



Analytische Wavelet-Packets zur Separation unbekannter Sprachsignale

Analytic Wavelet-Packets for Separation of Unknown Speech Signals

Andreas Sandmair, Mario Lietz, Fernando Puente León, Karlsruher Institut für Technologie

Zusammenfassung Die Separation unbekannter Signale ist im Bereich der Sprachsignalverarbeitung von besonderer Bedeutung. Die Trennung erfolgt bei Verwendung mehrerer Mikrofone durch Auswertung der resultierenden Laufzeiten zwischen den einzelnen Quellen und Sensoren. Diese Laufzeitdifferenzen führen im Frequenzbereich zu definierten Phasendifferenzen zwischen den Sensorsignalen. Durch die statistische Analyse der Phasenwerte ist eine Rekonstruktion der Signale unter bestimmten Rahmenbedingungen möglich. Zur Berechnung der spektralen Koeffizienten wird normalerweise die Kurzzeit-Fourier-Transformation (STFT) verwendet. Durch den Einsatz analytischer Wavelet-Packets (AWP) kann die Leistungsfähigkeit derartiger Separationsalgorithmen hinsichtlich Zeitdauer und Genauigkeit verbessert werden. Nach einer Beschreibung des Separationsproblems im Allgemeinen und der Unterschiede der beiden Zeit-Frequenz-Darstellungen wer-

den die Vorteile der Wavelet-Packets gegenüber der Kurzzeit-Fourier-Transformation erläutert. ▶▶▶ **Summary** The separation of a mixture of unknown speech signals is a demanding research topic in the area of signal processing. Recording the signals with two sensors at different positions, the resulting time delay of arrival can be used to reconstruct the original signals. In the frequency domain the delay is related to characteristic phase differences between the sensor signals. Under certain assumptions the original signals can be reconstructed based on a statistical analysis of the phase values. To this end, the signals have to be transformed to the time-frequency domain due to the non-stationary character of speech signals. Usually, the short-time Fourier transform (STFT) is used to calculate the spectral coefficients. Using analytic wavelet packets (AWP) the performance of the separation algorithm with regard to both computation time and accuracy can be improved.

Schlagwörter Analytische Wavelet-Packets, Blind-Source-Separation, Kurzzeit-Fourier-Transformation, Zeit-Frequenz-Darstellungen ▶▶▶ **Keywords** Analytic wavelet packets, blind source separation, short-time Fourier transform, time-frequency representations

1 Einleitung

Das Hören ist, neben dem Sehsinn, einer der wichtigsten Sinne der Menschen zur Wahrnehmung ihrer Umgebung. Im Gegensatz zur visuellen Wahrnehmung hat der Hörsinn jedoch einen klaren Vorteil: er ist nicht richtungsselektiv. Durch die Anatomie des Kopfes und die neuronale Verarbeitung im Gehirn kann der Mensch Schallquellen lokalisieren und einzelne Quellen adaptiv verstärken oder dämpfen [1]. Im Zusammenspiel

mit dem Sehsinn bildet er somit die Grundlage für die sichere Bewegung des Menschen in seinem Umfeld. Durch Schädigungen des Gehörs wird der Mensch jedoch erheblich in seinen Möglichkeiten eingeschränkt. Die Verwendung von Hörgeräten kann diese Verluste in begrenztem Maße ausgleichen. Bei einer großen Anzahl an Quellen oder diffus einfallendem Störschall stoßen aber auch die modernsten Geräte an ihre Grenzen [2]. In derartigen Situationen, in der Literatur als

„Cocktail-Party-Problem“ bekannt [3], müssen für die technische Weiterverarbeitung der Signale die einzelnen Schallquellen getrennt werden. Durch die Verwendung mehrerer Mikrofone und die Auswertung der resultierenden Richtungsabhängigkeit ist prinzipiell eine Trennung der Signale möglich. Dabei ist eine statistische Betrachtung der Signale notwendig, da keine Informationen über die Sprecher vorhanden sind. Im Folgenden wird ein Ansatz zur Trennung unbekannter Signale vorgestellt. Um einen Überblick über die Problematik im Gesamten zu erhalten, sei auf [4] verwiesen. Der Fokus dieser Arbeit liegt jedoch auf der Transformation in den Zeit-Frequenz-Bereich. Alternative Verfahren werden in Abschnitt 3 vorgestellt. Anschließend werden notwendige Modifikationen erläutert und die unterschiedlichen Methoden hinsichtlich ihrer Leistungsfähigkeit zur Separation von Signalen bewertet.

2 Trennung unbekannter Signale

Der Mensch nutzt zur Separation von Signalen die Laufzeitdifferenz und die Amplitudendifferenz zwischen seinen beiden Ohren. Signifikante Amplitudenunterschiede treten – bedingt durch die Anatomie des menschlichen Kopfes und der Ohren – durch Abschattungseffekte auf. Für eine Nutzung der Amplitudendifferenz im technischen Bereich müssen somit Störkörper zwischen den Sensoren platziert werden. Im vorgestellten Ansatz werden zur Trennung der Signale nur die Laufzeitdifferenzen zwischen den Sensoren verwendet.

Sind mehrere Sprecher oder Störquellen simultan aktiv, kommt es zu einer Überlagerung der Signale. Mathematisch lässt sich ein derartiges Szenario entweder als gewichtete Summe der zeitverzögerten, gedämpften Quellsignale $s_i(t)$ oder als Faltung des Quellsignals mit einer spezifischen Raumimpulsantwort a_{ji} beschreiben [4]. Für den j -ten Sensor gilt bei M Quellen und L unterschiedlichen Ausbreitungswegen:

$$x_j(t) = \sum_{i=1}^M \sum_{k=1}^L a_{ji}^k s_i(t - \delta_{ji}^k) \quad (1)$$

$$= \sum_{i=1}^M a_{ji}(t) * s_i(t). \quad (2)$$

Diese Darstellung ist notwendig, um den Einfluss der endlichen Ausbreitungsgeschwindigkeit und der Reflexionen zu berücksichtigen. Für eine Rekonstruktion der ursprünglichen Signale sind mindestens zwei Sensoren notwendig. Durch die Auswertung der Laufzeitdifferenzen können unter bestimmten Rahmenbedingungen die Originalsignale rekonstruiert werden.

Die Vorgehensweise bei der Rekonstruktion wird in Bild 1 kompakt dargestellt. Nach der Aufnahme der Signale folgt die Transformation in den Zeit-Frequenz-Bereich (ZFB). Aus der Faltung im Zeitbereich wird eine Multiplikation im Frequenzbereich. Das Signal kann jedoch aufgrund der Instationarität des Zeitsignals nur

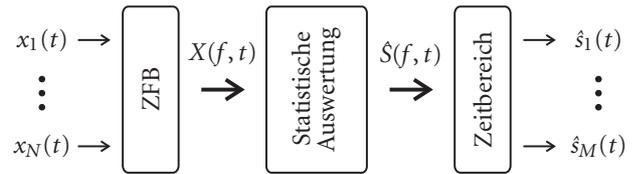


Bild 1 Verarbeitungsschritte bei der Trennung unbekannter Signale.

abschnittsweise in den Frequenzbereich transformiert werden. Das zeit- und frequenzabhängige Sensorsignal kann durch

$$X_j(f, t) = \sum_{i=1}^M A_{ji}(f) \cdot S_i(f, t) \quad (3)$$

beschrieben werden. Durch die Auswertung der Laufzeitdifferenzen zwischen den Sensoren lassen sich die Originalsignale rekonstruieren. In der Zeit-Frequenz-Ebene entspricht der Laufzeitunterschied einer spezifischen Phasendifferenz $\Delta\varphi(f)$ in jedem Frequenzband. Für die Annahme eines Szenarios ohne Reflexionen (siehe Bild 2) berechnet sich die Phasendifferenz $\Delta\varphi(f)$ aus dem Laufzeitunterschied

$$\Delta t = \frac{|d_{S1}^{X1} - d_{S1}^{X2}|}{c_{\text{Schall}}} = \frac{\Delta d}{c_{\text{Schall}}} \quad (4)$$

zu

$$\Delta\varphi(f) = 2\pi f \Delta t. \quad (5)$$

Die Phasendifferenz ist jedoch aufgrund der Periodizität der harmonischen Schwingungen nur im Bereich von $-\pi$ bis π eindeutig. Überschreiten die Werte von $\Delta\varphi(f)$ diesen Bereich, werden sie auf den entsprechenden Wert innerhalb des Bereiches abgebildet. Die unterschiedlichen Ausbreitungswege für ein Szenario mit zwei Quellen und zwei Sensoren sind in Bild 2 dargestellt. Zur Rekonstruktion der Originalsignale werden die Phasendifferenzen in jedem Frequenzband statistisch analysiert. Die Ergebnisse der statistischen Auswertung liefern eine Vorschrift zur Rekonstruktion der Signale. Für weiterführende Informationen sei auf [4; 5] verwiesen.

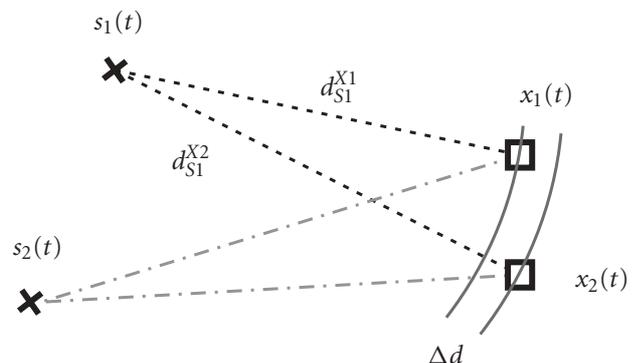


Bild 2 Szenario ohne Reflexionen. Die Kreissegmente markieren Strecken gleicher Entfernung von der Quelle $s_1(t)$.

Im Anschluss muss das Signal wieder in den Zeitbereich transformiert werden. Dafür findet die entsprechende inverse Transformation Verwendung.

3 Zeit-Frequenz-Analyse

Da sich aufgrund der Instationarität bei Sprachsignalen der Frequenzgehalt mit der Zeit ändert, ist eine Analyse weder ausschließlich im Zeit- noch im Frequenzbereich sinnvoll. Durch Transformation stationärer – bei Sprachsignalen ca. 20–30 ms langer – Abschnitte erhält man Darstellungsformen, die Änderungen des Frequenzgehaltes eines Signals zeitlich aufgelöst wiedergeben. Von besonderem Interesse sind die Kurzzeit-Fourier-Transformation (STFT) und die Wavelet-Transformation (WT) sowie deren Erweiterungen.

3.1 Kurzzeit-Fourier-Transformation

Bei der klassischen Fourier-Transformation wird das Innenprodukt der zu analysierenden Funktion $x(t)$ mit der komplexen harmonischen Schwingung $\exp(j2\pi ft)$ gebildet. Um den zeitveränderlichen Charakter nichtstationärer Signale zu berücksichtigen, wird bei der STFT das Signal $x(t)$ nur während des um τ verschobenen Zeitfensters $\gamma^*(t-\tau)$ betrachtet. Dieser Term wird der Fourier-Transformation unterworfen:

$$F_x^\gamma(\tau, f) = \langle x(t), \gamma(t-\tau) \exp(j2\pi ft) \rangle_t \quad (6)$$

$$= \int_{-\infty}^{\infty} x(t) \gamma^*(t-\tau) \exp(-j2\pi ft) dt.$$

Dabei ist das Fenster derart zu wählen, dass der Frequenzgehalt des Signals in diesem Bereich als nahezu stationär angenommen werden kann [6].

Neben der oben beschriebenen Interpretation lässt die Schreibweise als Innenprodukt eine weitere Deutung zu. Mit Hilfe des Innenprodukts wird die Ähnlichkeit des Signals $x(t)$ mit einem zeit- und frequenzverschobenen Fenster im Bereich um τ und f untersucht. Dies führt zu einer äquidistanten Aufteilung der Zeit-Frequenz-Ebene bei konstanter Zeitdauer und Bandbreite. Die Anwendung der inversen Fourier-Transformation liefert das Ursprungssignal $x(t)$. Aus der kontinuierlichen Transformation kann die diskrete Implementierung hergeleitet werden.

3.2 Wavelet-Transformation und Wavelet-Packets

Eine gleichmäßige Aufteilung der Zeit-Frequenz-Ebene ist nicht immer vorteilhaft. In vielen Signalen weisen hochfrequente Signalanteile eine starke Zeitvarianz auf, tieffrequente Anteile jedoch nicht. Für eine angepasste Analyse derartiger Signale wurde die Wavelet-Transformation entwickelt. Anstatt ein Fenster zeitlich und im Frequenzbereich zu verschieben, wird das sog. Mutterwavelet $\Psi(t)$ um die Zeit b verschoben und um

den Faktor a skaliert:

$$W_x^\Psi(a, b) = \langle x(t), \Psi_{a,b}(t) \rangle_t \quad (7)$$

$$= \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \Psi^*\left(\frac{t-b}{a}\right) dt. \quad (8)$$

Dadurch ändern sich Zeitdauer und Bandbreite des Vergleichssignals $\Psi_{a,b}(t)$ in Abhängigkeit von der mittleren Frequenz, die wiederum vom Skalierungsfaktor a abhängt [6].

Für niedrige Frequenzen ergibt sich somit eine hohe Frequenzauflösung, für hohe Frequenzen eine niedrige Auflösung. Die Zeitauflösung ist hingegen im hohen Frequenzbereich deutlich schärfer. Mittels der inversen Transformation können die Signale wieder in den Zeitbereich transformiert werden.

Bei zeitdiskreten Signalen lässt sich die diskrete Wavelet-Transformation (DWT) mit Hilfe von Multiratenfilterbänken realisieren. Die Berechnung der Wavelet-Koeffizienten erfolgt durch eine rekursive Zerlegung der Koeffizienten $c_k(m)$ mittels Faltung mit den Impulsantworten eines Tiefpassfilters g_{TP} bzw. eines Bandpassfilters g_{BP} . Im Anschluss daran erfolgt ein Downsampling. Die Wavelet-Funktionen werden nicht mehr explizit benötigt, sondern leiten sich aus den Filtern ab. Durch die Filterung der Signale wird die Zeitauflösung in jeder Filterstufe um den Faktor 2 verringert, die Frequenzauflösung entsprechend erhöht. Aus diesem Grund ist ein Downsampling (\downarrow) im Anschluss an die Filterung möglich. Die Koeffizienten $c_0(m)$ entsprechen dem Eingangssignal $x(t)$ zu den Abtastzeitpunkten $m \cdot t_A$. Die Vorgehensweise bei der Zerlegung der Koeffizienten ist in Bild 3 illustriert.

Die Rekonstruktion erfolgt mittels einer Filterbank mit inversen Filtern und entsprechenden Upsamplingoperatoren. Die Implementierung als Multiratenfilter ist besonders effizient. Die Anzahl der mathematischen Operationen ist von der Ordnung n [7].

Die Wavelet-Transformation ist jedoch nur eine Untermenge einer deutlich vielseitigeren Transformation – der Wavelet-Packet-Transformation. Bei letzterer Transformation werden in jeder Filterstufe alle Äste der Filterbank aufgespalten. Man erhält einen vollständigen Baum und somit eine redundante Darstellung des Signals, denn die Koeffizienten jeder Filterebene enthalten die vollständigen Signalinformationen. Jedem Knoten kann eine definierte Mittenfrequenz, Bandbreite und

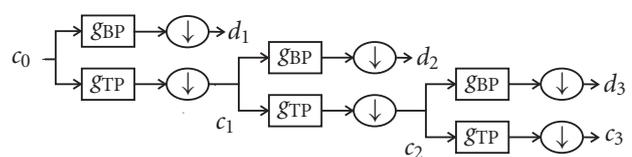


Bild 3 Multiratenfilter mit drei Filterstufen zur Realisierung der diskreten Wavelet-Transformation.

Zeitdauer zugeordnet werden. Durch die Auswahl bestimmter Knoten kann die Zeit-Frequenz-Ebene beliebig unterteilt werden. Die Auswahl der Knoten erfolgt mittels einer Minimierung des Approximationsfehlers, der die sogenannte „Beste Basis“ findet [8]. Der Vorteil dieses Verfahrens ist die optimale Anpassung der Frequenzachse an das Eingangssignal.

3.3 Analytische Wavelet-Packets

Die diskrete Wavelet-Transformation besitzt aufgrund der rein reellen Basisfunktionen einige signifikante Nachteile. Die Transformation ist verschiebungsvariant und es treten unerwünschte Oszillationen bei den Koeffizientenverläufen auf. Erklären lassen sich diese Nachteile durch die Oszillationen der reellen Bandpasssignale (Bild 4) [9]. Durch Erweiterung der DWT zur *Dual-Tree Complex Wavelet Transform* (DTCWT) [10] können die beschriebenen Probleme reduziert werden. Aus den reellen Wavelet-Funktionen ψ^{Re} wird mittels der Hilbert-Transformation der imaginäre Anteil ψ^{Im} berechnet. Damit ergeben sich die analytischen Wavelets zu

$$\psi^{\text{C}} = \psi^{\text{Re}} + j \psi^{\text{Im}} \quad (9)$$

Die Verwendung komplexer Wavelets liefert die komplexen Koeffizienten

$$d_k^{\text{C}}(m) = d_k^{\text{Re}}(m) + j d_k^{\text{Im}}(m) \quad (10)$$

Aufgrund der Linearität der obigen Gleichung können die Koeffizienten des Real- und Imaginärteils in zwei voneinander unabhängigen Filterbänke berechnet werden. Die Filter der imaginären Filterbank resultieren aus der Hilbert-Transformation der Filter der reellen Bank. Bei der Bestimmung der Filterbänke muss berücksichtigt werden, dass sich die Filter der ersten Stufe von den folgenden Filtern unterscheiden. Die Berechnung der DTCWT liefert Ergebnisse, die ähnliche Eigenschaften aufweisen wie die Resultate der Fourier- bzw. Kurzzeit-Fourier-Transformation. Der Betrag entspricht der komplexen Einhüllenden und weist keine Oszillationen auf, und die Phase zeigt ein nahezu lineares Verhalten. Der Aufwand zur Berechnung der Transformation ist doppelt so groß wie bei der Berechnung der DWT [11]. In Bild 4 sind reelle und analytische Basisfunktionen dargestellt. Die Ähnlichkeit zur Fourier-Transformation lässt sich auch aus dem Versatz zwischen reeller und imaginärer Impulsantwort von $\frac{\pi}{2}$ erkennen.

Die analytischen Wavelets können zu analytischen Wavelet-Packets erweitert werden. Diese Erweiterung ermöglicht, wie bereits in Abschnitt 3.2 für die reellen Wavelet-Packets beschrieben, eine signaladaptive Analyse des Eingangssignals $x(t)$. Eine einfache Bestimmung der Filterbänke wie bei der DTCWT ist jedoch nicht möglich. Zur Berechnung der analytischen Koeffizien-

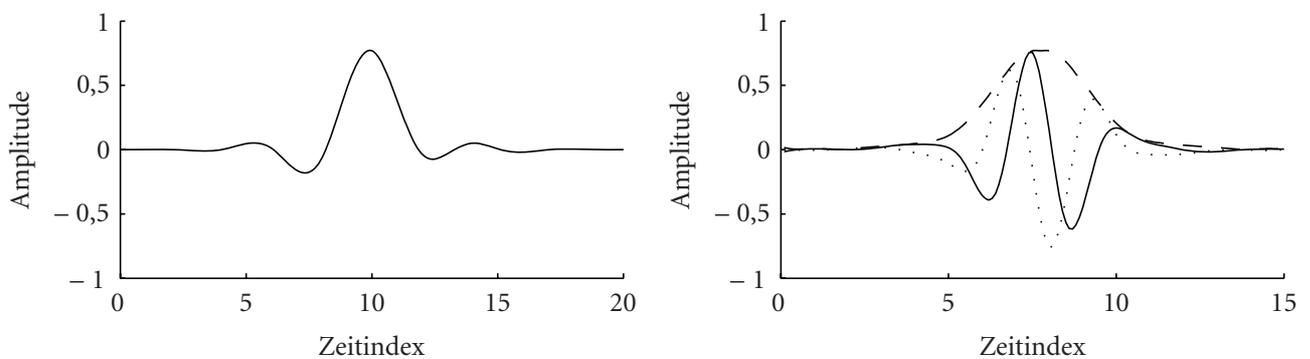


Bild 4 Impulsantworten reeller (links) und analytischer (rechts) Wavelets. Die drei Kurven im rechten Bild stellen den Betrag (—), den Realteil (---) und den Imaginärteil (···) des komplexen analytischen Wavelets dar.

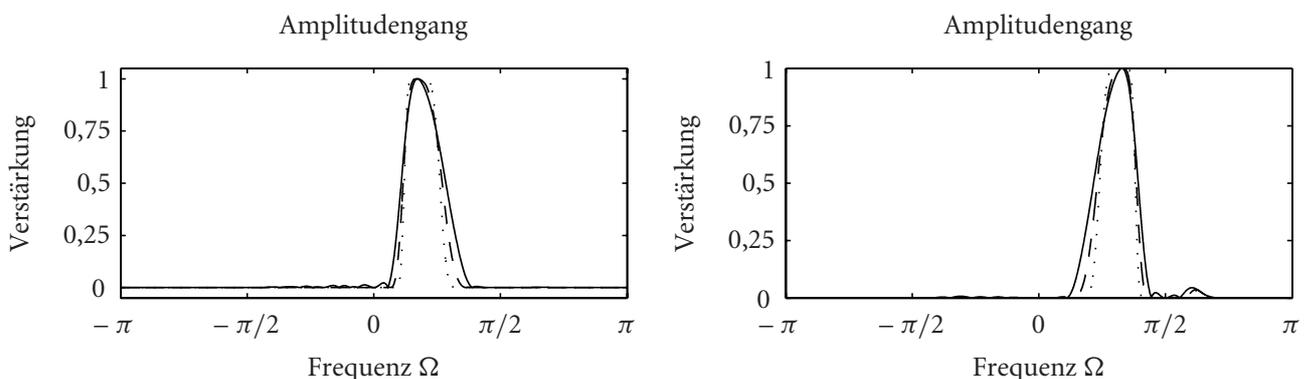


Bild 5 Amplitudengänge für zwei Pfade der Multiratenfilterbank in der dritten Stufe; durchgezogen: Q-Shift 14 (—), gestrichelt: Q-Shift 32 (---), gepunktet: Q-Shift 128 (···).

ten ist eine nichtintuitive Anordnung der Filter in der Baumstruktur notwendig [9; 12]. In Bild 5 sind stellvertretend für die analytischen Wavelet-Packets zwei Amplitudengänge dargestellt. Die analytische Form der Verläufe ist klar erkennbar. In der Arbeit werden die von Selesnick und Kingsbury entwickelten Q-Shift-Filter verwendet [11].

4 Verwendung zur Trennung unbekannter Signale

Bei der Separation unbekannter Signale können durch die statistische Auswertung der Phaseninformation die Originalsignale unter bestimmten Randbedingungen rekonstruiert werden. Aufgrund der oben beschriebenen Erweiterung der Wavelet-Transformation um einen imaginären Anteil ist ein Einsatz der DTCWT bzw. der analytischen Wavelet-Packets zur Transformation der Signale möglich. Im Folgenden werden nur die analytischen Wavelet-Packets betrachtet. Durch deren Einsatz können die Vorteile der adaptiven Signalanalyse genutzt werden.

Bevor die analytischen Wavelet-Packets mit der Kurzzeit-Fourier-Transformation verglichen werden, muss jedoch der Einfluss der Filterlänge betrachtet werden. Eine genauere Analyse der in Bild 5 dargestellten Amplitudengänge zeigt bereits ein Problem der Realisierung als Filterbank. Auch außerhalb des Bandes maximaler Amplitude gibt es Bereiche mit geringer Verstärkung. Treten signifikante Signalanteile in den Frequenzen der Nebenmaxima auf, werden diese dem falschen Frequenzband zugeordnet. Bei der Phaseninformation kommt es ebenfalls zu diesem Zuordnungsfehler. Eine Dämpfung der Nebenmaxima ist aus diesem Grund wünschenswert. Durch Erhöhung der Filterlänge kann dieser Effekt vermindert werden. Mit steigender Filterlänge erhöht sich auch die Symmetrie der Amplitudengänge.

Diese Zuordnungsfehler können bei der statistischen Analyse der Phaseninformation und der darauf basierenden Rekonstruktion zu Fehlern führen. Eine Abschätzung des Einflusses kann man durch die Betrachtung ei-

nes einfachen Szenarios (eine Quelle, zwei Sensoren) ohne Reflexionen gewinnen. Die Werte für $\Delta\varphi(f)$ müssen theoretisch eine erkennbar lineare Struktur aufweisen. Die statistische Analyse des Phasenverlaufs ist in Bild 6 dargestellt. Es werden normierte Histogramme von $\Delta\varphi(f_k)$ in jedem Frequenzband f_k berechnet und entlang der Frequenzachse aufgetragen. Die Histogramme werden für zwei unterschiedliche Filterkonfigurationen berechnet. Die Probleme der Zuordnung sind in den Bildern klar erkennbar. Insbesondere bei kurzen Filterlängen wirken sich diese Fehler deutlich aus. Bei der Verwendung der Filter sollte aus diesem Grund darauf geachtet werden, dass die Filter ausreichend lang sind. Abhängig von der Filterlänge erhöht sich zudem die Berechnungszeit der Transformation.

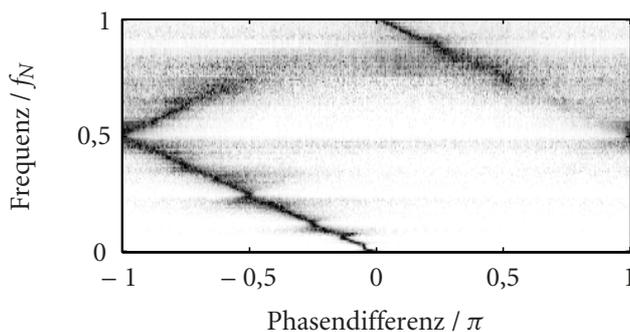
Die Nachteile der Multiratenfilterbank kommen bei der Kurzzeit-Fourier-Transformation nicht zum Tragen. In Bild 7 ist der etwas ausgeprägtere Verlauf der Phase bei der STFT für ein Szenario mit zwei Quellen erkennbar. Unter Berücksichtigung der aktuellen Betrachtungen scheint die Kurzzeit-Fourier-Transformation deutlich besser zur Separation der Signale geeignet zu sein.

Bei genauerer Betrachtung der analytischen Wavelet-Packets ergeben sich jedoch folgende Vorteile:

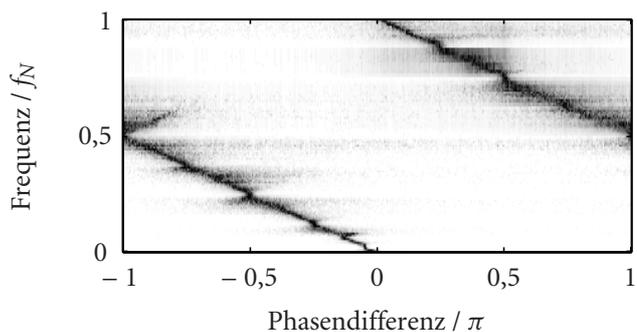
- adaptive Analyse des Signals,
- verbesserte Möglichkeiten zur Filterung.

Auf die Verfahren zur Filterung soll im Weiteren nicht genauer eingegangen werden. In Bild 8 sind die Histogramme der Phasendifferenzen zweier mit farbigem Rauschen überlagerter Signale dargestellt. Die lineare Struktur bleibt bei Verwendung der AWP in allen Frequenzbändern erhalten. Die verbesserten Filteralgorithmen [9] basieren auf den Ergebnissen der adaptiven Analyse des Signals.

Der signifikanteste Vorteil im Bereich der Separation liegt in der Adaptivität. Unter Vorgabe einer bestimmten Filtertiefe wird das Zeitsignal analysiert und mittels Basiswahl die optimale Aufteilung der Zeit-Frequenz-Ebene bestimmt. Auch wenn die theoretische Filtertiefe



(a) AWP – Kleine Filterlänge



(b) AWP – Große Filterlänge

Bild 6 Histogramme der Phasendifferenzen $\Delta\varphi(f_k)$ für kurze (links) und lange Filterlängen (rechts) in der ersten Filterstufe.

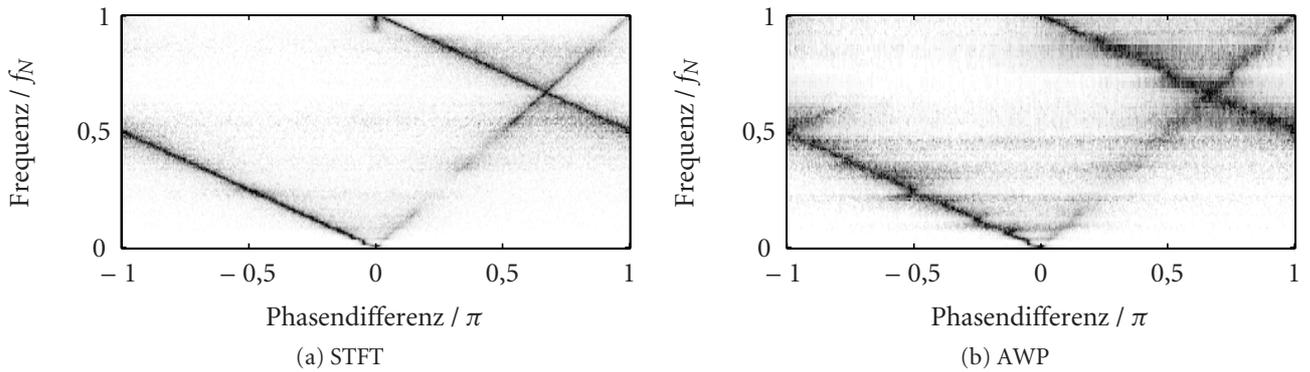


Bild 7 Histogramme der Phasendifferenzen $\Delta\varphi(f_k)$ nach Analyse der Zeitsignale mit unterschiedlichen Verfahren.

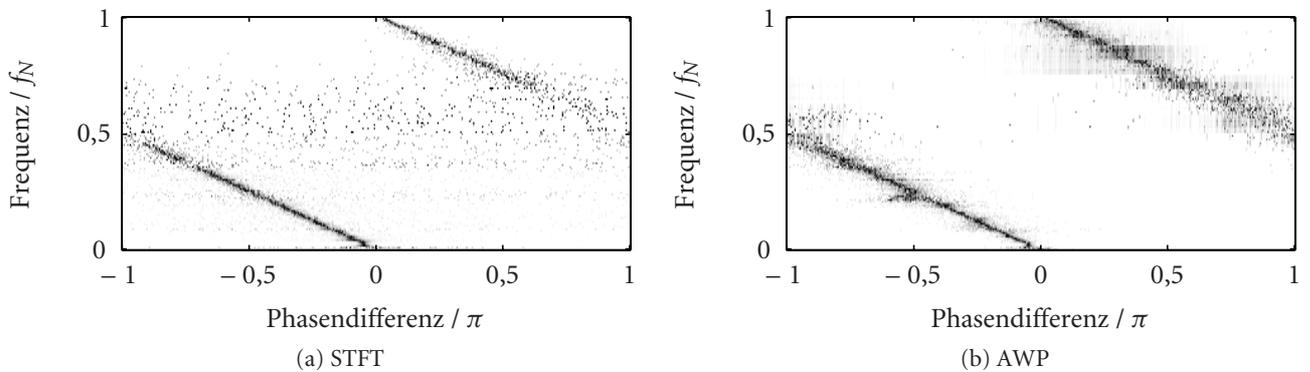


Bild 8 Histogramme der Phasendifferenzen $\Delta\varphi(f_k)$ nach Filterung der Zeit-Frequenz-Darstellung mit entsprechenden Filteralgorithmen. Das Originalsignal ist mit farbigem Rauschen überlagert.

der Frequenzaufteilung der STFT entspricht, wird das Signal normalerweise in weniger Bänder unterteilt. Für die Darstellung eines beliebigen Sprachsignals werden bei einer Filtertiefe von 10 Stufen bei der STFT stets 512 Frequenzbänder benötigt. Bei der Analyse des Signals mit den analytischen Wavelet-Packets wird das Signal jedoch in deutlich weniger Bänder (in diesem Beispiel in 230 Bänder) zerlegt. Summiert man die Koeffizientenbeträge auf, erhält man den in Bild 9 dargestellten Verlauf. Dieser zeigt anschaulich, welche Beiträge die Koeffizienten zum Signal liefern. Die reduzierte Auflösung der Frequenzebene erleichtert die statistische Analyse der

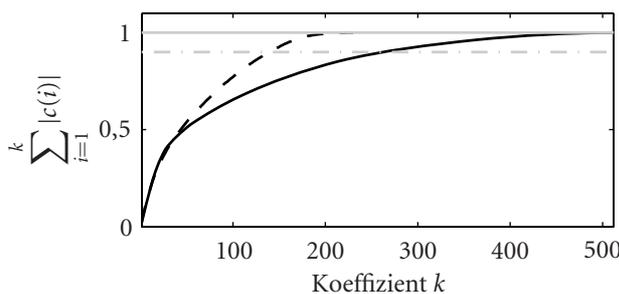


Bild 9 Kumulierte Summe der Koeffizientenbeträge, normiert auf 1. Die Koeffizienten sind absteigend nach Energiegehalt sortiert.

Signale. Es müssen weniger Frequenzbänder ausgewertet werden. Dadurch verkürzt sich die Berechnungszeit.

Die Ergebnisse bei der Trennung unbekannter Signale werden mit Hilfe der spektralen Verzerrung (*spectral distortion*) bewertet. Diese ist durch

$$SD = E_t \left\{ \sqrt{\frac{1}{F} \sum_{f=0}^{F-1} (20 \log |S(f, t)| - 20 \log |\hat{S}(f, t)|)^2} \right\} \quad (11)$$

definiert. $S(f, t)$ und $\hat{S}(f, t)$ sind die spektralen Koeffizienten des ursprünglichen und des rekonstruierten Signals, F ist die Anzahl der Frequenzbänder.

Um die Verwendbarkeit der Zeit-Frequenz-Darstellungen zur Rekonstruktion bestimmen zu können, wurden aus einer beliebigen Überlagerung zweier Sprachsignale die Originalsignale rekonstruiert. In Bild 10 sind die Originalsignale $s_i(t)$, die Mischsignale $x_i(t)$ und die rekonstruierten Signale $\hat{s}_i(t)$ dargestellt. Um den Einfluss des statistischen Verfahrens auf die Rekonstruktion zu reduzieren, wurde der Algorithmus 50-mal auf die Mischsignale angewendet. Die spektralen Verzerrungen wurden nach jedem Durchlauf berechnet. Der Mittelwert ist in Tabelle 1 eingetragen. Dieser lässt einen Rück-

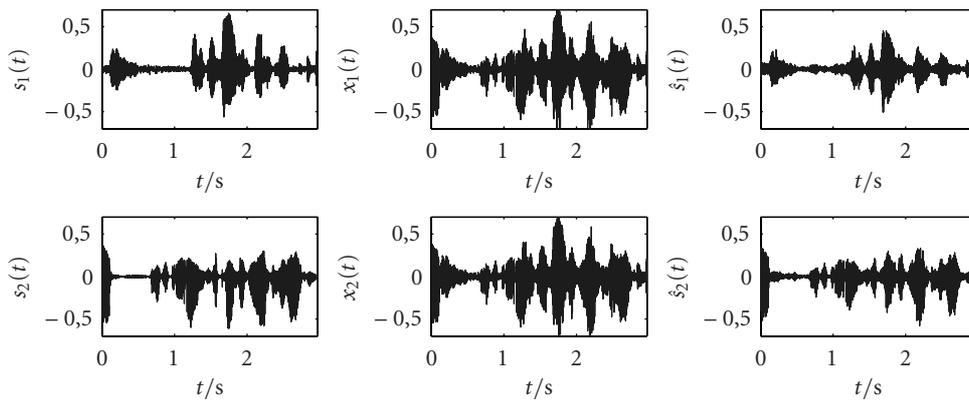


Bild 10 Originalsignale $s_i(t)$, Mischsignale $x_i(t)$ und rekonstruierte Signale $\hat{s}_i(t)$.

Tabelle 1 Spektrale Verzerrung der rekonstruierten Signale in Abhängigkeit der entsprechenden Originalsignale.

	STFT	AWP
rekonstruiertes Signal A	7,2297 dB	4,8455 dB
rekonstruiertes Signal B	7,7530 dB	5,5440 dB

schluss auf die Eignung der Zeit-Frequenz-Darstellungen zur Rekonstruktion der Signale zu. Eine Bewertung der Absolutwerte ist nicht sinnvoll, da diese vom Algorithmus zur Separation der Signale abhängen. Es ist jedoch erkennbar, dass die Resultate bei der Analyse mit AWP eine geringere Verzerrung aufweisen. Der Grund für diese Verbesserung liegt wiederum in der Adaptivität. Der Großteil der Energie konzentriert sich in weniger Koeffizienten. Werden diese bei der Rekonstruktion präzise zugeordnet, ist eine gute Rekonstruktion möglich. Durch die geringere Anzahl an Frequenzbändern treten zudem weniger Fehler bei der Rekonstruktion auf.

5 Zusammenfassung

Bei der Separation unbekannter Sprachsignale ist aufgrund der realen Umgebungseigenschaften der Übergang in den Zeit-Frequenz-Bereich notwendig. Bisher wurde stets die Kurzzeit-Fourier-Transformation zur Berechnung der spektralen Koeffizienten verwendet. Die Verwendung analytischer Wavelet-Packets bringt jedoch Vorteile mit sich. Durch eine signaladaptive Analyse können die Ergebnisse der statistischen Signalauswertung verbessert werden. Zudem werden die Signale bei gleicher Mindestfiltertiefe bei den analytischen Wavelet-Packets in weniger Frequenzbänder aufgeteilt. Der Faktor liegt bei etwa 0,5. Die Rechenzeit bei der Rekonstruktion verringert sich ebenfalls in dieser Größenordnung. Es muss jedoch berücksichtigt werden, dass die Berechnungszeit für die AWP höher ist als für die STFT.

Literatur

- [1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, 1997.
- [2] V. Hamacher und U. Rass, "Hightech im Ohr: physikalische und technische Grundlagen moderner Hörgeräte", *Physik in unserer Zeit* 37(2), 2006.
- [3] S. Haykin und Z. Chen, "The cocktail party problem", *Neural Computation* 17(9), S. 1875–1902, 2005.
- [4] P. O'Grady, B. Pearlmutter und S. Rickard, "Survey of sparse and non-sparse methods in source separation", *International Journal of Imaging Systems and Technology* 15(1), S. 18–33, 2005.
- [5] J.-F. Cardoso, "Blind signal separation: statistical principles", *Proceedings of the IEEE* 86, S. 2009–2025, Oct 1998.
- [6] U. Kiencke, M. Schwarz und T. Weickert, *Methoden der Signalverarbeitung*, Oldenbourg Wissenschaftsverlag, 2008.
- [7] A. Graps, "An introduction to wavelets", *IEEE Computational Science & Engineering* 2(2), S. 50–61, 1995.
- [8] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.
- [9] T. Weickert, *Nichtstationäre Filterung mit Hilfe Analytischer Wavelet Packets am Beispiel von Sprachsignalen*, Universitätsverlag Karlsruhe, 2009.
- [10] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals", *Journal of Applied and Computational Harmonic Analysis* 10(3), S. 234–253, 2001.
- [11] I. Selesnick, R. Baraniuk und N. Kingsbury, "The dual-tree complex wavelet transform", *IEEE Signal Processing Magazine* 22(6), S. 123–151, 2005.
- [12] T. Weickert, C. Benjaminsen und U. Kiencke, "Analytic Wavelet Packets – Combining the Dual-Tree Approach With Wavelet Packets for Signal Analysis and Filtering", *IEEE Transactions on Signal Processing* 57(2), S. 493–502, 2009.

Manuskripteingang: 3. Februar 2010, zur Veröffentlichung angenommen: 8. Februar 2010



Dipl.-Ing. Andreas Sandmair ist wissenschaftlicher Mitarbeiter am Institut für Industrielle Informationstechnik des Karlsruher Instituts für Technologie. Auf dem Gebiet der Signalverarbeitung forscht er an statistischen Verfahren zur Trennung unbekannter Sprachsignale und Echowunderdrückung.

Adresse: Karlsruher Institut für Technologie, Institut für Industrielle Informationstechnik, Hertzstraße 16, 76187 Karlsruhe,
E-Mail: sandmair@kit.edu



Mario Lietz ist Diplomand am Institut für Industrielle Informationstechnik des Karlsruher Instituts für Technologie. In seiner Studienarbeit untersuchte er die Nutzbarkeit der Analytischen Wavelet-Packets zur Separation überlagerter Sprachsignale.

Adresse: Karlsruher Institut für Technologie, Institut für Industrielle Informationstechnik, Hertzstraße 16, 76187 Karlsruhe,
E-Mail: mario.lietz@student.kit.edu



Prof. Dr.-Ing. Fernando Puente León leitet das Institut für Industrielle Informationstechnik am Karlsruher Institut für Technologie. Hauptarbeitsgebiete: Mess- und Automatisierungstechnik, Signalverarbeitung, automatische Sichtprüfung und Bildverarbeitung, Mustererkennung, Informationsfusion und Datenübertragung.

Adresse: Karlsruher Institut für Technologie, Institut für Industrielle Informationstechnik, Hertzstraße 16, 76187 Karlsruhe,
E-Mail: puente@kit.edu



Kein schlechtes Gewissen mehr



Werner Sesink

Einführung in das wissenschaftliche Arbeiten

Mit Internet – Textverarbeitung – Präsentation

7. Auflage 2007 | 262 S. | Broschur | € 22,80
ISBN 978-3-486-58191-1

In dieser Einführung werden die formalen und technischen Aspekte wissenschaftlichen Arbeitens nicht nur dargestellt, sondern auch im Hinblick auf ihre Bedeutung für ein engagiertes Studium relativiert. Ziel ist es, dem Studenten eine sichere Haltung gegenüber den Anforderungen des wissenschaftlichen Arbeitens zu geben. So kann er sich selbst ein begründetes Urteil bilden über den Sinn und Unsinn der vielen Standards, Gepflogenheiten und Vorschriften.

Dieses Buch richtet sich an Studierende geistes- und sozialwissenschaftlicher Fachrichtungen sowie an Oberstufenschülerinnen und -schüler.

Oldenbourg

Bestellen Sie in Ihrer Fachbuchhandlung oder direkt bei uns:
Tel: 089/45051-248,
Fax: 089/45051-333
verkauf@oldenbourg.de



Prof. Dr. Werner Sesink lehrt Allgemeine Pädagogik mit dem Schwerpunkt Bildung und Technik an der TU Darmstadt.



150 Jahre
Wissen für die Zukunft
Oldenbourg Verlag