A GROUND-TRUTH EXPERIMENT ON MELODY GENRE RECOGNITION IN ABSENCE OF TIMBRE

José M. Iñesta, Pedro J. Ponce de León, José L. Heredia-Agoiz
Universidad de Alicante, Spain
Departamento de Lenguajes y Sistemas Informáticos
{inesta,pierre}@dlsi.ua.es

ABSTRACT

Music genre or style is an important metadata for music collections and database organization. Some authors claim for the need of having ground truth studies on this particular topic, in order to compare results with them and lead to sound conclusions when analyzing software performances. When dealing with digital scores in any format, timbrical information is not always available or trustworthy so we have avoided this information in our computer models, using only melodic information. The main goal of this work is to assess the human ability for recognizing music genres in absence of timbre in order to assess comparatively the performance of computer models for this task.

For this, we have experimented with fragments of melodies in absence of accompainment and timbre, as our computer models do. For this particular paper we have worked with two well-stablished genres in the music literature, like classical and jazz music.

A number of analyses in terms of age, group, education, and music studies of the people subjected to the tests have been performed. The results show that, on average, the error rate was about 18%. This value shows the base line to be improved for computer systems in this task without using timbrical information.

I. INTRODUCTION

Music genre or style is an important metadata for music collections and database organization. A number of computer systems have been published that are able in some degree to categorize music data both from audio (Soltau et al., 1998; Tzanetakis & Cook, 2002; Zhu et al., 2004) or scores (Cruz et al., 2003; McKay & Fujinaga, 2004; Pérez-Sancho et al., 2005) in digital formats. Recently, papers have appeared trying to combine the best of both worlds (Lidy et al., 2007). On the other hand, some authors claim for the need of having ground truth studies on this particular topic in order to compare results and lead to sound conclusions when analyzing software performances (Craft et al., 2007; Lippens et al., 2004).

When dealing with digital scores in any format (MIDI, MusicXML,...), timbrical information is not always available or trustworthy because it depends on good sequencing practices. So we have avoided this information for our computer models (Ponce de Leon & Iñesta, 2007), focusing only on the information coded by the notes in the melody. Under these conditions, some important questions arise: is a particular success rate in automatic genre classification good or bad? what is the human ability for recognizing the music genre of a melody just from the notes in the score? what remains of genre when no timbrical information is provided?

Genre classification is of a hierarchical nature, so experiments should be placed in a given level of the hierarchy. It is non sense to classify between the whole classical music domain and a particular subgenre like, for example, hip-hop. On the other hand, genre labels are inherently subjective and influenced by a number of cultural, art, and market trends, therefore perfect results can not be expected (Lippens et al., 2004).

In the cited paper, the authors design a set of experiments for compare the results obtained by automatic computer models and by humans. For that, they utilized fragments of 30 seconds of 160 commercial recordings from classical, dance, pop, rap, rock, and 'other' (none of the previous labels). Those fragments were classified by a number of pattern recognition algorithms using different features extracted from the audio. They were also presented to a set of 27 human listeners that were asked to choose a musical genre out of the 6 possibilities given above. In summary, the results reported a 65% of correct classification by the computer against a 88% for the human listeners. These results show that there is still a gap with human abilities when dealing with the audio data, where all the musical information (melody, harmony, rhythm, timbre, etc.) is present. This is no surprise, since the data were presented in the way humans use to enjoy music, so our abilities to perform this task (in spite of subjectivity and other considerations) have been trained for years and we have a huge background knowledge compared to the training set used by the machine. Thus we are in a clear dominant position when competing against those artificial intelligence models.

The main goal of this work is to compete with a machine model in equal conditions. For this, we have exper-

imented presenting humans the same information available for the computer counterpart: fragments of melodies without accompaniment and timbrical information. This way, we can have a ground-truth reference on the human ability for recognizing music genres in absence of timbre in order to assess comparatively the performance of computer models for this task. For this particular paper we have worked with two well-stablished genres in the music literature, like classical and jazz music.

II. METHOD

A. Subjects

The melodic fragments were presented to 149 subjects (109 male and 40 female) classified into 3 groups: A) professional musicians (performers and professors), B) amateurs (both musicians and music lovers), and C) a control group composed of people with no particular relation to music practice. Table 1 shows the statistics on the subjects to whom the test was applied.

Table 1. Statistical profile of the people subjected to the test.

Group	Number	Male	Female	Age
Profess.	29	18	11	28.3 ± 8.0
Amateurs	57	46	11	27.2 ± 6.3
Control	63	42	21	29.3 ± 6.2

Despite the uneven distribution of people by sex (106 male, 43 female), no bias was detected in the answers according to this variable.

The minumum age was 9 years old and the maximum was 60. The overall average was of 28.1 years with a standard deviation of ± 9.

According to the level of studies, two criteria were adopted: classification under their general studies and their specific music studies. Different categories were established:

General studies (number of subjects):

- 1: Elementary education (6)
- 2: Secondary education (75)
- 3: Graduate studies (12)
- 4: Master (43)
- 5: Doctorate (13)

Music studies (number of subjects):

- 0: No studies (43)
- 1: Self-trained (48)
- 2: Not-finished conservatory (12)
- 3: Conservatory elementary degree (9)
- 4: Conservatory intermediate degree (19)
- 5: Conservatory high degree (8)
- 6: Musicology (10)

B. Melodies

Concerning to the music data, a set of 40 melody fragments (20 of classical music and 20 jazz pieces), were synthesized using sinusoidal waves (just a fundamental frequency without timbrical relation among spectral components). They were cut from the respective MIDI sequences by an expert. The durations were in average 19.4 ± 4.2 seconds in a range [12,30] (33 ± 32 [12,62] beats, 8.4 ± 8.0 [3,16] bars). In terms of number of notes, the range was [17,171] averaging 46.

The classical fragments covered a wide range of periods from Baroque (Haendel, Bach, Vivaldi,...) to Classical (Mozart, Paganini, Beethoven,...) and Romantic (Schumann, Schubert, Mendelssohn, Brahms,...). Jazz fragments were standards from a variety of styles like Pre-Bop, Bop, Bossa-nova, or fusion (Charlie Parker, Thelonious Monk, Antonio Carlos Jobim, Wayne Shorter,...).

All the fragments were pre-classified by an expert according to their *a priori* difficulty for being classified. For that, melodic, harmonic, and rhythmic aspects of the melodies were taken into account. Also their general public popularity was considered for assigning a difficulty degree for each fragment. For jazz, 5 were considered 'easy', 8 'intermediate', and 7 'difficult', and for classical, it was 11, 6, and 3 respectively.

When presented to the subject (just once) he or she must identify whether the melody belongs to a classical or jazz piece. The fragments were randomly ordered for presentation, using always the same ordering.

III. RESULTS

A number of analyses in terms of age, group, education, and music studies have been performed. Also, the difficulty level of the fragments, according to the *a priori* classification explained above, have been taken into account. The results (see Table 2) show that, on average, the error rate was 16.2%, although it ranged from 5.9% for the professionals to 19.2% for the control group. Note that there were no significant differences between amateurs and control. Only professional musicians performed much better than the other groups, showing much higher classification skills.

Table 2. Error percentages in terms of group of people.

	control	amateurs	professionals
Error %	19.2	18.0	5.9

The *a priori* difficulty of the fragments was clearly reflected in the ability for recognizing the genre (see Table 3) increasing from 3.5% for the easy ones to 23.2% for those considered difficult. Note that the error rate for difficult fragments is more than twice that for the intermediate ones (10.8%).

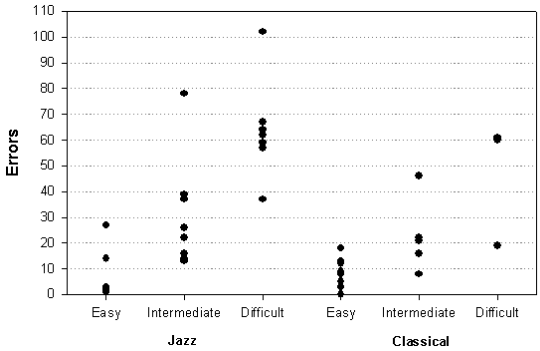

Figure 1. Number of errors as a function of the difficulty of the fragments.

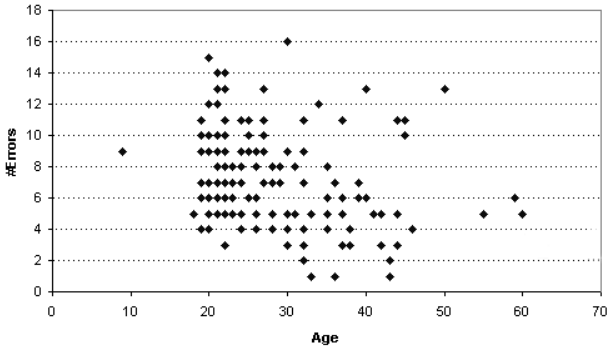

Figure 2. Number of errors as a function of the age.

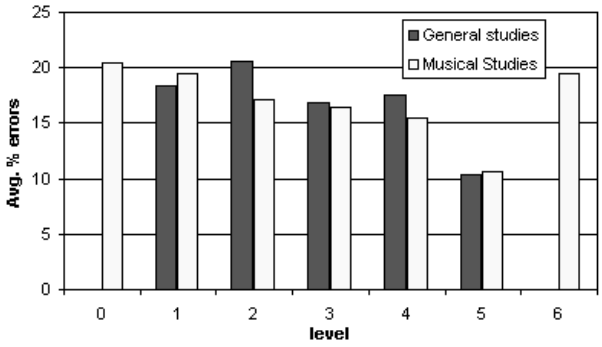

Figure 3. Number of errors as a function of both the general and musical level of studies.

This fact can also be seen in the distribution of errors for fragments (see figure 1). All subjects gave the correct answer (zero errors) for Haendel's "Fireworks - La Rejouissance" (an 'easy' fragment). For Jazz, just one error was committed for Telonious Monk's "Well You Needn't" (an 'easy fragment too). In contrast, the maximum number of errors (120, a 68.5% of the total number of tests) were made for "Young and Foolish" by Horwitt and Hague, while in classical music Schubert's Symphony no. 4 in C minor "Tragic" received 61 misclassifications (40.9% of the tests).

The number of answers that classified the fragments as classical was 56.4% (43.6% for jazz). This bias is due to the fact that, in general, people is more familiar with classical tunes and tend to think that an unknown fragment is classical only because they are usually more exposed to this genre.

Table 3. Error percentages in terms of the difficulty levels assigned.

(%)	easy	interm.	difficult
Jazz	3.5	14.3	27.8
Classical	3.5	9.3	21.0
Average	3.5	10.8	23.2

A negative correlation with age and general studies ($r = -0.28$) was observed (see figure 2). This suggests that people's experience is important in this ability. This is not surprising because through their lives people hear music and, even if they are not experts, they accumulate arguments in order to decide which kind of music they are hearing.

The evolution of the error as a function of study levels is also an important issue (see figure 3). Note that the error percentages are lower for higher levels of general studies (dark columns in the graph). More interesting and significant is to see what happens for different music studies (light columns). The higher the music studies the lower the error rate, and this tendency is neat in the graph. But there is the remarkable exception of musicologists, that performed clearly poorer than the average, showing a professional bias. We can consider (speaking in terms of a classification system) that they are 'overtrained'. Their high level of musical knowledge leads them to match fragments with some theme in their knowledge.

Most of the professionals were involved in works related to classical music, so they bias their decisions in this direction. Control subjects answered classical in 55% of the queries, 56% for amateurs, while professionals did it 58% of times.

IV. CONCLUSIONS

The results show that people are able to distinguish quite well between classical and jazz melodies when a timbre-less fragment is presented (roughly 4 out of 5 fragments were correctly classified). In other, less structured, experiments no one have had difficulties when these same fragments were presented with the timbrical information (utilizing a synthesizer and the whole MIDI sequences). This suggests that something about music genre remains in just the melody notes without timbre, at least between well stablished genres like classical and jazz music. This is more doubtful if we need to distinguish between closer genres where timbre is a key feature, like for example pop and rock.

For use in computer music information retrieval experiments as a base line, a 16% of error shows the performance level to be improved if authors claim to report good

results for this task without using timbrical information.

ACKNOWLEDGMENT

This work is supported by the Spanish PROSEMUS project (TIN2006-14932-C02) an the research programme Consolider Ingenio 2010 (MIPRCV, CSD2007-00018).

REFERENCES

Craft, A. J. D., Wiggins, G. A., & Crawford, T. (2007). How many beans make five? the consensus problem in music-genre classification and a new evaluation method for single-genre categorisation systems. *Proceedings of the Int. Conf. on Music Information retrieval, ISMIR 2007* (pp. 73–76). Vienna, Austria.

Cruz, P. P., Vidal, E., & Pérez-Cortes, J. C. (2003). Musical style identification using grammatical inference: The encoding problem. *Proceedings of CIARP 2003* (pp. 375–382). La Habana, Cuba.

Lidy, T., Rauber, A., Pertusa, A., & Iñesta, J. (2007). Improving genre classification by combination of audio and symbolic descriptors using a transcription system. *Proceedings of the ISMIR* (pp. 61–66). Vienna, Austria.

Lippens, S., Martens, J., Leman, M., Baets, B., Meyer, H., & Tzanetakis, G. (2004). A comparison of human and automatic musical genre classification. *Proceedings of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP 2004* (pp. 233–236).

McKay, C., & Fujinaga, I. (2004). Automatic genre classification using large high-level musical feature sets. *Int. Conf. on Music Information Retrieval, ISMIR 2004* (pp. 525–530).

Pérez-Sancho, C., Iñesta, J. M., & Calera-Rubio, J. (2005). Style recognition through statistical event models. *Journal of New Music Research, 34*, 331–340.

Ponce de Leon, P. J., & Iñesta, J. M. (2007). A pattern recognition approach for music style identification using shallow statistical descriptors. *IEEE Transactions on Systems Man and Cybernetics C, 37*, 248–257.

Soltau, H., Schultz, T., Westphal, M., & Waibel, A. (1998). Recognition of music types. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-1998)* (pp. 1137–1140).

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10*, 205–210.

Zhu, J., Xue, X., & Lu, H. (2004). Musical genre classification by instrumental features. *Int. Computer Music Conference, ICMC 2004* (pp. 580–583).