



HAL
open science

Investigating the Shared Meaning of Metaphorical Sound Attributes: Bright, Warm , Round , and Rough

Victor Rosi, Olivier Houix, Nicolas Misdariis, Patrick Susini

► **To cite this version:**

Victor Rosi, Olivier Houix, Nicolas Misdariis, Patrick Susini. Investigating the Shared Meaning of Metaphorical Sound Attributes: Bright, Warm , Round , and Rough. *Music Perception*, 2022, 39 (5), pp.468-483. 10.1525/mp.2022.39.5.468 . hal-03687247

HAL Id: hal-03687247

<https://hal.science/hal-03687247v1>

Submitted on 3 Jun 2022

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.

INVESTIGATING THE SHARED MEANING OF METAPHORICAL SOUND ATTRIBUTES: BRIGHT, WARM, ROUND, AND ROUGH

VICTOR ROSI, OLIVIER HOUIX, NICOLAS MISDARIIS,
& PATRICK SUSINI
STMS (IRCAM-CNRS-SU), Paris, France

MUSIC OR SOUND PROFESSIONALS USE SPECIFIC TERMINOLOGY to communicate about timbre. Some key terms do not come from the sound domain and do not have a clear definition due to their metaphorical nature. This work aims to reveal shared meanings of four well-used timbre attributes: bright, warm, round, and rough. We conducted two complementary studies with French sound and music experts (e.g., composers, sound engineers, sound designers, musicians, etc.). First, we led interviews to gather definitions and instrumental sound examples for the four attributes ($N = 32$). Second, using an online survey, we tested the relevance and consensus on multiple descriptions most frequently evoked during the interviews ($N = 51$). The analysis of the rich corpus of verbalizations from the interviews yielded the main description strategies used by the experts, namely acoustic, metaphorical, and source-related. We also derived definitions for the attributes based on significantly relevant and consensual descriptions according to the survey results. Importantly, the definitions rely heavily on metaphorical descriptions. In sum, this study presents an overview of the shared meaning and perception of four metaphorical timbre attributes in the French language.

Received: July 15, 2021, accepted March 25, 2022.

Key words: timbre semantics, sound perception, sound expertise, free verbalization, expert communication

AS HUMAN BEINGS, WE ARE USED TO verbally characterizing what we perceive through our five senses, and hearing is no exception. One can say that an alarm is too loud, or that a baby's crying is unpleasant. Sound professionals such as composers or sound engineers use a rich and often metaphorical vocabulary to describe sounds and communicate in working contexts. By "metaphorical vocabulary" we designate "concepts from one domain,

the source domain, that are borrowed for the description of things in another domain, the target domain" (Löbner, 2013, p. 63). For example, frequently used attributes for describing sounds, such as "bright" or "soft," come from vision and touch respectively. To this day, we do not know to what extent the meaning and usage of such attributes are shared among sound experts.

The dimension targeted by this metaphorical vocabulary is called timbre. Grey (1977) investigated the perceptual aspects of timbre through multidimensional scaling applied to dissimilarity ratings on musical instrument sound samples. With a similar method, McAdams et al. (1995) identified three descriptors related to the perceptual dimensions of musical timbre: the spectral centroid, the logarithmic attack time, and the degree of spectral variation.

In a study comparing the vocabulary employed by musicians and nonmusicians when describing the sound samples used in the previously cited studies, Faure (2000) gathered free verbalizations to explain the semantic description underlying the perceptual dimensions of timbre. The results revealed that the perceptual dimensions of timbre were sometimes described through metaphorical adjectives like "rich," "round," "warm," or "bright." Several studies have focused on the relations between timbre-related adjectives. von Bismarck (1974) used a method based on semantic differentials (Osgood, 1964) like "bright-dark" to rate 35 sounds. The results revealed four semantic dimensions of timbre: "full-empty," "dull-sharp," "colorful-colorless," and "compact-diffused." Later, Kendall and Carterette (1993) used unipolar scales, or verbal attribute magnitude estimation (VAME), e.g., "bright-non bright." Either using VAME or bipolar semantic scales to evaluate sounds, multiple subsequent studies have investigated the semantic dimensions of timbre (Alluri & Toiviainen, 2010; Nykänen et al., 2009; Pratt & Doak, 1976; Stepánek, 2006; Traube, 2004; Zacharakis et al., 2014).

Another line of research on timbre semantics investigated the strategies of sound description to uncover the habits of use of timbre attributes. Porcello (2004) highlighted different kinds of description strategies

wielded by sound engineers for verbal sound description (e.g., “pure” metaphor, evaluation, association, etc.). Wallmark (2019a) determined categories employed in different orchestration treatises to define the timbre of musical instruments (e.g., acoustic, mimesis, matter, crossmodal correspondences, etc.). In both studies, the authors identified description strategies grouping metaphorical attributes of sound that do not always give explicit information about the source or acoustic features. With a similar purpose, some studies employing psycholinguistic methods proposed semantic categorizations of the discourse of musicians evaluating their instrument of expertise (Cheminée et al., 2005; Lavoie, 2014; Paté et al., 2015; Saitis et al., 2017). By getting as close as possible to the usual verbalization context of the participants, these methods aim to harness ecological conditions of sound descriptions. For example, Lavoie (2014) formulated definitions of timbral attributes as understood by guitar players. More recently, a study has provided a model for musicians’ shared mental representation of Western musical instruments, based on consensual and relevant verbal descriptions of sounds as imagined (Reymore & Huron, 2020).

Although the studies on timbre semantics mentioned above have explored many facets of timbre-related vocabulary in the Western culture, few studies have investigated the meanings of metaphorical sound attributes commonly used in professional communication. Based on the literature on timbre, Carron et al. (2017) built a lexicon of 35 words used by French speaking sound experts from different fields (e.g., composers, sound designers, sound engineers, etc.) collected through interviews and surveys. The lexicon¹ aims at enhancing and supporting communication in sound design collaborative sessions with definitions and sound samples. It is structured in three classes of general aspects (e.g., high-low, short-long, etc.), temporal morphology (e.g., crescendo-decrescendo, continuous-discontinuous, etc.), and timbre attributes (e.g., bright, nasal, warm, etc.). The last category includes metaphorical descriptions of timbre whose meanings will be studied throughout this paper.

The present study addresses the issue of revealing the meaning of metaphorical attributes of sound in the French language present in Carron’s lexicon. We focused on four attributes from the lexicon—namely bright, warm, round, and rough—to evaluate our methodology. The choice of these four attributes is explained below.

Brightness (*brillance*) is certainly one of the most studied perceptual dimensions in the literature since Helmholtz. It is often correlated to the spectral centroid (Alluri & Toiviainen, 2010; Disley et al., 2006; Schubert & Wolfe, 2006; Wallmark, 2019b), and represents one of the main dimensions of timbre in multiple studies (Alluri & Toiviainen, 2010; Kendall & Carterette, 1993; Pratt & Doak, 1976; Zacharakis et al., 2014). Therefore, it makes it an excellent reference for our study. Furthermore, some works suggest that brightness might not only be based on high spectral energy but also on other features like the attack time, and that it might also interact with the concept of timbral sharpness (Ilkowska & Miśkiewicz, 2006; Saitis & Siedenburg, 2020).

Warm (*chaud*) and round (*rond*) are two sound attributes that seem to share many similarities in the description of sound as observed in some works (Bernays & Traube, 2014; Carron, 2016; Zacharakis et al., 2014). Studies conducted by three of the authors of this article have observed difficulties in distinguishing the two terms in professional conversations between sound designers and industrial partners (Carron et al., 2015; Misdariis et al., 2019). Although not thoroughly documented, this issue is persistent in sound design workshops based on verbal descriptions of sound characteristics. Since the use of the two terms is also recurrent in other sound domains, we want to identify the similarities and possible differences between these two attributes.

Roughness (*rugosité*) is also an attribute that has largely emerged from the aforementioned research dealing with sound design applications. It is defined psychoacoustically as the proximity of frequencies in critical bands (Pressnitzer & McAdams, 1999; Terhardt, 1974; Helmholtz, 1877/1954) producing the sensation of sound modulation. Even with an explicit scientific meaning, it is not clear whether experts like musicians refer to this definition to identify a rough sound.

Despite numerous insights on the general meaning of the four attributes, it remains unclear whether the usage and verbal descriptions of such attributes are consensual or generalizable among sound experts with different profiles. Therefore, this work is based on interviews and an online survey with sound experts from different fields. During the interviews, we asked participants to verbally define each attribute and to extend their definitions by selecting exemplary sound samples from a pre-defined sound library (Study 1). Then, the resulting descriptions were submitted via an online survey to an audience of experts to explicitly assess their consensus and relevance in relation to the definition of the four attributes (Study 2). Thus, Study 2 consists in

¹ <https://speak.ircam.fr/en/lexique/lexique-ircam/>

determining to what extent the descriptions obtained in Study 1 for each attribute are relevant and shared among participants. In sum, by combining the two studies, this work, provides a methodology for assessing the shared meaning of widely used timbre attributes, and a consensual definition for each of them.

Study 1: Interviews With Sound Experts

Interviews were designed to address two goals: first, to obtain rich definitions with corresponding sound samples for the four attributes, and second, to reveal the sound description strategies used by experts to define the selected attributes. During the interviews, we also asked participants to illustrate their definitions with sound samples taken from a database of musical instruments.

METHOD

Participants

Thirty-two French-fluent sound and music experts participated in the interviews (male: 23, females: 9, median age: 38.5, age range: 27–69). We selected a panel of experts that work in diverse audio fields to best represent the richness of description of the four attributes. The panel was mainly constituted of composers (10), sound engineers (7), classical musicians (7) and sound designers (6). See Supplementary Material² for the full presentation of the professional profiles.

Sound Corpus & Apparatus

To provide experts with sounds samples that could illustrate their definitions, the choice of a sound corpus was crucial. It had to be large, diverse, and easy to access. Therefore, we chose a corpus of musical instruments, showing multiple kinds of Western instruments, and playing techniques. The corpus of sounds was the result of the merger of the Studio-Online Library (SOL) (Ballet et al., 1999) and the Vienna Symphonic Library (VSL³). In addition to the usual instruments of strings, woodwinds, and brass, we added tonal keyboards (glockenspiel, vibraphone, xylophone, and marimba), an accordion, and a piano. For each instrument, we had a set of playing techniques, ranging from standard techniques (e.g., *pizzicato*, *flutterzunge*), to more contemporary ones (e.g., multiphonics, *Bartók pizzicato*). The instruments displayed variations in dynamics (from *piano* to *forte*) and pitch (octaves of C). Similarly to McAdams et al. (2017), and to avoid any potential bias

created by intervals, we only presented octaves of C (except for multiphonics). Besides, some studies have observed an influence of pitch on the appreciation of timbre (Allen & Oxenham, 2014; Alluri & Toiviainen, 2010; Marozeau et al., 2003; McAdams et al., 2017; Siedenburg et al., 2021). For comfort reasons and to exclude the loudness as a main factor, we normalized the loudness of each sound sample (-23 LUFS) following the EBU norm on loudness (R-128). The loudness normalization was not noticed by the participants, except for one who felt the normalization denaturalized the sounds.

Interviews were led by the first author and lasted about two hours. They took place either in the IRCAM studios or at the participant's home or workplace. The setup was composed of a Max/MSP interface providing easy access to the sound corpus. Participants listened to sounds via open headphones Sennheiser HD 650. Each interview was recorded with a SHURE MV5 microphone.

Interview Procedure

During the interview, the four attributes were studied sequentially. The order of presentation followed a Klein four-group permutation (Klein, 1884) to avoid any order effect bias.

Generally speaking, the design of an interview depends on the information to be extracted. Therefore, we designed a semi-directed interview in which some questions expect a certain type of response (e.g., selecting sound samples), while others leave more room for free verbalization (e.g., giving definitions), which is recommended for semantic study, as it has been done in formerly cited studies (Cheminée et al., 2005; Lavoie, 2014; Porcello, 2004; Reymore & Huron, 2020; Saitis et al., 2017).

The setup of an interview with experts often creates a hierarchy or asymmetrical interaction between both parties that could bring some sort of bias. This may come from the expert's assumption of lack of knowledge on the part of the interviewer, resulting in a lack of richness in the data collected. The status of the interviewer is thus defined as co-expert (Bogner et al., 2009; Van Audenhove, 2007). As a co-expert, the interviewer has similar knowledge of the technical terminology used by the expert, which allows for more depth in the conversation. To ensure clarity and relevance of answers, the interviewer must use a common vocabulary with all participants.

Before beginning the interview, both the corpus of sounds and the questions of the interview were introduced to the experts.

² <https://doi.org/10.5281/zenodo.6378886>

³ <http://www.vsl.co.at>

1. What is the context and frequency of use of the studied attribute?
2. How do you define the studied attribute?
3. Can you find at least three corresponding sound samples?
4. Can you find at least three sound samples in opposition to the studied attribute?
5. How do you define the opposite of the studied attribute?
6. Is there any affect related to the attribute under study and its opposite?

The first question was designed to obtain an overview of the context of use of each term as well as an indication of the frequency of use. In the end, it mainly tended to give a more ecological context to experts for formulating a definition.

For the second question, participants were asked to define the attribute. The interviewer helped the participants develop their responses by directing them to acoustic aspects of sound while trying to avoid definitions related to affect. The issue of affect was dealt with at the end of the questionnaire.

In the third question, participants were tasked to select three sound examples that corresponded to their definition of the attribute being studied. If necessary, the interviewer could help the expert find sounds, based on the sound descriptions given in the previous question (e.g., “low-pitched,” “strings,” “not too loud,” etc.). The request for sound samples was not too restrictive as it was sometimes simpler for participants to select a playing technique or an instrument rather than a specific sound.

In a second part of the interview, we discussed the opposite concept to the studied attribute. The objective was to refine the answers given to the second and third questions. The fourth question had the same purpose as the third question but with sound samples in opposition with the definition of the studied attribute. Then, in the fifth question, participants tried to define the type of sounds opposite of the term studied.

Finally, the sixth question was the opportunity to question the presence of affect in the meaning of the studied attribute. It was also a way to remove any characterizations strongly related to affect from the second and fifth questions because of their lack of acoustic information. The answers to this question were used as complementary information for interpretation in the rest of this paper.

ANALYSIS

After manually transcribing the interviews, we analyzed the descriptions given in the second and fifth questions.

The verbalizations were filtered according to three basic steps of Natural Language Processing (NLP): (1) Tokenization of the text data with the *nlk toolbox*.⁴ (2) Removal of stop words. (3) Lemmatization of the tokenized text, based on an adapted version of Sagot’s lexicon (Sagot, 2010). In the end, we obtained the lemma/interviewee frequency (i.e., the number of participants who cited a lemma for the definition of each attribute).

In an investigation of timbral attribute queries for sound effect libraries, Pearce et al. (2017) kept only relevant units of verbal description by following a few steps of manual filtering of their text data. We proposed a similar process that was run and reviewed by the four authors. Each ambiguous verbal unit was inspected according to its context in the sentence it is extracted from. One lemma was removed if its meaning was inconsistently identified more than 50% of the time. For instance, there was confusion about whether the term “aspect” was to be used to describe the metaphorical aspect of the sound or the fact that the sound had multiple aspects. Finally, if lemmas shared the same concept and root, they were grouped under the most frequent lemma out of the two. For instance, “bright” and “brightness” were grouped under the lemma “bright” rather than “brightness.” Moreover, we did not consider the hapaxes for analysis.

Inspired by the literature that focused on the vocabulary employed by sound professionals (Carron, 2016; Faure, 2000; Porcello, 2004; Wallmark, 2019a), we encoded the verbal data into categories of description strategies. The purpose was to better visualize the verbal data and to report the strategies most used to define each of the attributes.

As explained by Saitis et al. (2017) in a study on the evaluation of violin quality by professional violinists, there are two opposed perspectives regarding the qualitative analysis. Some believe that the researcher should analyze all data without any assumptions, while others think the researcher should enter the field with their hypotheses in mind (Strauss & Corbin, 1994). Here, we followed a hypothetico-deductive method and considered both prior knowledge from semantic timbre literature and information emerging from our corpus to create some of the categories of description strategies. Some of these categories emerged naturally from the transcriptions, such as the description of the source with musical instruments, and the playing technique, or spectral and temporal descriptions (all categories are reported in Table 1).

⁴ <https://www.nltk.org/>

RESULTS

We mainly observe descriptions that are either acoustic, source-related, or metaphoric. It is worth noting that in the French language, there can be confusion when classifying/lemmatizing descriptions of spectrum or pitch. For instance, *aigu*, and *haut*, will both describe high pitch or high frequencies. The same goes for *grave*, *bas* that designate either low frequencies or low pitch. *Basse* is more ambiguous as it can describe the bass clarinet, the bass guitar, or low frequencies. Considering source-related descriptions, the fact that experts mentioned instruments like the clarinet or the percussion is vastly influenced by the sound corpus. Finally, we counted numerous metaphorical descriptions such as pure (*pur*), full (*plein*), pleasant (*agréable*), and aggressive (*agressif*) that do not explicitly designate a physical aspect of the sound. Lists of the most occurring lemmas are reported for all attributes in the Supplementary Materials.²

Description Strategies

The categories of description strategies, organized in three classes of acoustic, metaphoric, and source-related descriptions, are summarized in Table 1 along with examples. In total, we proposed 10 description strategies distributed over the three classes, to structure the verbalizations. For both acoustic and metaphorical categories, we have relied on a synthesis of the most recurrent semantic categories in different works on

TABLE 1. Description Strategies (Left Column) Along with Samples From Most Occurring Lemmas (Right Column)

Acoustic	
Spectral	high (<i>aigu</i>), harmonic (<i>harmonique</i>), low (<i>grave</i>)
Temporal	attack (<i>attaque</i>), sustained (<i>entretenu</i>), steady (<i>stable</i>)
Dynamic	<i>forte</i> , <i>piano</i> , <i>crescendo</i>
Sound specific	nasal (<i>nasal</i>), resonant (<i>résonnant</i>), noisy (<i>bruité</i>)
Source related	
Excitation mode	rub (<i>frotter</i>), vibrato, breathing (<i>souffler</i>)
Source	string (<i>corde</i>), voice (<i>voix</i>), clarinet (<i>clarinette</i>)
Metaphoric	
CMC	warm (<i>chaud</i>), harsh (<i>dur</i>), clear (<i>clair</i>)
Matter	round (<i>rond</i>), full (<i>plein</i>), organic (<i>organique</i>)
Effect	enveloping (<i>enveloppant</i>), itchy (<i>qui gratte</i>)
Affect	pleasant (<i>agréable</i>), aggressive (<i>agressif</i>)

Note: CMC = Crossmodal correspondences.

timbre (Carron, 2016; Faure, 2000; Porcello, 2004; Wallmark, 2019a). Source-related description were also inspired by research on environmental sound identification (Houix et al., 2012).

The first class gathers all the **acoustic** descriptions of sounds. There are **temporal** and **spectral** descriptions, but also **dynamic and intensity** aspects of sound, along with all of the lexical fields that are explicitly related to **sound**.

The second class collects the references to the **source**. It also corresponds to the causal listening evoked by Carron et al. (2017). There was information on the **source** mainly represented by naming the instruments present in the sound corpus. There were also characterizations of the **excitation mode or the playing technique**.

The third class groups all of the **metaphoric** aspects of sound. The **crossmodal correspondence (CMC)** category that was extracted from previous studies contains descriptions related to other senses, such as sight, touch, and taste. A second metaphorical sub-category groups lemmas describing sound like **matter**, as specifically introduced by Wallmark (2019a). It shows descriptions of sound's shape, density, or material. The third metaphorical category groups all the descriptions of sound having an **effect** on the listener and its surroundings. The last category contains **affect, emotional value, and judgment** related to sounds. This category is present in all the studies cited above. The sixth question on the questionnaire was intended to prevent affect-related characteristics for the second question, but participants used this type of description anyway.

In order to test the validity of the description strategies, we performed an interrater agreement measure, as achieved by Wallmark (2019a). The four authors sorted the 50 top lemmas of both the second and the fifth question into the 10 categories. We noted incidental disagreement caused by the polysemy of some metaphorical items in the list, but we always considered the context of the word and the definition from *Trésor de la langue française* database⁵ to conclude on each classification. The measure of interrater agreement, Fleiss' κ , got a score of $\kappa = .069$, which reflects a substantial agreement (Landis & Koch, 1977). We then refined the categories and their definitions by collectively sorting the top 50 words one more time. Ultimately, we manually classified the lemmas cited by at least two experts in the categories.

Figure 1 presents the percentage of participants using the different description strategies. We noted that the

⁵ <http://atilf.atilf.fr/>

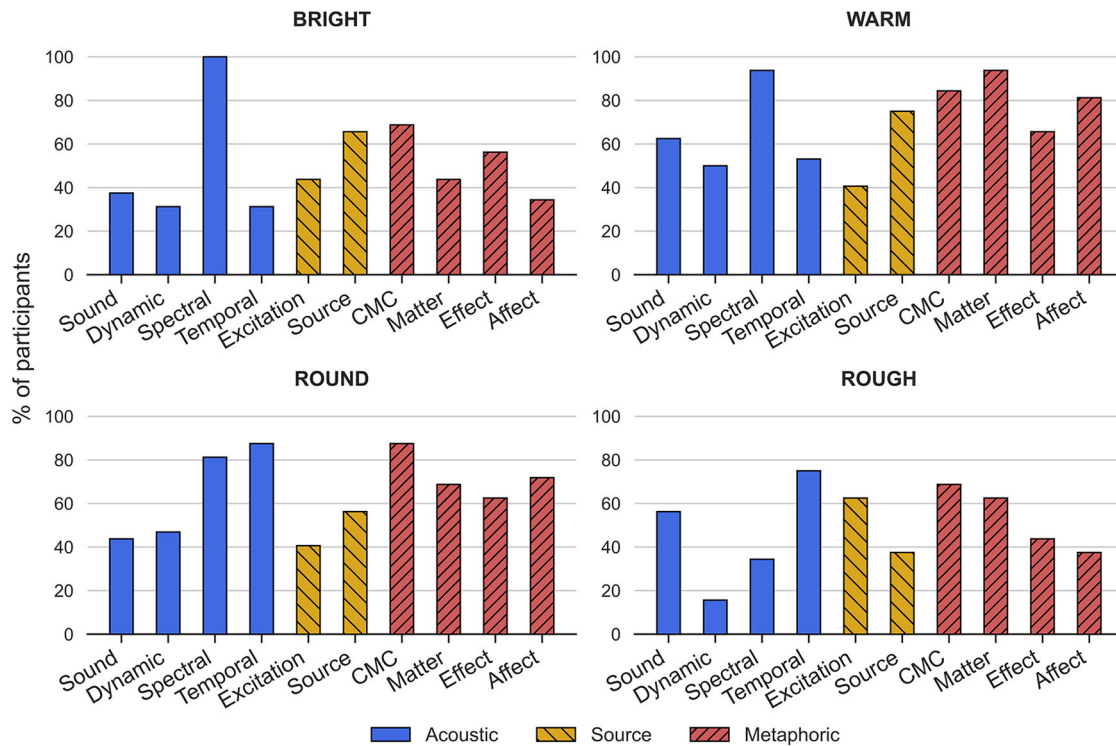


FIGURE 1. Percentage of participants using the different description strategies to define the four attributes in the second question.

acoustic aspects of bright were almost exclusively described through spectral features. To a lesser extent, the same is true for round and warm; but for round, there are also many descriptions of temporal characteristics of sounds. For both warm and round, there are many metaphorical descriptions. Finally, there are fewer descriptions for rough, which is associated more frequently with the mode of excitation than the source, unlike the other three attributes.

Verbal Descriptions and Sound Samples

Table 2 reports the descriptions most cited by the experts during the interviews when answering the second and fifth questions, along with the most frequently selected sound samples (third and fourth question). The descriptions are organized in the three classes of acoustic, source-related, and metaphoric descriptions (respectively in the first three columns). For each description, we indicated the corresponding category from Table 1. The number of participants that cited a description either in the second or the fifth question are displayed in parenthesis in the table. We only presented descriptions evoked by at least 20% of the participants for the term (+), and the opposite (-). Some descriptions were grouped, e.g., homogeneous

(*homogène*)/balanced (*équilibré*), if they were judged semantically closed according to the online dictionary of synonyms created by the Crosslanguage Research Centre on Meaning in Context (CRISCO⁶). We grouped descriptions that were expressed negatively in one question with corresponding descriptions expressed positively in the other question. For instance, the description “lots of high-frequency spectral content” for brightness was grouped with “few high-frequency components” for the opposite of brightness. The most frequently selected sound samples appearing in Table 2 were chosen according to the nature of the instrument and the playing technique. See Supplementary Material² to listen to the sound samples presented in Table 2.

DISCUSSION

The descriptions and sound samples cited by the experts allowed us to make multiple connections with results obtained in the literature on timbre semantics. Interestingly, although our study is in French, many of our results coincide with the literature on timbre semantics in English.

⁶ <http://crisco.unicaen.fr/>

TABLE 2. Descriptions Cited by (N) Experts for Each of the Four Terms Organized in the Three Classes of Description Strategies Along with the Most Frequently Selected Sound Samples

	Acoustic	Source	Metaphoric	Sound samples	
BRIGHT					
(+)	Mainly high-frequency components (25) Medium/high pitch (10) Sharp/strong/not soft attack (12)	<i>Spectral</i> <i>Temporal</i>		Lexical field of light (18) <i>CMC</i>	Glockenspiel - hard stick (29) Trumpet - <i>ordinario</i> (18)
	Mainly low-frequency components (7) Low pitch (10) Dull (<i>sourd</i>) (9)	<i>Spectral</i> <i>Sound</i>	Muffled (11) <i>Excitation</i>	Matt (11) Dull (<i>terne</i>) (7) <i>Matter</i> <i>Affect</i>	Tuba - <i>ordinario</i> (15) Marimba - <i>soft stick</i> (12)
WARM					
(+)	Mainly low/mid-low-frequency components (28) Low pitch (10) Harmonic richness (9) Soft/little/not sharp/not harsh attack (12)	<i>Spectral</i> <i>Temporal</i>	Voice (13) Strings (8) <i>Source</i> <i>Excitation</i> Vibrato (10) Breathed (9)	Round (21) Pleasant (19) <i>Matter</i> <i>Affect</i>	Bass clarinet - <i>ordinario</i> (22) Cello – <i>ordinario</i> (18) French horn – <i>ordinario</i> (11)
	Mainly high-frequency components (19)	<i>Spectral</i>		Soft (13) Enveloping (9) <i>CMC</i> <i>Effect</i>	
(-)			Cold (11) Aggressive (11) <i>CMC</i> <i>Affect</i>	Glockenspiel (16) Piccolo – <i>ordinario</i> (12) Accordion – <i>ordinario</i> (12)	
ROUND					
(+)	Soft/slow/little/not harsh/not sharp attack (23) Resonant (14) Mainly low/mid-low-frequency components (17)	<i>Temporal</i> <i>Spectral</i>		Warm (20) Soft (13) Balanced / Homogeneous (10) <i>CMC</i>	Double bass - <i>pizzicato</i> (19) Marimba - soft stick (18) Bass tuba - <i>ordinario</i> (17)
	Mainly high-frequency components (11)	<i>Spectral</i>	Brassy (8) <i>Excitation</i>	Full (10) Pleasant (8) <i>Matter</i> <i>Affect</i>	
(-)			Rough (13) Harsh (7) Aggressive (12) <i>CMC</i> <i>Affect</i>	Brass - brassy (19) Strings - Bartok <i>pizzicato</i> (13)	
ROUGH					
(+)	Noisy (14) Unstable/irregular/ with variations (16)	<i>Sound</i> <i>Temporal</i>	Rubbing/ friction (13) <i>Excitation</i>	Texture/asperities/ graininess (10) <i>CMC</i>	Winds - <i>flatterzunge</i> (19) Bassoon - multiphonic (18) Strings - <i>Sul ponticello</i> (17)
				Smooth (25) Pure (11) Round (10) <i>CMC</i> <i>Matter</i>	Accordion - <i>ordinario</i> (18) Vibraphone (18)

Note: Sub description strategies are indicated in italic next to the descriptions.
Excitation = Excitation mode. *Sound* = Sound specific descriptions.

Coherently with research on timbral brightness, the great majority of experts evaluated brightness as being linked to a strong high-frequency spectral energy. As observed in previous studies (Alluri & Toiviainen, 2010; Marozeau et al., 2003; Schubert & Wolfe, 2006), the experts evoked the influence of high pitch on brightness. Several participants mentioned that a sound with a sharp attack is perceived as brighter, as was presumed but not proven in Saitis and

Siedenburg (2020) in a study based on a pairwise comparison experiment. This may actually be due to a strong high-frequency spectral energy in the attack of the sound. Unlike the “bright-dark” semantic scales often used in the literature (Alluri & Toiviainen, 2010; Solomon, 1958; von Bismarck, 1974), the opposition to brightness here is more consistently expressed by terms like muffled, muted or dull, like in Pratt and Doak (1976).

Most selected sound samples were high-pitched instruments played on their high register and rather loudly, in accordance with the spectral description of brightness. The choice of glockenspiel sounds played with hard sticks corroborates the potential relation between brightness and a fast attack time.

Warmth and roundness seem to be comparable attributes as they share many descriptions, but with some subtle differences. Participants evoked substantial low-frequency components for the definition of warmth and roundness coherently with studies involving the two attributes (Disley & Howard, 2004; Zacharakis et al., 2014). However, the number of overtones in a round sound was a point of disagreement among the participants, while some associate roundness with spectral richness, others imagine a sound poor in overtones. Concerning temporal aspects, the descriptions of the attack also appeared in both the definitions of warmth and roundness (no attack, little attack, soft attack, not a hard attack). However, roundness was more often described by the quality of the attack than warmth. Bernays and Traube (2014) also noted a relationship between the nature of the attack and roundness in an experiment where pianists rated music recording rounder if the speed of the attack on the keys was slower.

Consistently with the acoustical descriptions from both the second and the fifth question, the round sound samples were quiet, low pitched, temporally stable, and with a soft attack. In addition, impact sounds such as the double bass playing *pizzicato* or the marimba enclosed a long resonance. Moreover, the opposition of the double bass *pizzicato* and the Bartók *pizzicato* in the selected sound samples confirms the importance of the attack for the roundness. As suggested by the source-related description of warmth, the selected cello sound displayed a strong *vibrato*. Importantly, in the evaluation of the “warm-cold” semantic scale with sounds, Eitan and Rothschild (2011) also correlated *vibrato* with the sensation of a warm sound. Finally, the breathy sound mentioned in the source-related category echoes the selected sound of bass clarinet which, when played *piano*, lets us hear the air coming out of the mouthpiece.

Many of the descriptions for warmth and roundness were metaphorical. Several participants evaluated that a warm sound was also a soft sound, similarly to Eitan and Rothschild (2011) that measured a positive correlation between the semantic scales “warm-cold” and “soft-hard.” In addition, we observed similar results to those of Zacharakis et al. (2014), as some participants contrasted round and rough, and others noted similarities between warm, round and soft.

Despite the design of our questionnaire, warm and round were often described through affect concepts in the second question. Hence, the experts characterized warm sounds, and to a lesser extent, round sounds, as pleasant and not aggressive. This result echoes findings in research treating valence and timbre. Valence has been depicted as being dependent on characteristics similar to observed descriptions of warm and round in this work, namely relatively long sounds with little energy and long transients (Eerola et al., 2012), and energy in the low spectrum (Wallmark et al., 2019c). However, contrary to our results, McAdams et al. (2017) observed a correlation of the perceived valence on musical instrument sounds with a strong high-frequency spectral content. This opposition highlights the possible variability in affect judgments on sounds, as it has been observed in a preference-based sound quality assessment (Susini et al., 2004).

From the verbal results, roughness is related to noise, temporal patterns, or instability. Furthermore, the metaphorical descriptions represented by the lexical field of touch were a large part of the data. While it is unclear to what extent the auditory definition of roughness relates to the sul ponticello sound, multiphonics match the dissonance evoked by Helmholtz, and that both *flutterzunge* and multiphonics produce the typical envelope fluctuation of psychoacoustical roughness (Pressnitzer & McAdams, 1999).

In sum, the interviews offered a great diversity of verbalizations with representative sound samples for the four attributes. We observed quite different description strategies for the four attributes that allowed to establish consistent links with the literature on timbre semantics. While bright, warm and round seem to be spectrally and temporally related, this is not the case for rough whose spectral definition is almost nonexistent. Warm and round retain many similarities both in their description strategies and in their acoustic definitions. Finally, the nature of the selected sounds highlighted certain aspects of sound over others (e.g., loud instruments denote high-frequency spectrum for bright, *flutterzunge* denotes temporal variation for rough).

Despite valuable insights on the descriptions strategies and meaning of the four terms, the results presented in Table 2 did not quite take advantage of the diversity of verbalizations as we had to summarize or group some concepts. Moreover, numerous metaphorical descriptions were difficult to interpret, and some participants had sometimes opposite points of view on them (e.g., richness for round). Therefore, in a similar fashion as some timbre semantics studies (Faure, 2000; Reymore & Huron, 2020), we sought to estimate which

characteristics are the most important and relevant. To do so, in a second part that consists of an online survey, we focused on the level of consensus and relevance of descriptions given by the interview participants for the four attributes.

Study 2: Online Survey

The goal of this survey is to find a way to select and rank the most relevant information contained in the verbalizations obtained during the interviews. For each of the four attributes, we built a corpus gathering the descriptions made up of the lemmas most frequently used for the second and fifth questions (see *Interview Procedure*). To investigate this question, we asked sound professionals to evaluate how one item of the corpus relates to the corresponding attribute, as part of an online survey. We also wish to evaluate presumably similar descriptions (e.g., “a sound with a soft attack,” “a sound with a slow attack,” “a sound without an attack”) in order to derive the most relevant and consensual version.

METHOD

Participants & Apparatus

Fifty-two sound experts participated in the survey. Similar to the interviews, all participants had one or multiple professional activities related to sound or music. Among the population of participants, 17 also participated in the interviews. They were mainly sound engineers (20), classical musicians (12), and sound designers (8). See Supplementary Material² for the full presentation of the professional profiles.

We designed the survey with an online Javascript tool for psychology experiments called *Lab.js* (Henninger et al., 2019). The survey was deployed on JATOS (Lange et al., 2015) and available on all kinds of web browsers.

The Phrases Corpus

As we wanted to study the verbalizations obtained in the interviews, we extracted the descriptions from the responses to the second and fifth questions (cf. Study 1)—when participants were asked to define a sound attribute and its opposite. We selected phrases that included the most occurring lemmas for each of the two questions (i.e., a lemma was selected if it was evoked by at least three persons). We discarded all the descriptions including names of instruments as participants of the survey could not listen to any sounds. We filtered the original corpus of descriptions to make the online survey feasible in a reasonable amount of time following three rules:

- Discard metaphoric concepts evoked by only one person.
- Homogenize the description of the spectrum with quantifiers, e.g., the sentences “a sound loaded with high harmonics” and “a sound with a lot of high-frequency components” become “a sound with a lot of high harmonics/components.”
- Eliminate a description from one of the questions that is opposed to another one from the other question, e.g., “a sound that is expressive” (second question) and “a sound that is not expressive” (fifth question).

By the end of the procedure, there were 45 phrases for bright, 67 for warm, 68 for round, and 34 for rough. Note that the corpus of descriptions is different for each attribute as it is based on the verbalizations proper to each attribute. However, some descriptions are common to more than one attribute (e.g., “rich,” “smooth,” “low,” etc.).

Questionnaire

The questionnaire invited participants to evaluate the adequacy and relevance of sound descriptions with the four attributes presented in a randomized order. When starting the experiment, they had to explain in which context they would use each of the four attributes. The idea behind this question is to get closer to the real context of use to enhance the reflection of the participant.

The form was composed of two questions:

1. According to you, the description “X” is:
Accurate/Vague/Incomprehensible
2. According to you, a [sound attribute] sound is X?
2A) Not relevant (Yes/No)
2B) Strongly disagree-Strongly agree (Likert scale)

An example of both question for bright is:

1. According to you, the description “a sound with a soft attack” is: . . .
2. According to you, a bright sound is a sound with a soft attack?

The descriptions were not originally formulated by the participants of the online survey, so it could be difficult for them to relate it to a specific attribute. To address that issue, we first asked participants to express their degree of comprehension of the description (question 1). Then, participants proceeded to the second question only if they answered “accurate” or “vague” to the first question. The question 2 is two-fold. First, participants indicated if the description was not relevant to the attribute (question 2A). Second, if they felt the description was relevant, they indicated how well it

matched or did not match the attributes on a 5-point Likert scale ranging from *strongly disagree* to *strongly agree* (question 2B). The additional information on relevance was motivated by Faure et al. (1996), that evaluated the relevance of a group of descriptions with a collection of sounds. An example of the interface in French is reported in the Supplementary Materials.²

ANALYSIS

In order to select descriptions that were familiar, relevant, and with a clear trend with respect to each attribute, we applied statistical tests to the answers of the three questions (i.e., 1, 2A, and 2B) sequentially. First, for question 1, to test whether a description's meaning is significantly familiar (i.e., "accurate"/"vague") or not (i.e., "incomprehensible") we used a chi-square test of homogeneity (1, $N = 51$, $p = .05$). Second, for question 2A, we used a similar chi-square test (1, $N = 51$, $p = .05$) to select the significantly relevant descriptions. Lastly, we applied a Wilcoxon signed-rank test to the Likert scale results of question 2B, to evaluate the central tendencies of a description whether it was matching the attribute, mismatching, or neutral. Only descriptions with significant tendencies from the Likert scale midpoint were retained. In other words, a description was not selected if it was not judged significantly familiar, and the Wilcoxon test was considered only if the description was judged significantly relevant.

Because tests were applied sequentially, the probability of type 1 errors is multiplicative and is generally low ($p < .05^3$). Thus, corrections for multiple comparisons by the number of tested descriptions (e.g., the Bonferroni correction) would not affect the results and were considered unnecessary here. See Supplementary Material² for more information on the statistical analysis of Study 2.

RESULTS

Figure 2 reports the most significant descriptions from the survey, hence giving a general meaning of the four attributes. Translations were formulated by the authors on the basis of the literature on sound semantics, and are therefore not perfectly accurate but rather an aid to understanding. See Supplementary Material² for the original versions of the descriptions in French.

Importantly, all the descriptions in agreement with the attribute under study are from the second question of the interviews and those in opposition from the fifth question. The statistical analysis revealed large consensus on the meaning of many attributes across participants. Table 3 reports the number of relevant phrases with a consensual meaning and their distribution into

the three classes of description strategies, namely, acoustic, metaphoric, and source related. In sum, we observe strong shared meanings for the four attributes with still many metaphorical descriptions.

DISCUSSION

The most important information emerging from the interviews is found in the expression of a strong consensus in the survey results, specifically on metaphorical descriptions. The shared meanings expressed through metaphorical descriptions may be due to their lexical relationship with the studied attribute instead of an acoustic description. For instance, descriptions such as "comfortable," "pleasant," or "enveloping" might simply be the depiction of a pleasant warm feeling uncorrelated with acoustic features.

Interestingly, the absence of audio material in the online survey might have changed the results of the relevance of some of the descriptions. The most glaring example of this phenomenon is the temporal description strategy, widely used in the interviews, that almost disappeared in the final results of the survey. To a lesser extent, the same observation could be made considering the source-related descriptions. These observations lead us to reflect upon the optimal conditions for collecting sound descriptions. While some studies have relied on listening support (e.g., Disley et al., 2006; Faure, 2000), others have done without it (e.g., Carron et al., 2017; Reymore & Huron, 2020). Findings in both cases reveal a consistent use of the descriptions employed. Our approach, which includes both types of verbalization context (with or without listening), allows us to gain insight into which strategies are dependent on the context of verbalization (e.g., temporal, source-related) and which are less dependent (e.g., spectral, affect).

In the end, the results for brightness are very similar to the ones obtained in the interviews with regard to its association with a high-frequency spectral content. Moreover, we observed a certain opposition of brightness with round and warm, based on their spectral descriptions. Overall, we noted a substantial consensus on the descriptions, with clear tendencies toward the meaning of brightness.

We noted many shared metaphorical descriptions for warmth and roundness. For instance, both were strongly associated with concepts like "full," "pleasant," or "soft" that were already emerging in the interviews results. While roundness is opposed to roughness, warmth has no significant trend with roughness. This absence of link between warm and rough is expressed by the fact that warm is opposed to the term "smooth" which is itself opposed to rough. Another distinction

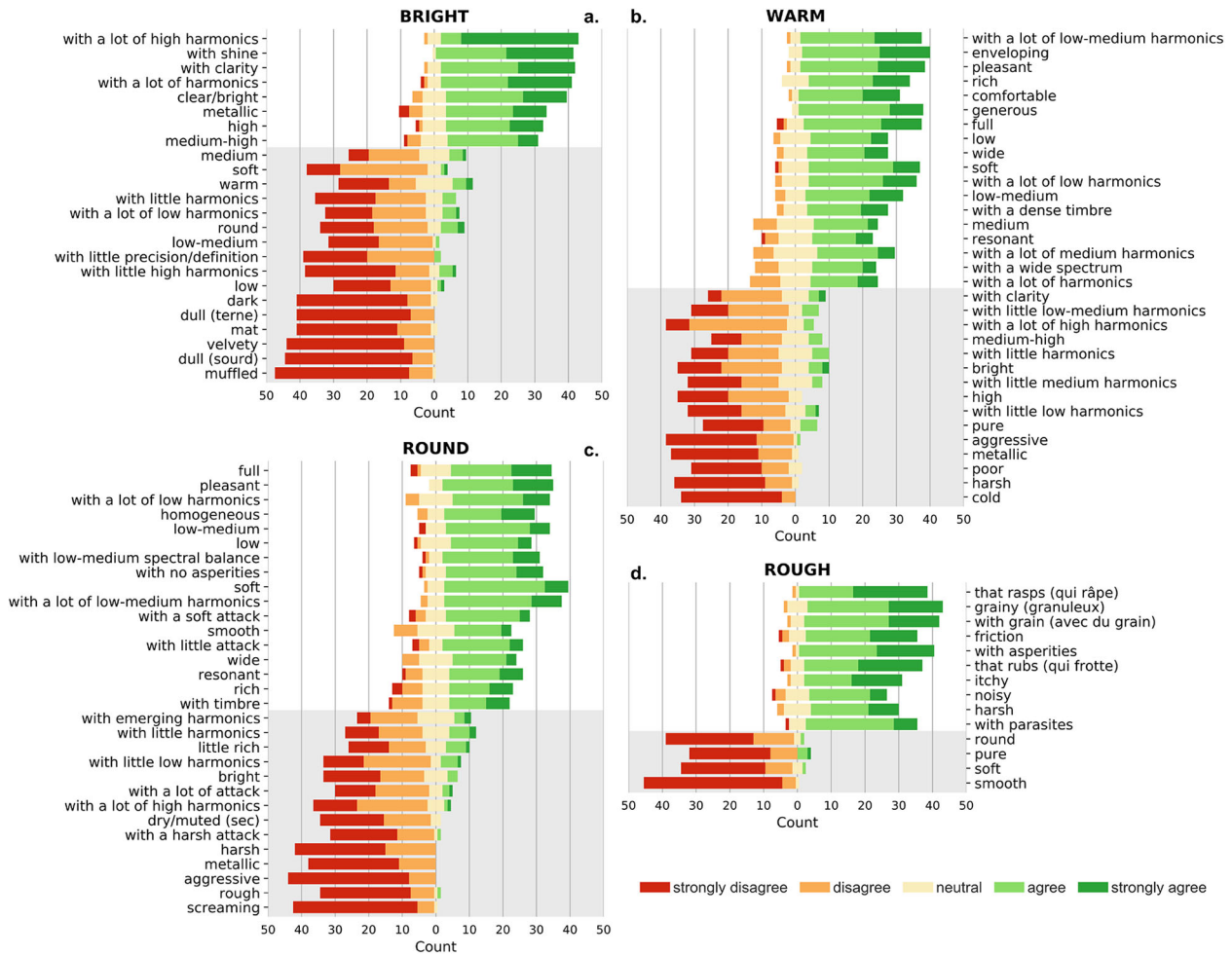


FIGURE 2. Relevant descriptions and distribution of answers on the Likert scales obtained through the online survey for (a) bright, (b) warm, (c) round, (d) rough. The grey area gathers the descriptions in mismatch with the attribute. Some ambiguous descriptions in English are followed with a French translation in parenthesis.

TABLE 3. Number of Relevant and Consensual Phrases Compared to the Total Number of Phrases, Along with the Distribution of These Phrases Into the Three Classes of Description Strategies Established in the First Study

Attribute	Total	Relevant & Consensual	Acoustic	Metaphoric	Source related
Bright	45	24	11	11	2
Warm	67	33	17	15	1
Round	68	31	15	15	1
Rough	34	14	2	9	3

that may exist between the definitions of roundness and warmth is the relevance of the description 'rich' that is more important and consensual for the description of warmth than for the description of roundness. Interestingly, the discrepancy between the relevance and tendency of "rich" and "with a lot of harmonics" in the results of warmth and roundness may suggest that

richness does not depend essentially on spectral features as it was mentioned by some participants of the interview. These results are consistent with a study on the richness of violin timbre that have also evaluated a correspondence of timbral richness with nonspectral aspects such as warmth, vibrato, or the ability of a violinist to play a wide variety of different sounds (Saitis

et al., 2019). Incidentally, the absence of a trend of “with a lot of harmonics” and roundness echoes the little consensus we noted on the relation between “spectral richness” and roundness during the interviews. While the description of the attack for roundness was very prominent in the interviews, it appeared diminished in the survey results (e.g., “a sound with a soft attack,” “a sound with little attack”). However, it remains more relevant than for warmth. Surprisingly, recurring source-related descriptions for warmth from the interviews such as “breathy” or “vibrating” did not appear in the results.

Overall, the results regarding roughness were consistent with the interview findings. The dominant descriptions were related to the source and the lexical field of touch. Acoustic descriptions were limited to the association of roughness with ‘noisy’ and the presence of ‘parasites’ in the sound.

General Discussion

With this study, we wanted to reveal the shared meaning and to clarify the definition of four well-used timbre attributes, bright, warm, round, and rough. To do this, we employed a methodology consisting of two complementary studies, interviews, and an online survey with a population of experts. The first is qualitative and allowed us to extract a rich vocabulary with various semantic characteristics, while the second statistically evaluated the consensus and relevance in a corpus of descriptions previously obtained. Consequently, we got three different outputs to understand the meaning of each attribute: free verbalizations structured in categories of description strategies (Figure 1, Table 2), sound samples (Table 2) and semantic portraits (Figure 2). We observed consistent descriptions across studies for warm, round, and bright that are in line with findings in the literature on timbre semantics. Furthermore, the overall results allowed us to highlight interactions between the four attributes, such as an opposition between bright on one side and warm and round on the other. Importantly, rough has no connection with bright but is opposed to round.

Due to the lack of richness in sound-exclusive terminology, sound or music experts borrow their vocabulary from other sensory domains or metaphors for sound description. The attributes of brightness and roundness are derived from the sense of sight, while the attributes of warmth and roughness are derived from the sense of touch. Switching from one sensory modality to another with the same term necessarily implies multiple meanings for a term that seem to overlap in our results. The

example of roundness is very illustrative to that matter. One can argue that the shape of a round object has a kind of perfection, homogeneity, and purity. These three words were found in the definitions of a round sound during the interviews. It is difficult to understand whether the person is referring to acoustic features, or visual characteristics, hence inferring a crossmodal representation of the attribute. This raises a question, “Is the perception of timbre linked to the visual perception of shapes?” Such phenomena of audiovisual crossmodal correspondences have been observed between pitch and shape (Marks, 1987), and between word morphologies and shapes (i.e., the “bouba-kiki” effect). Moreover, a study revealed that judgments of roundedness and pointedness on pseudowords recordings are based on analogous sound and visual properties: smoother and more continuous, for roundedness, and disrupted, discontinuous, and strident, for pointedness (McCormick et al., 2015). Further research is needed to establish whether such sound-to-shape mappings are based on more general cognitive correspondences.

In the end, the results account for the meaning given to timbral attributes in two different situations, thus specifying the shared meaning of each attribute within an expert population. The novelty of our approach lies in the quantification of the consensus on the descriptions of the four attributes obtained in a conversational context. The results show that the conditions of evaluation influence the meaning of a timbral attribute. In particular, the context of the interviews favored temporal descriptions that were not retained in the online survey.

Nonetheless, despite these different evaluation conditions, the outcomes of the interviews and the online survey show many similarities. Among the three main types of description strategies, acoustic and metaphorical descriptions seemed to be the most suitable for defining the attributes. Interestingly, the largest consensus involves descriptions semantically untied to sound (e.g., “full,” “rich,” “pleasant”). Thus, according to the survey results, a big part of the consensual descriptions for roughness are metaphorical, and the affect-related descriptions highly express a shared understanding of the terms round and warm.

Crucially, these results raise the question of our ability to formulate definitions for such perceptual attributes. In a study investigating the formalism of definitions for sensory descriptors, Giboreau et al. (2007) recommend avoiding ambiguous definition items. However, meeting this constraint would have been tedious given the polysemy and complexity of some of the most relevant and consensual descriptions according to the survey (e.g., “pleasant,” “full,” “rich,” etc.). With the summaries,

we have added terms that are either complementary or opposite to the attribute being studied. Nevertheless, inspired by the three types of results obtained, free verbalizations, sound examples (Table 2), and the results of the online survey (Figure 2), we derived definitions expressing the shared meaning of each attribute:

A *bright* sound has most of the spectral energy in the high frequencies. It is often a high-pitched sound, with clarity, definition, and similarities with a metallic sound. (*Non-bright*: Muffled, Dull, Velvety, Matte, Dark.)

A *warm* sound encloses substantial spectral energy in the low-mid frequencies. It is a rather low pitch sound. Temporally, it has a rather soft attack. A warm sound is pleasant, enveloping, and rich. (*Non-warm*: Cold, Harsh, Poor, Metallic, Aggressive.)

A *round* sound has a soft attack. It has a spectral balance localized in the low-mid range and is rather low-pitched. It is full, pleasant, and homogeneous. (*Non-round*: Screaming, Rough, Aggressive, Metallic, Harsh.)

A *rough* sound relates to a sound of friction. Listening to a rough sound feels raspy and itchy to the ear. It is a sound with grain, which has temporal disturbances and can be noisy. (*Non-rough*: Smooth, Soft, Pure, Round.)

In sum, the meaning of timbral brightness is in line with previous research. The importance of the attack was briefly addressed during the interview but then was evaluated as not relevant in the survey. It is therefore difficult to conclude on the importance of the attack time.

Round and warm remain at the end of the study quite similar. They enclose common spectral and temporal definitions and are both positively weighted with affect or axiological adjectives, i.e., adjectives that enclose emotional reactions or value judgments (Kerbrat-Orecchioni, 2009).

In contrast to warmth, participants seemed to emphasize the temporal definition for roundness. Besides, the opposition of warmth and roughness with “pure” and the opposition of roundness with roughness may associate a stable and monotonic temporal envelope to round sounds while it is not a necessary condition for warmth. Thus, the definition of roundness could be considered as a more restrictive form of warmth from a temporal point of view.

For roughness, the survey results mainly display high consensus in metaphorical descriptions. It is relatively paradoxical with regard to the scientific definition of the term which seems clear and simple, and the choice of sound examples which seem to confirm it.

Additionally, we note that brightness and roughness seem to present two different dimensions from a semantic point of view. Brightness interacts a lot with roundness and warmth, notably from a spectral point of view, but it does not interact with roughness on an acoustic level. This echoes the results obtained by Zacharakis et al. (2014), who identified luminance and texture as two semantic axes of timbre in a interlanguage study involving English and Greek sound descriptions. To take the results further, we could imagine a cross-language study in the future (e.g., French - English), questioning the similarities and differences in the definition of these attributes.

Conclusion

The starting point of our research was to question the consensus on the definition of timbre attributes well known in sound and music expert circles. Hence, this research aimed to understand the meaning of four timbre attributes and the way experts describe them. We approached the question with interviews with experts and an online survey. Our results included rich descriptions and representative sound samples, an assessment of the consensus and relevance of these descriptions to the meaning of the four attributes, and definitions for each of them.

The characterization of each attribute is divided into three main categories of acoustic, metaphorical, and source-related descriptions. In the end, between the interviews and the online survey, the representation of each of the terms is robust and relies mainly on metaphorical and acoustic descriptions. Through our results, we were able to summarize the consensus regarding the meaning of the four attributes. But we note that the ambiguity of metaphorical descriptions makes the task of formulating definitions tedious.

A limitation to our approach is the difficulty to study the influence of the professional profile of the experts who participated in both parts of the study as initially intended. In a subsequent work, we will use a subjective judgment method to evaluate a sound corpus to extract the most relevant acoustic features that correlate with the definition of the studied attributes. This study will allow us to investigate potential variations in the perception of the four attributes between participants with different professional profiles. Ultimately, we will consolidate the representation that one can make of these attributes, on perceptual, acoustical, and linguistic levels.

Author Note

This research was supported by the *Fonds K pour la musique*. The authors wish to thank Pablo Arias (STMS-Lund University) for key discussions and valuable advice on style throughout the writing of this paper. They also thank the reviewers, Kai Siedenburg, Lindsey

Reymore, and the third anonymous reviewer for their relevant and comprehensive remarks on the manuscript.

Correspondence concerning this article should be addressed to Victor Rosi, STMS (IRCAM-CNRS-SU) UMR9912, 1 place Igor Stravinsky 75004 Paris, France
E-mail: victor.rosi@ircam.fr

References

- ALLEN, E. J., & OXENHAM, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *Journal of the Acoustical Society of America*, 135(3), 1371–1379.
- ALLURI, V., & TOIVIAINEN, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception*, 27(3), 223–242.
- BALLET, G., BORGHESI, R., HOFFMANN, P., & LÉVY, F. (1999). Studio online 3.0: An internet "killer application" for remote access to ircam sounds and processing tools. *Journées d'Informatique Musicale*.
- BERNAYS, M., & TRAUBE, C. (2014). Investigating pianists' individuality in the performance of five timbral nuances through patterns of articulation, touch, dynamics, and pedaling. *Frontiers in Psychology*, 5.
- BOGNER, A., LITTIG, B., & MENZ, W. (2009). Introduction: Expert interviews—An introduction to a new methodological debate. In Bogner, A., Littig, B., Menz, W. (Eds.), *Interviewing Expert. Research Methods Series* (pp. 1–13). Palgrave Macmillan, London.
- CARRON, M. (2016). *Méthodes et outils pour définir et véhiculer une identité sonore: Application au design identitaire de la marque SNCF*. [Doctoral dissertation, Paris 6]. DOI: 10.13140/RG.2.1.1018.2005
- CARRON, M., DUBOIS, F., MISDARIIS, N., & SUSINI, P. (2015). Définir une identité sonore de marque: Méthodologie et outils. *Acoustique et Techniques*, 81, 20–26.
- CARRON, M., ROTUREAU, T., DUBOIS, F., MISDARIIS, N., & SUSINI, P. (2017). Speaking about sounds: A tool for communication on sound features. *Journal of Design Research*, 15(2), 85–109.
- CHEMINÉE, P., GHERGHINOIU, C., & BESNAINOU, C. (2005). Analyses des verbalisations libres sur le son du piano versus analyses acoustiques. *Colloque Interdisciplinaire de Musicologie (CIM05)*. Montreal (Québec), Canada.
- DISLEY, A. C., & HOWARD, D. M. (2004). Spectral correlates of timbral semantics relating to the pipe organ. *Speech, Music and Hearing*, 46, 25–39.
- DISLEY, A. C., HOWARD, D. M., & HUNT, A. D. (2006). Timbral description of musical instruments. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC 09)* (pp. 61–68). Bologna, Italy: Alma Mater Studiorum
- EROLA, T., FERRER, R., & ALLURI, V. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception*, 30(1), 49–70.
- EITAN, Z., & ROTHSCHILD, I. (2011). How music touches: Musical parameters and listeners' audio-tactile metaphorical mappings. *Psychology of Music*, 39(4), 449–467.
- FAURE, A. (2000). *Des sons aux mots, comment parle-t-on du timbre musical?* [Doctoral dissertation, Ecole des Hautes Etudes en Sciences Sociales].
- FAURE, A., MCADAMS, S., & NOSULENKO, V. (1996). Verbal correlates of perceptual dimensions of timbre. In B. Pennycook & E. Costa-Giomi (Eds.), *Proceedings of the 4th International Conference on Music Perception and Cognition (ICMPC 04)*, (pp. 79–84), Montreal, Canada: McGill University
- GIBOREAU, A., DACREMONT, C., EGOROFF, C., GUERRAND, S., URDAPILLETA, I., CANDEL, D., & DUBOIS, D. (2007). Defining sensory descriptors: Towards writing guidelines based on terminology. *Food Quality and Preference*, 18(2), 265–274.
- GREY, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5), 1270–1277.
- HENNINGER, F., SHEVCHENKO, Y., MERTENS, U., KIESLICH, P. J., & HILBIG, B. E. (2019). Lab.js: A free, open, online study builder. *Behavior Research Methods*, 1–18.
- HOUIX, O., LEMAITRE, G., MISDARIIS, N., SUSINI, P., & URDAPILLETA, I. (2012). A lexical analysis of environmental sound categories. *Journal of Experimental Psychology: Applied*, 18(1), 52–80.

- ILKOWSKA, M., & MIŚKIEWICZ, A. (2006). Sharpness versus brightness: A comparison of magnitude estimates. *Acta Acustica United with Acustica*, 92(5), 812–819.
- KENDALL, R. A., & CARTERETTE, E. C. (1993). Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck's Adjectives. *Music Perception*, 10(4), 445–467.
- KERBRAT-ORECCHIONI, C. (2009). *L'énonciation: De la subjectivité dans le langage*. Armand Colin.
- KLEIN, F. (1884). *Vorlesungen über das ikosaeder und die auflösung der gleichungen vom fünften grade*. BG Teubner.
- LANDIS, J. R., & KOCH, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363–374.
- LANGE, K., KÜHN, S., & FILEVICH, E. (2015). "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of Web Servers Supporting Online Studies. *PLOS ONE*, 10(6), e0130834.
- LAVOIE, M. (2014). *Conceptualisation et communication des nuances de timbre à la guitare classique*. [Doctoral dissertation, McGill University].
- LÖBNER, S. (2013). *Understanding semantics*. Routledge.
- MARKS, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3), 384–394.
- MAROZEAU, J., DE CHEVEIGNÉ, A., McADAMS, S., & WINSBERG, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America*, 114(5), 2946–2957.
- McADAMS, S., DOUGLAS, C., & VEMPALA, N. N. (2017). Perception and modeling of affective qualities of musical instrument sounds across pitch registers. *Frontiers in Psychology*, 8.
- McADAMS, S., WINSBERG, S., DONNADIEU, S., DE SOETE, G., & KRIMPHOFF, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58(3), 177–192.
- MCCORMICK, K., KIM, J., LIST, S. M., & NYGAARD, L. C. (2015). Sound to meaning mappings in the bouba-kiki effect. *CogSci*, 2015, 1565–1570.
- MISDARIIS, N., SUSINI, P., HOUIX, O., ROQUE, R., CERLES, C., LABEL, E., ET AL. (2019). Mapping sound properties and oenological characters by a collaborative sound design approach-towards an augmented experience. In R. Kronland-Martinet, S. Ystad, M. Aramaki (Eds.) *Perception, Representations, Image, Sound, Music, CMMR2019*. Lecture Notes in Computer Science(), vol 12631. Springer, Cham.
- NYKÄNEN, A., JOHANSSON, Ö., LUNDBERG, J., & BERG, J. (2009). Modelling perceptual dimensions of saxophone sounds. *Acta Acustica United with Acustica*, 95(3), 539–549.
- OSGOOD, C. E. (1964). Semantic differential technique in the comparative study of cultures 1. *American Anthropologist*, 66(3), 171–200.
- PATÉ, A., CARROU, J.-L. L., NAVARRET, B., DUBOIS, D., & FABRE, B. (2015). Influence of the electric guitar's fingerboard wood on guitarists' perception. *Acta Acustica United with Acustica*, 101(2), 347–359.
- PEARCE, A., BROOKES, T., & MASON, R. (2017). Timbral attributes for sound effect library searching. In *Proceedings of the 2017 AES International Conference on Semantic Audio*, June 22–24, Erlangen, Germany.
- PORCELLO, T. (2004). Speaking of sound: Language and the professionalization of sound-recording engineers. *Social Studies of Science*, 34(5), 733–758.
- PRATT, R. L., & DOAK, P. E. (1976). A subjective rating scale for timbre. *Journal of Sound and Vibration*, 45(3), 317–328.
- PRESSNITZER, D., & McADAMS, S. (1999). Two phase effects in roughness perception. *Journal of the Acoustical Society of America*, 105(5), 2773–2782.
- REYMORE, L., & HURON, D. (2020). Using auditory imagery tasks to map the cognitive linguistic dimensions of musical instrument timbre qualia. *Psychomusicology: Music, Mind, and Brain*. 30(3), 124–144.
- SAGOT, B. (2010). The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, et al. (Eds.), *7th International Conference on Language Resources and Evaluation (LREC 2010)*. La Vallette, Malte.
- SAITIS, C., FRITZ, C., & SCAVONE, G. (2019). Sounds like melted chocolate: How musicians conceptualize violin sound richness. In M. Kob (Ed), *Proceedings 2019 International Symposium on Musical Acoustics*, (pp. 50–57), Detmold, Germany.
- SAITIS, C., FRITZ, C., SCAVONE, G. P., GUASTAVINO, C., & DUBOIS, D. (2017). Perceptual evaluation of violins: A psycholinguistic analysis of preference verbal descriptions by experienced musicians. *Journal of the Acoustical Society of America*, 141(4), 2746–2757.
- SAITIS, C., & SIEDENBURG, K. (2020). Brightness perception for musical instrument sounds: Relation to timbre dissimilarity and source-cause categories. *Journal of the Acoustical Society of America*, 148(4), 2256–2266.
- SCHUBERT, E., & WOLFE, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *Acta Acustica United with Acustica*, 92(5), 820–825.
- SIEDENBURG, K., JACOBSEN, S., & REUTER, C. (2021). Spectral envelope position and shape in sustained musical instrument sounds. *Journal of the Acoustical Society of America*, 149(6), 3715–3726.

- SOLOMON, L. N. (1958). Semantic approach to the perception of complex sounds. *Journal of the Acoustical Society of America*, 30(5), 421–425.
- STEPÁNEK, J. (2006). Musical sound timbre: Verbal description and dimensions. In V. Verfaillie (Ed.), *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, 121–126. Montreal, Canada: McGill University.
- STRAUSS, A., & CORBIN, J. (1998). *Basic of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Sage Publications.
- SUSINI, P., MCADAMS, S., WINSBERG, S., PERRY, I., VIEILLARD, S., & RODET, X. (2004). Characterizing the sound quality of air-conditioning noise. *Applied Acoustics*, 65(8), 763–790.
- TERHARDT, E. (1974). On the perception of periodic sound fluctuations (roughness). *Acta Acustica United with Acustica*, 30(4), 201–213.
- TRAUBE, C. (2004). *An interdisciplinary study of the timbre of the classical guitar* [Doctoral dissertation, McGill University].
- VAN AUDENHOVE, L. (2007). Expert interviews and interview techniques for policy analysis. *Vrije University, Brussel*. Retrieved May, 5, 2009. https://www.researchgate.net/publication/228795228_Expert_Interviews_and_Interview_Techniques_for_Policy_Analysis
- VON BISMARCK, G. (1974). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acta Acustica United with Acustica*, 30(3), 146–159.
- HELMHOLTZ, H. L. F. VON (1954). On the sensations of tone (A. J. Ellis, Trans.). Dover. (Original work published 1877)
- WALLMARK, Z. (2019a). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music*, 47(4), 585–605.
- WALLMARK, Z. (2019b). Semantic crosstalk in timbre perception. *Music and Science*, 2, 1–18.
- WALLMARK, Z., FRANK, R. J., & NGHIEM, L. (2019c). Creating novel tones from adjectives: An exploratory study using FM synthesis. *Psychomusicology: Music, Mind, and Brain*, 29(4), 188–199.
- ZACHARAKIS, A., PASTIADIS, K., & REISS, J. D. (2014). An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception*, 31(4), 339–358.