Uncovering the Meaning of Four Semantic Attributes of Sound : Bright, Rough, Round and Warm

Victor Rosi, Olivier Houix, Nicolas Misdariis, Patrick Susini

UNCOVERING THE MEANING OF FOUR SEMANTIC ATTRIBUTES OF SOUND : BRIGHT, ROUGH, ROUND AND WARM

Victor Rosi[1] **Olivier Houix**[1] **Nicolas Misdariis**[1] **Patrick Susini**[1]

[1] Sound perception & design group, STMS-Lab (IRCAM-CNRS-SU),
Place Igor Stravinsky, Paris

victor.rosi@ircam.fr

ABSTRACT

Bright (brillant), round (rond), warm (chaud) and rough (rugueux) are four terms vastly used in the French language for sound description in sound creation processes such as music performance, orchestration, sound engineering or sound design, yet they lack formal, standardised definitions. Therefore, the intent of the present study is to obtain definitions and matching sound samples of these four terms as they are interpreted by different sound professionals. This work was organized around individual interviews with 32 sound professionals (musicians, composers, sound designers, acousticians...), during which they were asked to give definitions of the four terms and to discuss their opposite concepts. The analysis of the interview verbatims through qualitative analysis and standard NLP (Natural Language Processing) methods allows us to unravel the various sound description strategies. Based on literature review on timbre semantics we reveal relevant categories for the clustering of the descriptions across the four terms. These categories mainly group acoustic, source-related and metaphorical descriptions produced by the sound professionals. The experts were also asked to choose sound samples from a musical instrument database to complete the definition of each word. By analyzing the descriptions and comparing them to the selected sound samples, we formulate definitions of the four terms along with a few hypotheses on the acoustic correlates involved, thus assessing the potential plurality of perception among the experts.

1. INTRODUCTION

As suggested by Lemaitre et al. [1], environmental sound identification is mostly related to listeners' expertise in the field of sound, and to their capacity to identify the sounds. In other words, experts use acoustic characteristics more than non-experts, and more so for sounds that are not easily identifiable. Apart from pitch, duration, and intensity, sounds can be described via their timbral aspects. *Warm*, *sharp*, *smooth*, *round*, *bright*, *rough*, *nasal* are all frequently used in many languages. It is often considered in western cultures that the trumpet produces arguably one of the brightest sounds among the instruments of an orchestra, and that a bright sound has a high spectral centroid. While this affirmation about brightness seems to be generally assumed by the entire community of sound professionals when describing different kind of sounds [2] [3], there are timbre-related terms that do not have an obvious semantic nor acoustic definition which would enable clear intersubjective comprehension and usage. Revealing the meaning of timbre-related terms would facilitate communication in numerous fields [4]. It is therefore crucial to assess the richness of perception of this vocabulary from a population of users.

Several studies discuss the semantic of words used by sound professionals or musicians to describe timbre in particular situations. Disley [5], gave account of seven terms often shared among a panel of musicians for the description of the pipe organ's timbre *clear, thin, warm, bright, flutey, full* and *balanced*. In a study on piano, Bernays and Traube [6] elaborate the most relevant terms for sound description that seem to rely on two dimensions: percussion and resonance. Finally, their results included five words, *bright, dry, dark, round* and *velvety*. With different methods, many other studies show similar results related to how experts speak about timbre for specific sound sources such as the classical guitar [7], the violin [8], or the piano [9].

Several papers argue that semantic description of timbre seem to be associated to specific perceptual dimensions of timbre. Von Bismarck [10], in his study on timbre semantics, identified four great structures of timbre as *dull-sharp*, *full-empty*, *colorful-colorless* and *compact-diffused*. In the 1960s, in his *Traité des objets sonores*, Schaeffer [11] proposed similar dimensions that describes only sound features called *mass*, *harmonic timbre* and *grain*. More recently, Zacharakis [12], as a result of a VAME [1] procedure determined three principal dimensions providing a good general structure for timbre, *luminance*, *texture* and *mass*.

In spite of the fact that many of the perceptual dimensions identified across the studies are similar, these dimensions don't clearly include a metaphorical description employed by experts. In an anthropological research on sound professions, Porcello [13] noted five strategies for timbre

[1] VAME : Verbal Attribute Magnitude Estimation

description used by producers and sound engineers with some being metaphorical, related to affect or imitations. Faure [14], in her thesis work assessed the semantic categories in a population of musicians and naive listeners and found nine semantic categories, some related to sound's physical aspects and others more metaphorical. In his thesis, Carron [15], parsed the literature on timbre semantics and identified three kinds of descriptions : The causal description, related to the source and the excitation mode, the contextual description corresponding to metaphorical and emotional aspects of sounds and the reduced description that deals with sound physical properties. In a semantic analysis of an orchestra treatises corpus, Wallmarck [16] identified seven semantic categories used by the authors while describing the instruments' timbre.
The semantic categorisations of the description of timbre of the presented research show similar results such as the affect-emotion-judgement category, the crossmodal correspondance (CMC) category, the imitation category, and the "reduced listening" category.

The present study aims to understand the use and the definition of four terms selected from the sound lexicon developed by Carron [4] as they are cited in numerous studies for sound description. *Brightness* (*brillance*) is defined in several surveys, and seems to be well understood semantically and acoustically. It would therefore make a great point of reference for our study. Besides, some works point out that brightness doesn't only rely on high spectral energy but also on other spectral details, and that it also has a certain correlation with *sharpness* ([17]). *Warm* (*chaud*) and *round* (*rond*) are two concepts often used in the French language for timbre description that have several similarities of use as observed by [15] and [7]. They are difficult to interpret because there is no clear opposite term to *warm* nor *round*, thus there is no apparently clear dimension for those terms. *Roughness* (*rugosité*) is defined psychoacoustically as the proximity of frequencies in critical bands of the inner ear [18] producing the sensation of a modulation of sound. In addition, the concept of *graininess* (*granulosité*) in French language is used for sound description, and seems to interfere with the definition of Roughness. This work will be based on interviews with sound professionals from different fields. *Bright*, *warm*, *round* and *rough* are the four terms studied during the interviews. The goal is to give definitions, or semantic portraits, for each term, illustrated with sound samples from a musical instrument database that enable speculation about the acoustic features involved in the perception of each word.

2. METHODS

2.1 Expert Panel

32 French-fluent sound experts participated in the interviews, 23 males and 9 females with an average age of 38 years ranging from 27 years to 69 years. Among the 32 participants, 3 are not French natives but have been working in France for more than 10 years. The choice of the panel was led by the idea of acquiring a potential diversity of perception based on the diversity of professional experiences, and by the idea that experts have a more coherent speech about sound than non-expert listeners. Hence, they are led to communicate about sound as part of their professional practice. We tried to obtain close to equal proportions of composer, sound engineer, musicians and sound designers. Many of these experts have multiple hybrid profiles, for example two of the experts are both sound engineers and sound designers.

2.2 Sound dataset & Apparatus

Choosing a sound dataset for the interview was crucial as the extracted sound samples should represent the perception of the studied terms. An environmental sound library would be a good type of sound dataset, although this kind of sound promotes *everyday listening* ([19]) which produces a sound description strongly related to the nature of the sound source. A dataset of synthetic sounds could not be an option either, since the purpose of the subsequent study is to obtain the acoustic features involved in our perception of the four terms. Therefore we chose to use a dataset composed of 26 musical instrument sounds with different playing techniques. We arbitrarily discarded playing techniques that seemed redundant with the standard playing style (called *ordinario* for this dataset), or that were not relevant with regards to timbre description. The purpose of the selection was to have a proportion of sounds easy to manage for both the interviewer and the interviewee through a shared interface. Therefore, we used the sound library recorded at IRCAM from the project Studio-Online Library (SOL) [20] mixed with parts of the Vienna Symphonic Library (VSL[2]) for some instruments that we judged to be valuable but missing in the SOL. To enlarge the dataset composed of strings, woodwinds and brass, we added tonal keyboards as percussive instruments. The rule was to always have a distinctive pitch for each sound sample (except for the multiphonic playing technique). The complete dataset of instruments is reported in Tab. 1. Each instrument has a set of playing technique options, at different pitches and different intensities, in the ranges specific to each instrument. So there were standard playing techniques, with others more contemporary, depending on the instruments, like multiphonics and *flatterzunge* for the wind instruments, or *Bartók pizzicato* for the string instruments.
We selected only octaves of C[3] in order to reduce the amount of sound samples. We motivate the diversity of pitch both with the fact that instrument timbre varies with register, and that pitch information for timbre-related terms was mentioned in Carron's interview. [4] with sound experts.
For comfort reasons, we equalized the loudness of each sound sample following the EBU norm on loudness (R-128), setting all the sounds to -23 LUFS. One of the participants felt strained by this normalization as they felt

[2] http://www.vsl.co.at
[3] from 65 Hz to 4186 Hz depending on the instrument

the sounds were subsequently de-naturalized.

Strings	Woodwinds	Brass	Percussions	Other
Violin	Concert flute	Trumpet	*Glockenspiel*	Guitar
Viola	*Piccolo flute*	Trombone	*Xylophone*	Harp
Cello	*Alto flute*	French horn	*Vibraphone*	Accordion
Double bass	Oboe	Bass tuba	*Marimba*	*Piano*
	English horn			
	Bb Clarinet			
	Bass clarinet			
	Alto Saxophone			
	Bassoon			
	Contrabassoon			

Table 1: Instruments from the SOL and VSL which compose the dataset. In italic : sound samples from the VSL

The 32 interviews lasted approximately two hours and took place in quiet environments such as in a studio at IRCAM, at the expert's home, or at the expert's workplace. The setup was composed of an interface coded in Max/MSP giving easy access for listening to the sounds as they were sorted by instrument families, instruments and playing techniques. Interviewer and interviewee listened to the sounds through open headphones Sennheiser HD 650 allowing to listen to sound samples and to discuss them without having to take on and off the headphones, making the interview more comfortable. Every interview was recorded via the Max/MSP interface by using a SHURE MV5 microphone.

2.3 Interview procedure

2.3.1 Global process

The four terms were discussed sequentially with each of the participants. The order of presentation of the terms followed a Klein Four group permutation. Hence, one interview presented the term in the order of the first line, while the next expert interview would follow a permutation. This interview was designed as a semi structured qualitative interview, while some questions expect a specific type of answers (sound samples), other questions expect free verbalisations.

2.3.2 Questionnaire

There are six main questions composing the body of the interview:

- Q1 : What is the context and frequency of use of the studied term ?
- Q2 : Can you give a definition of the studied term ?
- Q3 : Considering your definition, can you try to find at least three sound samples that correspond to the studied term ?
- Q4 : Can you try to find at least three sound samples that correspond to the opposite of the studied term ?
- Q5 : Can you define the opposite of the studied term ?
- Q6 : Is there any affect related to the perception of this term and its opposite ?

The interview would start after a presentation of the questionnaire along with quick overview of the sound dataset.
The purpose of Q1 was to give a contextual framework to the expert so that they can more easily relate to each term. It also gives valuable quantitative information about the frequency of use of the term.
In Q2, the interviewee was invited to define the term. We try to obtain information on more acoustical aspects and we avoid mentioning any affect-related description of sound, as it is treated in Q6. For Q2 and Q5, the researcher would participate by facilitating the expression of the interviewee with follow-up questions and probes.
Q3 was the opportunity for the expert to find sounds related to the terms in the database. If needed, the researcher helped the expert to spot related sounds, based on elements from the definition previously established. The interviewee had to select at least three different sounds if possible.
In the second part of the questionnaire, we investigated the opposite concept of the studied Term. The purpose is to refine the answers from Q2 and Q3 and to evaluate a potential semantic polarity between two terms. Q4 has the same process as Q3, only this time we selected sounds for the opposite term naming it yet. Q5 is focused on the definition of the opposite concept of the term.
Finally Q6 interrogates the presence of affect elements in the definition of the term and the opposite concept.
At the end of Q3 and Q4, the interviewer asked the interviewee if there were a sound missing for the sound description of the term. Q1, Q4, Q5 and Q6 will be addressed in a subsequent work, as we only focused our analysis on Q2 and Q3 for this paper.

2.4 Analysis

As evoked by Saitis [8] in their study on perception of violin timbre, there are two opposed points of view regarding qualitative analysis. Some believe that the researcher should analyse all data without any assumptions, while others think the researcher should enter the field with their hypotheses in mind [21]. We followed a hypothetico-deductive method and considered prior knowledge and assumptions of semantic timbre literature to create categories of description strategies structuring the following analysis of the interview transcripts. Some categories emerged after analysis of the verbatims. For example, one of these categories is related to the dataset of musical instrument sounds and their features. After transcribing the interviews, we extracted the definitions of the investigated terms. To this purpose, we focused on the analysis of Q2. Answers got through basic steps of natural language processing (NLP) : First we tokenized (i.e. slicing sentences into words) the transcripts of the text with the *nltk toolbox*. Secondly we discarded the stop words. Thirdly we lemmatized the tokenized text, based on an adapted version of Sagot's lexicon [22], lemmatization is the process of putting a word into its canonical form so it can be grouped with other words.

Finally, We were able to assess the lemma/interviewee frequency (i.e. the number of interviewees using one lemma for each definition).

2.4.1 Manual filtering

In order to only keep relevant units of verbal description in their investigation of timbral attribute queries for sound effect libraries, Pearce et al. [23] followed several steps of manual filtering. We propose a similar process that was run and reviewed by the four authors. Each ambiguous verbal unit was thus inspected in terms of its context. One lemma was removed if its meaning was inconsistent more than 50% of the time. For instance, there was confusion about whether the term "aspect" was to be used to describe the aspect of the sound, or the fact that the sound had multiple aspects. "Aspect" was thus removed. If a lemma possesses multiple meanings and one of them is substantially present in the corpus, we would keep it and only discard the number of occurrences of the word in irrelevant contexts. One example is the term "variation" which, in most cases, metaphorically defines the sound's temporal shape but is sometimes employed as a syntactic tool. Finally, if lemmas shared the same concept and the same root, they were grouped together under the most frequent lemma of the two. For instance, "bright" and "brightness" were grouped together under the lemma "bright". Moreover, in order to drastically reduce the corpus, we did not consider the hapaxes [4] for analysis.

2.4.2 Interpretation of the results

The information presented in the the lists of lemmas corresponds to the vocabulary used for the definition of the four terms. We were able to give semantic portraits of the four terms, through a manual excavation of the syntactic context of those words. There were metaphorical adjectives like "high", "big", "sharp" or "generous" and sometimes dimensions of sounds such has "frequency", "harmonics" or "attack". A substantial part of the definitions given by the experts were phrased negatively (e.g. "not loud", "not sharp"...). In those cases, it was crucial to check the context of each lemma. For example if the term "high" (pitch or frequency) is cited by a great amount of experts for *bright*, *round* and *warm*, it doesn't mean that they all have high fundamental frequencies or high spectral centroid. Sometimes some words were used with quantifiers like "a little", "few", "not too much" etc. We did not employ these because of their diversity and the fact that we do not have a fully automatic approach towards text mining.

The sound samples collected in Q3 had four features: Instrument, playing technique, pitch and intensity. For the results, we focused on the instrument and the playing technique. If an instrument/playing-technique combination were to be selected by more than a third of the population of experts, then it was considered as a good example for the studied term. Pitch and intensity were also considered if they were obviously related to some of the relevant verbal descriptions.

[4] verbal units that appears only in one source (one interviewee) in the raw data

3. RESULTS

Here are the 50 most frequently used terms for defining the four words (Q2) after manual filtering in Tab. 2

Term	Rank	W/I (%)	Term	Rank	W/I (%)
warm* *chaud*	1	100	blow *souffle*	15	34
bright *brillant*	2	97	rub *frotter*	16	31
round *rond*	2	97	warm* *chaleureux*	16	31
high *aigu*	3	88	aggressive *agressif*	17	28
harmonic *harmonique*	4	84	voice *voix*	17	28
rough *rugueux*	5	81	harsh *dur*	18	28
attack *attaque*	5	81	music *musique*	18	28
low** *grave*	6	69	percussion *percussion*	19	25
spectrum *spectre*	7	66	resonant *résonant*	19	25
medium *médium*	8	56	clarinet *clarinette*	19	25
timbre *timbre*	9	53	musical *musical*	19	25
frequency *fréquence*	9	53	hertz *hertz*	19	25
string *corde*	9	53	pur *pur*	19	25
loud *fort*	9	53	dynamic *dynamique*	19	25
soft *doux*	10	50	bow *archet*	20	22
full *plein*	11	47	homogeneous *homogène*	20	22
rich *riche*	12	41	nuance *nuance*	20	22
low** *bas*	12	41	sustained *entretenu*	20	22
matter *matière*	13	38	color *couleur*	20	22
noise *bruit*	13	38	balanced *équilibré*	20	22
low** *basse*	13	38	temporal *temporel*	20	22
smooth *lisse*	14	34	enveloping *enveloppant*	20	22
high *haut*	15	34	variation *variation*	20	22
piano *piano* (Musical term)	15	34	vibrer *vibrating*	20	22
clear *clair*	15	34	stable *stable*	20	22

Table 2: Top 50 relevant words used for defining the four words. W/I : Word/Interviewee frequency. * words that communicate different concepts in French language but have the same translation in English.

The top 50 words used for the definition are ranked according to the word/interviewee frequency. For instance, 100% of the experts employed the term "warm", which is not surprising as it is one of the four investigated words. We can find acoustic information in the 50 most-used terms on spectral, temporal or loudness aspects. It is worth noting that there are many ways to indicate spectral or pitch information in French. For instance *aigu*, and *haut*, will both describe high pitch or high frequencies. The same is true for *grave*, *bas*. *Basse* is more ambiguous as it can describe the bass clarinet or low frequencies. The fact that experts mentioned instruments like *clarinette* or *percussion* is vastly influenced by the sound dataset, which is only composed of musical instruments. Many words imply a metaphorical description such as *pur*, *plein*, and *agressif*, among others.

3.1 Description strategies

Informed by the literature on semantic analysis of timbre [13–16], and our assumptions, we proposed 10 description strategies in order to analyse the definitions. Coding and re-coding a corpus is a common step present in many qualitative studies like Weston's study [24] or in Saitis's study [8] in the field of semantic timbre. In order to test the validity of the description strategies, an inter-rater analysis was performed. The four authors sorted the 50 top lemmas present in Tab. 2 into the 10

categories. Then, we computed Fleiss' kappa, an index that evaluates the inter-rater agreement in the sorting task. Since a big part of the timbre language is metaphorical and can be interpreted through different modalities, there was some disagreement but we suggested to always consider the context of the word and the definition from *Trésor de la langue française* database [5] to make a decision. Nonetheless, Kappa got a score of $\kappa = .69$, which reflects a substantial agreement according to Landis and Koch's interpretation on Kappa values. We then refined the categories and their definitions by collectively sorting the top 50 words one more time.

The 10 categories of description strategy were grouped in three greater classes:
The first class gathers all **acoustic aspects** of sounds like the discourse related to reduced listening proposed by Carron [4]. 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 **the sound domain** such as "resonant", "noisy", or "nasal".

The second class groups all of the **metaphorical aspects** of sound. Hence, there is the **crossmodal correspondance (CMC)** category that you can find in other works [14–16] that contains lexical fields related to other senses, such as sight, touch and taste. For example, "soft", "harsh", "clear"...
A second metaphorical sub-category group of lemmas related to **matter**'s characteristics is another category present in Wallmark's work [16] that discusses the material, shape and density of sound as matter or as an object. For example, "round", "full", "organic"...
The third category is the one related to **affect, emotional value and judgement** vocabulary employed for sound description, present in all the studies cited above. Although Q6 was supposed to concentrate all descriptions on affect, experts would still use this dimension to give the definitions in Q2. For example, "aggressive", "warm"* (see Tab. 2), "rich", "pleasant"...
The last metaphorical category groups all the vocabulary that evokes sound's action having an **effect** on the listener, the surroundings, or its behaviour. For example, "enveloping", "scratching".
The third class collects the references to the sound source. It corresponds to the **causal** listening evoked by Carron [4] and Gaver [19]. There were information on **the source** mainly represented by naming the instruments present in the dataset. For example, "clarinet", "percussion", "voice".
There were also characterisations of the **excitation mode or the playing technique**. For example, "bow", "rub", "impact" ...

3.2 Definitions and sound samples

Figure 1 presents how the experts spoke about the terms. We observe that the acoustic aspects of *Bright*, were almost exclusively described through spectral features, while *round* is defined both spectrally and temporally.

[5] http://atilf.atilf.fr/

Acoustic	Sound vocabulary, Dynamic, Spectrum, Temporality
Source	Source, Excitation mode/Playing technique
Metaphore	CMC, Matter (Material, shape, density), Effect, Affect

Table 3: Categories of description strategies built for the coding of the transcripts.

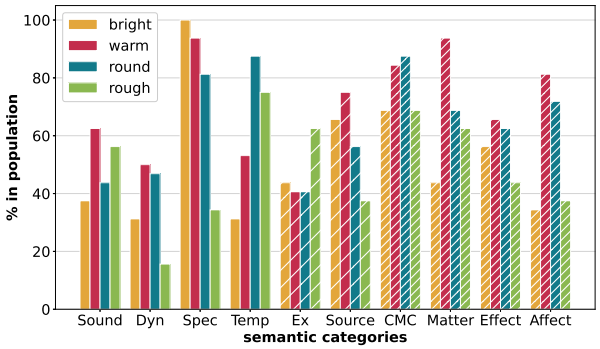

Figure 1: Proportion of experts that expressed information in each of the 10 categories for the four terms. Normal : Acoustic ; single stripe : Source ; double stripe : Metaphoric

In order to orient the definitions toward acoustic features of sound, we started by describing the acoustic semantic categories first and then joined the related metaphorical description (e.g. "a soft attack" - "harmonic richness"...). Then, considering its relevance, we proceeded to purely metaphoric descriptions, references to the source or the excitation mode. All percentage values in the definitions correspond to the proportion of experts among the total number of experts that evoked a semantic category or concept. We also reported the relevant sound samples for each term, in order to relate them to the descriptions from Q2. At the end of the description, we tried to summarize the relevant information for each term.

3.2.1 Bright

When defining brightness, a large part of the expert population (97%) gave spectral information. 90% of the population indicated that a *bright* sound has substantial spectral content located in the high frequencies. More generally, it means that a *bright* sound gives a sensation of richness in high frequencies. There was sometimes confusion between whether the fundamental frequency or the spectral content was high. Some experts would say that a low pitched sound is difficult to identify as *bright* (38%). Even though the temporal aspects don't seem to be very important to the majority of the population (see Fig. 1, 25% of the experts considered that the presence of a rather "sharp/fast" attack adds a feeling of brightness. That can be related to an excitation mode cited by 22%

of the experts such as ”articulated”, ”impact sound”, or ”tongued”. For the mentioned sound sources, 31% of the experts referred to metallic sources (glockenspiel, brass instruments, metal, concert flute...). On a metaphorical level, 69% of the experts defined sound through the CMC category mostly in the lexical field of light. Examples: ”clear”, ”halo”, ”shine”, ”luminous”.

As for the sound samples, coherently to the verbal description, most of the sounds chosen were played by high pitched instruments (trumpet, glockenspiel, piccolo, violin) on high register. If an instrument played louder, more harmonics are perceived, which seemed to influence the selection of sounds. 90% of the experts estimated that the sound of the glockenspiel played with hard sticks was a *bright* sound. The fact that they explicitly chose glockenspiel sound with the hard stick corroborates the fact that brightness is related to the attack of sound. The second kind of sound most selected was the trumpet (75%) played *ordinario* [6] mostly in the high register (C5-C6) and in loud dynamics (*forte-fortissimo*).

To summarize: A *bright* sound has most of the spectral energy in the high frequencies. It is often a high-pitched sound. It can be composed with a sharp attack.

3.2.2 Warm

When investigating the term *warm*, we first notice the fact that it is semantically comparable to *round*. The two terms were compared either in the definition of *warm* or in the definition of *round* by the entire panel of experts. Some of the experts would even accidentally use one in the place of the other during the interviews. As for *bright*, a large part of the experts (94%) described *warm* through spectral aspects. 34% of the population evaluated a *warm* sound as ”low-pitched”. Since rich timbre is correlated with spectral variation by some studies in the English language ([25]), we chose to compare the term ”rich” to spectral information. The harmonic content was therefore sometimes described as ”rich” (28%), with spectral energy mainly located in the ”low” and ”medium” frequencies and less information in the high frequencies (87%). When referring to the temporal aspects (66%), the experts mostly qualified the attack as being ”not sharp”, or ”soft”. Considering the sound sources, they would give examples in the low string instruments like the cello or the double bass and some of them would evoke a ”breathed” component in the sound (22%). As for the excitation mode, the idea of a ”vibrated” sound is sometimes evoked as enhancing a *warm* aspect in sounds (22%). *Warm* is strongly related to affect and behavioral dimensions (by more than 80% of the population) in a very metaphorical way. Hence it is described as enveloping (30%), lively (13%) and compared to a human behavior or the human voice (13%). A *warm* sound is also positively described by 34% of the experts as ”comfortable”, ”pleasant”, ”reassuring”, ”generous”, ”giving happiness”...

[6] ordinario: standard playing technique

The most cited *warm* sound is the bass clarinet (69%), played *ordinario*, medium or low pitched (C2-C3), in small dynamics (*pianissimo-mezzoforte*). These sounds appear to have most of the spectral content close to the fundamental harmonic, with a breathing noise on the notes played *pianissimo*. 56% of the experts indicated that for the cello played on a C3 or C4 *ordinario* (i.e. with a vibrato for the strings) *mezzoforte* or *piano* was a *warm* sound. It is low to medium pitched with a certain timbral or harmonic richness, the *vibrato* enhancing the perceived expressive aspects in the sound. We could imagine that both the breathy sound of the bass clarinet and the *vibrato* of the cello might add a feeling of richness that amplifies the *warm* quality of a sound. Finally, the sound of a French horn played *ordinario* between C2 and C5 in a *mezzoforte* or *pianissimo* dynamic was elected as *warm* by 34% of the experts.

To summarize: A *warm* sound seems to be a low-pitched or mid-low-pitched sound. It gives a feeling of spectral richness in the mid-low frequencies. It has a rather soft attack and it is a fairly pleasant sound for the listener, giving a sensation of envelopment.

3.2.3 Round

Round, as previously stated, was always compared to *warm*. 91% of experts gave a spectral description of *round*. While most of the experts said that the spectral content of a *round* sound is located in the mid-low frequencies, others said that it could also have energy in the high frequencies. The idea of richness is evoked by fewer experts than for *warm* (16%) and it is both used positively and negatively. Nonetheless 38% of the experts preferred the idea of a *round* sound in the mid-low frequencies with little spectral energy in the high frequencies, 16% using the term ”balanced” to this end. 32% of experts described a *round* sound or its spectrum as ”full”, which can be correlated to the fundamental frequency or the number of overtones ([12]). *Round* was also described through temporal aspects by 84% of the panel. More specifically, 70% of the experts focused on the nature of the attack, or the transient. Some experts would say that a *round* sound has no or little attack, some others would describe it metaphorically (CMC, Affect) like ”soft, ”not sharp”, ”not harsh” or ”not aggressive”. It was physically described as ”slow” or ”long”. Some experts described the release of a *round* sound as being ”long” or ”soft” explaining that a *round* sound are quite resonant. Another temporal aspect cited by some experts (28%) was temporal stability. Experts would describe a *round* sound as stable, with an absence of ”parasitic noise”. It can be related to metaphorical description of a *round* sound such as ”perfect”, ”soft”, ”pure” or ”homogeneous” (25%). Hence a *round* sound seems to carry little noise or variations over time. Dynamic was discussed by 50% of experts, and 19% of experts identified a round sound as not being very loud. As for *warm*, there were many positive descriptions in the

domain of affect, but to a lesser extent than *warm*. 34% of the experts thought a *round* sound to be ”pleasant”, ”reassuring”, ”expressive”, ”ideal”, ”generous” and ”not aggressive”.

The first sound to be selected by 60% of the experts was a *pizzicato* of doublebass, played C2 or C3 with an intensity ranging from *pianissimo* to *mezzoforte*. It is a low pitched and resonant sound with a majority of spectral energy around the fundamental frequency, and a soft attack. It doesn't have any special harmonic or dynamic variation. Even though it is one of the shortest sounds of the dataset, it matches the temporal and spectral definition. The marimba played *pianissimo* with soft sticks in a pitch interval from C2 to C5 was mentioned by 56% of the population as a *round* sound. It has similar spectral and temporal properties as the *pizzicato* of the double bass. 53% of the experts found the sound of the high register (C3-C4) of a bass tuba played *piano* to be *round*. It is also a long and stable sound, with a soft attack. The French horn, playing in the same register and in the same dynamic, is cited by 34% of the population. Here we have two kinds of sounds, one is damped, resonant and low pitched, the other is sustained and stable. They all have a soft attack that could be related to the attack time and a spectral energy located in the mid-low frequencies.

To summarize: A *round* sound has a soft attack, and is temporally stable. It tends to also have a soft release or a long resonance. A *round* sound is spectrally perceived as full with a spectral balance located in the mid-low frequencies.

3.2.4 Rough

From Figure 1, *rough* didn't have as clear a path of expression on the acoustic dimensions as the other terms. Nonetheless, it is possible to relate temporal aspects (62%) to elements from metaphorical categories such as Excitation mode (56%), CMC (72%) and Matter (62%).
Starting with the temporal and timbral aspects, a *rough* sound was defined as ”unstable”, ”irregular”, ”with variations”, ”noisy” (34%), ”with parasitic noise” (38%), or with a sensation of a pulse (28%). It can be linked to more metaphorical descriptions: via the Action/Behavior category, ”scratching”, ”shredding”, or ”random” (34%), via the CMC category, ”asperities”, ”texture”, ”resistance” or ”graininess” (32%) and via the Matter or affect category, ”not pure”, ”not homogeneous”, ”raw”, or ”not clean” (32%). Graininess was a concept mentioned by some experts (16%) as a different kind of roughness. However, the difference of perception between those terms remains unclear. Roughness was also expressed through the excitation mode of the source, like ”rubbing” or ”friction” (41%).

Unlike the three other terms, it seemed that the discriminant feature for *rough* sound sample selection wasn't the instrument but rather the playing technique. As it is not really related to the spectrum nor the pitch, several instruments with the same playing technique would produce a comparable sensation of roughness. Hence, 81% of the population evaluated a wind instrument playing a *flatterzunge* to be *rough*. The trumpet playing a *flatterzunge*, selected by 32% of the experts, might present a rougher *flatterzunge* than other instruments. The contrabassoon is interesting as its lowest note produced a similar effect of rapid pulses as the *flatterzunge*. It was mentioned by 41% of the experts as either being played *ordinario*, or *flatterzunge*. The second most frequently occurring type of *rough* sound was the multiphonics produced by the woodwinds (66%). The instrument producing the most *rough* multiphonics according to the panel were mainly the bassoon (56%) and the clarinet (25%). As there is a variety of multiphonics, it is notable that those mentioned presented the same pulsing effect as the *flatterzunge* paired with an inharmonicity, due to the multiple notes played at the same time. Finally, 41% of the experts chose string instruments playing *sul ponticello*, echoing the ”rubbing” excitation mode that was cited above. When playing *sul ponticello* a string instrument player rubs the string with the bow very close to the bridge, thus producing a friction sensation of instability of the harmonics and a noisy sound. It is unclear how the psychoacoustical definition of roughness relates to the *sul ponticello* sound, while it seems to match the sounds of *flatterzunge* and multiphonics.

To summarize: A *rough* sound is temporally unstable, it presents fast temporal variations that can bring some sort of noise. It gives a rubbing/scratching sensation.

4. GENERAL DISCUSSION

The goal of this study was to obtain definition and related sound samples for four terms frequently used for timbre description through the process of several expert interviews. The definitions of the terms were close to those obtained by Carron [15] in his sound lexicon. With the definitions, we get new clues on the acoustic features involved in the definition of the terms. They also display the variety of description strategies employed by sound experts in the French language. As we wanted to follow a methodology with few assumptions, we started with free verbalization (Q2, Q5). It allowed us to have new insights on the underlying acoustic features. In future work, Q5 will be investigated in order to reinforce the definition obtained in Q2, through an online survey presenting both verbalisations from Q2 and Q5.
While *bright* got a clear definition, *warm* and *round* seem to share similarities in their definitions, but they were better differentiated with the selection of sound samples, an observation that begs for further investigations. The same goes for *rough* that was mainly defined through CMC aspects but with very eloquent corresponding sound samples. This point suggests that verbal descriptions might not be a good way of defining certain sound attributes, and that giving a definition through sound samples is a better approach.

Finally the next step of this study will be to reveal the underlying acoustic features of these definitions.

5. REFERENCES

[1] G. Lemaitre, O. Houix, N. Misdariis, and P. Susini, "Listener expertise and sound identification influence the categorization of environmental sounds.," *Journal of Experimental Psychology: Applied*, vol. 16, no. 1, p. 16, 2010.

[2] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2902–2916, 2011.

[3] N. Misdariis, A. Minard, P. Susini, G. Lemaitre, S. McAdams, and E. Parizet, "Environmental sound perception: Metadescription and modeling based on independent primary studies," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2010, pp. 1–26, 2010.

[4] M. Carron, T. Rotureau, F. Dubois, N. Misdariis, and P. Susini, "Speaking about sounds: a tool for communication on sound features," *Journal of Design Research*, vol. 15, no. 2, pp. 85–109, 2017.

[5] A. C. Disley and D. M. Howard, "Spectral correlates of timbral semantics relating to the pipe organ," *Speech, Music and Hearing*, vol. 46, pp. 25–39, 2004.

[6] M. Bernays and C. Traube, "Verbal expression of piano timbre: multidimensional semantic space of adjectival descriptors," in *Proceedings of the International Symposium on Performance Science (ISPS2011)*, pp. 299–304, European Association of Conservatoires (AEC) Utrecht, Netherlands, 2011.

[7] C. Traube, *An interdisciplinary study of the timbre of the classical guitar*. McGill University, 2004.

[8] C. Saitis, B. L. Giordano, C. Fritz, and G. P. Scavone, "Perceptual evaluation of violins: A quantitative analysis of preference judgments by experienced players," *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 4002–4012, 2012.

[9] P. Cheminée, C. Gherghinoiu, and C. Besnainou, "Analyses des verbalisations libres sur le son du piano versus analyses acoustiques," in *Colloque interdisciplinaire de musicologie (CIM05), Montreal (Québec), Canada*, 2005.

[10] G. von Bismarck, "Timbre of steady sounds: A factorial investigation of its verbal attributes," *Acta Acustica united with Acustica*, vol. 30, no. 3, pp. 146–159, 1974.

[11] P. Schaeffer, "Traité des objets sonores," *Éditions du Seuil Paris*, 1966.

[12] A. Zacharakis, K. Pastiadis, and J. D. Reiss, "An interlanguage study of musical timbre semantic dimensions and their acoustic correlates," *Music Perception: An Interdisciplinary Journal*, vol. 31, no. 4, pp. 339–358, 2014.

[13] T. Porcello, "Speaking of sound: language and the professionalization of sound-recording engineers," *Social Studies of Science*, vol. 34, no. 5, pp. 733–758, 2004.

[14] A. Faure, *Des sons aux mots, comment parle-t-on du timbre musical?* PhD thesis, 2000.

[15] M. Carron, F. Dubois, N. Misdariis, and P. Susini, "Définir une identité sonore de marque: méthodologie et outils," 2015.

[16] Z. Wallmark, "A corpus analysis of timbre semantics in orchestration treatises," *Psychology of Music*, vol. 47, no. 4, pp. 585–605, 2019.

[17] M. Ilkowska and A. Miśkiewicz, "Sharpness versus brightness: A comparison of magnitude estimates," *Acta Acustica united with Acustica*, vol. 92, no. 5, pp. 812–819, 2006.

[18] E. Terhardt, "On the perception of periodic sound fluctuations (roughness)," *Acta Acustica united with Acustica*, vol. 30, no. 4, pp. 201–213, 1974.

[19] W. W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological psychology*, vol. 5, no. 1, pp. 1–29, 1993.

[20] G. Ballet, R. Borghesi, P. Hoffmann, and F. Lévy, "Studio online 3.0: An internet" killer application" for remote access to ircam sounds and processing tools," *Journée d'Informatique Musicale (JIM)*, 1999.

[21] A. Strauss and J. M. Corbin, *Grounded theory in practice*. Sage, 1997.

[22] B. Sagot, "The lefff, a freely available and large-coverage morphological and syntactic lexicon for french," 2010.

[23] A. Pearce, T. Brookes, and R. Mason, "Timbral attributes for sound effect library searching," in *Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio*, Audio Engineering Society, 2017.

[24] C. Weston, T. Gandell, J. Beauchamp, L. McAlpine, C. Wiseman, and C. Beauchamp, "Analyzing interview data: The development and evolution of a coding system," *Qualitative sociology*, vol. 24, no. 3, pp. 381–400, 2001.

[25] R. A. Kendall and E. C. Carterette, "Verbal attributes of simultaneous wind instrument timbres: Ii. adjectives induced from piston's" orchestration"," *Music Perception: An Interdisciplinary Journal*, vol. 10, no. 4, pp. 469–501, 1993.