

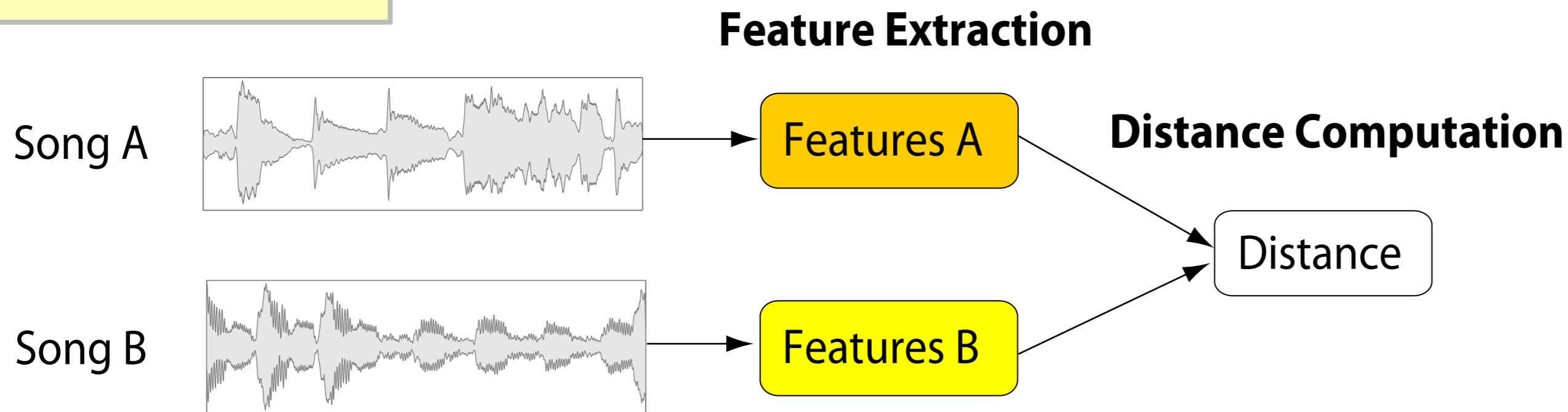
Audio-Based Music Similarity and Retrieval: Combining a Spectral Similarity Model with Information Extracted from Fluctuation Patterns



Elias Pampalk

National Institute for Advanced Industrial Science and Technology (AIST)
IT, AIST, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

Basic Concept



1103 numbers are extracted from each audio file (wav format, 22kHz mono). The computation time to extract features per file is less than 2 seconds. Each distance computation takes about 0.1 milliseconds (on a 2GHz Centrino).

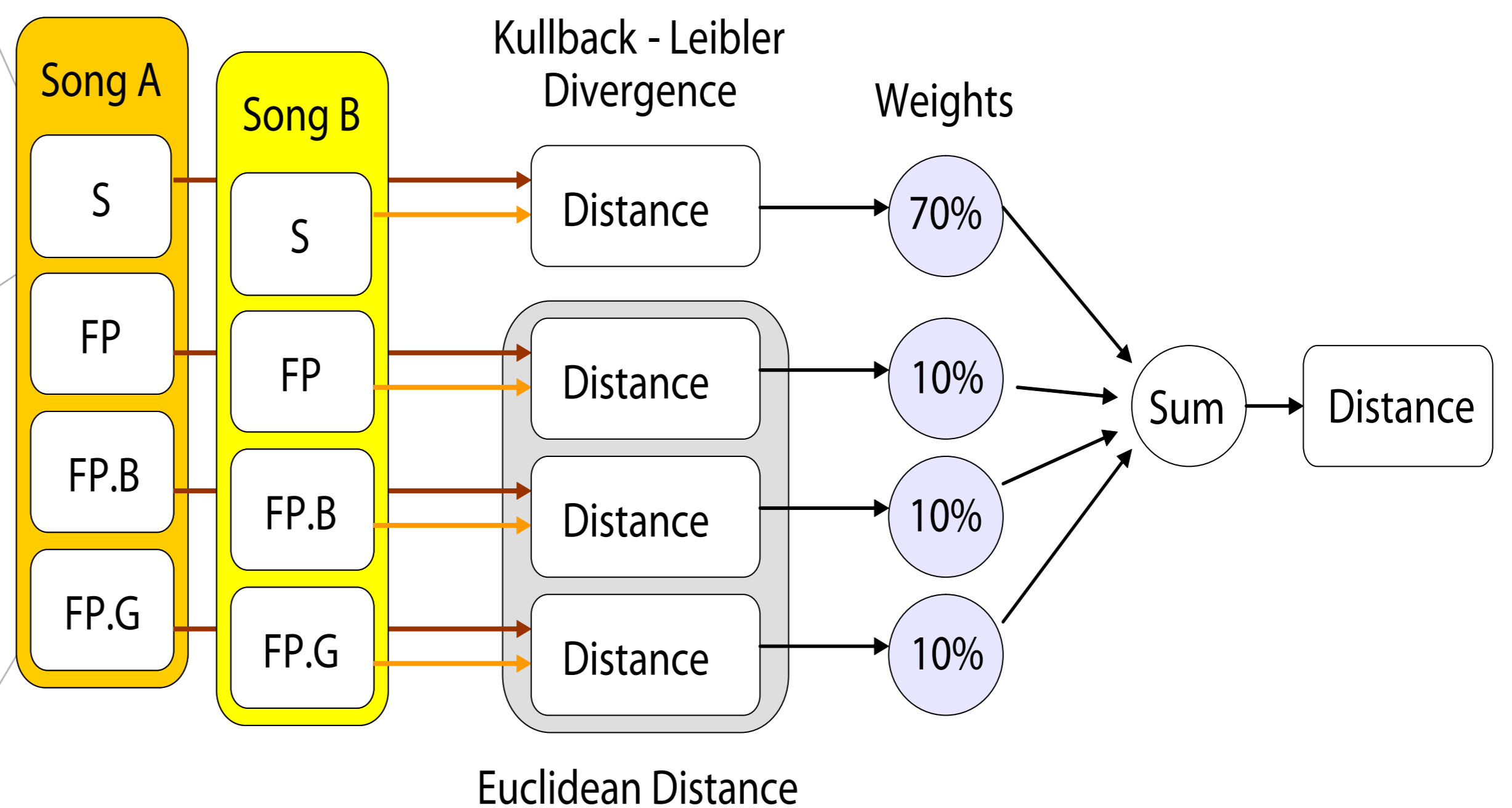
Spectral Similarity (G1)

For each song the average spectral shape is computed, and its variance (using MFCCs). These spectral shapes are compared using a symmetric form of the Kullback-Leibler Divergence. To some extent the spectral shape is related to the perception of timbre.

Fluctuation Patterns (FPs)

For each song the modulation of the loudness amplitude in different frequency bands is computed. Based on this, further information is extracted (FP.B, FP.G). To some extent the fluctuation patterns describe beats and other periodicities in the song.

Combination of Similarity Measures (G1C)



Optimization of the Weights

The combination weights were optimized using two music collections and a genre-based evaluation. Special precautions were taken to avoid overfitting. The best combination was evaluated using independent music collections and genre data. In addition a listening test was conducted which confirmed that the improvements compared to using only spectral similarity are significant.

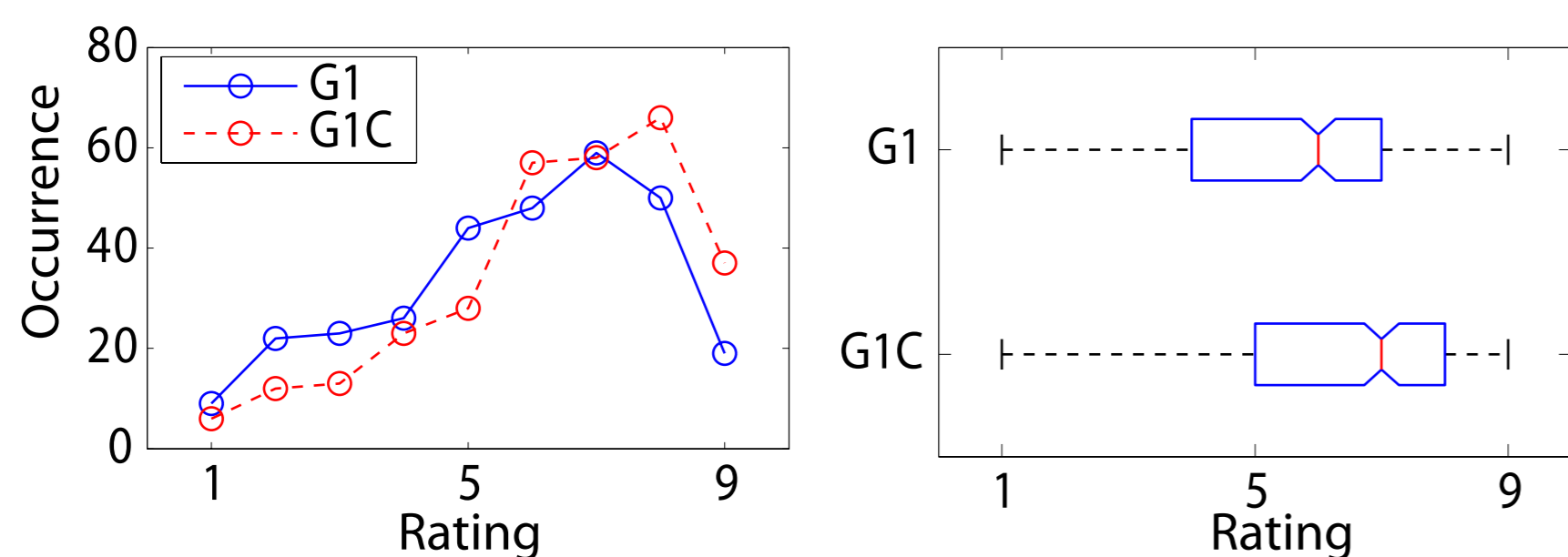


Figure 2.38 (Page 85): Histogram and box plots of all ratings comparing G1 to G1C.

| Rank | Noisiness | Perc. | FP | FP Gravity | FP Bass | FP DLF | Spec. Sim. | DB-MS | | DB-L | | Score |
|------|-----------|-------|----|------------|---------|--------|------------|-------------|-------------|-------------|-------------|-------|
| | | | | | | | | G1 | G30S | G1 | G30S | |
| 1 | | | 10 | 10 | 10 | | 70 | 67.4 | 67.4 | 32.4 | 35.2 | 6.14 |
| 2 | | 10 | 10 | 10 | 10 | | 60 | 67.1 | 66.4 | 33.0 | 34.6 | 5.83 |
| 3 | | 10 | | 10 | 10 | | 70 | 66.8 | 66.4 | 31.8 | 34.7 | 5.46 |
| 4 | | | 10 | 10 | | | 80 | 67.4 | 65.7 | 32.1 | 34.4 | 5.44 |
| 5 | | | 10 | 10 | 20 | | 60 | 66.1 | 66.9 | 31.5 | 34.9 | 5.42 |
| 6 | | 10 | 20 | | 10 | | 60 | 65.7 | 66.4 | 32.6 | 34.5 | 5.36 |
| 7 | 10 | | 10 | 10 | 10 | | 60 | 63.9 | 66.1 | <u>33.6</u> | <u>35.6</u> | 5.35 |
| 8 | | 10 | | 10 | | | 80 | 66.8 | 66.1 | 31.8 | 34.1 | 5.26 |
| 9 | 10 | | 20 | 10 | | 10 | 50 | 64.9 | 66.1 | 32.7 | 35.1 | 5.25 |
| 10 | | | 10 | 10 | 10 | 10 | 60 | 67.2 | 66.8 | 30.9 | 33.9 | 5.25 |
| 11 | | | | 10 | 10 | | 80 | <u>68.2</u> | 66.7 | 31.0 | 32.9 | 5.25 |
| 25 | 10 | | 20 | 10 | | | 60 | 64.1 | 65.2 | 32.7 | <u>35.6</u> | 4.92 |
| 515 | | | | 30 | | 20 | 50 | 66.0 | <u>68.4</u> | 26.5 | 29.5 | 3.15 |
| 2666 | | | | | | | 100 | 62.8 | 62.4 | 27.6 | 25.0 | 0.00 |

Table 2.10 (Page 72): Top 10 combinations. The last line is the baseline using only spectral similarity. The first column is the rank. All values (other than the ranks) are given in percent. Values marked with a line below and above are the highest accuracy achieved for the specific combination of similarity measure and collection.

For details see:

E. Pampalk, Computational Models of Music Similarity and their Application in Music Information Retrieval, Doctoral Dissertation, TU Wien, Austria, March 2006.