Steinberg VST AmbiDecoder

Immersives Monitoring mit Kopfhörer über Binauralisierung

Autor: Peter Kaminski



3D-Audio ist auf jeden Fall die Zukunft des Tons aber nicht jedem Produzierenden steht auch ein entsprechendes Abhörsystem mit neun oder mehr Lautsprechern zur

Verfügung. Insofern ist ein Monitoring über Kopfhörer mittels binauraler Wiedergabe eine Alternative, zudem man so auch mobil an Produktionen arbeiten kann. Steinberg stellte 2020 System-Ergänzung für Cubase Pro oder Nuendo vor, mit dem man Ambisonics-Produktionen über Kopfhörer binaural abhören kann. System deshalb, weil es aus mehreren Komponenten besteht und zwar einer Software, einem VST-Plug-In und auch einer oder mehrerer ermittelter und gespeicherter Außerohr-Übertragungsfunktionen (HRTF).

Binaurales Hören

Zunächst einmal ein paar Punkte zum binauralem Hören. Ich möchte hier einmal versuchen, dass möglichst einfach zu erklären. Der AmbiEncoder ist eine Lösung, die das Abhören einer 3D-Audio-Mehrkanalproduktion, bzw. Aufnahme über Kopfhörer ermöglicht und zwar so, dass die virtuelle akustische Positionen der einzelnen Schallquellen in der Produktion bei der Kopfhörerwiedergabe an der gleichen Stelle lokalisiert werden. Es wird also eine 3D-Audio-Lautsprecherwiedergabe über Kopfhörer simuliert. Diesen Prozess nennt man Binauralisierung.

Eigentlich hören wir ja nicht mit den Ohren sondern mit dem Gehirn. Die beiden Ohren sind lediglich unsere Sensoren, bzw., die Schallwandler. Wo und in welcher Entfernung Objekte wahrgenommen werden hängt von verschiedenen Parametern ab, wie Laufzeit, bzw. Laufzeitunterschiede zwischen den beiden Ohren und auch das wahrgenommene Frequenzspektrum spielt eine große Rolle bei der Ortung. Die Wahrnehmung ist eine sehr subjektive und zudem sehr individuelle Angelegenheit. Die Prinzipien sind bei allen Personen gleich aber im Detail gibt es Unterschiede, die durch die individuelle Kopfform (Schallbeugung am Kopf und Abstand der Ohren) und durch die individuellen Formen der Ohrmuscheln und des Gehöreingangs (verursachen Unterschiede in der spektrale Wahrnehmung) begründet sind. Wenn man also zum Beispiel eine Ambisonics-Mehrkanalaufnahme mittels einer binauraler "Simulation" über Kopfhörer wiedergeben möchte, dann muss diese Simulation einmal die grundsätzlichen Prinzipien simulieren, aber auch eben die personenabhängigen individuellen Eigenschaften. Letztere werden bestimmt durch die sogenannte Außerohrübertragungsfunktion (Abk. engl.: HRTF für Head-Related Transfer Function). Mit dieser Übertragungsfunktion werden die individuellen Eigenschaften beschrieben.

Wenn man die "Signalkette" Schallquelle zum Gehirn genauer betrachtet, dann spielen auch noch die Übertragungseigenschaften des verwendeten Kopfhörertyps mit seinen individuellen Parametern eine Rolle. Diese Einflussnahme ist aber kleiner als die individuellen Eigenschaften des Hörenden. Daher ignoriert man diese Einflussnahme einfach häufig bei der Binauralisierung von 3D-Audio. Weiter sind linke und rechte Kopfhälfte, wie auch die Ohrmuscheln, nicht hundertprozentig spiegelsymmetrisch. Auch hier ist es so, dass man zur Vereinfachung aber annehmen kann, dass dies so ist. Wie weit die Symmetrie von Kopf und Ohrmuschel ausgeprägt ist, ist von Mensch zu Mensch auch wieder sehr unterschiedlich. Individuell unterschiedlich ist auch die Sensibilität wie stark man eine Ortungs- oder

Klangminderung wahrnimmt, wenn man bei der Simulation nicht die eigene, sondern eine ähnliche oder gänzlich andere HRTF nutzt.

Jetzt zum Punkt wie wird denn die HRTF erfasst und in welcher Form erfolgt die Beschreibung. Die Erfassung der HRTF kann über zwei Ansätze erfolgen und zwar über eine Messung, in dem im Gehörgang jeweils ein Miniaturmikrofon eingebracht wird und dann über Testsignale eine Vermessung erfolgt. Eine andere Methode ist die, dass man Hörmuschel/n und ggf. Kopfform geometrisch ermittelt und dann daraus die resultierende HRTF erstellt.

Bei letzterem Ansatz gibt es zwei Varianten, die beide über ein Foto/Video erfolgen. Bei einem wird die Geometrie des Kopfes inklusive Ohrmuscheln und Gehöreingang aus dem Foto ermittelt und daraus dann die HRTF berechnet. Die andere Variante ist in eine Datenbank mit Kopf/Ohr-Geometrien und mit den dazugehörigen HRTFs vorzuhalten und durch einen Vergleich die beste Übereinstimmung der Geometrien zu ermitteln und dann eben diese dazugehörige HRTF zu nutzen.

Die Daten, die die individuelle HRTF beschreiben, müssen natürlich als Datei hinterlegt werden. Das verwendete Dateiformat kann dabei entweder proprietär, also System-bezogen, oder in einem Standardformat abgelegt werden. Als Standardformat hat sich das sogenannte SOFA-Dateiformat etabliert. SOFA (Spatially Oriented Format for Acoustics) wurde von dem Standardisierungsgremium der Audio Engineering Society definiert und ist in dem Standard AES69-2015 beschrieben. Weitere Informationen findet man auch auf der Web-Site www.sofaconventions.org. Das SOTA-File trägt in der Regel die Dateiendung ".sofa" und für die verschiedenen Abtastraten gibt es einzelne SOFA-Dateien, was in der Regel im Dateinamen auch vermerkt ist.

Installation und Konzept

Nun wieder zurück zum VST AmbiDecoder, den man in Zusammenhang mit Cubase Pro ab Version 10.5 oder Nuendo ab Version 10.3, unter den Betriebssystemen Windows 10 (64 Bit) und macOS (High Sierra oder Catalina), installiert und nutzen kann. Die Lizenzierung erfolgt nicht über den sonst für Steinberg üblichen eLicenser sondern über einen Freischaltcode. Man benötigt für die Einrichtung weiter noch ein Smartphone mit hochauflösender Kamera. Insgesamt lässt sich mit der Lizenz AmbiDecoder auf fünf Rechnern installieren mit bis zu fünf verschiedenen Nutzern bzw. HRTFs.

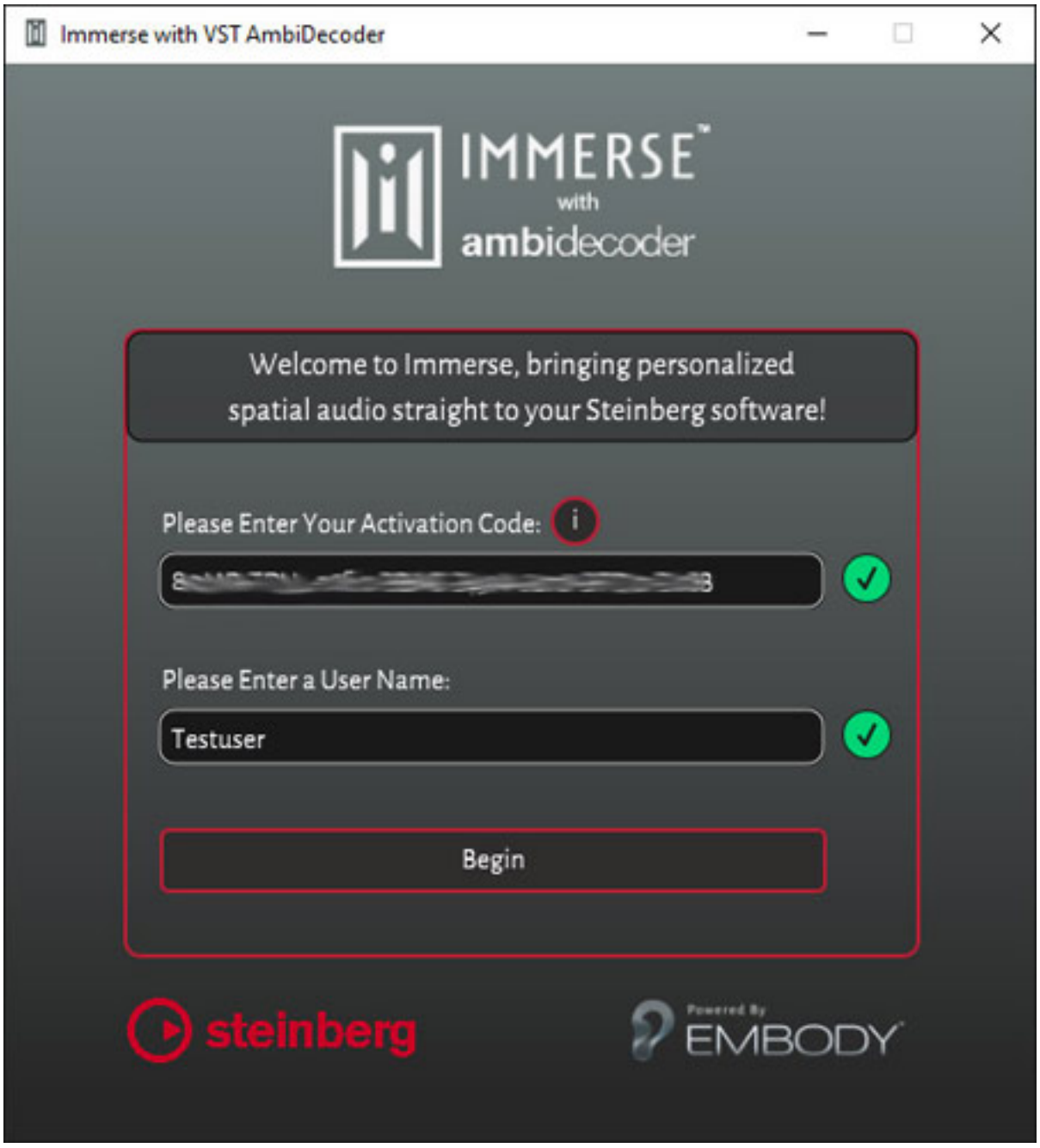
Die HRTF wird über das Erstellen eines Fotos von dem rechten Ohr berechnet. AmbiEncoder nutzt diese HRTF dann intern, bietet aber keine Möglichkeit die HRTF als SOFA zu speichern, kann aber extern verfügbare SOFA-Dateien importieren und nutzen. AmbiDecoder konvertiert Ambisonics-3D-Audio zu Binaural. Über einen kleinen Umweg kann man es aber zum Beispiel auch zum Abhören über Kopfhörer von Dolby Atmos nutzen. Dazu später mehr.

Das Erstellen der individuellen Außenohrübertragungsfunktion für den AmbiDecoder

wird über ein Technologie der Firma Embody aus Kalifornien durchgeführt. Dabei kommt sogenannte Machine-Learning-Algorithmen zum Einsatz - eine Form der künstliche Generierung von Wissen aus gesammelten Erfahrung, manchmal auch unter dem Stichwort "Expertensysteme" bekannt - ich durfte als Diplom-Arbeit mal selber eines entwickeln. Basis ist dabei eine Datenbank mit gemessenen HRTFs. Der Algorithmus analysiert das Foto und nutzt das gelernte Wissen eine HRTF auf Basis der Eigenschaften des Ohrs zu prognostizieren. Auch der Abstand der Ohren wird dabei abgeschätzt um die Interaural Time Differece (ITD), also die zeitliche Differenz des Schalleintreffens zwischen den beiden Ohren, zu berechnen und im Algorithmus zu berücksichtigen. Daher benötigt man bei der Erstellung der HRTF über dieses Verfahren auch kein kompletten 3D-Scan des Kopfes. Es wird bei dem Verfahren davon ausgegangen, dass beide Ohren spiegelsymmetrisch geformt sind.

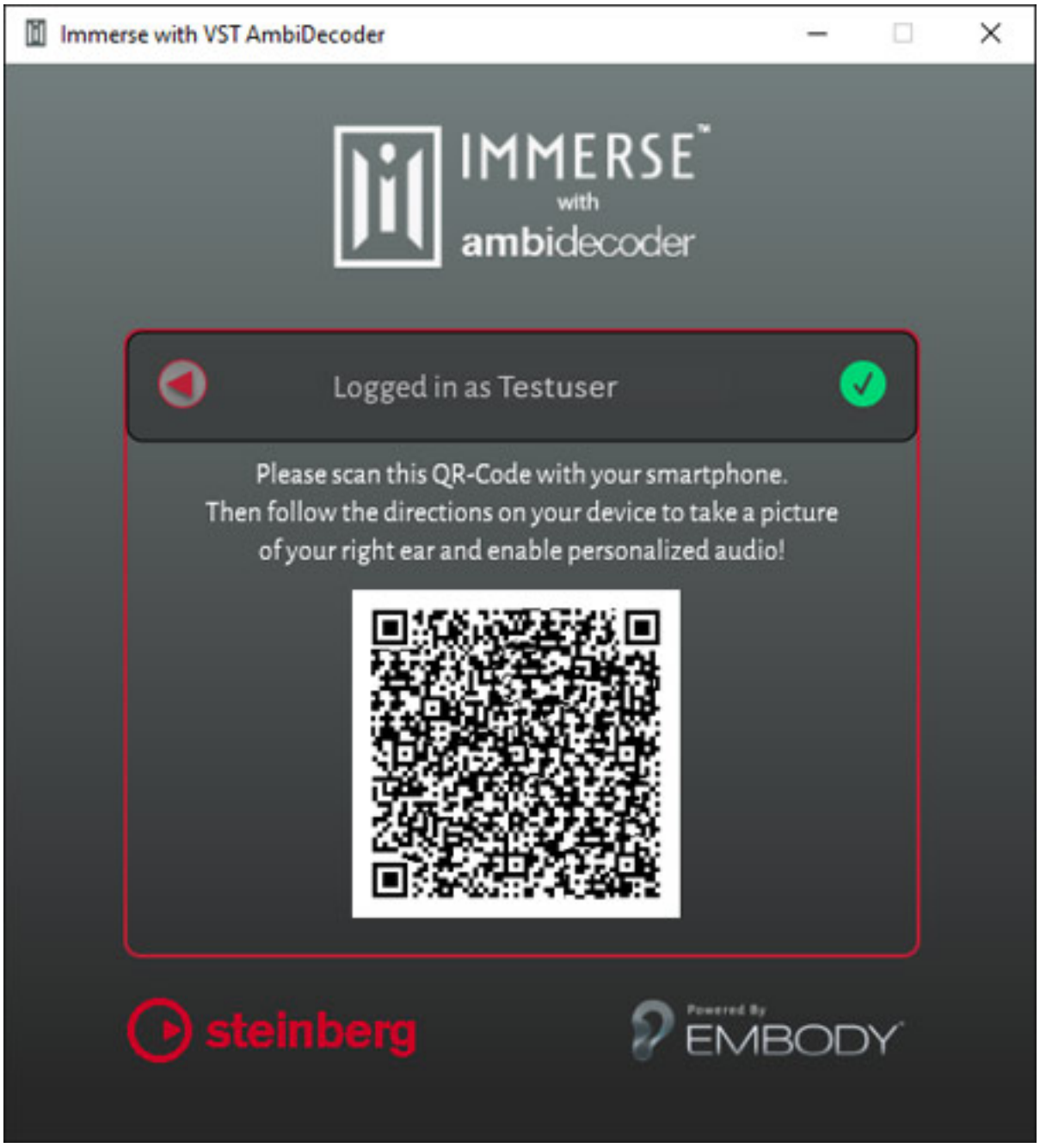
Vorbereitung

Wir möchten hier einmal die Programminstallation und den Prozess der Erstellung der HRTF über das Verfahren von Embody aufzeigen.



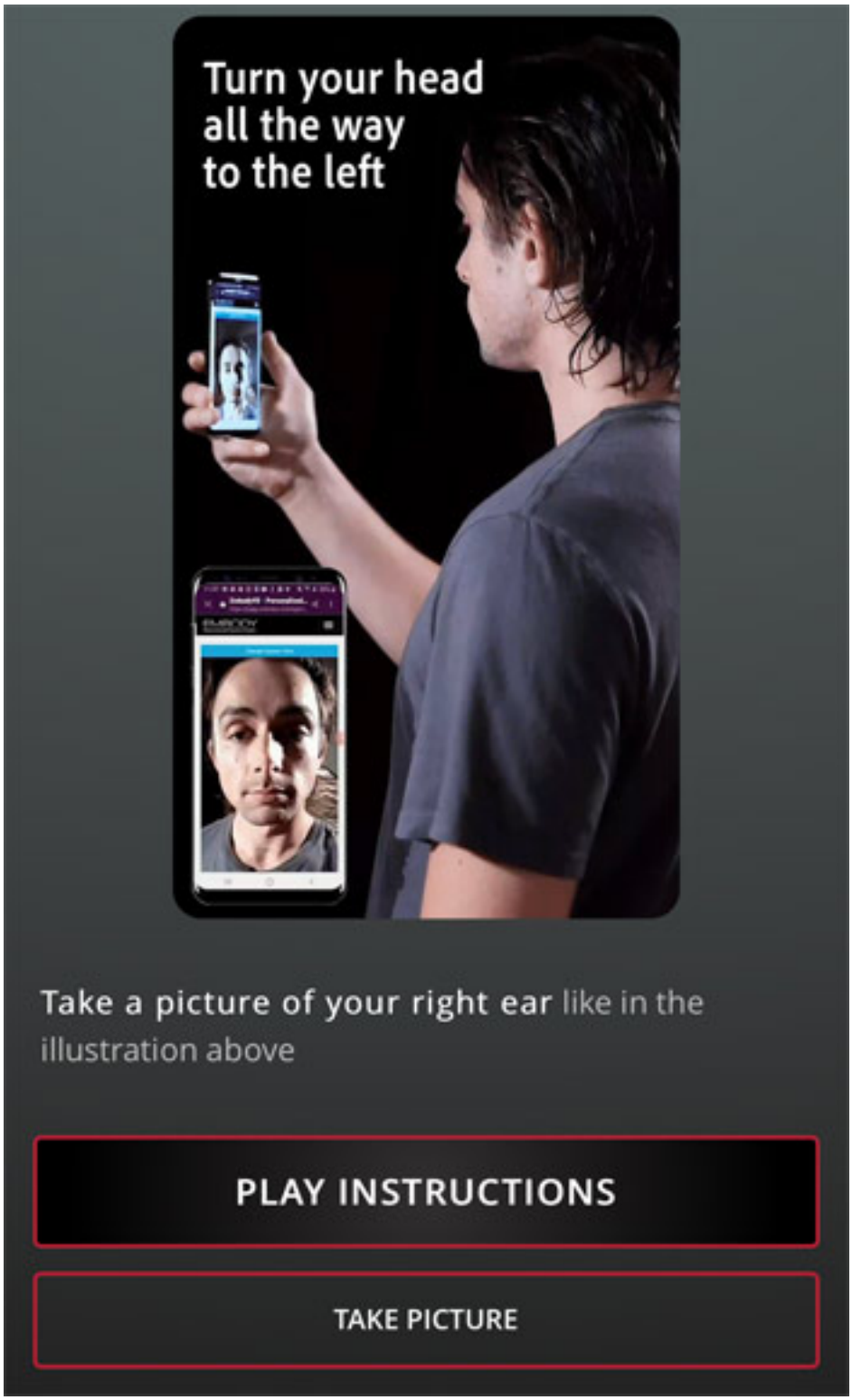
Nach der Installation muss man sein Activation Code eingeben und einen Anwendernamen festlegen.

Phoca PDF

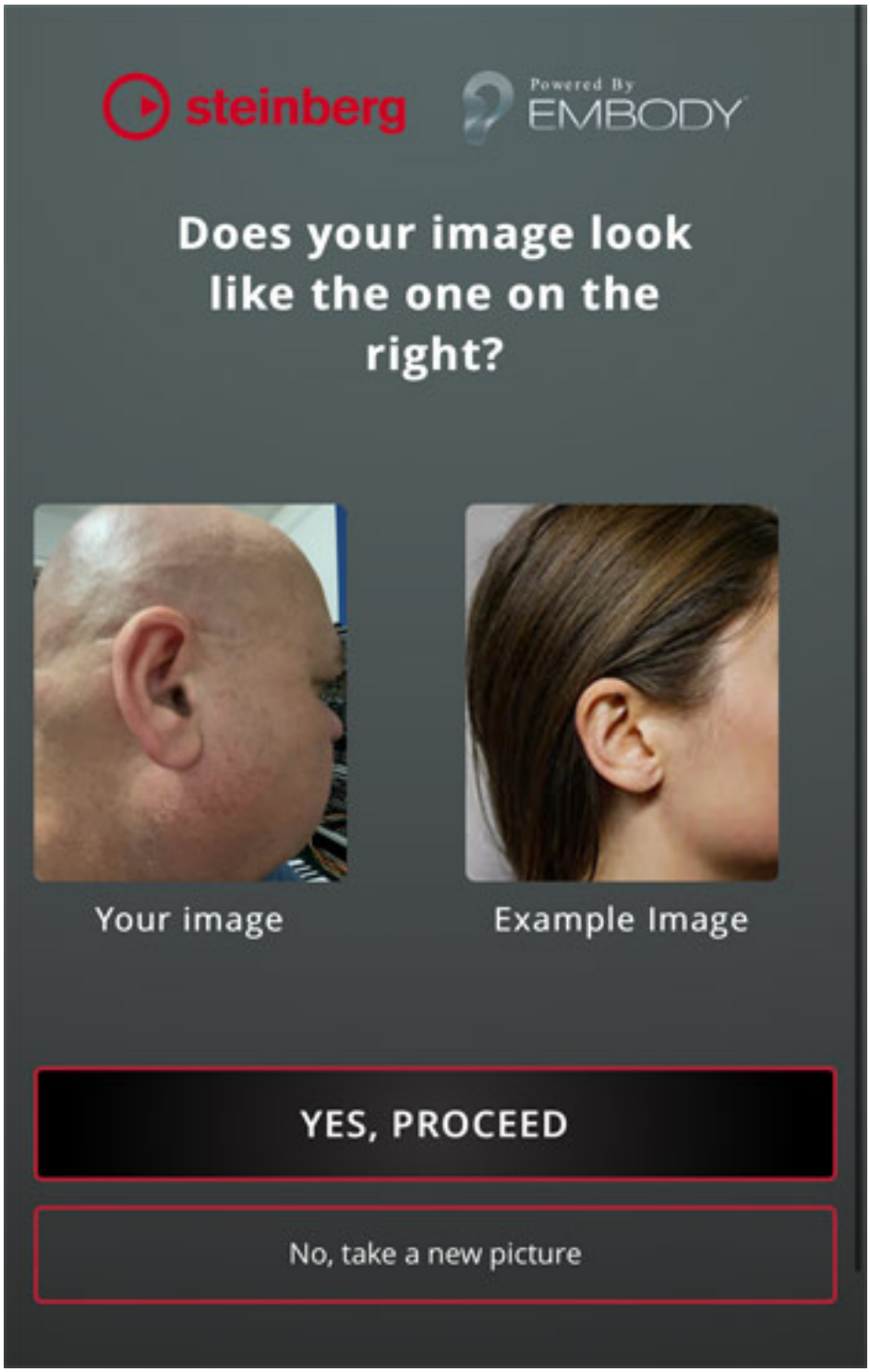


Man erhält nun einen QR-Code (der Code oben in der Abbildung ist optisch verfremdet). Den muss man mit seinem Smartphone in der Foto-App anvisieren und es sollte dann in der Regel zu einer Seite geleitet werden, mit der man das Foto für die HRTF-Ermittlung durchführt und übermittelt.

Phoca PDF



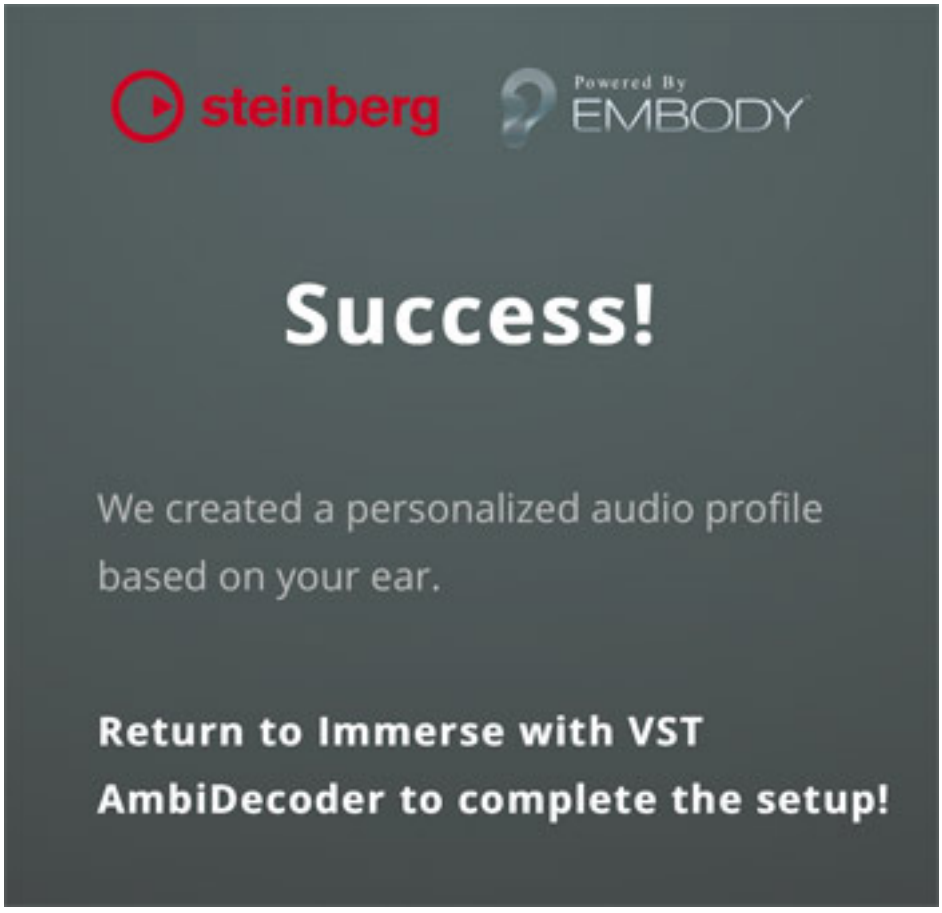
Man kann hier sich anschauen wie man das Foto machen soll. Es ist das rechte Ohr zu fotografieren und zwar soll man die Kamera von der Aufsicht dann zur Seite schwenken und über ein Antippen auf den Touchscreen wird das Foto ausgelöst.



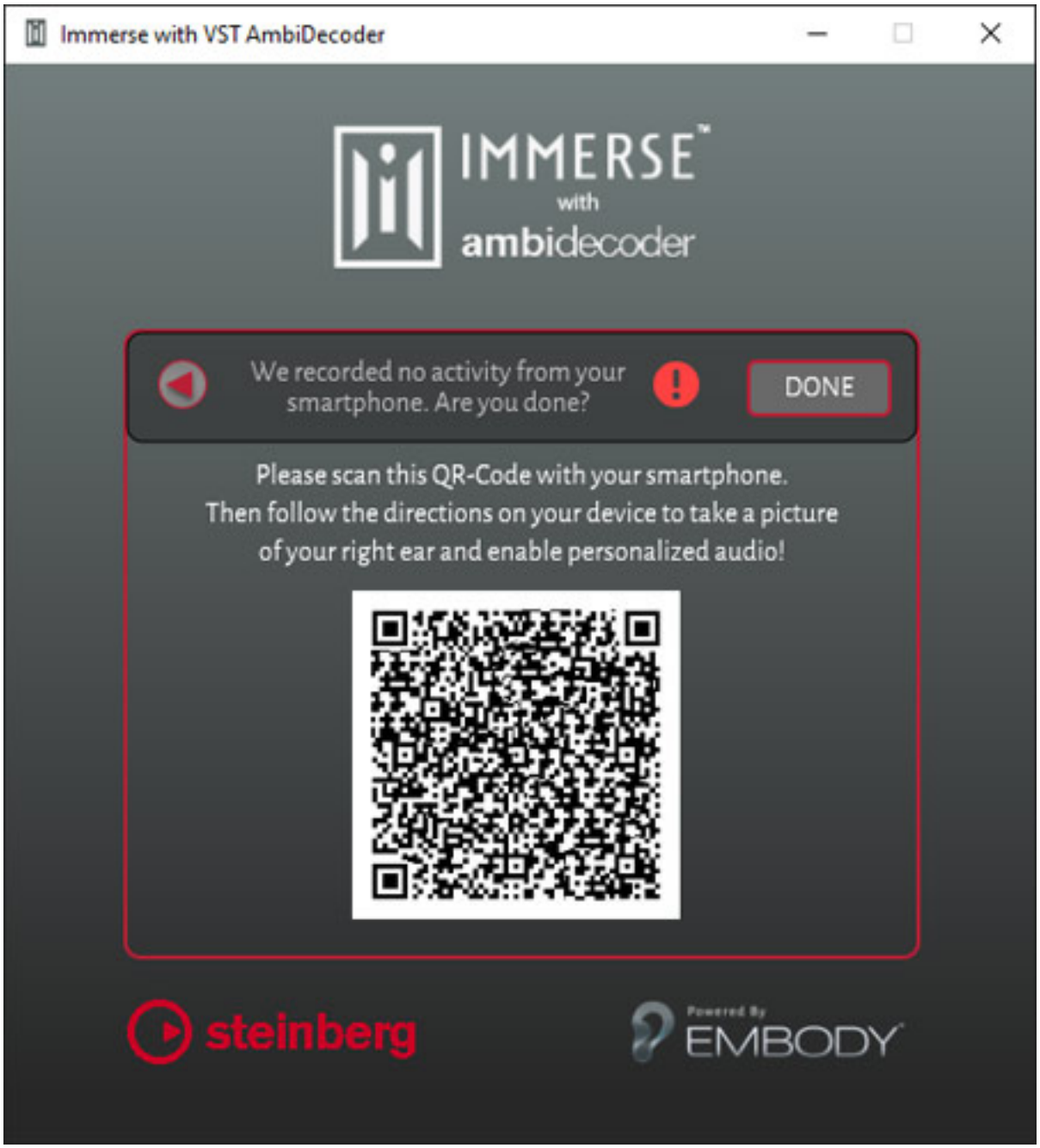
Man kann es dann auf der linken Seite sehen. Rechts ist ein Beispielfoto zu sehen. Wenn es noch nicht den Vorstellungen entspricht, kann man über "No, take a new picture" einen weiteren Versuch starten.



Wenn man dann mit "Yes Proceed" das Foto bestätigt wird der externe Processing-Vorgang über ein Ohrsymbol bestätigt.

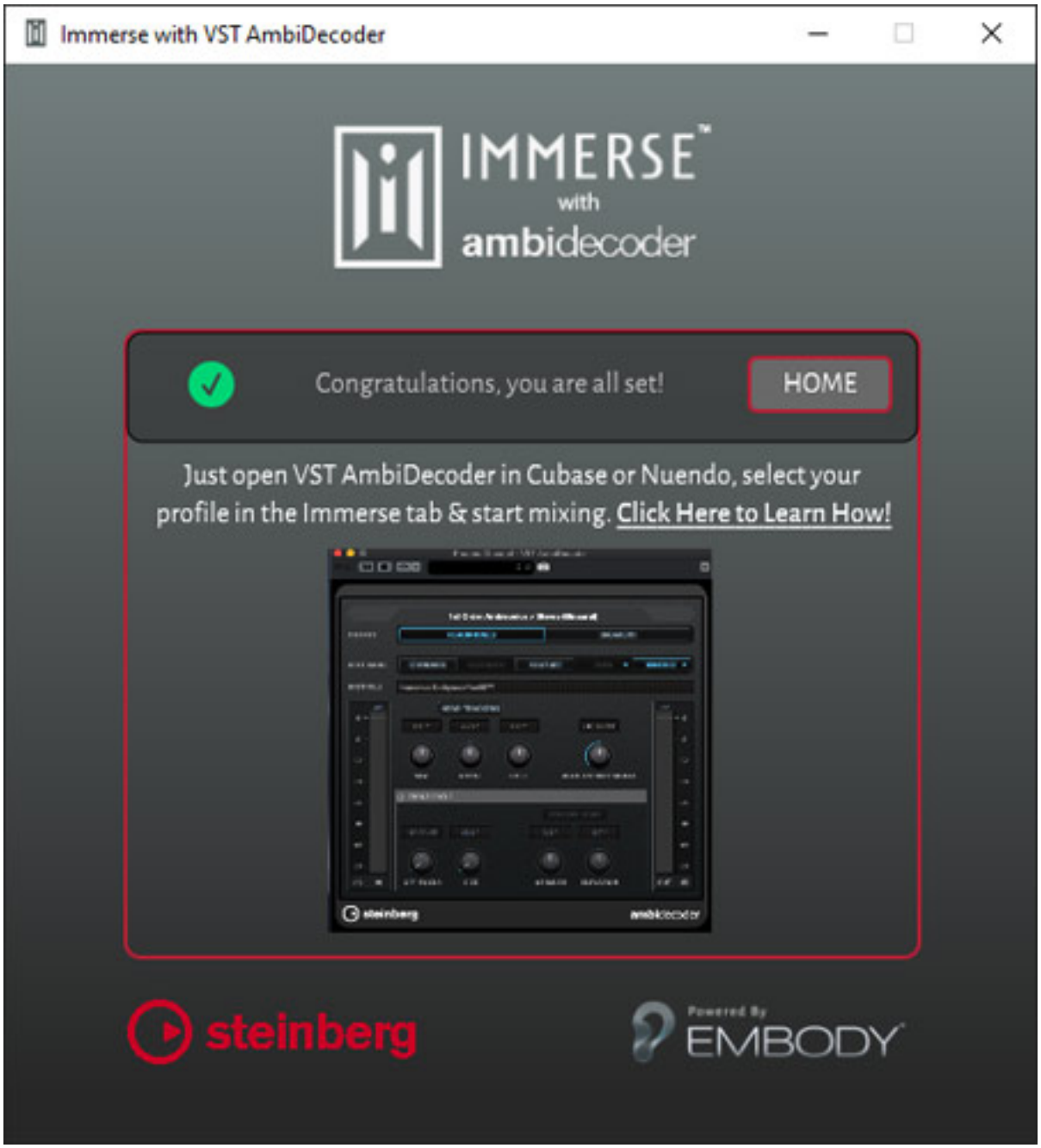


Wenn die Bearbeitung nach einigen Sekunden abgeschlossen ist gibt es einen Hinweis (s. Abb. oben).

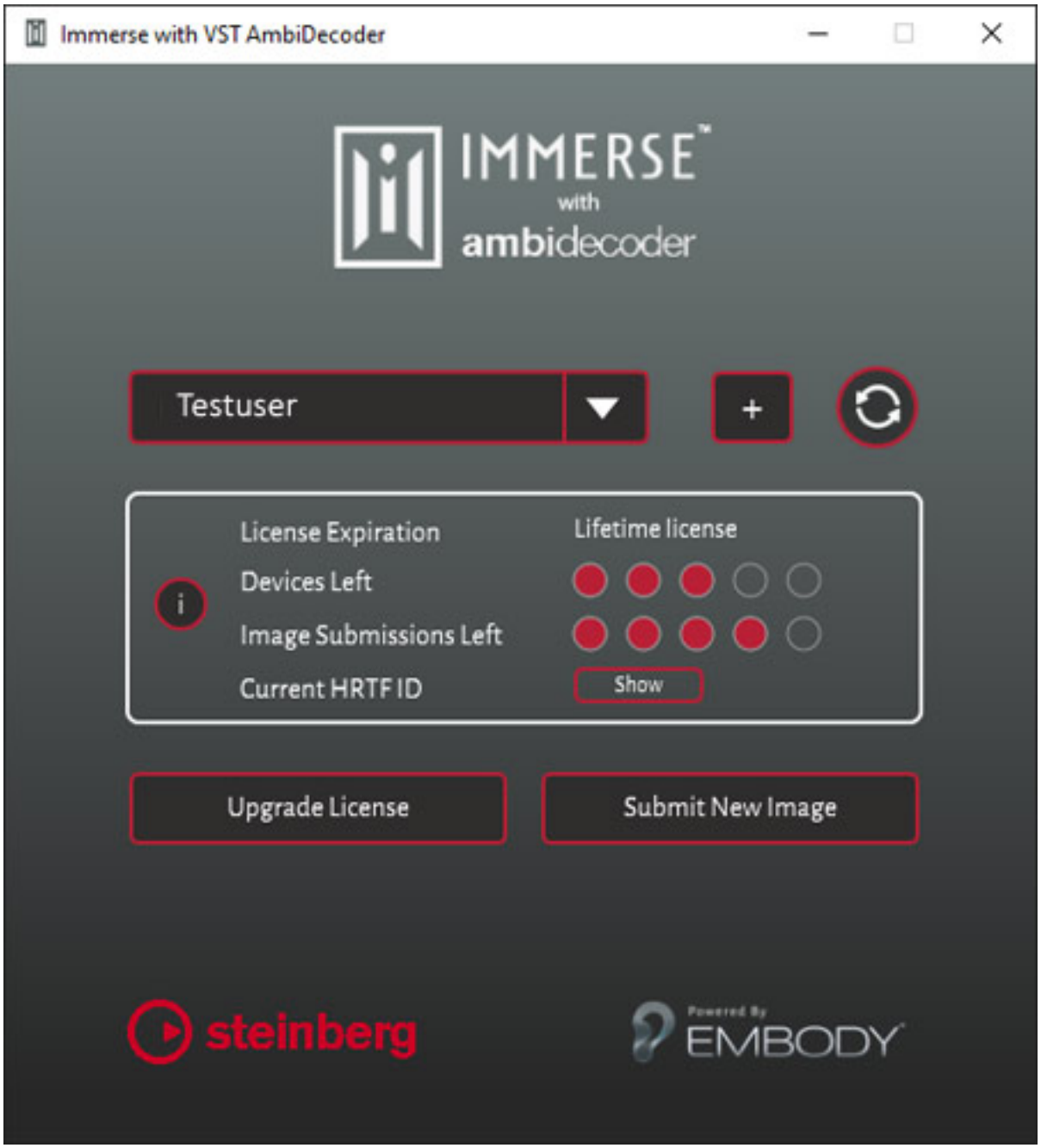


In der Softwaren wird dann folgender Screen angezeigt und man muss ihn mit "DONE" bestätigen.

Phoca PDF



Es wird dann der Abschluss der HRTF-Ermittlung mitgeteilt (s. Abb. oben).



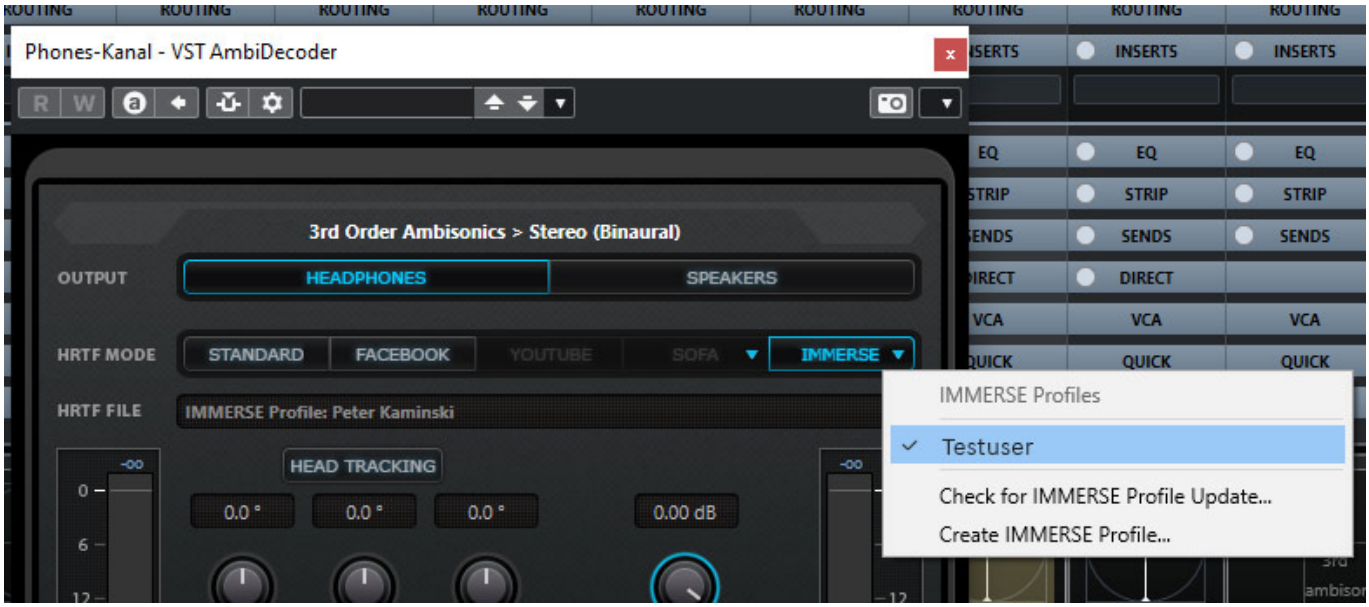
Nach Anklicken von Home sieht man den Verwaltungsdialog wo die Anzahl der noch freien Geräte und HRTFs, bzw. Personen sichtbar ist. Wie man auf diesem Beispiel sieht wurden zwei Geräte genutzt und eine HRTF. Über diesen Dialog kann man auch später noch weitere HRTFs für andere Benutzer auf den benutzen Workstations anlegen.

Bedienung

Phoca PDF



Der VST AmbiDecoder wird in Nuendo als Plug-In im Control Room eingebunden. Dazu muss das Plug-In (s. Abb. oben) in einem Kopfhörerpfad eingefügt sein.



Nach dem Öffnen des Plug in muss man als erstes den HRTF-Modus auswählen. Hier stehen einem mehrere Optionen bereit. STANDARD ist eine HRTF auf Basis des Neumann Kunstkopfes KU 100 und FACEBOOK und YOUTUBE die jeweilige HRTF, wobei die YouTube-HRTF, wie von YouTube spezifiziert, nur Ambisonics erster Ordnung unterstützt. Um die zuvor mit der App angelegte HRTF zu nutzen muss man IMMERSE anwählen und den entsprechenden Anwender auswählen (s. Abb. oben). Alternativ kann man auch eine eigene HRTF über das Importieren einer SOFA-Datei nutzen.



Es ist auch möglich den AmbiDecoder mit einem Headtracker zu koppeln um bei Kopfbewegung die binaurale Ausgabe entsprechend der Kopfposition zu korrigieren. Wenn des Headtracking-Device erkannt wurde, wird dies durch einer Aktivierung des FOLLOW HEAD Buttons symbolisiert. Es lassen sich auch verschiedene Parameter einstellen und das Headtracking kann auch über ein virtueller Taster deaktiviert werden.



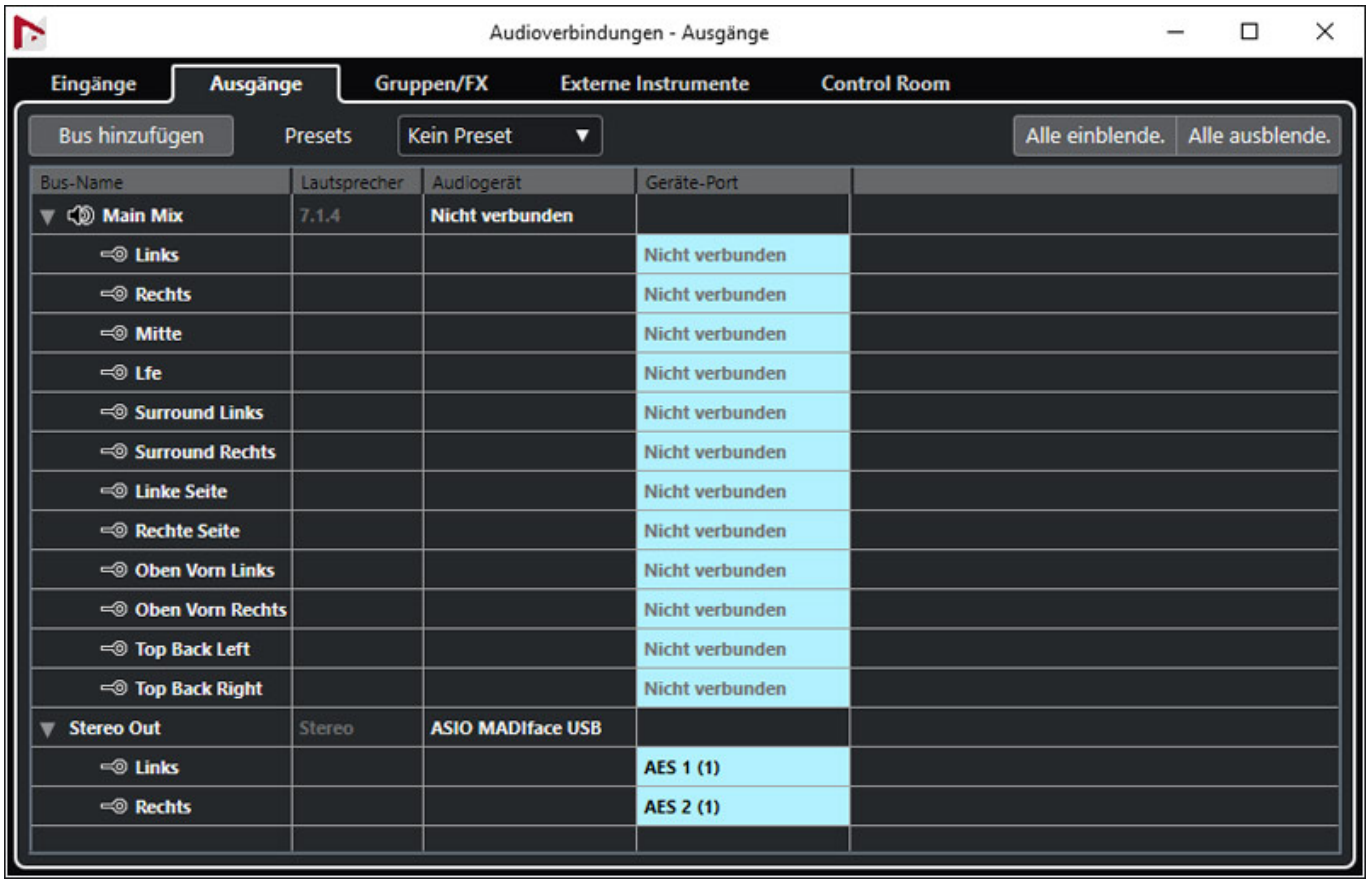
Übrigens gibt es oben ikm Kopf des Plug-Ins eine Umschaltmöglichkeit zwischen binauralem abhören über Kopfhörer und Lautsprecherwiedergabe. Hier wird dann bei Anwahl von Lautsprecherwiedergabe ein nicht-binauraler Stereo-Downmix ausgegeben. Das ist extrem nützlich denn binauralisiertes Audio über Lautsprecher ist auch als Kompromiss nicht brauchbar und da ist der Anwender mit einem klassischen Stereo-Downmix deutlich besser bedient.

Die Ortungsinformationen vorne/hinten und oben/unten gehen dabei zwar verloren, aber dafür ist der Klang deutlich angenehmer. Man möchte ja nicht immer binaural

abhören, bzw. braucht man die 3D-Ortung ja auch nicht in allen Produktionsprozessstufen, wie gegebenenfalls beim Aufnehmen von Audio oder beim Einstellen von Plug-Ins in einem Kanal etc. Bei langen Session tut es ja auch mal gut ohne Kopfhörer zu hören, denn dies ist auf Dauer anstrengend.

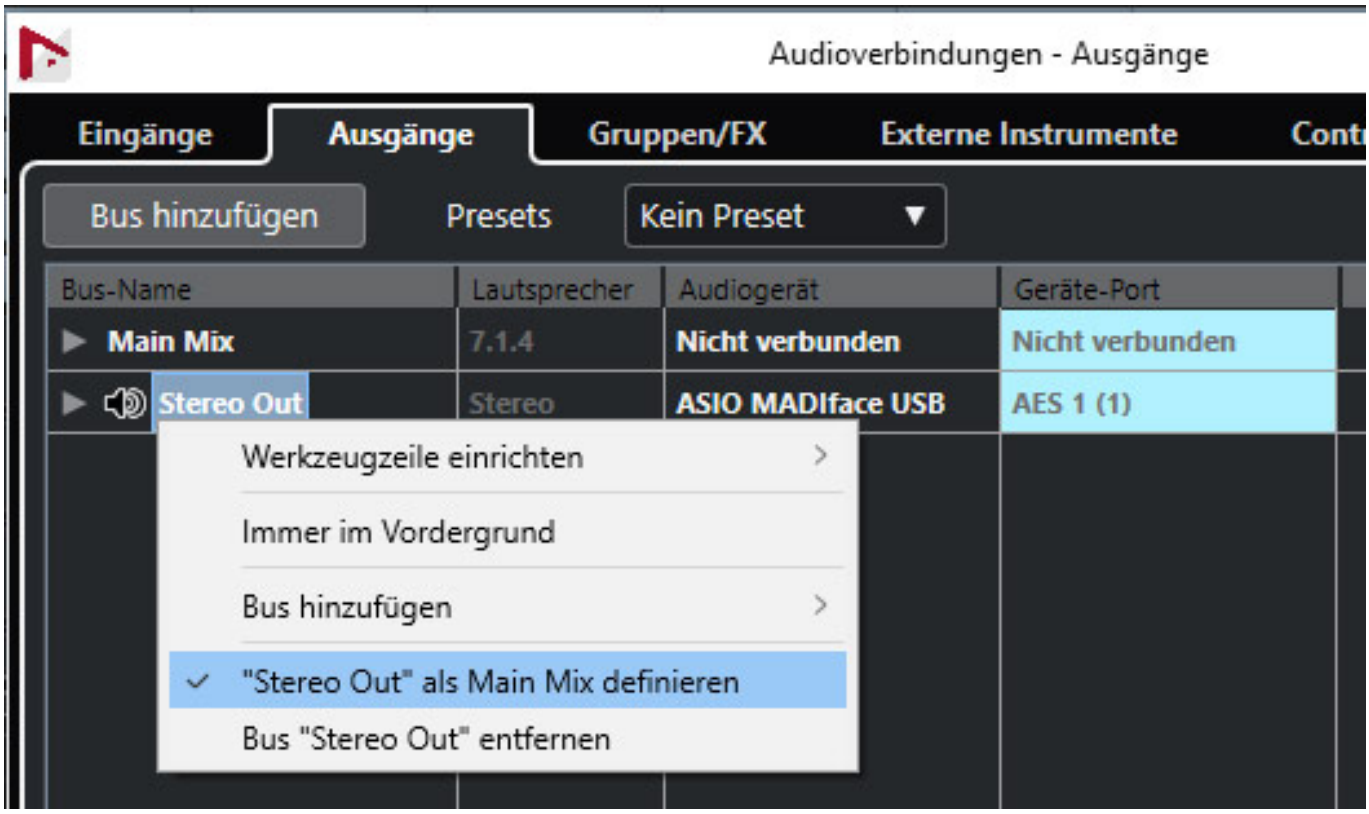
AmbiDecoder für Dolby Atmos

Der AmbiDecoder lässt sich nicht nur für Ambisonics einsetzen sondern auch für das binaurale Monitoring von Dolby Atmos. Wie das geht wenn man ausschließlich binaural abhören möchte, werden wir einmal an dieser Stelle genauer aufzeigen, da dies so explizit nicht in der Bedienungsanleitung beschrieben ist.

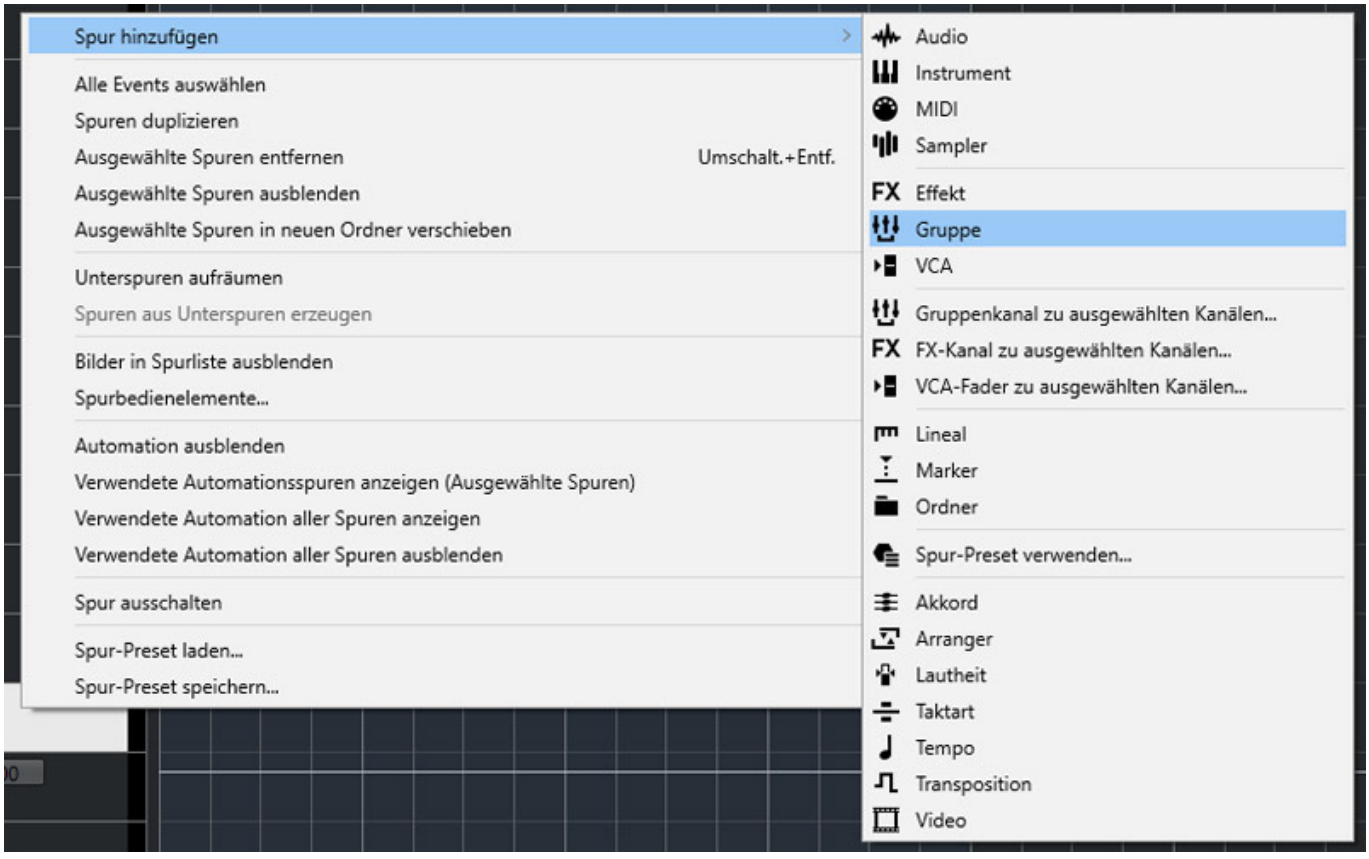


Man muss dazu als erstes bei den Audioverbindungen einen Stereoausgang einrichten, falls man diesen noch nicht eingerichtet, bzw., ggf. gelöscht hat (s. Abb. oben).

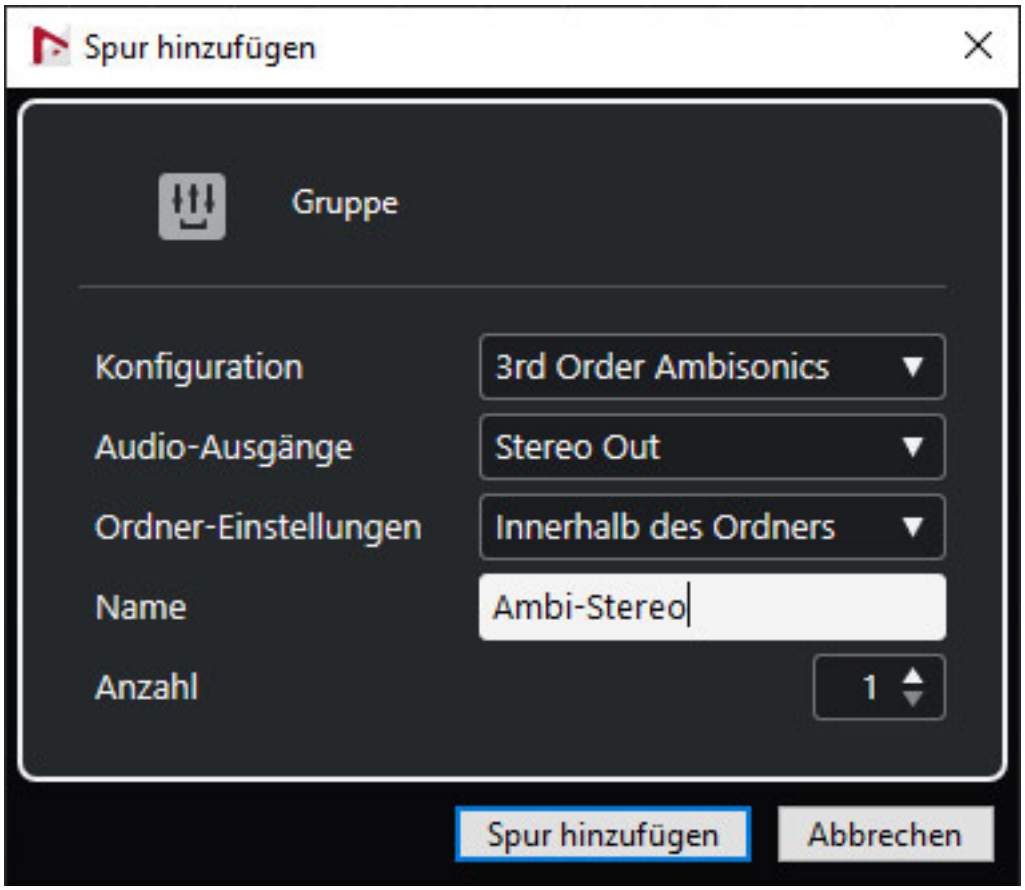
Phoca PDF



Wichtig ist, dass man diesen als Main Mix deklariert (s. Abb. oben).



Nun legt man eine neue Gruppenspur an (s. Abb. oben).



Diese sollte als Ambisonics dritter Ordnung konfiguriert werden und dem Stereo Out als Ausgang zugewiesen sein (s. Abb. oben).

Phoca PDF



Beim vorhandene 7.1.4 Bus in dem der Dolby Atmos Renderer eingebunden ist, wird ein Pre-Fader Insert eingerichtet (s. Abb. oben) und aktiviert.

Phoca PDF



Wenn das passiert ist, kann man den AmbiDecoder aufrufen. Dieser ist nun von Nuendo automatisch in der Ambisonics auf Stereo Gruppe eingeschliffen. Um ihn aufzurufen muss man ein Doppelklick auf den Panner-Bereich (siehe roten Pfeil) des entsprechenden Kanals durchführen. Am Stereoausgang lässt sich nun Dolby Atmos binauralisiert über Kopfhörer abhören. Ggf. muss man bei Bedarf noch im Control Room ein Kopfhörerausgang einrichten (über Audioverbindungen > Control Room).

Praxis

Für das Erstellen des optimalen Fotos von dem Ohr dürften mehrere Versucher erforderlich sein, zumindest war es bei unseren Testpersonen so. Das ist aber kein Problem, denn man kann das Ganze ja beliebig oft wiederholen.

Hier noch einige Anmerkung. Bei einer Deinstallation von AmbiDecoder und erneutem Installieren wird auch auf dem gleichen Rechner ein Device heruntergezählt. Das sollte man also vermeiden. Weiter findet man die AmbiDecoder-Applikation zur Verwaltung der Devices und User bei Windows-Betriebssystem im Programm-Menü nicht unter Steinberg oder AmbiDecoder sondern unter dem Eintrag "Embody". Da sucht man unter Umständen erst mal wenn man viele Programme auf der DAW installiert hat.

Anzumerken ist auch noch, dass sich das Ergebnis was man abhört als Datei über Menüs und Dialog File > Export > Audio Mixdown exportieren lässt. Dabei sollte man sich genau überlegen welche HRTF man nutzt. Ggf. eignet sich hier eine standardisierte HRTF, wie die vom Facebook, besser als die eigene HRTF.

Jetzt stellt sich die Frage nach dem Klang. Hier muss man deutlich sagen, dass die Wahrnehmung nicht nur subjektiv, sondern in diesem Fall von Hörer zu Hörer auch sehr unterschiedlich sein kann. Das Gesagte bezieht sich also auf meinen persönlichen Eindruck. Dazu muss ich sagen, dass ich schon sehr lange mit HRTFs experimentiere und dabei festgestellt habe, dass bei Abweichungen von meiner eigenen HRTF ich sofort Klangfärbungen und Ortungsschwierigkeiten wahrnehme. Es gibt Personen, deren Gehör da toleranter ist.

Bei Anwahl von "Standard" (Neumann KU100) kann man schon die Richtung der Schallquelle gut orten. Der Klang ist schon gefärbt, besonders wenn sich die virtuelle Schallquelle vorne befindet, wo für mich der Sound leicht nasal klingt. Die Anwahl der Facebook-HRTF führt zu einem besseren allgemeinen Klang aber dafür ist die Vorne/Hinten und Oben/Unten Ortung deutlich eingeschränkter möglich. Bei Anwahl des von mir erstellten Embody-Profils ergibt sich gegenüber den beiden Standard-HRTFs beim Monitoring über Kopfhörer eine deutliche Verbesserung was Klang und Ortung angeht. Hier kann man auf jeden Fall auch schon eine Vormischung durchführen oder ein Binauralen-Mix erstellen, wobei ich, wie schon erwähnt, für eine binaurale Veröffentlichung von gemischtem Material eher eine der beiden Standard-HRTFs nutzen würde.

Was auffällt ist, dass die Daten der HRTF bei einem SOFA-Import nicht normalisiert werden. So ist der Ausgangspegel bei Verwendung von externen HRTFs zum Teil im Pegel sehr unterschiedlich. Wenn man denn nur seine eigene zu Monitorzwecken geladen hat und nutzen möchte ist dieses aber in der Praxis nicht relevant.

Ich habe auch individuelle SOFA-Files mit meiner eigenen HRTF geladen. Hierzu muss man sagen, dass es hier unter Nuendo-Versionen vor der 11.0.20 (die voraussichtlich Ende April 2021 erscheinen wird) Probleme mit dem Laden von SOFA-Dateien gab, die aber mit dieser Version behoben sind. Wir hatten den Test schon mit einer Betaversion der 11.0.20 durchgeführt und da lief alles einwandfrei.

Steinberg wird in dem nächsten Update von AmbiDecoder auch noch toleranter gegenüber falschen Metadatenbeschreibungen in den SOFA-Dateien sein. Leider kommt auch das vor, wie wir beim Test mit externen HRTFs feststellen mussten.

Nicht alle die SOFA-Dateien generieren, halten sich auch an den Standard und da kann es beim Laden zu Fehlinterpretationen oder eben Fehlern kommen, so dass die SOFA-Datei nicht nutzbar ist. Hier wird Steinberg einige Dinge implementieren, um die Dateien dann trotzdem nutzen zu können, zumindest bei bekannten Problemen mit SOFA-Dateien - sehr lobenswert denn das eigentliche Problem liegt ja nicht beim AmbiDecoder oder Nuendo sondern bei den Anbietern die ihre SOFA-Dateien nicht dem Standard gemäß generieren.

Grundsätzlich ist es so, dass man mit Verfahren, die deutlich aufwendiger die Kopfgeometrie vermessen oder eben über akustischer Messungen am eigenen Kopf, eine HRTF erzeugt werden kann, die noch bessere Klang- und Ortungsergebnisse erzielt. Mich hat aber auch schon erstaunt, wie gut das von Embody erstellte HRTF-Profil beim AmbiDecoder lediglich über ein Foto eines Ohrs, funktioniert. Wer einen noch besseren Klang und Ortung haben will, der muss halt deutlich tiefer in die Tasche greifen und sich eine individuelle HRTF über aufwendigere Verfahren erzeugen lassen. Mit der Methode lediglich ein Ohr zu fotografieren, und daraus alle Geometrien den Kopfes und des Ohrs abzuleiten, wird man meiner Meinung nach kaum bessere Ergebnisse erzielen können, als mit dem Verfahren von Embody.

Fazit

Der Preis für den VST AmbiDecoder liegt bei knapp unter 190 Euro. Es gibt auch eine Möglichkeit AmbiDecoder für den Zeitraum von 14 Tagen zu testen. Der Preis ist, wenn man auch noch einbezieht, dass man dafür mehrere individuelle HRTF-Profile bekommt, absolut angemessen. Ein aufwendige individuelle Messung oder andere aufwendigere Verfahren kosten pro Person in der Regel ein Vielfaches.

Wenn man Ambisonics- oder Dolby Atmos-Produktion mit Hilfe von AmbiDecoder und einem individuellen Embody-Profil in Nuendo mischt, sollte man das Endprodukt auf jeden Fall noch einmal in einem Studio mit Immersive-Lautsprecher-Setup abhören und ggf. die Mischung dann nachkorrigieren. Das ist aber eher eine grundsätzliche Empfehlung beim binauralen Mischen von Immersive-Audio-Produktionen, egal welche HRTF oder welches HRTF-Verfahren man nutzt. Das Monitoring von 3D-Audio über ein Immersive-Lautsprecher-Setup hat immer Vorteile im Bereich der Lokalisierung und des Klangs gegenüber einem binauralen Kopfhörer-Monitoring.

AmbiDecoder für Nuendo und Cubase Pro von Steinberg ist auf jeden Fall ein empfehlenswertes Werkzeug, um sich an immersiven Audioprojekte heranzutasten, ohne zum Beispiel ein 7.1.4-Lautsprecher-Setp aufbauen zu müssen. Durch die Integration von Dolby Atmos in Nuendo 11 und AmbiDecoder kommt jetzt jeder Nuendo-Anwender in den Genuss, mit 3D-Audio-Projekte zu arbeiten, bzw., diese zu erstellen. Ich kann jedem Tonschaffenden nur empfehlen, sich mit diesem Thema auseinander zu setzen. Immersives Audio ist die Zukunft.

www.steinberg.net