Extraktion und Vergleich musikalischer Rhythmen aus Aufnahmen von Klanghölzern

Günther Wirsching und Carolin Straub

Katholische Universität Eichstätt-Ingolstadt
guenther.wirsching@ku.de

Kurzfassung: Die Güte der menschlichen Reproduktionsfähigkeit kurzer musikalischer Rhythmen von Klanghölzern wird in dieser Arbeit mathematisch in einer Weise modelliert, dass die Ergebnisse für psychologische Untersuchungen anwendbar sind. Zur Lokalisierung der Onsets verwenden wir gleitende Durchschnitte des quadrierten Signals sowie Maximumsbestimmungen auf geeignet ausgewählten Signalsegmenten. Dadurch wird jedem Ton genau ein Zeitpunkt – der Zeitpunkt des entsprechenden Onsets – zugewiesen. Wir erhalten so zu jedem Rhythmus einen Punkt in einem reellen Vektorraum, dessen Dimension der Anzahl der Töne entspricht. Wir definieren den Reproduktionsfehler als den euklidischen Abstand des Ausgangsrhythmus zur Bahn der Reproduktion, die durch die Operation der von Dilatationen und Translationen erzeugten Gruppe auf diesem Vektorraum gegeben ist. Dieses Vorgehen modelliert den psychologischen Sachverhalt, dass Rhythmen mit leicht unterschiedlichen Tempi als derselbe Rhythmus wahrgenommen werden (McAuley, 2010). Unterscheidet sich die Tonanzahl des Ausgangsrhythmus von derjenigen der Reproduktion, so werden entsprechend viele Töne des Rhythmus mit der geringeren Tonanzahl als mehrere gleichzeitige Töne betrachtet.

Die gesamte Vorgehensweise realisiert einen präzisen Vergleich der zeitlichen Abweichung zweier musikalischer Rhythmen vor dem Hintergrund musikpsychologischer Anwendung. Damit ist es möglich, bei derartigen Experimenten, die v.a. in der Musikpsychologie durchgeführt wurden (z.B. Tillmann et al., 2011; Fitch & Rosenfeld, 2007; Bernardi et al., 2013), Messergebnisse mit höherer mathematischer Präzision zu gewinnen. Zur Verfeinerung der Modellierung wäre eine zusätzliche Betrachtung der Amplitude von Interesse, um Akzente im Ausgangsrhythmus mit berücksichtigen zu können.

1 Problemformulierung und Voraussetzungen

In der psychologischen Literatur ist der Begriff *Rhythmus* meist durch die Verarbeitungsprozesse im Gedächtnis definiert (z.B. Rammsayer, 2000; Snyder, 2000). Da wir jedoch keine Gedächtnisprozesse sondern den Gegenstand Rhythmus betrachten, wird in Anlehnung an McAuley (2010) folgende musikpsychologisch orientierte, operationalisierte Rhythmusdefinition verwendet (Straub, 2015):

> Rhythmus ist die zeitliche, iterierende Abfolge von Tönen und Stille, wobei ein Ton ein Audiosignal mit nur einer Frequenz (Werner, 2012) ist und Stille nur Signale beinhaltet, die unter 20 Hz oder über 20 kHz liegen.

Die Einschränkung der Frequenzen ist dadurch motiviert, dass solche Frequenzen vom menschlichen Ohr nicht wahrnehmbar sind (Rousseau & Saint-Aubin, 2012).

Das Ziel dieser Untersuchung ist eine Quantifizierung der Güte einer Reproduktion. Zum Vergleich verschiedener Reproduktionen eines vorgegebenen Rhythmus hinsichtlich ihrer zeitlichen Abweichungen definieren wir ein Maß für den Reproduktionsfehler. Unsere Analyse basiert auf den folgenden Voraussetzungen:

1. *Nur ein Set von Klanghölzern* sowohl für den vorgegebenen Rhythmus als auch zum Reproduzieren. Den Versuchspersonen wurde die richtige Handhabung der Klanghölzer vor dem Experiment erläutert.

2. *Vernachlässigung der Lautstärke beim Rhythmusvergleich.* Die Amplitude der Signale wird nur für die Extraktion der Onsets verwendet und ansonsten ignoriert. Bei der Aufnahme der Vorgaberhythmen wurde jedoch darauf geachtet, dass möglichst alle Töne gleich laut sind und keine Akzente entstehen.

3. *Konstanthalten des Tempos bei 120 bpm.* Das entspricht sowohl dem durchschnittlichen spontan produzierten Tempo als auch dem in der Wahrnehmung präferierten Tempo (Fischinger & Kopiez, 2009).

4. *Abtastrate 44.1 kHz.* Das ist die übliche Abtastrate bei Tonaufnahmen. Historisch lässt sie sich auf eine weit verbreitete Verarbeitung von Videosignalen zurückführen (Rousseau & Saint-Aubin, 2012).

5. *Nur Viervierteltakt, und nur ein Takt.* Der Grund hierfür ist, dass die Verarbeitung im Kurzzeitgedächtnis mit einer Aufrechterhaltung von durchschnittlich drei bis fünf Sekunden (Snyder, 2000) im Zentrum des Anwendungsproblems steht.

6. *Mindestens drei und höchstens neun Töne.* Diese Einschränkungen verbessern die numerische Handhabbarkeit.

7. *Geringhalten des Umgebungsrauschens.* Bei den Aufnahmen wurde geachtet, die Hintergrundgeräusche möglichst gering zu halten.

8. *Vernachlässigung der Bekanntheit.* Da die Rhythmen nur aus einem Takt bestehen, und eine konstante Tonhöhe haben, wird angenommen, dass diese im Wesentlichen unbekannt sind.

9. *Beschränkung der Stichprobe auf Studentinnen und Studenten.* Als Stichprobe wurden Studenten und Studentinnen der Katholischen Universität Eichstätt-Ingolstadt herangezogen. Dadurch wird das Alter eingeschränkt auf junge Erwachsene (Durchschnittsalter 24.4 Jahre). Unter den Versuchspersonen befinden sich sowohl aktive Musiker als auch Menschen mit wenig oder kaum Musikerfahrung. Der Einfluss der Kultur wurde konstant gehalten, da alle Probanden in Deutschland aufgewachsen waren. Dies ist wichtig, da die Konstruktion der Rhythmen an westlichen Rhythmusstrukturen orientiert ist.

2 Technische Analyse

In diesem Abschnitt wird das hergeleitete Modell mathematisch analysiert.

2.1 Lokalisierung der Onsets

Mit *Onset* wird der genaue Beginn eines Tons bezeichnet. Aufgrund von Einschwing- und Ausklangvorgängen ist dieser im Allgemeinen schwer bestimmbar (Fricke & Louven, 2009). Da die Zeitabstände zwischen den Onsets im Zentrum dieser Arbeit stehen, ist es erforderlich, jeden Onset auf dieselbe Art und Weise zu bestimmen. Zur Lokalisierung der Onsets verwenden wir das quadrierte Signal, das der Signalenergie entspricht (Meyer, 2014). Jedes Signal wird gemäß Collecchia (2012) in die Umrisse der einzelnen Töne – „Amplitude Envelopes" – unterteilt. Als Onset eines Tons wird das Maximum des entsprechenden Umrisses verwendet.

2.1.1 Gleitende Durchschnitte

Da das Signal stark oszilliert, werden gleitende Durchschnitte gemäß Schlittgen (2008) berechnet: Es sei

$$x = (x_0, \ldots, x_{n-1})^T \in \mathbb{R}^n$$

der Vektor mit den n quadrierten Abtastwerten eines aufgezeichneten Signals. Zur Bildung der gleitenden Durchschnitte wählen wir als Intervallänge

$d :=$ Anzahl der Abtastwerte pro halber Wellenlänge des Haupttons der Klanghölzer.

Damit wird sichergestellt, dass die Durchschnitte jeweils näherungsweise über eine komplette Periode des quadrierten Abtastsignals gebildet werden. Im Fall unserer Klanghölzer lag der Hauptton bei ca. 50 Hz. Bei einer Abtastfrequenz von 44.1 kHz ergibt sich $d \approx 441$; im Folgenden setzen wir $d := 441$. Damit ergibt sich zur Berechnungen der gleitenden Durchschnitte

$$y_i := \frac{1}{441} \sum_{u=i}^{i+440} x_u, \qquad i \in \{0, \ldots, n-441\}.$$

2.1.2 Extraktion der Onsets

Zur Extraktion der Töne aus den gleitenden Durchschnitten $(y_0, \ldots, y_{n-441})$ zerlegen wir diese in entsprechend viele Segmente mittels zweier Schranken. Die erste Schranke $s > 0$ filtert diejenigen Abtastwerte heraus, die weit genug oberhalb des Hintergrundrauschens liegen. Bei unseren Daten hat sich

$$s := \frac{1}{45} \cdot \max\{y_0, \ldots, y_{n-441}\}$$

als günstig erwiesen. Wir erhalten damit einen Vektor

$$\tilde{y}_i := \begin{cases} y_i & \text{falls } y_i \geqslant s, \\ 0 & \text{sonst,} \end{cases} \qquad \text{für } i \in \{0, \ldots, n-441\}.$$

Die zweite Schranke $t > 0$ in Abszissen-Richtung verhindert, dass ein Onset mehrfach gezählt wird. Technisch ist das so realisiert, dass zwei Indizes i, j nur dann zu verschiedenen Onsets gezählt werden, wenn $|i-j| \geqslant t$ ist. In unserem Fall hat sich

$$t := 4 \cdot d = 1764$$

als günstig erwiesen. Als Grundlage für die Bestimmung der Onsets extrahieren wir aus dem Vektor $\tilde{y} = (\tilde{y}_0, \ldots, \tilde{y}_{n-441})^T$ Bereiche $(\tilde{y}_i, \ldots, \tilde{y}_{i+k})^T$ mit den Bedingungen

a) $\tilde{y}_i \neq 0 \neq \tilde{y}_{i+k}$,

b) höchstens $t-1$ aufeinanderfolgende Werte $\tilde{y}_j, \tilde{y}_{j+1}, \ldots$ dürfen verschwinden.

Gemäß dieser Konstruktion entspricht ein *Ton* der Klanghölzer genau einem solchen Bereich. Wir erhalten so zu einem Signal mit r Tönen eine Folge $B_0, \ldots, B_{r-1}$ aufeinanderfolgender Bereiche, deren Indexmengen sich nicht überlappen. Als *Onset* des Tons zu einem Bereich $B_j := (\tilde{y}_i, \ldots, \tilde{y}_{i+k})^T$ definieren wir den kleinsten Index des Maximums dieses Bereichs,

$$m_j := \arg\max(B_j) \quad \in \quad \{i, \ldots, i+k\}.$$

Dadurch gelingt es, jedem Ton genau einen Zeitpunkt – den Zeitpunkt des entsprechenden Onsets – zuzuweisen. Wir erhalten zu jeder Aufnahme A einen Vektor

$$m := (m_0, \ldots, m_{r_A-1})^T \in \mathbb{R}^{r_A},$$

dessen Dimension r_A die Anzahl der Töne in der Aufnahme ist. Ebenso extrahieren wir aus einem vorgegebenen Rhythmus, der aus r_0 Tönen besteht, einen Punkt

$$v := (v_0, \ldots, v_{r_0-1})^T \in \mathbb{R}^{r_0}.$$

2.2 Vergleich zweier Rhythmen

Wir vergleichen nun einen vorgegebenen Rhythmus $v \in \mathbb{R}^{r_0}$ mit einer Reproduktion $m \in \mathbb{R}^{r_A}$, indem wir zunächst den Fall $r := r_0 = r_A$ untersuchen, und anschließend die anderen Fälle darauf zurückführen.

2.2.1 Gleiche Tonanzahl

Sei $v \in \mathbb{R}^r$ ein vorgegebener Rhythmus. Um bei der Abstandsbestimmung von Verzögerungen, Streckungen und Stauchungen abzusehen, betrachten wir die Gruppe $\mathbb{G}$, die von den Translationen $t \mapsto t + \beta$ der Zeitachse (Verzögerungen) und den Dilatationen $t \mapsto \alpha \cdot t$ (Streckungen und Stauchungen) erzeugt wird. Diese Gruppe ist das semidirekte Produkt der multiplikativen Gruppe $(\mathbb{R}^\times, \cdot, 1)$, mit der Bezeichnung $\mathbb{R}^\times := \mathbb{R} \setminus \{0\}$, und der additiven Gruppe $(\mathbb{R}, +, 0)$. Sie besteht aus der Grundmenge

$$\mathbb{G} = \{(\alpha, \beta) : \alpha \in \mathbb{R}^\times, \beta \in \mathbb{R}\} = \mathbb{R}^\times \ltimes \mathbb{R}$$

mit der Verknüpfung

$$(\alpha, \beta) \odot \left(\widetilde{\alpha}, \widetilde{\beta}\right) = \left(\alpha\widetilde{\alpha}, \alpha\widetilde{\beta} + \beta\right).$$

Die Gruppe $\mathbb{G}$ operiert auf $\mathbb{R}^r$ durch

$$(\alpha, \beta) \cdot (m_0, \ldots, m_{r-1})^T := (\alpha m_0 + \beta, \ldots, \alpha m_{r-1} + \beta)^T.$$

Der *Reproduktionsfehler* von m bzgl. v ist nun definiert als der euklidische Abstand von v zur Bahn von m bzgl. dieser Operation:

$$\varepsilon_v(m) := \min\{\|v - (\alpha, \beta) \cdot m\|_2 : (\alpha, \beta) \in \mathbb{G}\}.$$

Für die euklidische Norm eines Vektors $x \in \mathbb{R}^r$ gilt die Formel

$$\|x\|_2 = \sqrt{\sum_{i=0}^{r-1} x_i^2}.$$

Weil die Funktion

$$f(\alpha,\beta) := \|v-(\alpha,\beta)\cdot m\|_2$$

wegen der Dreiecksungleichung konvex ist, besitzt sie höchstens ein Minimum. Zur Bestimmung des Minimums ist es günstiger, stattdessen den quadrierten Abstand

$$g(\alpha,\beta) := \|v-(\alpha,\beta)\cdot m\|_2^2 = \sum_{i=0}^{r-1}(v_i-\alpha m_i-\beta)^2$$

zu verwenden. Das Nullsetzen des Gradienten von g führt auf das lineare Gleichungssystem

$$\nabla g(\alpha_o,\beta_o)=0 \quad\Longleftrightarrow\quad \underbrace{\begin{pmatrix}\|m\|_1 & r\\ \|m\|_2^2 & \|m\|_1\end{pmatrix}}_{:=A}\cdot\begin{pmatrix}\alpha_o\\ \beta_o\end{pmatrix} = \underbrace{\begin{pmatrix}\|v\|_1\\ v^Tm\end{pmatrix}}_{:=b}, \tag{1}$$

wobei

$$\|v\|_1 = \sum_{i=0}^{r-1}|v_i| = \sum_{i=0}^{r-1}v_i, \qquad \|m\|_1 = \sum_{i=0}^{r-1}|m_i| = \sum_{i=0}^{r-1}m_i.$$

(Man beachte $v_i \geqslant 0$ und $m_i \geqslant 0$.) Es gilt

$$-\det A = r\|m\|_2^2 - \|m\|_1^2 = r\sum_{i=0}^{r-1}m_i^2 - \left(\sum_{i=0}^{r-1}m_i\right)^2 = \sum_{i=0}^{r-2}\sum_{j=i+1}^{r-1}(m_i-m_j)^2 > 0,$$

wie sich durch Nachrechnen ergibt. Daher ist das lineare Gleichungssystem (1) eindeutig lösbar.

2.2.2 Ungleiche Tonanzahlen

Sind die Tonanzahlen r_0 und r_A der beiden zu vergleichenden Rhythmen verschieden, so wird die oben beschriebene Vorgehensweise in einen kombinatorischen Algorithmus eingebettet. Hierfür bestimmen wir zunächst das Maximum $r := \max\{r_0, r_A\}$.

Im Fall $r_A < r_0$ werden soviele Komponenten des Vektors von $m = (m_0,\ldots,m_{r_A-1})$ mehrfach aufgezählt, dass ein Vektor

$$\widetilde{m} := (\widetilde{m}_0,\ldots,\widetilde{m}_{r-1})$$

entsteht. So kann z. B. für $r_A = 3$ und $r = 4$ ein Vektor m mit den Komponenten

$$m_0 \quad < \quad m_1 \quad < \quad m_2$$

zu einem Vektor $\widetilde{m}$ mit den Komponenten

$$\widetilde{m}_0 \quad < \quad \widetilde{m}_1 \quad = \quad \widetilde{m}_2 \quad < \quad \widetilde{m}_3$$

erweitert werden. Um die Anzahl der möglichen Erweiterungen von r_A Komponenten auf r_0 Komponenten zu ermitteln, stellen wir fest, dass in die Ungleichungskette der r_A Komponenten des Vektors m zur Erweiterung auf die r_0 Komponenten des Vektors $\widetilde{m}$ insgesamt $r_0 - r_A$ mal ein „=“-Zeichen eingefügt werden muss. Aus der Kombinatorik ist bekannt, dass die Anzahl der Möglichkeiten, $r_0 - r_A$ Objekte („=“-Zeichen) aus $r_0 - 1$ Objekten (Paaren $(\widetilde{m}_i, \widetilde{m}_{i+1})$) auszuwählen, durch den Term

$$r_{0,A} := \binom{r_0-1}{r_0-r_A} = \binom{r_0-1}{r_A-1}$$

gegeben ist. Unser Algorithmus berechnet nun für jede dieser $r_{0,A}$ Möglichkeiten wie oben beschrieben den Reproduktionsfehler von $\widetilde{m}$ bzgl. v, und wählt dann eine Möglichkeit mit minimalem Reproduktionsfehler aus.

Im Fall $r_0 < r_A$ ist die Vorgehensweise analog; hier ist v der „kürzere“ Vektor, und es sind $r_A - r_0$ „=“-Zeichen einzufügen, was analog zu oben auf

$$r_{A,0} := \binom{r_A - 1}{r_0 - 1}$$

verschiedene Arten geht. Als Reproduktionsfehler nehmen wir auch in diesem Fall das Minimum der Reproduktionsfehler bzgl. der verschiedenen Erweiterungen.

2.3 Zur numerischen Zuverlässigkeit

Um herauszufinden, wie zuverlässig eine numerische Lösung von (1) ist, berechnen wir die Kondition der Matrix A bzgl. der Norm

$$\|A\|_\infty := \max\{\|m\|_1 + r, \|m\|_2^2 + \|m\|_1\} = \|m\|_2^2 + \|m\|_1.$$

Damit gilt (Schwarz & Köckler, 2011)

$$\operatorname{cond}(A) = \|A\|_\infty \cdot \|A^{-1}\|_\infty = \frac{(\|m\|_2^2 + \|m\|_1)^2}{|\det A|} = \frac{(\|m\|_2^2 + \|m\|_1)^2}{\sum_{i=0}^{r-2}\sum_{j=i+1}^{r-1}(m_i - m_j)^2}. \tag{2}$$

Wir schätzen zunächst den Nenner dieses Bruchs nach unten ab. Bei unseren Daten ist $r \geqslant 3$. Setzt man $a := m_0 < b := m_1 < c := m_{r-1}$, dann folgt

$$\sum_{i=0}^{r-2}\sum_{j=i+1}^{r-1}(m_i - m_j)^2 \geqslant (a-c)^2 + (a-b)^2 + (b-c)^2 \geqslant \frac{3}{2}(a-c)^2.$$

Weil der Mindestabstand zwischen dem ersten und dem letzten Ton in unseren Daten mindestens 1.0 s, also 44 100 Abtastwerte, beträgt, folgt daraus für den Nenner

$$\sum_{i=0}^{r-2}\sum_{j=i+1}^{r-1}(m_i - m_j)^2 \geqslant \frac{3}{2}\cdot(44\,100)^2 \geqslant 1.95\cdot 10^9. \tag{3}$$

Zur Abschätzung des Zählers von (2) schätzen wir den Ausdruck $\|m\|_2^2 + \|m\|_1$ nach oben ab, indem wir die maximale Gesamtdauer x und den minimalen Abstand zweier Töne α berücksichtigen. Da die Abweichung der Tonzahlen zwischen Vorlage und Reproduktion in unserem Datensatz höchstens 1 beträgt, erhalten wir damit die folgenden Abschätzungen für die Onsets:

$$m_{r-1} \leqslant x - \alpha, \qquad m_{r-k} \leqslant x - (k-1)\alpha \quad \text{für } k \in \{2,\ldots,r-1\}.$$

In unseren Daten ist die Gesamtdauer kleiner als 2.5 s und die kürzeste reproduzierte Note entspricht knapp 78 ms, also

$$x = 110\,250 \text{ Abtastwerte}, \qquad \alpha = 3\,421 \text{ Abtastwerte}.$$

Daraus ergibt sich die folgende Abschätzung:

$$\|m\|_2^2 + \|m\|_1 \leqslant 1.24\cdot 10^{11}.$$

Mit (2) und (3) erhalten wir die Abschätzung der Kondition:

$$\operatorname{cond}(A) \leqslant \frac{\left(1.24\cdot 10^{11}\right)^2}{1.95\cdot 10^9} \approx 7.9\cdot 10^{12}.$$

Weil die Maschinengenauigkeit bei Python $\varepsilon \approx 10^{-16}$ beträgt, folgt daraus für den Fehler bei exakter Rechnung

$$\operatorname{cond}(A)\varepsilon \quad \lessapprox \quad 8\cdot 10^{-4}.$$

Daher sollten bei der Lösung des linearen Gleichungssystems (1) ungefähr zwei bis drei Dezimalstellen korrekt sein, was für unsere Anwendungen ausreicht.

3 Psychologischer Anwendungsbezug

Mit Hilfe der bestimmten minimalen Abweichungen ist es möglich, Reproduktionen hinsichtlich ihrer Güte zu vergleichen. Exemplarisch wird hier ein Schlagzeuger mit vorwiegend guten bis sehr guten Reproduktionsergebnissen (vgl. Tabelle 1) vorgestellt. Abbildung 1 zeigt diese Ergebnisse.

Tabelle 1 - Abweichungen A aller 15 Reproduktionen eines Schlagzeugers von den entsprechenden Ausgangsrhythmen (in ms).

Rhythmen	**1**	**2**	**3**	**4**	**5**	**Durchschnitt**
A	21.2	11.4	639.9	12.2	80.9	153.1
Rhythmen	**6**	**7**	**8**	**9**	**10**	**Durchschnitt**
A	15.0	726.5	12.4	18.8	0.44	155.4
Rhythmen	**11**	**12**	**13**	**14**	**15**	**Durchschnitt**
A	32.1	46.3	32.2	30.4	82.4	44.7

Notiz. Gesamtvarianz $s^2 = 49\,951$, gesamte mittlere Abweichung $\bar{x} = 117.7$.

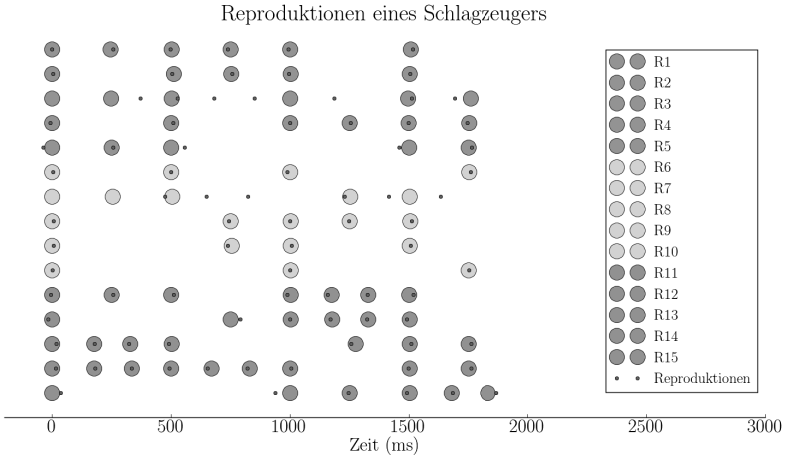

Abbildung 1 - Extrahierte Zeitwerte aller Rhythmen (verschiedenfarbige große Punkte) mit den minimal abweichenden Reproduktionen eines Schlagzeugers (rote kleine Punkte).

Alle Rhythmen bis auf drei und sieben wurden gut bis sehr gut reproduziert. Dies wird deutlich durch kleine minimale Abweichungen vom entsprechenden Vergleichsrhythmus (vgl. Tabelle 1). In Abbildung 1 sind die Töne der Reproduktionen (kleine rote Punkte) dann nahe der entsprechenden Töne der Vergleichsrhythmen (große Punkte). Bei Rhythmus drei und sieben ist eine deutlich größere minimale Abweichung zu beobachten. Graphisch äußert sich das durch eine entsprechend größere Entfernung der Reproduktion vom Vergleichsrhythmus. Dieses schlechtere Ergebnis hinsichtlich der Reproduktionsgüte ist dadurch erklärbar, dass beide Reproduktionen einen Ton mehr aufweisen als der jeweilige Ausgangsrhythmus.

Literatur

Bernardi, N. F., Schories, A., Jabusch, H.-C., Colombo, B. & Altenmüller, E. (2013). Mental practice in music memorization: An ecological-empirical study. *Music Perception*, *30* (3), 275-290.

Collecchia, R. (2012). *Numbers & notes: An introduction to musical signal processing*. Portland: Perfectly Scientic Press.

Fischinger, F. & Kopiez, R. (2009). Wirkungsphänomene des Rhythmus. In H. Bruhn, R. Kopiez & A. C. Lehmann (Hrsg.), *Musikpsychologie. Das neue Handbuch.* Reinbek: Rowohlt.

Fitch, W. T. & Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Perception*, *25* (1), 43-58.

Fricke, J. P. & Louven, C. (2009). Psychoakustische Grundlagen des Musikhörens. In H. Bruhn, R. Kopiez & A. C. Lehmann (Hrsg.), *Musikpsychologie. Das neue Handbuch* (S. 413-436). Reinbek: Rowohlt.

McAuley, J. D. (2010). Tempo and Rhythm. In M. R. Jones, R. R. Fay & A. N. Popper (Hrsg.), *Music Perception* (S. 165-200). New York: Springer Science+Business Media.

Meyer, M. (2014). *Signalverarbeitung*. Wiesbaden: Springer Vieweg.

Rammsayer, T. (2000). Zeitwahrnehmung und Rhythmuswahrnehmung. In K. Müller & G. Aschersleben (Hrsg.), *Rhythmus. Ein interdisziplinäres Handbuch* (S. 83-106). Bern: Huber.

Rousseau, C. & Saint-Aubin, Y. (2012). *Mathematik und Technologie*. Berlin Heidelberg: Springer-Verlag.

Schlittgen, R. (2008). *Einführung in die Statistik. Analyse und Modellierung von Daten*. München: Oldenbourg Verlag.

Schwarz, H. R. & Köckler, N. (2011). *Numerische Mathematik.* Wiesbaden: Vieweg + Teubner Verlag.

Snyder, B. (2000). *Music and memory*. Cambridge: The MIT Press.

Straub, C. (2015). *Extraktion und Vergleich musikalischer Rhythmen aus Aufnahmen von Klanghölzern.* (Bachelorarbeit)

Tillmann, B., Stevens, C. & Keller, P. E. (2011). Learning of timing patterns and the development of temporal expectations. *Psychological Research*, *75* (3), 243-258.

Werner, M. (2012). *Digitale Signalverarbeitung mit MATLAB*. Wiesbaden: Vieweg + Teubner Verlag.