How Musical Instrumentation Affects Perceptual Identification of Musical Genres

by Sofia Brené <sbrene@kth.se> and Carl Thomé <cthome@kth.se>
Abstract

A listening experiment was conducted to investigate which musical instruments are the most important for defining certain musical genres. 66 participants genre classified a series of audio samples, with the same songs recurring both with full instrumentation and partial instrumentation. The report used the collected genre classifications to clarify the relationship between certain musical genres and song instrumentation. A numerical analysis of the classifications, in the context of genre traditions and conventions, show that certain traditions hold true, while others do not. The most and least defining instrumentation for each genre was determined and discussed.
Sammanfattning

Contents

Statement of Collaboration ................................................................. 6

Introduction ......................................................................................... 7
  Problem Statement ........................................................................... 8

Background ........................................................................................... 9
  Previous Research ............................................................................. 9

Musical Characteristics of Genres ......................................................... 10
  Blues ................................................................................................. 10
  Classical ............................................................................................ 10
  Country ............................................................................................. 10
  Electronic .......................................................................................... 11
  Jazz .................................................................................................... 11
  Metal .................................................................................................. 11
  Pop/Rock ........................................................................................... 11
  Rap ..................................................................................................... 11
  Reggae .............................................................................................. 11

Method ................................................................................................. 12
  Designing the Listening Experiment ................................................. 12
  Data Collection and Statistical Analysis .......................................... 16
    Constructing Genre Classifications ................................................ 17
    Determining the Most Defining Instrumentation per Genre ............ 17
    Determining the Listeners' Genre Classification Certainty .............. 18

Results ................................................................................................. 19
  Listening Experiment Demographic ............................................... 19
  Genre Classification Diagrams ......................................................... 19
  The Most and Least Defining Instrumentation per Genre ................. 28
  Listeners' Genre Classification Certainty .......................................... 30

Discussion .......................................................................................... 31
  How Genre Classification Relates to Musical Instrumentation .......... 31
  The Genre Concept ........................................................................... 33
  Experiment Conditions and Possible Sources of Error ..................... 33
    Environmental Conditions ............................................................. 33
Statement of Collaboration

- Sofia Brené wrote the literary comparison in the background section and provided references in the report.
- Carl Thomé built the web-based listening experiment, analyzed data, constructed result diagrams and tables, and wrote the introduction, method, results, analysis of the results in the discussion and the conclusion.
- Data collecting and writing the discussion about the experiment conditions was shared equally.
Introduction

This section provides a historical context for the report and declares the problem statement.

As technological advances during the 20th century made it possible to store musical performances in various types of data formats such as vinyl discs, magnetic tape and the more recent digital audio formats, music rapidly became an integral part of everyday life in the modern world.

The increased availability increased both music consumption and music production, and never before has there been as many recording artists as there are today. The huge influx of available music has made the importance of being able to selectively filter and search through music collections all the more important. Music recommendation services stating “If you like this artist you might also like…” or “What’s your music listening mood?” in order to help humans navigate an increasingly crowded music domain have become commonplace, and the scientific field these tools rely upon is called Music Information Retrieval (MIR).

MIR is about using audio features and meta information in order to make predictions about different musical aspects. It can be related to both high-level descriptions such as predicting genre, music similarity, musical moods, as well as more specific things like melodic recognition and retrieval, or tempo estimation. Advanced signal processing methods are often used for computing audio features. For mapping features to descriptions machine learning methods or statistical inference methods are commonly used.

A key notion of MIR is to automate the tasks that have traditionally been performed by humans, such as A&R\(^1\) divisions signing new artists in trending music genres, or radio program directors targeting a niche of listeners by playing songs with a shared musical context. In order to achieve the same functionality by programmatically analyzing audio features there has to be some measure of success, and as music is an art form it is often thought of as being subjective and up to individual interpretation. This poses a problem, and even though there are hard metrics for audio, it is not as obvious when it comes to describing music’s emotional content.

\(^1\) A&R - Artists and repertoire is the division of a record label or music publishing company that is responsible for talent scouting and overseeing the artistic development of recording artists and/or songwriters.
One of the most long withstanding and well-known tools for describing music has been the genre classification concept. The idea that music can be sorted into groups based on a range of different qualities, such as musical theory, historical or geographical proximity between music artists, mood similarity and so on. The genre concept is a tool for identifying pieces of music as belonging to a shared tradition or set of conventions, but there are no strict rules as to what the set of conventions might entail. This makes MIR difficult because there are no obvious mappings between audio features and music genres.

In order for MIR technology to advance it is therefore important to be able to describe what constitutes a music genre - a very broad and difficult question to answer. A fraction of this question is to ask which musical instruments are the most important for listeners when genre classifying music, by comparing how humans classify songs when they hear the full instrumentation of a song, or just a partial instrumentation with soloed tracks.

For Metal music it is possible that a blaring drum kit is the most important instrument, while for Jazz music the brass instruments, and the complexity with which they are played, might be more important. Perhaps for Pop/Rock music a vocal track with a strong melodic hook is the key defining property. This report attempts to clarify these relationships between the musical instruments and the genres with a listening experiment. The listening experiment consisted of human participants classifying song samples into genres by rating a series of audio samples, with the same songs occurring both fully instrumented and partially instrumented. Finally the resulting genre classification ratings were compared. The difference in genre classification between the full mix rating and the soloed instruments serves as a basis for discussing which musical instruments seem to define a particular genre most/least.

**Problem Statement**

In order to clarify which musical instruments are the most and least defining for certain musical genres this report has investigated if songs are classified as the same genres when listeners hear the fully instrumented song mix, or just partially instrumented submixes of the same songs.

---

2 hook - a short phrase used in popular music to make the song appealing. The hook is often found in the chorus.
Background

An overview of previous research experiments with similar problem statements follow. Also, because knowledge of the genres is necessary to appreciate the report results, a quick walkthrough of each genre’s musical characteristics is presented here.

Previous Research

There have been several studies trying to define musical genres. Since it is a very hard, almost impossible task defining a genre people doing research in this area have done different studies in getting closer to the actual answer to this question.

The determination of musical genres is in fact a non-trivial question and interdisciplinary studies are therefore investigated in previous researches. There have been other attempts figuring this out, as defining a genre only by hearing the vocals or just the unpitched percussion instruments.

A survey by N. Scaringella and G.Zoia [1] reviewed typical extraction techniques used in music information retrieval for different music elements as timbre, melody/harmony and rhythm. The conclusion of their experiment resulted in finding that the investigation of categorizing music is evolving from purely objective machine calculations to techniques where preliminary knowledge and learning phases etc. plays a very significant role in the performance and results.

Another similar study is made by G.Tzanetakis and P.Cook [2] where they believe automatic classification of musical genres can replace the importance of human users in this process of musical genre annotation and would bring a valued addition to musical information retrieval systems.

By implementing two graphical user interfaces browsing as well interacting with audio collection the automatic hierarchical genre classification has been developed.

Kosinas [3] paper is an overview of music genre classification where signal processing, pattern classifications and disquisitions from areas as human sound perception are treated. She also presents her development MUGRAT, which is a prototype system for the
recognition of musical genres. This system is using a subset of the features proposed by G.Tzanetakis and P.Cook.

The system extracts an amount of features from the given sound which is also important in the human music genre recognition and can be distinguished in two categories: features related to the musical texture and features related to the rhythm/belatedness of the sound.

There are many studies and methods related to the analysis of music audio signals and it is important to keep developing modules for content-based music information retrieval systems since it will facilitate music genre classification.

Even if the music genre is a somewhat ambiguous descriptor it is still very widely used to categorize large collections of digital music [8][9][11].

Musical Characteristics of Genres

Blues
Marked by the frequent occurrence of blue notes, and a basic form of a 12-bar chorus consisting of a 3-line stanza with the second line repeating the first. Percussion usually plays a shuffle rhythm. [8]

Classical
Loosely defined as what popular music is not - characterized by the use of orchestra instruments (violins, oboes, timpani, etc.), opera singing and a lack of the verse/chorus/bridge form commonly used in popular music. [9]

Country
Simple in form and harmony, accompanied by (usually) vibrato-free vocals, acoustic or electric guitar, banjo, violin, and harmonica. [8]

---

blue note - a note sung or played at a slightly lower pitch than that of the major scale, for expressive purposes.
12-bar - A bar is a way of dividing beats in music, and blues songs are structured in a 12-bar format.
stanza - a grouped set of lines
Electronic
Often features an overly beat quantized rhythm (restricted by a 16-note grid within the composing machine) and synthesized melodic sounds generated with oscillators. [10]

Jazz
Complex styles, generally marked by intricate, propulsive rhythms, polyphonic ensemble playing, improvisatory, virtuosic solos, melodic freedom, and a harmonic idiom ranging from simple diatonicism through chromaticism to atonality. [8]

Metal
Loud and harsh sounding rock music with a straight beat, heavily distorted electric guitars and growl/scream singing techniques. [8]

Pop/Rock
A blend of rhythm-and-blues and country-and-western focusing on harmonized vocal melodies and repeating choruses, usually accompanied by electric guitars, an electric bass guitar and a western drum kit. [8]

Rap
An insistent, recurring beat pattern provides the background and counterpoint for a rapid, slangy, and often-boastful rhyming pattern intoned by one or several vocalists. [8]

Reggae
Blends blues, calypso and rock, characterized by a strong syncopated rhythm called the skank, an offbeat staccato rhythm usually played on an electric guitar. Also, the percussion often plays triplet ghost notes. [8]

---

6ghost note - a musical note with a rhythmic value, but no discernible pitch when played.
Method

A description on how the relationship between instrumentation and genre classification was investigated follows.

In order to clarify which musical instruments are the most important when humans classify songs into genres a listening experiment was conducted. Steps taken:

1. Designed a survey in the form of a web-based listening experience.
2. Let listeners genre classify audio samples in the web-based listening experience.
3. Performed statistical analysis on the collected data and constructed result diagrams and tables.

Designing the Listening Experiment

The web-based listening experiment was constructed in HTML5/PHP/CSV and designed iteratively in an agile process with user testing. User feedback was collected and design improvements were implemented accordingly. Refer to Appendix 1 for design impacting quotes from the usability testing.

The listening experiment consisted of a series of audio samples that the listeners rated (figure A) with a set of musical genres, with a low value indicating the listener did not believe the sample to be part of that genre, and a high value meaning the listener believed the audio sample to be part of that genre.
Figure A - Screenshot of the web-based listening experiment. The stepless sliders were designed to be an intuitive way for participants to genre classify audio samples.

There were nine genres in the experiment. [5] Two songs were chosen per genre to minimize errors from atypical song selections. All audio data were provided by a karaoke song database [6] (figure B) that allowed muting of individual instruments so that source separation would not have to be performed which otherwise might have introduced a possible measurement error in the experiment.
Each of the eighteen songs (appendix 4) were sliced into ten second samples with the audio software REAPER [7] (figure C), and further divided into four separate audio samples by soloing instruments on the song provider website and creating specific submixes. Again, no source separation had to be performed as the song provider offered master tracks. The four submixes were:

1. The full mix instrumentation.
2. Soloed vocal tracks (including any background vocals).
3. Soloed pitched instruments (ex: piano, guitar, organ, violin, etc.).
4. Soloed unpitched percussion instruments (ex: drum kits, timpani, side beats, sound effects, etc.).
Figure C - The audio software (REAPER) used to slice the songs into ten-second audio samples. Note that all songs used in the experiment were provided as master tracks so no post-process audio separation had to be performed (i.e. no source separation problems were present in the experiment).

User testing found that a total of 72 ten-second audio samples made the listening experiment too tedious so the test size was reduced to a fourth of the original length, by randomly selecting 18 audio samples from the full audio sample set instead. The number of participants was quadrupled accordingly, to 66 respondents. In short, each participant listened to a random selection of 18 audio samples out of the 72. The only constraint made to the shuffle ordering was that the full mix of a song should not directly precede a submix of the same song, as usability tests found such playlists to be confusing.

The shuffle ordering was implemented in PHP, which was also used as a session handler during the experiment (figure D). The responses provided by users in the HTML5 frontend were outputted as a CSV file, for the statistical analysis stage. An example of a user’s
responses as a CSV file is available in Appendix 2, along with the source code for the web-based listening experiment (provided in Appendix 3).

**Figure D** - The web-based listening experience was written in HTML5 with PHP as a session handler, outputting user scores as CSV files.

**Data Collection and Statistical Analysis**

The web-based listening experiment was conducted at various locations. Participants could partake in the experiment at whichever location they preferred. No particular target demographic was sought, and instead the listeners were asked to rank their personal knowledge and familiarity with each musical genre, before rating the audio samples. The genre familiarity numbers were used to scale the audio sample ratings.
The listeners’ answers were combined in a spreadsheet with which:

1. Genre classification diagrams were constructed.
2. The distance between the full mix genre classification and the submixes’ genre classifications were calculated in order to determine which instrumentation is the most defining per genre.
3. The listeners’ genre slider usage was analyzed in order to determine listeners’ certainty about genre classifying the audio samples.

**Constructing Genre Classifications**
For each pair of songs selected for a particular genre, for each genre, for each submix (including the full mix) the average score for all listeners was calculated, with each individual listener’s score weighed with that particular listener's self-rated genre familiarity. Each genre slider for rating the audio samples had a value range between 0-100 (inclusive). The genre familiarity sliders used the same value range and sliders.

**Determining the Most Defining Instrumentation per Genre**
With each genre as a dimension in a 9-dimensional vector space, the difference between the full mix’s genre classification and each submix's genre classification was calculated as the Euclidean distance (figure a) between the points in the vector space.

\[
d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.
\]

*Figure a - Euclidean distance for a n-dimensional vector space.*

Consider the full mix genre classification as the ground truth. Then a short distance from the full mix to a submix would imply that the submix’s instrumentation is of particular importance for defining the genre, while a long distance between the full mix and a submix implies that the instrumentation in the submix is of less importance when genre classifying the samples. This is under the natural assumption that the full mix genre classification is the most indicative of which genre a song belongs to.
Determining the Listeners’ Genre Classification Certainty

In order to measure how certain the listeners were when genre classifying the audio samples the $L^2$ Norm (figure b) was calculated for the genre classification averages (for each submix and the full mix).

$$\| x \| := \sqrt{x_1^2 + \cdots + x_n^2}.$$

*Figure b - the $L^2$ norm for a real-valued vector $x$ in a $n$-dimensional vector space.*

The resulting values serve as a measure of how well the songs were selected for each genre, as well as a measure of how easy it is to genre classify certain instruments per genre. The input vector was normalized (i.e. it sums to one) so the $L^2$ value goes from 0 to 1, and will be close to 1 only if the vector was far away from origo in a particular dimension. Translated in the context of the listening experiment this means that large $L^2$-Norm values indicate that users were certain about which genre an audio sample should be classified as, while a low $L^2$-Norm value indicate the listener being uncertain and that several, all, or no genre sliders were used when rating the sample.
Results

This section goes through the report results, including the demographic of its participants, their average genre ratings of the audio samples, formatted as genre classification diagrams, the distances between the full mix and the submixes, and the genre classification certainty.

Listening Experiment Demographic

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of listeners</td>
<td>66</td>
</tr>
<tr>
<td>Number of male listeners</td>
<td>43</td>
</tr>
<tr>
<td>Number of female listeners</td>
<td>23</td>
</tr>
<tr>
<td>Number of other listeners</td>
<td>0</td>
</tr>
<tr>
<td>Average age of listeners [years]</td>
<td>24</td>
</tr>
<tr>
<td>Listener age standard deviation [years]</td>
<td>7</td>
</tr>
<tr>
<td>Oldest listener [years]</td>
<td>56</td>
</tr>
<tr>
<td>Youngest listener [years]</td>
<td>14</td>
</tr>
</tbody>
</table>

Genre Classification Diagrams

The first diagram includes participant scores for the full set of songs, while the succeeding nine diagrams are genre specific and only count scores for the pair of songs corresponding to the genre.
The participants generally favored the *Pop/Rock* and *Electronic* sliders when classifying the audio samples (Figure 1), and felt most sure about choosing high values for those two genres throughout rating all the audio samples. However, the participants used all genre sliders. There is no obvious difference between any of the submixes or the full mix, but when filtering out songs the result becomes clearer (Figure 2-10).
Figure 2 - genre classification diagram for the Blues songs.

(Figure 2) For the two Blues songs the results indicate that while it is easy to classify music as Blues when all instruments play together, it gets more difficult to distinguish between Blues and Pop/Rock if there are no percussion instruments.

Figure 3 - genre classification diagram for the Classical songs.

(Figure 3) Classical music was consistently classified correctly by the participants, with the greatest genre ambiguity showing up for soloed percussion instruments.
(Figure 4) It appears as though Country music is heavily defined by its vocals and the pitched instruments, with the percussion instruments being close to impossible to classify as Country by the participants.
By removing the percussion from the song, and only leaving the vocals, the samples were perceived to be closer to Pop/Rock than Electronic music. As soon as the percussion or the pitched instruments were audible the two songs were much more easily classified as Electronic music by the participants.

(Figure 5) Figure 5 - genre classification diagram for the Electronic songs.
Figure 7 - genre classification diagram for the Jazz songs.

(Figure 6) By muting the pitched instruments the genre classification became more unclear and the results show participants confusing Jazz with both Blues and Pop/Rock music.
Figure 7 - genre classification diagram for the Metal songs.

(Figure 7) The participants easily classified the two songs as Metal when they could hear the pitched instruments, or the vocals, but when only the percussion was audible the participants believed the songs to be Pop/Rock songs.

Figure 8 - genre classification diagram for the Pop/Rock songs.

(Figure 8) Pop/Rock was easily classified by all participants.
Figure 9 - genre classification diagram for the Rap songs.

(Figure 9) The most definitely scored submix for Rap was the soloed vocal group, and the full mix.
(Figure 10) The soloed percussion mix was perceived as *Electronic* music to the same extent as *Reggae*, while the vocals and pitched submixes were easily classified as *Reggae*. 
The Most and Least Defining Instrumentation per Genre

(Table A) *Country* percussion was shown to be classified very differently from the full mix. *Reggae* and *Pop/Rock* vocals were shown to be classified very similarly to the full mix. Overall for all audio samples the vocals were the most important instrumentation, but only slightly more so than the pitched instruments and the unpitched percussion instruments. The percussion instrument genre classifications are the most different from the full mix with an average value for all genre classifications of 0.44 (almost double compared to the pitched instruments, and more than double compared to the vocals).

<table>
<thead>
<tr>
<th></th>
<th>Pitched</th>
<th>Percussion</th>
<th>Vocals</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.13</td>
<td>0.13</td>
<td>0.09</td>
</tr>
<tr>
<td>Electronic</td>
<td>0.44</td>
<td>0.41</td>
<td>0.19</td>
</tr>
<tr>
<td>Reggae</td>
<td>0.25</td>
<td>0.56</td>
<td>0.04</td>
</tr>
<tr>
<td>Rap</td>
<td>0.58</td>
<td>0.65</td>
<td>0.11</td>
</tr>
<tr>
<td>Country</td>
<td>0.11</td>
<td>1.03</td>
<td>0.23</td>
</tr>
<tr>
<td>Jazz</td>
<td>0.12</td>
<td>0.30</td>
<td>0.45</td>
</tr>
<tr>
<td>Pop-rock</td>
<td>0.14</td>
<td>0.21</td>
<td>0.05</td>
</tr>
<tr>
<td>Blues</td>
<td>0.32</td>
<td>0.17</td>
<td>0.32</td>
</tr>
<tr>
<td>Classical</td>
<td>0.11</td>
<td>0.44</td>
<td>0.14</td>
</tr>
<tr>
<td>Metal</td>
<td>0.16</td>
<td>0.51</td>
<td>0.12</td>
</tr>
<tr>
<td>Average</td>
<td>0.24</td>
<td>0.44</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table A - the Euclidean distance between the full mix and each sub mix, for different sets of audio samples filtered by genre. Low values are similar to the full mix. High values are different from the full mix. The last row displays the average per column. Values below 0.1 are highlighted in green, values above 0.9 are highlighted in red.
<table>
<thead>
<tr>
<th>Genre</th>
<th>Most defining</th>
<th>Least defining</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>Vocals</td>
<td>Pitched, Percussion</td>
</tr>
<tr>
<td>Blues</td>
<td>Percussion</td>
<td>Pitched, Vocals</td>
</tr>
<tr>
<td>Classical</td>
<td>Pitched</td>
<td>Percussion</td>
</tr>
<tr>
<td>Country</td>
<td>Pitched</td>
<td>Percussion</td>
</tr>
<tr>
<td>Electronic</td>
<td>Percussion</td>
<td>Vocals</td>
</tr>
<tr>
<td>Jazz</td>
<td>Pitched</td>
<td>Percussion</td>
</tr>
<tr>
<td>Metal</td>
<td>Vocals</td>
<td>Percussion</td>
</tr>
<tr>
<td>Pop/Rock</td>
<td>Vocals</td>
<td>Percussion</td>
</tr>
<tr>
<td>Rap</td>
<td>Vocals</td>
<td>Percussion</td>
</tr>
<tr>
<td>Reggae</td>
<td>Vocals</td>
<td>Percussion</td>
</tr>
</tbody>
</table>

*Table B - the most and least defining instruments per genre (i.e. the maximum/minimum Euclidean distance from table A).*
Listeners’ Genre Classification Certainty

(Table C) *Rap* vocals, pitched *Country* instruments, and *Pop/Rock* vocals were all the most easily and least-ambiguous genre classified audio samples in the experiment. Overall, all values are fairly large, with the smallest value being pitched instruments for all genres at 0.39. This is indicative of a high degree of listener genre classification certainty. Put simply: listeners’ genre classified the audio samples using fairly few sliders with fairly large values.

The most ambiguous instrumentation for listeners on average, for all genre classifications, was the percussion instruments. The full mix was shown to be the least ambiguous, while pitched instruments and vocals lie in between the full mix and the percussion.

<table>
<thead>
<tr>
<th></th>
<th>Pitched</th>
<th>Percussion</th>
<th>Vocals</th>
<th>Full Mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.39</td>
<td>0.44</td>
<td>0.44</td>
<td>0.41</td>
</tr>
<tr>
<td>Electronic</td>
<td>0.84</td>
<td>0.81</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>Reggae</td>
<td>0.61</td>
<td>0.57</td>
<td>0.77</td>
<td>0.78</td>
</tr>
<tr>
<td>Rap</td>
<td>0.49</td>
<td>0.57</td>
<td>0.95</td>
<td>0.87</td>
</tr>
<tr>
<td>Country</td>
<td>0.90</td>
<td>0.56</td>
<td>0.81</td>
<td>0.99</td>
</tr>
<tr>
<td>Jazz</td>
<td>0.81</td>
<td>0.55</td>
<td>0.51</td>
<td>0.73</td>
</tr>
<tr>
<td>Pop-rock</td>
<td>0.88</td>
<td>0.80</td>
<td>0.93</td>
<td>0.97</td>
</tr>
<tr>
<td>Blues</td>
<td>0.56</td>
<td>0.63</td>
<td>0.59</td>
<td>0.66</td>
</tr>
<tr>
<td>Classical</td>
<td>0.89</td>
<td>0.56</td>
<td>0.84</td>
<td>0.92</td>
</tr>
<tr>
<td>Metal</td>
<td>0.71</td>
<td>0.68</td>
<td>0.65</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>0.71</strong></td>
<td><strong>0.62</strong></td>
<td><strong>0.72</strong></td>
<td><strong>0.78</strong></td>
</tr>
</tbody>
</table>

Table C - the $L^2$-Norm values for each instrumentation mix, for different sets of audio samples filtered by genre. High values indicate listener certainty, low values indicate genre classification ambiguity. The last row displays the average value per column. Values above 0.9 are highlighted in green, values below 0.1 are highlighted in red.
Discussion

Numerical results are discussed in the context of the musical characteristics presented in the report background. Also, experiment conditions and sources of error are debated.

How Genre Classification Relates to Musical Instrumentation

The average $L^2$-Norm value for the full mix was the largest average for all mixes (Table C), which reinforces the natural assumption (Determining the Most Defining Instrumentation per Genre) that using the full mix as ground truth when determining how instrumentation defines genres is a sensible approach, and the results indicate that certain musical traditions are alive and well.

For example, the genre classification for Metal (Figure 7) seem to correspond with the tradition in the genre of using heavily distorted guitar sounds and specific vocal techniques such as growl singing. Likewise, for Country music the results are that if the pitched instruments are audible (Figure 4) the classification became less ambiguous, and considering the genre’s use of plucked instruments (banjo, steel-string guitar, mandolin, etc.), this makes sense.

Furthermore it appears that the Reggae skank was an important audio feature for listeners to perceive songs as Reggae. Listeners easily classified the two Reggae songs correctly only when they could hear that feature (Figure 10). Another preconception that appears true is that Rap music is almost entirely defined by its vocal style of rapping. The vocal submix, as well as the full mix (which of course includes the rapping vocals) was by far the easiest for listeners to classify (Figure 9). Considering the Rap genre does not really define any traditional instrumentation apart from the vocals it makes sense that the pitched and unpitched soloed instrumentation were hard for participants to classify correctly. Surely there are subgenres within Rap that often use the samesynths and drum machines, like the Roland TR-808\(^7\), but not to an enough extent for it to show up in the experiment results. The two Rap songs used were fairly traditional with vinyl scratching noises for example, but the songs also featured funk-style electric guitar, which might have confused the listeners.

---

\(^7\)Roland TR-808 - one of the first programmable drum machines.
Moving on, the Electronic genre has a tradition of applying vocal effects (such as the vocoder\(^8\)) to vocals, but far from all Electronic hit songs use such effects, and perhaps even more common is the tradition of employing a fairly standard Pop/Rock vocal on top of the Electronic instrumentation. This tradition corresponds with the participants rating of the audio samples. (Figure 5). Most listeners believed the Electronic songs to be Pop/Rock songs when the vocals were soloed!

For Blues it seems to have been important to have an audible shuffle rhythm, but the listeners still perceived each submix as Blues to some extent (Figure 2), and the same goes for Classical music for which the listeners identified the songs as Classical fairly easily but were unsure when they only heard the timpani accompaniment (Figure 3). However slight of an insecurity, it is an expected result considering percussion instruments are sparsely used in Classical pieces, with many famous compositions not even featuring percussion at all.

When it comes to Jazz music, the listening experiment shows that a key property is the pitched instruments (Figure 6) - an expected result considering the use of brass instruments in the genre such as saxophones and trumpets. The brass instruments often play the lead melody in complicated modal scales and when the listener could hear the melodic component of the song they were certain it was Jazz, but not otherwise.

In fact, throughout all these results there seems to be a trend that that percussion instruments are the most difficult for listeners to classify. Both the L\(^2\)-Norm values and the Euclidean distances between submixes and the full mix consistently show that percussion is the most uncertain classification (Table C), and the furthest from the full mix (Table A).

An underlying reason might be that the selected genres in the experiment are not characterized by rhythm to the same extent as harmonies, scales, lyrical content, and so on. However Blues is heavily characterized by a shuffle rhythm, Pop/Rock by a straight 4/4-drum beat, and Electronic music by synthesized drums. Still, most of the experiment’s genres are focused on vocals and pitched instruments (Musical Characteristics), for example:

- Classical’s operatic vocals;
- Country’s melodramatic lyrics;
- Metal’s raspy and/or growl singing, and:
- Jazz’s virtuosic saxophone playing;
- Metal’s distorted guitar sounds;
- Classical’s violin harmonies;
- Reggae’s skank guitar.

---

\(^8\)vocoder - an analysis/synthesis system, used to reproduce or manipulate vocals.
Still, the experiment results seem to indicate that rhythm is not as important for defining genres as the melodic and lyrical components. Considering several of the genres in the experiment tend to use similar rhythms, usually following a 16-note grid and playing a fairly simplistic 4/4 or ¾ beat throughout the entire song, it is not an unsurprising result.

**The Genre Concept**

Overall the genre concept is flawed, considering songs might be a mix of several genres, the possibility for extending the genre domain infinitely by adding sub genres, and how genres transform over time according to the most up-to-date and popular recording artist (a classic example being the rapid shift in rock music in the early 90’s when the popularity of Nirvana and grunge sent glam bands to the background of rock⁹). It is however a traditional music classification tool that is still used today and the report results strengthen the idea that there is actual merit to genre classifying songs. Considering how the genres’ musical characteristics and the experiment participants’ perception correlate quite often the genre concept is still a useful tool.

**Experiment Conditions and Possible Sources of Error**

**Environmental Conditions**

All participants did the experiment at a location of their own choosing, which means that their listening devices were of varying audio fidelity. Classifying the samples when listening with high-quality loudspeakers vs. low-quality headphones might impact the listeners’ perception. Also, since the experiment was web-based, internet connectivity could have been an issue. If a listener had to wait for audio samples to buffer perhaps their attentiveness would have been reduced. Inviting listeners to a controlled lab environment might therefore have been preferable.

**Demographic**

According to the age and gender distribution the majority of the experiment participants were males, 25 years of age. This might have impacted the results, and even though the survey did not have a particular target audience it is still important to be aware that the experiment might turn out differently depending on the listeners’ background, as music

listening is possibly an inherited and taught discipline rather than something humans are born with. More studies, targeting specific demographics, could be conducted.

**Song Selection**

The represented songs play a significant role in the experiment’s outcome so a careful selection process was used. Overall, the chosen songs are likely typical for each genre, in accordance with each genre’s musical characteristics, as the songs were selected from the song provider by genre. Also, to safeguard against poor song choices, or the song provider having incorrect genre metadata, more than one song were selected for each genre, and an average result for both songs was used in the results. It would probably have been even better to include more songs per genre, but the listening experiment had time constraints and participant attentiveness was deemed more important.

Another factor to consider is whether the listeners were already familiar with the songs or not, and whether that affected how they scored the submixes. If they could remember the original song mix, perhaps they subconsciously included the missing instruments when genre classifying a submix sample. All songs from the song provider were karaoke versions of famous hit songs, selected due to availability even though more obscure songs might have been preferable. However because the songs were all imperfect karaoke replicas of the original recordings perhaps the likelihood of listeners instantly identifying a song when only hearing a submix was somewhat alleviated.

One of the poorer song selections in the experiment was the fact that Reggae drums were incorrectly classified as Electronic drums (Figure 10). Reggae is after all heavily characterized by triplets and ghost notes on the snare’s rim and tom-toms, but unfortunately one of the selected Reggae songs featured a low-quality “synthesized-sounding” drum kit, and this might have affected listeners’ perception. It is possible that if the experiment were repeated with different Reggae songs the outcome would be different.
**Genre Selection**

During the usability testing some participants requested genres not included in the survey (quote 6, Appendix 1). Excluding genres might have introduced confusion about how to classify the songs and it would be interesting to conduct similar listening experiments targeted at specific genres and subgenres.

Although having a lot of genres (in the hundreds) would increase the computational complexity and make it more difficult to present the results in diagrams, it would still be entirely doable and likely preferable. A free form text input for listener’s custom genres should have been included in the survey, and if nothing else it could have provided a measurement of how often listeners were uncomfortable with the fixed genre sliders.

**Survey Instructions to the Listener**

The usability testing reinforced that the test was conducted in a proper manner (quote 1, 2 and 7, Appendix 1). The length of the survey was appropriate and there was no particular difficulty in understanding how to use the stepless sliders for classifying audio samples. However, the experiment instructions provided to the listeners before rating the audio samples were difficult to make understandable, and there was probably room for improvement. For example, there could have been clearer instructions regarding the length of the test and that it was not time limited.
**Conclusion**

Songs are often classified as the same genres when only part of the instrumentation is audible, but not always. Overall, the least defining and most genre ambiguous instruments are the percussion instruments, while the melodic components (i.e. vocals and pitched instruments) are the most genre defining. The most and least defining instrumentation by genre ([Table B](#)) reflects this with 8 out of 9 genres featuring *Percussion* as the least defining instrumentation. Therefore, when attempting to build automatic genre classification systems, with for example machine learning methods, it might be best to spend resources on extracted audio data focused on recognizing pitched instruments and vocal styling’s, rather than rhythm and percussion instruments.
References


[7] REAPER is a digital audio workstation software: a complete multitrack audio and MIDI recording, editing, processing, mixing, and mastering environment - http://www.reaper.fm/


Appendices

1. Responses from the User Testing

The following quotes account for the design impacting critique that was brought up during the user testing of the listening experiment.

1. “It looks good, no difficulty in understanding, as a participant, at all.”
2. “At first I was a little jumbled up that the form was not using a numbered scale - the habit of "a number between one to five". But then when I tried rating the first song sample there was absolutely nothing wrong. The slider works perfectly okay.”
3. “Maybe the survey instructions should explain that one might think that a song is not any of the displayed genres.”
4. “It should be clarified that you can have multiple selections on each question.”
5. “Is the experiment time limited?”
6. “I’m lacking a genre.”
7. “The survey is otherwise of an appropriate length (there is no need to sigh).”
8. “I thought it was difficult in the beginning since I thought there was a right or wrong.”
## 2. Example CSV Answer File from the Listening Experiment

Each listener’s listening experiment result was outputted in the CSV format below. All responses are available for download upon request.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Song</th>
<th>Classical</th>
<th>Rap</th>
<th>Electronic</th>
<th>Metal</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>24</td>
<td>09_Unpitched Percussion Instruments.mp3</td>
<td>0 0 0 0</td>
<td>0</td>
<td>0</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>06_Pitched Instruments.mp3</td>
<td>0 63 0 0</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>03_Vocals.mp3</td>
<td>0 34 0 0</td>
<td>100</td>
<td>0</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>05_Pitched Instruments.mp3</td>
<td>0 100 0 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>03_Unpitched Percussion Instruments.mp3</td>
<td>0 0 22 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>14_Full Mix.mp3</td>
<td>0 42 0 0</td>
<td>100</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>04_Unpitched Percussion Instruments.mp3</td>
<td>0 0 0 0</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>14_Vocals.mp3</td>
<td>0 65 0 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>16_Unpitched Percussion Instruments.mp3</td>
<td>0 9 11 0</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>01_Vocals.mp3</td>
<td>0 75 0 50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>18_Full Mix.mp3</td>
<td>0 100 0 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>11_Full Mix.mp3</td>
<td>0 100 0 0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>04_Pitched Instruments.mp3</td>
<td>0 0 0 0</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>03_Pitched Instruments.mp3</td>
<td>0 0 0 0</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>04_Vocals.mp3</td>
<td>0 84 0 39</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>02_Unpitched Percussion Instruments.mp3</td>
<td>0 100 0 0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>01_Pitched Instruments.mp3</td>
<td>0 100 0 0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>24</td>
<td>15_Full Mix.mp3</td>
<td>100 0 0 0</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1393514140-1393519058.csv
3. Listening Experiment Source Code

The following source code is a PHP script with inline HTML5 and CSS that was used to conduct the listening experiment. Audio files were loaded from the web server's file system, and are available upon request, along with the REAPER project and audio editing settings (beware: it is a fairly large download).

```php
<?php session_start(); // Load visitor session cookie.

// Global constants and install.
define("SITE_TITLE", "Listening Experiment");
define("SITE_DESCRIPTION", "Identify and determine music genres");
define("TEST_INSTRUCTIONS", "This is a listening experiment asking how much a song sounds like a music genre. There will be several song samples and you will decide how much you perceive the sample to sound like the genres. Note that it is perfectly fine to combine several genres, or even claim that a song fits into all, or none, of the music genres. It is really up to you to decide. The test will take around ten minutes.");
define("TEST_INSTRUCTIONS_GENRE_FAMILIARITY", "How familiar are you with each genre? The more to the right you set each slider, the more you believe you know about that genre - such as recalling famous songs and artists, or describing the genre's distinguishing features.");
define("TEST_FINISHED_TITLE", "Thanks for your help!");
define("TEST_FINISHED_MESSAGE", "Thank you for taking part in this survey. Your time and effort is highly appreciated.");
define("WARNING_COOKIES_DISABLED", "This website uses cookies. Please configure your web browser to allow cookies.");
define("WARNING_HTML5_AUDIO_DISABLED", "This website uses HTML5 audio. Please use a web browser that supports the audio feature.");
define("SUBMIT_TEST", "Submit answers");
define("CONTINUE_TEST", "Next song");
define("MAX_NUMBER_OF_SONG_SAMPLES", 18);
define("AUDIO_DIRECTORY", __DIR__./"audio/");
define("ANSWERS_DIRECTORY", __DIR__./"answers/");
if (!is_dir(AUDIO_DIRECTORY)) mkdir(AUDIO_DIRECTORY, 0700); // Don't forget to manually chose some audio files.
if (!is_dir(ANSWERS_DIRECTORY)) mkdir(ANSWERS_DIRECTORY, 0777);

global $parameters;

// Session handler.
if (!isset($_SESSION['active'])) : // New session.
    // Claim session as active.
    $_SESSION['active'] = true;
    $_SESSION['session_started'] = time();

    // Shuffle genre order.
    $_SESSION['parameters'] = $parameters;
```
shuffle($_SESSION['parameters']);

// Prepare survey fields.
$_SESSION['survey_fields'] = array_merge(array("Gender", "Age", "Song"),
$_SESSION['parameters']);
$_SESSION['survey_records'] = array();

// Create list of audio files.
$_SESSION['audio_files'] = array();
$fi = new FilesystemIterator(AUDIO_DIRECTORY, FilesystemIterator::SKIP_DOTS);
foreach ($fi as $file) $_SESSION['audio_files'][] = $file->getFilename();

// Shuffle audio files order.
shuffle($_SESSION['audio_files']);

// Minimize how often samples from the same song are adjacent.
$len = count($_SESSION['audio_files']);
if ($len > 2) for ($i = 0; $i < $len - 2; $i++) {
  $s1 = $_SESSION['audio_files'][$i];
  $s2 = $_SESSION['audio_files'][$i+1];
  $s3 = $_SESSION['audio_files'][$i+2];
  if (explode('_', $s1)[0] == explode('_', $s2)[0]) {
    $_SESSION['audio_files'][$i+1] = $s3;
    $_SESSION['audio_files'][$i+2] = $s2;
  }
}

// Only use the first number of songs.
$_SESSION['audio_files'] = array_slice($_SESSION['audio_files'], 0, MAX_NUMBER_OF_SONG_SAMPLES);
else: // Ongoing session.

// If a survey answer has been provided.
if (isset($_POST, $_POST[$_POST[0]])) {
  // Create records array.
  $r = array($_POST['gender'], $_POST['age'], $_POST['song']);
  foreach ($_SESSION['parameters'] as $parameter)
    $r[] = $_POST[$parameter];

  // Store records array in the session cookie.
  $_SESSION['survey_records'][$_POST['step']-1] = $r;
}
endif;

// Store results as a CSV file.
function save_answers() {
  $file_name = $_SESSION['session_started'].'-'.time().'.'.$csv';
  $file_contents = "";
  foreach ($_SESSION['survey_fields'] as $field)
    $file_contents .= $field.'    ';
  $file_contents .= PHP_EOL;
  foreach ($_SESSION['survey_records'] as $record) {
    foreach ($record as $field_value)
      $file_contents .= $field_value.'    ';
  }
}
```php
$file_contents .= PHP_EOL;
}

file_put_contents(ANSWERS_DIRECTORY.$file_name, $file_contents);

<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="utf-8">
<title><?= SITE_TITLE ?></title>
<style>
* { margin:0; padding:0; font:17px serif; }

body { background-color:#f0f0f0; padding:2%; }

h1 { font:bold 300% serif; color: rgb(255,100,100); }

h2 { font:175% serif; color:rgba(0,0,0,0.25); }

p { line-height:150%; }

header, nav, article, footer { margin:0 auto; width:480px; }

header { margin-bottom:1%; }

article { padding:40px; background:white; border:1px solid #ccc; box-shadow:0px 0px 20px rgba(0,0,0,0.2); }

div.block { margin-bottom:20px; }

audio { width:100%; }

form { width: 100%; }

form .input { width:100%; margin:1% 0; clear:both; }

form .range { font-size:0; width: 98%; padding:0% 1%; border-radius:20px; border:1px solid #fff; color:rgba(0,0,0,0.5); background: -moz-linear-gradient(left, rgba(255,255,255,0) 0%,
```
rgba(255,255,255,0.5) 50%, rgba(148,255,0,1) 100%); background: -webkit-gradient(linear, left top, right top, color-stop(0%, rgba(255,255,255,0)), color-stop(50%, rgba(255,255,255,0.5)), color-stop(100%, rgba(148,255,0,1))); background: -ms-linear-gradient(left, rgba(255,255,255,0) 0%, rgba(255,255,255,0.5) 50%, rgba(148,255,0,1) 100%); background: -o-linear-gradient(left, rgba(255,255,255,0) 0%, rgba(255,255,255,0.5) 50%, rgba(148,255,0,1) 100%); background: linear-gradient(to right, rgba(255,255,255,0) 0%, rgba(255,255,255,0.5) 50%, rgba(148,255,0,1) 100%); filter: progid:DXImageTransform.Microsoft.gradient(startColorstr='#00ffffff', endColorstr='#94ff00', GradientType=1 );

form .range:hover {
  color:rgba(0,0,0,0.9);
  border:1px solid #ccc;
}

form .range input {
  border:none;
  cursor:pointer;
}

form label {
  display:block;
  width:24%;
  float:left;
}

form input, form select {
  display:inline;
  width:75%;
  border:1px solid #ccc;
  height:100%;
  font-family:sans-serif;
}

form option {
  font-family:sans-serif;
}

form input[type="submit"] {
  display:block;
  clear:both;
  float:none;
  min-height:40px;
  border-radius:20px;
  border:none;
  margin:10px 0 0 auto;
  cursor:pointer;
  color:rgba(0,0,0,0.5);
  font-size:150%;
  font-family:sans-serif;
  text-shadow:1px 1px 0px rgba(255,255,255,0.25);
}

form input[type="submit"]:hover {
  background: rgba(148,255,0,1);
}

</style>
</head>
<body>
<?php if ($_SESSION['active'] && !isset($_POST['step'])) : ?>

<?php if ($_SESSION['active'] && !isset($_POST['step'])):

<h1><?php echo SITE_TITLE ?></h1>
<h2><?php echo SITE_DESCRIPTION ?></h2>

</body>
<h2>Instructions</h2>
<p><?= TEST_INSTRUCTIONS ?></p>

<form method="POST">
<h2>About You</h2>
<div class="input">
<label for="age">Age: </label>
<input type="number" name="age" id="age" min="0" max="123" />
</div>

<h2>Your Genre Familiarity</h2>
<p><?= TEST_INSTRUCTIONS_GENRE_FAMILIARITY ?></p>

<?php
echo '<input type="hidden" name="song" value="Genre Familiarity" />
<input type="hidden" name="step" value="0" />
<input type="submit" value="Begin listening test" />';
?>
</form>

<?php elseif ($_SESSION['active'] && $_POST['step'] < MAX_NUMBER_OF_SONG_SAMPLES) : ?>
<header>
<h1><?= SITE_TITLE ?></h1>
<h2><?= SITE_DESCRIPTION ?></h2>
</header>
<article>
<h2>Sample: <?= htmlspecialchars(($_POST['step']+1)/.MAX_NUMBER_OF_SONG_SAMPLES); ?></h2>
<input type="hidden" name="age" value="<?php echo htmlspecialchars($_POST['age']); ?>" />
<input type="hidden" name="gender" value="<?php echo htmlspecialchars($_POST['gender']); ?>" />
<input type="hidden" name="song" value="<?php echo htmlspecialchars($_SESSION['audio_files'][$_POST['step']]); ?>" />
<input type="hidden" name="step" value="<?php echo htmlspecialchars($_POST['step'] + 1); ?>" />
<input type="submit" value="<?php echo ($_POST['step'] == MAX_NUMBER_OF_SONG_SAMPLES-1) ? SUBMIT_TEST : CONTINUE_TEST; ?>" />
</article>
</article>
<?php else : if ($_SESSION['active']) save_answers(); $_SESSION['active'] = false; ?>
  <header>
    <h1><?php echo TEST_FINISHED_TITLE ?></h1>
    <h2><?php echo TEST_FINISHED_MESSAGE ?></h2>
  </header>
<?php endif; ?>
</body>
</html>
4. Songs

All songs used in the web-based listening experiment, grouped by genre metadata.

- Electronic
  - Calvin Harris - I Need Your Love
  - Deadmau5 - I Remember
- Reggae
  - Toots & The Maytals - Do The Reggay
  - Bob Marley - Buffalo Soldier
- Rap
  - NWA - Fuck Tha Police
  - Public Enemy - Fight The Power
- Country
  - Hank Williams, Sr. - I Saw The Light
  - Country Standards - Cotton-Eyed Joe
- Jazz
  - Ella Fitzgerald - Cheek To Cheek
  - The Andrews Sisters - Chattanooga Choo Choo
- Pop/Rock
  - Bon Jovi - Livin' On A Prayer
  - U2 - With Or Without You
- Blues
  - B.B. King - Rock Me Baby
  - The Jimi Hendrix Experience - Red House
- Classical
  - Andrea Bocelli - Nessun Dorma
  - Luciano Pavarotti - Granada
- Metal
  - Slipknot - Psychosocial
  - Metallica - Enter Sandman