Uncovering the meaning of four semantic attributes of sound: Bright, Rough, Round and Warm

Victor Rosi[1†], Olivier Houix[1], Nicolas Misdariis[1] and Patrick Susini[1]
[1] Sound perception and sound design group, STMS Lab (IRCAM-CNRS-SU), Paris, France
[†]Corresponding author: rosi@ircam.fr

Introduction

Several studies discuss the semantics of words used by sound professionals or musicians to describe timbre in particular situations, such as sound engineering or instrument playing. The present study aims to understand the use and the definition of four terms selected from the sound lexicon developed by Carron et al. (2017) as they are cited in numerous studies for sound description. 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, standardized definitions. This work is based on interviews with sound professionals from these different fields. The goal is to get definitions, or semantic portraits, for each word with corresponding sound samples from a musical instrument dataset.

Method

We organized individual interviews with 32 French-fluent sound professionals (musicians, composers, sound designers, acousticians...), during which the four terms were discussed sequentially with the participants. The study of one term had two main parts, during the first part, the interviewees were asked to give a definition of the studied term, then they selected sound samples from a musical instrument dataset that match their perception of the term. During the second part, the interviewees first chose sound samples that were opposed to the studied term and then tried to define the opposite concept. The sound dataset was mainly composed of the Ircam Studio-Online Library (SOL) mixed with parts of the Vienna Symphonic Library (VSL) for additional instruments. The dataset was presented to the participants through a Max/MSP interface they could manipulate.

After the transcription of the interviews, the definitions of the studied terms were processed with basic NLP (Natural Language Processing) steps: tokenization, lemmatization and filtering of the stop words. We assessed the lemma/interviewee frequency for each term (i.e. the number of interviewees using one lemma for each definition).

Results

Informed by the literature on semantic analysis of timbre (Wallmark, 2018; Carron et al., 2016; Porcello, 2004; Faure, 2000), and the experimental raw data, we proposed 10 categories in order to structure the data and to better compare the description strategies for the four terms. These categories are organized in three greater categories. The first one groups all the acoustic, the second one groups all the information on the source, and finally the third category gathers the metaphorical descriptions of sound. The categories were validated with an inter-rater agreement measure with the four authors (Fleiss' kappa $\kappa = 0.69$, $p<0{,}001$). The categories along with verbal examples used for the description of the terms extracted from the corpus are presented in Table 1.

Table 1 – Categories of description strategies with verbal examples translated in English along with the original verbatims in French.

Acoustic	
Spectral	high-pitch (*aigu*), harmonics (*harmoniques*), medium (*medium*)
Temporal	attack (*attaque*), release (*décroissance*), steady (*stable*)
Dynamic	*forte, piano, crescendo*
Sound specific semantic	nasal (*nasal*), resonant (*résonnant*), noisy (*bruité*)
Source related	
Source	trumpet (*trompette*), voice (*voix*), orchestra (*orchestre*)
Excitation mode	rub (*frotter*), *vibrato*, breathing (*souffler*)
Metaphoric	
Crossmodal correspondance (CMC)	warm (*chaud*), harsh (*dur*), clear (*clair*)
Matter (shape, density, material)	round (*rond*), full (*plein*), organic (*organique*)
Effect	enveloping (*enveloppant*), scratching (*qui gratte*), straightforward (*franc*)
Affect	pleasant (*agréable*), aggressive (*agressif*), comforting (*réconfortant*)

The description strategies most used for each term helped to shape the definitions. For instance, we noted that all the participants defined *Bright* through spectral descriptions, while *Rough* was more often explained metaphorically through analogies to the sense of touch, with temporal aspects or with definitions of an excitation mode. Finally, *Round* and *Warm* shared many similarities in their metaphorical and spectral descriptions, although, *Round* seems to be substantially more described by temporal aspects compared to *Warm*.

From these results, by parsing the context of the most frequently occurring lemmas in each definition, and with a comparison with the most elected sound samples we were able to summarize semantic portraits for each term:

- A *bright* sound has most of the spectral energy in the high frequencies. It is often a high-pitched sound that can be composed with a sharp attack.
- 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. A *warm* sound has a rather soft attack. It is a fairly pleasant sound that gives a sensation of envelopment.
- 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.
- A *rough* sound is temporally unstable; it presents fast temporal variations that can bring some sort of noise. It gives a rubbing/scratching sensation.

While these free verbalizations allowed us to dig deep into the nature of the four types of sounds, it makes the work of the researcher tedious and conducive to interpretation as there are many syntactic elements, negations or quantifiers to take into account in the definitions. Because of the diverse and sometimes conflicting nature of the definitions given by the experts, it is necessary to homogenize, reduce and hierarchize the information gathered in order to formulate both accurate and detailed definitions.

In a second study, a corpus of phrases was extracted from the verbatims and related the most occurring lemmas selected from the definitions of each term to their oppositions gathered in the interviews. As part of an online experiment currently in progress, we want to ask a bigger population of sound professionals the degree of familiarity that they associate with the phrases of the corpus and the relevance of this phrase to the definition of the associated term. There are other subsequent goals to this study: first, clustering some presumably similar descriptions (e.g. *"a sound with a soft attack", "a sound with a slow attack", "a sound*

without an attack"), enabling a disambiguation and reduction of the information. secondly, we expect to better understand the use of certain metaphorical description such as *rich*, or *full.*

Discussion

This two-part study allows us to report the variety of description strategies employed by sound experts in the French language. The methodology employed could be used for the study of other semantically ambiguous terms related to sound. Finally, the definitions formulated and the sound samples will be incorporated in the sound lexicon (*SpeaK*) currently in development, following Carron's work. One of the purposes of this lexicon is to enable better communication about timbre descriptions of sound in a sound design process.

The next step of our study is to connect the definitions of the four terms, *bright, warm, round* and *rough,* with their acoustic characterizations. Firstly, we chose to annotate the sound dataset used during the interviews (~600 sounds) with the four terms. To that end, we are currently adapting a novel annotation experiment called *Best-Worst Scaling* (BWS). During a BWS procedure, at each trial, participants are asked to elect the best and the worst items along a latent subjective dimension in a tuple of *N* items. At the end of the procedure, scores are computed for each item using for instance a simple counting method that results in a ranking of all the items. For example with the term *Bright,* we could gather scores from the most to the least *bright* sound of our dataset.

Previous studies (Hollis & Westbury, 2018; Kiritchenko & Mohammad, 2017) have adapted this method, originally designed for small datasets, to semantic research with many-item datasets. The comparison with procedures using rating scales showed that BWS gives better consistency. BWS seems to have the perks of pairwise comparison without its time consuming downfall that prevents from using such methodology in many-item paradigms.

Following this step, we imagine a feature extraction procedure in the form of a machine learning experiment whose purpose will be to obtain the salient acoustic correlates responsible for the ranking of the sounds along each studied term. Ultimately, we wish to create a validation experiment that will confront the previously obtained definitions and the sound samples elected by the experts with designed sounds depending on the result of the machine learning experiment. These results aim to propose a methodology in order to define terms frequently used for timbre description. This approach could be used for other terms and languages.

References

Carron, M., Rotureau, T., Dubois, F., Misdariis, N., & Susini, P. (2016). Speaking about sounds: A tool for communication on sound features. *J Design Res 15*(2), 85–109.

Faure, A. (2000). *Des sons aux mots, comment parle-t-on du timbre musical?* (PhD Thesis). Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris.

Porcello, T. (2004). Speaking of Sound: Language and the Professionalization of Sound-Recording Engineers. *Social Studies of Science*, *34*(5), 733–758. https://doi.org/10.1177/0306312704047328

Wallmark, Z. (2018). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music*, *21, 1-21*. https://doi.org/10.1177/0305735618768102

Hollis, G., & Westbury, C. (2018). When is best-worst best? A comparison of best-worst scaling, numeric estimation, and rating scales for collection of semantic norms. *Behavior Research Methods*, *50*(1), 115-133. https://doi.org/10.3758/s13428-017-1009-0

Kiritchenko, S., & Mohammad, S. M. (2017). Best-Worst Scaling More Reliable than Rating Scales : A Case Study on Sentiment Intensity Annotation. *ArXiv:1712.01765 [Cs]*. http://arxiv.org/abs/1712.01765

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling : Theory, methods and applications*. Cambridge: Cambridge University Press.