

## Pro-Seminararbeit

# MFCC Abtastfenster und deren Auswirkung

vorgelegt von

David Kessler

MIN-Fakultät

Fachbereich Informatik

Studiengang: Informatik

Matrikelnummer: 7296634

Abgabedatum: 30.08.2020

Erstgutachter: Dr. Timo Baumann

### **Eidesstattliche Erklärung**

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit im Bachelorstudien-  
gang Informatik selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel —  
insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen — benutzt habe.  
Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als  
solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem  
anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf  
dem elektronischen Speichermedium entspricht.

Hamburg, den 26.08.2020

\_\_\_\_\_  
Vorname Nachname

### **Veröffentlichung**

Ich stimme der Einstellung der Arbeit in die Bibliothek des Fachbereichs Informatik zu.

Hamburg, den 26.08.2020

\_\_\_\_\_  
Vorname Nachname

# Inhaltsverzeichnis

Abbildungsverzeichnis	v
<b>1 Einleitung</b>	<b>1</b>
<b>2 Begriffserklärung</b>	<b>1</b>
2.1 Begriffe . . . . .	1
2.2 Abtastfenster . . . . .	2
<b>3 Auswirkung der Fenster auf das Sprachsignal</b>	<b>6</b>
3.1 Phonemen Vergleich . . . . .	7
3.2 Dynamic Programming distance . . . . .	7
<b>4 Deutung</b>	<b>8</b>
<b>5 Ausblick</b>	<b>8</b>
Quellen	10



# Abbildungsverzeichnis

2.1	Rechteck(uniform) Fenster . . . . .	2
2.2	Triangular(Bartlett) Fenster . . . . .	3
2.3	Hanning Fenster . . . . .	3
2.4	Hamming Fenster . . . . .	4
2.5	Parzen Fenster . . . . .	5
2.6	Welch Fenster . . . . .	5
2.7	Blackman Fenster . . . . .	6
2.8	Flat top Fenster . . . . .	6
3.1	Phonemen Vergleich . . . . .	7
3.2	Dynamic Programming distance . . . . .	8
4.1	Vergleich Phonemen(Grün) und Dynamic Programming distance(Rosa) . . . .	9



# 1 Einleitung

Mithilfe von Spracherkennung ist es möglich Sprache zu analysieren. Dabei gibt es unterschiedliche Techniken. Zum Beispiel ist es möglich, mit der Mel Frequency Cepstral Coefficients die Frequenzen eines Audiosignals nach der Melskala zuzuordnen. Hierbei wird die Fourier Transformation angewendet. Durch Abtastfenster wird das Signal untersucht und die ermittelten Frequenzen ausgegeben. Die Abtastfenster gibt es in verschiedenen Varianten. Je nach Signal und Fenster unterscheiden sich die ermittelten Frequenzen. Diese Proseminararbeit befasst sich mit den unterschiedlichen Abtastfenstern und ihre Auswirkungen auf das Frequenzresultat. Dabei werden unter anderem folgende Fenster untersucht und die Versuchsergebnisse verglichen: Rectangular, Triangular, Hamming, Hanning, Parzen, Welch, Blackman und Flat top.

## 2 Begriffserklärung

### 2.1 Begriffe

#### MFCC

Bei der MFCC wird ein Signal mit der Fourier-Transformation in die einzelnen Frequenzen zerlegt und diese an der Melskala ausgerichtet. Das Ergebnis ist die Aufteilung der Frequenzen, die für den Menschen unterscheidbar sind.[15]

#### Fourier-Transformation

Bei der Fourier-Transformation wird ein Signal in das Ursprungsspektrum zerlegt. Dabei wird das Signal mit Abtastfenster abgetastet. Da natürliche Signale nicht gleichmäßig sind und nicht wiederholbar sind, werden spezifische Abtastfenster verwendet. Diese minimieren die fehlenden Übergänge. Das Ergebnis ist ein Spektrum, das die Frequenzen in Dezibel anzeigt. [8, 13]

#### Side Lobes

Unter Side Lobes versteht man die ermittelten Amplituden von der Fourier-Transformation außerhalb der Haupt Amplitude (Main Beam). Diese fallen entsprechend dem Signal und Fenster unterschiedlich aus.[11]

## 2 Begriffserklärung

### Main Lobe Width

Dies bezeichnet die Länge(Dauer) der Haupt Amplitude vom Ergebnis der Fourier-Transformation.

### Dynamic Time Warping

Mithilfe von Dynamic Time Warping ist es möglich Datensätze unterschiedlicher Länge zu vergleichen. Dies ist für die unterschiedlichen Aussprachen von Wörtern wichtig. Zum Beispiel kann „Hallo“ kurz oder lang gezogen werden. Eine Erkennung bzw. Vergleichen ist nur möglich, wenn der gespeicherte Datensatz und das Eingesprochen angepasst werden.[2]

### Dynamische Programmierung

Unter Dynamische Programmierung versteht man das Aufteilen von Rechenoperationen. Diese bilden am Ende ein Gesamtergebnis.[1]

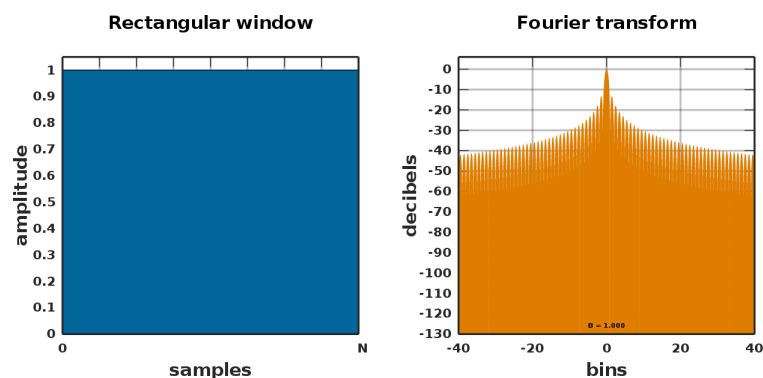
### Dynamic Programming distance

Die Dynamic Programming distance ist der Abstand zweier Werte. Beziehungsweise die Anzahl der Schritte, um ein Wert in den anderen umzuwandeln.[14]

## 2.2 Abtastfenster

### Rechteck(uniform) Fenster

Dieses Fenster ist ein einfaches Rechteck. Die Side Lobes bieten meistens eine ungenau Auflösung. Dementsprechend eignet es sich hauptsächlich für Impulse.[6, 8]



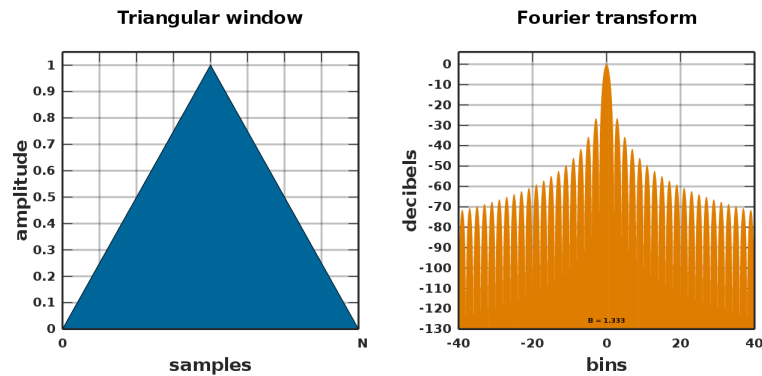
Quelle: [https://commons.wikimedia.org/wiki/File:Window\\_function\\_and\\_frequency\\_response\\_-\\_Rectangular.svg](https://commons.wikimedia.org/wiki/File:Window_function_and_frequency_response_-_Rectangular.svg)

Abbildung 2.1: Rechteck(uniform) Fenster



### Triangular(Bartlett) Fenster

Das Triangular Fenster ist ein einfaches Dreieck. Es bietet keine optimale Auflösung der Mainlobe und Side Lobes. Dafür bietet es eine gute Rauschanalyse.[7, 8]

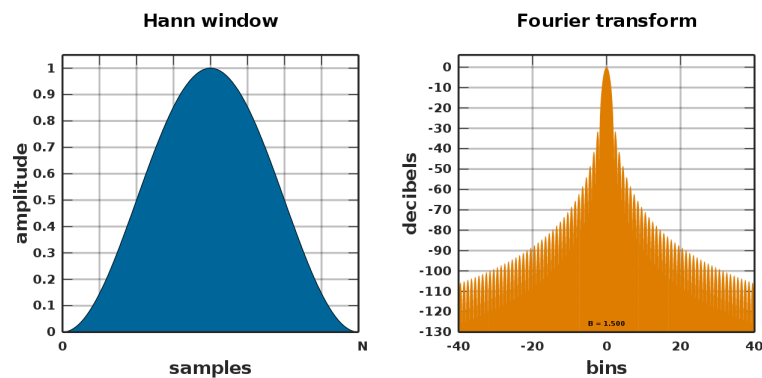


Quelle: [https://en.wikipedia.org/wiki/File:Window\\_function\\_and\\_its\\_Fourier\\_transform\\_%E2%80%93\\_Triangular\\_\(n\\_%3D\\_0...N\).svg](https://en.wikipedia.org/wiki/File:Window_function_and_its_Fourier_transform_%E2%80%93_Triangular_(n_%3D_0...N).svg)

Abbildung 2.2: Triangular(Bartlett) Fenster

### Hanning Fenster

Dieses Fenster bietet eine gut erkennbare Side Lobes Auflösung und Main Lobe Width. Dementsprechend wird es als Standard Fenster für ungenaue Signale eingesetzt.[5, 8]



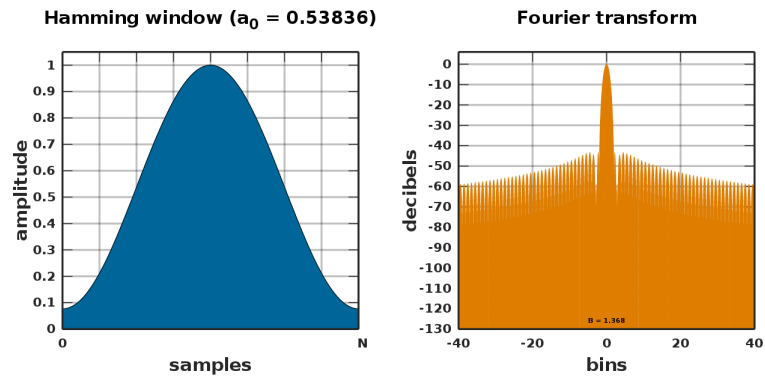
Quelle: [https://en.wikipedia.org/wiki/File:Window\\_function\\_and\\_its\\_Fourier\\_transform\\_-\\_Hann\\_\(n\\_%3D\\_0...N\).svg](https://en.wikipedia.org/wiki/File:Window_function_and_its_Fourier_transform_-_Hann_(n_%3D_0...N).svg)

Abbildung 2.3: Hanning Fenster

## 2 Begriffserklärung

### Hamming Fenster

Es ist ähnlich zum Hanning Fenster aufgebaut. Jedoch besitzt es keine null Amplitudenwerte. Demzufolge sind die ersten Side Lobes besser aufgelöst als die folgenden und die Main Lobe Width ist genauer. Das Fenster eignet sich gut für eng beieinanderliegenden Signalen.[4, 8]

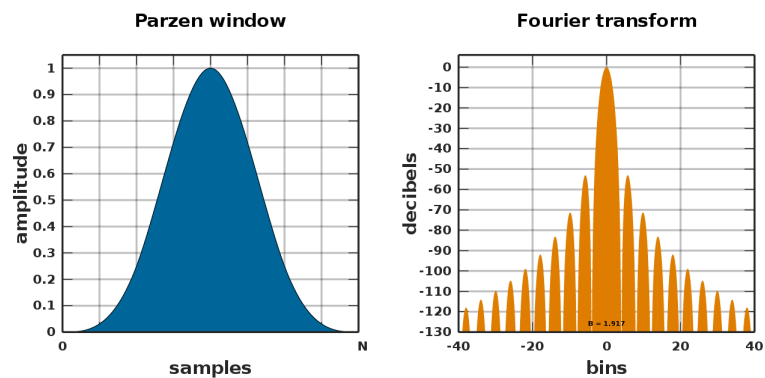


Quelle: [https://commons.wikimedia.org/wiki/File:Window\\_function\\_and\\_frequency\\_response\\_-\\_Hamming\\_\(alpha\\_%3D\\_0.53836,\\_n\\_%3D\\_0...N\).svg](https://commons.wikimedia.org/wiki/File:Window_function_and_frequency_response_-_Hamming_(alpha_%3D_0.53836,_n_%3D_0...N).svg)

Abbildung 2.4: Hamming Fenster

### Parzen Fenster

Folgendermaßen ist das Fenster aufgebaut. Es eignet sich hauptsächlich für Datainterpolation.[10]

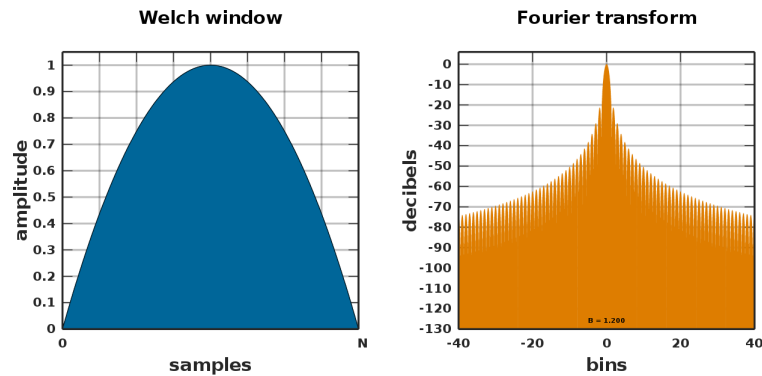


Quelle: [https://commons.wikimedia.org/wiki/File:Window\\_function\\_and\\_frequency\\_response\\_-\\_Parzen.svg](https://commons.wikimedia.org/wiki/File:Window_function_and_frequency_response_-_Parzen.svg)

Abbildung 2.5: Parzen Fenster

### Welch Fenster

Das Welch Fenster ist eine Parabel. Dieses eignet sich somit gut für Polynome.[12][12]

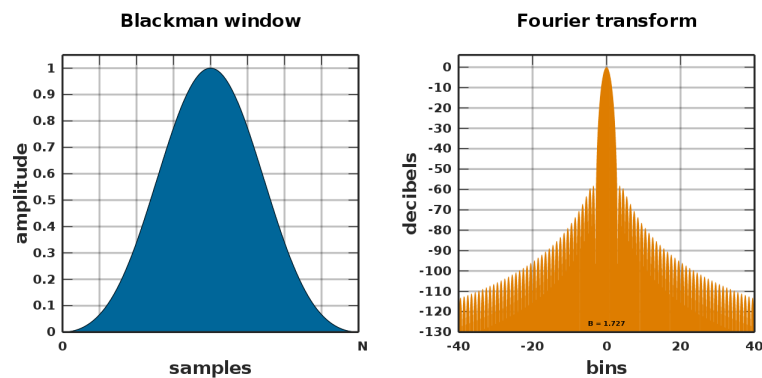


Quelle: [https://commons.wikimedia.org/wiki/File:Window\\_function\\_and\\_frequency\\_response\\_-\\_Welch.svg](https://commons.wikimedia.org/wiki/File:Window_function_and_frequency_response_-_Welch.svg)

Abbildung 2.6: Welch Fenster

### Blackman Fenster

Dieses Fenster ist ähnlich wie das Hanning Fenster aufgebaut. Es bietet eine genaue Amplituden Darstellung mit wenigen Side Lobes. Dieses Fenster eignet sich gut für die Audioanalyse.[3, 8]

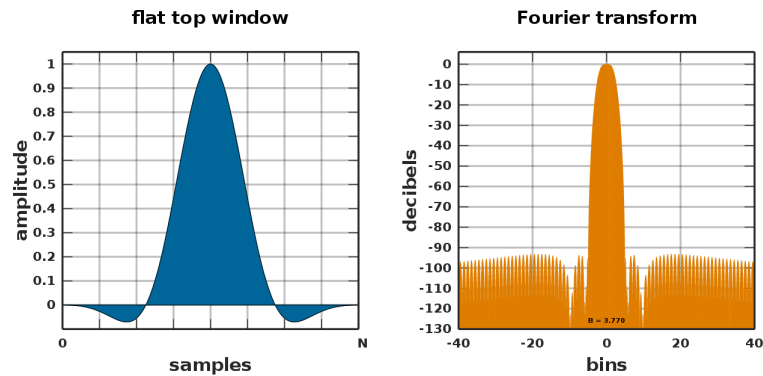


Quelle: [https://commons.wikimedia.org/wiki/File:Window\\_function\\_and\\_its\\_Fourier\\_transform\\_%E2%80%93\\_Blackman\\_\(n\\_%3D\\_0...N\).svg](https://commons.wikimedia.org/wiki/File:Window_function_and_its_Fourier_transform_%E2%80%93_Blackman_(n_%3D_0...N).svg)

Abbildung 2.7: Blackman Fenster

## Flat top Fenster

Das Flat top Fenster ist eine Amplitude, die ins Negative und Positive reicht. Demzufolge sind die Side Lobes am geringsten und die Main Lobe Width am größten. Es bietet dabei keine optimale Rauscherkennung. Somit eignet sich das Fenster zum Kalibrieren von Aufnahmegeräten.[8, 9]



Quelle: [https://commons.wikimedia.org/wiki/File:https://commons.wikimedia.org/wiki/File:Window\\_function\\_and\\_frequency\\_response\\_-\\_flat\\_top.svg](https://commons.wikimedia.org/wiki/File:https://commons.wikimedia.org/wiki/File:Window_function_and_frequency_response_-_flat_top.svg)

Abbildung 2.8: Flat top Fenster

## 3 Auswirkung der Fenster auf das Sprachsignal

Die unterschiedlichen Abtastfenster erzeugen unterschiedliche Ergebnisse. Es stellt sich die Frage wie ein optimaler Vergleich möglich ist. Hierfür werden zwei Möglichkeiten näher erläutert.

### 3.1 Phonemen Vergleich

Die einfachste Möglichkeit ist die Erfassung von Phonemen. Dies geschieht mithilfe der Ermittlung der Frequenzen (Peaks). Durch den Vergleich zur Originalanzahl der Phoneme eines Satzes kann die Differenz und somit die Genauigkeit erfasst werden. Der Beispielsatz „She had your dark suit in greasy wash water all year“ besitzt 37 Phoneme. Bei einer Analyse dieses Satzes werden folgende Werte erfasst:

Aus der Tabelle ist erkennbar, dass das Triangular und das Hamming Fenster die 37 Peaks besitzt. Die restlichen Fenster sind entweder zu sensitiv oder ungenau.[16]

#	Fenster Funktion	Peaks Count(Phoneme)
1	Rectangular	35
2	Triangular	37
3	Parzen	36
4	Welch	38
5	Hann	38
6	Hamming	37
7	Blackman	38
8	Nuttall	38
9	Flat top	36
10	Gaussian	38
11	Tukey	38
12	Planck-taper	36
13	Slepian	36
14	Kaiser	35

Quelle: [16, S.455]

Abbildung 3.1: Phonemen Vergleich

### 3.2 Dynamic Programming distance

Bei dieser Methode wird die Dynamic Programming distance ermittelt. Diese erfasst den Abstand zum hinterlegten (trainierten) Audiosignal. Als Beispiel wird das Wort „Hallo“ verwendet und mit vier unterschiedlichen Sprechweisen verglichen die ermittelten Werte sind folgende:

Name des Fensters Fenster Funktion	Dynamic Programming distance			
	Sprecher 1	Sprecher 2	Sprecher 3	Sprecher 4
Hamming	503.0783	606.4834	481.5268	558.8639
Hann	515.0137	613.8545	490.7390	568.1450
Rectangular	419.3535	522.4511	443.5564	501.7295
Kaiser	419.8803	523.4796	442.3071	501.9276
Blackman	543.5555	636.8195	509.1822	584.2372
Bartlett	499.2801	601.6781	476.4476	556.2980
Gaussian	516.4941	617.3757	489.9639	566.0629

Quelle: [14, S.6]

Abbildung 3.2: Dynamic Programming distance

Die Tabelle zeigt deutlich, dass das Rectangular Fenster die minimale Distanz aufweist. Gefolgt vom Kaiser Fenster. Das Hamming Fenster besitzt durchgehend eine der höchsten Distanzen.[14]

## 4 Deutung

Der Vergleich der Abtastfenster gibt unterschiedliche Ergebnisse. Es ist gut erkennbar, dass je nach Fokus des Vergleichs unterschiedliche Ergebnisse entstehen. Beim Phonem Vergleich schneidet das Hamming Fenster am besten ab. Das Rectangular Fenster ist dagegen am ungenauesten. Beim Vergleich der Dynamic Programming distance ist das Rectangular Fenster am genauesten und das Hamming Fenster ungenau. Die unterschiedlichen Ergebnisse können an den unterschiedlichen Audiofiles liegen. Der Phonemen Vergleich verwendet einen vollständigen Satz. Der Dynamic Programming distance dagegen nur ein Wort. Dies kann man auch Visuell erkennen:

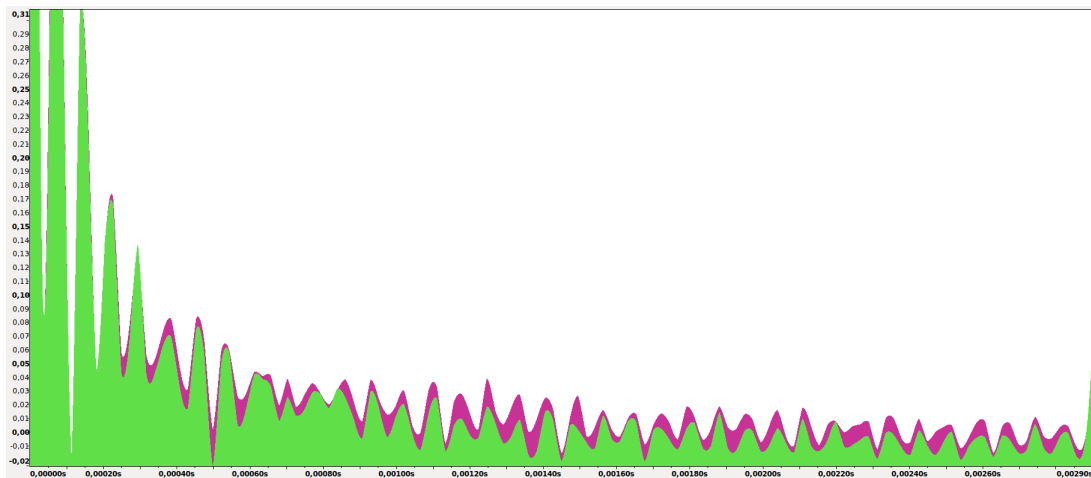


Abbildung 4.1: Vergleich Phonemen(Grün) und Dynamic Programming distance(Rosa)

## 5 Ausblick

Beide Methoden bringen unterschiedliche Ergebnisse hervor. Dabei treten auch Widersprüche auf. Für einen weiteren Vergleich sollten die Audiosignale gleich bzw. ähnlich komplex sein, um dieselben Ausgangsbedingungen zu erzeugen. Erst danach kann ein genauer Vergleich gezogen werden.

# Quellen

- [1] Sebastian Peuser (2012). *Dynamische Programmierung*. [Online]. 26. Aug. 2020. URL: <https://code-line.de/dynamische-programmierung/> (besucht am 26. 08. 2020).
- [2] Jeremy Zhang (2020). *Dynamic Time Warping*. [Online]. 26. Aug. 2020. URL: <https://towardsdatascience.com/dynamic-time-warping-3933f25fcdd> (besucht am 26. 08. 2020).
- [3] Julius Orion Smith. (2020). *Blackman window*. [Online]. 26. Juli 2020. URL: [https://ccrma.stanford.edu/~jos/sasp/Three\\_Term\\_Blackman\\_Harris\\_Window.html](https://ccrma.stanford.edu/~jos/sasp/Three_Term_Blackman_Harris_Window.html) (besucht am 26. 08. 2020).
- [4] Julius Orion Smith. (2020). *Hamming window*. [Online]. 26. Juli 2020. URL: [https://ccrma.stanford.edu/~jos/sasp/Generalized\\_Hamming\\_Window\\_Family.html](https://ccrma.stanford.edu/~jos/sasp/Generalized_Hamming_Window_Family.html) (besucht am 26. 07. 2020).
- [5] Julius Orion Smith. (2020). *Hann window*. [Online]. 26. Juli 2020. URL: [https://ccrma.stanford.edu/~jos/sasp/Hann\\_Poisson\\_Window.html](https://ccrma.stanford.edu/~jos/sasp/Hann_Poisson_Window.html) (besucht am 26. 08. 2020).
- [6] Julius Orion Smith. (2020). *Rectangular window*. [Online]. 26. Juli 2020. URL: [https://ccrma.stanford.edu/~jos/sasp/Rectangular\\_Window.html](https://ccrma.stanford.edu/~jos/sasp/Rectangular_Window.html) (besucht am 26. 07. 2020).
- [7] Julius Orion Smith. (2020). *Triangular window*. [Online]. 26. Juli 2020. URL: [https://ccrma.stanford.edu/~jos/sasp/Bartlett\\_Triangular\\_Window.html](https://ccrma.stanford.edu/~jos/sasp/Bartlett_Triangular_Window.html) (besucht am 26. 08. 2020).
- [8] Pico Technology. (2020). *Fast Fourier Transforms explained*. [Online]. 26. Aug. 2020. URL: <https://www.picotech.com/download/brochures/fast-fourier-transforms-explained.pdf> (besucht am 26. 08. 2020).
- [9] recordingblogs. (2020). *Flat top window*. [Online]. 26. Aug. 2020. URL: <https://www.recordingblogs.com/wiki/flat-top-window> (besucht am 26. 08. 2020).
- [10] recordingblogs. (2020). *Parzen window*. [Online]. 26. Aug. 2020. URL: <https://www.recordingblogs.com/wiki/parzen-window> (besucht am 26. 08. 2020).
- [11] recordingblogs. (2020). *Sidelobe falloff*. [Online]. 26. Aug. 2020. URL: <https://www.recordingblogs.com/wiki/sidelobe-falloff> (besucht am 26. 08. 2020).
- [12] recordingblogs. (2020). *Welch window*. [Online]. 26. Aug. 2020. URL: <https://www.recordingblogs.com/wiki/welch-window> (besucht am 26. 08. 2020).

## Quellen

- [13] National Instruments (2016 Dec. 07). *Understanding FFTs and Windowing*. [Online]. 26. Aug. 2020. URL: <https://download.ni.com/evaluation/pxi/Understanding%20FFTs%20and%20Windowing.pdf> (besucht am 26.08.2020).
- [14] S. M. S. Alam und S. Khan. "Response of different window methods in speech recognition by using dynamic programming". In: *2014 International Conference on Electrical Engineering and Information Communication Technology*. 2014, S. 1–6.
- [15] C. Kim und R. M. Stern. "Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24.7 (2016), S. 1315–1329.
- [16] Aparna Ramdoss. "Role of Windowing Techniques in Speech Signal Processing For Enhanced Signal Cryptography". In: Sep. 2017.