

Extraktion von Parametern zur Klangsynthese aus aufgenommenen Signalen

Forschungspraktikum

Lehrstuhl für Multimediakommunikation und Signalverarbeitung

Universität Erlangen-Nürnberg

Richard Füg

14. Oktober 2013

1 Einleitung

Durch Physical Modeling mit der Functional Transformation Method (FTM) ist es möglich die Saitenschwingungen eines Instruments allein durch die physikalischen Parameter der Saite und ihrer Anregung mit vertretbarer Rechenkomplexität in Echtzeit zu simulieren. Für viele Instrumente und Materialien sind diese Parameter bekannt und können entsprechenden Nachschlagwerken entnommen werden. Möchte man allerdings den Klang eines speziellen (z.B. mit sehr alten Saiten bestückten) oder sehr exotischen (z.B. Kanun) Instruments simulieren, ist man auf aufwändige Messungen der physikalischen Parameter angewiesen. Um dieses Problem zu umgehen, wurde während des Forschungspraktikums eine Methode entwickelt die zur FTM Synthese einer Saitenschwingung notwendigen Parameter aus einer Aufnahme einer realen Saitenschwingung zu extrahieren. Der prinzipielle Vorgang ist in Abbildung 1 dargestellt.

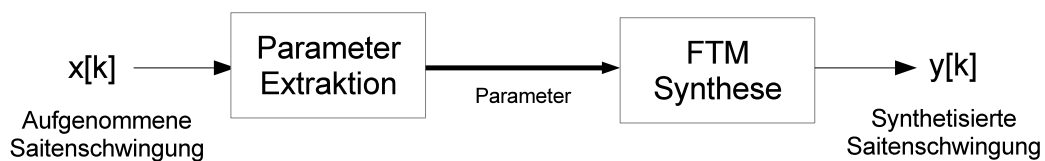


Abbildung 1: Prinzipieller Aufbau der Problemstellung

2 Überblick über die FTM Synthese

In diesem Abschnitt wird ein kurzer Überblick über die wesentlichen Aspekte der FTM Synthese einer Saitenschwingung gegeben, die für die spätere Parameterextraktion relevant sind. Eine vollständige Beschreibung findet sich in [2]. Die Notation in dieser Arbeit orientiert sich soweit möglich an dieser Darstellung.

Das der FTM Synthese zu Grunde liegende Saitenmodell wird beschrieben durch die partielle Differenzialgleichung

$$\rho A \frac{\partial^2 y}{\partial t^2} + EI \frac{\partial^4 y}{\partial x^4} - T_s \frac{\partial^2 y}{\partial x^2} + d_1 \frac{\partial y}{\partial t} + d_3 \frac{\partial^3 y}{\partial t \partial x^2} = f_e(x, t), \quad 0 < x < l \quad (1)$$

wobei x die Ortsvariable (beschränkt durch 0, den Anfang der Saite, und l , die Länge der Saite), t die Zeitvariable, y die Auslenkung der Saite und $f_e(x, t)$ die Kraftdichte der Anregung repräsentiert. ρ die Dichte, A die Querschnittsfläche, E der Elastizitätsmodul, I das Trägheitsmoment, T_s die Spannung, d_1 die frequenzunabhängige und d_2 die frequenzabhängige Dämpfung stellen die konstanten Parameter der Saite dar. Im Folgenden werden diese Konstanten zu

$$a_{02} = -\frac{T_s}{\rho A}, \quad a_{04} = \frac{EI}{\rho A}, \quad a_{10} = \frac{d_1}{\rho A}, \quad a_{12} = \frac{d_3}{\rho A}, \quad b_0 = \frac{1}{\rho A}$$

zusammengefasst. Mit Hilfe der Laplace Transformation und der Sturm-Liouville Transformation (STL) kann man aus dieser Differentialgleichung eine multidimensionale Übertragungsfunktion mit der diskreten Ortsfrequenzvariable μ und der kontinuierlichen komplexen Frequenzvariable s ableiten:

$$\bar{G}_e(\mu, s) = \frac{b_0}{s^2 + c_1(\mu)s + c_2(\mu)} = \frac{b_0}{(s - s_{\infty\mu})(s - s_{\infty\mu}^*)}. \quad (2)$$

Das Nennerpolynom lässt sich dabei entweder durch die Variablen

$$c_1(\mu) = a_{10} - a_{12}\gamma_\mu^2 \quad (3)$$

$$c_2(\mu) = a_{04}\gamma_\mu^4 - a_{02}\gamma_\mu^2 \quad (4)$$

$$\gamma_\mu = \frac{\mu\pi}{l} \quad (5)$$

oder in Linearfaktoren zerlegt mit

$$s_{\infty\mu} = \sigma_\mu + j\omega_\mu \quad (6)$$

$$\sigma_\mu = -\frac{1}{2}c_1(\mu) = -\frac{1}{2}(a_{12}\gamma_\mu^2 - a_{10}) \quad (7)$$

$$\omega_\mu^2 = \left[a_{04} - \frac{1}{2}a_{12}^2 \right] \gamma_\mu^4 + \left[-a_{02} + \frac{1}{2}(a_{10}a_{12})^2 \right] \gamma_\mu^2 - \left[\frac{1}{2}a_{10} \right]^2 \quad (8)$$

darstellen. Um die spätere Darstellung zu vereinfachen, können Gleichung (7) und (8) mit den modifizierten Konstanten

$$\tilde{a}_{02} = -\frac{T_s}{\rho A} \cdot \left(\frac{\pi}{l}\right)^2, \quad \tilde{a}_{04} = \frac{EI}{\rho A} \cdot \left(\frac{\pi}{l}\right)^4, \quad \tilde{a}_{12} = \frac{d_3}{\rho A} \cdot \left(\frac{\pi}{l}\right)^2$$

zu

$$\sigma_\mu = -\frac{1}{2}(\tilde{a}_{12}\mu^2 - a_{10}) \quad (9)$$

$$\omega_\mu^2 = \left[\tilde{a}_{04} - \frac{1}{2}a_{12}^2 \right] \mu^4 + \left[-\tilde{a}_{02} + \frac{1}{2}(a_{10}\tilde{a}_{12})^2 \right] \mu^2 - \left[\frac{1}{2}a_{10} \right]^2 \quad (10)$$

umgeformt werden.

Zur Realisierung dieser Übertragungsfunktion mit einem Computer wird die Frequenzvariable s mit Hilfe der impulsinvarianten Transformation in den Z-Bereich transformiert. Dies führt zur diskreten multidimensionalen Übertragungsfunktion

$$\overline{G}_e^d(\mu, z) = \frac{b_1^d(\mu)z}{z^2 + c_1^d(\mu)z + c_0^d(\mu)}. \quad (11)$$

wobei die Koeffizienten durch

$$b_1^d(\mu) = T b_0 \frac{\sin(\omega_\mu T)}{(\omega_\mu T)} \quad (12)$$

$$c_1^d(\mu) = -2 \exp(\sigma_\mu T) \cos(\omega_\mu T) \quad (13)$$

$$c_0^d(\mu) = \exp(2\sigma_\mu T) \quad (14)$$

gegeben sind und $f_s = \frac{1}{T}$ der Abtastfrequenz entspricht. Die inverse Z-Transformierte von $\overline{G}_e^d(\mu, z)$

$$\overline{g}_e^d(\mu, k) = \frac{b_0 T}{\Omega_\mu} e^{\sigma_\mu^d (k-1)} \sin(\Omega_\mu k), \quad k \geq 1 \quad (15)$$

$$\sigma_\mu^d = \sigma_\mu T, \quad \Omega_\mu = \omega_\mu T \quad (16)$$

stellt eine diskrete, mit der Abklingrate σ_μ^d abklingende Sinusschwingung der Frequenz Ω_μ dar.

Das synthetisierte Signal $y[k]$ erhält man nun indem man das Produkt

$$\overline{Y}_e^d(\mu, z) = \overline{G}_e^d(\mu, z) \overline{F}_e^d(\mu, z), \quad (17)$$

aus der Übertragungsfunktion $\overline{G}_e^d(\mu, z)$ mit $\overline{F}_e^d(\mu, z)$, der Z- und Sturm-Liouville-Transformierten von $f_e(x, kT)$, bildet und anschließend die inverse STL und Z-Transformation bildet:

$$y[k, x_p] = \mathcal{Z}^{-1} \left\{ \sum_{\mu=1}^{M_s} \frac{1}{N_\mu} K(\mu, x_p) \overline{Y}_e^d(\mu, z) \right\}, \quad K(\mu, x) = \sin\left(\mu\pi \frac{x}{l}\right). \quad (18)$$

x_p stellt dabei den Ort der Tonabnahme an der Saite und M_s die Anzahl der zu synthetisierenden Partialschwingungen dar. Bei Annahme einer impulsartigen Anregung der Saite an genau der Stelle x_e mit der Auslenkung f ergibt sich

$$y[k, x_p] = \mathcal{Z}^{-1} \left\{ \sum_{\mu=1}^{M_s} \frac{1}{N_\mu} K(\mu, x_p) \overline{G}_e^d(\mu, z) \cdot f \cdot K(\mu, x_e) \right\} \quad (19)$$

$$= \mathcal{Z}^{-1} \left\{ \sum_{\mu=1}^{M_s} a(\mu, x_p) \overline{G}_e^d(\mu, z) \right\} = \sum_{\mu=1}^{M_s} a(\mu, x_p) \overline{g}_e^d(\mu, k) \quad (20)$$

als Ausgangssignal. Die Saitenschwingung besteht also aus der gewichteten Summe von abklingenden Sinusschwingungen unterschiedlicher, von μ abhängender, Frequenzen und Abklingraten.

3 Theoretische Vorüberlegungen

Wie im letzten Abschnitt dargestellt, besteht die Saitenschwingung aus einer gewichteten Summe von abklingenden Sinusschwingungen, den Partialschwingungen, unterschiedlicher Frequenzen ω_μ und Abklingraten σ_μ . ω_μ und σ_μ sind durch die Gleichungen (9) und (10) mit den Koeffizienten a_{10} , \tilde{a}_{12} , \tilde{a}_{02} und \tilde{a}_{04} vollständig bestimmt. Somit ist ein Ziel der Parameterextraktion diese Koeffizienten zu schätzen. Dazu bietet es sich an die Frequenzen $\hat{\omega}_\mu$ und Abklingraten $\hat{\sigma}_\mu$ der ersten M_e Partialschwingungen des Eingangssignal zu messen und schließlich aus diesen Werten durch Polynomfitting die geschätzten Koeffizienten a_{10}^{fit} , \tilde{a}_{12}^{fit} , \tilde{a}_{02}^{fit} und \tilde{a}_{04}^{fit} zu erhalten. In der darauf folgenden FTM Synthese können dann $b_1^d(\mu)$, $c_1^d(\mu)$ und $c_0^d(\mu)$ aus ω_μ^{fit} , σ_μ^{fit} und T auch für $\mu > M_e$ bestimmt werden. b_0 ist zwar nicht bestimmbar, aber für alle Partialschwingungen eine Konstante und somit bis auf eine lineare Verstärkung irrelevant für das synthetisierte Ausgangssignal.

Die Gewichtungen $a(\mu, x_p)$ der Partialschwingungen sind im Modell nicht von den physikalischen Eigenschaften der Saite sondern nur von Art und Position der Anregung sowie von der Position der Tonabnahme abhängig. Diese Parameter müssen somit nicht notwendigerweise geschätzt werden, da sie beliebig berechnet und dann sogar während der Synthese variiert werden können. Trotzdem kann die Schätzung von $a(\mu, x_p)$ sinnvoll sein, da dadurch Effekte, die nicht im Modell erfasst sind (z.B. Resonanzen eines Instrumentes), implizit mit erfasst werden können. Da die Abnahmeposition x_p in diesem Fall durch die Position des Mikrofons oder Tonabnehmers fest ist gilt $a(\mu, x_p) = a_\mu$. Möchte man also alle relevanten FTM Parameter einer Aufnahme extrahieren, sind die zu messende Werte:

- Frequenzen ω_μ
- Abklingraten σ_μ
- Gewichtungen der Partialschwingungen a_μ

4 Aufbau des Systems

Abbildung 2 zeigt das Blockdiagramm des aus den vorherigen Überlegungen hervorgegangenen Systems zur Parameterextraktion. Gemessene Parameter sind jeweils mit $\hat{\cdot}$ und durch Funktionsfitting erhaltene Werte jeweils durch ein hochgestelltes *fit* gekennzeichnet. Alle Frequenzen und Abklingraten werden in ihrer kontinuierlichen Form angegeben. Die Umrechnungen zu diskreten Größen kann einfach nach den Formeln (16) erfolgen.

Zu Beginn wird eine Schätzung des Leistungsdichtespektrums des gesamten Signals vorgenommen. Dieses wird benötigt um mit der MAT (Median-Adjustive Trajectories) Methode

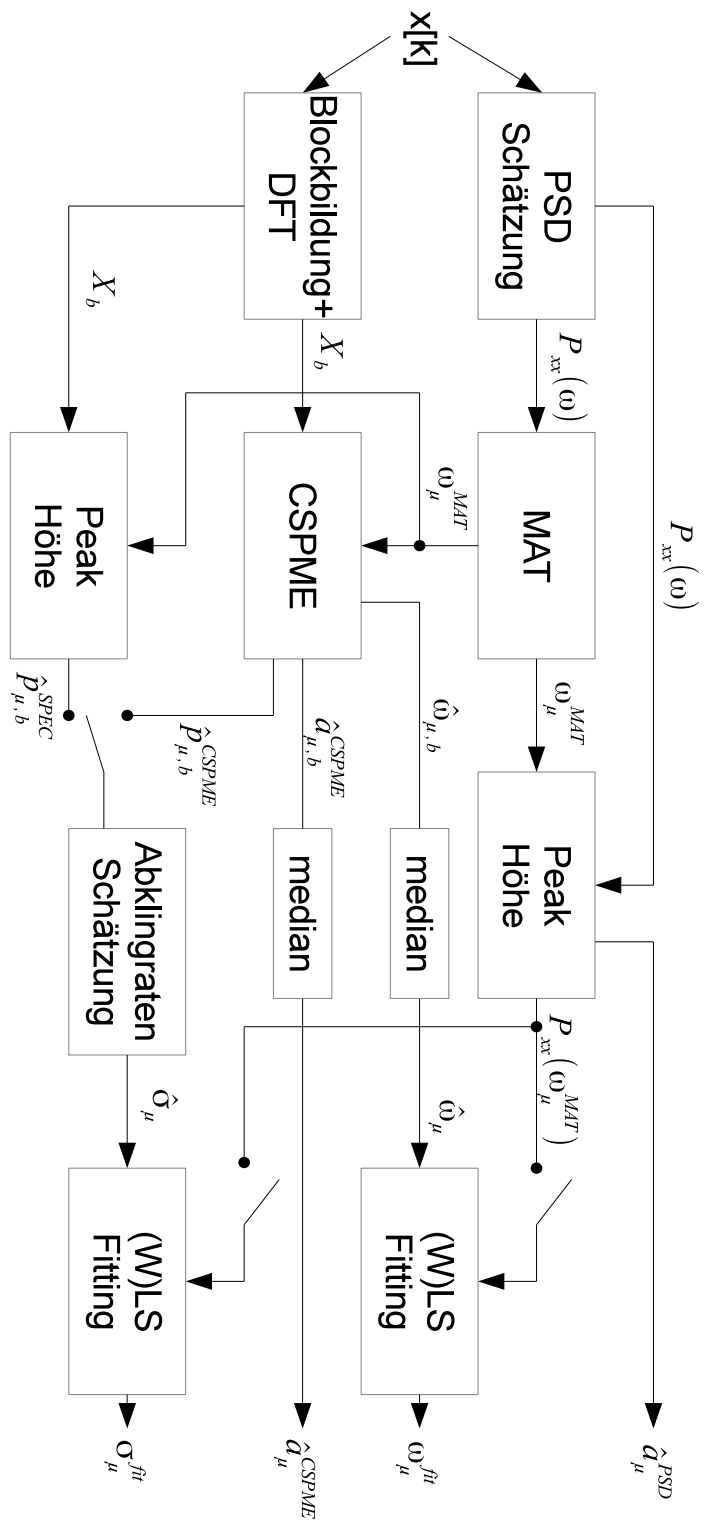


Abbildung 2: Aufbau des Systems zur Parameterextraktion

die ungefähre Frequenz der Partialschwingungen zu ermitteln. MAT wird in [1] (Kapitel 3.3.1) ausführlich behandelt, weshalb hier nur sehr grob auf das Prinzip eingegangen wird. Durch geschicktes Peak Picking wird eine Schätzung der Grundfrequenz ω_0 sowie der Inharmonizität β ermittelt. Mit Hilfe dieser beiden Werte kann dann durch die Formel

$$\omega_\mu^{MAT} = \mu\omega_0^{MAT} \sqrt{1 + \beta^{MAT} \mu^2} \quad (21)$$

die ungefähre Frequenz und damit Position der Peaks der Partialschwingungen im Spektrum ermittelt werden. Für eine tatsächliche Messung der genauen Frequenzen ist diese Methode allerdings zu ungenau. Mit Hilfe von ω_μ^{MAT} (oder alternativ dem später berechneten ω_μ^{fit}) kann anschließend an MAT eine Messung der Gewichtung \hat{a}_μ erfolgen:

$$\hat{a}_\mu^{PSD} = \frac{T\omega_\mu^{MAT}}{b_0T} \sqrt{P_{xx}(\omega_\mu^{MAT})} \sim \omega_\mu^{MAT} \sqrt{P_{xx}(\omega_\mu^{MAT})} \quad (22)$$

Die Skalierung (aus Gründen der Übersichtlichkeit nicht im Blockdiagramm eingezeichnet) ist notwendig, da die Partialschwingungen, wie in Gleichung ersichtlich (15), schon von sich aus eine Gewichtung besitzen, die in die Messung einfließt. b_0 kann vernachlässigt werden, da es unabhängig von μ und somit nur eine lineare Verstärkung aller Partialschwingungen darstellt.

Die eigentliche Frequenzmessung findet nach einer Blockbildung, Fensterung und DFT des Eingangssignals statt. Dabei wird ein Cosinusfenster zweiter Ordnung gewählt wobei Fensterlänge und Überlappung der Blöcke in Anlehnung an [1] (Kapitel 3.1.3) in Abhängigkeit von der Grundfrequenz des Eingangssignals gewählt wird. Die Blöcke X_b mit $1 \leq b \leq N_{blocks}$ werden dann verwendet, um jeweils für jeden Block mittels CSPME (Complex Spectral Phase Magnitude Evolution) [1] eine Messung der Frequenzen der Partialschwingungen $\hat{\omega}_{\mu,b}$, und eine zur obigen Methode alternative Messung der Gewichtungen $\hat{a}_\mu, \hat{a}_{\mu,b}^{CSPME}$ zu erhalten. CSPME, das eine Weiterentwicklung des ursprünglich nur zur Frequenzbestimmung geeigneten CSPE (Complex Spectral Phase Evolution) [3] darstellt, ist ein Verfahren um Frequenz, Abklingrate, Phase und (Start-)Amplitude einer komplexen Exponentialschwingung durch Auswertung des Phasenunterschieds zwischen einem Signal und einer verzögerten Version dieses Signals exakt zu bestimmen. Die Genauigkeit des Verfahrens nimmt in Anwesenheit von Störungen oder zusätzlichen Schwingungen ab, jedoch gibt es zwei Möglichkeiten diesem Effekt entgegen zu wirken. Im Falle mehrerer Schwingungen wird dazu die Phase an dem DFT Wert ausgewertet, der der ungefähren Frequenz der genau zu messenden Schwingung am nächsten liegt, da dort die Einflüsse der restlichen Schwingungen oder Störungen am geringsten ist. Diese Eigenschaft kann außerdem dazu genutzt werden bereits gemessene Schwingungen zu synthetisieren und vom Signal abzuziehen, was die Störungen für die folgenden Messungen reduziert. Die ungefähren Frequenzen der Schwingungen sind in diesem System durch MAT bekannt, was sich in Kombination mit CSPME in [1] bewährt hat. Somit erhält man abgesehen von den Frequenzen $\hat{\omega}_{\mu,b}$ in jedem Block für jede Partialschwingung eine Messung der Abklingrate $\hat{\sigma}_{\mu,b}$ und die Startamplitude (in Block b) $\hat{p}_{\mu,b}^{CSPME}$. Die Messungen der Abklingrate $\hat{\sigma}_{\mu,b}$ stellte sich als zu unzuverlässig für die Berechnung von σ_μ^{fit} heraus. Deshalb wird auf diese Möglichkeit im Weiteren nicht mehr eingegangen. Trotzdem kann $\hat{\sigma}_{\mu,b}$ benutzt werden, um mittels CSPME die Gewichtung der Partialschwingungen $\hat{a}_{\mu,b}^{CSPME}$ zu messen. Dafür wird für jede Partialschwingung $\hat{p}_{\mu,b}^{CSPME}$ mit $\hat{\sigma}_{\mu,b}$ und der vergangenen Zeit zurückgerechnet:

$$\hat{a}_{\mu,b}^{CSPME} = \frac{T\omega_\mu}{b_0T} \hat{p}_{\mu,b}^{CSPME} \cdot e^{-(b-1) \cdot (hopLength)\hat{\sigma}_{\mu,b}} \sim \omega_\mu \hat{p}_{\mu,b}^{CSPME} \cdot e^{-(b-1) \cdot (hopLength)\hat{\sigma}_{\mu,b}}. \quad (23)$$

Die zusätzliche Gewichtung mit $\omega_\mu e^{T\sigma_\mu}$ erfolgt analog zu Gleichung (22).

Sowohl von $\hat{a}_{\mu,b}^{CSPME}$ als auch von $\hat{\omega}_{\mu,b}$ wird anschließend der Median über alle Blöcke gebildet um fehlerhafte Messungen auszuschließen.

$$\hat{\omega}_{\mu} = \text{median}_b \{ \hat{\omega}_{\mu,b} \} \quad (24)$$

$$\hat{a}_{\mu}^{CSPME} = \text{median}_b \{ \hat{a}_{\mu,b}^{CSPME} \} \quad (25)$$

Für die Schätzung der Abklingraten wird versucht die Änderung der Amplituden der Partialschwingungen über die Zeit zu verfolgen und daraus die Abklingrate zu berechnen. Für die Amplitudenschätzungen gibt es zwei Möglichkeiten. Entweder man bedient sich der ohnehin aus der CSPME vorhandenen $\hat{p}_{\mu,b}^{CSPME}$ oder man berechnet in jedem Block für jedes ω_{μ}^{MAT} die Höhe der Peaks $\hat{p}_{\mu,b}^{SPEC}$ im Spektrum. Abgesehen von der Herkunft der Amplitudenschätzungen verläuft alles Weitere identisch. Daher wird der folgende Ablauf nur noch für $\hat{p}_{\mu,b}$ beschrieben. Für die μ -te Partialschwingung berechnet man dann $\hat{\sigma}_{\mu}$ indem man die Steigung einer Gerade ermittelt, die im Least Squares Sinne durch $\ln(\hat{p}_{\mu,b})$ verläuft. Da die Qualität der Messungen von $\hat{p}_{\mu,b}$ mit der Zeit abnehmen und schließlich im Rauschen verschwinden, ist es sinnvoll nicht alle N_{blocks} Werte für jede Partialschwingung zu verwenden. Um die Anzahl der zu verwendenden Werte zu bestimmen sind verschiedene Methoden möglich. Um das Konzept einfach zu halten, wird ein Kriterium verwendet, dass für alle $\hat{p}_{\mu,b}$ das Maß

$$\alpha_{\mu,b} = 20 \lg(\hat{p}_{\mu,b}/\hat{p}_{\mu,1}) \text{ dB} \quad (26)$$

bildet und dann jeweils für jede Partialschwingung die Amplitudenschätzungen aus den ersten $b_{max,\mu}$ Blöcken für die die Bestimmung von σ_{μ} benutzt, so dass

$$b_{max,\mu} = \max[2, \max(\{b : \alpha_{\mu,i} > -60\text{dB} \quad \forall i < b\})] \quad (27)$$

gilt.

Schließlich kann mit $\hat{\omega}_{\mu}$ und $\hat{\sigma}_{\mu}$ ein Least Squares (LS) oder wahlweise Weighted Least Squares (WLS) Polynomfitting ω_{μ}^{fit} und σ_{μ}^{fit} berechnet werden, woraus sich dann die gesuchten Koeffizienten a_{10}^{fit} , \tilde{a}_{12}^{fit} , \tilde{a}_{02}^{fit} und \tilde{a}_{04}^{fit} ergeben. Als Gewichte für das WLS Verfahren wird dabei $P_{xx}(\omega_{\mu})^w$ verwendet, da davon auszugehen ist, dass die Messungen für Partialschwingungen mit hoher Amplitude genauer sind. Der Exponent w ist weiterhin eine Möglichkeit um hohe Werte noch stärker und kleine Werte noch kleiner zu gewichten. Für das WLS Fitting von $\hat{\omega}_{\mu}$ stellte sich in ersten Experimenten $w = 5$ als eine gute Wahl heraus, während für $\hat{\sigma}_{\mu}$ $w = 2$ gute Ergebnisse liefert.

5 Evaluation

Um die Qualität der extrahierten Parameter zu beurteilen, wurden zwei Methoden verwendet, die in Abbildung 3 und 8 veranschaulicht werden. Bei der ersten Methode wird zuerst ein Signal $x[k]$ mit bekannten physikalischen Parametern einer Saitenschwingung erzeugt. Diesem Signal werden dann die Parameter extrahiert womit das Signal $y[k]$ resynthetisiert werden kann. Durch Vergleich der Signale und der bekannten sowie extrahierten Parameter kann anschließend eine Aussage über die Qualität der Parameterextraktion gemacht werden. Bei der zweiten Methode wird die Aufnahme einer natürlichen Saitenschwingung analysiert, synthetisiert und anschließend verglichen.

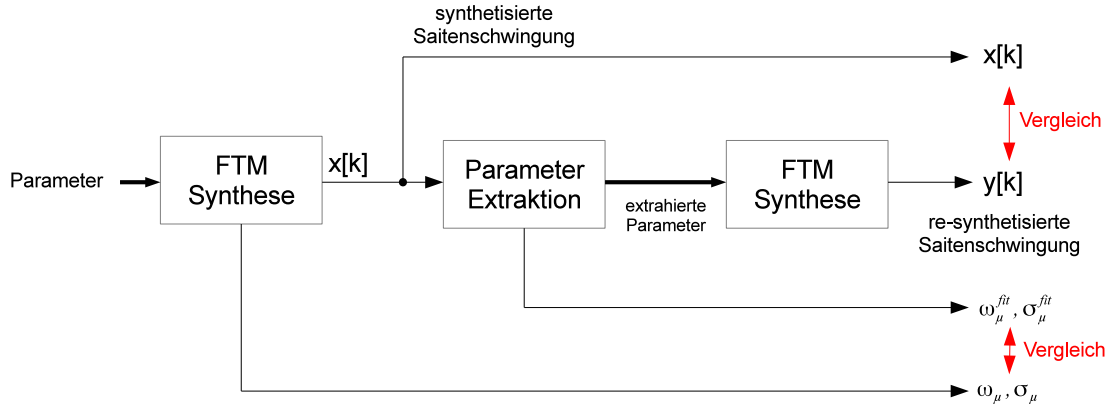


Abbildung 3: Parameterextraktion für eine synthetische Saitenschwingung

5.1 Evaluation der Parameterextraktion für eine synthetische Saitenschwingung

Die physikalischen Parameter einer Nylonseite für die Synthese wurden wie in [2] gewählt, wobei der Ort der Anregung zu $\frac{x_e}{l} = \frac{1}{13}$ und der Ort der Tonabnahme zu $\frac{x_p}{l} = \frac{1}{17}$ gewählt wurde. Lediglich diese beiden Werte wurden auch bei der Resynthese als bekannt voraus gesetzt. Damit war die Extraktion von a_μ nicht notwendig. Sowohl für die Synthese als auch für die Resynthese wurde die Anzahl der zu synthetisierenden Partialschwingungen $M_s = 30$ verwendet. Für höhere oder niedrigere M_s änderten sich die Ergebnisse nicht wesentlich.

5.1.1 Qualität der Schätzung von ω_μ

Abbildung 4a) zeigt die exakten Werte für ω_μ (in der Abbildung für bessere Sichtbarkeit quadriert), die Messungen $\hat{\omega}_\mu$, sowie ω_μ^{fit} für LS und WLS Fitting. Abbildung 4b) zeigt die relativen Fehler von $\hat{\omega}_\mu$, sowie ω_μ^{fit} für LS und WLS Fitting zu den exakten Werten auf einer logarithmischen Skala. Es ist deutlich zu erkennen, dass die Messungen für kleine μ eine sehr hohe Genauigkeit besitzen, die ab ungefähr $\mu = 7$ deutlich abnimmt. Für das LS Fitting werden alle Messungen gleich gewichtet, was dazu führt, dass die ungenauen Messungen für große μ die Genauigkeit der gefitteten Funktion deutlich senken. Besonders problematisch ist, dass bereits für die ersten Partialschwingungen und somit auch für die Grundfrequenz der Fehler in der Größenordnung von 100% liegt. Beim WLS Fitting werden die gemessenen Werte abhängig von ihrer Peakhöhe im Leistungsdichtespektrum gewichtet. Da die Peakhöhe der Partialschwingungen mit steigendem μ im Allgemeinen abnehmen, werden die Messungen für niedrige μ stärker gewichtet als für hohe μ . Dies führt zu einer sehr genauen Schätzung von ω_μ .

5.1.2 Qualität der Schätzung von σ_μ

Die Schätzung von σ_μ basiert auf der Messung von Amplituden für jede Partialschwingung in jedem Block. Wie in 2 dargestellt, gibt es zu diesem Zweck die Möglichkeit die Amplitudenwerte $\hat{p}_{\mu,b}^{CSPME}$ über CSPME oder die Peakhöhe $\hat{p}_{\mu,b}^{SPEC}$ des Spektrum zu erhalten. Abbildung 5 zeigt, dass beide Varianten eine gute Genauigkeit bieten. Jedoch gibt es bei Verwendung

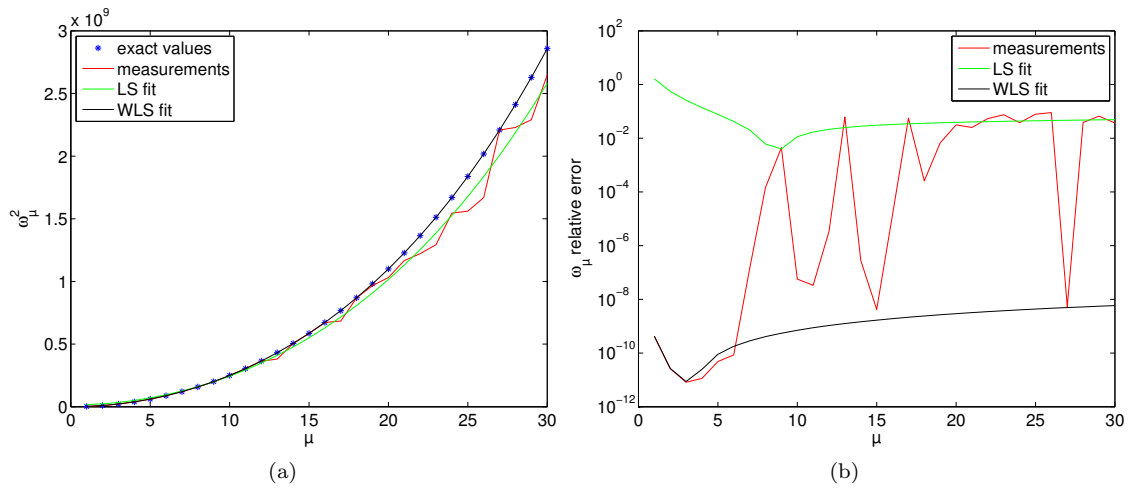


Abbildung 4: Parameterextraktion für eine synthetische Saitenschwingung: Schätzungen für ω_μ (a) und relative Fehler der Schätzungen (b)

der CSPME Amplituden für höhere μ starke Ausreißer. Im Allgemeinen scheint die Methode der Peakhöhe das robustere Verfahren zu sein. Daher werden sämtliche Schätzungen für σ_μ im Folgenden auf diese Weise erstellt.

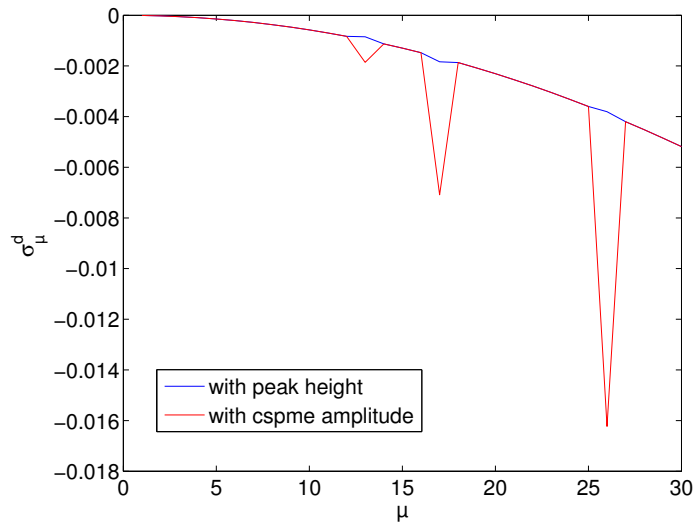


Abbildung 5: Parameterextraktion für eine synthetische Saitenschwingung: Vergleich der Genauigkeit für $\sigma_\mu^d = \sigma_\mu T$ bei Berechnung mittels CSPME Amplituden oder Peakhöhe des Spektrums

Abbildung 6a) zeigt die exakten Werte für σ_μ , die Messungen $\hat{\sigma}_\mu$, sowie σ_μ^{fit} für LS und WLS Fitting. Abbildung 6b) zeigt die relativen Fehler von $\hat{\sigma}_\mu$, sowie σ_μ^{fit} für LS und WLS Fitting zu den exakten Werten auf einer logarithmischen Skala. Da bereits die Messungen bis

auf wenige Ausreißer sehr genau sind liefert sowohl LS als auch WLS gute Ergebnisse, wobei WLS im direkten Vergleich noch ein mal deutlich genauer ist.

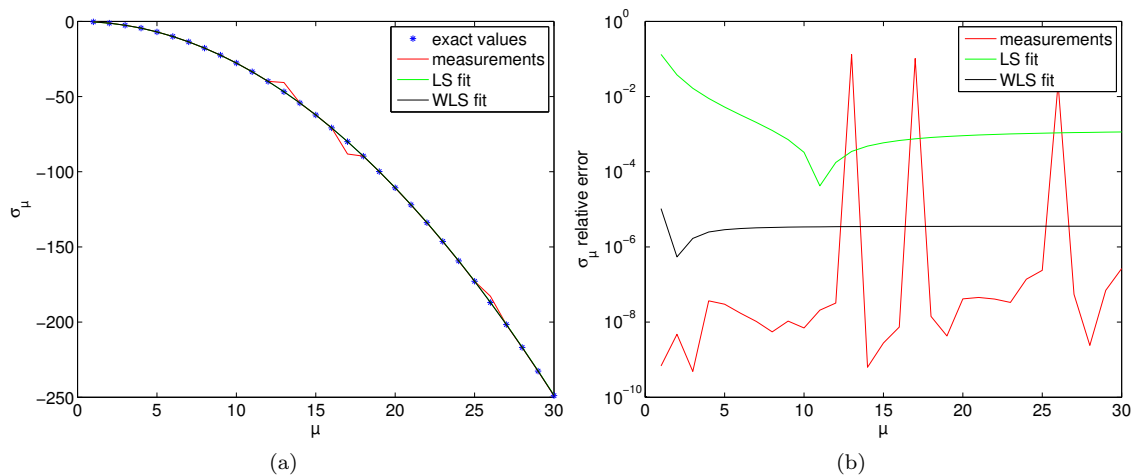


Abbildung 6: Parameterextraktion für eine synthetische Saitenschwingung: Schätzungen für σ_μ (a) und relative Fehler der Schätzungen (b)

5.1.3 Vergleich der Leistungsdichtespektren

Abbildung 7 zeigt die Leistungsdichtespektren der Signale $x[k]$ und $y[k]$ (bei Verwendung von ω_μ^{fit} und σ_μ^{fit} mit WLS) und deren Unterschied. Die Leistungsdichtespektren selbst sind für das Auge quasi identisch und auch der Unterschied zwischen den beiden Leistungsdichtespektren ist gering. Tatsächlich beträgt die Varianz des Fehlersignals $e[k] = x[k] - y[k]$ nur $1.042 \cdot 10^{-7}$. Damit wurde gezeigt, dass das System für synthetische Eingangssignale eine gute Qualität der Parameterextraktion erreichen kann.

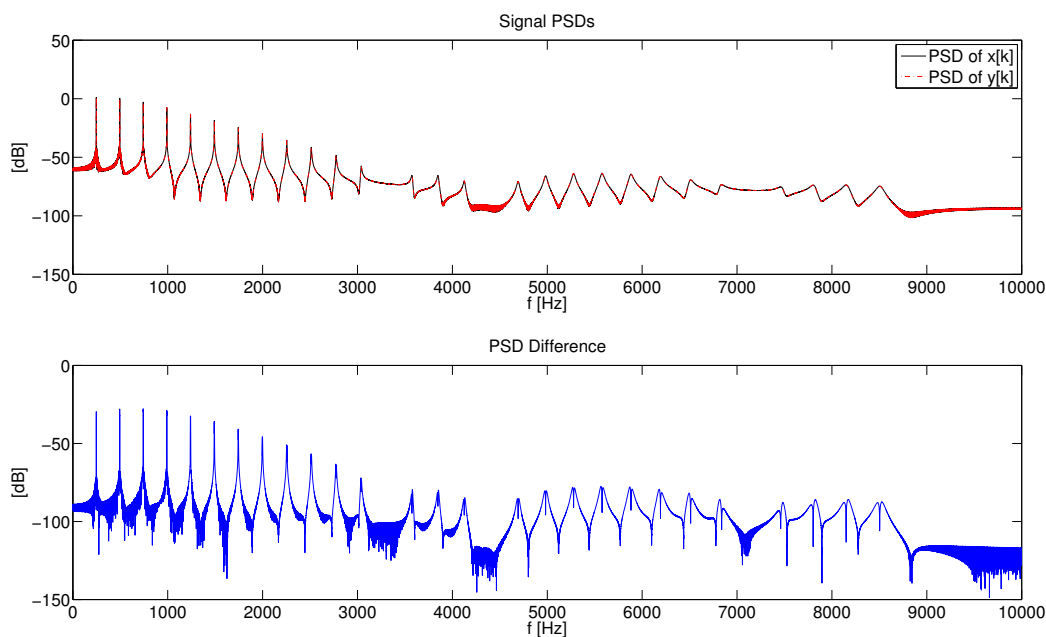


Abbildung 7: Parameterextraktion für eine synthetische Saitenschwingung: Vergleich der Leistungsdichtespektren der Signale $x[k]$ und $y[k]$ (oben) und Unterschied der Leistungsdichtespektren (unten)

5.2 Evaluation der Parameterextraktion für eine aufgenommene Saitenschwingung

Um die Qualität der Parameterextraktion für eine aufgenommene Saitenschwingung zu testen, wurde eine Aufnahme der H-Saite einer akustischen Gitarre mit Nylonsaiten verwendet. Zur Analyse wurde nur die erste Sekunde der Aufnahme genutzt. Es wurde $M_e = M_s = 50$ gewählt. Abgesehen davon war die Vorgehensweise wenn nicht anders angegeben nicht anders als für die Analyse des synthetischen Signals im vorherigen Abschnitt.

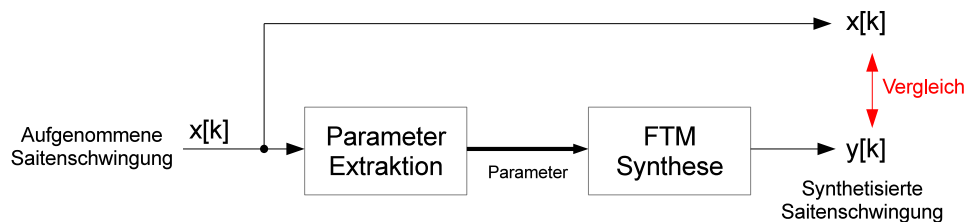


Abbildung 8: Parameterextraktion für eine aufgenommene Saitenschwingung

5.2.1 Qualität der Schätzung von ω_μ

Abbildung 9a) zeigt die Messungen für $\hat{\omega}_\mu$ sowie ω_μ^{fit} für LS und WLS Fitting. In Abbildung 9b) sind die relativen Abweichungen der beiden gefitteten Funktionen zum gemessenen $\hat{\omega}_\mu$ dargestellt. Zwar verläuft die LS Kurve allgemein näher an den Messungen, dafür zeigen sich bei den wichtigen ersten Partialschwingungen große Abweichungen. Die WLS Kurve dagegen zeigt für niedrige $\mu < 8$ bessere Genauigkeit, die für höhere μ allerdings eine schlechtere Genauigkeit als LS bietet. Ein guter Kompromiss mit WLS könnte durch weiterer Experimente mit dem Gewichtungs-Exponenten w erreicht werden.

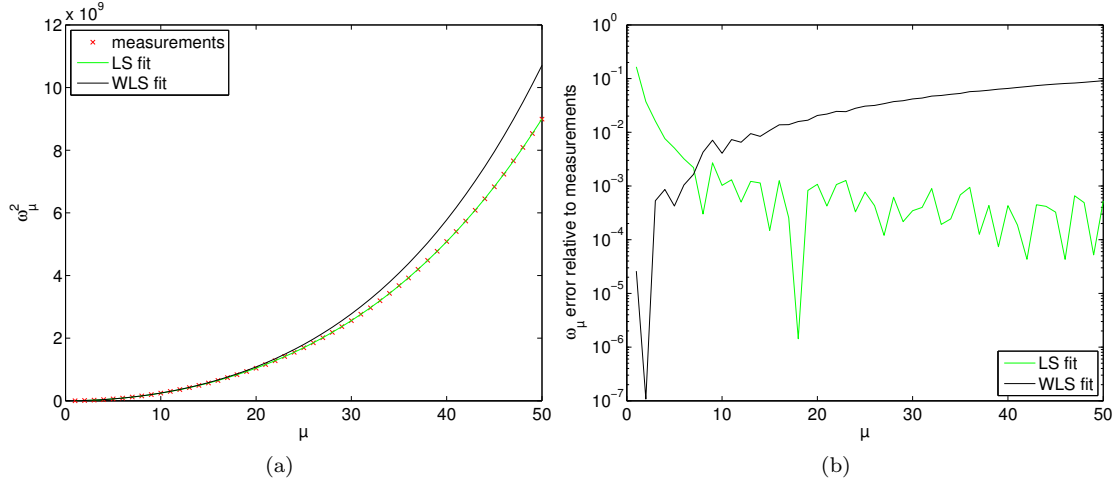


Abbildung 9: Parameterextraktion für eine aufgenommene Saitenschwingung: Schätzungen für ω_μ (a) und relative Fehler der Schätzungen in Bezug auf die Messungen (b)

5.2.2 Qualität der Schätzung von σ_μ

Abbildung 9 zeigt die Messungen $\hat{\sigma}_\mu$ sowie σ_μ^{fit} für LS und WLS Fitting. Deutlich erkennbar ist, dass schon die Qualität der Messungen mangelhaft ist. Für $\mu < 10$ ergibt sich noch der erwartete, abfallende Verlauf. Für größere μ steigt $\hat{\sigma}_\mu$ sogar über 0, was einer aufklingenden Partialschwingung entsprechen würde. Durch die schlechte Qualität der Messungen zeigt auch σ_μ^{fit} für LS Fitting kein realistisches Ergebnis. Für WLS zeigt sich ein realistisches, jedoch aufgrund der Qualität der Messungen wahrscheinlich ebenfalls ungenaues Resultat.

5.2.3 Qualität der Schätzung von a_μ

Abbildung 11 zeigt einen Vergleich der Messungen \hat{a}_μ^{CSPME} und \hat{a}_μ^{PSD} für a_μ . Die Messungen wurden jeweils auf $\hat{a}_\mu = 1$ normiert. \hat{a}_μ^{CSPME} scheint im Gegensatz zu \hat{a}_μ^{PSD} größere Störungen aufzuweisen. Da die \hat{a}_μ^{PSD} Messungen bei den von MAT berechneten Frequenzen im Leistungsdichtespektrum getätigt werden, kann es jedoch gerade bei hohen Frequenzen zu großen Ungenauigkeiten kommen, die sich z.B. im Fehlen von Partialschwingungen äußern.

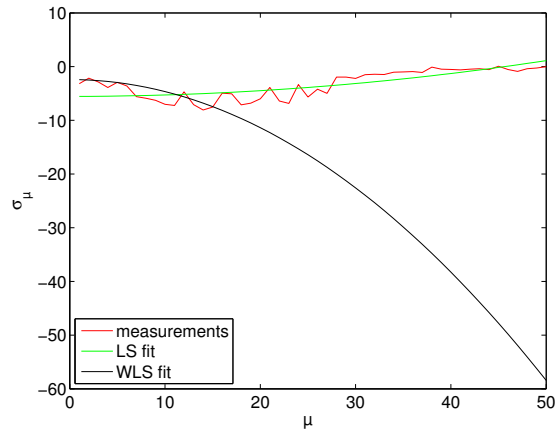


Abbildung 10: Parameterextraktion für eine aufgenommene Saitenschwingung: Messungen und Schätzungen für σ_μ

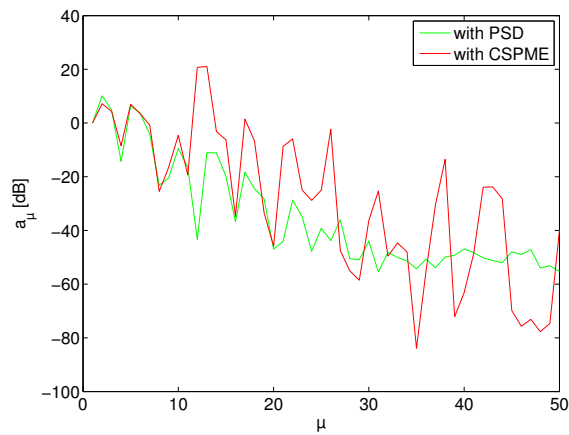


Abbildung 11: Parameterextraktion für eine aufgenommene Saitenschwingung: Messungen für a_μ

5.2.4 Vergleich der Leistungsdichtespektren

Abbildung 12 zeigt das Leistungsdichtespektrum des Eingangssignals und des synthetisierten Signals im Vergleich. Dabei wurde \hat{a}_μ^{PSD} und für ω_μ^{fit} bzw. σ_μ^{fit} jeweils die WLS Schätzungen für die Partialschwingungen genutzt. Abbildung 13 a) und b) zeigen vergrößerte Ausschnitte verschiedener Frequenzbereiche. Es zeigt sich, dass die Leistungsdichtespektren für tiefe Frequenzen große Ähnlichkeiten aufweisen. Lediglich die Sub-Partialschwingung bei $\approx 123\text{Hz}$ unter der eigentlichen Grundschwingung fehlt. Dies ist darauf zurückzuführen, dass sie von von MAT nicht erkannt und somit gänzlich ignoriert wurde. Für die höheren Frequenzen in 13b) zeigen sich die Abweichungen der WLS Schätzung in ω_μ^{fit} von den exakten ω_μ durch eine Verschiebung der Peaks. Auch die Peakgewichtungen sind in diesem Frequenzbereich schon sehr ungenau.

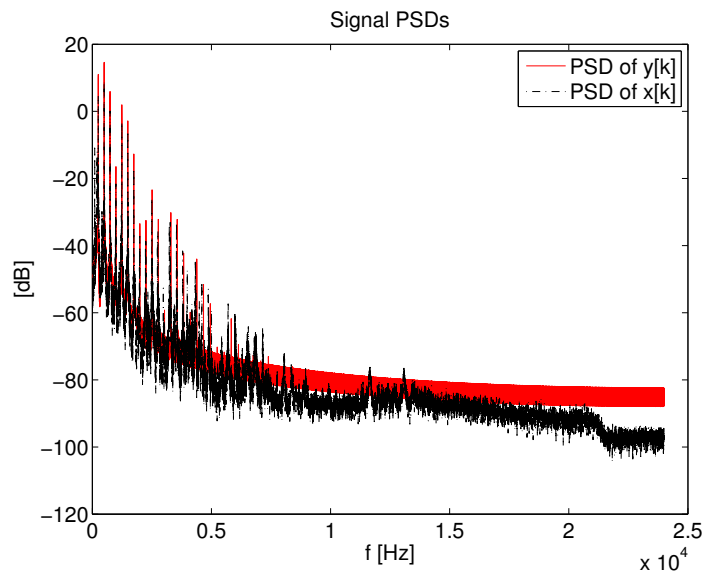


Abbildung 12: Parameterextraktion für eine aufgenommene Saitenschwingung: Vergleich der Leistungsdichtespektren der Signale $x[k]$ und $y[k]$ über das gesamte Spektrum

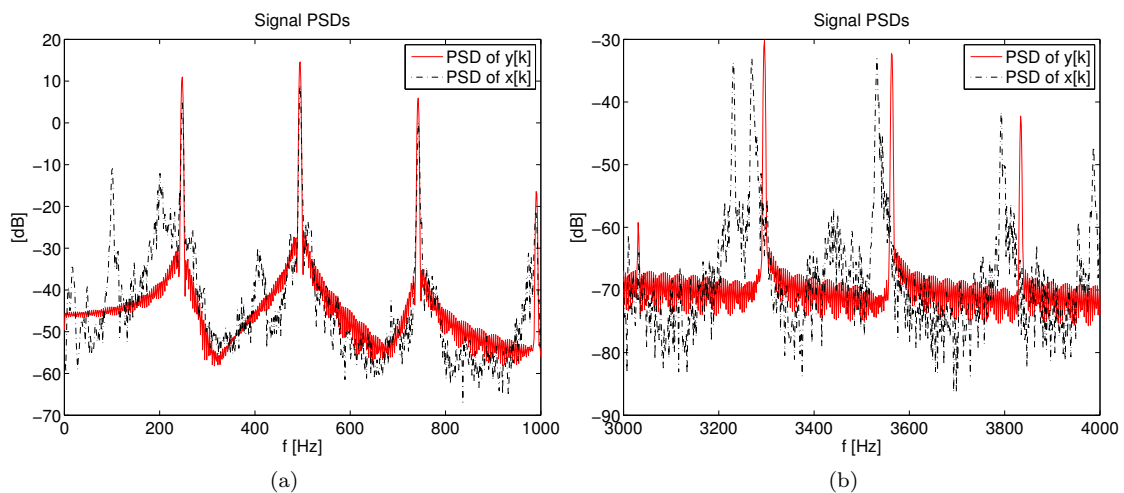


Abbildung 13: Parameterextraktion für eine aufgenommene Saitenschwingung: Vergleich der Leistungsdichtespektren der Signale $x[k]$ und $y[k]$ in unterschiedlichen Frequenzbereichen

5.2.5 Hörvergleich

In informellen Hörvergleichen zeigt sich, dass die Qualität der Parameterextraktion nicht ausreicht, um eine zufriedenstellende Ähnlichkeit des Klangs zu erreichen. Zwar sind die Gemeinsamkeiten zu erkennen, jedoch machen sich die Fehler in der Schätzung von σ_μ und a_μ deutlich in dumpferem Klang oder unterschiedlicher Ausklangzeit der Schwingung bemerkbar. Tests mit anderen Eingangssignalen zeigten außerdem, dass die Qualität der extrahierten

Parameter und somit des synthetisierten Signals stark von der Qualität (Hall, Rauschlevel, Anregungsart) des Eingangssignals abhängt.

6 Zusammenfassung

Es wurde ein System erarbeitet, um die Parameter aus einer aufgenommenen Saitenschwingung zu extrahieren, die für eine spätere Synthetisierung notwendig sind. Die prinzipielle Funktionalität wurde dabei mit synthetischen Eingangssignalen verifiziert und verschiedene Methoden abgewägt. Für Aufnahmen von echten Saitenschwingungen konnte allerdings keine vollständig überzeugende Qualität der Parameterextraktion erreicht werden. Die Gründe dafür können in der Qualität der verwendeten Aufnahmen, nicht dem verwendeten Modell entsprechenden Saitenschwingungen und vor allem auch fehlender Robustheit der verwendeten Extraktionsmethode vermutet werden. Um die Ergebnisse zu verbessern, bieten sich deshalb folgende weitere Arbeiten an:

- Experimente zur Optimierung des Exponenten w für die Gewichte beim WLS Fitting
- Bestimmung eines guten Wertes für die Anzahl der zu analysierenden Partialschwingungen M_e
- Verbesserung der Robustheit der Schätzung für σ_μ
- Verbesserung der Parameterextraktion durch Mittelung über mehrere Aufnahmen des gleichen Instruments
- Bestimmung von Anschlags- und Tonabnahmeposition aus a_μ
- Erweiterung des Modells (z.B. auslenkungsabhängige Saitenspannung, Korpusmodell, ...)

Literatur

- [1] Matthieu Hodgkinson. *Physically Informed Subtraction of a String's Resonances from Monophonic, Physically Informed Subtraction of a String's Resonances from Monophonic, Discretely Attacked Tones : Discretely Attacked Tones : a Phase Vocoder Approach*. PhD thesis, National University of Ireland, Maynooth, 2012. [4](#), [4](#)
- [2] Rudolf Rabenstein and Lutz Trautmann. Digital sound synthesis of string instruments with the functional transformation method. *Elsevier*, 2002. [2](#), [5.1](#)
- [3] Kevin M. Short and Ricardo A. Garcia. Signal analysis using the complex spectral phase evolution (cspe) method. In *AES 120th Convention*, 2006. [4](#)