

Entwicklung und Validierung des Balance Assessment-Features von Starkey



Justin R. Burwinkel¹, Kristen K. Steenerson², Majd Srour¹, Christy Cloninger¹ und Chris Howes¹

Hintergründe

Der Zusammenhang zwischen Hörproblemen und erhöhtem Sturzrisiko ist in der Literatur gut dokumentiert. Eine der Studien ergab, dass Personen, die in audiologischer Behandlung sind, ein höheres Sturzrisiko haben als ihre Altersgenossen (Criter & Honaker, 2016). Lin & Ferrucci (2012) berichteten über einen signifikanten Zusammenhang zwischen dem Schweregrad der Hörbeeinträchtigung und den Berichten über Stürze – selbst unter Berücksichtigung demografischer, kardiovaskulärer und vestibulärer Aspekte.

Man geht davon aus, dass das individuelle Sturzrisiko multifaktoriell bedingt ist und zahlreiche verhaltensbedingte, physiologische und pathologische Mechanismen eine Rolle spielen. Studien zufolge besteht ein Zusammenhang zwischen Hörschäden und Stürzen, einschließlich komorbider vestibulärer, neuronaler oder kardiovaskulärer Erkrankungen. Auch genetische Einflüsse und geteilte Aufmerksamkeitsressourcen für die Fortbewegung und die Aufrechterhaltung des Gleichgewichts werden genannt, ebenso wie körperliche Gebrechlichkeit, welche durch soziale Isolation, Depression, körperliche Inaktivität und kognitive Beeinträchtigung noch verstärkt werden könnte (für eine Übersicht siehe Agmon et al.).

Obwohl bekannt ist, dass ein Zusammenhang zwischen fortgeschrittenem Alter, einer vorliegenden Hörminderung und dem Sturzrisiko besteht, werden in HNO-Kliniken oder anderen Gesundheitseinrichtungen, die von älteren Menschen häufiger aufgesucht werden, routinemäßige Sturzrisikobewertungen nur unzureichend durchgeführt (Howland et al., 2018; Patterson & Honaker, 2014). Derzeit werden Sturzrisikountersuchungen hauptsächlich von Hausärzten und Gerontologen im Rahmen der jährlichen Medicare Vorsorgeuntersuchungen (Annual Wellness Visits, AWW) durchgeführt.

Diese Besuche werden jedoch im Rahmen des Gesundheitssystems nicht ausreichend genutzt, da Berichten zufolge weniger als einer von fünf älteren Amerikanern diese kostenlosen Leistungen pro Jahr *in Anspruch nimmt*. (Centers for Medicare & Medicaid Services, 2017). Da Hörgeräteakustiker häufig ältere Patienten mit erhöhtem Sturzrisiko behandeln, sind sie gut positioniert, um Untersuchungen und Bewertungen des Sturzrisikos in ihre Routinepraxis zu integrieren. Diese zusätzlichen Schritte entsprechen einem umfassenderen Ansatz in der Patientenversorgung und bieten die Möglichkeit, die Behandlungsergebnisse deutlich zu verbessern, insbesondere da an anderer Stelle möglicherweise keine Sturzrisikobewertungen für den Patienten durchgeführt wurden.

Darüber hinaus ergab eine US-amerikanische Studie, für die Hausärzte befragt wurden, dass zwar 96 % der Hausärzte der Meinung sind, dass alle älteren Erwachsenen auf ihr Sturzrisiko hin untersucht werden sollten, aber nur 14 % mit der von den Centers for Disease Control and Prevention (CDC) herausgegebenen Broschüre „Stopping Elderly Accidents, Deaths & Injuries (STEADI)“ vertraut sind; die Broschüre liefert einen praktischen Rahmen für die wirksame Umsetzung der gemeinsam von der American Geriatrics Society (AGS) und der British Geriatrics Society (BGS) veröffentlichten Leitlinien zur Sturzprävention (Eckstrom et al., 2017) bietet. Angesichts dieser Herausforderung besteht ein dringender Bedarf an zugänglichen Instrumenten, die dazu beitragen können, Lücken im Sturzrisikomanagement zu schließen, insbesondere in Umgebungen, in denen regelmäßige, umfassende Bewertungen nicht konsequent durchgeführt werden.

¹Starkey

²Stanford University School of Medicine, Department of Otolaryngology-Head and Neck Surgery

Das Balance Assessment-Feature in der My Starkey Mobile App erfüllt diesen Bedarf, indem es eine benutzerfreundliche Möglichkeit bietet, das Sturzrisiko regelmäßig zu überprüfen und wichtige Aspekte des Gleichgewichts, des Gangs und der Kraft zu überwachen. Durch die Integration der STEAD-Initiative der CDC mit fortschrittlicher Sensortechnologie und künstlicher Intelligenz (KI) bietet diese Funktion Ärzten ein zuverlässiges Tool zur Unterstützung des proaktiven Sturzrisikomanagements zwischen den Praxisbesuchen. Regelmäßige Selbstbewertungen ermöglichen es dem Einzelnen, ein größeres Bewusstsein für seine Gesundheit zu entwickeln und veränderbare Risikofaktoren anzugehen, bevor sie eskalieren, was möglicherweise zu besseren Behandlungsergebnissen und weniger sturzbedingten Verletzungen führt.

Balance Assessment-Feature auf einen Blick

Das Balance Assessment-Feature in der My Starkey-App stellt einen bedeutenden Fortschritt dar; das Tool hilft Hörsystemträgern, Veränderungen des eigenen Gleichgewichtszustands zu überwachen und Risikofaktoren zu verstehen, die sie betreffen können. Das Feature basiert auf international anerkannten medizinischen Richtlinien zur Sturzprävention und hält sich eng an die von den CDC empfohlenen Protokolle für Vorsorgeuntersuchungen und Funktionsbeurteilungen. Das Tool beginnt mit einer elektronischen Version des Zwölf-Punkte-Fragebogens „*Stay Independent*“ („Unabhängig bleiben“) der STEADI-Initiative. Im Anschluss werden die eingebetteten Bewegungssensoren und die künstliche Intelligenz der Starkey-Edge-AI-Hörsysteme genutzt, um die Leistung des Nutzers mithilfe funktionaler Untersuchungen zu bewerten.

Während der Nutzer die verschiedenen Aufgaben durchläuft, werden die Echtzeitdaten der in den Hörgeräten implementierten Bewegungssensoren von KI-Algorithmen ausgewertet, die selbstständig eine Bewertung vornehmen und die Ergebnisse aussagekräftig darstellen. Die Algorithmen sind so konzipiert, dass sie die Bewegungen des Nutzers während der Tests ähnlich wie ein ausgebildeter Arzt bewerten. Sie wurden in mehreren Studien des Stanford-Medicine-Departments für HNO-Heilkunde und Kopf- und Halschirurgie unabhängig validiert, wie in diesem Whitepaper beschrieben.

Zum ersten Mal können Edge-AI-Hörsysteme die Leistung der Hörsystemnutzer in allgemein

empfohlenen Tests zur Bewertung des funktionalen Gangs, der Kraft und des Gleichgewichts älterer Menschen selbstständig bewerten. Dieser innovative Ansatz hilft bei der Identifizierung von Personen, die möglicherweise Gleichgewichtseinschränkungen haben, die weitere Untersuchungen rechtfertigen. Eine benutzerfreundliche Oberfläche, die mithilfe von Beiträgen und Feedback mehrerer Gruppen von potenziellen Nutzern und Experten aus dem klinischen Bereich entwickelt wurde, führt die Hörsystemträger mit schriftlichen, illustrierten und auch Audioanweisungen effektiv durch jede Phase der Bewertung.

Forschungsstudien

Das Feature zur Gleichgewichtsbeurteilung wurde im Rahmen einer mehrjährigen Zusammenarbeit zwischen Starkey und Stanford Medicine entwickelt, wie in Abbildung 1 dargestellt (*siehe oben auf der nächsten Seite*). Ursprünglich sammelte Starkey hierfür Trainingsdaten von 71 Teilnehmern im Alter von 47 bis 99 Jahren (Durchschnittsalter: 79,2 Jahre, Standardabweichung (SD): 9,9) zur Erstellung von Algorithmen für eine autonome Bewertung von Gang-, Kraft- und Gleichgewichtstests unter Verwendung von Bewegungssensordaten aus Forschungshörgeräten (*Burwinkel et al., 2022*). Anschließend wurde eine App zur Erfassung anonymisierter Daten entwickelt, die es den Stanford-Medicine-Wissenschaftlern ermöglichte, jeden Test einzeln anhand der Bewertungsprotokolle der STEADI-Initiative zu analysieren – unabhängig von der automatischen Bewertung der App. Die Genauigkeit der Algorithmen wurde später durch den Vergleich der Analysen der Wissenschaftler und der automatischen App-Ergebnisse bewertet.

Auf der Grundlage dieser Bewertungen wurden die Algorithmen von Starkey schrittweise verfeinert und formative Studien zur Benutzerfreundlichkeit mit Prototypversionen des Tools durchgeführt. Das Stanford-Medicine-Team untersuchte die Machbarkeit von Fernbewertungen und verglich die Leistung der Gleichgewichtsbewertungsfunktion bei persönlichen, fernüberwachten und unabhängig durchgeführten Tests zu Hause. Die letzte Entwicklungsphase umfasste eine Feldvalidierungsstudie, in der die Leistung der Teilnehmer in überwachten Labortests mit der Leistung bei unbeaufsichtigter Nutzung zu Hause verglichen wurde. Die möglichen Leistungsunterschiede zwischen beaufsichtigten und unbeaufsichtigten Bewertungen wurden von Medizinerinnen untersucht.

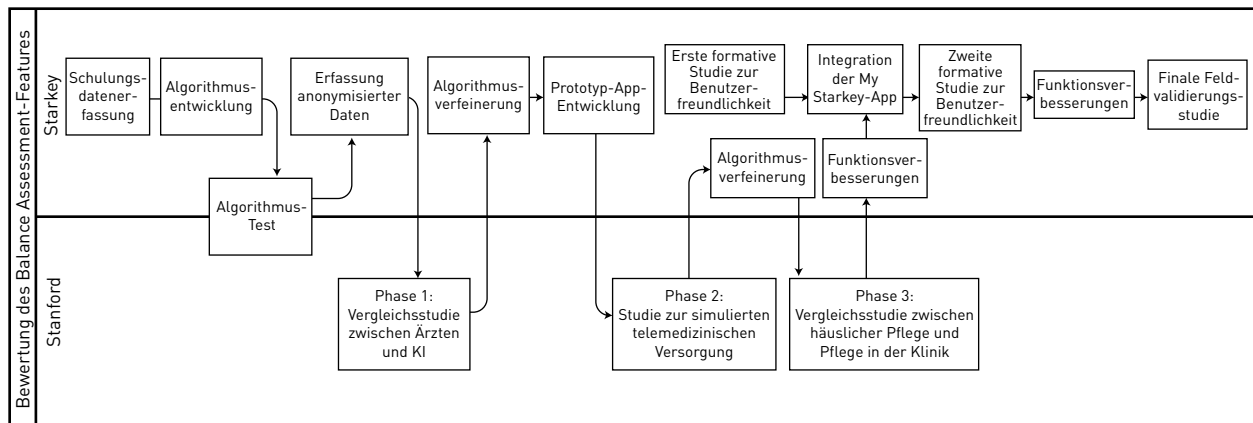


Abbildung 1: Forschungs- und Entwicklungsprozess der Balance Assessment-Funktion, eine Zusammenarbeit zwischen Starkey und Stanford Medicine. Dies umfasste die Entwicklung von Algorithmen, die Validierung von Modalitäten für Ärzte und Verwaltung, Gebrauchstauglichkeitsstudien und eine abschließende Feldvalidierung, bei der überwachte Laborversuche mit unbeaufsichtigter Anwendung zu Hause verglichen wurden.

Stanford-Medicine-Studie

Stanford Medicine führte eine dreiphasige Forschungsstudie zur Bewertung der Genauigkeit und Zuverlässigkeit der Algorithmen des Balance-Assessment-Features (zur Bewertung verschiedener Sturzrisikofaktoren) im Vergleich zu den Bewertungen erfahrener Ärzte durch – unter Verwendung der funktionellen Gang-, Kraft- und Gleichgewichtstestverfahren der STEADI-Initiative. Die Bewertungsprotokolle der STEADI-Initiative umfassen die Durchführung von drei Tests: ein vierstufiger Gleichgewichtstest, der Test, wie oft ein Proband sich in 30 Sekunden hinsetzen und wieder aufstehen kann und die Zeitmessung, wie lange der Proband zum Aufstehen und Losgehen benötigt (Timed-up-and-go-Test (TUG)). Jeder dieser Tests ist für seine empirische Gültigkeit und die Vorhersagegenauigkeit zur Ermittlung des Sturzrisikos bei älteren Erwachsenen bekannt.

In Phase 1 der Stanford-Medicine-Studie wurden die Beobachtungen der Ärzte mit den autonomen Bewertungen früherer Algorithmusversionen des Balance-Assessment-Features verglichen. In Phase 2 wurde die Umsetzung des Features in der Telemedizin untersucht. In Phase 3 wurde die Wirksamkeit unbeaufsichtigte Tests zu Hause bewertet und untersucht, wie gut der Forschungsprototyp des Balance-Assessment-Features außerhalb einer kontrollierten Laborumgebung funktioniert.

Phase 1

In der ersten Phase der Studie wurde ein anonymisierter Vergleich zur Bewertung des Sturzrisikos durchgeführt, bei dem sowohl die Einschätzungen der Ärzte als auch die autonomen Bewertungen der mit Bewegungssensoren und KI-Technologien ausgestatteten Hörgeräte berücksichtigt wurden.

Die Studie wurde in einem tertiären Fachzentrum durchgeführt und umfasste 250 Teilnehmer im Alter von 55 bis 100 Jahren, bei denen ein erhöhtes Sturzrisiko aufgrund ihres fortgeschrittenen Alters und/oder einer Vorgeschichte mit Instabilität oder Stürzen festgestellt wurde (Durchschnittsalter: 78,4 Jahre, SD: 9,6).

Verfahrensablauf

Die Teilnehmer unterzogen sich den Funktionstests der STEADI-Initiative der CDC zu Gang, Kraft und Gleichgewicht, während sie beidseitig mit Bewegungssensoren ausgestattete Hörgeräte trugen. Jeder Versuch wurde unabhängig voneinander von den Algorithmen der Hörgeräte und drei Ärzten interpretiert: einer war während der Tests anwesend und zwei überprüften später die Videoaufzeichnungen derselben Tests. Die Ärzte kannten die Ergebnisse der autonomen Bewertung der Hörgeräte nicht. Zu den wichtigsten gemessenen Ergebnissen gehörten sowohl die algorithmischen als auch die manuell ermittelten Ergebnisse aus den Beobachtungen der drei Ärzte bei den Testversuchen mit vierstufigem Gleichgewichtstest, 30-Sekunden-Stuhlstandtest und TUG-Test.

Ergebnisse

Insgesamt konnten die Wissenschaftler festhalten, dass bereits eine gute Übereinstimmung zwischen der Bewertung der Ärzte und der automatischen Bewertung des frühen Balance Assessment-Algorithmus vorlag, wenn es um die Bewertung des Sturzrisikos gemäß den Protokollen der STEADI-Initiative ging. Es gab keine statistisch signifikanten Unterschiede zwischen den Interpretationen der Ärzte und den Algorithmen des Balance Assessment-Features für den vierstufigen Gleichgewichtstest und den TUG-Tests ($p > 0,05$).

Lediglich beim 30-sekündigen Test zum Aufstehen und Hinsetzen wurde ein signifikanter Unterschied beobachtet ($t = 10,13$, $p < 0,05$), mit einer durchschnittlichen Differenz von -0,8 Wiederholungen. Die Übereinstimmung der Bewertungen unter den Ärzten war ausgezeichnet.

Diskussion

Während die frühen Algorithmen des Balance Assessment-Features insgesamt genaue Ergebnisse lieferten, wurde durch die konsequente Unterstreichung des 30-Sekunden-Stuhlstandtests eine Diskrepanz deutlich, die die Wissenschaftler darauf zurückführten, dass das Balance Assessment-Feature die Bewertungsrichtlinien der CDC zu früh anwandte. Insbesondere weisen die CDC die Ärzte an, Stuhlstände, die vor Ablauf der Testzeit eingeleitet werden, als vollständige Stände zu zählen, eine Überlegung, die ursprünglich nicht in den Bewertungsalgorithmus für Hörgeräte für diesen Test aufgenommen wurde. Diese Erkenntnis machte eine Verfeinerung des Algorithmus für die folgenden Phasen erforderlich, um die Bewertung des Balance Assessment-Features besser an die CDC-Standards anzupassen.

Phase 2

Die Phase 2 der Stanford-Medicine-Studie baute auf den Forschungsergebnissen der Phase 1 auf. Es erfolgte eine telemedizinische Bewertung des Sturzrisikos anhand der Bewertungsprotokolle der STEADI-Initiative. Diese Phase konzentrierte sich auf die Durchführbarkeit einer dezentralen Auswertung funktioneller Gang-, Kraft- und Gleichgewichtstests. Phase 2 wurde im selben Fachzentrum unter Verwendung derselben anonymisierten Datenerfassungs-App wie in Phase 1 durchgeführt. Die Untersuchung umfasste 50 Teilnehmer im Alter von 57 bis 98 Jahren, die aufgrund ihres fortgeschrittenen Alters und/oder einer Vorgeschichte mit Instabilität oder Stürzen ein erhöhtes Sturzrisiko aufwiesen (Durchschnittsalter: 78,2 Jahre, SD: 8,1).

Verfahrensablauf

In dieser zweiten Phase unterzogen sich die Teilnehmer den Funktionstests der STEADI-Initiative, während sie beidseitig mit Bewegungssensoren ausgestattete Hörgeräte trugen. Anders als in Phase 1 wurden die überwachten Tests nun aus der Ferne durchgeführt, wobei ein Arzt über einen sicheren Videoanruf Anweisungen und Aufsicht erbrachte.

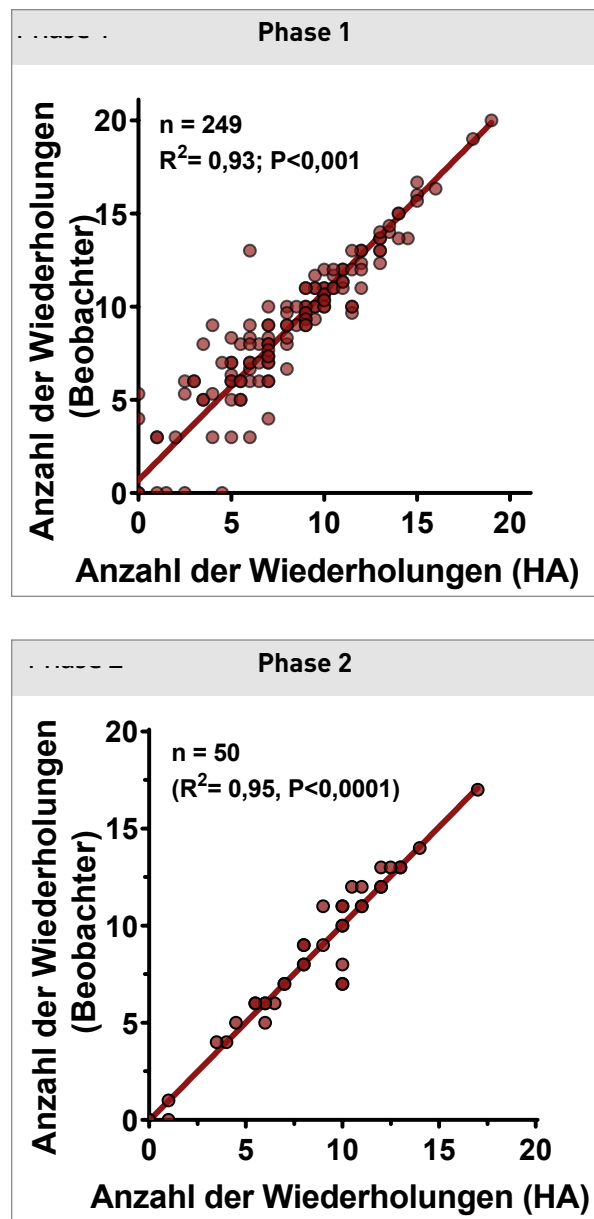


Abbildung 2: Vergleich der Algorithmusleistung der Balance Assessment-Funktion im 30-Sekunden-Stuhlstandtest über zwei Phasen hinweg. Das obere Diagramm stellt die Daten der Phase 1 dar ($n = 249$, $r^2 = 0,93$; $p < 0,001$) und das untere Diagramm die Daten der Phase 2 ($n = 50$, $r^2 = 0,95$; $p < 0,0001$). Die Y-Achsen zeigen die von den Ärzten beobachteten Gesamtstände, während die X-Achsen die Gesamtstände anzeigen, die von der Balance Assessment-Funktion aufgezeichnet wurden. Die engere Ausrichtung in Phase 2 deutet auf eine verbesserte Algorithmusgenauigkeit hin.

Ein Wissenschaftler war im Labor anwesend, um jeden Teilnehmer aus Sicherheitsgründen zu überwachen, beteiligte sich aber ansonsten nicht an der Durchführung oder Leitung der Tests. Jeder Teilnehmer führte einen Versuch für jede Aufgabe durch, der von den Algorithmen der Hörgeräte ausgewertet und von einem Arzt bewertet wurde, der Zugriff auf die Videoaufzeichnungen der Versuche hatte. Der Arzt kannte die Bewertung der Hörgeräte nicht, um einen unvoreingenommenen Vergleich zwischen dem Ergebnis des Algorithmus und seiner eigenen Einschätzung zu gewährleisten. Zu den wichtigsten gemessenen Ergebnissen gehörten sowohl die algorithmischen Bewertungen als auch die Bewertungen der klinischen Beobachtung für den vierstufigen Gleichgewichtstest, den 30-Sekunden-Stuhlstandtest und den TUG-Test.

Ergebnisse

Die Ergebnisse aus Phase 2 stimmten weitgehend mit denen aus Phase 1 überein; sie zeigten eine gute Gesamtübereinstimmung der Balance-Assessment-Feature-Algorithmen sowie der Bewertung der Leistung im Gang-, Kraft- und Gleichgewichtstest durch den Arzt gemäß den Bewertungsprotokollen der STEADI-Initiative.

4-stufiger Gleichgewichtstest: Eine etwas bessere Übereinstimmung zwischen den Algorithmen des Balance Assessment-Features und dem Arzt wurde bei den Gleichgewichtspositionen „Seite an Seite“, „Zehe zum Spann“ und „Einbeinstand“ beobachtet, während der „Tandem-Stand“ eine schlechtere Übereinstimmung zeigte. Diese Unterschiede waren relativ minimal und könnten auf die erhöhte Varianz zurückzuführen sein, die sich aus der geringeren Stichprobengröße in Phase 2 im Vergleich zu Phase 1 ergibt.

30-Sekunden-Stuhlstandtest: Der Algorithmus des Balance Assessment-Features zur Bewertung des 30-Sekunden-Stuhlstandtests zeigte eine signifikante Verbesserung von Phase 1 ($n = 249$, $r^2 = 0,93$; $p < 0,001$) auf Phase 2 ($n = 50$, $r^2 = 0,95$; $p < 0,001$), wie in Abbildung 2 (siehe vorherige Seite) dargestellt. In Phase 2 war der Unterschied zwischen den Ergebnissen des Algorithmus und den Beobachtungen des Arztes statistisch nicht mehr signifikant ($n = 50$, $t = 0,24$, $p > 0,05$).

Dieses Ergebnis unterstreicht die Wirksamkeit der zwischen Phase 1 und Phase 2 vorgenommenen Algorithmusverfeinerungen, die zu einer besseren Abstimmung zwischen dem Algorithmus und den Zählsummen des Arztes führen.

Timed Up and Go (TUG)-Test: In Phase 1 zeigten die Algorithmen des Balance Assessment-Features keinen signifikanten Unterschied zu den klinischen Bewertungen des TUG-Tests ($n = 220$, $t = 0,94$; $p > 0,05$). In Phase 2 zeigten die Algorithmen jedoch einen signifikanten Unterschied ($n = 48$, $t = 2,51$; $p < 0,05$), wobei im Allgemeinen langsamere TUG-Zeiten aufgezeichnet wurden. Dieser Unterschied deutet darauf hin, dass die dezentrale Durchführung des Tests in Phase 2 möglicherweise zu einer zusätzlichen Variabilität oder zu Herausforderungen geführt hat, die sich auf die Leistung der Algorithmen im Vergleich zur direkten Beobachtung durch den Arzt ausgewirkt haben. Dennoch scheint diese Auswirkung in erster Linie die Spezifität des Tests beeinflusst zu haben (d. h. seine Fähigkeit, Nicht-Sturzfälle auszuschließen) und nicht seine Sensitivität (d. h. seine Fähigkeit, potenzielle Sturzfälle zu erkennen), was darauf hindeutet, dass das Balance Assessment-Feature für die Überwachung der TUG-Testleistung eines Nutzers ein brauchbares Hilfsmittel blieb und trotz der beobachteten Unterschiede weiter erforscht werden sollte. Abbildung 3 (siehe nächste Seite)

Diskussion

Phase 2 der Stanford Medicine-Studie zeigte die Machbarkeit von Fernbewertungen zur genauen Beurteilung von Gang, Kraft und Gleichgewicht bei älteren Erwachsenen mit Sturzrisiko. Die dezentrale Durchführung dieser Aufgaben erwies sich auch als sicher, da keine Stürze während der Durchführung der Geh-, Kraft- und Gleichgewichtstests der Teilnehmer auftraten.

Ähnlich wie in Phase 1 war auch die Genauigkeit der Fernbewertung der Algorithmen im Allgemeinen hoch. Die Übereinstimmung zwischen den Algorithmen des Balance Assessment-Features und den Beobachtungen des Arztes war bei den meisten Aufgaben hoch.

Insbesondere verbesserte sich die Genauigkeit des Krafttests von Phase 1 zu Phase 2 aufgrund der Verfeinerungen des Algorithmus, die zwischen den Phasen 1 und 2 vorgenommen wurden. Allerdings wurden bei den Gleichgewichtsaufgaben und dem Gangtest einige erhöhte Abweichungen beobachtet, was wahrscheinlich auf die geringere Stichprobengröße und die potenziellen Herausforderungen von Fernbewertungen zurückzuführen ist. Diese Ergebnisse verdeutlichten, dass klare Anweisungen visuell und verbal gegeben werden müssen.

Insgesamt hat Phase 2 gezeigt, dass die dezentrale Bewertung von Gang, Kraft und Gleichgewicht sowohl genau als auch sicher sein kann und eine solide Grundlage für weitere Verbesserungen des frühen Balance Assessment-Features bietet.

Phase 3

In der dritten Phase der Stanford-Medicine-Studie sollte das Sturzrisiko der Teilnehmer anhand der Protokolle der STEADII Initiative in häuslicher Umgebung unter Verwendung bilateraler Hörgeräte, die mit Bewegungssensoren und KI-Technologien ausgestattet sind, bewertet werden. Diese Studie umfasste 50 Teilnehmer im Alter von 56 bis 97 Jahren (Durchschnittsalter = 76,0, SD = 8,3), bei denen ein erhöhtes Sturzrisiko auf Grundlage selbst eingeschätzter Kriterien gemäß STEADI-Initiative „Stay Independent“ („Unabhängig bleiben“) festgestellt wurde.

Verfahrensablauf

Die Teilnehmer absolvierten drei Versuche zu Hause: einen Testversuch, gefolgt von zwei weiteren Versuchen zur Bewertung. Jeder Test umfasste den vierstufigen Gleichgewichtstest, den 30-Sekunden-Stuhlstandtest und den TUG-Test, wobei die Bewertungen mithilfe der in den Hörgeräten integrierten Technologien durchgeführt wurden.

Ergebnisse

Die Tests zu Hause ergaben eine unterschiedliche Übereinstimmung zwischen den Versuchen für den 4-stufigen Gleichgewichtstest, die von 76 % bis 87 % für die drei Standpositionen reicht, die in den Bestanden/Nicht bestanden-Kriterien der STEADI-Initiative verwendet werden, nachdem Versuche entfernt wurden, bei denen aufgrund eines Nutzerfehlers mit der Forschungsversion der Anwendung keine Bewertung abgegeben wurde.

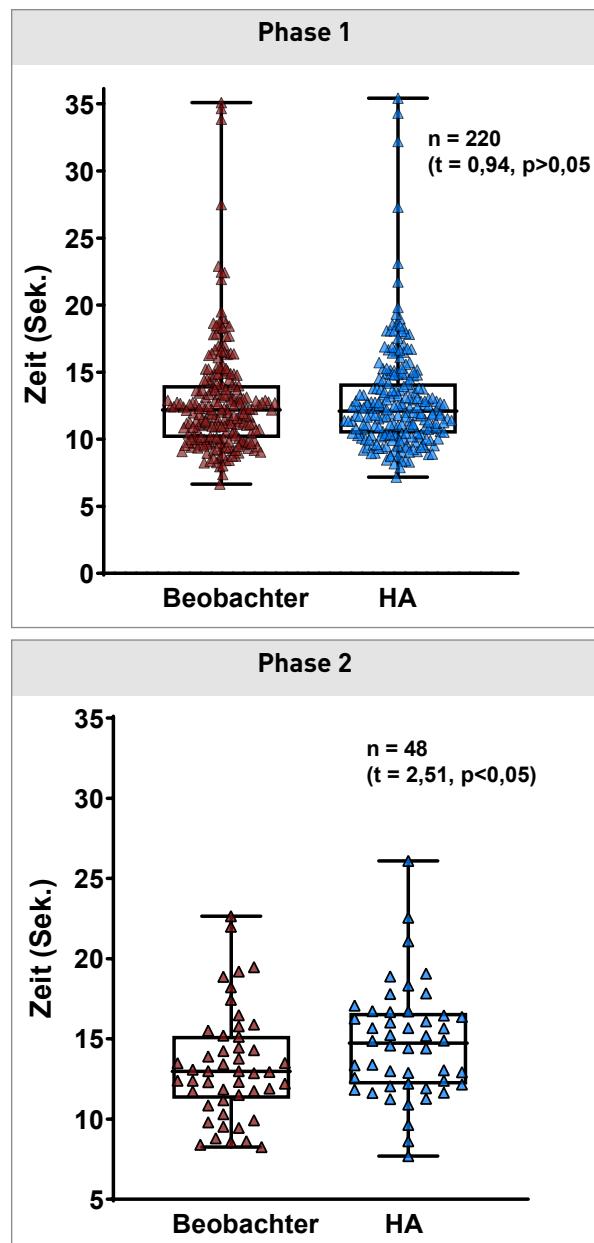


Abbildung 3: Vergleich der von Ärzten (Beobachter) aufgezeichneten TUG-Testzeiten und der Algorithmen (HA) der Balance Assessment-Funktion während Phase 1 (links) und Phase 2 (rechts). Jeder Boxplot stellt die Verteilung der in Sekunden gemessenen TUG-Zeiten dar, wobei einzelne Datenpunkte über die Plots verteilt sind. Die Boxplots heben die zentrale Tendenz und Variabilität der TUG-Zeiten für beide Phasen hervor und veranschaulichen den deutlichen Unterschied in der Leistung des Algorithmus, der in Phase 2 beobachtet wurde, im Vergleich zur engeren Ausrichtung in Phase 1.

Der 30-Sekunden-Stuhlstandtest zeigte zunächst eine moderate Test-Retest-Zuverlässigkeit. Es kam jedoch zu ähnlichen „Fehlinterpretationen“, als bei einem Versuch der Teilnehmer null Stände aufgezeichnet wurden, während bei einem anderen Versuch ein Messwert aufgezeichnet wurde. Die Korrektur dieser Fehlmessungen zeigte eine hohe Zuverlässigkeit ($n = 41$, $R2 = 0,83$, $p < 0,001$) für den 30-Sekunden-Stuhlstandtest. Ebenso zeigte der TUG-Test eine mäßige Zuverlässigkeit, da etwa 20 % der Versuche falsch interpretiert wurden. Nach Korrektur auf Fehlanzeigen verbesserte sich die Zuverlässigkeit signifikant ($n = 40$, $R2 = 0,84$, $p < 0,001$) für den TUG-Test.

Diskussion

In Phase 3 wurden große Herausforderungen bei der Durchführung von Sturzrisikobewertungen zu Hause mithilfe einer mobilen Anwendung festgestellt. Die Wissenschaftler bewerteten einen beachtlichen Prozentsatz der Versuche als „Fehleinschätzungen“, die sich darauf zurückführen lassen, dass Nutzer Schwierigkeiten bei der Anwendung hatten. Einzelberichte von Teilnehmern unterstrichen die Bedenken hinsichtlich der Benutzerfreundlichkeit der Forschungsversion der mobilen Anwendung.

Trotz dieser Herausforderungen zeigte die Studie, dass die Beurteilungsprotokolle der STEADI-Initiative mit Hörgeräten zu Hause sicher durchgeführt werden können. Wichtig ist, dass es bei der Selbsteinschätzung keine Berichte über Stürze oder Gleichgewichtsstörungen gab, was darauf hindeutet, dass das Protokoll für die Anwendung zu Hause durch diese Bevölkerungsgruppe geeignet ist.

Formative Tests zur Gebrauchstauglichkeit

Das Balance-Assessment-Feature wurde in einem umfassenden Prozess entwickelt, in den das Feedback von Nutzern sowie Ärzten einfluss. Das Ziel von Starkey war es, ein benutzerfreundliches, effektives und zuverlässiges Tool zu entwickeln, mit dem Hörgeräteträger ihr Gleichgewicht beurteilen und ihre Unabhängigkeit bewahren können. Dieser Prozess umfasste mehrere Stufen von formativen Gebrauchstauglichkeitstests (Usability-Tests) – zusätzlich zu den Leistungstests des Algorithmus –, um das Design und die Funktionalität des Features auf Grundlage der Erkenntnisse und Erfahrungen von Nutzern zu verfeinern.

Verfahrensablauf

Vor der Integration in die My Starkey-App wurde das Design des Balance Assessment-Features – einschließlich der Benutzeroberfläche, des Navigationsflusses und der Lehrinhalte – durch direkte Beobachtungen und strukturierte Interviews mit potenziellen Nutzern bestimmt. Frühe Gebrauchstauglichkeitstests zielten darauf ab, die Vorlieben der Nutzer zu verstehen, Hindernisse bei der Erledigung von Aufgaben zu identifizieren und sicherzustellen, dass den Nutzern klare Anweisungen gegeben wurden.

Nach der Entwicklung eines Prototyps der App konnte das Stanford-Medicine-Department für HNO-Heilkunde und Kopf- und Halschirurgie Studien durchführen, in denen die Wirksamkeit des frühen Features in der Telemedizin und in häuslichen Umgebungen bewertet wurde. Diese Studien, bei denen Hörgeräte und der Prototyp des Balance Assessment-Features zur Vervollständigung der STEADI-Initiative-Protokolle verwendet wurden, ergaben mehrere Bereiche, in denen Verbesserungen möglich waren. Dazu gehörten die Reduzierung von Fehlern durch die Nutzer, die zu nicht bewerteten Versuchen geführt hatten, sowie die Verbesserung der Bewertung von Teilversuchen. Probleme wie Fehlinterpretationen aufgrund bestimmter Verhaltensweisen der Nutzer, Smartphone-App zu Beginn der Gleichgewichtstests, wurden ebenfalls identifiziert und führten zu Anpassungen der Benutzeroberfläche und Verfeinerungen des Algorithmus für eine verbesserte Stabilität und Fehlerbehandlung.

Phase 4

Eine Reihe formativer Gebrauchstauglichkeitstests wurde mit einer Nutzerversion des Balance-Assessment-Features durchgeführt, das in die My-Starkey-App integriert war. Diese Phase umfasste eine heterogene Gruppe von 15 Hörgeräteträgern mit unterschiedlicher Erfahrung im Umgang mit Smartphoneanwendungen. Die Teilnehmer trugen Edge-AI-Hörgeräte und nutzten die Funktion Balance Assessment in einer simulierten Wohnzimmerumgebung; für diese Anwendung füllten sie eine elektronische Version des Fragebogens „*Stay Independent*“ sowie funktionale Gang-, Kraft- und Gleichgewichtstests aus.

Zudem wurde ihnen eine Auswahl an Messwerkzeugen zur Verfügung gestellt, die sie für die Abmessung des Gehwegs für den TUG-Test verwenden konnten. Die Einhaltung der Aufgaben wurde während der gesamten Studie überwacht.

Ergebnisse

Die Rückmeldungen aus der zweiten Runde des formativen Gebrauchstauglichkeitstests zeigten eine deutliche Verbesserung der Benutzerfreundlichkeit. Die meisten Teilnehmer fanden die Anweisungen einfach zu befolgen und verstanden die Ergebnisse, wobei viele die Bereitschaft zum Einsatz der Funktion zu Hause zum Ausdruck brachten. Es wurden jedoch drei Schlüsselbereiche für weitere Verbesserungen identifiziert:

1. Die Teilnehmer wünschten sich mehr Hintergrundinformationen zum Zweck der einzelnen Übungen und zu den durchgeführten Messungen.
2. Die anschaulichen Anweisungen zu den einzelnen Übungen wurden von den Nutzern nicht immer beachtet.
3. Die Anweisungen für den 30-Sekunden-Stuhlstandtest und die TUG-Tests waren für einige noch unklar, was zu Missverständnissen führte.

Diskussion

Die formativen Gebrauchstauglichkeitstests waren für die Verfeinerung des Balance Assessment-Features von entscheidender Bedeutung. Anfängliche Probleme, wie nicht bewertete Versuche aufgrund von Nutzerfehlern, wurden durch algorithmische Anpassungen und klarere Anweisungen gelöst. Obwohl erhebliche Fortschritte erzielt wurden, bestanden weiterhin einige Herausforderungen hinsichtlich der Benutzerfreundlichkeit – insbesondere im Zusammenhang mit der Vermittlung von Anweisungen und dem bereitgestellten Hintergrund für jeden Test. Diese fortbestehenden Herausforderungen unterstrichen die Bedeutsamkeit des kontinuierlichen Einholens von Benutzerfeedback, um die Wirksamkeit des Features zu optimieren. Dieser iterative Verfeinerungsprozess war unerlässlich, um das Feature zur Gleichgewichtsbewertung in der abschließenden Feldvalidierungsstudie einsetzen zu können. Diese zielte darauf ab, die Stabilität der

Funktion in unbeaufsichtigten, realen Umgebungen zu testen.

Abschließende Feldvalidierung

Anhand der abschließenden Feldvalidierungsstudie sollten die Benutzerfreundlichkeit und die Effektivität des Balance-Assessment-Features sowohl in überwachten als auch in unbeaufsichtigten Umgebungen bewertet werden. Das Hauptziel der Studie bestand darin, festzustellen, ob die Funktion zuverlässige und konstante Ergebnisse in verschiedenen Umgebungen liefert und ob die von Ärzten durchgeführten Demonstrationen die Benutzerfreundlichkeit oder die Leistung der Teilnehmer beeinträchtigen.

Verfahrensablauf

14 Teilnehmer wurden aus der Starkey-Forschungsdatenbank rekrutiert und in zwei Gruppen eingeteilt: Eine Gruppe (n = 7) erhielt eine persönliche Demonstration zur Nutzung des Balance-Assessment-Features von einem Arzt, während der anderen Gruppe (n = 7) lediglich gezeigt wurde, wie man innerhalb der App auf das Feature zugreift. Die zweite Gruppe wurde gebeten, den Anweisungen selbstständig zu folgen.

Beide Gruppen absolvierten beaufsichtigte Gleichgewichtsbewertungsversuche in einer Laborumgebung. Die eingeführte Gruppe absolvierte diese Tests, bevor sie zu Hause die Bewertungen vornahm, während die Gruppe ohne Demonstration die überwachten Tests nach der Rückkehr von ihrer Feldstudie durchführte. Dies ermöglichte die Bewertung, ob die vorangegangene Demonstration die Leistung der Teilnehmer und das Verständnis der Funktion während der unbeaufsichtigten Verwendung zu Hause beeinflusste.

Bemerkenswert ist, dass ein Teilnehmer aus jeder Gruppe (n = 2) die Tests zu Hause nicht absolviert hat: Ein Teilnehmer hat sich versehentlich mit Videos zur Gleichgewichtsübung in einem anderen Abschnitt der My Starkey App beschäftigt, während der andere vergessen hat, auf die Funktion zur Gleichgewichtsbewertung zuzugreifen. Ein dritter Teilnehmer (n = 1) gab an, die Home-Tests abgeschlossen zu haben, aber seine App-Nutzungsdaten fehlten für diese Tests in der Cloud, was eine Einbeziehung dieses Teilnehmers in die Analyse ausschließt.

Quantitative Daten wurden aus den Ergebnissen des Gleichgewichtstests, des 30-sekündigen Tests zum Aufstehen und Hinsetzen und des TUG-Tests sowie den Antworten der einzelnen Teilnehmer auf den Fragebogen „*Stay Independent*“ („Unabhängig bleiben“) erhoben. Die Ergebnisse der überwachten als auch der nicht überwachten Versuche wurden mit dem t-Test von Welch analysiert. Die Gesamtergebnisse wurden mit dem Chi-Square-Unabhängigkeitstest verglichen, um die Konsistenz und Zuverlässigkeit der Balance-Assessment-Versuche zu beurteilen. Die Leistungsunterschiede zwischen der eingeführten Gruppe (n = 5) und der Gruppe ohne Demonstration (n = 6) wurden ähnlich analysiert.

Ergebnisse

Die Ergebnisse der Feldvalidierungsstudie zeigten bei den meisten Messungen keine signifikanten Unterschiede bei den Ergebnissen für Gleichgewicht, Stuhlstand und TUG oder bei den Antworten zu „*Stay Independent*“ zwischen unbeaufsichtigten und beaufsichtigten Versuchen oder zwischen den Gruppen mit und ohne Demonstration. Wie in Tabelle 1 (*siehe nächste Seite*) zusammengefasst, zeigten statistische Analysen mit dem t-Test und dem Chi-Square-Test von Welch konsistente Testergebnisse über die meisten Messungen hinweg, was auf keine nennenswerten Unterschiede in der Leistung aufgrund der Testumgebung oder der Anweisungsmethode hindeutet. Es wurde jedoch ein signifikanter Unterschied zwischen den Gruppen mit und ohne Demonstration in den Ergebnissen des Fragebogens „*Stay Independent*“ („Unabhängig bleiben“) beobachtet (Welch t-Test: $t = -2,93$, $p = 0,01$; Chi-Square: $\chi^2 = 7,37$, $p = 0,01$). Diese Beobachtung stammte wahrscheinlich aus unkontrollierten intrinsischen Merkmalen der Teilnehmer, wobei die eingeführte Gruppe eine höhere Anzahl von selbst identifizierten Sturzrisikofaktoren wie frühere Stürze, Gleichgewichtsprobleme oder Bedenken hinsichtlich der Gehstabilität berichtete als die Gruppe ohne Demonstration. Dies scheint eher mit den von den Teilnehmern selbst wahrgenommenen Schwachstellen zusammenhängen als nur mit der Unterweisungsmethode.

Diskussion

Die abschließende Feldvalidierungsstudie unterstrich die Stabilität des Balance Assessment-Features sowohl in überwachten als auch in unbeaufsichtigten Umgebungen und bestätigte seine Zuverlässigkeit für die dezentrale Bewertung in jeder der vier Teilaufgaben. Obwohl Unterschiede bei der Varianz zwischen den Probanden beobachtet wurden – insbesondere bei unbeaufsichtigten Versuchen und unter den Teilnehmern in der Gruppe ohne Demonstration im Vergleich zu denjenigen, die anbietergeführte Demonstrationen erhielten – erreichten diese Varianzen keine statistische Signifikanz. Die Funktion lieferte konsistent genaue Ergebnisse, ohne dass Stürze oder Sicherheitsvorfälle gemeldet wurden, was die Sicherheit bei der Verwendung in unbeaufsichtigten Umgebungen erhöht.

Obwohl die Aufgabenleistung weitgehend konsistent war, hatten einige Teilnehmer Schwierigkeiten, die Funktion selbstständig zu Hause aufzurufen und zu nutzen, was darauf hindeutet, dass Erstnutzer von zusätzlicher Schulungsunterstützung profitieren könnten. Diese Ergebnisse unterstreichen den potenziellen Vorteil der Anleitung durch den Hörakustiker während der ersten Nutzung, insbesondere für Nutzer, die weniger vertraut mit digitalen Tools sind.

Tabelle 1: Zusammenfassung der statistischen Tests (Welch-t-Test und Chi-Quadrat-Test) zum Vergleich von überwachten und unüberwachten Versuchen sowie von Gruppen mit und ohne Demonstration in Bezug auf Gleichgewichts-, Stuhlstand-, TUG- und Fragebogenwerte zur Unabhängigkeit. Signifikante Unterschiede wurden nur im Fragebogen „Stay Independent“ beim Vergleich der Gruppen mit und ohne Demonstration festgestellt.

Messung	Gegenüberstellung	Test	Test-Statistik	p-Wert	Signifikanter Unterschied? (p ≤ 0,05)
Gleichgewichts-bewertung	Unüberwachte vs. überwachte Bewertung	Welchs t-Test	t = -1,05	0,3	Nein
	Unbeaufsichtigtes vs. überwachtes Ergebnis	Chi-Square	x ² = 0,04	0,85	Nein
	Bewertung mit vs. ohne Demonstration	Welchs t-Test	t = 1,35	0,19	Nein
	Ergebnis mit vs. ohne Demonstration	Chi-Square	x ² = 1,99	0,16	Nein
Stuhlstand-Bewertung	Unüberwachte vs. überwachte Bewertung	Welchs t-Test	t = 1,06	0,95	Nein
	Unbeaufsichtigtes vs. überwachtes Ergebnis	Chi-Square	x ² = 0,05	0,83	Nein
	Bewertung mit vs. ohne Demonstration	Welchs t-Test	t = 1,71	0,9	Nein
	Ergebnis mit vs. ohne Demonstration	Chi-Square	x ² = 0,96	0,33	Nein
TUG-Bewertung	Unüberwachte vs. überwachte Bewertung	Welchs t-Test	t = 1,70	0,1	Nein
	Unbeaufsichtigtes vs. überwachtes Ergebnis	Chi-Square	x ² = 0,56	0,45	Nein
	Bewertung mit vs. ohne Demonstration	Welchs t-Test	t = 1,47	0,15	Nein
	Ergebnis mit vs. ohne Demonstration	Chi-Square	x ² = 1,39	0,24	Nein
Stay Independent-Fragebogen	Unüberwachte vs. überwachte Bewertung	Welchs t-Test	t = -0,24	0,81	Nein
	Unbeaufsichtigtes vs. überwachtes Ergebnis	Chi-Square	x ² = 0,02	0,9	Nein
	Bewertung mit vs. ohne Demonstration	Welchs t-Test	t = -2,93	0,01	Ja
	Ergebnis mit vs. ohne Demonstration	Chi-Square	x ² = 7,37	0,01	Ja

Zusammenfassung

Die Balance Assessment-Funktion in der My Starkey-App stellt einen bedeutenden Fortschritt bei der Verbesserung und Verwaltung des Gleichgewichts bei älteren Erwachsenen mit Hörverlust dar. Durch die Integration der STEADI-Initiative der CDC mit fortschrittlicher Sensortechnologie und KI hilft diese Funktion Hörsystemträgern, durch personalisierte, benutzerfreundliche Bewertungen, die ihre Geh-, Kraft- und Gleichgewichtsfähigkeiten messen, auf einfache Weise ein gutes Gleichgewicht zu erreichen und aufrechtzuerhalten. Diese Bewertungen können gesündere Lebensgewohnheiten fördern, indem sie das Bewusstsein für veränderbare Risikofaktoren für Stürze schärfen, eine größere Unabhängigkeit unterstützen und letztlich zu einer Verringerung des Sturzrisikos beitragen.

Häufige Bewertungen können es Nutzern ermöglichen, Änderungen ihres Gleichgewichts frühzeitig zu erkennen, sodass sie rechtzeitig Maßnahmen ergreifen und potenzielle Probleme beheben können, bevor sie sich verschlimmern. Für diejenigen, die möglicherweise keine regelmäßigen Beurteilungen von Gesundheitsdienstleistern erhalten, bietet diese Funktion ein zuverlässiges und zugängliches Tool, um das Gleichgewicht von zu Hause aus zu überwachen und sie aktiv am Gesundheitsmanagement zu beteiligen. Zusammen helfen diese Vorteile den Nutzern, sich ihres Gleichgewichtszustands bewusst zu bleiben und fundierte Entscheidungen zu treffen, die das Sturzrisiko verringern können.

Die in diesem White Paper vorgestellten Studien bestätigen, dass das Balance Assessment Feature auch bei unbeaufsichtigtem Einsatz sehr effektiv ist und sowohl Sicherheit als auch Zuverlässigkeit zeigt. Die Einbeziehung von Hörakustikern bleibt jedoch unerlässlich, um Nutzern dabei zu helfen, das volle Potenzial der Funktion auszuschöpfen, und um eine genaue Bewertung zu gewährleisten und zeitnahe Gespräche über Maßnahmen zu ermöglichen, wenn Gleichgewichtsprobleme auftreten. Diese Partnerschaft zwischen Technologie und Hörgeräteakustikern stellt sicher, dass die Nutzer nicht nur von den Funktionen der My Starkey-App profitieren, sondern auch die notwendige Unterstützung erhalten, um ihr Gleichgewicht proaktiv und effektiv zu verwalten.

Anerkennungen

Wir danken Matthew Fitzgerald, Robert K. Jackler, Bryn Griswold, Soumya Venkitakrishnan, Deborah M. Kado, David A. Fabry, Achintya K. Bhowmik, Amit Shahar, Matan Sivan, Roy Rozenman, Andy Lin und Archelle Georgiou für ihre unschätzbaren Beiträge zur Entwicklung des Balance Assessment-Features sowie zur durchdachten Gestaltung und Durchführung der in diesem White Paper vorgestellten Forschungsstudien.

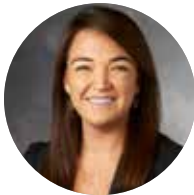
Quellennachweise

1. Agmon, M., Lavie, L. und Dumas, M. (2017). The Association between Hearing Loss, Postural Control, and Mobility in Older Adults: A Systematic Review. *Journal der American Academy of Audiology*, 28(6), 575-588. <https://doi.org/10.3766/jaaa.16044>
2. Burwinkel, J. R., Sivan, M., Rozenman, R., & Georgiou, A. (April 2022). Preliminary evaluation of ear-wearable postural stability monitoring technology. EU Falls Festival, Leuven, Belgien.
3. Centers for Medicare & Medicaid Services. (2017). Beneficiaries Utilizing Free Preventive Services by State, 2016. Centers for Medicare & Medicaid Services. <https://downloads.cms.gov/files/Beneficiaries%20Utilizing%20Free%20Preventive%20Services%20by%20State%20YTD%202016.pdf>
4. Criter, R. E., & Honaker, J. A. (2016). Audiology patient fall statistics and risk factors compared to non-audiology patients. *International Journal of Audiology*, 55(10), 564-570. <https://doi.org/10.1080/14992027.2016.1193235>
5. Eckstrom, E., Parker, E. M., Lambert, G. H., Winkler, G., Dowler, D., & Casey, C. M. (2017). Implementing STEADI in Academic Primary Care to Address Older Adult Fall Risk. *Innovation in Aging*, 1(2), igx028.
6. Howland, J., Hackman, H., Taylor, A., O'Hara, K., Liu, J., & Bruschi, J. (2018). Older adult fall prevention practices among primary care providers at accountable care organizations: Pilotstudie. *PLoS One*, 13(10), e0205279. <https://doi.org/10.1371/journal.pone.0205279>
7. Lin, F. R., & Ferrucci, L. (2012). Hearing Loss and Falls Among Older Adults in the United States. *Archives of Internal Medicine*, 172(4), 369. <https://doi.org/10.1001/archinternmed.2011.728>
8. Patterson, J. N., & Honaker, J. A. (2014). Survey of Audiologists' Views on Risk of Falling Assessment in the Clinic. *Journal of the American Academy of Audiology*, 25(4), 388-404. <https://doi.org/10.3766/jaaa.25.4.10>

Autorenbiografien



Justin R. Burwinkel, Au.D., ist Senior Research Audiologe bei Starkey. Die Forschung von Dr. Burwinkel hat zur Entwicklung verschiedener Nicht-Audio-Sensoranwendungen für den Einsatz mit Hörgeräten beigetragen. Er hat zahlreiche Patente für Erfindungen im Zusammenhang mit dem Sturzrisikomanagement, künstlichen Intelligenzanwendungen in der Audiologie und Fortschritten bei der Hörsystemkonnektivität erhalten. Dr. Burwinkel erwarb sowohl seinen B.S. als auch seinen Au.D. an der Universität von Cincinnati. Derzeit lehrt er als Associate Professor im Salus am Drexel University Osborne College of Audiology und arbeitet mit dem FETCHLAB der Universität Cincinnati zusammen.



Kristen K. Steenerson, MD, ist Vorstandszertifizierte Neurologin mit Stipendienausbildung in vestibulärer Neurologie. Sie absolvierte ihr Cum Laude-Studium am Claremont McKenna College, erhielt ihren MD an der University of Utah und absolvierte ein Residenzjahr in der Mayo Clinic in Arizona in Neurologie. Dort entdeckte sie den ungedeckten Bedarf an Behandlungen für Gleichgewichtsstörungen und Schwindel, was sie dazu motivierte, ein Fellowship in Otoneurologie am Barrow Neurological Institute zu absolvieren. Sie kam mit Positionen in den Bereichen Hals-Nasen-Ohren-Heilkunde und Neurologie an die Stanford University, um gemeinsam die Verbindung von Innenohr- und Hirnerkrankungen zu untersuchen.



Majd Srour ist Senior Software Engineer II bei Starkey, wo er mehr als sechs Jahre in der Advanced Development-Abteilung gearbeitet hat und sich auf die Entwicklung von Proof-of-Concepts, Prototypen und neuen Technologien konzentriert, die jetzt in die Hörsysteme des Unternehmens integriert werden. Bevor er zu Starkey kam, sammelte Majd wertvolle Erfahrungen in der Arbeit für Start-ups und Unternehmen, darunter die Intel Corporation.



Christy Cloninger, Ph.D., ist Senior Human Factors Engineer bei Starkey. Dr. Cloninger erwarb ihren B.A. in Psychologie am Grinnell College und ihren M.S. und Ph.D. in Neurobiologie und Anatomie am University of Rochester Medical Center. Darüber hinaus verfügt sie über acht Jahre Erfahrung im Bereich Human Factors und führt formative und summative Studien durch, um die sichere und effektive Verwendung von Medizinprodukten und Konsumgütern zu gewährleisten.



Chris Howes ist Senior Software Product Manager bei Starkey. Chris arbeitet seit über 25 Jahren bei Starkey in allen Bereichen der Produktforschung und -entwicklung. Sein Hauptfokus liegt auf dem Design und der Entwicklung von mobiler und Cloud-Software für alle Aspekte der Interaktion mit Hörsystemen. Chris besitzt mehrere erteilte Patente für eine Vielzahl von Konzepten und wurde 2016 zum Starkey Erfinder des Jahres gekürt.