

Über dieses Dokument

Inhalt

Das vorliegende Dokument ist die zweite von vier Application Notes über die Durchführung von Hörversuchen. Es beschreibt die gängigsten Testarten, die zur Durchführung von Hörversuchen verwendet werden. In einer Zusammenfassung sind Vor- und Nachteile der unterschiedlichen Testarten aufgelistet.

1. Ranking _____ 2
2. Paarvergleich _____ 3
3. Semantisches Differenzial _____ 5
4. Kategoriale Bewertung _____ 6
5. Simultane kategoriale Bewertung mit Referenz _____ 7
6. Zusammenfassung der Vor- und Nachteile der verschiedenen Testarten __ 9
7. Realisierung von Hörversuchen mit individuellen Testdesigns _____ 10

Zielgruppe

Der nachfolgende Text wendet sich insbesondere an (potenzielle) Anwender¹ des ArtemiS SUITE Jury Testing Module SQala, die sich über die in dieser Software verfügbaren Testarten informieren möchten.

Fragen?

Sie haben Fragen? Wir freuen uns über Ihre Rückmeldungen!

Fragen zum Inhalt dieses Dokument: Imke.Hauswirth@head-acoustics.com

Technische Fragen zu unseren Produkten: SVP-Support@head-acoustics.com

Hörversuche durchführen – Teil 2

Die im folgenden beschriebenen Tests stellen die gängigsten, etablierten Testarten im Bereiche Geräuschqualität und Benchmarking dar.

Alle diese Testarten können mit dem ArtemiS SUITE Jury Testing Module SQala durchgeführt werden. Für die Durchführung von anderen Testarten kann die API-Schnittstelle ASX03 verwendet werden.



¹ Im nachfolgenden Text wird verallgemeinernd das generische Maskulinum verwendet. Dies soll ausschließlich der besseren Lesbarkeit dienen. Selbstverständlich möchten wir gleichermaßen alle Geschlechter ansprechen bzw. einbeziehen.

1. Ranking

Rangreihenfolge bestimmen

Beim Ranking wird der Teilnehmer gebeten, N Geräusche entsprechend einem Kriterium (z. B. Lästigkeit) in eine Rangreihenfolge von 1 bis N einzuteilen. Diese Aufgabe wird schwieriger, je mehr Geräusche zu sortieren sind. Aus diesem Grund sollte für ein solches Ranking nicht zu viele Geräusche angeboten werden. In den meisten Fällen können Teilnehmer bis zu sechs Geräusche verlässlich² in eine Rangreihenfolge bringen.

Abbildung 1 zeigt eine Möglichkeit, wie die Benutzeroberfläche für einen Ranking-Versuch aussehen kann. Durch Anklicken der Wiedergabe-Schaltflächen kann der Teilnehmer das Geräusch abspielen und diese z. B. per Drag & Drop in der Rangfolge verändern.



Abbildung 1: Ranking-Hörversuch in SQala

Rangreihenfolge ohne Information über Abstände

Mit dieser Testmethode kann auf einfache und unkomplizierte Weise der erste Eindruck, z. B. die Kunden-Präferenz, geprüft werden. Ein Rankingtest kann innerhalb sehr kurzer Zeit durchgeführt werden. Außerdem kann sich der Teilnehmer alle Geräusche direkt nacheinander anhören und vergleichen. Ein Nachteil dieser Bewertungsmethode ist, dass der Versuchsleiter ausschließlich Informationen über die Reihenfolge nicht aber über die Abstände zwischen den Geräuschen erhält. Es kann z. B. keine Aussage darüber gemacht werden, ob der Abstand zwischen dem ersten Rang und dem zweiten Rang genauso groß ist wie der zwischen dem fünften und sechsten Rang.

² Verlässlich heißt in diesem Zusammenhang, dass bei einer Wiederholung des Hörversuchs der Teilnehmer wieder dieselbe Rangreihenfolge mit nur geringen Abweichungen wählt.

2. Paarvergleich

Hörversuch mit
Paarvergleichen

Beim Paarvergleich werden den Teilnehmern jeweils zwei Geräusche nacheinander dargeboten. Jeder Teilnehmer muss diese entsprechend einem vorgegebenen Kriterium (z.B. Lautheit) bewerten. Dazu stehen ihm zwei bzw. drei Urteilmöglichkeiten zur Verfügung:

A > B	• Beispiel: A lauter als B
B > A	• Beispiel: B lauter als A
A = B	• Beispiel: A und B gleich laut

Abbildung 2 zeigt ein Beispiel für die SQala-Benutzeroberfläche bei einem Paarvergleich. Mit den oberen Schaltflächen kann die Wiedergabe der Geräusche gestartet werden und die Schaltflächen darunter dienen der Urteilsabgabe.

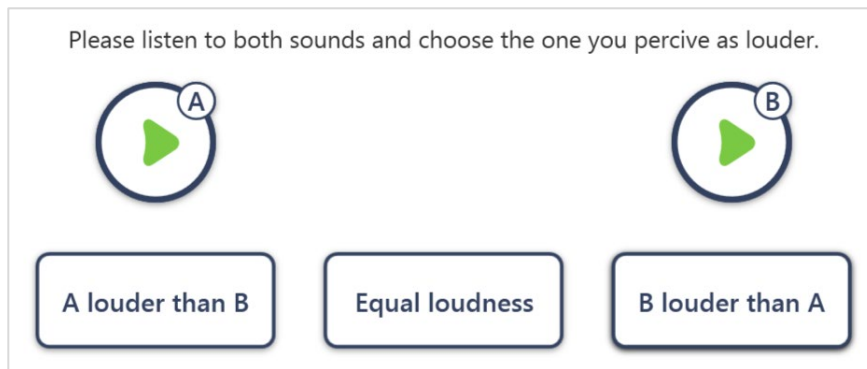


Abbildung 2: Paarvergleich-Hörversuch in SQala

Forced Choice

Ein unsicherer Teilnehmer kann dazu neigen, der Entscheidung auszuweichen und eventuell häufig die Antwortmöglichkeit A = B auswählen, falls diese angeboten wird. Dies kann die Aussagekraft des Tests reduzieren. Um dies zu vermeiden, kann die Testvariante „forced choice“ eingesetzt werden. Bei dieser Testvariante wird eine Entscheidung erzwungen, indem nur die beiden Antwortmöglichkeiten A > B und B > A angeboten werden. Es muss allerdings beachtet werden, dass diese Variante den Teilnehmer unter Druck setzt, weil auch dann eine Entscheidung erzwungen wird, wenn der Teilnehmer keinen Unterschied hört. Der Teilnehmer muss in diesem Fall ein Urteil abgeben, das nicht seiner Wahrnehmung entspricht. Durch eine geeignete Testeinführung und Instruktion können beide Effekte reduziert werden.

Anwendungsbereich für den
Paarvergleichstest

Der Paarvergleichstest eignet sich, um kleine Unterschiede in sehr ähnlichen Geräuschen aufzuspüren. Das menschliche Gehör ist z. B. in der Lage, den Schalldruckpegel im Kurzzeitgedächtnis festzuhalten, so dass in einem Paarvergleich mit schnell aufeinander folgenden Geräuscharbietungen auch geringe Pegeländerungen aufgespürt werden können.

Im Alltag, z. B. bei der Bewertung von Fahrzeuginnengeräuschen, besteht die Möglichkeit eines direkten Vergleiches jedoch nicht. Diese Geräusche können nur nacheinander, mit einem gewissen zeitlichen Abstand voneinander bewertet werden und das akustische Langzeitgedächtnis des Menschen speichert in erster Linie

Geräuschmuster ab. Aus diesem Grund wird bei Geräuschen, die nicht direkt aufeinander folgend dargeboten werden, weniger der absolute Pegel beurteilt, sondern die Geräuschcharakteristik basierend auf den enthaltenen Geräuschmustern. Vor der Durchführung eines Hörversuchs muss also die Zielsetzung festgestellt werden:



- Ist es wichtig, kleine Unterschiede in Geräuschen zu finden?
- Oder sollen Geräusche bewertet werden, die in der Praxis nicht direkt verglichen werden können?

Auf der Basis dieser Entscheidung kann dann beschlossen werden, ob die Methode des Paarvergleichs geeignet ist.

Testdauer beim Paarvergleichstest

Bei einem Hörversuch mit Paarvergleichen sollte der Versuchsleiter darauf achten, nicht zu viele Geräusche bewerten zu lassen. Bei einer großen Geräuschanzahl steigt die Testdauer bedingt durch die vielen Paarkombinationsmöglichkeiten sehr stark an. Dies ist insbesondere dann der Fall, wenn sowohl die Paarung A – B als auch die Paarung B – A getestet werden soll, um mögliche Reihenfolgeeffekte zu untersuchen. Mit SQala besteht die Möglichkeit in den Testeigenschaften einzustellen, ob die volle Matrix (A – B und B – A) oder nur die halbe Matrix (A – B oder B – A) beurteilt werden soll.

Auswertung der Ergebnisse

Bei einem Paarvergleich erhält der Versuchsleiter für jedes Geräusch eine Angabe darüber, wie oft dieses anderen Geräuschen vorgezogen wurde. Bei der Auswertung der Ergebnisse des Paarvergleichshörversuchs kann somit eine Rangreihenfolge der Geräusche bestimmt werden. Eine Aussage darüber wieviel besser ein Geräusch im Vergleich zu den anderen ist, kann nicht ohne weiteres getroffen werden. Es existieren aber statistische Methoden, um die Ergebnisse entsprechend auszuwerten und so auch quantitative Aussagen zu treffen.

3. Semantisches Differenzial

Beurteilung mit semantischem Differenzialen

Die Verwendung eines semantischen Differenzials zur Beurteilung ermöglicht eine sehr differenzierte Untersuchung der Geräuschbeispiele. Während sich die Teilnehmer bei den oben beschriebenen Testmethoden auf ein vorgegebenes Beurteilungskriterium konzentrieren soll, können bei dieser Testmethode mehrere Attribute eines Geräusches untersucht werden. Der Teilnehmer bewertet das dargebotene Geräusch auf mehreren bipolaren Skalen, deren Enden mit einem Adjektiv und dessen Antonym (Gegenteil) beschriftet sind. Die verwendeten Skalen sind häufig sieben- oder neunstufig.

In Abbildung 5 ist ein Beispiel für ein semantisches Differenzial mit einer siebenstufigen Skala und 4 Attributpaaren dargestellt.

Please listen to the sound and evaluate the following sound attributes:

cheap	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	expensive
sporty	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unsporty
steady	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	uneven
unpleasant	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	pleasant

Abbildung 3: Beispiel für ein semantisches Differenzial unter Verwendung der SQuala-Benutzeroberfläche

Anwendungsbereich für das semantische Differenzial

Mit Hilfe dieser Testmethode erhält der Versuchsleiter ein detailliertes Geräuschprofil, das weit mehr Informationen beinhaltet als die Aussage, dass ein Geräusch dem anderen vorgezogen wird und in welchem Ausmaß. Durch die Bewertung auf mehreren Skalen ist es einfacher, Korrelationen mit den Ergebnissen von messtechnischen Analysen zu finden. Dies ermöglicht es herauszufinden, warum ein Geräusch abgewertet wurde und welcher Aspekt eines Geräusches verändert werden muss, um die Geräuschqualität zu verbessern.

Die Beurteilung eines Geräusches mit einem semantischen Differenzial ist zeitaufwendiger als andere Testmethoden. Die Anzahl der Geräusche sowie die Anzahl der Beurteilungskriterien darf nicht zu groß gewählt werden, da sonst die Konzentration der Teilnehmer zum Ende des Hörversuchs nachlässt. Es hat sich gezeigt, dass die Zahl der Attributpaare 8-12 nicht überschreiten sollte.

Auswahl der Attributpaare

Bei der Auswahl der Attributpaare ist aus mehreren Gründen große Sorgfalt nötig. Wenn für alle Geräusche innerhalb eines Hörversuchs dieselben Beurteilungen abgegeben wurden, kann es sein, dass für den Test unpassende Attribute gewählt wurden.

Außerdem ist darauf zu achten, dass die Attributpaare unterschiedliche Aspekte des Geräusches abfragen. Ist dies nicht der Fall korrelieren die Antworten der unterschiedlichen Attributpaare gegebenenfalls stark. In einem solchen Fall wäre vielleicht die Abfrage nur eines dieser Attribute (z.B. in einem Kategorialtest) ausreichend gewesen, da die Abfrage der anderen stark korrelierenden Attributpaare nur wenig neue Informationen ergibt. Weiterhin beeinflusst die Auswahl des Antonyms die Beurteilung. Das folgende Beispiel soll dies verdeutlichen: Zu dem Attribut „alt“ kann das Antonym „jung“ ebenso wie „neu“ verwendet werden. In vielen Fällen wird die bipolare Skala „alt – jung“ zu anderen Ergebnissen führen als die Skala „alt – neu“. Die Zielsetzung des Hörversuchs muss also bei der Auswahl der Antonyme berücksichtigt werden.

Internationale Hörversuche

Bei der Durchführung von Hörversuchen mit semantischen Differenzialen, an denen Teilnehmer aus verschiedenen Ländern teilnehmen, empfiehlt es sich, jeden Teilnehmer in seiner Muttersprache zu befragen. Um die Ergebnisse der verschiedenen Teilnehmer vergleichen zu können, muss beachtet werden, dass die verschiedenen Attributpaare in der Übersetzung auch tatsächlich dasselbe bedeuten.

4. Kategoriale Bewertung

*Hörversuch mit
Kategorialbewertung*

Bei der kategorialen Bewertung wird dem Teilnehmer jeweils ein Geräusch vorgespielt, das entsprechend einem vorgegebenen Kriterium auf einer mehrstufigen Skala bewertet werden soll. Im DEGA-Kompendium zur Durchführung von Hörversuchen wird z. B. für die Beurteilung der Intensität die folgende fünfstufige Skala empfohlen:

gar nicht – kaum – mittelmäßig – ziemlich – außerordentlich

In Abbildung 3 ist ein Beispiel für die Bewertung der Schärfe mit einer Kategorialskala dargestellt.

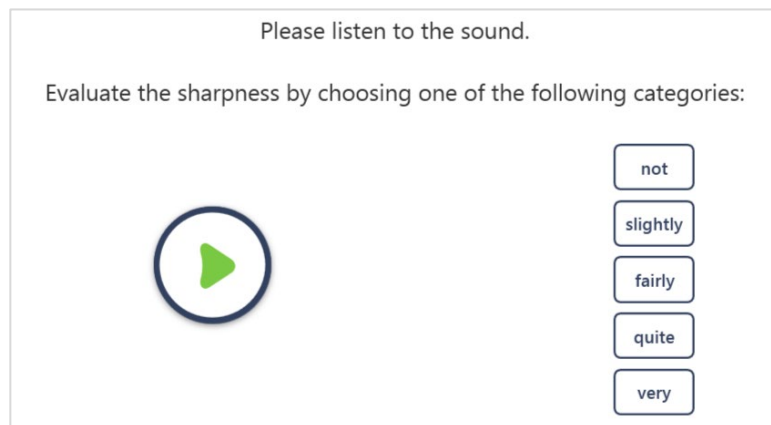


Abbildung 4: Beispiel für eine Kategorien-Skala erzeugt mit SQala

Intervallskalierte Daten

Bei der Erstellung einer Kategorialskala muss darauf geachtet werden, dass die einzelnen Kategorien als äquidistant verstanden werden können. Nur so erhält der Versuchsleiter intervallskalierte Daten, die es ihm ermöglichen, den einzelnen Kategorien entsprechende Zahlenwerte zuzuordnen. Die Übertragung der Beurteilungen durch die Teilnehmer in Zahlenwerte erlauben es, im weiteren Verlauf, z. B. eine

Korrelationsanalyse mit den Ergebnissen von messtechnischen Analysen durchzuführen.

Verzerrungseffekte und Gegenmaßnahmen

Bei der Bewertung auf einer Kategorienskala kann es zu unterschiedlichen Verzerrungen kommen:

- Ein Geräusch, das nachfolgend auf ein als besonders scharf empfundenen Geräusch beurteilt werden soll, wird vielleicht anders bewertet, als wenn es direkt nach einem nicht scharfen Geräusch beurteilt wird. Die Beurteilung eines Geräusches wird also eventuell durch das vorhergehende Geräusch beeinflusst. Dieser Effekt wird als Kontext-Effekt bezeichnet und kann durch wiederholte Beurteilung der Geräusche in randomisierter Reihenfolge umgangen werden.
- Außerdem verzichten die Teilnehmer häufig auf die Verwendung der Extreme. Stattdessen tendiert sie zu Antworten im zentralen Bereich der Skala. Die Teilnehmer möchte so verhindern, dass sie im Testverlauf von einem Geräusch extremer Ausprägung überrascht wird, für das aber dann keine Kategorie mehr zur Verfügung steht. Dieser Effekt kann durch ein ausreichendes Training verhindert werden. Während dieses Trainings können den Teilnehmern die Geräusche mit maximaler Ausprägung vorgestellt werden, so dass diese bekannt sind und entsprechend bewertet werden können.
- Zudem wird die Skala unter Umständen von den Teilnehmern unterschiedlich ausgenutzt. Dies kann durch eine nachträgliche Normalisierung (Bezug der Daten auf den Mittelwert) ausgeglichen werden.

5. Simultane kategoriale Bewertung mit Referenz

Hörversuch mit simultaner Kategorialbewertung

Bei der simultanen Kategorialbewertung kann sich der Versuchsteilnehmer mehrere Geräusche anhören und vergleichen. Die Beurteilung wird dann für jedes Geräusch auf einer separaten Kategorienskala abgegeben. Weiterhin besteht bei diesem Hörversuch die Möglichkeit, ein Referenzgeräusch anzubieten, das eine bestimmte Ausprägung des zu bewertenden Geräuschattribut aufweist.

In Abbildung 4 ist ein Beispiel für eine simultanen Kategorialbewertung der Lautheit von fünf Geräuschen dargestellt. Darüber hinaus hat der Teilnehmer die Möglichkeit die Geräusche mit einer Referenz zu vergleichen, die im Vergleich zu den anderen Geräuschen die höchste Lautheit aufweist.

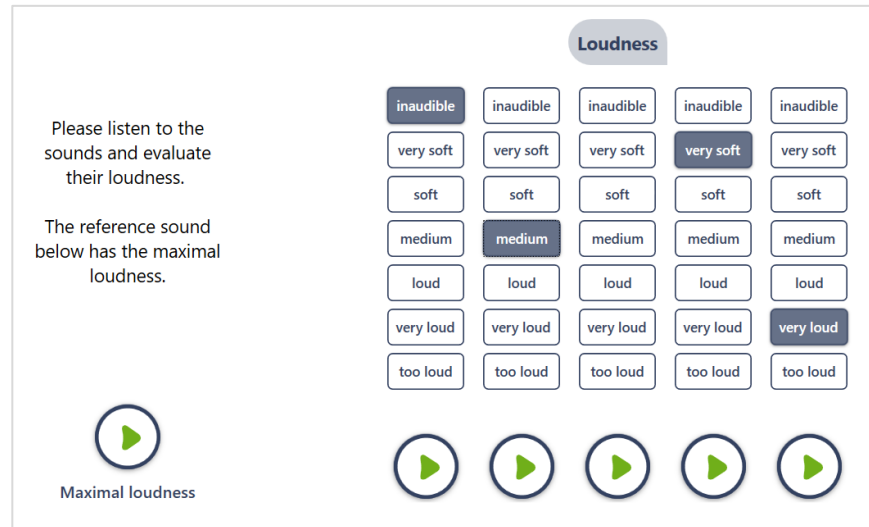


Abbildung 5: Beispiel für einen Hörversuch mit simultaner kategorialer Bewertung mit SQala

Auswertung der Hörversuchsergebnisse

Wurde für die Beurteilung eine Kategorienskala mit äquidistanten Kategorien verwendet, können die Hörversuchsurteile in Zahlenwerte umgewandelt und z. B. für eine Korrelationsanalyse mit messtechnischen Analysen genutzt werden.

Anwendungsbereich für die simultane Kategorialbewertung

Der Vorteil der simultanen Kategorialbewertung liegt in der Möglichkeit, Geräusche vor der Beurteilung mit einem Referenzgeräusch und mit anderen zu bewertenden Geräuschen direkt zu vergleichen. Auf diese Weise können auch sehr spezielle Geräuschphänomene, wie Wummern oder Getrieberasseln, untersucht werden. Insbesondere die Möglichkeit ein Referenzgeräusch immer wieder anzuhören, erlaubt es auch Nicht-Experten, die mit dem Geräuschphänomen eigentlich nicht vertraut sind, dieses dennoch zu bewerten.

Wie beim Paarvergleichstest wird allerdings durch den direkten Vergleich eine Versuchssituation geschaffen, die nicht unbedingt mit der Alltagssituation vergleichbar ist. Im Alltag kann häufig kein direkter Vergleich stattfinden. Vor der Durchführung eines Hörversuchs muss also die Zielsetzung festgelegt werden. Dann lässt sich beurteilen, ob die simultane Kategorialbewertung die geeignetste Methode für die zu untersuchenden Geräusche und die teilnehmenden Personen ist.

6. Zusammenfassung der Vor- und Nachteile der verschiedenen Testarten



Vor- und Nachteile
Rangreihenfolge



Rangreihenfolge

Vorteile: <ul style="list-style-type: none">• Intuitive, unkomplizierte Beurteilung• Die Beurteilung ist in den meisten Fällen sehr zuverlässig.• Schnelle Durchführung	Nachteile: <ul style="list-style-type: none">• Nur Rangfolge, keine Aussage zu den Abständen der einzelnen Positionen möglich
--	--



Vor- und Nachteile
Paarvergleich



Paarvergleich

Vorteile: <ul style="list-style-type: none">• Intuitive Beurteilung• Die Beurteilung ist in den meisten Fällen sehr zuverlässig.• Kleine Unterschiede in ähnlichen Geräuschen werden festgestellt.	Nachteile: <ul style="list-style-type: none">• Nur durchführbar mit wenig Geräuschen, ansonsten starker Anstieg der Versuchsdauer• Nur Rangfolge, Aussage zu den Abständen nur mittels statistischer Berechnungen möglich
---	---


Vor- und Nachteile
semantisches Differenzial



semantisches Differenzial


Vorteile: <ul style="list-style-type: none">• Urteile sind intervallskaliert• Die Ergebnisse können z. B. für eine Korrelationsanalyse mit messtechnischen Analysen verwendet werden.• Detaillierte Informationen über die Geräuschein-schaften werden erhoben.	Nachteile: <ul style="list-style-type: none">• Verschiedene Verzerrungs-effekte müssen berücksichtigt werden, um valide Test-ergebnisse zu erhalten.• Zeitaufwendig, nur anwend-bar mit überschaubarer Geräuschanzahl.
--	--

*Vor- und Nachteile
Kategorialbewertung*




Kategorialbewertung

Vorteile: <ul style="list-style-type: none">• Bei Verwendung einer entsprechenden Skala sind die Urteile intervallskaliert.• Die Ergebnisse können z. B. für eine Korrelationsanalyse mit messtechnischen Analysen verwendet werden.• Relativ kurze Testdauer auch bei einer größeren Geräuschanzahl.	Nachteile: <ul style="list-style-type: none">• Verschiedene Verzerrungseffekte müssen berücksichtigt werden, um valide Testergebnisse zu erhalten.• Für ungeübte Beurteiler unter Umständen nicht intuitiv.
--	---




*Vor- und Nachteile simultane
Kategorialbewertung*



simultane Kategorialbewertung

Vorteile: <ul style="list-style-type: none">• Bei Verwendung einer entsprechenden Skala sind die Urteile intervallskaliert.• Auch Bewertungen von sehr speziellen Geräuschphänomenen können abgefragt werden.• Relativ kurze Testdauer auch bei einer größeren Geräuschanzahl.	Nachteile: <ul style="list-style-type: none">• Die Bewertungssituation ist für einen ungeübten Teilnehmer komplex.• Das Testdesign ist eher für erfahrene Hörversuchsteilnehmer geeignet.
---	---



7. Realisierung von Hörversuchen mit individuellen Testdesigns

*Individuelle Testdesigns
realisieren*

Die API-Schnittstelle ASX03 in SQala bietet die Möglichkeit, eigene Versuchsdesigns zu erstellen, die über die hier vorgestellten Testarten hinausgehen. SQala ist somit individuell erweiterbar und erlaubt so auch die Umsetzung von Testdesigns, die in der Standardsoftware nicht verfügbar sind.

Um Erweiterungen mit der API-Schnittstelle zu erstellen, sind Programmierkenntnisse in C# erforderlich. Die Durchführung von Hörversuchen mit individuellem Testdesign ist zunächst auf den lokalen Modus beschränkt.

➔ Weiter zur [dritten Application Note über Hörversuche](#) mit grundlegenden Informationen zum Testablauf