

Implementierung und Vergleich zweier Sprachcodierungsstrategien für Cochlea-Implantate

Sean-Patrik Cretti¹, Lewei He², Prof. Dr. Lilia Lajmi³

¹ Ostfalia Hochschule für angewandte Wissenschaften, 38302 Wolfenbüttel, E-Mail: se.cretti@ostfalia.de

² Ostfalia Hochschule für angewandte Wissenschaften, 38302 Wolfenbüttel, E-Mail: le.he@ostfalia.de

³ Ostfalia Hochschule für angewandte Wissenschaften, 38302 Wolfenbüttel, E-Mail: l.lajmi@ostfalia.de

Einleitung

Cochlea-Implantate ermöglichen gehörlosen und stark schwerhörigen Menschen wieder zu hören. Dabei werden Schallwellen in elektrische Impulse umgewandelt, die die Hörnerven in der Cochlea stimulieren. Im Jahr 2011 schätzte die Deutsche Cochlea Implantat Gesellschaft e.V. die Zahl der Cochlea-Implantat-Träger auf ca. 300.000 Menschen weltweit ein. In Deutschland nutzten ca. 30.000 Menschen Cochlea-Implantate. Durch eine bessere medizinische Versorgung steigt das Durchschnittsalter der Menschen. Das führt dazu, dass die Zahl der älteren Menschen ansteigt. Damit einher geht die steigende Anzahl an Menschen, die an Altersschwerhörigkeit leiden [1]. Da Cochlea-Implantate nicht nur für vollständig ertaubte Menschen geeignet sind, sondern auch für Menschen mit einem gewissen Grad an Schwerhörigkeit in Frage kommen, ist zu erwarten, dass die Zahl der Cochlea-Implantat-Träger in Zukunft steigen wird. Für ein gutes Sprachverstehen mittels Cochlea-Implantaten sind mehrere Faktoren wichtig. Zum einem ist die Anzahl der implantierten Elektroden relevant. Zum anderen spielt auch die genaue Platzierung der Elektroden in der Cochlea eine Rolle [2]. Ist das Cochlea-Implantat erfolgreich implantiert worden, so ist es wichtig, dass es an die Bedürfnisse des Patienten angepasst wird. Eine weitere wesentliche Rolle für das Sprachverstehen spielt die Sprachcodierungsstrategie. Sie ist die Software des Cochlea-Implantats. Diese Software verarbeitet das von einem Mikrofon aufgenommene Audiosignal, steuert die Elektroden in der Cochlea an und stimuliert somit die Hörnerven [3]. Während das Sprachverstehen in ruhiger Umgebung bei vielen Patienten gut ist, ist es in lauter Umgebung wie z.B. in Bars oder Restaurants für viele Cochlea-Implantat-Träger problematisch ein Gespräch zu führen (Cocktailparty-Effekt). Hier kann das Verbessern vorhandener Sprachcodierungsstrategien, sowie die Entwicklung neuer Algorithmen das Hörerlebnis von Cochlea-Implantat-Trägern verbessern.

Cochlea-Implantat (CI)

Zum Aufbau des Cochlea-Implantats: Es besteht aus einem äußeren Teil und einem inneren Teil. Das äußere Teil besteht aus einem Gehäuse. Darin befinden sich ein Mikrofon, ein Sprachprozessor sowie Batterien. Das Gehäuse sitzt hinter dem Ohr des Patienten und wird von einem Bügel gehalten. Damit über ein Kabel verbunden ist eine Sendespule. Diese wird mittels eines Magneten, der sich im inneren Teil des Cochlea-Implantats befindet an der Seite des Kopfes platziert.

Das innere Teil besteht neben dem bereits erwähnten

Magneten aus einer Empfangsspule, weiterer Elektronik zur Erzeugung der Impulse, sowie dem Elektrodenarray. Der genaue Aufbau sowie die Positionierung der Elemente des Cochlea-Implantats im Gehör des Patienten ist in Abbildung 1 gezeigt.

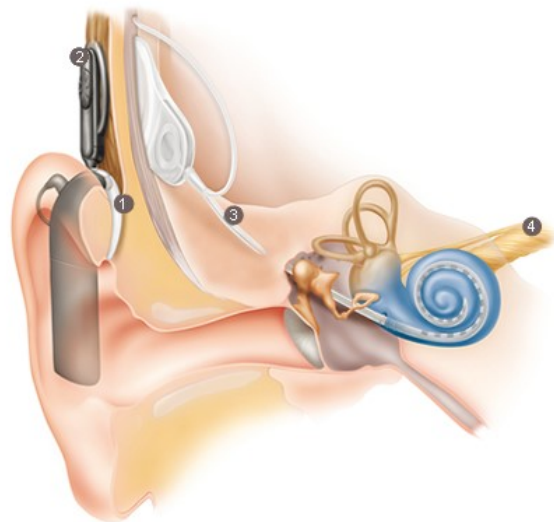


Abbildung 1: Darstellung eines Cochlea-Implantats [4]

1. Prozessor, zur Verarbeitung der Audiosignale, mit Mikrofon zum Erfassen der Schallwellen und Batterie/Akku für die Spannungsversorgung.
2. Sendespule, überträgt die Signale vom Prozessor an das Implantat.
3. Implantat, eine Empfangsspule nimmt das Signal entgegen. Die Elektronik generiert aus den Informationen elektrische Impulse.
4. Elektrodenarray, stimuliert die Hörnerven in der Cochlea.

Zur Funktionsweise: Die Schallwellen werden von dem Mikrofon aufgenommen und in ein elektrisches Signal gewandelt. Dieses elektrische Signal wird anschließend vom Sprachprozessor verarbeitet. Die Art der Verarbeitung hängt vom Cochlea-Implantat ab, da verschiedene Hersteller verschiedene Sprachcodierungsstrategien verwenden. Bekommt ein Patient ein Cochlea-Implantat eingesetzt, so muss es nach der Implantation noch durch das sogenannte Mapping an den Patienten angepasst werden. Diese patientenspezifischen Daten sind ebenfalls in dem Sprachprozessor gespeichert.

Nachdem das Signal vom Sprachprozessor verarbeitet wurde, werden die zu übertragenden Informationen an die

Sendespule weitergegeben. Diese übermitteln die Informationen dann per Induktion an die Empfangsspule. Außerdem wird durch die Sendespule die benötigte Energie für das Implantat an die Empfangsspule übertragen. Diese leitet die Informationen an die Elektronik weiter. In der Elektronik werden dann aufgrund der übermittelten Informationen die elektrischen Impulse erzeugt und über das Elektrodenarray ausgegeben. Durch die Impulse werden die Hörnerven an den Positionen der Elektroden stimuliert.

Algorithmen

In den folgenden Kapiteln werden die Sprachverarbeitungsstrategien Continuous Interleaved Sampling und Psychoacoustic Advanced Combinational Encoder beschrieben. Außerdem wird die Erweiterung dieser Strategien durch die Ergänzung von virtuellen Kanälen erklärt. Damit normalhörende die Sprachverständlichkeit der Algorithmen beurteilen können, müssen die elektrischen Impulse in ein Audiosignal gewandelt werden. Der dafür genutzte Vocoder wird ebenfalls in diesem Abschnitt erläutert.

Continuous Interleaved Sampling (CIS)

Bei der CIS-Strategie [5] werden alle Elektroden in der Cochlea des Patienten einzeln angesteuert. Das bedeutet, dass immer nur eine Elektrode mit einem Stromimpuls Nervenfasern in der Cochlea stimuliert. Dadurch sollen Interaktionen zwischen benachbarten Elektroden verhindert werden. In Abbildung 2 ist dies beispielhaft dargestellt.

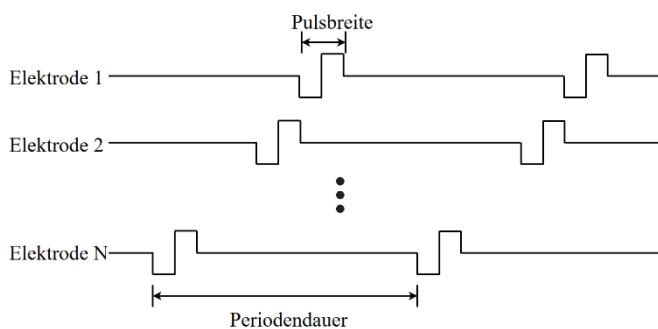


Abbildung 2: Biphase-Pulssequenz

Es ist zu erkennen, dass zu jedem Zeitpunkt der Pulssequenz nur ein Impuls mit einer definierten Pulsbreite gleichzeitig aktiv ist.

Anhand von Abbildung 3, welches das Blockschaltbild für den CIS-Algorithmus zeigt, wird im folgendem der genaue Ablauf beschrieben.

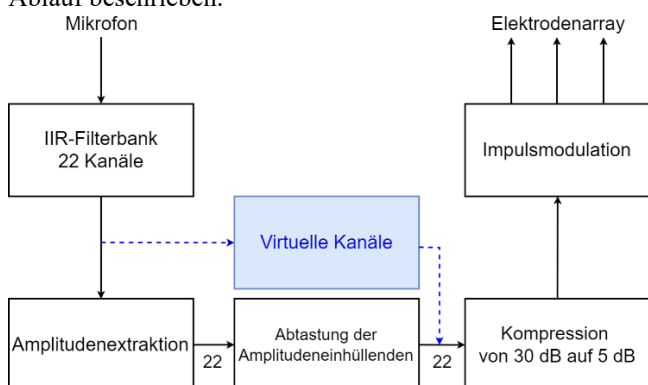


Abbildung 3: Blockschaltbild für CIS

Die wert- und zeitkontinuierlichen Schallwellen werden von einem Mikrophon aufgenommen und mit Hilfe eines A/D-Wandlers in ein digitales, zeit- und wertdiskretes, elektrisches Signal umgewandelt. Anschließend wird das Signal in 22 Kanäle eingeteilt. Diese Einteilung in die verschiedenen Kanäle erfolgt durch eine IIR-Filterbank, die aus 22 Bandpassfiltern besteht. Anschließend wird in dem Block Amplitudenextraktion von dem bandpassgefilterten Zeitsignal die Einhüllende berechnet. Diese wird danach unterabgetastet, sodass es für jeden Impuls einen Amplitudenwert gibt. Mittels einer logarithmischen Kompressionsfunktion wird für die Kanäle der Dynamikbereich der akustischen Amplitude (ca. 30dB bei Sprache) auf ca. 5dB komprimiert und somit an den wahrnehmbaren Dynamikbereich von CI-Trägern angepasst. Dies ist normalerweise ein patientenspezifischer Vorgang. Zum Schluss wird die komprimierte Einhüllende mit den Impulsen moduliert und an die Elektroden gesendet. Die CIS-Strategie kann um virtuelle Kanäle erweitert werden. Dies wird durch den blauen Pfad im Blockschaltbild dargestellt. Durch virtuelle Kanäle kann die Frequenzauflösung bis zu einem gewissen Grad erhöht werden. Die genaue Funktionsweise von virtuellen Kanälen wird im späteren Verlauf dieser Arbeit erklärt.

Psychoacoustic Advanced Combinational Encoder (PACE)

Bei der PACE-Strategie [6] werden nicht alle Elektroden für jedes Frame aktiviert. Mit Hilfe eines psychoakustischen Modells werden bestimmte Elektroden ausgewählt. Abbildung 4 stellt das Blockschaltbild der PACE-Strategie grafisch dar.

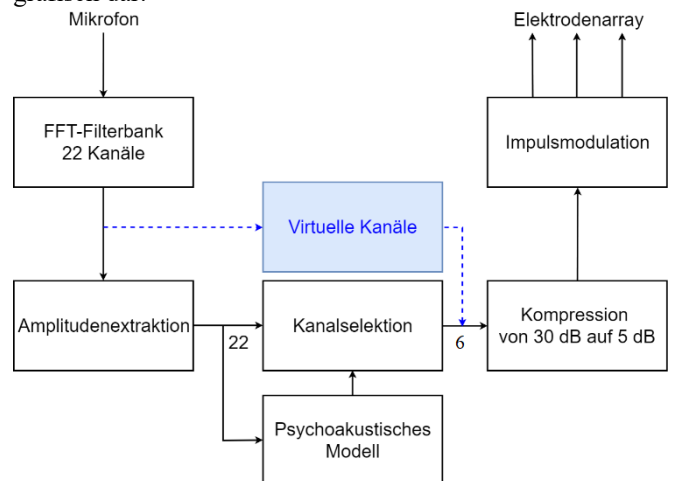


Abbildung 4: Blockschaltbild der PACE-Strategie

Simultan zur Vorgehensweise des CIS-Algorithmus, werden die Schallwellen zunächst diskretisiert. Anschließend wird das digitale Audiosignal mit einer FFT-Filterbank in 22 Kanäle zerlegt. Auch der Block Amplitudenextraktion funktioniert wie in CIS beschrieben. Danach werden mit dem psychoakustischen Modell die 6 Kanäle selektiert, deren Informationsgehalt für das menschliche Gehör am höchsten ist. Als Kriterium dafür wird die spektrale Maskierung berücksichtigt. In Abbildung 5 ist die Selektion der Kanäle beispielhaft dargestellt.

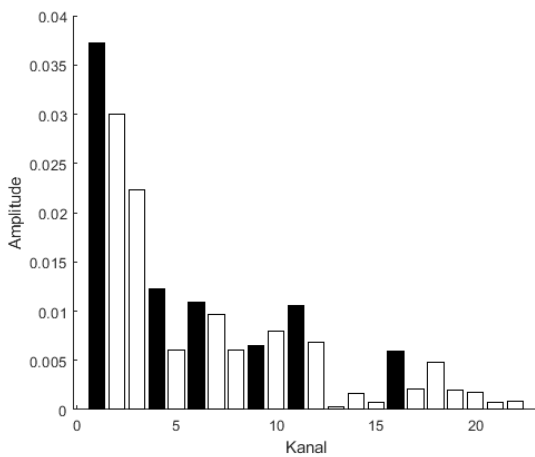


Abbildung 5: Kanalselektion auf Grund des psychoakustischen Modells

Die schwarz markierten Kanäle werden vom Algorithmus ausgewählt. Es ist zu sehen, dass einige Kanäle mit hohen Amplituden nicht ausgewählt werden. Der Grund dafür ist, dass die Kanäle von den bereits ausgewählten Kanälen maskiert wurden und dadurch nicht mehr von Menschen wahrnehmbar sind. Nachdem die Kanäle ausgewählt wurden, werden die Amplituden wie bei der CIS-Strategie komprimiert, mit elektrischen Impulsen moduliert und an das Elektrodenarray gesendet. Auch bei dieser Strategie kann das Modell um virtuelle Kanäle erweitert werden. Dies ist mit dem blauen Pfad in Abbildung 4 dargestellt. Im folgenden Kapitel wird die Funktionsweise von virtuellen Kanälen erklärt.

Virtuelle Kanäle

Durch die Anzahl an Elektroden ist die spektrale Auflösung stark limitiert. Trotz dieser Limitierung ist es Cochlea-Implantat-Trägern möglich Sprache in ruhiger Umgebung zu verstehen. Gibt es allerdings viele Nebengeräusche, wie z.B. in einer Bar, ist eine höhere spektrale Auflösung hilfreich. Instrumente erkennen und Melodien wahrnehmen ist ebenfalls nur mit einer erhöhten spektralen Auflösung möglich [7]. Townshend et al. berichteten 1987, dass zusätzliche Tonhöhen von Patienten wahrgenommen werden, wenn zwei benachbarte Elektroden gleichzeitig angesteuert werden [8]. Welche Tonhöhe wahrgenommen wird, kann durch den Strom, der durch die Elektroden fließt, gesteuert werden. Die wahrgenommene Lautstärke bleibt dabei gleich, so lange die Summe, der durch die beiden Elektroden fließenden Ströme konstant ist [7]. Um diese Stromsteuerung zu realisieren, musste jedoch eine neue Generation von Cochlea-Implantaten entwickelt werden. Denn bei den bisherigen Modellen war es nicht möglich zwei Elektroden gleichzeitig mit Strom zu versorgen. Die neu entwickelten Cochlea-Implantate versorgen jede Elektrode mit einer eigenen Stromquelle. So ist es möglich zwei Elektroden gleichzeitig anzusteuern und die Stromstärke zu variieren. In Abbildung 6 ist das Prinzip der Stromsteuerung dargestellt.

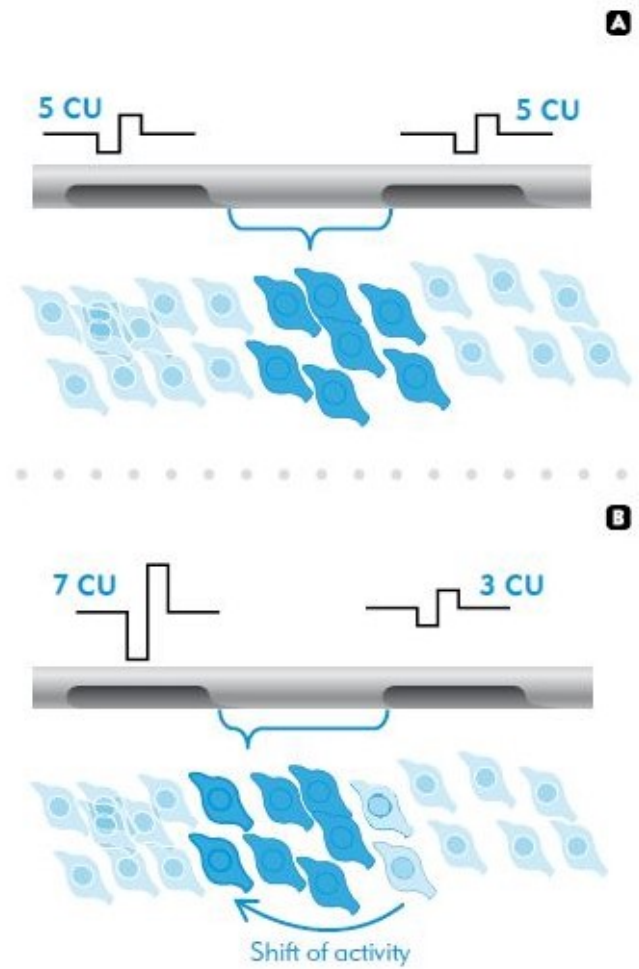


Abbildung 6: Prinzip der Stromsteuerung [7]

In Teil A der Abbildung ist zu sehen, wie zwei benachbarte Elektroden jeweils mit 5 Stromeinheiten (Current Units, CU) angesteuert werden. Dadurch werden nicht, wie vielleicht zu erwarten wäre, die jeweiligen Nervenfasern unter den Elektroden aktiviert. Stattdessen werden die Nervenfasern in der Mitte zwischen den beiden Elektroden aktiviert. So nimmt der Patient einen anderen Ton wahr. Wenn nun das Stromstärkeverhältnis zwischen den Elektroden verändert wird, ändert sich auch der Ort, an dem die Nervenfasern aktiviert werden. Dies ist in Teil B der Abbildung zu sehen. Nun wird die erste Elektrode mit 7 CU angesteuert und die zweite Elektrode mit 3 CU. Dadurch verschiebt sich die wahrgenommene Tonhöhe in Richtung der ersten Elektrode.

Synthese

Für die Validierung unserer Implementierungen wird aus den Impulsen ein Audiosignal synthetisiert, sodass normalhörende die Audioqualität bzw. Sprachverständlichkeit beurteilen können.

Dafür wurde ein sinusoidaler Vocoder verwendet. Die Berechnung zur Generierung des Audiosignals ist in Formel (1) dargestellt [9].

$$s(n) = \sum_{k=1}^N A_k(n) \cdot \sin(2\pi f_k n) \quad (1)$$

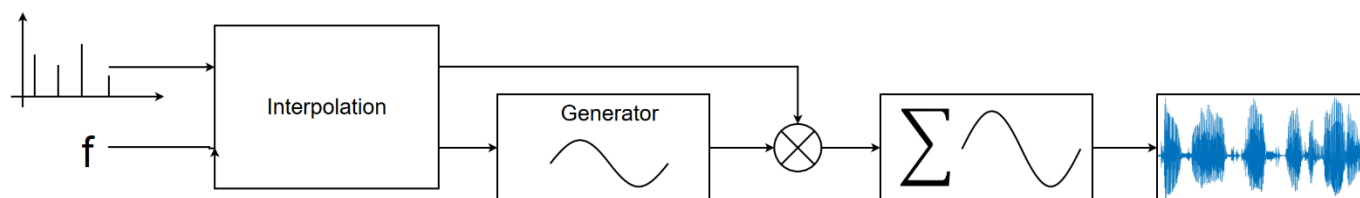


Abbildung 7: Blockschaltbild des sinusoidalen Vocoders

A_k entspricht der Amplitude des k -ten Kanals. Bei der PACE-Strategie werden nicht alle Elektroden angesteuert. Die Amplitude für die Elektroden, die nicht angesteuert werden, beträgt 0. Der Parameter f_k entspricht der Mittelfrequenz des jeweiligen Kanals. N ist die Anzahl der Kanäle. Um Sprünge zwischen den Frames zu vermeiden wird die Amplitude zwischen den Frames interpoliert. In Abbildung 7 ist der soeben beschriebene Vocoder in Form eines Blockschaltbildes dargestellt.

Fazit und Ausblick

Damit die in dieser Arbeit vorgestellten Algorithmen mit normalhörenden getestet werden können, mussten die elektrischen Impulse mittels eines Vocoders in Audiosignale gewandelt werden. Die Qualität der Audiosignale ist mit dem verwendeten Vocoder allerdings zu gut, sodass Ergebnisse von Sprachverständlichkeitstests nicht sehr stark mit denen von CI-Trägern korrelieren. Um die Sprachverständlichkeit der verschiedenen Sprachcodierungsstrategien mittels Normalhörenden zu bewerten, ist der in dieser Arbeit vorgestellte und genutzte Vocoder daher nur begrenzt geeignet.

Es bedarf eines Vocoders, der wichtige Eigenschaften des Gehörs berücksichtigt, und somit die Sprachverständlichkeit des synthetisierten Audiosignals besser mit dem Hörempfinden von Cochlea-Implantat-Trägern korreliert. Dafür kann der Vocoder beispielsweise um eine Vorstufe erweitert werden. Eine solche Vorstufe sollte das durch die Impulse sich ausbreitende elektrische Feld simulieren. Dadurch kann bestimmt werden, in welchem Bereich der Cochlea Hörnerven durch einen Impuls aktiviert werden. So wird das Übersprechen von Elektroden berücksichtigt. Außerdem ist das Einbeziehen der Refraktärzeit (Erholungszeit) der Hörnerven sinnvoll. Durch diese Erweiterung ist es möglich, dass Hörnerven trotz eines elektrischen Impulses nicht feuern, da sie sich noch von der Aktivierung erholen müssen. Eine solche Vorstufe wurde in [10] beschrieben und realisiert.

Literatur

- [1] Huhnd, L. E.: Einfluss der Cochlea Implantat Versorgung auf Lebensqualität, Sprachverstehen, Tinnitus und psychische Komorbiditäten unter besonderer Berücksichtigung älterer Patienten. Klinik für Hals-, Nasen-, Ohrenheilkunde der Medizinischen Fakultät Charité – Universitätsmedizin Berlin (2011)
- [2] Hochmair, I. et al.: Deep electrode insertion and sound coding in cochlear implants. *Hearing Research* 322 (2015), 14-23
- [3] Nägeli, A. M.: Vergleichende Studie einer neuen Sprachcodierungsstrategie für Cochlea-Implantate. Medizinische Fakultät der Universität Zürich (2013)
- [4] Cochlear, URL: <https://www.cochlear.com/de/startseite/hoeren-und-hoerverlust/hoeren-und-hoerverlust/behandlungsmoeglichkeiten/cochlea-implantate>
- [5] Wilson, B. S.: Better speech recognition with cochlear implants. *Nature* Vol. 352 No. 6332 (1991), 236-238
- [6] Nogueira, W. et al.: A Psychoacoustic „NofM“-Type Speech Coding Strategy for Cochlear Implants. *EURASIP Journal on Applied Signal Processing* 18 (2005), 3044-3059
- [7] Advanced Bionics, URL: https://advancedbionics.com/content/dam/advancedbionics/Documents/Global/en_ce/Professional/Technical-Reports/Sound-Processing/AB_HiRes_Fidelity_120_Sound_Processing_report.pdf
- [8] Townshend, B. et al.: Pitch perception by cochlear implant subject. *The Journal of the Acoustical Society of America* Vol. 82 (1987), 106-115
- [9] Vondrášek, M.: Real Time Simulations for Cochlear Implants. *Analysis of Biomedical Signals and Images* 20, 129-132
- [10] El Boghdady, N. et al.: A neural-based vocoder implementation for evaluating cochlear implant coding strategies. *Hearing Research* 333 (2016), 136-149