HARMONIC COMPATIBILITY BASED ON TONAL INTERVAL VECTORS

Miguel Pérez Fernández
miguel_on_94@hotmail.com

ABSTRACT

In this work we explored the Tonal Inverval Vectors (TIV), a measure for harmonic compatibility for music excerpts. We will evaluate the TIV and compare it with another dissonance model, using a set of music fragments ready for music production. After calculate the most compatible mixes, we conducted an user-based test to measure the quality of the results. The results aren't strong enough and further experiments should be conducted. However we obtained certain directions to continue with the exploring harmonic compatibility measures. All the code can be found in a github repository with instructions to run it [1] .

1. INTRODUCTION

Harmonic compatibility can be defined in many ways. Two musical excerpts that sound one after another can be more or less harmonically compatible, being more compatible if they fit well one after another. We can also refer to harmonic compatibility as the capability of two pieces of music to sound together, and produce a pleasant sound. Those compatibilities are also usually described as *sequential compatibility* (Horizontal compatibility) and *simultaneous compatibility* (Vertical compatibility). In this work we will refer to the *vertical compatibility* when the *harmonic compatibility* expression appears, unless it's explicitly referred as *sequential compatibility*. The approaches to measure harmonic compatibility in the actual State of the Art (SoA) can be classified into two main groups: Chroma-distance based measures and Spectrum based measures.

2. RELATED WORK

In this section, we will give an overview to the *Chroma-distance* and the *Spectrum* based measures.

2.1 Chroma-distance

The *Automashupper* algorithm [6] is a tool to create mashups. Mashup is a type of music made by blending two or more prerecorded songs. The blending could be by overlapping certain parts of the song, or switching between sections of both tracks. The main problem to address for this music creations is to find song fragments that fit well when mixing. To solve this *harmonic compatibility* problem the authors use Chroma vectors sync with the beat of the songs. By doing 2D convolutions between Chroma vectors of the songs, the authors are able to find the beat and pitch offset that produces the biggest harmonic compatibility. This work also introduces a concept that doesn't exist in the other methods mentioned here in the SoA: Spectral Balance. To enforce the creation of spectral balanced mixes, the authors divide the spectrum into three different regions. The regions are: low-band $f < 220Hz$, mid-band $220Hz < f < 1760Hz$, high-band $1760Hz < f$.

By using the perceptual loudness for each or those three regions, and finding the beat offset that makes the spectrum the flattest possible. By summing both features (spectral and harmonic) the authors created a score for a given pair of songs:

$$M_n(k) = w_H M_{H,n}(k) * w_L M_{L,n}(k)$$

Where $M_n(k)$ is the compatibility given beat offset k, $M_{H,n}(k)$ is the harmonic component of *mashability* for song n with beat offset k and $M_{L,n}(k)$ is the spectral balance component of the *mashability*. w_H and w_L are the weights for the harmonic and spectral components, being $w_H = 1$ for harmonic component and $w_L = 0.2$ for spectral component.

However, this process does not take into account movements that occur within the musical phrase. This may result in moving audio fragments in your melody overlapping. This mixing is usually unpleasant and can look like random notes. To solve this problem the authors of [11] create a measure of *harmonic change rate*, based on how Chromas stay stable along every beat of the audio excerpt. The audios can be classified as *totally unstable* (1) or *totally stable* (0). The algorithm by the authors in [11] tries to match together those musical phrases that have a *high harmonic change rate* with those of *low harmonic change rate*.

2.2 Spectrum-based

In his work [8], Gebhartdt propose a method to mix sound excerpts trying to minimize the **roughness** of the result. Roughness is the sensation arisen when two frequencies very close in the spectrum, creates a sensation of amplitude modulation. This sensation is usually perceived as uncomfortable for most people. Based on the work of Hutchinson & Knopoff [10], and Plomp & Levelt [13], this roughness model is used along with a *Pitch commonality* model,

[1] https://github.com/migperfer/MIR-UPF

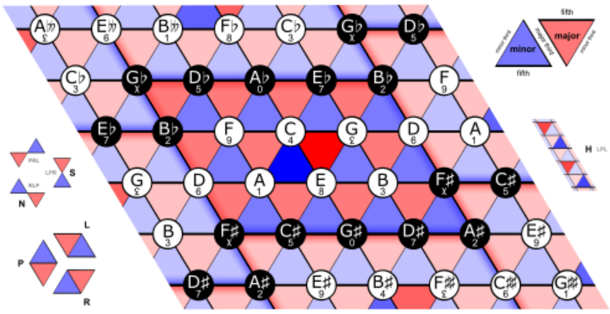

Figure 1. Tonnetz space. Source: Wikipedia

based on Parncutt & Strasburger work [12] to assist harmonic mixing. The authors extract 20 sines from each audio excerpt using the package *Spectral Modeling Synthesis Tools* [15], and then apply the roughness model to compute the overall roughness. In addition to this to ensure the maximum harmonic compatibility between the two musical excerpts, the roughness is calculated through 97 shifts (48 onward and 48 downward around a "no shift"). The results shows that mixes with smaller **roughness** values, were more consonant to the listeners. Some models are derived from experiments with dyads and user evaluation, like the one proposed in Plompt & Levelt [14]. This model is based on the roughness created by two tones when they're in the same critical band. There is a implementation of this algorithm in the Essentia library [5].

3. TONAL INTERVAL VECTORS

Approaches using raw Chroma doesn't take into account the aspects of human perception regarding intervals. In a Chroma vector notes are arranged by their closeness in frequency, but not by their relevance in terms of harmony. Models like the Tonnetz defined by Euler [7] describe a space in which each pitch class is closer to other pitch classes harmoniously more important. To address this lack of perceptual analysis in harmonic compatibility models, the Tonal Interval Vector (TIV) feature is used in works as [9] [3] and [2]. TIVs are defined in the following manner:

$$T(k) = w_a(k) \sum_{n=0}^{N-1} \bar{c}(n) e^{\frac{-j2\pi kn}{N}}, 1 \leq k \leq 6 \in \mathbf{Z}$$

Where $\bar{c}(n)$ is a normalized version of the original chroma vector $c(n)$ divided by the DC component: $T(0) = \sum_{n=0}^{N-1} c(n)$. This normalization allows to compare musics with different hierarchical levels of tonal pitch [9]. w_a is a weight value that adjusts to the contribution of each interval (k). According to the similarity with the Discrete Fourier Transform we can define therefore [2]:

$$|T(k)| = \sqrt{\Re\{T(k)\}^2 + \Im\{T(k)\}^2}$$

$$\phi(k) = tan^{-1} \frac{\Im\{T(k)\}}{\Re\{T(k)\}}$$

In his work [9], presented harmonic indicators based on TIVs:

- Dissonance

$$D = 1 - \frac{||T(k)||}{||w_(k)||}$$

- Perceptual relatedness

$$R_{i,j} = \sqrt{\sum_{k=1}^{M} |T_i(k) - T_j(k)|^2}$$

From the dissonance equation, we can therefore measure the dissonance between two overlapping tracks:

$$D_{i,j} = 1 - \frac{||a_i T_i(k) + a_j T_j(k)||}{(a_i + a_j)||w_a(k)||}$$

A *small-scale* measure between sound i and j is provided in [9] given the previous dissonance and relatedness:

$$H_{i,j} = \bar{R}_{i,j} \bar{D}_{i,j}$$

Where $\bar{R}_{i,j} \bar{D}_{i,j}$ are the normalized version of $R_{i,j}$ and $D_{i,j}$. The lower is the H value, the more compatible are the two songs.

4. METHODOLOGY

We selected a subset of music fragments from the audio database [2]. This collection of audio is part of the looperman website [1]. The subset consists in music excerpts that are on 140bpm with a 32-beats long duration. The final length of this subset is 682 samples. We calculate the beatwise TIV for the target audio excerpt and all the candidates in the subset. By using the *small scale compatibility* defined in [4] we calculate the compatibility between two musical excerpts for each beat. We asses then the overall compatibility between two audios by taking the sum across the beat dimension. Besides the beatwise version of the TIV we also calculate another single TIV for the whole audio excerpt, and a TIV for every frame in both audios.

5. EVALUATION

Using the Jupyter notebook attached in the repository we select randomly a target audio excerpt. After computing all the compatibilities as stated in the *Methodology* section, we present the user with the ten most compatible mixes. After hearing the ten mixes for each version of the TIVs, the user is presented with a form that allows to rate the mixes with an integer in the (0,5) range. We then retrieve the following information:

- The target audio randomly selected
- The score for the mixes
- The small scale compatibility for each mix

[2] https://drive.google.com/open?id=1mS0G_Gk4v6IHe2E2IZ2Z6dcE-lcTuZx3

	TIV(beatwise)	TIV(whole)	TIV(framewise)	Dissonance
Mean rating	1.68	1.08	2.19	1.08
Rating variance	1.94	1.08	2.01	1.31

Table 1. Mean ratings for each algorithm in a 0-5 scale

	TIV(beatwise)	TIV(whole)	TIV(framewise)	Dissonance
Correlation	-0.21	-0.18	0.06	0.21
P-value	0.17	0.23	0.68	0.16

Table 2. Correlation between harmonic compatibility and user rating

An audio excerpt is selected randomly from the subset created for the experiment. In this sense when rating all mixes, if the candidate audio is mainly percussion the user is asked to rate the mix with a *-1* instead of the (0,5) normal range.

In addition to the TIV methods, the model from [14] implemented in Essentia is included in the experiments. We calculate the framewise roughness and then mean across the frame dimension for the resulting mix of both audios.

There is a particular difference between these two methods. The *small scale compatibility* measure presented ensures that the highest compatibility for some audio will be the audio itself. In the other hand, the dissonance method will calculate the dissonance of the mix, which may lead to mixes on which the best candidate is not the target audio itself.

For this experiment we didn't apply any pitch shift because the pitch shifting algorithm itself could affect the consonance of the resulting mix. We calculate only the harmonic compatibility for the original audio, unlike works as Automashupper [6].

6. RESULTS

We asked four people with musical to fill the survey described in section 5. In total we collected 4 different cases, consisting each case in 40 mixes:

- 10 most compatible mixes according to beatwise TIVs.
- 10 most compatible mixes according to a single TIV for the whole audio.
- 10 most compatible mixes according to framewise TIVs.
- 10 most compatible mixes according to dissonance measure from Essentia.

The results in the table 1 shows the mean score given by users. Those results are from the 10 most harmonically compatible mixes for each target song and algorithm. We also analysed the correlation between the measures of harmonic compatibility and ratings given by the user. The results for this correlation can be seen in table 2.

7. DISCUSSION

P-values in table 2 shows values too high to ensure a valid correlation. Most compatible sounds should get a TIV small scale compatibility near 0. In this sense the expected correlation should be negative since smaller *H* should get higher ranking scores. Negative correlations in beatwise and whole TIVs correspond with the expected behaviour. However the framewise correlation is very weak despite the high user ratings. During experiments apart from the survey we found that framewise TIV usually matches audios with high ratings of silence. This could be explained by the fact that silence contains TIVs close to $0 + 0i$ for every k. This can minimize $\bar{R}$ value of the *small scale compatibility*. In the other hand beatwise TIV has stronger correlation than whole TIV, with a smaller p-value. The lack of time resolution for TIVs for the whole audio can lead to sometimes dissonant intervals in the audio. Essentia's dissonance shows a similar p-value and opposite direction to TIV(Beatwise). Essentia's dissonance has the lowest ratings and also a relatively small variance. The positive correlation doesn't correspond to the expected negative correlation, since less dissonant sounds should have a value close to 0. In our experiments, we also found that most Dissonant matches were low-harmonic leading voices like high-pitched sine-waves, or triangles. Essentia's dissonance is based on Plompt & Levelt's work [14] about dissonance as result of interaction in critical bands (CB) in the ear. The algorithm seems to select relatively low harmonic content since it will create less dissonance in the auditory system, despite music excerpts not being related in terms of tonality.

8. CONCLUSIONS

Here we explored the Tonal Interval vectors as a measure for harmonic compatibility for audio excerpts. We calculated TIVs in three different time resolutions: Framewise, beatwise and whole excerpt. We also calculated the framewise dissonance implemented in Essentia. The correlation and p-values of measures with ratings are not strong enough to prove the models to be successful. However they allow to continue working with certain directions. TIV time resolution should be wide enough to capture enough tonal information of the audio, but not so small that it fails to follow the movements in the voice leading. Essentia's dissonance fails to capture tonality information because prioritises avoiding interactions inside CB.

9. FUTURE WORK

The amount of surveys we collected may affect our results in table 2. More discussions and conclusions could be drawn with more observations. Beatwise TIVs were successful to follow the melody on the music, but fail to capture dissonances created by roughness. A composite model of beatwise TIVs and CB interference can lead to more interesting results. Instead of using the Plompt & Levelt model [14] the analysis can be done with other models as the one from Hutchinson & Knopoff [10]. This last model has been already proven to assist music mixing [8].

10. REFERENCES

[1] Looperman: https://www.looperman.com/.

[2] Gilberto Bernardes, Diogo Cocharro, Marcelo Caetano, Carlos Guedes, and Matthew E.P. Davies. A multi-level tonal interval space for modelling pitch relatedness and musical consonance. *Journal of New Music Research*, 45(4):281–294, 2016.

[3] Gilberto Bernardes, Matthew Davies, and Carlos Guedes. A perceptually-motivated harmonic compatibility method for music mixing. 09 2017.

[4] Gilberto Bernardes, Matthew E P Davies, and Carlos Guedes. A Hierarchical Harmonic Mixing Method BT - Music Technology with Swing. pages 151–170, Cham, 2018. Springer International Publishing.

[5] Dmitry Bogdanov, Nicolas Wack, Emilia Gómez, Sankalp Gulati, Perfecto Herrera, Oscar Mayor, Gerard Roma, Justin Salamon, José Zapata, and Xavier Serra. Essentia: An open-source library for sound and music analysis. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, page 855–858, New York, NY, USA, 2013. Association for Computing Machinery.

[6] Matthew E.P. Davies, Philippe Hamel, Kazuyoshi Yoshii, and Masataka Goto. AutoMashUpper: Automatic creation of multi-song music mashups. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 22(12):1726–1737, 2014.

[7] Leonhard Euler. *Tentamen novae theoriae musicae ex certissimis harmoniae principiis dilucide expositae*. ex typographia Academiae scientiarum, 1739.

[8] Roman Gebhardt, Matthew E.P. Davies, and Bernhard Seeber. Harmonic mixing based on roughness and pitch commonality. *DAFx 2015 - Proceedings of the 18th International Conference on Digital Audio Effects*, pages 1–8, 2015.

[9] Carlos Guedes. *A Hierarchical Harmonic Mixing Method*, volume 11265. Springer International Publishing, 2018.

[10] William Hutchinson and Leon Knopoff. The acoustic component of western consonance. *Interface*, 7(1):1–29, 1978.

[11] Chuan-Lung Lee, Yin-Tzu Lin, Zun-Ren Yao, Feng-Yi Lee, and Ja-Ling Wu. Automatic mashup creation by considering both vertical and horizontal mashabilities. In *ISMIR*, pages 399–405, 2015.

[12] Richard Parncutt and Hans Strasburger. Applying psychoacoustics in composition: Harmonic progressions of non-harmonic sonorities. *Perspectives of New Music*, 32:88–129, 07 1994.

[13] R. Plomp and W. J. M. Levelt. Tonal consonance and critical bandwidth. *The Journal of the Acoustical Society of America*, 38(4):548–560, 1965.

[14] R. Plomp and W. J. M. Levelt. Tonal consonance and critical bandwidth. *The Journal of the Acoustical Society of America*, 38(4):548–560, 1965.

[15] Xavier Serra and Music Technology Group. Spectral Modeling Synthesis Tools.