

FIOhrin als Konzept für die Kommunikation in Zeiten sozialer Distanz und darüber hinaus

Dr.-Ing. Hauke Krüger, 31.5.2020

Die Beschulung der Kinder zuhause (das sog. *Home Schooling*) ist in Zeiten der Corona Pandemie zu einer der größten Herausforderungen unserer Zeit geworden: Nachdem die Schulkinder über viele Wochen weitestgehend sich selbst und der Obhut der Eltern zuhause überlassen wurden, sind sie aktuell in kleinen Gruppen entweder im Präsenzunterricht in der Schule oder zuhause. Aufgrund der so entstandenen Engpässe an Personal und Klassenräumen müssen Lehrer die Inhalte mehrfach vermitteln. In der Konsequenz besuchen die Kinder nur für einzelne Tage in der Woche Lehrveranstaltungen. Bereits jetzt ist gut ersichtlich, dass dieser Zustand auf lange Sicht katastrophale Auswirkungen auf den Wissensstand der Kinder haben und sich die Schere zwischen Schüler*innen, die dem Lernstoff in der Schule gut folgen können und solchen, die Schwierigkeiten haben, immer weiter öffnen wird.

Sofern dieser Zustand nur kurzfristig vorherrscht, sind die Lücken unter Umständen wieder schließbar. Im Moment deutet jedoch alles daraufhin, dass die Einschränkungen des Lehrbetriebs auch nach den Sommerferien weiter Bestand haben werden. Um zu verhindern, dass sich in der nächsten Zeit Rückstände bei der Ausbildung unserer Kinder ansammeln und sie wertvolle Zeit in ihrer schulischen Entwicklung verlieren, müssen innovative Konzepte der Lehre bei Einhaltung der Hygienevorgaben definiert und so schnell wie möglich umgesetzt werden.

Ein Baustein zur Wiederaufnahme einer nachhaltigen Lehrmethodik können hybride Konzepte sein, bei denen Kinder in Präsenz (in der Schule) und Kinder daheim gemeinsam am Unterricht teilnehmen. Hierfür müssen die technischen Randbedingungen jedoch zunächst geschaffen werden. Ein wichtiger Baustein ist die Einbindung von Anwendungen zur Videotelefonie, um die Kinder zuhause und die in der Schule über eine Internetverbindung miteinander zu verbinden. Leider hat die Technologie insbesondere bzgl. der Verständlichkeit Nachteile, die diesen Ansatz erschweren. Um an dieser Stelle eine Verbesserung der Kommunikationsmöglichkeiten zu erreichen, schlagen wir die FIOhrin Technologie vor.

In diesem Whitepaper wird im Folgenden zunächst kurz der Ansatz eines hybriden Ausbildungskonzepts in Zeiten des Corona bedingten Home Schoolings erläutert. Anschließend wird die grundsätzliche Idee der FIOhrin Technologie beschrieben und worin der prinzipielle Vorteile gegenüber dem heute typischen Einsatz von System zur Internet basierten Kommunikation besteht. Abschließend wird dann erörtert, inwieweit die vorgestellte FIOhrin Technologie trotz der innovativen Herangehensweise durch geschickte Konfiguration in verfügbaren Anwendungen zur Internet-Telefonie **bereits heute** mit eingebunden werden kann.

Internet basierte Kommunikation als hybrides Konzept für das Home Schooling und darüber hinaus

Bis vor kurzem kamen Internet Kommunikationssysteme (kurz: IP-Telefonie) in erster Linie bei Firmenmeetings zum Einsatz. Und selbst dabei war häufig eine Tendenz zu erkennen, trotz eines immer stärker werdenden Bewusstseins der Notwendigkeit von Energieeinsparung zur Verlangsamung des globalen Klimawandels lieber auf die andere Seite der Erde zu fliegen als von zuhause aus über eine Internetverbindung mit Kunden oder Auftraggebern zu kommunizieren.

Und dann kam die Corona-Krise: Die Menschen haben aus der Not der Pandemie heraus gelernt, sich auf die Eigenschaften von Internet Kommunikationslösungen einzustellen. Anfangs erschien es ermüdend und anstrengend, und eine hohe Gesprächsdisziplin ist erforderlich, um mit den technischen Randbedingungen dieser Technologie umzugehen. Aber die Menschen haben festgestellt, dass es durchaus ein zielführender Ansatz ist.

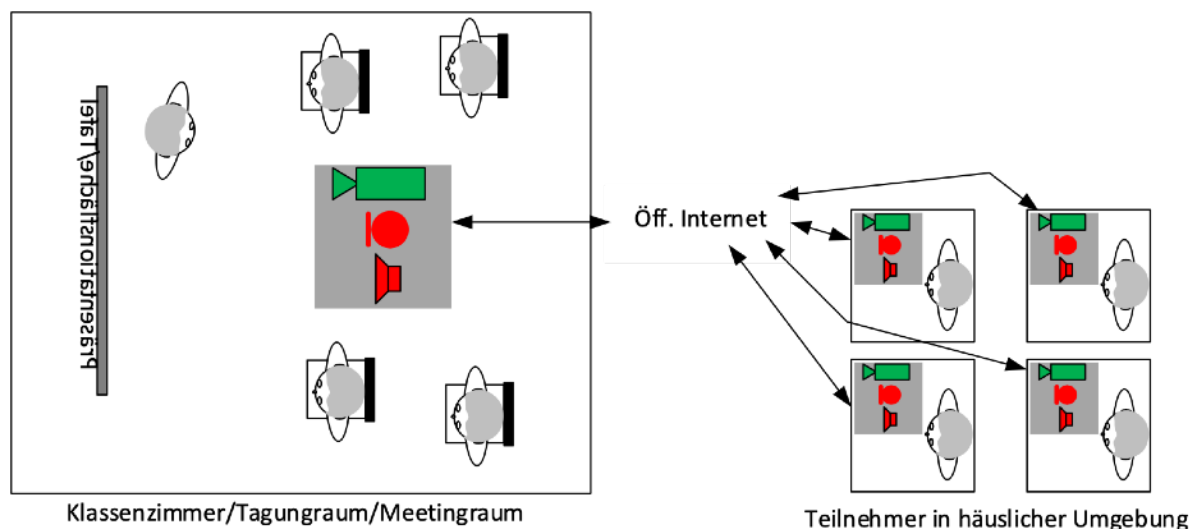
Nun sind die ersten Lockerungen in der Corona Krise eingetreten und die Menschen passen sich abermals an:

- Mitarbeiter großer Firmen kehren an ihre übliche Wirkungsstätte zurück,
- Schüler gehen wieder in die Schule und
- Fachsitzungen (z.B. Therapiegruppen-Meetings, Workshops, Tagungen) finden wieder statt, jedoch mit einer geringen Anzahl von Teilnehmern.

Was bleibt, sind jedoch die Hygieneregeln, die die Menschen auf unabsehbare Zeit weiterhin werden einhalten müssen. Um diese einzuhalten, werden zahlreiche Kompromisse eingegangen:

- Mitarbeiter besuchen ihre Büros im Schichtbetrieb: die eine Hälfte der Mitarbeiter ist zuhause, die andere im Büro.
- Schüler gehen zurzeit in kleinen Gruppen in die Schule. Während die eine Gruppe am Präsenzunterricht teilnimmt, müssen die anderen Schüler zuhause weiter Home Schooling betreiben. Die Lehrer haben ein erhöhtes Arbeitsaufkommen, da sie den Lehrstoff somit mehrfach präsentieren müssen. In der Konsequenz besuchen die Kinder nur an einzelnen Tagen ihre Schule.
- An Therapiegruppensitzungen oder ähnlichen Veranstaltungen kann aufgrund der Hygienemaßnahmen (Abstandsregeln) nur ein Bruchteil der Menschen teilnehmen, die gerne teilnehmen möchten.

Um an dieser Stelle eine Verbesserung zu erreichen, wird zunehmend der Ruf nach hybriden Konzepten lauter, bei dem Menschen vor Ort und Menschen, die sich im häuslichen Umfeld befinden, **gleichzeitig** an Meetings, dem Schulunterricht oder Therapiegruppensitzungen teilnehmen. Die Methodik dahinter besteht darin, mittels Internet basierter Kommunikation eine gemeinsame virtuelle Umgebung zu schaffen, in der sich sowohl die Teilnehmer vor Ort als auch die Personen befinden, die von zuhause aus mittels Internetverbindung angebunden sind. Dieses Konzept ist in der folgenden Abbildung schematisch dargestellt:



Bei der Umsetzung wird das Geschehen im Klassenraum / dem Meeting-Raum oder in der Tagungsortlichkeit mit einer Kamera (Video) und einem Mikrofon (Audio) aufgezeichnet und an die Teilnehmer zuhause übertragen. Gleichzeitig werden auch von jedem Teilnehmer zuhause das Video und das Audiosignal aufgezeichnet und wiederum in das Klassenzimmer / den Meeting-Raum oder an

die Tagungsortlichkeit übertragen. Dort werden die Signale dann wiederum über einen Lautsprecher wiedergegeben.

Gegenwärtig werden für diese Art der Kommunikation gerne Internet Kommunikationssysteme kommerzieller und nicht-kommerzieller Anbieter wie z.B. *Microsoft Teams*, *Zoom*, *Jitsy* und viele mehr eingesetzt. Das Haupteinsatzgebiet dieser Anwendungen ist jedoch in erster Linie eine *Face-to-face* Kommunikation, bei der an jedem Endgerät eine Person vor der Kamera und in der Nähe des Mikrofons positioniert ist oder sogar ein Headset trägt.

Für eine Kommunikation im Sinne eines hybriden Konzepts wie zuvor beschrieben stoßen die verfügbaren Systems bislang jedoch in der Praxis nicht auf eine überwältigende Akzeptanz: Insbesondere bei der **Akustik der Gespräche**, also der wahrgenommen Audioqualität, gibt es einige kritische Eigenschaften, die einen weiträumigen Einsatz des hybriden Konzepts gegenwärtig einschränken:

- Mit dem Abstand des Sprechers zum Mikrophon wird der Signalpegel der Sprache, die an die angebundenen Zuhörer übertragen wird, immer geringer. Gleichzeitig wird die Sprache durch den Raumnachhall, also die Reflektion der akustischen Signale an den Wänden und Decken der Räume immer stärker überlagert, so dass es immer schwieriger wird, den Sprecher zu verstehen.
- Aktuelle Systeme zur Internet Kommunikation setzen eine automatische Lautstärkekontrolle ein, um die schwächeren Pegel kompensierend wieder zu verstärken. Dies wiederum führt jedoch auch zu einer Verstärkung der Hintergrundgeräusche und des Nachhalls.
- Gegen Nachhall und Störungen setzen die vorhandenen Lösungen wiederum Algorithmen zur Störreduktion (*Noise Reduction*), zur Nachhallbeseitigung (*De-reverberation*,) und zur Signalextraktion (*Beamforming*) ein. Diese Algorithmen wiederum greifen jedoch leider in vielen Fällen auch die Sprachanteile an, die zu einem Verständnis des Sprechers wichtig sind, so dass in der Summe die Sprachverständlichkeit eher leidet.
- Mit Anwesenheit der Zuhörer vor Ort steigt auch die Wahrscheinlichkeit für Störung durch die Menschen die präsent sind. Gerade in einer Schule ist ein hoher Umgebungsstörpegel durch die anwesenden Schüler gegeben. Auch hier greift die Störreduktion ein, die Verständlichkeit jedoch leidet.
- Zeitgleich sprechende Teilnehmer in der Präsenzsituation führen bei der Wiedergabe dazu, dass keiner der Sprecher mehr verstanden werden kann, da sich die Signale unkontrolliert überlagern. Das wird schlimmer je mehr Personen gleichzeitig sprechen. An dieser Stelle kann nur dadurch Abhilfe geschaffen werden, eine hohe Gesprächsdisziplin einzuhalten, so dass immer nur eine Person gleichzeitig spricht.
- Noch schwieriger wird es, wenn Menschen, die über das Internet angebunden sind, ebenfalls sprechen möchten und das Signal im Präsenzraum wiedergegeben wird. Auch eine hohe Gesprächsdisziplin ist hier unter Umständen nicht mehr ausreichend: Während die Menschen in der Präsenzsituation bei einer Wortmeldung noch kurzfristig einhalten können, wenn sie feststellen, dass gerade auch jemand anderes zum Sprechen ansetzt, wird dies bei einer Verbindung über das Internet erst stark verspätet festgestellt, da bei der Übertragung in der Regel eine hohe Übertragungsverzögerung gegeben ist.

Typischerweise werden solche Sitzungen als schwierig bzgl. der Verständlichkeit und ermüdend wahrgenommen. Insbesondere ist eine hohe Kommunikationsdisziplin die Voraussetzung für ein erfolgreiches Gespräch, und man stellt immer wieder in Sitzungen fest, dass alle Mikrofone bis auf das des Hauptsprechers abgeschaltet sind, um diese Disziplin einzuhalten. Soziales Feedback und

ähnliche Maßnahmen, die man aus der direkten Mensch-zu-Mensch Kommunikation kennt, fehlen und lassen solche Gespräche als unnatürlich und anstrengend erscheinen.

Verbesserung der wahrgenommenen Audio-Qualität durch Binauraltechnik

Im Umfeld der Binaurics Audio GmbH und am Institut für Kommunikationssysteme der RWTH Aachen wurden die Schwierigkeiten bei der heute weit verbreiteten Kommunikationstechnik bereits vor vielen Jahren erkannt und aus diesem Grunde an der Umsetzung des nächsten Schritts der Evolution von Kommunikationssystemen geforscht und folgende Fragestellung betrachtet: Wieso kann mit den herkömmlichen Ansätzen zur Telekommunikation niemals das gleiche Gefühl wie bei einer Präsenz vor Ort erreicht werden? Und was unterscheidet ein zwischenmenschliches Gespräch von einer Konferenzsituation über eine Internetverbindung?

Die Antwort ist im Grunde ganz einfach: während der Mensch **über zwei Ohren** verfügt, die in zwischenmenschlichen Gesprächen auch eingesetzt werden, wird in Internet basierten Telekonferenzen in der Regel nur ein **einkanaliges Signal** übertragen. Gerade für die Wahrnehmung der Räumlichkeit akustischer Szenarien und damit die Verortung von Audioquellen im Raum ist es in der Tat zwingend notwendig, auf **beide** Ohren zurückzugreifen. Dies wiederum ermöglicht es dem Gehirn, Signale, die aus einer bestimmten Richtung kommen, herauszufiltern und Störgeräusche aus anderen Richtungen zu unterdrücken. Diese Eigenschaft des Gehirns nennen wir „Brain Processing“. Dieses Brain Processing hat unterbewusst jeder schon einmal eingesetzt, der auf Veranstaltungen mit vielen Teilnehmern ein Gespräch mit einer Einzelperson geführt hat: die zum Teil sehr lauten Störungen aus der Umgebung werden „überhört“, während die Aussagen des Gesprächspartners gut wahrgenommen werden können. Die wahrgenommene Räumlichkeit in diesen Situationen ist dabei rein technisch gesehen wiederum auf die feinen Unterschiede, Abweichungen und Relationen der beiden Signale, die jeweils die Ohren des Zuhörers erreichen, zurückzuführen.

Im Gegensatz dazu wird bei der Wiedergabe eines einkanalig aufgezeichneten Signals jegliche Räumlichkeit bereits bei der Aufnahme verworfen. Alle Audioquellen kommen für den Zuhörer, dem das Signal zugeführt wird, folglich aus der gleichen zentralen Richtung. Die Fähigkeiten des Gehirns zum Brain Processing, oder genauer gesagt zur Trennung von Einzelsignalen und Unterdrückung von Störanteilen, kann somit aufgrund der Einschränkungen des Kommunikationssystems nicht zum Einsatz kommen. Dies gilt im Übrigen auch dann, wenn das Signal mit einem Stereokopfhörer beiden Ohren des Zuhörers zugeführt wird, da trotz Wahrnehmung mit beiden Ohren die Eigenschaften der Räumlichkeit aus dem ursprünglichen Signal nicht einfach wieder reproduziert werden können: zwischen dem Signal für das linke Ohr und dem für das rechte Ohr existieren keine feinen Unterschiede, die eine Verortung der ursprünglichen Audioquellen ermöglichen würden.

Um die Probleme zu beheben, die mit der fehlenden Räumlichkeit der Signale einhergehen, wurden in der Vergangenheit von vielen Firmen, die Internet basierte Kommunikationslösungen anbieten, die oben erwähnten Algorithmen zur Verbesserung der einkanaligen Signale entwickelt. Die Reichweite dieser Maßnahmen erscheint indes jedoch insgesamt limitiert. Als Konsequenz aus den beschriebenen Zusammenhängen erscheint es somit erfolgsversprechender, anstatt der Einführung und Verbesserung weiterer Signalverbesserungsmaßnahmen für einkanalige Signale lieber direkt die Signale **für beide Ohren** vom Ort des Geschehens an die Zuhörer zu übertragen. Letzterer kann dann zur Signalkonstruktion das zuvor erwähnte Brain Processing anwenden, da er wahrnehmungstechnisch das Gefühl bekommt, sich wirklich inmitten der ursprünglichen akustischen Szene zu befinden.

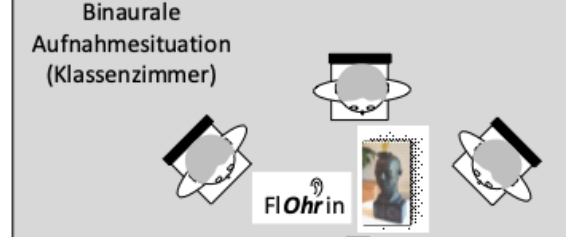
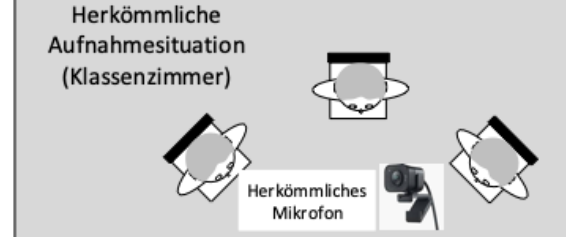
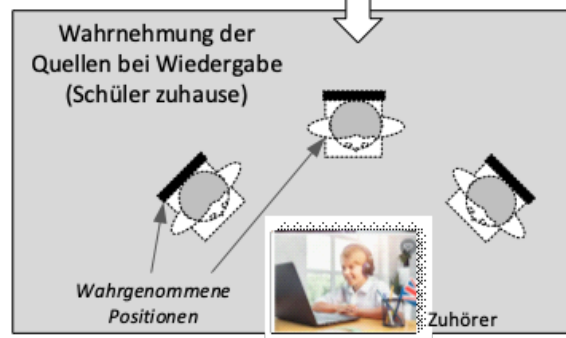
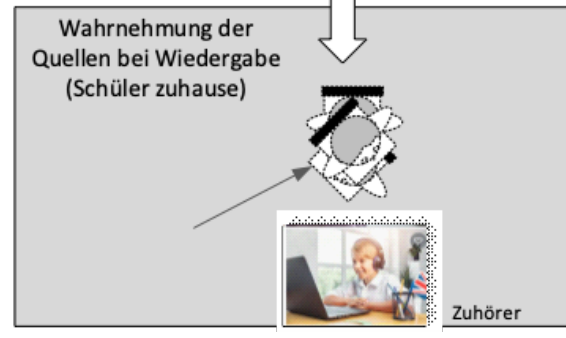
Einsatz der Binauraltechnik mit FIOhrin: Die Binaurale Kommunikation

Die Technik zur Aufnahme von Binauralsignalen ist prinzipiell seit vielen Jahren wohlbekannt: Gerade bei audiophilen Puristen und im Bereich der akustischen Messtechnik kommen sogenannte „Kunstköpfe“ seit vielen Jahren zum Einsatz, um Signale auch bzgl. der Räumlichkeiten so natürlich wie möglich aufnehmen und später wiedergeben zu können. Leider jedoch sind solche Lösungen in der Regel so teuer, dass ein Einsatz im privaten Umfeld oder dem Umfeld einer Schule nicht denkbar ist. Außerdem war bislang kein Internet basiertes Kommunikationssystem verfügbar, das eine Ende-zu-Ende Übertragung von zweikanaligen Signalen unterstützte.

Das Konzept von Binaurics Audio ist nun der erstmalige Aufbau eines Kunstkopfes namens FIOhrin, der auch für den Massenmarkt einsetzbar ist und somit die technologische Basis für eine neue Art der Kommunikation darstellt. Eine Kommunikation, bei der binaurale Technologie zum Einsatz kommt, nennen wir **Binaurale Kommunikation**. Grundlage der Lösung von Binaurics Audio sind dabei Standardkomponenten, die als Einzelprodukte des kommerziellen Massenmarktes verfügbar sind und deswegen recht preisgünstig erworben werden können. Diese Technologie kann aus diesem Grunde somit nun auch weiten Teilen der Bevölkerung zugänglich gemacht werden.

Eine zusammenfassende Gegenüberstellung der **Binauralen Kommunikation** und **herkömmlicher Kommunikation** ist in der folgenden Tabelle aufgelistet:

Binaurale Kommunikation	Herkömmliche Kommunikation
2 Mikrofone, platziert an der Position der Ohren von	1 Mikrofon, in der Regel z.B. als Bestandteil der Kamera
2 Audiokanäle (rechts & links)	1 Audiokanal
Wiedergabe über Stereokopfhörer	Wiedergabe über Lautsprecher
Binauralsignal erlaubt Rekonstruktion der Räumlichkeit bei Aufnahme	Keine Räumlichkeit

<p>Binaurale Aufnahmesituation (Klassenzimmer)</p> 	<p>Herkömmliche Aufnahmesituation (Klassenzimmer)</p> 
<p>Wahrnehmung der Quellen bei Wiedergabe (Schüler zuhause)</p>  <p>Räumliche Verortung der unterschiedlichen Sprecher bei Wiedergabe möglich!!</p>	<p>Wahrnehmung der Quellen bei Wiedergabe (Schüler zuhause)</p>  <p>Alle Sprecher erscheinen am gleichen Ort verortet.</p>
<ol style="list-style-type: none"> 1. Zuhörer kann einzelne Sprecher durch „Fokussierung“ herausfiltern, 2. Zuhörer kann Störgeräusche durch „Fokussierung“ unterdrücken, 3. verbesserte Verständlichkeit, 4. keine Ermüdung während der Unterhaltung. <p>„Unterhaltung als wäre man vor Ort“</p>	<ol style="list-style-type: none"> 1. Alle Sprecher erscheinen am gleichen Ort, eine Trennung ist nicht möglich, 2. Störende Wahrnehmung von Nachhall und Umgebungsgeräuschen, 3. schlechte Verständlichkeit, 4. stark ermüdende Kommunikation.

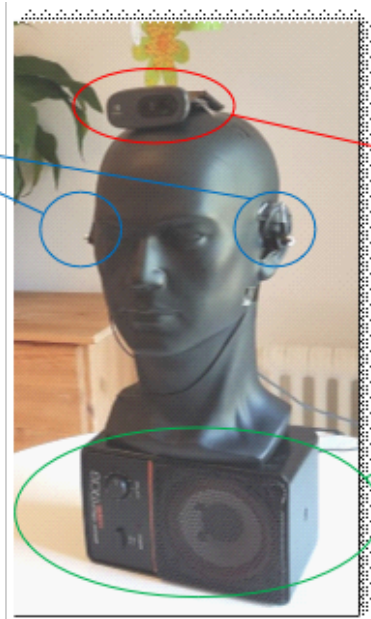
Vom Kunstkopf zum Endgerät für eine Hybride Internet basierten Binaurale Kommunikation

Obwohl die Kernfunktionalität des FIOhrin Konzepts die binaurale Aufnahmetechnik ist, ist das Gesamtkonzept jedoch viel mehr als nur ein Kunstkopf: Damit das System auch für heutige Video-Konferenzanwendungen zum Einsatz kommen kann, wurde der künstliche Kopf zur Aufnahme der Audiosignale um eine HD Kamera auf dem Kopf erweitert. Somit nehmen die angebundenen Konferenzteilnehmer auch den Blickwinkel von FIOhrin ein.

Um darüber hinaus die angebundenen Teilnehmer auch interaktiv einzubinden, verfügt FIOhrin zusätzlich über einen kleinen Lautsprecher, über den die Signale der angebundenen Zuhörer wiedergegeben werden können. Somit wird FIOhrin zu einem **audio-visuellen Avatar**, der stellvertretend für alle von zuhause aus an der Konferenz teilnehmenden Konferenzteilnehmer steht und z.B. in der Schule als *virtueller Mitschüler* eingebunden wird.

Die funktionalen Komponenten von FIOhrin sind in dem folgenden Bild noch einmal zusammengefasst dargestellt:

Binaurale Mikrofone –
externe Teilnehmer
hören mit den Ohren
von **FI Ohr** in



HD Kamera – externe
Teilnehmer haben den
gleichen Blickwinkel
wie **FI Ohr** in

Audioausgabe – FI Ohr in
verleiht externen
Zuhörern eine Stimme

Anbindung an heute verfügbare Videokonferenzsysteme

Damit FI Ohr in auch für eine bi-direktionale Kommunikation eingesetzt werden kann, müssen die aufgenommenen zweikanaligen Signale über eine Internetverbindung von einem zu den anderen Teilnehmern mithilfe einer Anwendung zur Internet Kommunikation übertragen werden, die somit über einen Ende-zu-Ende-Stereomodus verfügen muss. Es ist außerdem von Vorteil, wenn in dem eingesetzten Kommunikationssystem eine Option existiert, sämtliche Signalverarbeitungsalgorithmen, die – wie zuvor beschrieben – in erster Linie sinnvoll sind, wenn das Signal einkanalig übertragen werden muss, auszuschalten.

Beide Voraussetzungen erfüllt die aktuell häufig eingesetzte Kommunikationslösung von Zoom. Die Unterstützung für die zweikanalige Übertragung (Stereo Modus) scheint jedoch zum heutigen Zeitpunkt nicht besonders häufig um Einsatz zu kommen. Deshalb muss man momentan einige funktionale Einschränkungen in Kauf nehmen:

- 1) Binauralaufnahmen sowie die Deaktivierung der Algorithmen zur (einkanaligen) Signalverbesserung müssen in der Zoom Web-Umgebung zunächst freigeschaltet werden. Erst im Anschluss kann die Optionen auch im Zoom Client während des aktiven Gebrauchs entsprechend eingestellt werden.
- 2) Die Funktionalität zur Aufnahme zweikanaliger Signale scheint nur unter dem Betriebssystem Microsoft Windows zu funktionieren. Die Übertragung und Wiedergabe zweikanaliger Signale scheint jedoch auch unter Mac OS X zu funktionieren.
- 3) Die Übertragung und Wiedergabe zweikanaliger Signale scheint momentan nicht auf Smartphones zu funktionieren, sondern ausschließlich auf Laptops und Desktop PCs.
- 4) Für eine Wiedergabe der binauralen Signale ist **zwingend** ein Stereo-Kopfhörer zu verwenden.
- 5) **Achtung:** bei der Ankopplung von Bluetooth-Kopfhörern wird zwischen PC und Kopfhörer nur ein monaurales Signal übertragen (HFP, *Hands-free Profile*). **Wir raten aus diesem Grunde für diese Anwendung von einem Einsatz von Bluetooth Kopfhörern ab!**

Sofern eine weitere Verbreitung der Binauraltechnik stattfindet, ist es für die Firma Zoom vermutlich ein Leichtes, die Technik der zweikanaligen Wiedergabe von Audiosignalen auch auf Smartphones zu aktivieren.

Es ist durchaus möglich, dass auch andere Internet basierte Kommunikationslösungen die Übertragung von Stereosignalen unterstützen. In vielen Fällen jedoch ist die Methodik zur Aktivierung entweder nicht einfach zugänglich oder nicht ausreichend dokumentiert.

Einsatz der Technologie beim Webstreaming

Gerade im Umfeld einer Schule kann es sinnvoll sein, die angebundenen Zuhörer nicht über eine Audio-Verbindung einzubinden, sondern nur unidirektional als Web-Stream, der von den Zuhörern abonniert werden kann. Anfragen durch die Zuhörer können dann zum Beispiel mittels einer Chat Funktion an den Lehrer übermittelt werden, der dann die Fragen mit den Schülern klärt. Dies hätte den Vorteil, dass die Eingaben aus dem Publikum (hier: Schüler) deutlich besser in das didaktische Konzept des Lehrers eingearbeitet werden können.

Eine Lösung für das Web-Streaming ist zum Beispiel bei unserem Partner ferncast (www.ferncast.com) erhältlich.

Acknowledgement

Ein ganz herzlicher Dank geht an meine Freunde Jens Wiegert und Daniel Haupt für die Teilnahme an diversen Testsessions der Binauralen Kommunikation, zielführende Diskussionen, Anregungen und wertvolle Hinweise.