ZUM EINSATZ VON ELEKTROENZEPHALOGRAPHIE BEI DER MESSUNG DER WAHRNEHMUNG GESTÖRTER SPRACHE

Sebastian Möller[1], Jan-Niklas Antons[1], Sebastian Arndt[1], Anne K. Porbadnigk[2], Robert Schleicher[1]

[1]Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin
[2]Machine Learning Laboratory, TU Berlin und
GRK Sensory Computation in Neural Systems, BCCN, Berlin
sebastian.moeller@telekom.de

Abstract: In diesem Beitrag geben wir einen Überblick über Arbeiten zur Erfassung auditiv und audiovisuell wahrgenommener Störung von Sprachreizen mittels Elektroenzephalographie (EEG). Unser Augenmerk liegt dabei auf der Unterscheidung von bewussten und nichtbewussten Anteilen der Verarbeitung, um Aufschluss über die der Wahrnehmung und Beurteilung zugrunde liegenden Prozesse zu bekommen. Ziel ist die Entwicklung von Messverfahren, mit deren Hilfe Störungen von Sprachsignalen auch ohne direkte Befragung von Versuchspersonen quantifiziert werden können. Als erste Schritte auf dieses Ziel hin werden die Ergebnisse von zwei Experimenten beleuchtet, für die Ereigniskorrelierte Potenziale bei kurzen auditiv oder audiovisuell präsentierten Sprachlauten bzw. -silben analysiert werden. Es zeigt sich eine Veränderung des Eintrittszeitpunktes sowie der Amplitude einer bestimmten Komponente des Hirnstrom-Potenzials (P300) nach einem gestörten Stimulus gegenüber einem ungestörten Stimulus. In einer dritten Studie wird die Auswirkung von gestörten langen auditiven Sprachstimuli auf den Zustand des Zuhörers gezeigt; gemessen wurden hier Variationen in Frequenzbändern der neuronalen Aktivität. Bei diesen Sprachstimuli konnte eine erhöhte Aktivierung in Frequenzbändern, welche mit Ermüdung assoziiert werden, gemessen werden. Die entwickelten Verfahren werden in Bezug auf ihre Sensitivität und ihre praktischen Einsatzmöglichkeiten diskutiert.

1 Einleitung und Motivation

Trotz erheblicher Fortschritte bei den Technologien hat sich die Qualität übertragener Sprache in den letzten Jahren zu einem wichtigen Thema für die Anbieter von Telekommunikationsdiensten entwickelt. Dies liegt zum einen an der immer größer werdenden Vielfalt von Übertragungstechniken, Netzen und Endgeräten, die verschiedene Szenarien zum Teil sogar während einer einzigen Sprachverbindung ermöglichen, z.B. schmalbandige vs. breitbandige Übertragung, einohrige Handapparatbeschallung vs. beidohrige Headset-Beschallung, GSM-Netzübertragung vs. WLAN-Telefonie, auditive vs. audiovisuelle Darbietung. Zum anderen haben sich auch die Erwartungen der Nutzer an ein „normales“ Telefonat geändert: Nutzer, die beruflich viel über breitbandige Verbindungen telefonieren, haben einen anderen Qualitätsanspruch als solche, die viel über schmalbandige Handyverbindungen telefonieren. Da Qualität als Ergebnis eines Wahrnehmungs- und Beurteilungsprozesses verstanden werden muss (vgl. [1]), bei dem das Wahrgenommene („Hörereignis“) mit etwas Erwünschtem oder Erwartetem („Referenz“) verglichen wird, wirken sich sowohl das durch die Übertragungstechnik beeinflusste Hörereignis als auch die durch die Erfahrung und die Beurteilungssituation beeinflusste Referenz auf das Qualitätsurteil aus.

Derzeit wird die Qualität übertragener oder anderweitig gestörter Sprache normalerweise mit Beurteilungstests gemessen. Dabei werden Versuchspersonen verschiedene Stimuli meist in einer passiven Hör-, Seh- oder Hör- und Sehsituation (bei audiovisueller Übertragung)

vorgespielt, und die Versuchspersonen werden gebeten, für jeden Stimulus oder für jedes Stimuluspaar ein quantitatives Urteil auf einer Skala abzugeben (vgl. z.B. ITU-T Rec. P.800 [2]), ggf. kontinuierlich über einen längeren Zeitraum [3]. Solche Urteile bedingen neben sensorischen Wahrnehmungsprozessen auch einer Introspektion, nämlich Bewusstbarmachung und Einordnung der Qualität der vorgespielten Stimuli in den Zusammenhang der zuvor gemachten Erfahrungen. Der genaue Ablauf der Urteilsbildung und die dabei relevanten Prozesse sind jedoch bislang nur in Ansätzen erforscht.

Wir gehen derzeit davon aus, dass der Vergleich zwischen Wahrgenommenem und Referenz auf verschiedenen Ebenen stattfindet. Nach der peripheren auditorischen Verarbeitung im Hörorgan bzw. visuellen Verarbeitung im Sehorgan werden die eingehenden Nervenimpulse zunächst auf einer sensorischen Ebene analysiert. Dabei werden sensorische Merkmale aus den Eingangsinformationen extrahiert und mit einer (sensorischen) Referenz verglichen. Danach findet wahrscheinlich ein Vergleich auf einer höheren Objektebene statt; für diesen sind höhere Verarbeitungsstufen im Gehirn notwendig. Erst auf dieser Verarbeitungsstufe bildet sich das tatsächliche Wahrnehmungsereignis. Dieses wird dann in einer dritten Vergleichsstufe mit einer internen oder externen, durch die Beurteilungssituation getriggerten Referenz verglichen, wodurch sich das Qualitätsereignis bildet. Dieses ist allerdings nicht direkt zugänglich, sondern bedarf einer Beschreibung durch die wahrnehmende Person. Die einzelnen Verarbeitungsschritte sind bislang Hypothesen, da etablierte Messmethoden zu ihrer Bestimmung noch fehlen.

In diesem Betrag möchten wir Messverfahren diskutieren, welche in der Lage sind, einzelne an der Qualitätsbildung beteiligte Prozesse analysierbar zu machen. Im Gegensatz zum normalerweise angewandten Verfahren der Introspektion und anschließenden Beschreibung auf einer Bewertungsskala versuchen wir mit diesen Messverfahren, nicht bewusste und bewusste Prozesse der Probanden zu erfassen: Die einzige Aufgabe der Probanden ist die Quittierung wahrgenommener Störungen (d.h. das Zuhören bzw. Zusehen), auch wenn zur Analyse der Aussagekraft der Ergebnisse dieser Verfahren derzeit noch Qualitätsurteile von den Probanden mit erhoben werden (dies wird später überflüssig sein). Wir gehen davon aus, dass die Versuchsaufgabe und dadurch der Aufmerksamkeitsfokus der Probanden einen nicht unerheblichen Einfluss auf das Qualitätsereignis haben und dieses u.U. verfälschen kann; dies wird künftig durch eine solche rein „passive“ Messung vermieden.

In Abschnitt 2 gehen wir kurz auf das von uns zunächst gewählte Messverfahren ein. Darauf folgen in Abschnitt 3 und 4 Zusammenfassungen von Anwendungen dieses Messverfahrens zur Erfassung von Störungen auditiv bzw. audiovisuell dargebotener Sprache. Die Ergebnisse werden bzgl. ihrer Bedeutung für die Erforschung der Qualitätswahrnehmung sowie ihrer praktischen Anwendbarkeit diskutiert. Ein kurzer Ausblick in Abschnitt 5 zeigt laufende und zukünftige Arbeiten auf diesem Gebiet auf.

2 Messmethode

Bislang existieren kaum Verfahren zur gezielten physiologischen Untersuchung der an der Qualitätsbildung beteiligten neuronalen Prozesse. Eine Ausnahme bildet die Erfassung von frühen akustisch evozierten Potentialen im Hirnstamm, welche bspw. mit der auditorischen Hirnstamm-Reaktionsaudiometrie (Brainstem Evoked Response Audiometry, BERA) gemessen werden kann [4]; jedoch ist der Zusammenhang zwischen diesen Potenzialen und der Qualität unklar. In einer Reihe von Arbeiten ([5]-[9]) haben wir versucht, einige der oben beschriebenen Prozesse auf der Ebene des Muster- und Objektvergleiches durch Messung physiologischer Kennwerte erfassbar und später auch beschreibbar zu machen. Aufgrund der guten zeitlichen Auflösung und der vergleichsweise einfachen Durchführung bedienen wir uns dabei zunächst der Elektroenzephalographie (EEG).

Die Elektroenzephalographie umfasst die Messung von schwachen Potenzialunterschieden in den obersten Hirnschichten mittels auf der Kopfhaut angebrachter Elektroden. Diese Potentialunterschiede können auf zwei verschiedene Arten und Weisen analysiert werden [10]:

1. Zum einen die Analyse ereigniskorrelierter Potenziale (EKPs): Diese beschreiben Änderung von Hirnstrom-Potenzialen nach einem expliziten (Reiz-)Ereignis. Ein solches Ereignis besteht z.B. in der neuronalen Verarbeitung eines gestörten oder ungestörten Stimulus. Die Reaktion des Gehirns kann in einem Zeitfenster von typischerweise 0-1000 ms nach Einsatz des Triggers beobachtet werden. Man spricht dann bspw. von einer P300-Reaktion, wenn eine positive Potenzialänderung ca. 300 ms nach dem Trigger auftritt, oder von einer N400 bei entsprechender negativer Reaktion nach 400ms. Eine P300-Reaktion wird z.B. erwartet bei einem gestörten Stimulus in einer Reihe ansonsten ungestörter Stimuli (sog. „Oddball“ für „Ausreißer“).
2. Zum anderen die Analyse niedriger Frequenzen im Spontan-EEG, welche auf Aufmerksamkeit bzw. Wachheit hindeuten. Man teilt die interessierenden Frequenzbereiche verschiedene Bänder ein, bspw. das Delta-Band (1 - 4 Hz), das Theta-Band (4 - 8 Hz), oder das Alpha-Band (8 - 13 Hz).

In den bisherigen Untersuchungen haben wir uns zumeist auf die Messung von EKPs bei der gleichzeitigen Präsentation auditiver, visueller oder audiovisueller Sprachstimuli konzentriert.

Die Versuchsaufgabe besteht dabei aus einem Oddball-Paradigma: Die Versuchspersonen werden gebeten, aus einer Reihe von zwei oder mehr kurzen Stimuli herauszuhören, ob ein bestimmter Stimulus gestört ist oder nicht. Als Stimuli verwenden wir zumeist kurze Silben, wie bspw. /a/ oder /pa/ bis hin zu ganzen Worten, welche sich durch einen klar definierten Anfang auszeichnen und – im Falle visueller Darbietung – eine schnelle Änderung des Bildschirminhalts in der Bildmitte (z.B. durch das Öffnen des Mundes) aufweisen. Diese Stimuli werden durch verschiedene auditive oder visuelle Beeinträchtigungen gestört (signalkorreliertes Rauschen mittels einer Modulated Noise Reference Unit, Kodierverzerrungen mittels verschiedener Varianten des AMR-Codecs, *Blockiness* erzeugt durch quadratweise Mittelung von Pixelwerten), welche in ihrer Stärke individuell für jede Versuchsperson variiert werden können. Das EEG wird mit bis zu 64 Elektroden auf der Kopfhaut gemessen. Als Vergleichsgröße werden Beurteilungen der Versuchspersonen erhoben, entweder in Form von binären Entscheidungen (z.B. durch Drücken eines Knopfes für die Entscheidung gestört/ungestört) oder in Form von Bewertungen auf einer Skala nach ITU-T Rec. P.800 [2].

3 Messung der Wahrnehmung auditiv dargebotener gestörter Sprache

In ersten Studien [5][6][8] konnten wir typische EKP-Aktivitätsmuster bei oberschwellig gestörter Sprache beobachten, die bei einigen Probanden auch durch unterschwellige Störungen ausgelöst wurden. D.h. dass sich bei Probanden ähnliche Reaktionsmuster im Gehirn zeigten, obwohl die Stärke der präsentierten Störung (in diesem Fall der Laut /a/ gestört mit signalkorreliertem Rauschen) noch nicht ausreichend war, um bei den Versuchspersonen eine direkte Reaktion in Form eines bewussten Urteils hervorzurufen, dass der entsprechende Stimulus gestört war. Es lassen sich also mittels EEG vermutlich Aspekte der nicht bewussten Wahrnehmung von Störungen erfassen. Eine Analyse der EKPs zeigte, dass vor allem die Amplitude der P300-Reaktion bei gestörten Stimuli größer ist, und dass die Reaktion bei starken Störungen früher eintritt. Die verringerte Latenz kann als geringerer „neuronaler Effort“ bei der Störungsentdeckung verstanden werden.

3.1 Studie „Nicht-bewusste Störungsverarbeitung“

3.1.1 Teiluntersuchung 1: Laute

Methoden: Zehn Versuchspersonen nahmen an der ersten Teiluntersuchung teil. Als Reizmaterial wurde der Laut /a/ (200ms), gesprochen von einem männlichen Sprecher, verwendet. Eine ungestörte Referenz und vier gestörte Varianten wurden den Versuchspersonen vorgespielt. Als kontinuierliche Störung wurde signalkorreliertes Rauschen verwendet. Während der vor dem eigentlichen Test durchgeführten Kalibrierungsphase wurden für jede Versuchsperson vier Störungslevel ausgesucht; diese sollten folgenden Detektionsraten entsprechen: 100%, 75%, 50% und 0%. Im Median über alle Versuchspersonen wurden folgende SNR-Level verwendet: 28 dB, 24 dB, 21dB und 5 dB. Die ungestörte Referenz wurde in 70% und die vier gestörten Versionen sowie ein weiterer Kontrollreiz jeweils in 6% der Trials präsentiert. Aufgabe der Versuchspersonen war es, mittels Tastendruck zu kennzeichnen, ob der im aktuellen Trial präsentierte Reiz gestört war oder nicht. Je Versuchsperson wurden zwischen 2400 und 3600 Trials präsentiert; die Reihenfolge der Reize wurde randomisiert. Als subjektive Parameter dienten die Reaktionszeit sowie die Detektionsrate. Die EKP-Parameter P300-Latenz und P300-Amplitude wurden für jede Versuchsperson anhand der gemittelten ERP für jedes Reizlevel bestimmt. Eine Klassifizierung von Aktivierungsmuster erfolgte unter Verwendung einer Variante der Linear Discriminant Analysis (Shrinakge LDA).

Ergebnis: Für die subjektive Bewertung der Stimuli konnte ab dem Level von 21 dB eine signifikant schlechtere Bewertung der Stimuli gemessen werden ($p < 0.05$). Für die Amplitude und die Latenz der P300 fanden sich folgende zwei bedeutende Effekte: Erstens variierte die Latenz der P300-Amplitude mit der Störungsintensität, also je stärker die Störung, desto früher trat die maximale P300-Amplitude auf ($p < 0.05$). Zweitens kovariierte auch die maximale P300-Amplitude mit der Störungsintensität, denn je stärker die Störung, desto größer die Amplitude ($p < 0.05$). Unter Verwendung von Klassifikation mit Shrinkage LDA konnten wir Trials identifizieren, bei denen das Aktivierungsmuster einer Versuchsperson im Fall einer nicht erkannten unterschwelligen Störung dem einer bewusst identifizierten Störung ähnlich war. Um die Güte der Klassifizierung zu bestimmen wurde die Area Under the Curve (AUC) der Receiver Operating Characteristic (ROC) bestimmt (AUCb = Balanced Accuracy). Werte $AUCb > 0.9$ entsprechen einer exzellenten Klassifizierung und Werte $AUCb = 0.5$ entsprechen einer zufälligen Zuweisung von Klassenzugehörigkeiten. In diesem Teilversuch lag der gemittelte AUCb Wert bei $0.8 > AUCb > 0.56$ (Mittelung über jene Probanden, für die ausreichend Daten vorlagen).

Diskussion: Mit den Verhaltensdaten konnte die subjektive Detektionsschwelle bestimmt werden: Ab einem Störabstand von 21 dB nehmen die Versuchspersonen eine Störung wahr. Die Latenz der maximalen P300-Amplitude deuten wir als neuronalen Aufwand, der von den Versuchspersonen erbracht werden musste, um eine Störung zu identifizieren. Die Höhe der P300-Amplitude gibt Auskunft darüber, wie viel Aufmerksamkeit von der Versuchsperson durch die Reizpräsentation gebunden wurde. Die Klassifikation konnte zeigen, dass Versuchspersonen während einiger Trials die Störung wahrscheinlich als Wahrnehmungsereignis neuronal verarbeiteten, die entsprechende Stimulation aber eine interne Schwelle nicht überschritt und daher nicht in eine bewusste Störungsdetektion resultierten.

3.1.2 Teiluntersuchung 2: Wörter

Methoden: An der zweiten Teiluntersuchung nahmen neun Versuchspersonen teil. Als Reize dienten die Wörter /Haus/ und /Schild/, jeweils gesprochen von einem männlichen Sprecher und einer weiblichen Sprecherin. Diese wurden den Versuchspersonen ungestört (Referenz) und in vier gestörten Varianten vorgespielt. Als Störung wurde eine Bitratenreduzierung

verwendet (AMR-WB-Codec nach ITU-T Rec. G.722.2). Wie in Teiluntersuchung 1 wurden in einer Kalibrierungsphase für jede Versuchsperson vier Störungslevel ausgesucht; diese sollten folgende Detektionsraten entsprechen: 100%, 60%, 40% und 0%. Im Median über alle Versuchspersonen wurden folgende Bitraten-Level verwendet: 14.25, 12.65, 12.65 und 6.6 kbit/s. Das Versuchsdesign sowie die Analyse der Daten erfolgte in ähnlicher Weise wie für die erste Teiluntersuchung.

Ergebnis: Ab dem Level von 8.85 kbit/s wurde die Qualität der Reize als signifikant schlechter bewertet ($p < 0.05$). Im Kontrast zur ersten Teiluntersuchung konnte nur für die Amplitude der P300 ein bedeutender Effekt gefunden werden: Die Intensität der maximalen P300-Amplitude variierte mit der Störungsintensität. Je stärker also die Störung, desto größer die Amplitude ($p < 0.01$). Auch bei Reizen mit der Länge von Wörtern konnten mit Hilfe von Klassifikation (Shrinkage LDA) Trials identifiziert werden, bei denen das Aktivierungsmuster einer Versuchsperson im Fall einer nicht erkannten unterschwelligen Störung dem einer bewusst identifizierten Störung ähnlich war ($0.65 > AUCb > 0.53$; Mittelung über jene Probanden, für die ausreichend Daten vorlagen).

Diskussion: Die subjektive Detektionsschwelle in der zweiten Teiluntersuchung lag bei 8.85 kbit/s. Über die Größe der P300-Amplitude konnte auch für Reize in der Länge von Wörtern ermittelt werden, wie viel Aufmerksamkeit durch die Präsentation eines gestörten Reizes gebunden wurde. Die Klassifikation konnte erneut zeigen, dass Versuchspersonen teilweise eine Störung nicht bewusst wahrnahmen, diese jedoch neuronal verarbeitet wurde.

3.2 Studie „Langzeiteffekte von gestörten auditiven Reizen“

Die so identifizierten nicht bewussten Prozesse beziehen sich auf die kurzfristige Wahrnehmung von Störereignissen (Ereigniskorrelierte Potentiale). Bezüglich der längerfristigen Auswirkungen der Darbietung von gestörtem Audiomaterial ließ sich in [8] feststellen, dass diese in einem Zeitraum von 20 Minuten zu einer stärkeren Zunahme an Alpha-Wellen führten als das Anhören ungestörter Aufnahmen, d.h. dass die Probanden dadurch stärker ermüdeten.

Methoden: An der Studie nahmen 18 Versuchspersonen teil. Als Reiz wurde eine 40 Minuten lange Aufzeichnung mit vorgelesenen (männlicher Sprecher) Informationen über Sehenswürdigkeiten in Berlin verwendet. Diese wurde den Versuchspersonen in zwei Blöcken à 20 Minuten vorgespielt. Ein Block hatte die beste ungestörte Qualität (Breitband-Übertragung 50-7000 Hz), der andere wurde durch einen AMR-WB-Codec gestört (ITU-T Rec. G.722.2, 6.6 kbit/s). Die Reihenfolge, ob zuerst ein ungestörter oder gestörter Block präsentiert wurde, wurde randomisiert. Die Aufgabe der Versuchspersonen war es, während der Präsentation nach 9, 17, 25 und 33 Minuten auf einer Skala die Qualität zu beurteilen (Skala von exzellent (9) bis schlecht (1) mit Überlaufbereichen 0…10). Die Versuchspersonenurteile wurden zu sog. Mean Opinion Scores (MOS) für jeden Block in ungestörter und gestörter Variante gemittelt. Als EEG-Parameter wurde die Energie in den Frequenzbändern Theta (4-8 Hz) und Alpha (8-13 Hz) bestimmt.

Ergebnis: Die Analyse des MOS ergab eine signifikant bessere Bewertung für den ungestörten Stimulus ($p < 0.01$). Für die Energie im Thetaband konnte ein signifikanter Haupteffekt gefunden werden. Während der Präsentation des gestörten Stimulus war die Energie im Thetaband signifikant erhöht im Vergleich zum ungestörten Stimulus ($p < 0.05$). Zusätzlich war die Aktivität im Alphaband während der zweiten Blockhälfte im Vergleich zur ersten Hälfte des Blockes signifikant erhöht ($p < 0.05$).

Diskussion: Wie erwartet wurde die Qualität des gestörten Reizes als signifikant niedriger bewertet. Die erhöhte Aktivität im Thetaband ist ein Indikator für eine stärkere Ermüdung der Versuchsperson und damit womöglich für eine eingeschränkte Verarbeitung des Reizes, welche in unserer Studie auf die schlechtere Qualität des Reizes zurückzuführen ist. Eine

höhere Alphaaktivität gegen Ende der Blöcke ist auf die Ermüdung der Versuchspersonen aufgrund der mit der Bearbeitung der Aufgabe verbrachten Zeit zurück zu führen (Time-on-Task-Effekt).

4 Messung der Wahrnehmung audiovisuell dargebotener gestörter Sprache

Methoden: In weiteren Studien [7][11] wurden die Untersuchungen auf visuelle bzw. audiovisuelle Stimuli ausgeweitet. Dabei wurden Videoaufzeichnungen des Mundes beim Aussprechen der Silbe /pa/ mittels Blockbildung (Blockiness) im Videosignal (4 Qualitätsstufen) und signalkorreliertem Rauschen im Audiosignal (2 Qualitätsstufen) gestört. Der Stimulus wurde hierbei in einem Paarvergleich präsentiert. Zunächst wurde der ungestörte audiovisuelle Stimulus vorgespielt und darauf folgend der vermeintlich qualitativ schlechtere Stimulus. Die Aufgabe der Probanden war es dabei, nach Ende der Paarpräsentation Auskunft darüber zu geben, ob sie eine Störung im zweiten Teil wahrgenommen hatten oder nicht. Durch die Selbstauskunft wurde zunächst das Verfahren physiologischer Messungen im Zusammenhang mit der Erfassung von Qualitätsurteilen für audiovisuelle Reize überprüft. Anschließend an die EEG-Aufzeichnung wurde noch ein Standard-Videoqualitätstest mit Erfassung des MOS durchgeführt. An diesem audiovisuellen Versuch nahmen 13 Versuchspersonen teil.

Ergebnis: Die Ergebnisse des Versuches zeigen wie zuvor für auditive Reize, dass bei der bewussten Wahrnehmung von Störungen eine P300-Komponente mit höherer Amplitude auftritt als beim Referenzstimulus. Des Weiteren konnte gezeigt werden, dass je größer die hinzugefügte Videostörung ist, desto größer die gemessene P300-Amplitude, und desto früher erreicht diese auch ihr Maximum. Der Vergleich der EKPs zwischen den beiden Audio-Qualitätsstufen für jede Video-Qualitätsstufe zeigt hingegen keinen signifikanten Unterschied, was zum einen auf das asynchrone Starten von Audio- und Videosignal zurückzuführen sein kann, und zum anderen darauf, dass die Videostörung anscheinend schon so stark war, dass die Audioqualität keine entscheidende Rolle mehr für die Gesamtwahrnehmung des Stimulus spielte. Eine ANOVA für die P300-Amplitude ergab einen statistischen Haupteffekt für den Faktor der Video-Qualitätsstufe ($F(3,18) = 54.63$, $p \leq 0.01$, $\eta^2 = 0.90$). Dies lässt auf eine verstärkte kognitive Verarbeitung qualitativ schlechterer Stimuli schließen. Auch bei den ausgelassenen Störungsreizen konnte teilweise ein ähnliches Muster der Ereigniskorrelierten Potentiale gezeigt werden, was auf eine nicht bewusste Verarbeitung der Störung schließen lässt, die sich aber nicht bis auf die Verhaltensebene durchgesetzt hat.

Zwischen den erfassten Qualitätsurteilen (MOS) und der pro Versuchsperson über die einzelnen Qualitätsstufen gemittelten P300-Amplitude zeigte sich eine starke Korrelation beider Werte ($r = -0.87$, $p \leq 0.01$). Es konnte so ein sinnvoller Zusammenhang zwischen physiologisch gemessenen Parametern und in Standardqualitätstests ermittelten MOS aufgezeigt werden.

5 Ausblick

In diesem Beitrag wurden Messverfahren zur Analyse der Wahrnehmung gestörter Sprache diskutiert, welche auf Grundlage physiologischer Größen Aufschluss über die an der Qualitätsbildung beteiligten Prozesse liefern können. Erste Ergebnisse, welche in den Abschnitten 3 und 4 beschrieben wurden, belegen, dass sich mit Hilfe Ereigniskorrelierter Potenziale (EKPs) und einer Analyse der EEG-Signalenergie in Frequenzbändern nicht nur Prozesse der bewussten Wahrnehmung nachvollziehen lassen, sondern auch nicht bewusste Wahrnehmungen messbar sind. Dadurch werden Störungen, die unterhalb der Reaktionsschwelle eines expliziten, aktiven Messprozesses mit Introspektion und Befragung der Probanden liegen, erfassbar gemacht. Dies konnte zunächst nur für auditiv dargebotene

Sprache gezeigt werden, jedoch deuten die im auditiven wie auch im audiovisuellen Fall ähnlichen Aktivitätsmuster und daraus abgeleiteten Kenngrößen (hier insbesondere die P300) darauf hin, dass dies für unterschiedliche Sinnesmodalitäten möglich sein könnte.

Die Sensitivität der vorgestellten Methode ist bislang nicht besser als die eines Standard-Tests mit Introspektion und expliziter Beurteilung von Probanden. Bei entsprechender Anwendung von Methoden des maschinellen Lernens ist in Zukunft aber durchaus vorstellbar, so auch geringere, nicht bewusste Störungen detektierbar werden zu lassen, wie in den Studien [5][6][8] gezeigt werden konnte. Auch kann durch die physiologische Messung die längerfristige Auswirkung der Darbietung gestörter Sprache erfasst werden, ohne dass Probanden kontinuierlich Bewertungen abgeben müssen (vgl. die Methode nach [3]). Dem gegenüber steht bislang allerdings ein nicht unerheblicher Aufwand, sowohl finanziell als auch bzgl. der Versuchsdurchführung. So muss zunächst der Kontakt zwischen der Kopfhaut und einer Reihe von Elektroden mittels leitfähigem Gel hergestellt werden, für jeden Probanden müssen in einem Vorversuch individuelle Schwellwerte für verschiedene Störungsklassen bestimmt werden, und durch die notwendige häufige Messwiederholung summiert sich die Versuchszeit pro Proband auf derzeit 2-4 Stunden. Wir sind jedoch zuversichtlich, dass sich die Dauer des Versuches durch die Verwendung weniger, möglichst trocken mit der Kopfhaut in Kontakt gebrachter Elektroden verbunden mit einer intelligenten Auswertung der EEG-Signale bei gleicher Sensitivität noch deutlich reduzieren lässt.

Zudem erachten wir die Analyse multimodaler Stimuli als sehr aussichtsreich. Wir hoffen, durch die getrennte Manipulation der Audio- und der Videoqualität audiovisuell dargebotener Stimuli und Messung der nicht bewussten und bewussten Wahrnehmungsgrößen im EEG Aufschluss auf die Integration beider Modalitäten beim Qualitätsereignis zu bekommen. Diese Integration ist bislang zwar vielfach erforscht worden, jedoch meist mit unklarem Ergebnis. Sicher ist, dass sowohl der Inhalt des dargebotenen Medienmaterials als auch die Beurteilungssituation einen Einfluss darauf haben, welche Modalität (auditiv vs. visuell) die Qualität dominiert.

Darüber hinaus erproben wir andere physiologische Messgrößen, bspw. die Nahinfrarotspektroskopie (NIRS). Hierdurch erhoffen wir uns Aufschluss über Prozesse, welche mit dem EKP-Paradigma nicht oder nur schlecht erfasst werden können. Bei der Auswahl der Messgrößen ist neben den inhärenten Charakteristika der Messung (bspw. räumliche und zeitliche Auflösung des Verfahrens) auch auf den damit verbundenen apparativen Aufwand und dadurch bedingte Störeinflüsse zu achten. So erfordern bspw. fMRI-Messungen bislang sehr geräuschbehaftete Apparaturen, die die Erfassung subtiler auditiver Störungen derzeit unrealistisch erscheinen lassen.

Literatur

[1] U. Jekosch: Voice and Speech Quality Perception — Assessment and Evaluation, Springer Series in Signals and Communication Technology, Berlin, 2005.

[2] ITU-T Rec. P.800: Methods for Subjective Determination of Transmission Quality. International Telecommunication Union, Genf, 1996.

[3] ITU-T Rec. P:880: Continuous Evaluation of Time-varying Speech Quality. International Telecommunication Union, Genf, 2004.

[4] R.J. Roeser: Audiology: Diagnosis. Thieme, New York, 2007.

[5] J.N. Antons, B. Blankertz, G. Curio, S. Möller, A.K. Porbadnigk, und R. Schleicher: Subjective listening tests and neural correlates of speech degradation in case of signal-correlated noise, in: Audio Engineering Society (AES) 129th Convention, 2010.

[6] A.K. Porbadnigk, J.N. Antons, B. Blankertz, M.S. Treder, R. Schleicher, S. Möller, G. Curio: Using ERPs for Assessing the(Sub)Conscious Perception of Noise, in: Proc. of

the 32nd Annual Int Conf. of the IEEE Engineering in Medicine and Biology Society, pp.2690-2693 , 2010.

[7] S. Arndt, J.N. Antons, R. Schleicher, S. Möller, S. Scholler, und G. Curio: A physiological approach to determine video quality, in: Proc. 2011 IEEE International Symposium on Multimedia, S. 518–523, 2011.

[8] J.N. Antons, R. Schleicher, S. Arndt, S. Möller, A.K. Porbadnigk und G. Curio: Analyzing speech quality perception using electro-encephalography, Journal of Selected Topics in Signal Processing, angenommen zur Veröffentlichung, 2012.

[9] J.N. Antons, R. Schleicher, S. Arndt, S. Möller, und G. Curio: Too tired for calling? A physiological measure of fatigue caused by bandwidth limitations, Angenommen für Fourth International Workshop on Quality of Multimedia Experience (QoMEX), IEEE, 2012.

[10] M. Coles und M. Rugg: Electrophysiology of Mind: Event-Related Brain Potentials and Cognition, Kapitel "Event-related brain potentials: an introduction", Oxford University Press, 1995.

[11] S. Arndt, J.N. Antons, R. Schleicher, S. Möller, und G. Curio: Perception of low-quality videos analyzed by means of electroencephalography. Angenommen für Fourth International Workshop on Quality of Multimedia Experience (QoMEX), IEEE, 2012