DOI: 10.61782/fa.2025.0522

FORUM ACUSTICUM EURONOISE 2025

LISTENING AND TALKING ABOUT SOUND: EXPLORING EFFECTIVE DESCRIPTORS FOR EVERYDAY SOUNDS WITH NON-EXPERT LISTENERS

David Larsson Holmgren[1*] **Nicolas Misdariis**[2] **Sandra Pauletto**[1*]

[1] KTH Royal Institute of Technology, Sweden
[2] STMS Ircam - CNRS – SU, France

ABSTRACT

A key challenge in sonic interaction design is the lack of a shared vocabulary for describing sound in ways non-experts can easily understand. To address this, we investigate the intuitive use of a lexicon originally developed for communication between sound designers and stakeholders. Building on a pre-study with pre-labeled stimuli, this new study tests the descriptors with a larger, unlabeled dataset. Through online listening tests, participants categorized sounds using selected descriptors, and their responses were compared to expert labels. Our findings confirm that descriptors such as Rough, Smooth, Metallic are the most intuitively well understood; Dull, Warm, Round and Non Round are also relatively well understood. These results inform ongoing research on sound preferences and the development of tools for personalized sonic interactions, helping listeners articulate their preferences for user-centered sonic interactions.

Keywords: *listening, sound descriptors, non-expert listeners*

1. INTRODUCTION

Unlike the visual medium, which can be described by colors and brightness for which people have a widespread understanding, the auditory medium lacks a common vocabulary for conveying different acoustic characteristics in a way understandable to non-experts. Some adjectives are more common, but their definitions vary between individuals, which may cause misunderstandings when communicating about sound or designing sound-driven interfaces. This could limit the usefulness of sound as a communication tool, despite its information-carrying capacity.

Carron et al. [1] argue that research on verbal descriptions of sound and communication tools to indicate acoustic characteristics in design practice is lacking. Sound design processes involve stakeholders and other non-experts [2], where this vocabulary gap can lead to miscommunication [3, 4]. Additionally, audio interfaces use words to indicate adjustable parameters, and if these are unclear, the system can be difficult to use. To address this, Carron et al. [1] propose a lexicon of sound descriptors, developed through empirical research with sound designers and non-experts, presented as an online communication tool with sound examples and definitions of terms.

In this study, we use the descriptors from this lexicon without providing definitions or sound examples to assess whether these terms are understood by listeners without explicit experience in sound design (non-experts). A pre-study using sound examples from the lexicon showed that some descriptors are better understood intuitively than others [5]. This paper extends the pre-study, exploring how well these descriptors are understood with new unlabeled sounds and comparing the categorization by three experts and non-expert participants.

**Corresponding author*: *ndtho@kth.se & pauletto@kth.se.*

2. RELATED STUDIES

Several studies have introduced structured taxonomies and lexicons to describe sound, each adapted to a particular domain and application.

Everyday sounds refer to sounds that surround us in our daily lives. When we listen to and describe these sounds, we often adopt an ecological approach by identifying sound sources as well as the events and interactions that produce the sound. Referring to this as *everyday listening* [6], Gaver presents a framework to facilitate and organize the description of everyday sounds, which he calls the *Map of Everyday Sounds*. This is represented as a 2D-space, based on everyday listening and the physics of sound-producing events. Each physical state of matter (solids, liquids, gases) and corresponding basic-level events occupy their own corners of the space, with increasing complexity towards the center, including temporal patternings, compound, and hybrid events. Tan et al. [7] propose a new categorization scheme for urban soundscapes based on sound stimuli recorded in Singapore and evaluated through online tests. Similarly, Torija et al. [8] analyze various locations in Granada to classify and group them into distinct soundscape typologies.

Building on a previously developed sound lexicon [9] and research in perceptual audio evaluation [10], Pedersen and Zacharov [11], introduce a perceptual attribute lexicon presented as the hierarchical visual representation known as the *Sound Wheel*. Aiming to create a standardized lexicon for sound reproduction, using expert-selected attributes to characterize various audio products, The Sound Wheel is organized into a hierarchy with three main groups (*Artifacts*, *Timbre*, and *Spatial*), encompassing 40 attributes in total, with addition of spatial attributes [12, 13]. Building on Axelsson et al. [14] and the ISO 12913 series on soundscape [15], which present the circumplex structure of soundscape perception in a two-dimensional space based on *Pleasantness* (positive to negative emotions) and *Eventfulness* (calm to dynamic soundscapes), Aletta et al. [16] evaluate the cross-cultural validity of these descriptors. Eight core descriptors are assessed across 18 languages to ensure accurate translations from English related to pleasantness and eventfulness. Most languages retained the circumplex structure, but some showed varying confidence levels. Overall, these findings highlight the importance of descriptor choice and the potential issue of cross-cultural validity.

2.1 The Words4sound.speak Lexicon

In [1], Carron et al. create an online sound lexicon called *Words4sound.speak* [1] (hereafter referred to as W4S). The lexicon is designed to serve as a communicative tool for sound designers and non-experts involved in sound design practices [17]. The authors argue that a reduced mode of listening—referring to describing sounds independently of their producing event or source, as inspired by Schaeffer [18]—is critical for sound designers, and emphasize the need for all participants in design projects to be able to describe listened sounds consistently. A list of 41 descriptors was developed consisting of 19 pairs of antonyms and 3 individual descriptors, organized into three distinct typologies: **(1) General qualities:** relating to measurable quantities, such as frequency and sound intensity; **(2) Timbre:** referring to acoustic characteristics relating to sensory qualities; **(3) Morphology:** meaning variations of acoustic parameters in time. Each descriptor is represented by different sound examples and definitions. The lexicon was tested within a French industry setting [17]. It is important to note that the authors do not claim the lexicon to be universally applicable, but a practical tool for discussions on sound design regardless of expertise in sound. Given the focus on reduced listening and non-expert listeners, this lexicon is of particular relevance for our research.

2.2 Pre-study

A pre-study [5] was conducted to investigate the effectiveness of 19 selected descriptors from the W4S lexicon for non-expert listeners. The study used online listening tests with pre-labeled (along one descriptor) environmental sounds from the lexicon as test stimuli.

A cross-continental sample of 30 participants (15 females, 15 males) aged 21 to 78 (mean = 31.5, SD = 11.7) was recruited. Participants matched sounds to the descriptors they found most appropriate and provided feedback on clarity, explaining their choices. Then a second listening test further examined whether descriptors became clearer when presented alongside their antonyms. The analysis suggested that at least half of the descriptors were well understood. Sounds were often described using multiple descriptors, and the pairwise categorization condition improved clarity, helping participants more accurately assign even initially unclear descriptors (e.g. Dull/Bright, Round/Non-Round, Resonant/Dry).

[1] https://speak.ircam.fr

Forthermore, some descriptors remained difficult to interpret, including Rich/Non-Rich, Artificial/Natural, Non-Dynamic/Dynamic, and Nasal/Non-Nasal.

These findings informed the design of the present study by supporting the use of antonym pairs for descriptor presentation, and highlighting the need to test a larger sound corpus to further assess the usefulness and understandability of these descriptors.

3. METHOD

3.1 Aim of the study

Building on previous findings (see Section 2.2), the aim of this study is to identify which Timbre and General Qualities' descriptors from the W4S lexicon are intuitively understood by listeners who are not experts in sound. The descriptors used in this study are 8 antonyms (Smooth/Rough, Dull/Bright, Dry/Resonant, Natural/Artificial, Rich/Non Rich, Nasal/Non Nasal, Dynamic/Non Dynamic, Round/Non Round), and 3 single descriptors (Metallic, Strident, Warm). Their definitions and sound examples can be found on the SpeaK website [1].

Furthermore, we aimed to categorize 30 unlabeled everyday sounds using these descriptors. Ten sounds from the W4S-dataset, already labeled according to one descriptor, were added in this study as a control set. These are the W4S environmental sound examples for Rough, Smooth, Bright, Warm, Resonant, Metallic, Round, Strident, Rich, and Dynamic.

3.2 Stimuli

We chose the stimuli with the goal to capture a broad spectrum of everyday sounds, encompassing most categories within Gaver's taxonomy. Furthermore, additional categories such as *electric* and *digital* sounds were added. Another priority was to ensure diversity in timbre, considering sounds with varied spectral centroid, pitched versus noisy sounds, granular versus sustained envelopes, increasing and decreasing dynamics, as well as regular versus irregular temporal event distributions.

Stimuli were sourced from three primary datasets, the CLOSED-dataset [2], predominantly featuring environmental everyday sounds; the SkatVG-dataset [3], comprising a blend of environmental everyday sounds, sounds of mechanical interactions and digital user interface (UI) sounds; and the W4S examples as mentioned above. Fifteen sounds were collected from the sample library Splice [4], to cover missing sound categories. A final selection of 40 sounds were used in this study.

To ensure that the sound stimuli covered a broad spectrum, each sound was described by the first author according to the following labels:

- *Basic Level Source*: Classifying sounds into solids, liquids, gases, electric or digital.
- *Material*: The main material (when known) involved in sound producing event, e.g. wood, metal, water.
- *Basic Level Events*: Identifying fundamental events such as impacts, drips, or gusts.
- *Temporal Patterning*: Describing temporal patterns like walking, filling containers, or fire.
- *Compound Events*: Determining if a sound comprises multiple basic level events or not.
- *Hybrid Events*: Determining if a sound involves multiple basic level sources.
- *General Temporal Event Progression*: Categorizing the overall progression of events of the sound as fast, slow, faster (increasing over time) or slower (decreasing over time).
- *General Dynamic Progression*: Assessing whether the overall dynamics of the sound increase, decrease, or remain static.
- *Sonic Shape*: Characterizing the shape of the sound as granular, sustained, or a mixture of both.
- *General Temporal Event Distribution*: Analyzing the regularity of events occurrence, below the average frequency threshold of pitch perception of 20 Hz, as regular, irregular, or static.
- *Pitched Components*: Differentiating between sounds as containing pitched or noisy components, or both.

Each sound sample was edited using Audacity [5]. All sounds were trimmed to 5 s with a 250 ms fade-out. The loudness was normalized to -30.0 LUFS. All sounds were

[2] https://www.ircam.fr/projects/pages/closed

[3] https://skatvg.unipa.it

[4] https://splice.com/

[5] https://www.audacityteam.org/

exported to MP3 format at 320 kbps to facilitate faster loading in the online listening environment Gorilla.sc[6].

A table describing the stimuli is in Appendix A. The original sounds can be found in `https://tinyurl.com/9dtx5sd3`

3.3 Participants

Forty participants were recruited through the online platform Prolific[7]. The following screenings were applied: no hearing difficulties, half of the subjects should be male and half female. Participants were 20 male and 20 female with a mean age of 30 (SD = 9.5). Fourteen participants were from Europe, 14 from South Africa, 5 from North America, 1 from India. None of the participants reported hearing problems.

3.4 Test Design

The test started with an introduction, a consent form and a volume test where participants could set their headphones to a suitable level with the instruction to leave it unchanged throughout the test. Then each sound was presented with one question for all descriptors (see Fig. 1 for screenshot excerpt referring to 3 descriptors: Smooth, Rough and Metallic).

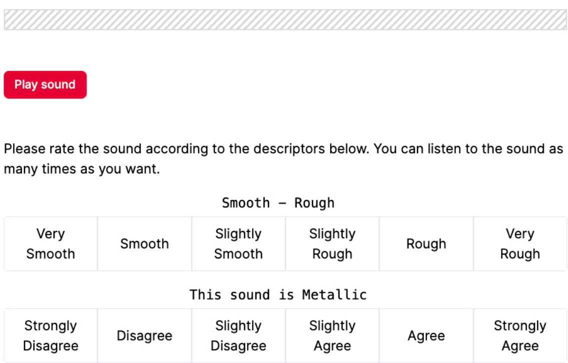

Figure 1. Excerpt from the online listening test displaying the six levels for the opposite descriptors *Smooth/Rough* and the single descriptor *Metallic*.

For each sound, the descriptors were presented as a pair of antonyms (e.g. Smooth/Rough) or as a single descriptor without explicit opposites, e.g. see Metallic as presented in Fig.1. For each line, participants were asked to rate the sound according to 6 levels. For example, for the pair Smooth/Rough the scale was "Very Smooth", "Smooth", "Slightly Smooth", "Slightly Rough", "Rough" and "Very Rough". For single descriptors like Metallic, the participants were asked to rate the following statement "This sound is Metallic" using the options: "Strongly Disagree", "Disagree", "Slightly Disagree", "Slightly Agree", "Agree", and "Strongly Agree". For each sound, there were 11 ratings, either pairs or single descriptors. Furthermore, for each sound, participants were encouraged to write additional descriptors they considered suitable, or "NONE" if they could not propose any. The order of the sounds and the list of descriptors were randomized across participants and stimuli.

After rating all sounds, participants were asked to select which descriptors they found confusing and which clear, and explain why. Finally, participants could provide feedback about the overall test.

4. RESULTS

Since the stimuli were unlabeled, we asked three experts, well-versed in the definitions of the descriptors, to label the stimuli independently. Then, when analyzing the results, we considered that when all three sound experts were in agreement about the labels attributed to a sound the labeling was correct. Additionally, the Fleiss Kappa shows agreement between the samples of the 3 experts for all the descriptors (K>.6 (substantial) for Smooth/Rough; K>.4 (moderate) for Dull/Bright, Rich/Non Rich, Metallic/Metallic NO, Dry/Resonant, Warm/Warm NO; K>.2 (fair) for Strident/Strident NO, Dynamic/Non Dynamic, Nasal/Non Nasal, Round/Non Round; K>.1 (slight) for Natural/Artificial). We then compared the labeling of the participants to the labeling of the experts, and identified which descriptors were used similarly by experts and non-expert participants.

To provide an initial overview, results were first analyzed condensing the data into a binary choice. For instance, selections such as "Very Smooth", "Smooth", and "Slightly Smooth" were all grouped under the descriptor "Smooth" for analysis. Similarly, for single descriptors like Metallic, any selection of "Slightly Agree", "Agree", or "Strongly Agree" categorized the sounds as Metallic. Only descriptors selected by at least 26 out of 40 participants ($p \leq 0.05$) were considered for labeling. the Fleiss Kappa shows agreement between the samples of

[6] `https://gorilla.sc/`
[7] `https://www.prolific.com`

Table 1. Labels agreement between 3 experts and 26 or more participants/subjects ($p \leq 0.05$).

Descriptor:	3 Experts Num sounds:	Subjects+3Experts Num sounds:
Dull	7	5 **(71%)**
Bright	20	9 (45%)
Smooth	12	10 **(83%)**
Rough	20	13 **(65%)**
Dry	15	3 (20%)
Resonant	7	3 (43%)
Natural	16	3 (20%)
Artificial	3	3 **(100%)**
——	——	——
Non Nasal	23	22 **(96%)**
Nasal	2	0 (0%)
Non Rich	20	4 (20%)
Rich	7	4 **(57%)**
Non Round	25	16 **(64%)**
Round	1	1 **(100%)**
Non Dynamic	13	1 (8%)
Dynamic	7	5 **(71%)**
——	——	——
Warm NO	32	22 **(69%)**
Warm	1	1 **(100%)**
Strident NO	28	6 (21%)
Strident	2	0 (0%)
Metallic NO	24	20 **(83%)**
Metallic	4	4 **(100%)**

the 40 participants for all the descriptors (K>.2 (fair) for Smooth/Rough and Metallic; K>.0 (slight) for other descriptors). Note that when in the lexicon there is no opposite descriptor, we indicate the disagreement with the single descriptor accompanied by a NO (e.g. Metallic NO).

Results show that there is agreement between experts and participants for 13 out of 22 descriptor choices (in bold in the table). We consider there to be agreement if all the 3 experts have labeled N sounds as, for example, Dull, and 26 or more participants have labeled at least half these sounds as Dull too.

It must be noted that the number of sounds labeled by the 3 experts with a particular descriptor varies from 1 sound to a maximum of 32. Additionally, descriptors choices defined by exclusion, i.e. as not being a particular descriptor (e.g. Metallic NO), tend to group more sounds because they are less narrowly defined.

Descriptors for which there is clear disagreement are: Bright, Dry, Resonant, Natural, Nasal, Strident, Strident NO, Non Rich, Non Dynamic. Note that Nasal, Resonant, Strident and Non Dynamic are indicated as confusing in the feedback given by participants (see Section 4.1).

When looking at the 10 control stimuli, the W4S sound examples for Bright, Warm, Dynamic, Smooth, Metallic, Resonant, Rich, Rough, Round, Strident, the experts are in full agreement for 9 out of 10 sounds, with the only disagreement for the Warm example, which only 2 experts labeled correctly. When looking at the agreement between 3 experts and participants, 4 (Warm, Bright, Rich and Strident) out of 10 sounds are not labeled according to the main descriptor they are meant to be an example of. Similar findings are reported in [19], where researchers found that despite the stable acoustic definitions of some descriptor (e.g. Bright) people's varying mental representations of the word might create low consensus. These descriptors correspond to many of those for which there is general disagreement or weak agreement between experts and participants (see Table 2).

Both participants and experts seem to agree that, in this dataset, only a few sounds can be described as: Artificial, Round, Warm, and Metallic. Experts agree that two sounds can be labeled as Nasal and Strident, while participants do not label any sound as Nasal or Strident.

4.1 Clarity of descriptors and additional descriptors

Regarding feedback on the descriptors, the majority of participants (23 out of 40) reported that they did not find any descriptors confusing. The most confusing descriptor (8 participants) is Nasal. Following closely are Non-Nasal and Round (6); Artificial and Resonant (4); and Non-Dynamic, Dynamic, Strident, and Non-Round (3). Seven comments were made concerning Nasal, all indicating that participants did not associate the term with non-vocal sounds, e.g. *"I don't have a sense of how nasal sounds like in [relation to] sounds that are not human speech"*. Similar comments were expressed regarding Non-Nasal. The term Artificial also caused confusion, with 4 participants indicating uncertainty about whether it referred to synthesized sounds or non-natural sounds.

One participant wrote: *"I wasn't sure whether this applied to computer-generated sounds or sounds generated by inorganic matter"*. Additionally, 4 comments address the descriptor Dynamic, with participants struggling to intuitively grasp its meaning. One participant wrote: *"[I was] having trouble intuitively understanding what it means in regard to sounds"*. Similar observations were made regarding Resonant. Two participants were also unfamiliar with the meaning of the word Strident. As for the descriptors considered clear, 23 out of 40 participants considered Smooth and Metallic clear, followed by Natural, Dull and Rough (20); Warm (19), and finally Bright and Artificial (18). Then followed by Dry (16), Strident (16), Nasal (15), Non Nasal (15), Round (15), Non Rich (14), Non Dynamic (13), Non Round (12). Nine out of 40 participants did not find any descriptors very clear.

The participants provided 119 interpretations of the descriptors, revealing several notable themes. Concerning Artificial, many participants contrast it with Natural. Some also associate it with machine-generated sounds, one participant stating: *"[Artificial sounds are] easy to recognize as they tend to be very mechanical"*. Regarding Natural, 4 participants associate it with sounds produced by organic sources. Emotions are mentioned in the interpretations of the descriptors Bright, Dull, Rich, Smooth, Strident, and Warm. Example of statements in this directions are: *"Bright: a sound that makes me feel happy"*, *"Dull: boring or [un]interesting"*, *"Rich: satisfaction"*, *"Smooth: the sound was enjoyable"*, *"Strident: a very irritating sound"*, and *"Warm: easy as the warm sounds are more peaceful and calm"*. Some interpretations revolve around acoustic parameters, such as frequency and amplitude: *"Bright: high-frequency sound"*, *"Dull: low-frequency sound"*, *"Nasal: high-frequency [sound that] sounds nasally"*, *"Rich: the sound had variety in tones, pitches"*, and *"Round: a sound [that has] a curve or without high spikes to hear"*. Associations with the material producing the sound are also common: *"Dry: [reminiscent of] sandpaper"*, *"Dull: in my understanding it means something that is solid, like percussing (banging) on a solid structure"*, *"Metallic: shares traits with the sounds that metal objects make"*, and *"Strident: [...] the sound of chalk on a whiteboard"*. Another theme is to use metaphors, such as: *"Warm: feels like home"*, *"Smooth: easy to the ears, like butter"*, *"Rough: feels like a bumpy road"*, and *"Rough: this is like the voice of a smoker"*. Looking at the descriptors suggested by the participants, they can be grouped into four different categories of words. The largest category is *ecological descriptions* of sounds, similar to the taxonomy of Gaver [6]. These include different ways of referring to the assumed sound-producing object, interaction or material of the sound. Examples of these descriptions are the object making the sounds (e.g., "motor", "water bowl", "rain"), the sound-producing action made with the object (e.g., "skateboard rolling", "sweeping rock floor", "melting"), and reference to the material producing the sound (e.g., "porcelain", "scratching wood", "water flows"). The second category of additional descriptors is *onomatopoeic* words, which imitate the sound they represent, e.g., "bang", "tick", "tock", "poing", "hum", "squeak". The third category includes words related to *acoustic* descriptions, such as "high pitch", "loud", and "soft". The fourth category involves *other* descriptors not specifically related to the acoustic properties or the sound-producing object, e.g., "jarring", "harsh", "sharp", "warbly", "shrill", "heavy", "hollow".

4.2 6-level ratings

The 6-level rating provides an insight on which sound was strongly considered by experts and participants as belonging to a descriptor, and which did not. For example (see Fig.2), the sound of electric sparks and the sound of a digital blip are labeled particularly confidently as Artificial and Smooth (blip sound) and Rough and Dry (electric sparks sound). Other descriptors are less clearly chosen, e.g. Bright for the blip sound. On the other hand, the train sound is overall considered Dynamic, Rich and Non Nasal, but not as strongly as in the above mentioned cases .

5. LIMITATIONS

The study has a number of limitations. For example, while participants were instructed to use headphones throughout the test and maintain a consistent sound level, differences in headphone systems and listening environments are likely to have affected the perception of the stimuli. While this is a limitation, it also suggests that the descriptors that gathered consensus are quite robust in regard to listening conditions. Moreover, the stimuli were compressed to facilitate loading in the online testing environment. To combat this, we used a relatively high-quality MP3 format and the effect of the compression was judged to be unperceivable, especially comparing to the potential effects of the uncontrolled testing environment. Finally, everyday sounds are normally experienced in interaction (while we produce them or someone else produces them),

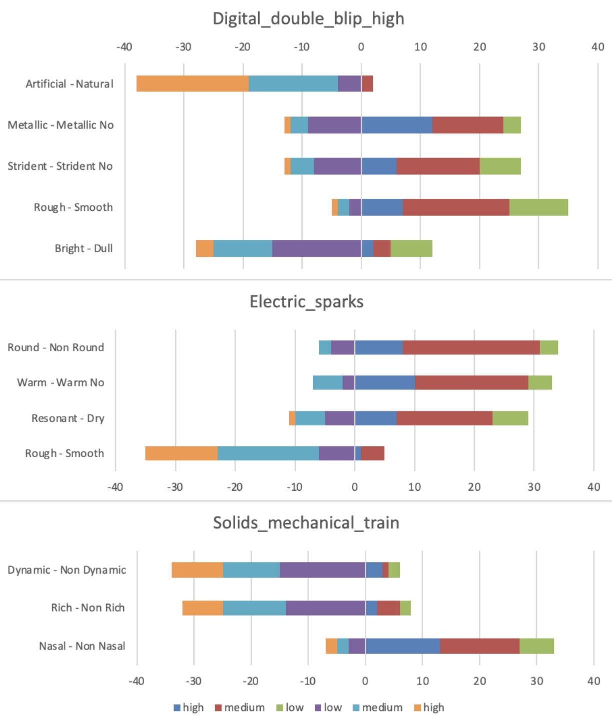

Figure 2. Example of 6-level ratings for 3 sounds - only agreed descriptors are shown

and not through headphones. This way of listening might have affected the categorization, something to take into consideration when attempting to generalize results.

6. DISCUSSION

In this study, we found agreement between participants and experts for the following descriptors: Smooth/Rough, Dull, Artificial, Metallic, Warm, Rich, Round/Non Round, Dynamic, as well as Non Nasal, Warm NO and Metallic NO. Comparing to the pre-study, we note that the pair Smooth/Rough is consistently understood, a result consistent with [19]. Metallic is the most clearly understood single descriptor in both studies, and Warm and Round/Non Round show relatively strong agreement. Additionally there is consensus for Dull, though less so for Bright, which is consistent with [19]. This study further confirms that Natural, Non Dynamic and Non Rich are not well understood by non experts. Differently from the pre-study, there is some consensus for Artificial, Dynamic and Rich, while Dry/Resonant, Bright, Strident and Nasal do not find high consensus.

Overall, this study confirms that some descriptors are generally well understood (e.g. Smooth/Rough, Metallic) and others much less. Additionally, it is possible that descriptors that provide clues about the material used to produce the sound (e.g. Metallic, Smooth, Rough) are used with more confidence when listeners feel that they can guess the cause of the original sound. Varied results were expected, since the stimuli are not distributed equally across descriptors. This means that if there is only one sound that could potentially fall under a descriptor, the likelihood of strong consensus is limited. Additionally, the descriptors defined by exclusion (e.g. Non Round, Non Nasal, Warm NO, Metallic NO) can achieve high agreement, as all the sounds that are not grouped under a specific descriptor can be categorized under a non-descriptor. These are, however, overall less useful sound descriptors. For example, if the aim is to use these descriptors as potential parameters in a synthesizer, the non-descriptors may simply represent the 0 level of a unipolar parameter (i.e. the unchanged original sound) rather than the specific characteristics of the sound. We also note that certain descriptors are often selected together, indicating that they may be perceived as having similar meanings. For example, the 5 sounds described by the participants as Round were also described as Smooth; most of the sounds (11 out of 13) labeled as Strident were also described as Rough; and most of the sounds (7 out of 8) identified as Rich were also described as Dynamic by the participants. This suggests potential semantic confusions. For example, it seems that Rich was interpreted as having "varying dynamics" rather than Rich in spectral content. Additionally, descriptors that have different meanings outside the audio context (e.g. Dry, Bright, Dull) might be particularly difficult to use by non-experts in a sound context.

Overall, we found that some of these descriptors are intuitively understood by non-experts and they could be used to develop intuitive sound interfaces, for instance. Others may need to be accompanied by an explanation and/or sound examples. Additionally, in some cases, alternative descriptors may need to be explored to better align with how non-experts perceive and describe sound. Finally, this study suggests that a sound categorization done by a limited number of experts familiar with the meaning of the descriptors would be effective, and align with non-expert judgments at least for the intuitively understood descriptors.

7. CONCLUSION

This study confirms that some of the W4S descriptors are intuitively understood by non-expert listeners, while others would likely require an accompanying explanation and examples to be useful for non-experts. This work contributes to establishing a widely understood sound lexicon: a useful tool for communication about sound and for the development of easy-to-use audio interfaces and personalization tools.

8. ACKNOWLEDGMENTS

Part of the *Personalizing Sonic Interactions* project (PI Pauletto, Ref. 2022-03421, Swedish Research Council).

9. REFERENCES

[1] M. Carron, T. Rotureau, F. Dubois, N. Misdariis, and P. Susini, "Speaking about sounds: a tool for communication on sound features," *J. of Design Research*, vol. 15, no. 2, 2017.

[2] W. Keiper, "Sound quality evaluation in the product cycle," *Acta Acustica united with Acustica*, vol. 83, no. 5, 1997.

[3] E. Özcan and R. Van Egmond, "Product sound design: An inter-disciplinary approach?," in *DRS International Conference*, 2009.

[4] M. Carron, F. Dubois, N. Misdariis, C. Talotte, and P. Susini, "Designing Sound Identity: Providing new communication tools for building brands" corporate sound"," in *Audio Mostly Conference*, pp. 1–8, 2014.

[5] D. Larsson Holmgren, N. Misdariis, and S. Pauletto, "Investigating the effectiveness of sound descriptors for non-expert listeners," in *Audio Mostly Conference*, 2024.

[6] W. W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological psychology*, vol. 5, no. 1, 1993.

[7] J. K. A. Tan, Y. Hasegawa, and S.-K. Lau, "A comprehensive environmental sound categorization scheme of an urban city," *Applied Acoustics*, vol. 199, p. 109018, Oct. 2022.

[8] A. J. Torija, D. P. Ruiz, and A. F. Ramos-Ridao, "Application of a methodology for categorizing and differentiating urban soundscapes using acoustical descriptors and semantic-differential attributes," *The Journal of the Acoustical Society of America*, vol. 134, pp. 791–802, July 2013.

[9] T. Holm Pedersen, *The Semantic Space of Sounds*. Jan. 2008.

[10] S. Bech and N. Zacharov, *Perceptual audio evaluation-Theory, method and application*. John Wiley & Sons, 2007.

[11] T. H. Pedersen and N. Zacharov, "The development of a sound wheel for reproduced sound," Audio Engineering Society, 2015.

[12] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, and S. Weinzierl, "A spatial audio quality inventory (SAQI)," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 984–994, 2014. Publisher: European Acoustics Association.

[13] N. Zacharov, T. Pedersen, and C. Pike, "A common lexicon for spatial sound quality assessment - latest developments," in *International Conference on Quality of Multimedia Experience (QoMEX)*, 2016.

[14] Ö. Axelsson, M. E. Nilsson, and B. Berglund, "A principal components model of soundscape perception," *The Journal of the Acoustical Society of America*, vol. 128, no. 5, pp. 2836–2846, 2010.

[15] "ISO/TS 12913-2:2018 acoustics — soundscape — part 2: Data collection and reporting requirements."

[16] F. Aletta, A. Mitchell, T. Oberman, J. Kang, S. Khelil, T. A. K. Bouzir, D. Berkouk, H. Xie, Y. Zhang, and R. Zhang, "Soundscape descriptors in eighteen languages: Translation and validation through listening experiments," *Applied acoustics*, vol. 224, 2024.

[17] P. Susini, O. Houix, N. Misdariis, and R. Barthélémy, "The speak methodology for sound design," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 270, pp. 9214–9223, Institute of Noise Control Engineering, 2024.

[18] P. Schaeffer, *Treatise on musical objects: An essay across disciplines*, vol. 20. Univ of California Press, 2017.

[19] V. Rosi, *The metaphors of sound: From semantics to acoustics. a study of brightness, warmth, roundness, and roughness*. PhD thesis, Sorbonne Université, 2022.

A. APPENDIX

Table 2. Final agreed labelling

Sound ID	**Labels**
gasses_gusts_fan	non nasal, metallic NO, natural
liquids_solids_electric_high	rough, warm NO
solids_electric_hum_impacts	dull, non nasal, strident NO
solids_rubbing	rough, dry, warm NO, metallic NO, non round
liquids_gasses_bubbles_oil	bright, non nasal, metallic NO, non round, natural
solids_gasses_deformation_wind	rough, non nasal, warm NO, metallic NO, non round
liquids_solids_splash_1	smooth, resonant, rich, metallic
liquids_gasses_solids_spray	bright, rough, warm NO, metallic NO, non round
solids_impacts_ratteling	bright, rich, natural, dynamic
solids_deformation_scraping_grating	rough, non nasal, warm NO, metallic NO
solids_deformation_crushing	rough, dry, warm NO, metallic NO, non round
electric_oscilating_tone_siren	smooth, non nasal, warm NO, metallic NO, dynamic
electric_sparks	rough, dry, warm NO, non round
digital_double_blip_high	bright, smooth, non rich, strident NO, metallic NO, artificial
liquids_ripple	bright, smooth non nasal, rich, strident NO, metallic NO
gasses_solids_whistle	warm NO, metallic NO, non round
digital_regular_complex_blips	warm NO, artificial
digital_double_blip_low	non rich, artificial
solids_plastic_switch	non nasal, non rich, warm NO, non round
solids_car_door	dull, non nasal, non round
gasses_explosion_big	non nasal, warm NO
solids_gasses_whooshing	smooth, metallic NO
gasses_wind_2	dull, strident NO
liquids_drip_2	bright, smooth, resonant, non nasal, metallic NO, dynamic
solids_scraping_knife	bright, non nasal, warm NO, non round
solids_scraping_sanding	rough, non nasal, warm NO, metallic NO, non round
solids_writing	non nasal, warm NO, metallic NO, non round
solids_rolling_2	rough, non nasal, warm NO, metallic
solids_clanking_metal	non nasal, warm NO, metallic, non round
solids_mechanical_train	non nasal, rich, dynamic
——Control —-	——
bright	non nasal, rough, warm no, metallic NO, non round
warm	dull, smooth, strident NO, metallic NO
dynamic	rough, non nasal, warm NO, metallic NO, **dynamic**
smooth	**smooth**
metallic	rough, non nasal, warm NO, **metallic**, non round
resonant	bright, smooth, **resonant**, non nasal
rich	warm NO, metallic NO
rough	dull, **rough**, metallic NO
round	smooth, non rich, warm, strident NO, **round**, non dynamic
strident	bright, non nasal, warm NO, non round