Aucouturier, Jean-Julien & Francois Pachet. Music Similarity Measures: What's the Use? Proceedings of the ISMIR International Conference on Music Information Retrieval (Paris, France), 2002, pp. 157-163.

> The authors discuss the context of music similarity research, describing the domain of Electronic Music Distribution and explaining the practical applications of the field. This is largely a technical paper which presents a computational method for calculating music similarity by using measurements of timbre similarity. The authors do discuss subjective evaluation of their technique, and admit that "the acceptance of the notion of timbral similarity by users is not always systematic." They point out the inherent difficulty of assessing music similarity based on one measure, and evaluate the question of what constitutes an 'interesting' result from their timbral algorithm. The last part of the paper describes the potential uses for their algorithms, including playlist generation and a function they call the 'aha slider'.

Berenzweig, Adam, Beth Logan, Daniel P.W. Ellis, & Brian Whitman. A Large-Scale Evaluation of Acoustic and Subjective Music Similarity Measures. Proceedings of the ISMIR International Conference on Music Information Retrieval (Baltimore, MD), 2003, pp. 99–105.

> This article addresses technical acoustic similarity, as well as subjective similarity, and suggests using the subjective measurements as a companion to the authors' technical algorithms. These subjective measures are described as a 'sanity check' against the more technical, quantitative methods, and methods for converting subjective judgments into useful measures are also discussed. The method they describe involves a web-based survey which asks users the question "which artist [of ten offered] is most similar to the target artist?" The responses are then formed into a matrix which contained the "number of times a given artist was chosen as most similar to a target as a proportion of the trials in which it could have been chosen." Even though the researchers collected 22,000 responses, the results reportedly provided only a 'sparse' similarity matrix.

Deliege, I. (2001). Introduction: Similarity Perception ↔ Categorization ↔ Cue Abstraction. (Cover story). *Music Perception*, *18*(3), 233. Retrieved from Academic Search Complete database.

> This article discusses the various different disciplines that have done studies in the area of music similarity, identifying psychology, music theory, and computation. The author discusses the music-theory based concept that music similarity is defined more by the human tendency to recognize differences between self-created groupings of music. She refers to this as the Gestalt principle. In the article, she is most concerned with the internal cognitive processes involved in the categorization process. She describes this categorization as an 'emergent process' in which cues are identified in the music and abstracted into a mental structure that allows the musical work to be categorized within the listeners own, self-generated table of groupings. Finally, specific studies are referenced that include subjective human listening experiments.

Ellis, Daniel, Brian Whitman, Adam Berenzweig, & Steve Lawrence. The Quest For Ground Truth in Musical Artist Similarity. Proceedings of the ISMIR International Conference on Music Information Retrieval (Paris, France), 2002, pp. 170-177.

> This article investigates different subjective measures of similarity in order to determine which characteristics would be best employed in computational methods in the future. The authors point out that current computational methods cannot be easily compared, since they accept different standards of what constitutes 'similarity', and see a need for a 'ground truth' of subjective similarity measurements in order to be able to make useful comparisons. Some of the different subjective measures they discuss include using data from a published music guide to create an Erdös matrix, using a P2P network to extract cultural similarity data, community metadata, and finally an actual web-based survey soliciting direct comparisons from users. In the end, all of these methods are compared, and while there is only a limited degree of agreement, the authors feel they did develop a usable matrix based on a combination of the methods.

Logan, Beth, Daniel P.W. Ellis, & Adam Berenzweig. Toward Evaluation Techniques for Music Similarity. Proceedings of the SIGIR Workshop on the Evaluation of Music Information Retrieval Systems (Toronto, Canada), 2003.

> The introduction points out many of the difficulties facing anyone trying to formulate a music similarity algorithm, and also addresses the lack of a common 'ground truth'. They also address various different methods of taking subjective music similarity measures and quantizing them for use in computational schemes. They also discuss the survey website constructed for their previous study (see above). This follow-up paper goes into a little more depth than their previous publication, but the work is clearly still in process.

Novello, Alberto & Martin F. McKinney. Assessment of Perceptual Music Similarity. Proceedings of the ISMIR International Conference on Music Information Retrieval (Vienna, Austria), 2007, pp. 111-112.

Novello, Alberto, Martin F. McKinney, & Armin Kohlrausch. Perceptual Evaluation of Music Similarity. Proceedings of the ISMIR International Conference on Music Information Retrieval (Victoria, BC (Canada)), 2006, PAP.

> These two papers constitute a very useful overview of previous research in subjective, perceptual music similarity, which leads them to the conclusion that the published research is fragmented or too specific for general application. They then present their own study, which was an attempt to evaluate whether people have a common perception of music similarity. Their conclusions suggest that musicians have a more common approach, while non-musicians vary considerably more. The second paper from the following year describes a different, larger study which confirmed their earlier results, though the larger scale of the experiment revealed more complexity in the data.

Pampalk, Elias, Arthur Flexer, & Gerhard Widmer. Improvements of Audio-Based Music Similarity and Genre Classification. Proceedings of the ISMIR International Conference on Music Information Retrieval (London, UK), 2005, pp. 628–633.

The authors discuss the evaluation techniques for computational music similarity algorithms, suggesting that genre categorization rather than listening tests provide the best solution for evaluation. The specifics of their computational system are addressed, and then the system is evaluated in the method they proposed. Essentially, their algorithm tests above the baseline, but they still identify a 'glass ceiling' beyond which current technical solutions cannot penetrate. They suggest that "combining audio-based approaches with information from different sources (such as the web), or modeling the cognitive process of music listening are more likely to help us get beyond the glass ceiling."

Schubert, E., & Stevens, C. (2006). The effect of implied harmony, contour and musical expertise on judgments of similarity of familiar melodies. *Journal of New Music Research*, *35*(2), 161-174. doi:10.1080/09298210600835000.

A music-theory based study with 72 participants, including both musicians and non-experts. This study directly solicited their judgments on whether simple melodies were similar, and then presented the melodies after they were put through various transformations involving both harmony and pitch displacement. Ultimately, they determine that there were substantial differences between the experts and non-experts, perhaps even in terms of cognitive architecture, and certainly in their subjective judgments.

Slaney, Malcolm, Kilian Weinberger, & William White. Learning a Metric for Music Similarity. Proceedings of the ISMIR International Conference on Music Information Retrieval (Philadelphia, USA), 2008, pp. 313-318.

These authors describe a number of computational methods for embedding music harvested from blogs (with accompanying metadata, descriptions, etc.) into a Euclidean metric space designed such that it is possible to determine two songs' similarity or dissimilarity by a distance measurement. The difficulty of assessing similarity based on limited data leads them to try six different ways for weighting the available data so that they can compare performance. They found that the best results came from the algorithms that provided greater weight to the metadata harvested with the music. This metadata, including the text of blog posts, is essentially a way to insert the human judgments into their computational scheme.

Timmers, R. (2005). Predicting the similarity between expressive performances of music from measurements of tempo and dynamics. *Journal of the Acoustical Society of America*, *117*(1), 391-399. doi:10.1121/1.1835504.

Describes the results of two studies performed with 40 participants (including both experts and non-experts) who evaluated the similarity between several performances of Mozart and Chopin. Various characteristics of these performances were rated, such as tempo, loudness, rubato, dynamics, articulation, emotion, etc. The author then discusses the contrast between the results and those of various computational methods, finding that some of the models had very little agreement.