,
 289 but as a result of the PCA, it is opposed to the *calm* emojis. Then, it could
 290 make it a good candidate to represent the verbal attribute of the diagonal
 291 axis *chaotic*. However, it also carries an important part of the variance
 292 represented by the third dimension.

293 Figure 8 shows the results of a principal component analysis on axes 2
 294 and 3 of the 12 attributes. 🙄 is not only opposed to 🤔 or 🤨 but is also
 295 projected on a third dimension. For the sake of the comparison with the 2-
 296 dimension representation of SATP experiment, 🙄 is therefore removed from
 297 the analysis. The analysis of the position of this emoji is discussed in more

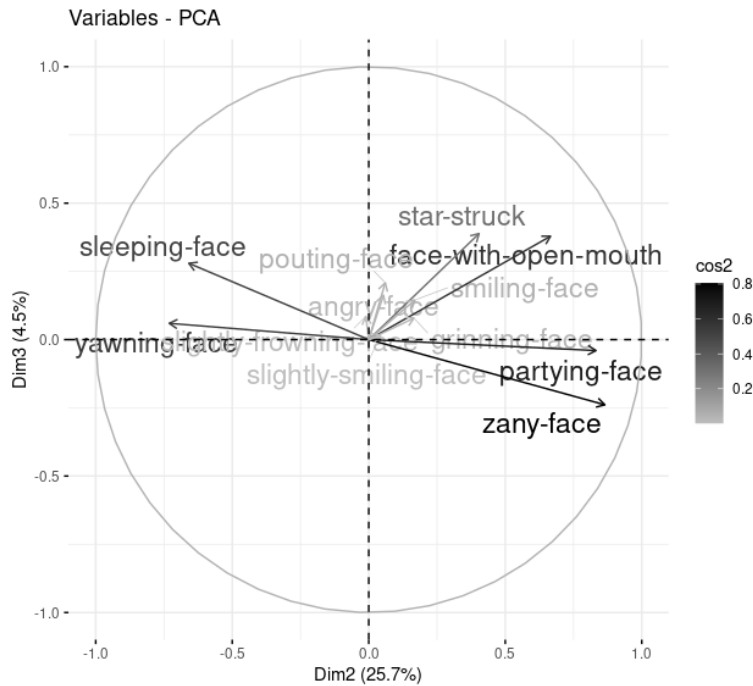


Figure 8: Importance (squared cosinus) of the 12 attribute scales in Component 2 and 3

298 detail in the discussion section 4. Then, in addition to *uneventful* and *boring*,
 299 the attribute *chaotic* has no equivalent emoji either.

300 3.2. Comparative analysis

301 The soundscape analysis, as proposed in ISO/TS 12913-3:2019, can also
 302 be performed by projecting the scores obtained for each of the 8 soundscape
 303 descriptors on the two main dimensions ("valence" and "eventfulness"). This
 304 section proposes to assess whether, despite the absence of equivalent facial
 305 emojis for each verbal attribute, it is possible to describe the sound environ-
 306 ments in terms of valence/arousal in a way that is equivalent to what can be
 307 done with the results of the experiment with the French verbal attributes.

308 The first step is to project the results from each of the adjectives on the
 309 two main dimensions called ISOPleasant and ISOEventful as described in
 310 Annex A of ISO/TS 12913-3:2019. The transformation is first applied to
 311 each individual response, as proposed as a correction to the ISO standard
 312 by Mitchell et al. (37). Then, the dataset is aggregated by sound recordings
 313 using the median value. As before, the aggregated dataset contains 27 sound
 314 objects characterized by 12 variables.

315 Transformation to ISOPleasant/ISOEventful was done based on the equa-
 316 tion 1 and 2 :

$$317 \quad ISOPleasant = (p - a) + \cos\left(\frac{\pi}{180}\right) * (ca - ch) + \cos\left(\frac{\pi}{180}\right) * (v - m) * \frac{1}{4 + \sqrt{32}} \quad (1)$$

$$ISOEventful = (e - u) + \cos\left(\frac{\pi}{180}\right) * (ch - ca) + \cos\left(\frac{\pi}{180}\right) * (v - m) * \frac{1}{4 + \sqrt{32}} \quad (2)$$

318 The correspondences between the parameters of the equation and verbal
 319 attributes and facial emojis are presented in table 2. Some verbal attributes
 320 have one or more easily determined equivalents. But, in other cases, where
 321 no equivalent could be easily found, the assumption of a perfect symmetry
 322 between the antonymous attributes is then made (e.g. Chaotic = 5 - Calm
 323 and so *chaotic* = 5 - 🤪). Beware, this assumption is strong and is discussed
 324 in the discussion section. When several equivalent emojis are possible, a short
 325 sensitivity analysis has been done by replacing one with the other but, as
 326 expected, no substantial difference can be observed in the results presented
 327 hereafter. The final choice was therefore made by simple expertise.

328 From this transformation, Figure 9 shows the difference in localization of

Table 2: Equation variable, English, French equivalents, and facial emoji for each sound-scape attribute.

Eq. parameters	English Verbal	French Verbal	Emojis
p	Pleasant	Plaisant, Agréable	😊
a	Annoying	Gênant, Dérangeant	😞
ca	Calm	Calme, Tranquille	😴
ch	Chaotic	Agité, Chaotique	5 - 🤯
v	Vibrant	Stimulant, Dynamique	🎉
m	Monotonous	Ennuyeux, Monotone	5 - 🙄
e	Eventful	Animé, Mouvemente	😄
u	Uneventful	Inerte, Amorphe	5 - 😐

329 the audio recordings between the experiment performed with the French ver-
 330 bal attributes and the facial emojis on the ISO Pleasant, ISO Eventful plane.

331 Overall, we observe a close positioning of the audio recordings according
 332 to the two methods. The sound objects placed in the upper left corner
 333 (chaotic) are nevertheless less present in the method using facial emojis.

334 Figure 10 shows the correlation between the position of audio recordings
 335 on the ISO Pleasant and ISO Eventful axes. The pleasant dimension is rep-
 336 resented in an equivalent way according to the two methods ($r_{\text{pearson}}=0.95$,
 337 $p<0.001$). The ISO Eventful axis also presents a high significant value of cor-
 338 relation between the two attributes scales, but its value is lower ($r_{\text{pearson}}=0.71$,
 339 $p<0.001$). The bivariate scatterplot shown in the lower left of figure 10 con-
 340 firms that the lowest correlation comes from a discrepancy on the description
 341 of the most eventful audio recordings between the two methods. Note that

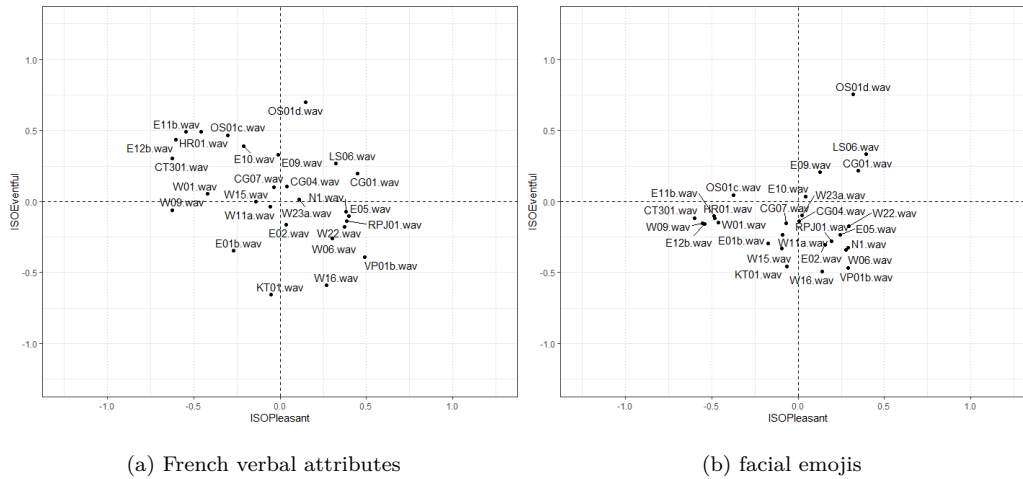


Figure 9: Position of the 27 audio files on the two main axes ISOPleasant and ISOEventful. The transformation is first applied to each individual response, and then the median is calculated at the sound recording level as proposed in Mitchell et al. (37)

342 the Spearman correlation on the ISOEventful axis is 0.74 ($p < 0.001$) and
 343 0.90 ($p < 0.001$) on the ISOPleasant axis. Also, if the dynamics of the link
 344 between the two methods on the ISOEventful axis is not totally linear, the
 345 classification of the audio recordings on this axis, is well maintained.

346 Finally, Figure 11 shows the initial principal component analysis consid-
 347 ering the selected attributes (8 verbal scales or 5 selected emojis scales) incor-
 348 porating the ISOPleasant and ISOEventful axes as complementary variables.
 349 For French verbal attributes, Figure 11 shows that the principal dimensions
 350 ISOPleasant and ISOEventful are very close to the attributes *pleasant* and
 351 *eventful*. However, we can also observe that the angle between the two ISO
 352 axes is slightly obtuse. For facial emojis, the axis between the two ISO
 353 dimensions is very close to being orthogonal.

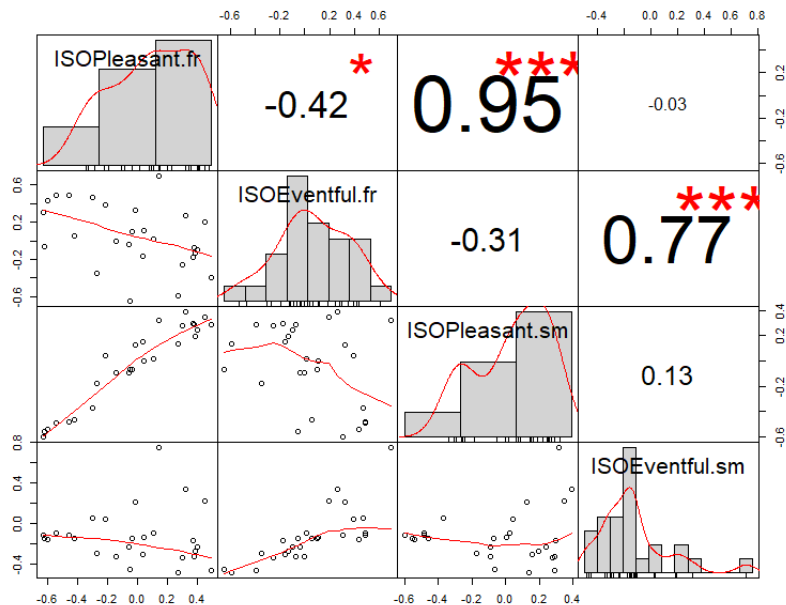
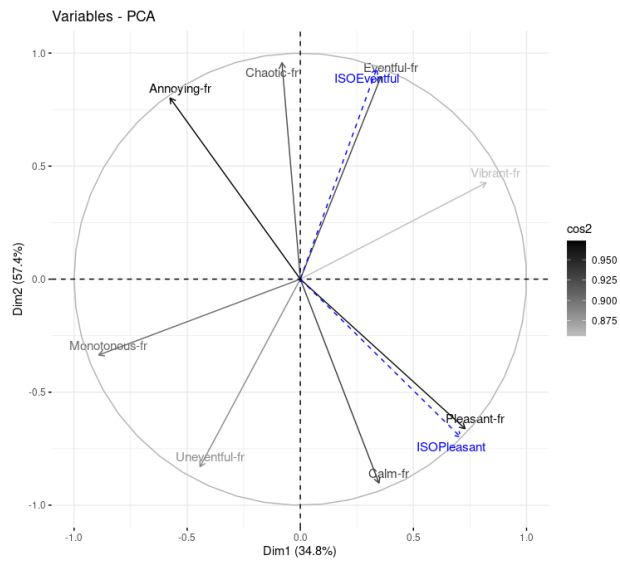


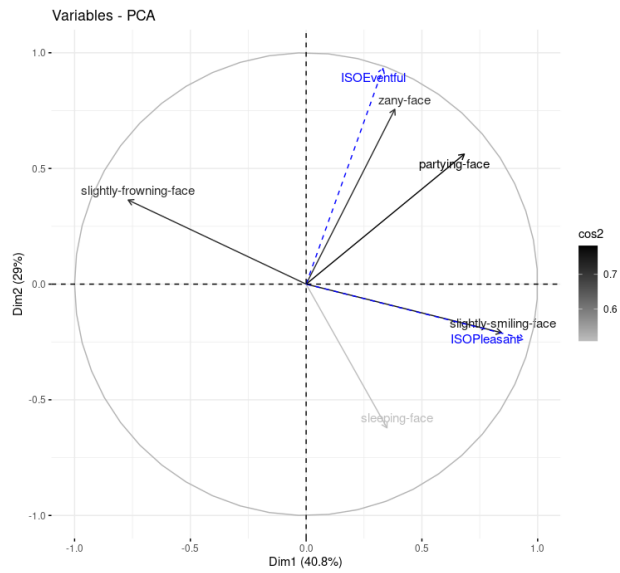
Figure 10: Correlation (Pearson) between the position by the two methods (verbal attributes or using emojis) of the audio files on the main axis ISO Pleasant and ISO Eventful. The distribution of each variable is shown on the diagonal. On the bottom of the diagonal: the bivariate scatter plots with a fitted loess regression red line are displayed. On the top of the diagonal: the value of the correlation plus the significance level as stars. p-values (< 0.001 , < 0.01 , < 0.05) = symbols (“****”, “***”, “**”, “*”).

354 4. Discussion

355 Based on the presented results, it appears that French translation is more
 356 suitable than emojis for describing the Soundscape using the ISO Pleasant
 357 and ISO Eventful variables, especially in simple evaluations that consider partic-
 358 ipants’ language proficiency (French students and native speakers). The
 359 results also suggest that emojis could be a viable option in situations in-
 360 volving illiterate populations (e.g. children) or cross-cultural comparisons



(a) French verbal attributes



(b) facial emojis

Figure 11: Projection of the ISOPleasant and ISOEventful axes (in blue) as complementary variables on the principal component analysis that considers the selected attributes (8 verbal scales or 5 selected emojis scales).

361 between language groups. In order to validate the interest in using facial
362 emojis within the standard ISO/TS 12913-3:2019, it would be necessary to
363 replicate this experiment on several territories and ages and evaluate the na-
364 tional/cultural/social influence on the interpretation of emojis as proposed
365 for example in (38). The results of the pre-study between Chile and France
366 gave a first glimpse of the small differences between the countries. Neverthe-
367 less, the participants have relatively similar languages (originally Latin) and
368 cultures. It would also be interesting to replicate the experiment presented
369 in this study with a more culturally diverse population or in relation to their
370 language skills.

371 Also, a limitation of the study addresses the fact that in the coming
372 years, the standard representation of emojis may change. However, the use
373 of human faces representing emotions in a similar context probably remains
374 something universal, as the studies presented in the introduction indicate.

375 Only the representations of the Twitter database have been tested. How-
376 ever, the conclusions would probably have been very similar to other databases
377 as the representations are very close.

378 The use of smileys can also pave the way for describing elements of the
379 sound environment using a standardized pictorial language beyond emotions.
380 For example, the emoji 🤫 explicitly evokes silence or the emojis 🐦, 🚗 or 🗣️
381 evoke classically assessed urban sound sources. Finally, this unicode database
382 is constantly growing, and if, for example, a facial emoji seems to be missing
383 to describe the *chaotic* dimension, new ones, not available at the time of
384 the experiment, could be coming up like the unicode U+1FAE8 (also called
385 shaking-head) and be appropriate for this attribute.

386 Also, Figure 7 shows that 🙄 carries the weight of a third axis of vari-
387 ability in responses. The shape of the emoji may evoke surprise. It can be
388 hypothesized that this emoji carries the dimension of familiarity with the
389 sound environment, the third dimension proposed in Axelsson’s study (39)
390 and in Viollon’s study (40). This hypothesis deserves to be explored in more
391 detail.

392 If it is difficult to explore certain dimensions of the circumplex circle of
393 affects with facial emojis, perhaps a bipolar scale with a smiley face on each
394 side could help the participant to better understand some representations as
395 EmojiGrid propose (28) or the affective slider (AS) (41). These choices have
396 sometimes been made by research teams and are consistent with Russell’s
397 original publication (1), although there is disagreement about the appropri-
398 ateness of using bipolar scales to represent emotions. Indeed, as presented
399 in (42; 43), emotions can be mixed, and two or more emotions of identical or
400 opposite valence can coexist together.

401 Finally, The use of facial emojis could therefore be an alternative to the
402 self-assessment manikin as already tested. Studies report a high degree of
403 equivalence between these tools, but also a gain in modernity with pictorial
404 visualizations more usual than the self-assessment manikin (44; 41).

405 5. Conclusion

406 This work shows that the use of facial emojis as a replacement for verbal
407 attributes can be a relevant solution as soundscape descriptors (at least for
408 5 out of the 8 ISO 12913 standard verbal attributes). In particular, there is
409 a strong and significant correlation between the use of pictorial descriptors

410 or French words and the placement of 27 tested sound recordings on the pro-
411 jected axes ISOPleasant, ISOEventful (according to ISO/TS 12913-3:2019).

412 The use of facial emojis could then allow illiterate populations (e.g. young
413 children) or those with unusual languages to use this means of communica-
414 tion when using a protocol close to the ISO 12913 standard and offers an
415 alternative to the self-assessment manikin.

416 More broadly, one may wonder if the use of facial emojis is not a more
417 cross-cultural language than words. Only a study extended to more countries
418 could answer this question, by choosing, for example, 3 or 4 of the countries
419 which present the most difference in the analysis which will integrate all the
420 languages having participated in the SATP project.

421 **Acknowledgement**

422 The authors would like to thank Mathieu Lagrange, theIRSTV research
423 federation, and the LS2N laboratory for having lent the laboratory of the
424 Ecole Centrale de Nantes for the duration of the study on French verbal
425 attributes.

References

- [1] J. A. Russell, A circumplex model of affect, *Journal of Personality and Social Psychology* 39 (1980) 1161–1178, place: US Publisher: American Psychological Association. doi:10.1037/h0077714.
- [2] Axelsson, M. E. Nilsson, B. Berglund, The Swedish soundscape-quality protocol, *The Journal of the Acoustical Society of America* 131 (4)

(2012) 3476–3476. doi:10.1121/1.4709112.

URL <https://asa.scitation.org/doi/10.1121/1.4709112>

- [3] J. Y. Jeon, J. Y. Hong, C. Lavandier, J. Lafon, Axelsson, M. Hurtig, A cross-national comparison in assessment of urban park soundscapes in France, Korea, and Sweden through laboratory experiments, *Applied Acoustics* 133 (2018) 107–117. doi:10.1016/j.apacoust.2017.12.016. URL <https://www.sciencedirect.com/science/article/pii/S0003682X1730912X>
- [4] C. Tarlao, D. Steele, P. Fernandez, C. Guastavino, Comparing soundscape evaluations in French and English across three studies in Montreal, in: *Internoise 2016*, 2016, pp. 6855–6861.
- [5] K. N. Watcharasupat, S. Jaratjarungkiat, B. Lam, S. Jitwiriyant, K. Akaratham, K. Ooi, Z.-T. Ong, T. Suthiwan, N. Pichetpan, M. Rojtinakorn, W.-S. Gan, Quantitative Evaluation Approach for Translation of Perceptual Soundscape Attributes: Initial Application to the Thai Language (Mar. 2022). doi:10.48550/arXiv.2203.12245. URL <https://arxiv.org/abs/2203.12245v3>
- [6] B. Lam, J. Chieng, K. N. Watcharasupat, K. Ooi, Z.-T. Ong, J. Y. Hong, W.-S. Gan, Crossing the linguistic causeway: A binational approach for translating soundscape attributes to Bahasa Melayu, *Applied Acoustics* 199 (2022) 108976. doi:10.1016/j.apacoust.2022.108976. URL <https://www.sciencedirect.com/science/article/pii/S0003682X22003504>
- [7] N. M. Papadakis, F. Aletta, J. Kang, T. Oberman, A. Mitchell, G. E. Stavroulakis, Translation and cross-cultural adaptation methodology

for soundscape attributes – A study with independent translation groups from English to Greek, *Applied Acoustics* 200 (2022) 109031. doi:10.1016/j.apacoust.2022.109031.

URL <https://www.sciencedirect.com/science/article/pii/S0003682X22004054>

- [8] M. M. Bradley, P. J. Lang, Measuring emotion: The self-assessment manikin and the semantic differential, *Journal of Behavior Therapy and Experimental Psychiatry* 25 (1) (1994) 49–59. doi:10.1016/0005-7916(94)90063-9.

URL <https://www.sciencedirect.com/science/article/pii/0005791694900639>

- [9] C. Sainz-de Baranda Andujar, L. Gutiérrez-Martín, J. Miranda-Calero, M. Blanco-Ruiz, C. López-Ongil, Gender biases in the training methods of affective computing: Redesign and validation of the Self-Assessment Manikin in measuring emotions via audiovisual clips, *Frontiers in psychology* 13 (2022) 955530, iSBN: 1664-1078 Publisher: Frontiers Media SA.

- [10] R. W. Backs, S. P. da Silva, K. Han, A Comparison of Younger and Older Adults' Self-Assessment Manikin Ratings of Affective Pictures, *Experimental Aging Research* 31 (4) (2005) 421–440, publisher: Routledge _eprint: <https://doi.org/10.1080/03610730500206808>. doi:10.1080/03610730500206808.

URL <https://doi.org/10.1080/03610730500206808>

- [11] J. D. Morris, Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response, *Journal of ad-*

vertising research 35 (6) (1995) 63–68, publisher: Advertising Research Foundation.

- [12] J. Redondo, I. Fraga, I. Padrón, A. Piñeiro, Affective ratings of sound stimuli, *Behavior Research Methods* 40 (3) (2008) 784–790. doi:10.3758/BRM.40.3.784.
URL <https://doi.org/10.3758/BRM.40.3.784>
- [13] J. Huang, D. Xu, B. S. Peterson, J. Hu, L. Cao, N. Wei, Y. Zhang, W. Xu, Y. Xu, S. Hu, Affective reactions differ between Chinese and American healthy young adults: a cross-cultural study using the international affective picture system, *BMC Psychiatry* 15 (1) (2015) 60. doi:10.1186/s12888-015-0442-9.
URL <https://doi.org/10.1186/s12888-015-0442-9>
- [14] M. Montefinese, E. Ambrosini, B. Fairfield, N. Mammarella, The adaptation of the Affective Norms for English Words (ANEW) for Italian, *Behavior Research Methods* 46 (3) (2014) 887–903. doi:10.3758/s13428-013-0405-3.
URL <https://doi.org/10.3758/s13428-013-0405-3>
- [15] J. Redondo, I. Fraga, I. Padrón, M. Comesaña, The Spanish adaptation of ANEW (Affective Norms for English Words), *Behavior Research Methods* 39 (3) (2007) 600–605. doi:10.3758/BF03193031.
URL <https://doi.org/10.3758/BF03193031>
- [16] A. P. Soares, M. Comesaña, A. P. Pinheiro, A. Simões, C. S. Frade, The adaptation of the Affective Norms for English Words (ANEW) for Eu-

ropean Portuguese, Behavior Research Methods 44 (1) (2012) 256–269.
doi:10.3758/s13428-011-0131-7.

URL <https://doi.org/10.3758/s13428-011-0131-7>

- [17] F. K. M. Stevens, D. T. Murphy, S. L. Smith, Soundscape Preference Rating Using Semantic Differential Pairs And The Self-Assessment Manikin: Sound and Music Computing 2016 (SMC2016), 2016, pp. 455–462.
URL http://quintetnet.hfmt-hamburg.de/SMC2016/wp-content/uploads/2016/09/SMC2016_proceedings.pdf
- [18] F. Stevens, S. L. Smith, D. T. Murphy, Soundscape Categorisation and the Self-Assessment Manikin: 20th Int. Conf. on Digital Audio Effects (DAFx17), 2017, pp. 481–488.
- [19] F. Stevens, T. D. Murphy, S. L. Smith, The self-assessment manikin and heart rate: responses to auralised soundscapes, in: Interactive Audio Systems Symposium, 2016.
- [20] P. Ekman, W. V. Friesen, Constants across cultures in the face and emotion., Journal of personality and social psychology 17 (2) (1971) 124, iISBN: 1939-1315 Publisher: American Psychological Association.
- [21] P. Ekman, Strong evidence for universals in facial expressions: a reply to Russell’s mistaken critique. ISBN: 1939-1455 Publisher: American Psychological Association (1994).
- [22] A. S. Cowen, D. Keltner, F. Schroff, B. Jou, H. Adam, G. Prasad, Sixteen facial expressions occur in similar contexts worldwide, Nature

- 589 (7841) (2021) 251–257, iSBN: 1476-4687 Publisher: Nature Publishing Group.
- [23] G. Kutsuzawa, H. Umemura, K. Eto, Y. Kobayashi, Classification of 74 facial emoji’s emotional states on the valence-arousal axes, *Scientific Reports* 12 (1) (2022) 1–10, iSBN: 2045-2322 Publisher: Nature Publishing Group.
- [24] S. R. Jaeger, D. Jin, G. S. Ryan, J. J. Schouteten, Emoji for Food and Beverage Research: Pleasure, Arousal and Dominance Meanings and Appropriateness for Use, *Foods* 10 (11) (2021) 2880, iSBN: 2304-8158 Publisher: MDPI.
- [25] J. J. Schouteten, F. Llobell, C. M. Roigard, D. Jin, S. R. Jaeger, Toward a valence \times arousal circumplex-inspired emotion questionnaire (CEQ) based on emoji and comparison with the word-pair variant, *Food Quality and Preference* 99 (2022) 104541, iSBN: 0950-3293 Publisher: Elsevier.
- [26] P. Ferré, J. Haro, M. Pérez-Sánchez, I. Moreno, J. A. Hinojosa, Emoji-SP, the Spanish emoji database: Visual complexity, familiarity, frequency of use, clarity, and emotional valence and arousal norms for 1031 emojis, *Behavior Research Methods* (2022) 1–19 iSBN: 1554-3528 Publisher: Springer.
- [27] A. Cherbonnier, N. Michinov, The Recognition of Emotions Conveyed by Emoticons and Emojis: A Systematic Literature Review, *Technology, Mind, and Behavior* 3 (2: Summer 2022) (Apr. 2022).

doi:10.1037/tmb0000067.

URL <https://tmb.apaopen.org/pub/7uq57c2p/release/1>

- [28] D. Kaneko, A. Toet, S. Ushiyama, A.-M. Brouwer, V. Kallen, J. B. van Erp, EmojiGrid: a 2D pictorial scale for cross-cultural emotion assessment of negatively and positively valenced food, *Food Research International* 115 (2019) 541–551, iISBN: 0963-9969 Publisher: Elsevier.
- [29] A. Toet, J. B. F. v. Erp, Affective rating of audio and video clips using the EmojiGrid, *Tech. Rep. 9:970*, F1000Research, type: article (Apr. 2021). doi:10.12688/f1000research.25088.2.
URL <https://f1000research.com/articles/9-970>
- [30] L. T. Silva, I. S. Oliveira, J. F. Silva, The impact of urban noise on primary schools. Perceptive evaluation and objective assessment, *Applied Acoustics* 106 (2016) 2–9, 00003. doi:10.1016/j.apacoust.2015.12.013.
URL <http://www.sciencedirect.com/science/article/pii/S0003682X15003679>
- [31] F. Aletta, T. Oberman, Axelsson, H. Xie, Y. Zhang, S.-K. Lau, S.-K. Tang, K. Jambrošić, B. De Coensel, K. Van den Bosch, P. Aumond, C. Guastavino, C. Lavandier, A. Fiebig, B. Schulte-Fortkamp, J. Sarwono, A. S. Sudarsono, A. Astolfi, K. Nagahata, J. Kang, Soundscape assessment: towards a validated translation of perceptual attributes in different languages, in: *Proceeding of Inter-Noise 2020*, Seoul, KR, 2020.
- [32] T. Oberman, A. Mitchell, F. Aletta, J. A. Almagro Pastor, K. Jambrošić, J. Kang, Soundscape Attributes Translation Project (SATP) Dataset, version Number: 1.2.1 Type: dataset (Oct. 2022). doi:10.5281/ZEN-

ODO.6914433.

URL <https://zenodo.org/record/6914433>

[33] Twemoji.

URL <https://twemoji.twitter.com/content/twemoji-twitter/en.html>

[34] Diana, home.unicode.org.

URL <https://home.unicode.org/>

[35] B. Lam, K. Ooi, Z.-T. Ong, K. N. Watcharasupat, T. Wong, W.-S. Gan, Assessment of a cost-effective headphone calibration procedure for soundscape evaluations, arXiv preprint arXiv:2207.12899 (2022).

[36] R: The R Project for Statistical Computing.

URL <https://www.r-project.org/>

[37] A. Mitchell, F. Aletta, J. Kang, How to analyse and represent quantitative soundscape data, JASA Express Letters 2 (3) (2022) 037201, publisher: Acoustical Society of America. doi:10.1121/10.0009794.

URL <https://asa.scitation.org/doi/10.1121/10.0009794>

[38] G. Kutsuzawa, H. Umemura, K. Eto, Y. Kobayashi, Age Differences in the Interpretation of Facial Emojis: Classification on the Arousal-Valence Space, Frontiers in psychology 13 (2022) 915550, iSBN: 1664-1078 Publisher: Frontiers Media SA.

[39] Axelsson, M. E. Nilsson, B. Berglund, A principal components model of soundscape perception, The Journal of the Acoustical Society of America 128 (5) (2010) 2836–2846. doi:10.1121/1.3493436.

- [40] S. Viollon, C. Lavandier, C. Drake, Influence of visual setting on sound ratings in an urban environment, *Applied Acoustics* 63 (5) (2002) 493–511. doi:10.1016/S0003-682X(01)00053-6.
URL <http://www.sciencedirect.com/science/article/pii/S0003682X01000536>
- [41] A. Betella, P. F. M. J. Verschure, The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions, *PLOS ONE* 11 (2) (2016) e0148037, publisher: Public Library of Science. doi:10.1371/journal.pone.0148037.
URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0148037>
- [42] J. T. Larsen, A. P. McGraw, J. T. Cacioppo, Can people feel happy and sad at the same time?, *Journal of personality and social psychology* 81 (4) (2001) 684, iISBN: 1939-1315 Publisher: American Psychological Association.
- [43] J. T. Larsen, A. P. McGraw, The case for mixed emotions, *Social and Personality Psychology Compass* 8 (2014) 263–274, place: United Kingdom Publisher: Wiley-Blackwell Publishing Ltd. doi:10.1111/spc3.12108.
- [44] S. R. Jaeger, C. M. Roigard, D. Jin, L. Vidal, G. Ares, Valence, arousal and sentiment meanings of 33 facial emoji: Insights for the use of emoji in consumer research, *Food Research International* 119 (2019) 895–907, iISBN: 0963-9969 Publisher: Elsevier.