Chapter 9

Timbre

Hanna Järveläinen Giovanni De Poli

9.1 Definition of timbre modelling

Giving a definition of timbre modelling is a complicated task. The meaning of the term "timbre" in itself is somewhat unclear. The American National Standards Institute (ANSI) defines timbre as "... that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different". This obviously leaves lots of space for imagination, stating only that anything which is not loudness or pitch belongs under the title "timbre". Sometimes timbre is referred to as "tone color". Many previous, informal definitions agree that timbre is something which gives a musical instrument its identity, so that it may be distinguished from other instruments or instrument families. Another aspect of timbre is tone quality: even though a sound maintains its identity under varying conditions, its quality may change in many ways. For instance, the voice of the same speaker sounds different heard acoustically from a short distance than heard over the telephone line.

The attributes of sound that make up a timbre are numerous and partly unknown. The desire to control timbre by a small number of parameters has lead to a hunt for the most important dimensions that contribute to the identity of a sound. This brings us near to the idea of modelling: learning to understand a complex system through a simplification. The modelling of timbre is focused around two viewpoints: modelling of perception and sound modelling. The perceptional viewpoint aims at understanding the mechanisms of our hearing system that are responsible for the timbre sensation. The sound modelling viewpoint includes analysis, automatic identification and classification of sounds as well as sound synthesis.

The afore mentioned identity and quality aspects relate timbre to the sound source. This is important for the control of the timbre of synthesized sounds. Once the identity of a musical instrument, for instance, is created by a sound model, its timbre can be controlled by a limited number of parameters in the range typical of that instrument class. However, virtual synthetic sounds don't belong to any well-defined timbre class. For the analysis of the timbre of these sounds, new schemes should be developed to create the concept of sound object which is not related to any physical sound source.

Another aspect to be considered is the musical, or compositional one. This calls for a higher-level cognitional view of timbre as an information carrier. Timbre has a special meaning for musical sounds, produced both acoustically and electronically.

9.2 Perception of timbre

Timbre is described as anything which is not pitch, loudness or direction. It is the only one of the four perceptual characteristics of sound that cannot be described by numbers. The further definition of timbre has been subject to speculations. One of the frequently given is that timbre is the feature of sound that allows us to identify it and distinguish it from other sounds with the same pitch and loudness [1]. Along with identifying the sound source, timbre specifies its quality: for instance, a single speaker can create many different vocal timbres. The ambiguity of timbre as a concept creates many methodological problems for measuring timbre perception. For a presentation of these issues, see [2].

What results as timbre is a combination of several perceptual dimensions, whose number and quality are partly unknown. The main factors are obviously related to the spectral content. The number and organization of the spectral components decides whether a sound is perceived as tonal or harmonic, for instance. The spectral envelope, i.e. the relative amplitudes of the components, is another important factor. Also dynamic (time-variant) attributes seem to be salient.

Along with spectral characteristics, it has been shown that also temporal events and modulations have a great effect on timbre. The attack transients of musical instrument sounds contribute to the characteristic timbre of each type of instrument so much that if they are absent, it is hard to recognize the instrument any more, see [3] for discussion. The same goes for instance for vibrato in singing voice synthesis. Even though the temporal events do not create a certain kind of timbre themselves, connected to the spectral characteristics they seem to reinforce or bring out the unique timbre of the sound source. It seems, however, that the attacks are mostly important for the identification of sound sources, while the attributes that are present throughout the sound help us judge the timbre in a more general way [3].

The uniqueness of the timbre of a certain type of instrument, for instance, leaves another open question in timbre perception. Even though the spectro-temporal characteristics vary as a function of pitch and loudness, the identity of the sound source is unchanged. In this sense, timbre should be considered in relation to a sound source, a feature of a sound object whose other features are pitch, loudness, and direction. This way the interaction between each of the features could be captured in source-specific models. This kind of object-based viewpoint should become attractive with the development of the structured representations of sound [4], [5], [6].

Along with the physical characteristics that cause a perceived timbre, we should consider the information-carrying capacity of timbre, i.e., the different aspects of timbre perception. The most important one was mentioned earlier, the uniqueness of timbre for different types of sound sources, which allows identifying the sound or placing it to a certain category. Changes in timbre can also carry musical information by means of affecting musical tension which seems to be related to roughness, a perceptual measure for the effects of beats [7], [8]. Another example are vowel sounds in speech. Timbral effects can also help us make conclusions about the shape of the surrounding space, such as a concert hall. These examples seem to engage also top-down processing, meaning that our brain is sensitive to the kind of spectral patterns that are common in our environment and can recognize those [9].

9.2.1 Timbre space

During the past 20 years, there has been remarkable interest to establish a kind of perceptual space for timbre, i.e., to find a small number of relevant perceptual dimensions of timbre. The features of timbre are presented in a low-dimensional space whose orthogonal dimensions represent the most prominent components. A timbre space is constructed by a perceptual study of the desired type of sounds, for instance musical instrument sounds. Listeners will judge the perceived similarity of each timbre against all others, and from the similarity matrix derived from the ratings, the euclidean distances between each of the timbres can be calculated and presented in what is called the timbre space. A short distance between two timbres corresponds to high similarity and a long distance to high dissimilarity. The method of measuring perceptual distances is called multidimensional scaling (MDS) [10], and it is an efficient way of reducing the dimensionality of the timbre features.

The results obtained by the multidimensional scaling technique differ to a certain extent at the number and quality of the perceptual dimensions. For instance, for orchestral instruments a three-dimensional timbre space was found, whose axes were spectral energy, synchronization of the transients, and a temporal attribute related to the beginning of the sound [11]. Expanding this material by hybrid sounds brought more diverse findings. One of these timbre spaces included the dimensions brightness, steepness of the attack, and the offset between the rise of the high and the low frequency harmonics [12]. Later on, the spectral flux, i.e., the variation of the spectral content with time, was also found to be important [13], but this wasn't confirmed by a later study [14].

An important point is that the studies mentioned above used almost entirely string and wind instruments. When the set of sounds was extended to percussive instruments [15], it was found that although two or three common dimensions could be found related to the spectral centroid and rise time, the results were generally context-dependent. This shows that a unitary timbre space is hardly likely to be found even for the timbres of musical instruments.

For this reason the instrument sound description in the context of MPEG-7 [16] is based on separate timbre spaces for harmonic and percussive sounds [17]. A number of physical attributes, or signal descriptors, that best correlate with the similarity ratings and the qualitative axes of the timbre space, were derived according to the results of Krumhansl [13] and McAdams [14] for harmonic sounds: Log-attack-time, spectral centroid, spectral spread, spectral variation, and spectral deviation. For percussive sounds, three descriptors were found based on the work of Lakatos [15]. They are the Log-attack-time, temporal centroid and spectral centroid.

9.2.2 Quantization of timbre sensations

A draw-back of the MDS technique is that it is non-metric. However, the timbre sensation can also be described on many metric levels. Semantic bipolar rating scales are commonly used to acquire subjective descriptions of timbre. Typically, the same timbre is judged on a number of scales, for instance dull-brilliant, cold-warm, pure-rich, dull-sharp, full-empty, and compact-scattered. After an analysis, some of the scales may be found not to be independent, but correlated with others. Usually the timbre sensation can be described by a few most important scales [2]. The problem is that the semantic scales are hard to relate directly to certain physical attributes of the sound.

Between the semantic descriptions and numeric spectral measures there are a number of psychoacoustic indices, developed for quantitative description of sound quality and perceptual response to sounds. Measures like sharpness, roughness, or fluctuation strength have been derived by matching the results of listening experiments to the physical characteristics of the sound. The sharpness measure, for instance, is connected to the perceived loudness measure suggested by Zwicker [18], which integrates the energy over the critical bands. Rather than for describing all kinds of timbres, these measures are used in specific tasks to compose more complex indices for instance for sound quality in vehicles or consumer products.

The physical parameters that have been found reveal that many perceptual timbre features correspond to physical characteristics other than spectral envelope. Various temporal and spectro-temporal measures are used to characterize timbre. For instance, the attack time and the different decay times between harmonics contribute to the timbre of musical instrument sounds. Measures related to the interaural time and level differences (ITD and ILD) or the proportion of reflections and direct sound, etc. are used to compose indices of concert hall acoustics, such as brightness, warmth, or envelopment.

An interesting development is the search for the representation of timbre in the primary auditory cortex (AI) [19]. It was found that some of the cells in the AI are specialized in extracting the spectro-temporal features, which in psychoacoustic studies have been found important to timbre perception, such as the local bandwidth and asymmetry of spectral peaks, and their onset and offset transition rates.

Bibliography

[1] A. Houtsma, “Pitch and timbre: Definition, meaning, and use,” *J. New Music Research*, vol. 26, pp. 104–115, 1997.

[2] J. M. Hajda, R. A. Kendall, E. C. Carterette, and M. L. Hashberger, “Methodological issues in timbre research,” in *Perception and Cognition of Music* (J. Deliège and J. Sloboda, eds.), ch. 12, pp. 253–306, East Sussex, UK: Psychology Press, 1997.

[3] P. Iverson and C. Krumhansl, “Isolating the dynamic attributes of musical timbre,” *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2595–2563, 1993.

[4] ISO/IEC, “ISO/IEC IS 14496-3 Information Technology – Coding of Audiovisual Objects, Part 3: Audio,” 1999.

[5] B. Vercoe, W. G. Gardner, and E. D. Scheirer, “Structured audio: Creation, transmission, and rendering of parametric sound representations,” *Proc. IEEE*, vol. 86, no. 5, pp. 922–940, 1998.

[6] E. D. Scheirer and J.-W. Yang, “Synthetic and SNHC audio in MPEG-4,” *Signal Processing: Image Communication*, vol. 15, pp. 445–461, 2000.

[7] D. Pressnitzer, S. McAdams, S. Winsberg, and J. Fineberg, “Roughness and musical tension of orchestral timbres,” in *Proc. 4th International Conference on Music Perception and Cognition*, pp. 85–90, 1996.

[8] D. Pressnitzer, S. McAdams, S. Winsberg, and J. Fineberg, “Perception of musical tension for non-tonal orchestral timbres and its relation to psychoacoustic roughness,” *Perception and Psychophysics: Human Perception and Performance*, vol. 62, pp. 66–80, 2000.

[9] A. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, Massachusetts: MIT Press, 1990.

[10] J. B. Kruskall, “Nonmetric multidinemsional scaling: A numerical method,” *Psychometrika*, vol. 29, pp. 115–129, 1964.

[11] J. Grey, “Multidimensional perceptual scaling of musical timbres,” *J. Acoust. Soc. Am.*, vol. 61, no. 5, pp. 1270–1277, 1977.

[12] D. Wessel, “Timbre space as musical control structure,” *Computer Music J.*, vol. 3, no. 2, pp. 45–52, 1979.

[13] C. Krumhansl, “Why is musical timbre so hard to understand?,” in *Structure and perception of electroacoustic sound and music* (S. Nielzenand and O. Olsson, eds.), Amsterdam: Elsevier, 1989.

[14] S. McAdams, S. Winsberg, G. de Soete, and J. Krimphoff, “Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes,” *Psychological research*, vol. 58, pp. 177–192, 1995.

[15] S. Lakatos, “A common perceptual space for harmonic and percussive timbres,” *Perception and psychophysics*, vol. 62, pp. 1426–1439, 2000.

[16] MPEG-7, “Overview of MPEG-7.” www.cselt.it/mpeg/standards/mpeg-7/mpeg-7.htm.

[17] G. Peeters, S. McAdams, and P. Herrera, “Instrument sound description in the context of MPEG-7.” Proc. Int. Computer Music Conf., Berlin, Germany, 2000.

[18] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*. Springer Verlag, New York, 1990.

[19] P. Ru and A. Shamma, “Representation of musical timbre in the auditory cortex,” *J. New Music Research*, vol. 26, no. 2, 1997.