WHITEPAPER

September 2026

Inhalt

1	Einführung	3
2	Physische Aufnahmereichweite	3
3	Aufzeichnungen in verschiedenen Umgebungen	5
4	Messergebnisse von Aufzeichnungen in verschiedenen Umgebungen	7
1	Anhang	9

1 Einführung

Die Festlegung eines definitiven maximalen Aufnahmebereichs für Audioaufnahmegeräte ist selten einfach, da die Leistung von der akustischen Umgebung abhängt. Zwar liefern technische Daten einen Grundbezugsrahmen, doch hängt der tatsächliche Aufnahmebereich von einem komplexen Zusammenspiel umgebungsbedingter und akustischer Faktoren ab, die bei jeder Installation einzigartig sind.

In diesem Whitepaper werden die Hauptfaktoren untersucht, die bestimmen, über welche Distanz ein Mikrofon Schall zuverlässig erfassen kann, wobei die Verständlichkeit von Sprache als Hauptmaßstab im Fokus steht. Zu den Hauptfaktoren zählen der Umgebungsgeräuschpegel, die spektralen Eigenschaften dieser Geräusche, die Nachhallzeit, die Richtcharakteristik des Mikrofons sowie das Eigenrauschen von Mikrofon und Aufzeichnungsgerät.

Um einen praktischen Bezugsrahmen zu erhalten, nutzen wir die menschliche Stimme als Referenzschallquelle. Der Schall wird gemessen und mit einem A-Filter bewertet, um eine bessere Korrelation zwischen dem gemessenen und dem wahrgenommenen Schallpegel zu erzielen. Das Signal-Rausch-Verhältnis (SNR) der aufgezeichneten Sprache wird als Differenz zwischen dem Stimmpegel und dem A-bewerteten Schallpegel in Dezibel berechnet *(siehe Anhang zum SNR)*. Empirische Untersuchungen zeigen, dass die Sprachverständlichkeit bis zu einem Signal-Rausch-Verhältnis (SNR) von etwa 0 dB (Rauschen und Sprache sind gleich laut) akzeptabel ist, vorausgesetzt, das Rauschen weist innerhalb des Sprachfrequenzbereichs einen relativ flachen Frequenzgang auf.

Durch das Verständnis dieser Variablen können Installateure und Systemplaner fundiertere Entscheidungen bei der Auswahl und Platzierung von Audiogeräten in realen Szenarien treffen.

2 Physische Aufnahmereichweite

Die Aufzeichnungsqualität eines Mikrofons hängt in der Regel von den Umgebungsgeräuschen ab. Die wichtigsten limitierenden Faktoren für die physische Aufnahmereichweite sind:

- **Schalldruckpegel:** Nicht alle Geräusche breiten sich über die gleiche Entfernung aus. Ein Schuss erzeugt zum Beispiel einen extrem hohen Schalldruckpegel und kann bis zu einem Kilometer weit zu hören sein. Eine menschliche Stimme ist jedoch leiser und verklingt mit zunehmender Entfernung schneller. Zur realistischen Schätzung des Aufnahmebereichs müssen Sie zunächst definieren, welche Art von Schall aufgezeichnet werden soll, damit der Pegel der Schallquelle die Obergrenze dessen vorgibt, was aufgenommen werden kann.
- **Umgebungsgeräusche:** Je lauter die Umgebung ist, desto schwieriger ist es für ein Mikrofon, das gewünschte Geräusch zu isolieren und aufzunehmen. Die Umgebungsgeräusche fallen in jeder Umgebung unterschiedlich aus und stehen in direkter Konkurrenz zum Zielgeräusch. Zu ihnen gehören unter anderem Wind-, Verkehrs-, Maschinen- und Publikumsgeräusche. Je höher das Grundgeräuschpegel am Einsatzort des Mikrofons ist, desto kleiner ist der effektive Aufnahmebereich. Das erklärt, warum dasselbe Mikrofon in einem ruhigen Korridor ganz anders funktioniert als an einer belebten Kreuzung.
- **Spektralprofil des Geräuschs:** Ebenso wie die Lautstärke eines Geräuschs ist auch seine Frequenz von Bedeutung. Die menschliche Sprache liegt hauptsächlich im mittleren Frequenzbereich, etwa zwischen 300 Hz und 3.400 Hz. Wenn Umgebungsgeräusche vor allem im sehr tiefen oder sehr hohen Frequenzbereich auftreten, beeinträchtigen sie die Sprachverständlichkeit in der Regel weniger. Wenn Umgebungsgeräusche sich jedoch mit demselben Frequenzbereich wie das Zielgeräusch überschneiden, wird es selbst bei mäßigen Schallpegeln schwierig, Sprache von diesen zu unterscheiden.
- **Nachhallzeit am Standort des Mikrofons:** Nachhall ist die Fortdauer des Schalls nach dem Verstummen der ursprünglichen Schallquelle, die durch Reflexionen an Wänden, Decken, Böden und anderen Oberflächen verursacht wird. Eine lange Nachhallzeit kann die Sprachverständlichkeit beeinträchtigen, was selbst in Situationen gilt, in denen der Gesamtgeräuschpegel und das Signal-Rausch-Verhältnis akzeptabel erscheinen. In Räumen mit starkem Nachhall, wie etwa großen, leeren Hallen oder gefliesten Räumen, trifft der Sprachschall mit einer unendlichen Anzahl von Raumreflexionen auf das Mikrofon, wodurch der Klang verschwimmt. Das erschwert es sowohl Personen als auch Aufzeichnungssystemen, einzelne Wörter deutlich zu unterscheiden.

- **Richtcharakteristik der Mikrofonkonfiguration:** Mikrofone unterscheiden sich darin, wie empfindlich sie auf Geräusche aus verschiedenen Richtungen reagieren. Ein ungerichtetes Mikrofon nimmt Schall aus allen Richtungen gleichermaßen auf, was auch bedeutet, dass es Umgebungsgeräusche und Nachhall aus allen Richtungen aufnimmt. Eine gerichtete Konfiguration, wie beispielsweise mehrere für die Strahlformung konfigurierte Mikrofone, konzentriert die Empfindlichkeit auf eine bestimmte Richtung, wodurch der Einfluss von Geräuschen und Nachhall aus anderen Richtungen wirksam reduziert wird. Das kann den nutzbaren Aufnahmebereich in lauten oder stark hallenden Umgebungen erheblich vergrößern, allerdings nur in der Richtung, in die die Mikrofonanordnung ausgerichtet ist.
- **Eigenrauschen von Mikrofonen und Aufzeichnungsgeräten:** Jedes Mikrofon und Audioaufzeichnungsgerät erzeugt ein geringes Maß an internem elektrischem Rauschen, das als Eigenrauschen bezeichnet wird. Es gibt die absolute Untergrenze für das vor, was das System aufnehmen kann. Ganz gleich, wie leise eine Umgebung auch sein mag: Geräusche, die unterhalb der Schwelle des Eigenrauschens liegen, lassen sich einfach nicht klar aufzeichnen. Beim Vergleich von Mikrofonen bedeutet ein niedrigeres angegebenes Eigenrauschen in der Regel, dass das Mikrofon leisere Geräusche besser erfasst, was den Aufnahmebereich vergrößern kann, wenn Lärmpegel und Nachhallzeit niedrig genug ausfallen.

Für eine vereinfachte Analyse können wir von einem idealisierten Szenario ausgehen, in dem es weder Geräusche noch Reflexionen oder Nachhall gibt, und die menschliche Stimme als Referenzschallquelle heranziehen. Unter diesen Umständen lautet die entscheidende Frage: Welches Signal-Rausch-Verhältnis (SNR) sollten wir als Grenzwert für einen akzeptablen Aufnahmebereich ansetzen?

Empirische Untersuchungen liefern uns eine klare Antwort. Wenn wir einen A-Bewertungsfilter auf das Systemrauschen anwenden und den durchschnittlichen RMS-Pegel (Root Mean Square - quadratischer Mittelwert) des Rauschens mit dem durchschnittlichen RMS-Pegel des Sprachschalls vergleichen, nimmt die Sprachverständlichkeit erst dann ab, wenn das Signal-Rausch-Verhältnis (SNR) unter 0 dB fällt. Daher verwenden wir ein Signal-Rausch-Verhältnis (SNR) von 0 dB als Grenzwert für die Aufzeichnung menschlicher Sprache. Dies gilt, solange das Rauschen einen relativ flachen Frequenzgang besitzt, ähnlich wie weißes Rauschen, und innerhalb des Sprachfrequenzbereichs liegt, was beim Eigenrauschen von Mikrofonen üblicherweise der Fall ist.

Ein digitales Mikrofon kann zum Beispiel eine Empfindlichkeit von -36 dBFSrms aufweisen *(siehe Anhang zu dBFSrms)*, wenn ein 1-kHz-Ton mit 94 dBspl wiedergegeben wird (voll ausgesteuertes digitalisiertes Sinussignal wird hier entsprechend Mikrofondatenblättern als 0 dBFSrms definiert) *(siehe Anhang zu dBspl)*. Das gilt für viele Produkte von Axis mit integrierten digitalen MEMS-Mikrofonen, wenn die Verstärkung des Mikrofons auf 0 dB gesetzt wird. Befindet sich dieses Mikrofon in einer sehr ruhigen Umgebung, in der der Umgebungsgeräuschpegel unter 10 dBspl(A) fällt *(siehe Anhang zu dBspl(A))*, wird das Eigenrauschen des Mikrofons zum dominanten Geräusch. In diesem Beispiel ist im Datenblatt des Mikrofons ein digitaler Ausgangspegel von -105 dBFSrms(A) angegeben. Das Signal-Rausch-Verhältnis (SNR) des Mikrofons ist die Differenz zwischen dem Referenzpegel und dem Eigenrauschen: -36 – (-105) = 69 dB.

Wir rechnen dieses elektrische Eigenrauschen in einen äquivalenten Schallpegel um: 94 – 69 = 25 dBspl(A).

Dann berechnen wir die maximale Entfernung, bei der wir menschliche Sprache mit einem Signal-Rausch-Verhältnis (SNR) von 0 dB aufzeichnen können. Gemäß ITU-Standard beträgt der nominelle durchschnittliche Sprachpegel (average speech level, ASL) einer menschlichen Stimme etwa 57 dBspl bei einer Entfernung von 1 m. Im freien Raum ohne reflektierende Objekte nimmt der Schalldruckpegel mit der Entfernung gemäß dem reziproken Abstandsgesetz ab:

$$dB = 20\log\frac{reference\ distance}{test\ distance}$$

.

Beträgt der Referenzabstand 1 m, muss der Sprachpegel auf 25 dBspl(A) sinken, um bei diesem Mikrofon ein Signal-Rausch-Verhältnis (SNR) von 0 dB zu erreichen, was einer Verringerung um 32 dB gegenüber dem Referenzpegel bei 1 m entspricht. Wenn wir die Formel umstellen, erhalten wir:

$$test\ distance = \frac{reference\ distance}{10^{\frac{-(57-noise\ level)}{20}}} = \frac{1}{10^{\frac{-32}{20}}} = 39.8\text{m}$$

.

Das bedeutet, dass dieses Mikrofon im freien Raum ohne Umgebungsgeräusche, Wind oder Nachhall theoretisch verständliche Sprache auf eine Entfernung von etwa 40 m aufnehmen kann. Das entspricht dem theoretischen maximalen Aufnahmebereich.

Umgebungsfaktoren haben jedoch einen größeren Einfluss auf den Rauschpegel als das Eigenrauschen des Mikrofons. In einem ruhigen Büro nach Feierabend, in dem die Lüftung noch läuft, liegt der Rauschpegel bei etwa 28–32 dBspl. Das allein kann den effektiven Aufnahmebereich um etwa 50 % verringern, wodurch sich die maximale Reichweite auf etwa 20 m reduziert. Bei einer Aufzeichnung im Freien in einer städtischen Umgebung sind wir mit einer Kombination verschiedenster Geräusche konfrontiert: Windgeräusche, Verkehrslärm, durch Menschen verursachte Geräusche usw. Ein ungerichtetes Mikrofon hat für alle Schallpegel und Schallspektren einen eigenen entsprechenden Aufnahmebereich. Hier beträgt der Bereich bei lauten Verkehrsgeräuschen maximal 2 bis 4 m.

Ein Aspekt, der in der Praxis beachtet werden muss, ist: Aufzeichnungen menschlicher Sprache, die im Bereich des Grenzwerts von 0 dBSNR gemacht wurden, erfordern eine erhebliche digitale Verstärkung, um hörbar zu sein, da der Pegel der aufgezeichneten Sprache so niedrig ist, dass er andernfalls bei der Wiedergabe unter unsere eigene Hörschwelle fiele. In den folgenden Aufzeichnungen wurden die Audiodateien um +40 dB verstärkt, um den geringen Schallpegel der entfernten Sprachquelle auszugleichen.

3 Aufzeichnungen in verschiedenen Umgebungen

Dieser Abschnitt befasst sich mit den Details verschiedener Aufzeichnungsumgebungen und veranschaulicht diese durch Screenshots aus den Videoaufzeichnungen sowie Sprach-Samples, die in verschiedenen Entfernungen aufgenommen wurden.

Beispiel 1: Der Aufzeichnungsort befindet sich im Freien an einer Autobahn, von der Verkehrsgeräusche zu hören sind. Der Umgebungsgeräuschpegel schwankt zwischen 62 und 68 dBspl[A] und zusätzlich treten gelegentlich Windgeräusche auf.



Abbildung 3.1 *Der Aufzeichnungsort befindet sich in der Nähe einer Autobahn (siehe roter Punkt).*



Abbildung 3.2 *Die maximale Aufnahmedistanz zu einer am Aufzeichnungsort verwendeten Body Worn-Kamera von Axis beträgt 4 m.*

Verwenden Sie zur Wiedergabe der Aufzeichnung vorzugsweise einen Kopfhörer.

Beispiel 2: Ein ruhiger Korridor mit einem Umgebungsgeräuschpegel von 32 dBspl[A]. Der Korridor in diesem Beispiel wirkt wie ein langer Schallleiter. Der Schall breitet sich im Korridor aus, wird von den reflektierenden Oberflächen zurückgeworfen und erreicht die Mikrofone lauter, als sie im freien Raum bei gleicher Entfernung wären.



Abbildung 3.3 *Die maximale Aufnahmedistanz beträgt 27 m.*

Verwenden Sie zur Wiedergabe der Aufzeichnung vorzugsweise einen Kopfhörer.

Beispiel 3: Eine große Empfangshalle mit einem Umgebungsgeräuschpegel von 42 dBspl[A]. Hier befindet sich schalldämmendes Material hinter den Diffusorplatten, wodurch die meisten Erstreflexionen unterbunden werden. Aus größerer Entfernung breitet sich der Schall im akustischen Kanal zwischen dem Boden und der reflektierenden Decke aus.

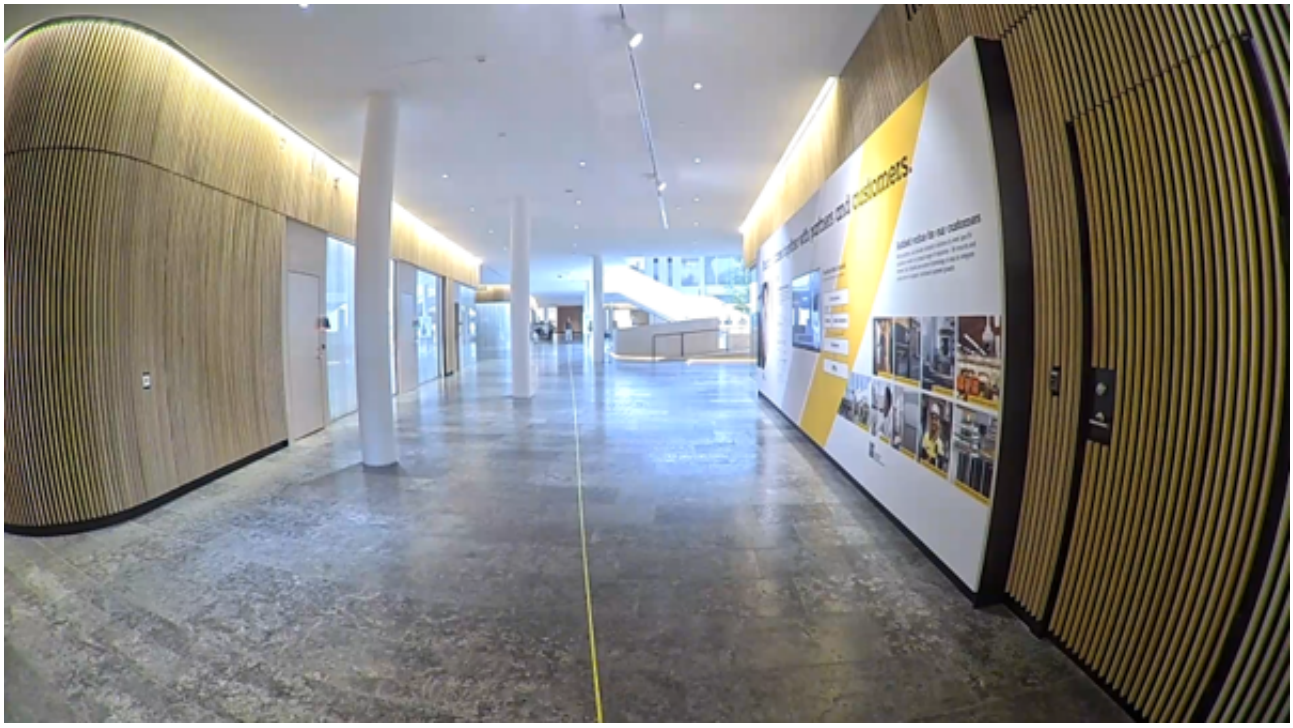


Abbildung 3.4 *Bei einer maximalen Aufnahmedistanz von 40 m ist der Sprecher kaum zu sehen, aber dennoch zu hören.*

Verwenden Sie zur Wiedergabe der Aufzeichnung vorzugsweise einen Kopfhörer.

4 Messergebnisse von Aufzeichnungen in verschiedenen Umgebungen

Im Fokus dieser Analyse steht die Aufnahme von Sprache. Lautere Geräusche wie Schüsse sind aus größerer Entfernung zu hören. Wir haben jeden Satz aus den Audioaufzeichnungen analysiert und gemessen und für jeden aufgezeichneten Satz das Signal-Rausch-Verhältnis (SNR) geschätzt.

Das Signal-Rausch-Verhältnis (SNR) wurde wie folgt geschätzt: Sprache + Rauschen - Umgebungsgeräusche. Genauer wäre die folgende Rechnung: Sprachpegel - Rauschpegel, doch es gibt keine Möglichkeit, den Sprachpegel ohne die Umgebungsgeräusche zu messen. Das bedeutet, dass die Genauigkeit des Signal-Rausch-Verhältnisses (SNR) abnimmt, wenn der Sprachpegel nahe am Rauschpegel liegt.

Aufzeichnungsort und Umgebungsgeräu-sche:	Im Freien mit Verkehrsgeräu-schen mit 61–68 dBspl[A]	In einem stillen Korridor mit 32 dBspl[A]	In einer großen Empfangshalle mit 42 dBspl[A]	Freier Raum, keine Reflexionen, absolute Stille
Distanz (m)	**SNR [dB] (geschätzt)**	**SNR [dB] (geschätzt)**	**SNR ([dB] geschätzt)**	**Theoretisches SNR = Sprachpegel Mikrofonrauschen [dB]**

1	2,4	29,4	21,5	32,0
2	1,0	25,6	17,2	26,0
4	0,5	23,4	14,2	20,0
8	0,4	18,9	11,2	13,9
16	Nicht hörbar	17,5	6,3	7,9
24	Nicht hörbar	14,8	5,4	4,4
32	Nicht hörbar	Korridor zu kurz	4,6	1,9
40	Nicht hörbar	Korridor zu kurz	4,7	0,0

Theoretisch sollte der Sprachpegel bei der doppelten Entfernung um -6 dB abnehmen. Das hieße auch, dass sich das Signal-Rausch-Verhältnis entsprechend verringern müsste. Aus dieser Tabelle geht hervor, dass das Signal-Rausch-Verhältnis im Korridor und in der Empfangshalle besser ausfällt als erwartet. Der Grund hierfür ist die Erhöhung der Schalldruckpegel durch die Reflexionen im Raum. Vor allem im Korridor wirken Wände, Boden und Decke als Schallleiter. Dadurch liegen die Lautstärke und das Signal-Rausch-Verhältnis bei einem Test in einem Korridor mit einer Entfernung von 24 m um etwa 10 dB über dem theoretischen Signal-Rausch-Verhältnis im freien Raum.

Sobald die Umgebungsgeräusche das akustische Äquivalent des Eigenrauschens eines einzelnen ungerichteten Mikrofons übersteigen, hängt der Aufnahmebereich ausschließlich von den Umgebungsgeräuschen und nicht vom Mikrofon selbst ab.

Anhang 1 Anhang

- **Schalldruckpegel in Dezibel (dBspl):** Der Schalldruckpegel in Dezibel (dB) bezogen auf 20 µPa. Bei der Messung von Rausch- und Sprachpegeln werden häufig zeitintegrierte RMS-Werte herangezogen. Diese werden manchmal als Schallleistungspegel bezeichnet.

$$RMS = \sqrt{\frac{1}{T2 - T1}\int_{T1}^{T2} [f(t)]^2 \, dt}$$

  Die beiden gängigsten Integrationszeiten sind „Fast RMS" mit einem Zeitfenster von 125 ms und „Slow RMS" mit einem Zeitfenster von 1 Sekunde. Bei der Analyse des Grundrauschens wird der Pegel häufig über einen längeren Zeitraum gemittelt.

- **Mit A-Bewertungsfilter gewichteter Schalldruckpegel (dBspl[A]):** Dieser Pegel wird manchmal als dB(A) angegeben. Es handelt sich um den A-gewichteten Wert des dBspl-Pegels. Die A-Bewertung ist eine Methode zur Filterung des Mikrofonsignals, um den gemessenen Schalldruckpegel besser an das menschlichen Gehör anzupassen. Grund hierfür ist die geringere Empfindlichkeit des menschlichen Gehörs bei tiefen Frequenzen unter 500 Hz und hohen Frequenzen über 10 kHz.

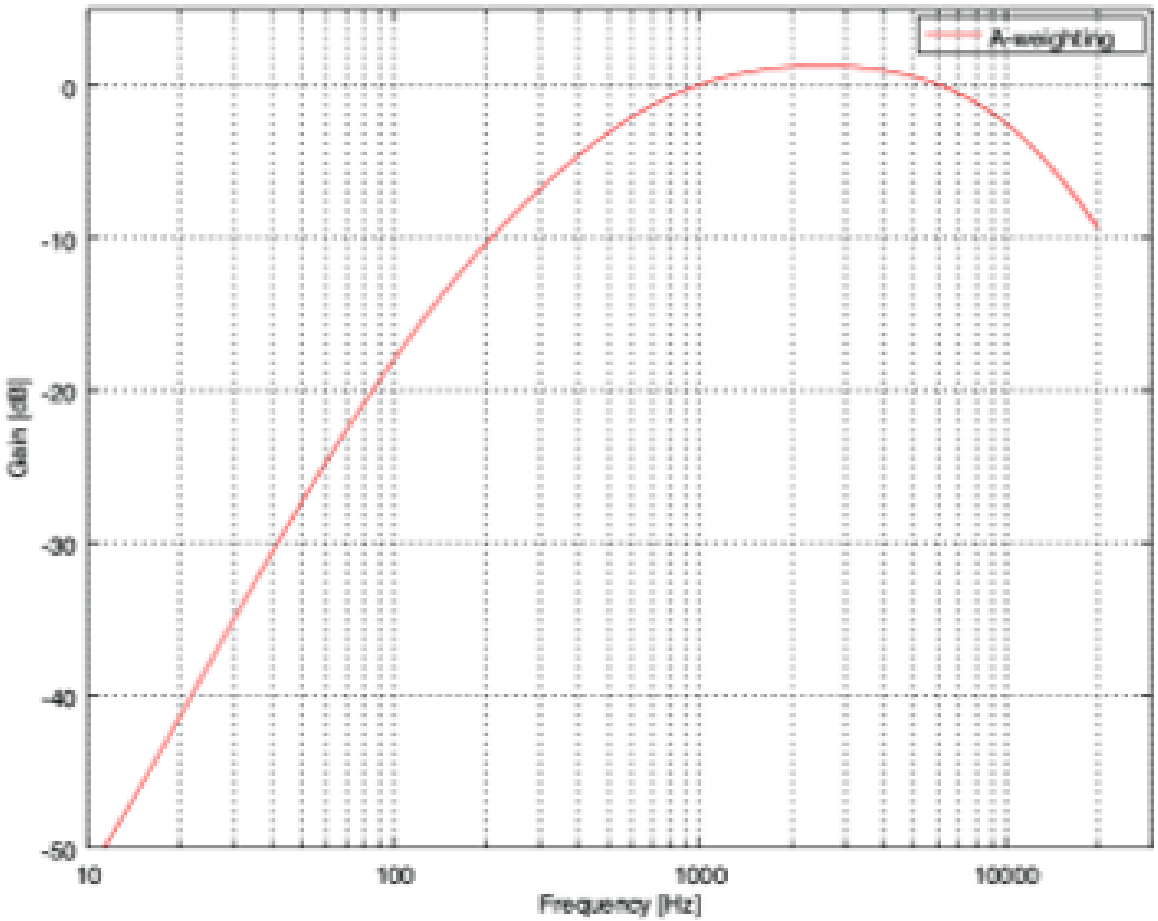


Abbildung .1 *A-Bewertungsfilter.*

- Der **Spitzenpegel innerhalb der digitalen Pegelskala (dBFSpeak/dBFSp)** bezeichnet eine digitale Audioamplitude im Verhältnis zum maximal digital darstellbaren Wert (Pegelskala). Eine Sinuswelle mit maximaler digitaler Amplitude, die nicht abgeschnitten wird, hat einen Wert von 0 dBFSpeak.

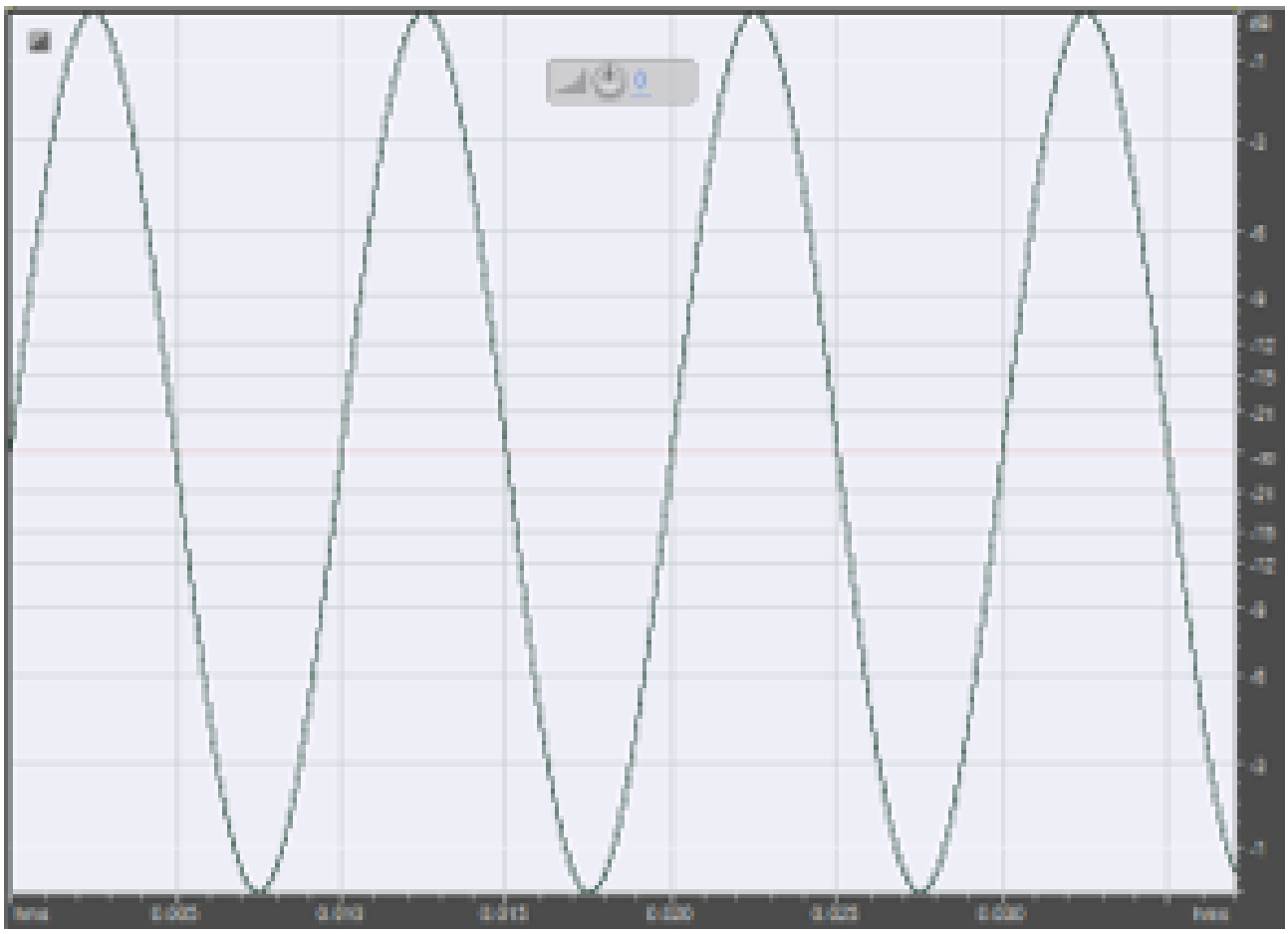


Abbildung .2 *Ein Beispiel für ein Signal mit einer Amplitude von 0 dBFSpeak.*

- **dBFSrms** bezeichnet den digitalen Audioleistungspegel eines digitalen Signals. Ein voll ausgesteuertes Sinussignal mit 0 dBFSpeak kann je nach Definition der Einheit unterschiedliche RMS-Pegel aufweisen. In Mikrofondatenblättern wird ein voll ausgesteuertes Sinussignal als 0 dBFSrms definiert, weswegen wir diese Definition auch in diesem Whitepaper verwenden. Im Allgemeinen ist die andere Definition eines voll ausgesteuerten Sinussignals von -3 dBFSrms mathematisch korrekter und steht eher im Einklang mit der Definition von RMS- und Spitzenpegeln im analogen Bereich. Nach dieser anderen Definition entspricht ein voll ausgesteuertes Rechtecksignal 0 dBFSrms.
- Die **Mikrofonempfindlichkeit** ist der Ausgangspegel eines Mikrofons bei einem akustischen Eingangssignal in Form eines sinusförmigen Tons mit einer Frequenz von 1 kHz bei einem Schalldruckpegel von 1 Pa, 0 dB a oder 94 dBspl (drei verschiedene Ausdrucksweisen für denselben Schalldruckpegel).
- **Signal-Rausch-Verhältnis (SNR):** Wenn Schallpegel in dB angegeben werden, berechnet sich das Verhältnis aus der Differenz zwischen dem Signalpegel und dem Rauschpegel.
- Das **Signal-Rausch-Verhältnis (SNR) eines Mikrofons** wird als Differenz des Ausgangspegels zwischen einem Testton von 1 kHz bei 94 dBspl und dem Eigenrauschen des Mikrofons in völliger Stille berechnet. Ein typisches Signal-Rausch-Verhältnis für ein digitales MEMS-Mikrofon beträgt 69 dBspl[A].
- Mit dem in diesem Whitepaper verwendeten Begriff **Signal-Rausch-Verhältnis der Sprache** ist das Verhältnis von menschlicher Sprache und dem A-bewerteten Rauschpegel bei Abwesenheit von Sprachschall gemeint. Das Signal-Rausch-Verhältnis wird in Dezibel gemessen. Bei der Berechnung des theoretischen Aufnahmebereichs für Sprache gehen wir von einem Signal-Rausch-Verhältnis der Sprache von 0 dB als Definition für das niedrigste akzeptable Signal-Rausch-Verhältnis aus. In diesem Zusammenhang kann das Rauschen sowohl durch Umgebungsgeräusche und elektrisches Rauschen als auch durch das Eigenrauschen des Mikrofons verursacht werden.
- **Rauschpegel eines Mikrofons:** Der digitale Rauschpegel eines Mikrofons wird in dBFSrms gemessen. In der Regel erreichen die besten digitalen Mikrofone ein theoretisches Grundrauschen von -105 dBFSrms[A].

- Das **akustische Äquivalent des Mikrofonrauschens** lässt sich als Testpegel von 94 dBspl minus des Signal-Rausch-Verhältnisses (SNR) des Mikrofons berechnen. Ein gängiges Beispiel hierfür wäre
  94 – 69 = 25 dBspl[A].

Über Axis Communications

Axis ermöglicht eine smartere und sichere Welt durch die Verbesserung von Sicherheit, Schutz, betrieblicher Effizienz und Geschäftsanalytik. Als Technologieführer im Bereich Netzwerk-Video bietet Axis Videosicherheits-, Zutrittskontroll-, Intercom- und Audiolösungen. Die branchenweit anerkannten Schulungen der Axis Communications Academy vermitteln fundiertes Expertenwissen zu den neuesten Technologien.

Das 1984 gegründete schwedische Unternehmen beschäftigt etwa 5.000 engagierte Mitarbeiter in über 50 Ländern und bietet mit Technologie- und Systemintegrationspartnern auf der ganzen Welt kundenspezifische Lösungen an. Der Hauptsitz ist in Lund, Schweden.

© 2026 — Axis Communications AB. AXIS COMMUNICATIONS, AXIS, ARTPEC und VAPIX sind in verschiedenen Jurisdiktionen eingetragene Marken von Axis AB. Alle anderen Marken sind Eigentum ihrer jeweiligen Inhaber.Wir behalten uns das Recht auf Änderungen ohne Vorankündigung vor.