**CHAPTER 3: METHODOLOGY**

# INTRODUCTION

This chapter presents a comprehensive methodology for developing a talking drums dataset for artificial intelligence pattern generation. The approach systematically leverages existing digital audio resources while maintaining the highest standards of cultural authenticity and technical rigor, following established frameworks in music information retrieval (Müller and Ewert 2015; Lerch 2012) and ethnomusicological research (Nettl 2005; Titon et al. 2017).

The methodology is designed to be both academically robust and practically implementable. Each technical process is explained in clear terms before presenting the mathematical foundation, ensuring accessibility to readers from diverse backgrounds while maintaining scholarly depth, as recommended by Creswell (2014) for mixed-methods research approaches.

The research adopts a resource-integration approach that curates, processes, and synthesizes available materials from multiple sources. This practical strategy enables immediate research implementation while establishing scalable methodologies for future dataset expansion, aligning with current trends in digital humanities (Moretti 2013; Hockey 2004) and computational ethnomusicology (Serra et al. 2013; Fuentes et al. 2007).

**Mathematical Framework Overview:**

The entire methodology can be represented as an optimization problem where we seek to maximize dataset quality , authenticity , and cultural preservation :

Subject to practical constraints:

This mathematical foundation guides every decision in the methodology, ensuring systematic and measurable progress toward our research goals.

# RESEARCH DESIGN AND APPROACH

## Resource-Integration Framework

Traditional ethnomusicological research typically involves extensive field recordings, which can be time-consuming and logistically challenging (Rice 2014; Barz and Cooley 2008). Our resource-integration approach offers a practical alternative by systematically identifying, evaluating, and curating existing digital audio materials, following principles established in digital library science (Arms 2000; Borgman 2015).

This framework operates on four fundamental principles:

1. **Systematic Resource Discovery**: We conduct comprehensive searches across digital platforms, academic repositories, and cultural archives. This process follows structured methodology that ensures complete coverage of available materials (Terras 2015; Zorich 2012).
2. **Quality-Driven Curation**: Every audio sample undergoes rigorous evaluation using both technical metrics and cultural authenticity criteria (Casey et al. 2008; Klapuri and Davy 2006).
3. **Standardized Processing**: All selected materials pass through a uniform preprocessing pipeline following established audio processing standards (Zölzer 2011; Orfanidis 2007).
4. **Comprehensive Annotation**: We implement multi-dimensional labeling systems based on established practices in music informatics (Downie 2003; Schedl, Gómez, and Urbano 2014).

**Mathematical Representation:**

Each resource is characterized by a feature vector:

Where:

* = Quality score (0 to 1, where 1 is highest quality)
* = Authenticity score (0 to 1, where 1 is most authentic)
* = Cultural relevance score (0 to 1, where 1 is most relevant)
* = License compatibility (0 or 1, binary indicator)
* = Technical suitability score (positive real number)

The resource selection process uses a weighted scoring function:

Resources with scores are included in the dataset. This mathematical approach ensures objective, reproducible selection criteria while balancing all important factors.

## Theoretical Framework Integration

Our methodology integrates three complementary theoretical approaches based on current research in computational musicology (Huron 2006; Temperley 2007):

**Digital Ethnomusicology Framework**

This approach utilizes digital tools and resources to study traditional music while maintaining cultural authenticity and respect for source communities (Seeger 2004; Post 2006). The mathematical model represents this as:

Where = Musical analysis, = Cultural context, = Preservation techniques, = Traditional knowledge.

**Music Information Retrieval (MIR) Framework**

We apply computational methods specifically adapted for African musical contexts (Rycroft 1961; Kubik 1999). The MIR framework is modeled as:

Where: = Feature extraction, = Audio analysis, = Rhythm recognition, = Classification methods.

**Cross-Cultural AI Development Framework (CCAI):**

This ensures AI systems can understand and generate culturally appropriate content (Barocas, Hardt, and Narayanan 2019; Liang et al. 2022):

Where: = Diverse datasets, = Multicultural models, = Cross-cultural validation, = Ethical considerations.

# SYSTEMATIC RESOURCE IDENTIFICATION AND ACQUISITION

## Digital Platform Survey Strategy

The resource discovery process follows a systematic approach to maximize the probability of finding high-quality talking drum samples, utilizing established methodologies in digital collection development (Lee 2002; Johnson 2018).

Our primary focus centers on platforms with substantial African music collections:

**Freesound.org Community Repository**

Freesound.org represents our primary source for community-contributed samples (Font, Roma, and Serra 2013). The platform contains over 700,000 audio samples with Creative Commons licensing, making it ideal for research applications.

**Search Strategy Implementation:**

We employ a systematic search methodology using multiple keyword combinations:

* Primary keywords: “talking drum”, “dundun”, “gangan”, “yoruba drum”
* Secondary keywords: “tama”, “sabar”, “djembe”, “west african drum”
* Tertiary keywords: “hourglass drum”, “african percussion”, “traditional drum”

The search effectiveness is measured mathematically as:

**Expected Yield Calculation:**

Based on preliminary searches and quality filtering, we estimate:

**Academic Digital Archives**

Several academic institutions maintain digitized ethnomusicological collections relevant to this research (Seaman 2011; Proffitt 2005):

* **British Library Sound Archive**: Comprehensive world music collection with professional-quality recordings
* **Smithsonian Folkways Recordings**: Extensive African music catalog with detailed cultural documentation (Goldsmith 2019)
* **Archive of World Music (Harvard)**: Academic collection with ethnomusicological context
* **UCLA Ethnomusicology Archive**: Research-quality recordings with scholarly annotations

## Quality Assessment Framework

Each audio sample undergoes comprehensive technical evaluation using established audio quality metrics (Thiede et al. 2000; Kabal 2002):

Audio Quality Assessment Metrics and Mathematical Formulations

| **Metric** | **Mathematical Formula** | **Minimum** | **Target** | **Weight** |
| --- | --- | --- | --- | --- |
| Signal-to-Noise Ratio |  | 35 dB | 50+ dB | 0.25 |
| Dynamic Range |  | 30 dB | 50+ dB | 0.20 |
| Frequency Response |  | 0.68 | 0.95+ | 0.20 |
| Total Harmonic Distortion |  |  |  | 0.15 |
| Spectral Bandwidth |  | Hz | Hz | 0.20 |

The composite quality score combines these metrics:

where represents the weight for metric and is the normalized score for that metric.

# COMPREHENSIVE AUDIO PROCESSING AND ENHANCEMENT

## Standardization and Preprocessing Pipeline

All audio samples undergo systematic preprocessing following established digital audio standards (Pohlmann 2005; Watkinson 2001):

**Technical Specifications:**

* Sample Rate: 44.1 kHz (CD quality standard)
* Bit Depth: 24-bit (professional quality)
* Format: WAV (uncompressed)
* Channels: Mono (focused on rhythmic content)

**Processing Pipeline Mathematical Formulation:**

The preprocessing pipeline applies sequential transformations:

Where:

**Noise Reduction Implementation:**

We employ spectral subtraction for noise reduction (Boll 1979; Berouti, Schwartz, and Makhoul 1979):

Where is the estimated clean signal, is the noisy signal, and is the estimated noise spectrum.

**Dynamic Range Compression:**

The compression function follows standard audio processing practices (Katz 2002):

Where is the compression threshold and is the compression ratio.

**Normalization:**

Peak normalization ensures consistent amplitude levels:

## Onset Detection and Segmentation

Precise onset detection forms the foundation for rhythmic analysis, utilizing established algorithms in music information retrieval (Bello et al. 2005; Dixon 2006).

**Multi-Method Onset Detection:**

We combine multiple onset detection methods for robustness:

The combined onset strength function is:

Weight optimization follows Bello et al. (2005):

# MULTI-DIMENSIONAL FEATURE EXTRACTION AND ANALYSIS

## Time-Domain Feature Analysis

Time-domain features capture temporal characteristics essential for understanding talking drum communication patterns (McAulay and Quatieri 1986; Quatieri 2002):

**Root Mean Square Energy:**

**Zero-Crossing Rate:**

**Attack and Decay Time Analysis:**

Following established practices in audio analysis (Peeters 2004), attack time is calculated as:

Decay time is similarly computed:

## Frequency-Domain Analysis

**Fundamental Frequency Estimation**

We employ the YIN algorithm (De Cheveigné and Kawahara 2002) for robust pitch detection:

**Difference Function:**

**Cumulative Mean Normalized Difference:**

**Fundamental Frequency:**

where is the first minimum of below the threshold.

**Spectral Features**

Following standard practices in music informatics (Tzanetakis and Cook 2002), we extract:

**Spectral Centroid:**

**Spectral Rolloff:**

**Spectral Bandwidth:**

## Rhythmic Feature Extraction

**Mel-Frequency Cepstral Coefficients (MFCCs)**

Following Logan (2000), MFCCs are computed using:

**Harmonic-Percussive Separation**

We utilize the method by Fitzgerald (2010) to separate harmonic and percussive components, with the ratio:

**Inter-Onset Interval Analysis**

For rhythm analysis, we compute inter-onset intervals (London 2012):

**Normalized Pairwise Variability Index (nPVI)**

Following Grabe and Low (2002):

**Microtiming Deviation**

Timing accuracy relative to ideal grid positions:

## Cultural and Linguistic Feature Analysis

**Speech Surrogate Pattern Recognition**

This novel component analyzes drum-speech correspondence, building on work by Carrington (1949; Gleick 2011):

**Tonal Contour Analysis:**

Fundamental frequency is converted to semitones for tonal analysis:

**Tonal Movement:**

**Direction Change Rate:**

**Syllable Rate Analysis:**

Communication patterns are analyzed through syllable-like segmentation:

# MACHINE LEARNING FRAMEWORK

## Model Architecture Design

Our generative framework utilizes state-of-the-art deep learning architectures, following recent advances in generative modeling (Goodfellow, Bengio, and Courville 2016; Kingma and Welling 2013).

**Cultural Distance Metric**

We define cultural similarity between samples using:

where represents cultural feature with weight .

**Support Vector Machine Implementation**

For classification tasks, we employ SVM with RBF kernel (Cortes and Vapnik 1995):

Optimization objective:

Subject to constraints following Schölkopf et al. (2000).

**Deep Learning Architecture**

For generative modeling, we implement a multi-class neural network with cross-entropy loss:

# VALIDATION AND EVALUATION FRAMEWORK

## Cross-Validation Strategy

We employ stratified k-fold cross-validation following Kohavi et al. (1995):

where is the model trained excluding fold , and is the test set for fold .

**Confidence Intervals:**

Following standard statistical practices (Dietterich 1998):

## Performance Metrics

We evaluate model performance using established metrics in machine learning (Hastie, Tibshirani, and Friedman 2009; Bishop 2006):

**Classification Metrics:**

**Generation Quality Assessment**

For generative models, we employ Fréchet Audio Distance (Kilgour et al. 2018):

where and are the mean and covariance of real and generated samples respectively.

# ETHICAL CONSIDERATIONS AND CULTURAL SENSITIVITY

Our research adheres to established ethical frameworks in cultural research (Brown 2003; Seeger 2004), ensuring:

* Respectful engagement with traditional knowledge systems
* Proper attribution of cultural sources
* Community benefit and knowledge sharing
* Protection of sacred or sensitive cultural content

These considerations are integrated throughout the methodology, ensuring that technological advancement serves cultural preservation and understanding.

# IMPLEMENTATION TIMELINE AND RESOURCE ALLOCATION

The project follows a structured 24-week implementation schedule:

Project Implementation Timeline

| **Phase** | **Duration** | **Key Activities** | **Mathematical Models** | **Expected Deliverables** |
| --- | --- | --- | --- | --- |
| Phase 1 | Weeks 1-6 | Resource Collection | optimization | Curated dataset (500+ samples) |
| Phase 2 | Weeks 7-12 | Processing & Analysis | Feature extraction pipeline | Processed audio + features |
| Phase 3 | Weeks 13-18 | Model Development | ML/DL architectures | Trained models |
| Phase 4 | Weeks 19-24 | Validation & Testing | Performance metrics | Validated system |

# EXPECTED OUTCOMES AND CONTRIBUTIONS

This methodology is expected to produce:

1. A comprehensive, culturally-authenticated talking drums dataset (500+ samples)
2. Novel computational methods for traditional music analysis
3. AI models capable of generating culturally-appropriate rhythmic patterns
4. Frameworks for ethical AI development in cultural contexts
5. Open-source tools for ethnomusicological research

The research contributes to both computational musicology and cultural preservation, establishing new standards for technology-assisted cultural research.

# CONCLUSION

This methodology chapter presents a comprehensive, culturally-sensitive approach to developing AI systems for traditional music analysis and generation. The framework balances technical rigor with cultural authenticity, ensuring that computational advancement serves cultural understanding and preservation.

The systematic approach, combining rigorous mathematical frameworks with respectful cultural engagement, creates a model for future research at the intersection of technology and traditional knowledge. The detailed implementation timeline and comprehensive evaluation framework ensure reproducible, reliable results that advance both scientific understanding and cultural appreciation.

Arms, William Y. 2000. *Digital Libraries*. Cambridge, MA: MIT Press.

Barocas, Solon, Moritz Hardt, and Arvind Narayanan. 2019. “Fairness and Machine Learning.” *Fairmlbook.org*.

Barz, Gregory F., and Timothy J. Cooley. 2008. *Shadows in the Field: New Perspectives for Fieldwork in Ethnomusicology*. 2nd ed. New York, NY: Oxford University Press.

Bello, Juan Pablo, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler. 2005. “A Tutorial on Onset Detection in Music Signals.” *IEEE Transactions on Speech and Audio Processing* 13 (5): 1035–47.

Berouti, Michel, Richard Schwartz, and John Makhoul. 1979. “Enhancement of Speech Corrupted by Acoustic Noise” 4: 208–11.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Boll, Steven F. 1979. “Suppression of Acoustic Noise in Speech Using Spectral Subtraction.” *IEEE Transactions on Acoustics, Speech, and Signal Processing* 27 (2): 113–20.

Borgman, Christine L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.

Brown, Steven C. 2003. *Research Ethics in Ethnomusicology: Directions and Challenges for a New Century*. *Ethnomusicology*. Vol. 47. 2. University of Illinois Press.

Carrington, John F. 1949. *The Talking Drums of Africa*. London, UK: Carey Kingsgate Press.

Casey, Michael A., Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. 2008. “Content-Based Music Information Retrieval: Current Directions and Future Challenges.” *Proceedings of the IEEE* 96 (4): 668–96.

Cortes, Corinna, and Vladimir Vapnik. 1995. “Support-Vector Networks.” *Machine Learning* 20 (3): 273–97.

Creswell, John W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th ed. Thousand Oaks, CA: Sage Publications.

De Cheveigné, Alain, and Hideki Kawahara. 2002. “YIN, a Fundamental Frequency Estimator for Speech and Music.” *The Journal of the Acoustical Society of America* 111 (4): 1917–30.

Dietterich, Thomas G. 1998. “Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms.” *Neural Computation* 10 (7): 1895–1923.

Dixon, Simon. 2006. “Onset Detection Revisited,” 133–37.

Downie, J. Stephen. 2003. “Music Information Retrieval.” *Annual Review of Information Science and Technology* 37 (1): 295–340.

Fitzgerald, Derry. 2010. “Harmonic/Percussive Separation Using Median Filtering,” 246–53.

Font, Frederic, Gerard Roma, and Xavier Serra. 2013. “Freesound Technical Demo,” 411–12.

Fuentes, Marcelo, Daniel McEnnis, Jared Shifrin, and George Tzanetakis. 2007. “Computational Ethnomusicology.” In *Proceedings of the 8th International Conference on Music Information Retrieval*, 499–500. Austrian Computer Society.

Gleick, James. 2011. *The Information: A History, a Theory, a Flood*. New York, NY: Pantheon Books.

Goldsmith, Jeff. 2019. “Smithsonian Folkways: The Encyclopedia of World Music and Culture.” *Music Reference Services Quarterly* 22 (1): 43–48.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Cambridge, MA: MIT Press.

Grabe, Esther, and Ee Ling Low. 2002. “Durational Variability in Speech and the Rhythm Class Hypothesis.” *Laboratory Phonology* 7: 515–46.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York, NY: Springer Science & Business Media.

Hockey, Susan. 2004. *The History of Humanities Computing: An Outline*. *A Companion to Digital Humanities*. Oxford, UK: Blackwell Publishing.

Huron, David Brian. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Johnson, Peggy. 2018. *Fundamentals of Collection Development and Management*. 4th ed. Chicago, IL: American Library Association.

Kabal, Peter. 2002. “An Examination and Interpretation of ITU-r BS. 1387: Perceptual Evaluation of Audio Quality.” *TSP Lab Technical Report*, 1–89.

Katz, Bob. 2002. *Mastering Audio: The Art and the Science*. Burlington, MA: Focal Press.

Kilgour, Kevin, Mauricio Zuluaga, Dominik Roblek, and Matthew Sharifi. 2018. “Fréchet Audio Distance: A Reference-Free Metric for Evaluating Music Enhancement Algorithms,” 353–59.

Kingma, Diederik P., and Max Welling. 2013. “Auto-Encoding Variational Bayes.” *arXiv Preprint arXiv:1312.6114*.

Klapuri, Anssi, and Manuel Davy. 2006. *Signal Processing Methods for Music Transcription*. New York, NY: Springer Science & Business Media.

Kohavi, Ron et al. 1995. “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection” 2: 1137–45.

Kubik, Gerhard. 1999. *Africa and the Blues*. Jackson, MS: University Press of Mississippi.

Lee, Stuart D. 2002. “Collection Development in the Digital Age.”

Lerch, Alexander. 2012. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Hoboken, NJ: Wiley-IEEE Press.

Liang, Percy, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yuhui Zhang, et al. 2022. “Holistic Evaluation of Language Models.” *arXiv Preprint arXiv:2211.09110*.

Logan, Beth. 2000. “Mel Frequency Cepstral Coefficients for Music Modeling.”

London, Justin. 2012. *Hearing in Time: Psychological Aspects of Musical Meter*. 2nd ed. New York, NY: Oxford University Press.

McAulay, Robert, and Thomas Quatieri. 1986. *Speech Analysis/Synthesis Based on a Sinusoidal Representation*. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. Vol. 34. 4. IEEE.

Moretti, Franco. 2013. “Distant Reading.”

Müller, Meinard, and Sebastian Ewert. 2015. *Music Information Retrieval: Recent Developments and Applications*. Vol. 8. 2-3. Foundations; Trends in Information Retrieval.

Nettl, Bruno. 2005. *The Study of Ethnomusicology: Thirty-One Issues and Concepts*. 2nd ed. Urbana, IL: University of Illinois Press.

Orfanidis, Sophocles J. 2007. *Introduction to Signal Processing*. Upper Saddle River, NJ: Prentice Hall.

Peeters, Geoffroy. 2004. “A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project.”

Pohlmann, Ken C. 2005. *Principles of Digital Audio*. 5th ed. New York, NY: McGraw-Hill.

Post, Jennifer C. 2006. *Ethnomusicology: A Research and Information Guide*. New York, NY: Routledge.

Proffitt, Merrilee. 2005. *Preservation in the Digital World*. Washington, DC: Council on Library; Information Resources.

Quatieri, Thomas F. 2002. *Discrete-Time Speech Signal Processing: Principles and Practice*. Upper Saddle River, NJ: Prentice Hall PTR.

Rice, Timothy. 2014. *Ethnomusicology: A Very Short Introduction*. Oxford, UK: Oxford University Press.

Rycroft, David K. 1961. “The Guitar Improvisations of Mwenda Jean Bosco.” *African Music* 2 (4): 81–98.

Schedl, Markus, Emilia Gómez, and Julián Urbano. 2014. *Music Information Retrieval: Recent Developments and Applications*. Vol. 8. 2-3. Foundations; Trends in Information Retrieval.

Schölkopf, Bernhard, Alex J. Smola, Robert C. Williamson, and Peter L. Bartlett. 2000. *New Support Vector Algorithms*. *Neural Computation*. Vol. 12. 5. MIT Press.

Seaman, David. 2011. *Digital Library Development: The Collections Perspective*. Washington, DC: Council on Library; Information Resources.

Seeger, Anthony. 2004. “Traditional Music and Music-Cultures: Attempting a Definition.” *British Forum for Ethnomusicology* 13 (1): 59–68.

Serra, Xavier, Perfecto Magas, Emmanouil Benetos, Magdalena Chudy, Simon Dixon, Arthur Flexer, Emilia Gómez, Perfecto Herrera, Joan Serra, and Gerhard Widmer. 2013. “Roadmap for Music Information Research.” *Creative Commons BY-NC-ND 3.0 License*.

Temperley, David. 2007. *Music and Probability*. Cambridge, MA: MIT Press.

Terras, Melissa. 2015. *Digitisation Perspectives*. London, UK: Facet Publishing.

Thiede, Thilo, William C. Treurniet, Roland Bitto, Christian Schmidmer, Thomas Sporer, John G. Beerends, and Catherine Colomes. 2000. “PEAQ-the ITU Standard for Objective Measurement of Perceived Audio Quality.” *Journal of the Audio Engineering Society* 48 (1/2): 3–29.

Titon, Jeff Todd, Timothy J. Cooley, David Locke, David P. McAllester, and Anne K. Rasmussen. 2017. *Worlds of Music: An Introduction to the Music of the World’s Peoples*. 6th ed. Boston, MA: Cengage Learning.

Tzanetakis, George, and Perry Cook. 2002. “Musical Genre Classification of Audio Signals.” *IEEE Transactions on Speech and Audio Processing* 10 (5): 293–302.

Watkinson, John. 2001. *The Art of Digital Audio*. 3rd ed. Oxford, UK: Focal Press.

Zölzer, Udo. 2011. *DAFX: Digital Audio Effects*. 2nd ed. Chichester, UK: John Wiley & Sons.

Zorich, Diane M. 2012. “Transitioning to a Digital World: Art History, Its Research Centers, and Digital Scholarship.”