ARCHIVES OF ACOUSTICS
34, 2, 137–155 (2009)

Multiple Dichotomies in Timbre Research

George PAPANIKOLAOU[(1)], Constantinos PASTIADIS[(2)]

Aristotle University of Thessaloniki
Thessaloniki 54124, Greece

[(1)] *Department of Electrical and Computer Engineering*
Laboratory of Electroacoustics and TV Systems
e-mail: pap@eng.auth.gr

[(2)] *Department of Musical Studies*
e-mail: pastiadi@mus.auth.gr

(*received November 28, 2008; accepted April 24, 2009*)

In this paper an overview of aspects, terminology and literature on contemporary research regarding timbre is presented. Timbre is a multidimensional entity, and research traces its multifaceted nature. The paper handles this structural complexity using a domain-task-results paradigm. Several domains of application are examined and various aspects of timbre questioning are outlined, although consideration of aspects in music and its contextual applications are postponed for a following detailed report for reasons of presentation compactness and extent. A self-evident differentiation of research categorization stems from the type of consideration of timbre as a perceptual attribute or as a manifestation of physical (either generative or modified after transmission) phenomena and processes. As more "axes" of differentiation also emerge, this work attempts to highlight issues that rise and propose possible research directions.

Keywords: timbre, sound quality, subjective judgements, objective measures, dissimilarity, MDS, dimensionality reduction, categorization, identification, classification, discrimination.

1. Introduction

The establishment of an ultimate definition of the term *timbre* appears as a still unsolved problem in acoustics. Historically, it has been addressed by many researchers and in several fields such as the physics of sound, music, speech, etc. It is a multidimensional attribute both in perceptual and acoustical terms [8]. Within such a lax framework, timbre may be considered as a duality between

physical and perceptual (auditory) or, alternatively, between objective and subjective, allowing operations such as identification or classification of sounds. From an objective point of view, timbral information is transmitted by sound sources and carried by sound waves. As such, a sound's amplitude in the time domain or, equivalently, its spectral evolution over time contain all necessary ingredients for subsequent timbre interpretations. Thus, a sound information decoding scheme, that consists of carefully selected indices and procedures, could ideally be used to define an appropriate acoustic space of timbre retrieval. From a subjective point of view, timbre encompasses all auditory attributes that facilitate unique sound recognition or identification. In general, unique subjective sound recognition/identification could be based on microscopic or macroscopic cues. For some categories of sounds with steady characteristics (such as stationary spectral constituents), timbral fingerprints must be abstracted beyond and apart from their pitch, loudness or duration through information decoding by human perception [34]. However, physical and functional complexity of biological systems still keep the discovery of such decoding schemes an intriguingly difficult problem, for which even partial solutions remain under investigation [50].

In addition to this type of dualism in its notion, timbre seems to reflect (and obtain) its multidimensional character from a wealth of attributes which are met in various types of analyses and applications. Timbral characteristics are exploited in several fields such as music, speech, biomedicine, machine diagnostics, or environmental acoustics, to mention a few, while research approaches employ various types of methodologies.

The present paper attempts a compact and multidisciplinary exploration of terminology, research methods, fields of application, descriptions, and models met in contemporary literature on sound timbre. Although such a literature review cannot be exhaustive, we believe that, by summarizing in this way, various aspects of timbre questioning are outlined from different points of view, and the duality mentioned above is highlighted. Initially, the paper presents a brief overview of timbre related terminology. This presentation of terms and definitions may serve as an introductory step to the following sections that consider timbre literature on a base of multiple types of categorization.

2. Definitions and terminology

As already mentioned, historically the term *timbre* has been connected to many kinds of meanings [33, 71, 75]. Since the Helmholtz era, the term has been debatedly assigned as *"quality of tone"* or *"sound quality"* or *"quality of sound"*. ASA's definition (later adopted by IEC) was an attempt to provide the scientific community with a more clear (but not quite clear) exposition of *what timbre is not,* namely neither loudness, pitch or duration. However, difficulties of such an abstracting approach met in the case of sounds with neither clear-cut pitch nor steady levels, or in the context of relative fields (music or music education

for example), left the problem of a proper definition unsolved. MARTIN [52] has even noted in his dissertation that "... (timbre) should be expunged from the vocabulary of hearing science". Despite the scientific dispute on the adoption of a formal definition, it seems that there is a more or less general acceptance of the functional aspect of timbre. That is, timbre (together with its acoustic descriptors, whichever they may be) enables operations such as classification, recognition, and identification of sounds or sound sources [21, 33, 41]. In a recent work [75] an attempt was made to systematize and clarify different aspects of the dualism and interplay between timbre's objective and subjective matter. Recently, the term *"sound quality"* has been re-introduced in various fields, albeit with a more lax relation to timbre, and with a significantly different meaning and implication compared to the historically assigned notion [36].

Timbre spaces are conceptions used to arrange "timbre" entities in a way that reflects some kind of measured or estimated similarity (or dissimilarity) [8]. Their major usefulness stems from their ability to offer visualization of timbral characteristics and to facilitate timbral structure investigations. In order for such a space to have a useful and interpretable form, its dimensionality must be kept somehow low (to be graphically displayed). However, the dimensions of the space may or may not have a direct and physical interpretation. For example, if the initial data represent vectors of some known acoustical features, such a dimensionality reduction will conclude with a compact set of dimensions that comprise combinations of acoustic parameters, thus gaining a possible physical interpretation. On the other hand, if the initial data represent perceptual judgments, one may result with a reduced set of dimensions, for which a physical interpretation must be sought within sets of acoustic correlates [21]. Such methods of dimensionality reduction are used in various applications such as data mining, visualization, structure detection, etc. Details may be found in many statistics and pattern recognition texts under the terms Principal Component Analysis (PCA), Factor Analysis (FA) and Multidimensional Scaling (MDS), although PCA and FA are not based on distance data but on feature vector representations of the studied objects. An interesting possibility arises when perceptual dimensions of timbre spaces are interpretable and correlate with certain acoustic descriptors (or their combinations); build an equivalent metric "physical" space for timbre, whose dimensions are these descriptors (or their combinations) [53, 70].

Multidimensional Scaling (MDS) is one of the most frequently used analysis frameworks in timbre perception [8, 26, 53, 54]. Applied in many fields (logistics, psychology, political sciences etc.), it facilitates the construction of structured spaces from dissimilarity or distance data. In timbre research, where usually no a priori knowledge about the nature of many inherent perceptual dimensions exists, it can be used to build perceptual spaces of reduced dimensionality for which physical interpretation is often possible. Several types of MDS models exist and have been applied to timbre research, according to the distance optimization strategy used, such as MDSCAL, INDSCAL, etc.

3. Fields of application: tasks and methods

The referral framework of research methods and tools actually reflects, at a significant degree, the underlying division of tasks and application domains with timbre related interest. It may be conceptually divided into two major and broad categories (1) those oriented at automated applications and are not based on human judgments, and (2) those oriented at perceptual investigations and are based on human judgments. The first category includes the almost inexhaustible reservoir of methods that employ objective measures, computational models, algorithmic and machine intelligence estimations, etc. in tasks such as sound or source (and properties) classification, recognition, identification, assessment, etc. [1, 41]. The second category is based on subjective and perceptual base judgments' processing in similar tasks. This second category can be subdivided into three general types [49, 84]: psychoacoustical methods, methods of cognitive psychology and methods of ecological psychology. A separate approach could include neurophysiological and neuroimaging techniques. The psychoacoustical methods try to discover the relation between perception and acoustics. Their experimental and analysis procedures may employ positioning on verbal (adjective) characterization of perceptual attributes and data reduction by means of factor analysis (FA), or dissimilarity rating followed by MDS techniques. On the other hand, cognitive psychology approaches deal with the ways that sensory information (coding signal properties) is represented and processed by the cognitive mechanisms. The so called "information processing" approach compares incoming sound properties to previous knowledge, and source recognition is performed by means of properties invariants. The "psycholinguistic approach" exploits the human linguistic representation capabilities in a two-stage scheme; initially, a direct linguistic identification and categorization of sound source is attempted based on the sound event's properties, and if this is not possible, the source is described by its acoustic properties. The "ecological psychology" approach deals directly with the sound sources' properties, and not the properties of the signals. Thus source identification is based on "source invariants" from its physical properties, and not on invariants of perceptually coded properties of sound signals. Ecological principles have been deployed in environmental sounds' characterizations, or "everyday listening" inferences on sound events recognition.

Not prematurely, one might ask if the first category that we mentioned is really addressing timbre, which, in strict sense, is a perceptual attribute. In this paper, we decide to widely expand our consideration into applications which in one or another way might have potential timbre perception implications. For example, one could consider the perceptual factors that enable a sonar operator to discriminate between biological and non-biological sounds, while at the same time systems for automatic classification of such sounds may be deployed. However, in this perspective, a major issue arises on the perceptual relevance of the employed objective descriptors. A few of them can be interpreted and appended to per-

ceptual quantities (for example, LPC coefficients reflect the formant structure of speech signals, enabling correlations to factors affecting speech sounds' timbre). But, other descriptors still lack such interpretation. Despite the gap between the two broad categories of our perspective, significant overlap and strong interactions are found between the approaches and algorithms used by both of them [42, 89]. For example, automatic sound classification may be greatly enhanced by cognitively inspired signal processing, and, as already mentioned, several features that are used in machine conducted operations are directly reflected on perceptual characteristics.

Applications with inherent timbre potential span a wide area of fields, including (among others) environmental acoustics, underwater acoustics, biomedicine, product/appliances design and operation, machinery diagnostics, voice and speech, singing, and of course music and its related branches (musical acoustics, music information retrieval, etc.).

3.1. Environmental acoustics

Environmental sounds include the variety of sound types answered in everyday life (animal sounds, household appliances sounds, traffic sounds, conversations, etc). Several works have studied the effect of acoustic factors pertaining to categorization and characterization/identification. In [27] the psycholinguistic approach is used for sounds in urban soundscapes. Moreover [18], urban soundscapes sounds are categorized into categories related to social activities. In [30] subjects' identification capabilities for various types of environmental sounds are examined after spectral manipulation. Listeners' effectiveness in correct identification differed according to the type of filtering used. Event-modulated noises (EMN) were also used to study the contribution of temporal factors, showing that identification of EMN were related with acoustic features like the envelope shape, periodicity, and the consistency of temporal changes across frequency channels. In [31] acoustical correlates in similarity and categorization tasks of environmental sounds are investigated using MDS. In [56] different timbre spaces were identified for each sound class of used sounds (impacts, motor, instrument-like). However, one acoustic feature (timbral brightness) was common to all sound classes.

3.2. Underwater acoustics

In the field of underwater acoustics, sound identification is crucial for applications like sonars, underwater acoustical imaging, etc. Following the initial broad categorization, both types of methods have been reported. In automatic classification of underwater recordings various classifiers and descriptors were employed [84]: statistical signal processing, time-frequency analysis (Wigner-Ville distribution, STFT, Wavelets and Wavelet Packets), fuzzy logic and neural network frameworks. Auditory modeling was employed in [63] and [5]. In [84] several

timbre related acoustic measures after MDS (spectral centroid, spectral variation, spectral deviation, spectral flux, temporal centroid, spectral spread, log-attack time), statistical measures of rhythmogram scales, statistical signal indices (ridge feature vector) and ERB spectral representations, are examined in automatic classification using fuzzy k-NN and Gaussian Mixture Models.

3.3. Biomedicine

In biomedicine, recognition, extraction and classification of transients, pulsation, and continuant sounds is of crucial diagnostic importance. [29] reports the use of homomorphic filtering and GAL (Grow and Learn) networks for segmentation and classification of Phonocardiogram (PCG) signals without using a reference signal, employing Daubechies-2 wavelet detail coefficients at the second decomposition level. In [2] significant results using wavelets, fractal dimensions and recurrence quantification analysis for feature extraction are reported and a neural network for discrimination and classification of pathological murmurs from physiological ones in PCG signals. In [16] PCG signal analysis is performed using FT, STFT, WD, CWT, DWT and PWT. In [85] WT and STFT features are used in an expert diagnosis system for Doppler Ultrasound cardiac signals. In [17] time domain, spectral and wavelet features, together with a neural network, were used in an autonomous system appropriate for long-term, unsupervised monitoring of bowel sounds. In [51] higher-order zero-crossings were employed for discrimination of bowel sound patterns between ascites patients and controls. In lung sound analysis, similar methods and tools have been employed. [38] used AR model coefficients, wavelet coefficients and crackle parameters as features, and k-NN and artificial neural network (ANN) classifiers for respiratory sounds.

3.4. Product quality and machine diagnostics

As mentioned earlier, the terminology has recently re-adopted "quality" as a notion of timbre, in domains such as voice/speech, which will be discussed later. Another field in which the term "quality" has gained ground is product design in consumer or industrial applications, as well as services. Moreover, the term is bound to overall product quality, thus signifying the impact of acoustical characteristics to a product's functional adequacy and commercial acceptance [36, 37]. During the last decade or so, a merging of methods and tools, traditionally used in timbre research, has been observed in the description of sound quality of products, and its quantified results were embedded in design and manufacturing process. Psychoacoustical methods like MDS or differentials, account for a significant amount of acoustic product evaluation [20, 80]. Moreover, the need for as accurate sound reproduction in listeners' ears as possible, in order to simulate the original acoustic field, has also highlighted the value of binaural signal recording and modeling in the assessment procedure. Actually,

the evaluation can be based both on direct subjective judgments or predicted through carefully selected acoustic features such as time-varying or stationary loudness, roughness, sharpness, fluctuation strength, tone-to-noise ratio, prominence ratio, etc [7, 13]. In [81] a subjective evaluation of air-conditioning sounds using MDS revealed the contribution of noise-to-harmonic ratio, spectral center of gravity and loudness in overall appliance rating. In a complementary fashion, [49] incorporate timbre analysis of car horns in warning signals messaging. Several more works can be found in the fields of transportation (e.g. vehicle noises), household appliances, industrial products, etc. A wavelet pre-processing neural network (WT-NN) is employed in [87] for sound-quality prediction of non-stationary vehicle noises. Similarly, [68] uses an FIR neural network system to predict sound quality judgments of environment, traffic, house appliances, and industry, based on one-third octave spectra.

In the related field of machine diagnostics, timbre analysis could also be beneficial. Fault diagnosis has traditionally been conducted using vibration monitoring, current monitoring, chemical monitoring, etc. Machinery sounds were usually considered as noise and treated accordingly. However, emitted sound analysis has, relatively recently, been introduced in the discussed framework. In [4] fault detection is demonstrated by means of simple procedures based on root mean square (RMS) of the sound signal, power spectrum density, STFT and Hilbert transform. In [67] an approach applying simple statistical criteria on individual power levels from a gammatone filterbank, simulating the human auditory perception, are used as features for nondestructive machine fault detection, for fault indication.

3.5. Voice/speech

Much of the attention paid to timbre perception has been focused in the fields of speech/voice and musical acoustics, with applications ranging from voice type and speaker recognition to singing and musical instruments related tasks. Even though we present only briefly contemporary trends related to timbre in the fields of voice and speech, these fields still hold the lion's share in the extent of discussion. We feel that a more detailed and expanded exposition of topics in the fields of singing and music would deserve a separate and dedicated work.

The voice is perhaps the most important tool for direct communication between humans (and also with animals). It constitutes the fundamental mechanism for carrying not only speech, but also a variety of related information such as identity (e.g. information regarding characteristics used to identify or recognize a speaker) and emotional state, while in many cases it readily reflects the physiological status of the speaker thus allowing the determination of possible pathologic or other biomedical conditions [3, 14, 65]. Although timbre in voice/speech is often considered at the segmental level, in a more extended consideration one could summarize timbre related tasks as:

- Speaker identification (e.g. determination or discrimination of speaking person's identity, gender, age, local accent, etc);
- Affective characterization (e.g. determination of speaker's emotional state);
- Specific voice quality related characterizations (e.g. voice type (register), vocal aesthetics, voice pathology, etc).

However, it must be pointed out that there is significant functional overlap of the above tasks and strong inherent relations between the characteristics used in each one of them. For example, the term *voice quality* has multiple interpretations depending on the domain that is being used and can encompass significant timbral quantities such as the voice register and the phonation type. These quantities, together with a combination of prosodic features, can enhance speaker identification or emotional characterization [11, 14, 43]. A further notice should be made; even more general speech perception tasks may be considered as timbre related. For example, judgments of vowel dissimilarity can be represented in a multidimensional perceptual space, whose dimensions correlate highly with speech formants [57, 66] characterizing the *vowel quality*.

In automatic speaker recognition, both segmental and supra-segmental features may be employed in models that use HMM and DTW for text-dependent recognition, and Vector Quantization (VQ), ergodic Hidden Markov Models (HMM), Gaussian Mixture Models (GMM) or long-term-statistics for text-independent recognition [22, 69]. Segmental spectral envelopes using linear prediction (LP), cepstrum, mel-frequency cepstrum (MFC) and perceptual linear prediction (PLP) have been utilized, while delta-cepstrum has been used to model spectral dynamics. Voice source information have been recently applied in the form of voice source cepstrum coefficients (VSCC) [28]. Finally, suprasegmental prosodic features such as F0 contours or energy measures have also been used, although segmental features have gathered most of the attention [22, 82]. Multidimensional scaling [47, 59] and semantic differentials [86] have been mainly used in voice quality/timbre research focused on speaker recognition. Although some differences at the talker sex level exist, the fundamental and formant frequencies appear as important cues for speaker identification and similarity judgments by humans at segmental level [47, 72], while temporal or glottal factors may affect judgments at phrase level [59, 60]. However, there is evidence that a Gestalt type of perception is involved in the features set, where the effect of each acoustic cue shall rather be assessed within the overall voice pattern [46]. Voiers [86] used factor analysis on 49 subjective bipolar descriptors of speaker characterization, concluding that 4 orthogonal factors (magnitude, clarity, roughness and animation) could be used to cluster voice samples. Speaker recognition accuracy by humans depends on factors like the familiarity of the heard voice or the duration of the material [62]. The neuronal aspects of speaker perception are not yet extensively explored. Both PET and fMRI studies regarding speaker identification are reported [3], showing significant activity in the right anterior temporal-lobe.

The vocal affective state of a speaker is an interesting aspect of timbre related functionality, and emotional information is included in voice quality considerations with significant impact on various speech applications [14], although some fundamental problems relating voice qualities and affective categorization exist [25]. Automatic emotion recognition from speech signals has been reported in several works, with prosodic features playing an important role. In [77] 20 pitch and energy related features were employed in the classification of 7 archetypal emotions (anger, disgust, fear, surprise, joy, neutral, sad). GMMs are used for classification in a global statistics framework of an utterance, while continuous HMMs are used to model temporal complexity using low-level instantaneous features. Recognition accuracy raises up to 86%. In [48], features such as pitch, log energy, formant, mel-band energies, MFCCs and added velocity/acceleration of pitch and MFCCs are used for emotion recognition with support vector machines (SVM), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and hidden Markov model (HMM) classifiers. In [83] used 16 LPC coefficients, 12 LPCC components, 16 LFPC components, 16 PLP coefficients, 20 MFCC components and jitter for vector quantization and subsequent classification with LDA, K-NN, and HMM for Mandarin utterances. Other similar features have also been employed [83]. In [9] successful application of acoustic features is reported in the identification of emotion or attitude. Neuroimaging techniques like PET and fMRI are also reported from the 1990s in the perception of manipulated emotional speech prosody, or nonverbal vocalizations [3].

Dealing with specific voice quality characterizations could be a miss of the inherent generalization of the notion of *voice quality.* Indeed, voice quality has no "generally accepted definition" [11, 12]. In this paper, voice quality and the general notion of voice timbre are treated as tightly interwoven if not coinciding. Moreover, such a treatment allows for the search of common grounds and connections between various domains of timbre related interests, singing and music included. A possible type of categorization regarding "quality" can refer to its function. In speech, voice quality serves for support of pragmatic language elements, affective and performance expression (prosodic features), identification of voicing operation characteristics, and contributes to speaker personalization. Similarly, in singing or music, the respective sound qualities contribute to musical structure, expression and of course instrument identification. In another categorization, voice quality can be regarded at segmental or suprasegmental levels, serving for different aspects of the above functioning. For example, in voice pathology, classification is often based on phoneme analysis, whereas in speaker identification, longer segments (phrases and sentences) are mostly used [22, 65]. Similarly, in singing and music, the expressive features are presented in higher level structural elements, whereas a singer's voice register or an instrument's register identification may be performed at either phrase or note level. Actually, this ubiquitous character of the term "quality" may pose significant problems in the determination of its perceptual dimensions and the search for their objective

(acoustical) counterparts in various applications (such as speech synthesis, voice pathology, telecommunications, etc).

The research on voice quality characterization (as well as in musical timbre) has regarded both sound generation and its perception levels [6, 10, 11, 15, 23, 40]. From the voice production's view, one could identify four general dimensions (not necessarily uncorrelated): phonation type (vocal fold, ventricular, vocal folds' turbulent noise, mixtures), aperiodicity (noise, jitter, shimmer), "pressure" (lax vs. tense quality) and "effort" (mainly loudness related) [12, 14]. Voicing registers (modal, vocal fry, falsetto) of vocal fold phonation are related to vocal folds vibration mechanisms and have been described using X-ray, EGG, glottal flow and radiated speech measures [6, 10, 11, 14, 24, 40]. In the fields of general voice quality description and voice pathology assessment or classification similar methods are reported for voice aperiodicities relating to breathiness, hoarseness, roughness, and multiphony [12, 24, 55, 58, 64, 65], although inconsistencies between studies may be found [58]. In [35], a method for automatic detection of vocal fry based on three acoustic measures (Power Peak (PwP), Intraframe Periodicity (IFP), and Interpulse Similarity Measure (IPS)) is reported. Also, in [9] some indices of vocal fry, pressed voice, aspiration, harsh-whispery voicing are also introduced. Automatic classification of pathological voices has been also widely studied. A collection of references may be found in [32, 73, 74, 76] and [65]. Voice qualities related to "pressure" dimension generally relate to the spectrum and function of vocal folds [15, 61]. Vocal "effort" (rather independent from "pressure") discriminates between "loud" and "weak" voices and is related to subglottal pressure, glottal tension, magnitude of flow and voice amplitude [14, 15]. Several acoustic features from glottal and radiated speech signals have been proposed for the description of the above dimensions. These include both time-domain and spectral parameters together with disturbance and perturbation indices. A collection of references may be found in [14, 65].

The studies on perception of voice quality have dealt (as already mentioned) both with semantic differentials and multidimensional scaling techniques. Voice register's relevance with perceptual characteristics such as pitch and loudness have been reported [11]. The perceptual rating of voice quality especially in the domain of clinical voice evaluation is equally important to objective descriptions that was mentioned earlier. Usually, judgments of various quality characterizations (e.g. breathiness) are made either on equal-appearing interval (EAI) scales or visual-analog (VA) scales [44]. Well known scales for vocal quality and function assessment include the Grade, Roughness, Breathiness, Asthenicity, Strain (GRBAS) scale, the Consensus Auditory Perceptual Evaluation–Voice (CAPE-V) scale, the Voice Related Quality of Life (V-RQOL) scale, the Iowa Patient's Voice Index (IPVI) scale, the Buffalo Voice Profile (BVP), and the Vocal Profile Analysis Scheme (VPA) [39, 88]. However, the validity and consistency of such judgments has been debated [43, 44]. A proposed model [46] for the control of factors that cause variability in listeners' judgments identifies four potential fac-

tors: instability of listeners' internal standards for different qualities, difficulties isolating individual attributes in complex acoustic voice patterns, measurement scale resolution, and the magnitude of the attribute being measured. Besides unidimensional scale ratings, MDS approaches are also reported. In [45], multidimensional and unidimensional ratings of breathiness and roughness in pathological voices are combined with acoustic features. The study shows that the multidimensional nature of voice quality greatly influences unidimensional ratings. Shrivastav [78], using MDS for the judgment of breathiness in pathological voices, proposed that listeners' variability could be minimized by averaging multiple takes of ratings for each listener. Hybrid techniques for assessment or classification of voice quality have been employed. In [23], assessment of pathological voices was performed by controlling a set of voice synthesis parameters, reminding of psychophysical methods used in psychoacoustics. In [19], a combination of acoustic and auditory-perceptual measure was used to increase the logistic regression-based classification accuracy for even a simple task such as normal vs. dysphonic discrimination. In [79], the partial loudness of a harmonic vocal signal and of the aspiration noise computed from an auditory modeling showed promising results in predicting perceptual ratings of breathiness.

4. Discussion

We feel that a dominant dichotomy emerges from the preceding research review; this of the perspective under which timbre is examined, namely either as a perceptual attribute or through an exploitation and interpretation of its physical correlates and their interactions with their perceptual counterparts. The examined subjects expand on the exploitation of this dual perspective within each application's aims and characteristics. Alternatively such a duality may be reflected by the orientation of research objectives either towards sound (and thus acoustic information) production/transmission characteristics, or perception.

Another clear distinction refers to the domain of application, such as biomedicine, speech/voice, etc. Such a distinction is also closely related to the type of sounds for which timbre is investigated; in underwater acoustics or biomedical applications transients are mostly answered, whereas both steady sounds and transients are of interest at other domains.

A third "axis" could represent the differentiation between machine-based (automated) tasks and human oriented (non-automated) tasks. For example, in speech/voice, timbre has important implications on tasks such as automatic speaker recognition, underwater events identification, etc. At the same time, subjective assessments, like quality evaluations, play an important role when the degree of a product's commercial acceptance is to be maximized or a human expert has to evaluate a speaker's vocal condition or a cardiac murmur's characteristics in clinical praxis. Often, within a domain of application, both kinds of tasks can

coexist (e.g. automated and human assessment in speech/voice pathology), with sharing of methods or characteristics and significant interplay between them. Actually, from an engineer's perspective, it seems natural to pay effort towards an automatization or machine mimicking of human activities and functions. The noted progress during the last years offers valid prospect that such attempts will push automated applications of timbre beyond chimeric limits. Within these tasks categories, one can discriminate several types of employed approaches; on one hand, machine-based tools include the wealth of objective acoustical features, distance measures, models, algorithms, and intelligent approaches (NN, fuzzy logic, etc), and on the other hand, manipulation, processing and inferencing from subjective and perceptual judgments employ psychoacoustical approaches, methods of cognitive psychology and methods of ecological psychology and neurophysiology/neuroimaging. Although there seems to exist a one-to-one correspondence between tasks and methods/tools (automated tasks employ machine-based methods and tools, while methods of perceptual evaluation are used to quantify mainly non-automated "human-based" tasks), in reality a significant mixing of approaches between tasks also takes place. Automated applications exploit results from perceptual investigations, while the latter may be inspired by the possibilities of information acquisition obtained by the former.

After the above schema, another categorization could be based on the kind of tasks, namely recognition, identification, discrimination, etc. At a following level, a significant factor of differentiation between studies on timbre is the nature of obtained results. For example, either verbal descriptors or objective acoustic features may be used directly to describe timbre entities, and significant effort is paid to quantify features and reduce the dimensionality of the features set. On the other hand, a more indirect approach, attempts to obtain timbre descriptions by dissimilarity evaluations (could be either perceptual or objective based on selected features). Eventually, the main objective is to build "timbre spaces" that represent timbre relationships in a valid manner. Although, generally, the term "timbre space" has been used to signify mainly a perceptual space, the correspondence between perceptual and objective measures is a major pursuit.

A major remark from the above review regards an observed imbalance in the use of perceptual approaches between domains of applications. Considerable work has been done both by perceptual studies and objective evaluation or automatic operations in fields like speech/voice. On the other hand, we believe that there remains plenty of room for studies of subjective evaluation in several other fields (such as biomedical applications). The possibilities that would arise from the synergistic addressing of recognition problems by field expertise, acoustics and perception would allow further improvement of automated methods and support understanding of cognitive and intelligent mechanisms in the accumulation of human knowledge of characteristic timbre patterns.

Such considerations also highlight the issue of perceptual relevance of employed objective measures in machine-based approaches, and, reversely, the de-

gree of contribution of auditory and perceptual models of timbre to efficiency improvement of machine-based approaches. Already, auditory inspired signal transformations are employed (such as Mel-frequency scalings). Could more elaborate models of peripheral or higher centers of audition provide additional support through the transformation of information? A definite answer would require both the existence of physiologically and mathematically supported implementations and their computational feasibility.

Similarly, what could be the contribution to machine-based approaches of dimensionality reduction offered by perceptual studies? Is it true that any kind of dimensionality reduction that inherently reduces variability actually also reduces the potential of information processing to be exploited by intelligent systems? Or, could perceptual parsimony enhance some automated tasks?

5. Conclusions

In this paper we presented a brief overview of aspects, terminology and literature on contemporary research regarding timbre.

Not only timbre is a multidimensional entity, but also research on timbre has a strongly "multidimensional" profile, that is timbre research may be conceptually charted within a space of several "axes" of multiple dichotomies.

Our approach follows a domain-task-results organization in order to systematize presentation. This work, as a glimpse on contemporary research, may prove useful as a literature collection and guideline for students and researchers. Several domains were examined, such as environmental acoustics, underwater acoustics, biomedicine, industrial product quality and machinery diagnostics, and voice/speech. Although timbre aspects could be of interest also in other fields, we feel that the examination, in the above domains, offers a compact display of the current research status. Of course, most of the discussed domains could actually require an individual thesis to produce a wider aperture on research literature. In any case, a brief presentation on such a multifarious subject could hardly escape an indicative character and in no case can be considered exhaustive.

Timbre studies in music and its related fields were deliberately omitted for brevity of presentation, within the limits of a published review paper. A more detailed and expanded exposition of issues in these applications, which, together with speech/voice, consist perhaps the biggest portion of timbre research subjects, would deserve a separate and dedicated work.

As the scope of this article was to explore and report on timbre research, it becomes obvious that several questions actually arise and new topics have to be addressed in future works for more versatile and robust models of timbre descriptions to be constructed and exploited in understanding of timbre function. On such a wide subject, S. Handel points "...(timbre) is the one I feel most uncomfortable about". Perhaps, this may be the curse of timbre, but also its inherent treasure.

References

[1] –, *Mpeg-7 overview*, Available from: http://www.chiariglione.org/MPEG/standards/mpeg-7/mpeg-7.htm

[2] Ahlstrom C., Hult P., Rask P., Karlsson J., Nylander E., Dahlström U., Ask P., *Feature extraction for systolic heart murmur classification*, Annals of Biomedical Engineering, **34**, 1666–77 (2006).

[3] Belin P., Fecteau S., Bédard C., *Thinking the voice: neural correlates of voice perception*, Trends in Cognitive Sciences, **8**, 129–135 (2004).

[4] Benko U., Petrovcic J., Juricic D., Tavcar J., Rejec J., Stefanovska A., *Fault diagnosis of a vacuum cleaner motor by means of sound analysis*, Journal of Sound and Vibration, **276**, 781–806 (2004).

[5] Bleeck S., Fox P.D., White P.R., O'meara N., *Auditory models and nonlinear filterbanks in underwater auralization*, The Journal of the Acoustical Society of America, **123**, 3344 (2008).

[6] Blomgren M., Chen Y., Ng M.l., Gilbert H.R., *Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers*, The Journal of the Acoustical Society of America, **103**, 2649–58 (1998).

[7] Bruel & Kjaer, *Bp-1589: pulse sound quality software – type 7698. The pulse application for analysing and improving sound quality*, Bruel & Kjaer, 2005.

[8] Caclin A., Mcadams S., Smith B.K., Winsberg S., *Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones*, The Journal of the Acoustical Society of America, **118**, 471–82 (2005).

[9] Carlos Toshinori Ishi, Hiroshi Ishiguro, Norihiro Hagita, *Evaluation of prosodic and voice quality features on automatic extraction of paralinguistic information*, Proceedings of Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on, 2006.

[10] Childers D.G., Ahn C., *Modeling the glottal volume-velocity waveform for three voice types*, The Journal of the Acoustical Society of America, **97**, 505–19 (1995).

[11] Childers D.G., Lee C.K., *Vocal quality factors: analysis, synthesis, and perception*, The Journal of the Acoustical Society of America, **90**, 2394–410 (1991).

[12] Childers D.G., *Speech processing and synthesis toolboxes*, WILEY, pub-WILEY 1999.

[13] Cox T., *Sound quality*, Available from: http://www.acoustics.salford.ac.uk/res/cox/sound_quality/

[14] D'alessandro C., *Voice quality in vocal communication*, Tutorial, Presented at: Interspeech, Antwerp, Belgium 2007.

[15] D'alessandro C., D'alessandro N., Le Beux S., Doval B., *Comparing time-domain and spectral-domain voice source models for gesture controlled vocal instruments*, Proceedings of 5th International Conference on Voice Physiology and Biomechanics, Tokyo, Japan 2006.

[16] Debbal S.M., Bereksi–Reguig F., *Computerized heart sounds analysis*, Computers in Biology and Medicine, **38**, 263–80 (2008).

[17] Dimoulas C., Kalliris G., Papanikolaou G., Petridis V., Kalampakas A., *Bowel-sound pattern analysis using wavelets and neural networks with application to long-term, unsupervised, gastrointestinal motility monitoring*, Expert Systems with Applications, **34**, 26–41 (2008).

[18] Dubois D., Guastavino C., *In search for soundscape indicators: physical descriptions of semantic categories*, Proceedings of Internoise 2006, Honolulu, Hawaii, USA 2006.

[19] Eadie T.L., Doyle P.C., *Classification of dysphonic voice: acoustic and auditory-perceptual measures*, Journal of Voice: Official Journal of the Voice Foundation, **19**, 1–14 (2005).

[20] Ellermeier W., Mader M., Daniel P., *Scaling the unpleasantness of sounds according to the btl model: ratio-scale representation and psychoacoustical analysis*, Acta Acustica united with Acustica, **90**, 101–107 (2004).

[21] Fiebrink R., Mcadams S., *Looking to a unified theory of timbre to improve timbral similarity systems in music information retrieval*, Course material, MUGS695: Timbre as a Form-Bearing Element in Music: Perceptual and Cognitive Issues, 2005.

[22] Furui S., *Recent advances in speaker recognition*, Pattern Recognition Letters, **18**, 859–872 (1997).

[23] Gerratt B.R., Kreiman J., *Measuring vocal quality with speech synthesis*, The Journal of the Acoustical Society of America, **110**, 2560–6 (2001).

[24] Gerratt B.R., Kreiman J., *Toward a taxonomy of nonmodal phonation*, Journal of Phonetics, **29**, 365–381 (2001).

[25] Gobl C., Ni Chasaide A., *The role of voice quality in communicating emotion, mood and attitude*, Speech Communication, **40**, 189–212 (2003).

[26] Grey J.M., *Multidimensional perceptual scaling of musical timbres*, The Journal of the Acoustical Society of America, **61**, 1270–7 (1977).

[27] Guastavino C., Dubois D., *From language and concepts to acoustics: how do people cognitively process soundscapes?*, Proceedings of Internoise 2006, Honolulu, Hawaii, USA 2006.

[28] Gudnason J., Brookes M., *Voice source cepstrum coefficients for speaker identification*, Proceedings of Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, 2008.

[29] Gupta C.N., Palaniappan R., Swaminathan S., *Classification of homomorphic segmented phonocardiogram signals using grow and learn network*, Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society, 2005.

[30] Gygi B., Kidd G.R., Watson C.S., *Spectral-temporal factors in the identification of environmental sounds*, The Journal of the Acoustical Society of America, **115**, 1252–1265 (2004).

[31] Gygi B., Kidd G.R., Watson C.S., *Similarity and categorization of environmental sounds*, Perception & Psychophysics, **69**, 839–855 (2007).

[32] Hadjitodorov S., Boyanov B., Teston B., *Laryngeal pathology detection by means of class-specific neural maps*, IEEE Transactions on Information Technology in Biomedicine, **4**, 68–73 (2000).

[33] Hajda J., *The effect of dynamic acoustical features on musical timbre*, [in:] Analysis, Synthesis, and Perception of Musical Sounds, Springer, New York 2007.

[34] Houtsma A.J.M., *Experiments on pitch perception: implications for music and other processes*, Archives of Acoustics, **32**, 475–490 (2007).

[35] Ishi C., Sakakibara K., Ishiguro H., Hagita N., *A method for automatic detection of vocal fry*, IEEE Transactions on Audio, Speech, and Language Processing, **16**, 47–56 (2008).

[36] Jekosch U., *Assigning meaning to sounds – semiotics in the context of product-sound design*, [in:] Communication Acoustics, Springer, Berlin/Heidelberg 2005.

[37] Jekosch U., *Basic concepts and terms of "quality", reconsidered in the context of product-sound quality*, Acta Acustica united with Acustica, **90**, 999–1006 (2004).

[38] Kahya Y.P., Yeginer M., Bilgic B., *Classifying respiratory sounds with different feature sets*, Proceedings of IEEE EMBS 28th Annual Conference, New York 2006.

[39] Karnell M.P., Melton S.D., Childes J.M., Coleman T.C., Dailey S.A., Hoffman H.T., *Reliability of clinician-based (grbas and cape-v) and patient-based (v-rqol and ipvi) documentation of voice disorders*, Journal of Voice, **21**, 576–590 (2007).

[40] Klatt D.H., Klatt L.C., *Analysis, synthesis, and perception of voice quality variations among female and male talkers*, The Journal of the Acoustical Society of America, **87**, 820–857 (1990).

[41] Kostek B., *Perception-based data processing in acoustics, applications to music information retrieval and psychophysiology*, Springer, Berlin/Heidelberg 2005.

[42] Kostek B., *"Computing with words" concept applied to musical information retrieval*, Electronic Notes in Theoretical Computer Science, **82**, 141–152 (2003).

[43] Kreiman J., Gerratt B.R., *Sources of listener disagreement in voice quality assessment*, The Journal of the Acoustical Society of America, **108**, 1867–76 (2000).

[44] Kreiman J., Gerratt B.R., *Validity of rating scale measures of voice quality*, The Journal of the Acoustical Society of America, **104**, 1598–608 (1998).

[45] Kreiman J., Gerratt B.R., Berke G.S., *The multidimensional nature of pathologic vocal quality*, The Journal of the Acoustical Society of America, **96**, 1291–302 (1994).

[46] Kreiman J., Gerratt B.R., Ito M., *When and why listeners disagree in voice quality assessment tasks*, The Journal of the Acoustical Society of America, **122**, 2354–64 (2007).

[47] Kuwabara H., Ohgushi K., *Experiments on voice qualities of vowels in males and females and correlation with acoustic features*, Language and Speech, **27 (Pt 2)**, 135–45.

[48] Kwon O., Chan K., Hao J., Lee T., *Emotion recognition by speech signals*, 2003.

[49] Lemaitre G., Susini P., Winsberg S., Mcadams S., Letinturier B., *The sound quality of car horns: a psychoacoustical study of timbre*, Acta Acustica united with Acustica, **93**, 457–468 (2007).

[50] Lewicki M.S., *Efficient coding of natural sounds*, Nature Neuroscience, **5**, 356–363 (2002).

[51] Liatsos C., Hadjileontiadis L., Mavrogiannis C., Patch D., Panas S., Burroughs A., *Bowel sounds analysis: a novel noninvasive method for diagnosis of small-volume ascites*, Digestive Diseases and Sciences, **48**, 1630–1636 (2003).

[52] Martin K.D., *Sound-source recognition: a theory and computational model*, Ph.D. thesis, Massachusetts Institute of Technology, 1999.

[53] Mcadams S., *Perspectives on the contribution of timbre to musical structure*, Computer Music Journal, **23**, 85–102 (1999).

[54] Mcadams S., Winsberg S., Donnadieu S., Soete G., Krimphoff J., *Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes*, Psychological Research, **58**, 177–192 (1995).

[55] Michaelis D., Fröhlich M., Strube H.W., *Selection and combination of acoustic features for the description of pathologic voices*, The Journal of the Acoustical Society of America, **103**, 1628–39 (1998).

[56] Minard A., Susini P., Misdariis N., Lemaitre G., Mcadams S., Parizet E., *Two-level description of environmental sounds: comparison and generalization of 4 timbre studies*, The Journal of the Acoustical Society of America, **123**, 3253 (2008).

[57] Moore B.C.J., *An introduction to the psychology of hearing*, 3rd edition, Academic Press, London 1989.

[58] Muñoz J., Mendoza E., Fresneda M.D., Carballo G., López P., *Acoustic and perceptual indicators of normal and pathological voice*, Folia Phoniatrica Et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics (IALP), **55**, 102–14 (2003).

[59] Murry T., Singh S., *Multidimensional analysis of male and female voices*, The Journal of the Acoustical Society of America, **68**, 1294–300 (1980).

[60] Necioglu B., Clements M., Barnwell T., Schmidt–Nielsen A., *Perceptual relevance of objectively measured descriptors for speaker characterization*, Proceedings of Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on, 1998.

[61] Nordstrom K.I., Driessen P.F., *Variable pre-emphasis lpc for modeling vocal effort in the singing voice*, Proceedings of 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal, Canada 2006.

[62] O'shaughnessy D., *Speech communication: human and machine*, Addison-Wesley Pub. Co, Reading, Mass 1990.

[63] Parks T., Weisburn B., *Classification of whale and ice sounds with a cochlear model*, Proceedings of Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on, 1992.

[64] Pastiadis C., Papanikolaou G., Printza A., *Application of glottal disturbogram as a novel tool for the description of vocal disturbances*, Hippokratia, **12**, 122–127 (2008).

[65] Pastiadis K., *Contemporary voice analysis techniques for the speech impaired*, Ph.D. thesis, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 2002.

[66] Pols L.C.W., Van Der Kamp L.J.T., Plomp R., *Perceptual and physical space of vowel sounds*, The Journal of the Acoustical Society of America, **46**, 458–467 (1969).

[67] Potočnik P., Govekar E., Grabec I., *Application of psychoacoustic filtering for machine fault detection*, Proceedings of 8th International Conference of the Slovenian Society for Non-Destructive Testing: "Application of Contemporary Non-Destructive Testing in Engineering", Portoroz, Slovenia 2005.

[68] Prante H.U., *Estimation of sound quality measures using fir neural networks*, Acta Acustica united with Acustica, **85**, 674–677 (1999).

[69] Reynolds D.A., *Speaker identification and verification using gaussian mixture speaker models*, Speech Communication, **17**, 91–108 (1995).

[70] Rioux V., Mcadams S., Susini P., Peeters G., *Wp2.1.5 psycho-acoustic timbre descriptors*, Perception et cognition musicales, Analyse et synthèse sonores, IRCAM, 2002.

[71] Risset J., Wessel D.L., *Exploration of timbre by analysis synthesis*, [in:] The Psychology of Music, Academic Press, San Diego 1999.

[72] Rosson M.B., Cecala A.J., *Designing a quality voice: an analysis of listeners' reactions to synthetic voices*, SIGCHI Bull., **17**, 192–197 (1986).

[73] Saenz-lechon N., Godino–Llorente J.I., Osma–Ruiz V., Blanco–Velasco M., Cruz–Roldan F., *Automatic assessment of voice quality according to the grbas scale*, Proceedings of Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE, 2006.

[74] Saenz–Lechon N., Godino–Llorente J.I., Osma–Ruiz V., Gomez–Vilda P., *Methodological issues in the development of automatic systems for voice pathology detection*, Biomedical Signal Processing and Control, **1**, 120–128 (2006).

[75] Sankiewicz M., Budzynski G., *Reflections on sound timbre definitions*, Archives of Acoustics, **32**, 591-602 (2007).

[76] Schlotthauer G., Torres M.E., Jackson–Menaldi C., *Automatic diagnosis of pathological voices*, Lisbon, Portugal 2006.

[77] Schuller B., Rigoll G., Lang M., *Hidden markov model-based speech emotion recognition*, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03), 2003.

[78] Shrivastav R., *Multidimensional scaling of breathy voice quality: individual differences in perception*, Journal of Voice: Official Journal of the Voice Foundation, **20**, 211–22 (2006).

[79] Shrivastav R., Sapienza C.M., *Objective measures of breathy voice quality obtained using an auditory model*, The Journal of the Acoustical Society of America, **114**, 2217–24 (2003).

[80] Susini P., Mcadams S., Winsberg S., *A multidimensional technique for sound quality assessment*, Acta Acustica united with Acustica, **85**, 650–657 (1999).

[81] Susini P., Mcadams S., Winsberg S., Perry I., Vieillard S., Rodet X., *Characterizing the sound quality of air-conditioning noise*, Applied Acoustics, **65**, 763–790 (2004).

[82] Terasawa H., Slaney M., Berger J., *Perceptual distance in timbre space*, Proc. of ICAD, **5**, (2005).

[83] Tsang–Long Pao, Yu–Te Chen, Jun–Heng Yeh, *Emotion recognition from mandarin speech signals*, Proceedings of Chinese Spoken Language Processing, 2004 International Symposium on, 2004.

[84] Tucker S., *An ecological approach to the classification of transient underwater acoustic events: perceptual experiments and auditory models*, Ph.D. thesis, Department of Computer Science, University of Sheffield, 2003.

[85] Turkoglu I., Arslan A., Ilkay E., *An expert system for diagnosis of the heart valve diseases*, Expert Systems with Applications, **23**, 229–236 (2002).

[86] Voiers W.D., *Perceptual bases of speaker identity*, Journal of the Acoustical Society of America, **36**, 1065–1073 (1964).

[87] Wang Y., Lee C., Kim D., Xu Y., *Sound-quality prediction for nonstationary vehicle interior noise based on wavelet pre-processing neural network model*, Journal of Sound and Vibration, **299**, 933–947 (2007).

[88] Webb A.L., Carding P.N., Deary I.J., Mackenzie K., Steen N., Wilson J.A., *The reliability of three perceptual evaluation scales for dysphonia*, European Archives of Oto-Rhino-Laryngology, **261**, 429–434 (2004).

[89] Zotkin D., Shamma S., Ru P., Duraiswami R., Davis L., *Pitch and timbre manipulations using cortical representation of sound*, Proceedings of Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on, 2003.