BIAŁA KSIĘGA

Wrzesień 2026

Spis treści

1	Wprowadzenie	3
2	Fizyczny zasięg nagrywania	3
3	Nagrywanie w różnych warunkach	5
4	Wyniki pomiarów uzyskane na podstawie nagrań wykonanych w różnych środowiskach	7
1	Dodatek	9

1 Wprowadzenie

Określenie definitywnego maksymalnego zasięgu zbierania dźwięku dla urządzeń do nagrywania audio rzadko jest proste, ponieważ ich wydajność zależy od warunków akustycznych otoczenia. Chociaż specyfikacje techniczne stanowią pewien punkt odniesienia, faktyczny zasięg nagrywania zależy od złożonego wzajemnego oddziaływania czynników środowiskowych i akustycznych, które różnią się w zależności od instalacji.

W niniejszej białej księdze omówiono główne czynniki decydujące o tym, z jakiej odległości mikrofon jest w stanie niezawodnie rejestrować dźwięk, ze szczególnym uwzględnieniem zrozumiałości mowy jako podstawowego punktu odniesienia. Czynniki te obejmują poziom szumów tła, charakterystykę spektralną tych szumów, czas pogłosu, kierunkowość mikrofonu oraz szum własny zarówno samego mikrofonu, jak i urządzenia nagrywającego.

Aby ustalić praktyczną bazę odniesienia, jako źródło dźwięku wykorzystujemy ludzki głos. Szumy mierzy się z użyciem filtra korekcyjnego A, co pozwala uzyskać lepszą korelację między wartością pomiaru a postrzeganym poziomem szumów. Stosunek sygnału do szumu (SNR) dla zarejestrowanej mowy oblicza się jako różnicę między poziomem głosu a poziomem szumów z korekcją A wyrażonymi w decybelach *(definicję SNR podano w Dodatku)*. Badania empiryczne wskazują, że zrozumiałość mowy jest akceptowalna nawet przy SNR wynoszącym około 0 dB (tzn. gdy szum i dźwięk są równie głośne), pod warunkiem że szum ma stosunkowo płaską charakterystykę częstotliwościową w zakresie częstotliwości mowy.

Zrozumienie tych zmiennych pozwala instalatorom oraz projektantom systemów podejmować bardziej świadome decyzje przy wyborze i rozmieszczaniu urządzeń audio w rzeczywistych warunkach.

2 Fizyczny zasięg nagrywania

Funkcjonalność mikrofonu w zakresie nagrywania zależy na ogół od poziomu szumów otoczenia. Głównymi czynnikami ograniczającymi fizyczny zasięg nagrywania są:

- **Poziom ciśnienia akustycznego:** nie wszystkie dźwięki rozchodzą się na taką samą odległość. Na przykład wystrzał z broni palnej wytwarza niezwykle wysoki poziom ciśnienia akustycznego i może być słyszalny z odległości nawet kilometra. Ludzki głos jest jednak cichszy i szybciej zanika wraz z odległością. Aby oszacować realistyczny zasięg zbierania dźwięku, należy najpierw określić rodzaj dźwięku, który ma być rejestrowany, tak aby poziom sygnału źródłowego wyznaczał górną granicę tego, co można nagrać.
- **Szum tła:** im głośniejsze otoczenie, tym trudniej mikrofonowi wyodrębnić i zarejestrować pożądany dźwięk. Szum tła różni się w zależności od otoczenia i bezpośrednio konkuruje z dźwiękiem docelowym. Obejmuje on między innymi odgłosy wiatru, ruchu ulicznego, maszyn oraz tłumu. Im wyższy poziom szumu tła w miejscu umieszczenia mikrofonu, tym mniejszy efektywny zasięg zbierania dźwięku. To wyjaśnia, dlaczego ten sam mikrofon może działać zupełnie inaczej w cichym korytarzu niż na ruchliwym skrzyżowaniu.
- **Profil spektralny szumu:** tak samo jak ważna jest głośność szumu, tak samo istotne są częstotliwości, na których występuje. Mowa ludzka mieści się głównie w zakresie średnich częstotliwości, czyli mniej więcej od 300 Hz do 3400 Hz. Jeśli szum tła koncentruje się na bardzo niskich lub bardzo wysokich częstotliwościach, jest mniej prawdopodobne, że wpłynie on negatywnie na zrozumiałość mowy. Jeśli jednak szum występuje na tym samym zakresie częstotliwości co dźwięk docelowy, odróżnienie mowy od szumu staje się trudne, nawet przy umiarkowanym poziomie szumów.
- **Czas pogłosu w miejscu umieszczenia mikrofonu:** pogłos to utrzymywanie się dźwięku po zaniku dźwięku ze źródła pierwotnego, spowodowane odbiciami od ścian, sufitów, podłóg i innych powierzchni. Długi czas pogłosu może pogorszyć zrozumiałość mowy, nawet w sytuacjach, w których ogólny poziom dźwięku i wartość SNR wyglądają na akceptowalne. W przestrzeniach o silnym pogłosie, takich jak duże puste sale lub pomieszczenia wyłożone płytkami, dźwięk mowy dociera do mikrofonu wraz z nieskończoną liczbą odbić, co powoduje jego rozmycie. Utrudnia to wyraźne rozróżnienie poszczególnych słów zarówno słuchaczom, jak i systemom nagrywającym.
- **Kierunkowość układu mikrofonów:** mikrofony różnią się czułością na dźwięki dochodzące z różnych kierunków. Mikrofon dookólny rejestruje dźwięk ze wszystkich kierunków w równym stopniu, co oznacza również, że wychwytuje szum tła i pogłos ze wszystkich stron. Układ kierunkowy, np. wiele mikrofonów skonfigurowanych z wykorzystaniem kształtowania wiązki, skupia czułość w konkretnym kierunku, co skutecznie redukuje wpływ szumu i pogłosu docierających z innych kątów. Może to znacznie zwiększyć

użyteczny zasięg zbierania dźwięku w hałaśliwych lub pogłosowych środowiskach, ale tylko w tym kierunku, w którym skierowany jest układ mikrofonów.

- **Szum własny mikrofonów i urządzeń nagrywających:** każdy mikrofon i urządzenie nagrywające generuje niewielką ilość wewnętrznego szumu elektrycznego, zwanego szumem własnym. Stanowi on absolutną dolną granicę dla dźwięków, które system jest w stanie zarejestrować. Niezależnie od tego, jak ciche jest otoczenie, dźwięków o natężeniu niższym od poziomu szumu własnego po prostu nie da się czysto nagrać. Przy porównywaniu mikrofonów niższa wartość szumu własnego zazwyczaj oznacza, że mikrofon lepiej rejestruje cichsze dźwięki, co może przełożyć się na większy zasięg zbierania dźwięku, o ile poziom szumów otoczenia i czas pogłosu są wystarczająco niskie.

Aby uprościć analizę, możemy zacząć od idealnego scenariusza, w którym nie ma szumów tła, odbić ani pogłosu, a jako referencyjnego źródła dźwięku używamy ludzkiego głosu. W tych warunkach kluczowe pytanie brzmi: jaki stosunek sygnału do szumu (SNR) powinniśmy przyjąć jako próg dla akceptowalnego zasięgu zbierania dźwięku?

Badania empiryczne dają nam jasną odpowiedź. Gdy zastosujemy filtr korekcyjny A do szumów systemu i porównamy średnią wartość skuteczną (RMS) poziomu szumów ze średnią wartością RMS mowy, zrozumiałość mowy zaczyna spadać dopiero wtedy, gdy SNR spada poniżej 0 dB. W związku z tym jako kryterium graniczne dla nagrywania ludzkiego głosu przyjmujemy SNR wynoszący 0 dB. Zasada ta sprawdza się, gdy szumy mają stosunkowo płaską charakterystykę częstotliwościową, zbliżoną do białego szumu i mieszczą się w zakresie częstotliwości mowy, co zazwyczaj ma miejsce w przypadku szumu własnego mikrofonu.

Na przykład mikrofon cyfrowy może mieć czułość wynoszącą -36 dBFS RMS *(definicję dBFS RMS podano w Dodatku)*, gdy odtwarzany jest sygnał o częstotliwości 1 kHz i poziomie 94 dBSPL (zdigitalizowana sinusoida o pełnej amplitudzie jest tutaj zdefiniowana jako 0 dBFS RMS zgodnie z kartami katalogowymi mikrofonów) *(definicja dBSPL znajduje się w Dodatku)*. Dotyczy to wielu produktów firmy Axis wyposażonych we wbudowane cyfrowe mikrofony MEMS, gdy wzmocnienie mikrofonu jest ustawione na 0 dB. Gdy mikrofon ten zostanie umieszczony w bardzo cichym otoczeniu, w którym poziom szumów tła spadnie poniżej 10 dBSPL(A) *(definicję dBSPL(A) podano w Dodatku)*, dominującym stanie się szum generowany przez mikrofon wewnętrzny. W tym przykładzie w karcie katalogowej mikrofonu podano, że poziom sygnału na wyjściu cyfrowym wynosi -105 dBFS RMS(A). SNR mikrofonu stanowi różnicę między poziomem odniesienia a podłogą szumową: –36 – (–105) = 69 dB.

Przeliczamy ten poziom szumu elektrycznego na równoważny poziom szumów akustycznych:
94 – 69 = 25 dBSPL(A).

Teraz obliczymy maksymalną odległość, z której możemy nagrywać ludzki głos przy SNR wynoszącym 0 dB. Zgodnie z normami ITU nominalny średni poziom mowy (ASL) ludzkiego głosu wynosi około 57 dBSPL z odległości 1 m. W wolnej przestrzeni, bez obiektów odbijających dźwięk, poziom ciśnienia akustycznego maleje wraz z odległością zgodnie z prawem odwrotnych kwadratów:

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

.

Jeśli odległość odniesienia wynosi 1 m, to aby osiągnąć w tym mikrofonie SNR na poziomie 0 dB, poziom mowy musi spaść do 25 dBSPL(A), co oznacza spadek o 32 dB w stosunku do poziomu przy odległości odniesienia wynoszącej 1 m. Po przekształceniu wzoru otrzymujemy:

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

.

Oznacza to, że w wolnej przestrzeni, przy braku szumów tła, wiatru i pogłosu, mikrofon ten może teoretycznie nagrywać zrozumiałą mowę z odległości około 40 m. Jest to teoretyczny maksymalny zasięg zbierania dźwięku.

Jednak czynniki środowiskowe mają większy wpływ na poziom szumów niż wewnętrzna podłoga szumowa mikrofonu. W cichym biurze po godzinach pracy, gdy systemy wentylacyjne nadal działają, poziom szumów tła wynosi około 28–32 dBSPL. Już taka wartość może zmniejszyć efektywny zasięg zbierania dźwięku o około 50%,

ograniczając maksymalną odległość do około 20 m. W przypadku nagrywania dźwięku na zewnątrz w środowisku miejskim, występuje kombinacja różnych szumów: odgłos wiatru, hałas uliczny, hałasy wynikające z działalności człowieka itd. W przypadku mikrofonu dookólnego każdy poziom szumów i każde widmo szumów ma swój odpowiedni zasięg zbierania dźwięku. W tym przypadku zasięg ten może wynosić maksymalnie 2–4 m, jeśli hałas uliczny jest głośny.

Jedną z praktycznych kwestii jest to, że nagrania ludzkiego głosu wykonane przy progu 0 dBSNR lub w jego pobliżu wymagają znacznego wzmocnienia cyfrowego, aby były słyszalne, ponieważ zarejestrowany poziom głosu jest tak niski, że w przeciwnym razie podczas odtwarzania znalazłby się poniżej progu słyszalności. W poniższych przykładach plików audio nagrania zostały wzmocnione o +40 dB, aby skompensować poziom głosu dochodzący z daleka.

3 Nagrywanie w różnych warunkach

W tej sekcji przedstawiono szczegółowe informacje na temat lokalizacji nagrywania wraz ze zrzutami ekranu z nagrań wideo oraz próbkami głosu zarejestrowanymi z różnych odległości.

Przykład 1: lokalizacja znajduje się na zewnątrz, a w tle słychać hałas ruchu pojazdów poruszających się autostradą. Poziom szumów tła waha się od 62 do 68 dBSPL[A], a ponadto słychać również szum wiatru.



Ilustracja 3.1 *Miejsce nagrywania znajduje się w pobliżu autostrady (patrz czerwona kropka).*



Ilustracja 3.2 *W miejscu nagrywania maksymalna odległość nagrywania wynosi 4 metry od kamery nasobnej Axis.*

Podczas odsłuchiwania nagrania najlepiej używać słuchawek.

Przykład 2: cichy korytarz o poziomie szumów tła wynoszącym 32 dBSPL[A]. W tym przypadku korytarz pełni rolę długiego falowodu akustycznego. Dźwięk rozchodzi się po nim, odbijając się od powierzchni odbijających, i dociera do mikrofonów z większym poziomem głośności niż miałoby to miejsce w wolnej przestrzeni przy tej samej odległości.



Ilustracja 3.3 *Maksymalny zasięg nagrywania wynosi 27 m.*

Podczas odsłuchiwania nagrania najlepiej używać słuchawek.

Przykład 3: duży hol o poziomie szumów tła wynoszącym 42 dBSPL[A]. Za panelami rozpraszającymi pokrywającymi ściany umieszczono elementy pochłaniające dźwięk, dzięki czemu większość wczesnych odbić zostaje wyeliminowana. Na większej odległości dźwięk rozchodzi się w kanale akustycznym między podłogą a odbijającym dźwięk sufitem.



Ilustracja 3.4 *Przy maksymalnej odległości nagrywania wynoszącej 40 m osoba mówiąca jest ledwo widoczna, ale nadal można ją usłyszeć.*

Podczas odsłuchiwania nagrania najlepiej używać słuchawek.

4 Wyniki pomiarów uzyskane na podstawie nagrań wykonanych w różnych środowiskach

Poniższa analiza koncentruje się na zbieraniu głosu. Głośniejsze dźwięki, takie jak strzały, będą słyszalne z większej odległości. Przeanalizowaliśmy i zmierzyliśmy każde zdanie z nagrań audio, a następnie oszacowaliśmy SNR dla każdego nagranego zdania.

SNR oszacowano jako: sygnał mowy + szum – szum tła. Bardziej poprawnym sposobem byłoby: poziom mowy - poziom szumów, jednak nie ma możliwości zmierzenia poziomu mowy bez uwzględnienia szumów. Oznacza to, że gdy poziom mowy jest zbliżony do poziomu szumów, dokładność SNR maleje.

Lokalizacja i poziom szumów tła:	Na zewnątrz przy hałasie ulicznym 61–68 dBSPL[A]	W cichym korytarzu 32 dBSPL[A]	W dużym holu 42 dBSPL[A]	Wolna przestrzeń, brak odbić, całkowita cisza
Odległość (m)	**SNR [dB] (szacowana wartość)**	**SNR [dB] (szacowana wartość)**	**SNR (szacowana wartość w [dB])**	**Teoretyczny SNR = poziom mowy – poziom szumów mikrofonu [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	Niesłyszalne	17,5	6.3	7,9
24	Niesłyszalne	14,8	5,4	4,4
32	Niesłyszalne	Korytarz zbyt krótki	4,6	1,9
40	Niesłyszalne	Korytarz zbyt krótki	4,7	0,0

Teoretycznie przy podwojeniu odległości poziom mowy powinien zmniejszyć się o 6 dB. Oznaczałoby to również, że SNR powinien odpowiednio spaść. Z tabeli wynika, że SNR w korytarzu i w holu jest lepszy od wartości oczekiwanej. Powodem jest fakt, że odbicia dźwięku w pomieszczeniu spowodowały wzrost poziomu ciśnienia akustycznego. Zwłaszcza w korytarzu ściany, podłoga i sufit pełnią dla dźwięku rolę falowodu. Dzięki temu poziom głośności i SNR są o około 10 dB lepsze niż teoretyczny SNR w wolnej przestrzeni, przy pomiarze przeprowadzonym w korytarzu z odległości 24 m.

Gdy poziom szumów tła przekracza akustyczny ekwiwalent szumu własnego pojedynczego mikrofonu dookólnego, zasięg zbierania dźwięku zależy wyłącznie od poziomu hałasu otoczenia, a nie od samego mikrofonu.

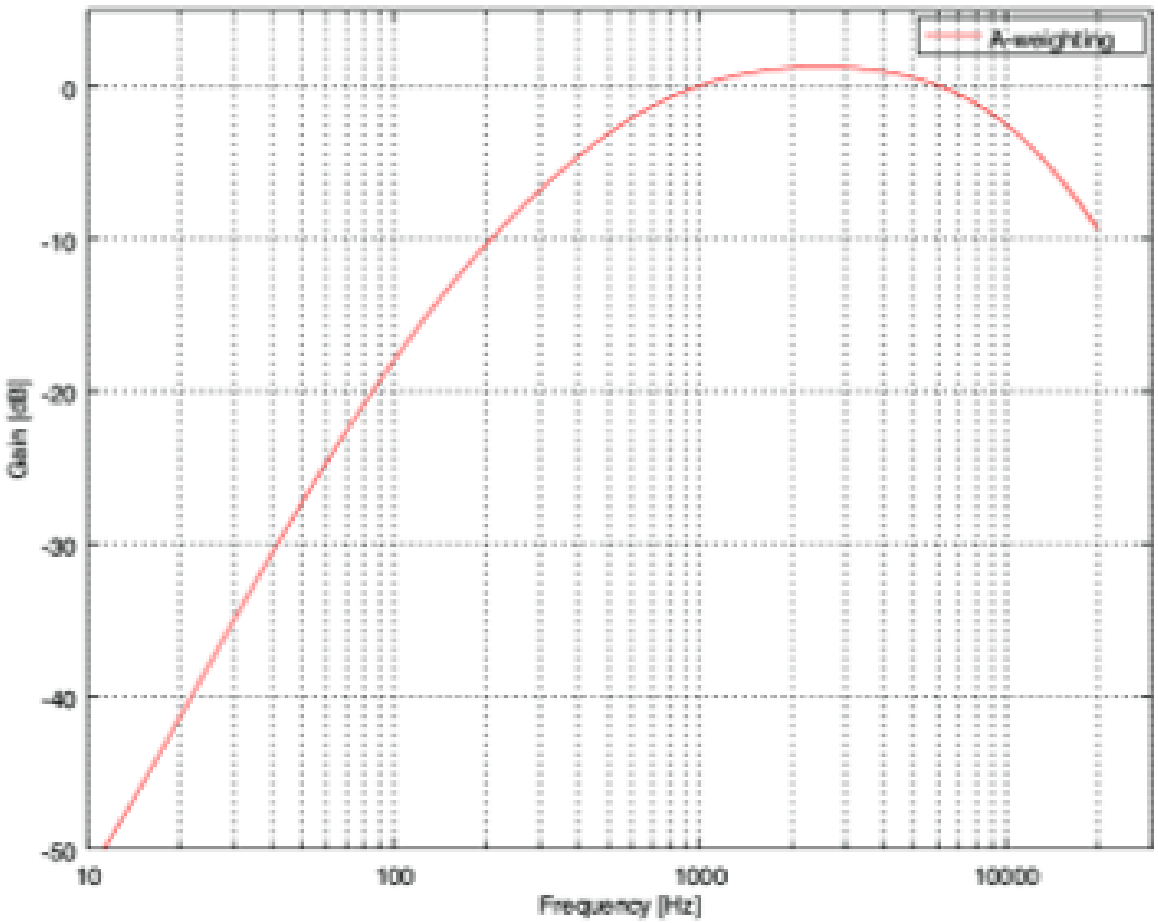
Dodatek 1 Dodatek

- **Poziom ciśnienia akustycznego w decybelach (dBSPL):** poziom ciśnienia akustycznego wyrażony w decybelach (dB) w odniesieniu do wartości 20 µPa. Podczas pomiaru poziomów szumów i głosu często stosuje się uśrednione w czasie wartości skuteczne (RMS). Parametr ten jest czasami nazywany poziomem mocy.

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

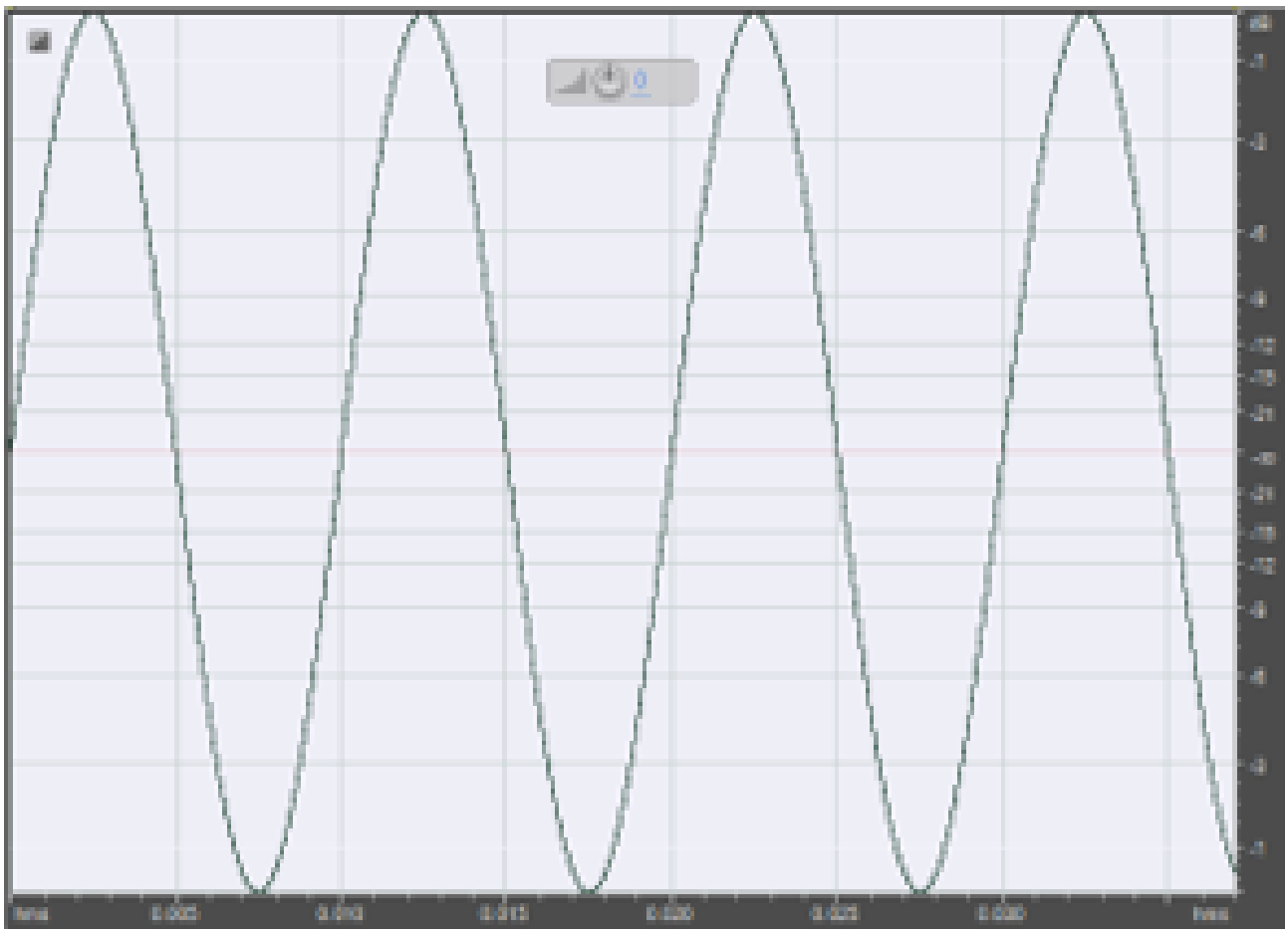
Dwa najczęściej stosowane czasy to „szybki RMS" z oknem czasowym wynoszącym 125 ms oraz „wolny RMS" z oknem czasowym wynoszącym 1 sekundę. W przypadku analizy podłogi szumowej często oblicza się średnią w dłuższym przedziale czasowym.

- **Poziom ciśnienia akustycznego skorygowany z użyciem filtra korekcyjnego A (dBspl[A]):** czasami zapisywany jako dB(A). Jest to wartość poziomu dBSPL z korekcją A. Korekcja A to metoda filtrowania sygnału z mikrofonu mająca na celu lepsze dostosowanie zmierzonego poziomu ciśnienia akustycznego do ludzkiego słuchu. Wynika to z faktu, że ludzkie ucho jest mniej czułe na częstotliwości bardzo niskie (poniżej 500 Hz) oraz wysokie (powyżej 10 kHz).



Ilustracja .1 *Filtr korekcyjny A.*

- **Szczytowa wartość w skali pełnej w decybelach (dBFSpeak/dBFSp)** odnosi się do amplitudy sygnału audio w formacie cyfrowym w odniesieniu do maksymalnej dostępnej wartości w reprezentacji cyfrowej (pełna amplituda). Sygnał sinusoidalny o maksymalnej amplitudzie cyfrowej bez przesterowania ma wartość szczytową równą 0 dBFSp.



Ilustracja .2 *Przykład sygnału o amplitudzie 0 dBFSpeak.*

- **dBFS RMS** określa skuteczny poziom cyfrowego sygnału audio. Sygnał sinusoidalny o pełnej amplitudzie i wartości 0 dBFSpeak może charakteryzować się różnymi poziomami RMS w zależności od definicji jednostki. W kartach katalogowych mikrofonów sygnał sinusoidalny o pełnej amplitudzie definiuje się jako 0 dBFS RMS i z tego samego względu definicję tę stosujemy w niniejszej białej księdze. Ogólnie rzecz biorąc, druga definicja sygnału sinusoidalnego o pełnej amplitudzie jako wynoszącego -3 dBFS RMS jest bardziej poprawna z matematycznego punktu widzenia i bardziej spójna z definicjami poziomów RMS i poziomów szczytowych w dziedzinie analogowej. Zgodnie z tą drugą definicją sygnał prostokątny o pełnej amplitudzie wynosi 0 dBFS RMS.
- **Czułość mikrofonu** to poziom jego sygnału wyjściowego, gdy na wejście akustyczne podawany jest sygnał sinusoidalny o częstotliwości 1 kHz i ciśnieniu akustycznym wynoszącym 1 Pa, 0 dB a lub 94 dBSPL (są to trzy sposoby wyrażenia tego samego poziomu ciśnienia akustycznego).
- **Stosunek sygnału do szumów (SNR):** gdy poziomy audio są wyrażane w dB, stosunek ten oblicza się jako różnicę między poziomem sygnału a poziomem szumów.
- **SNR mikrofonu** jest obliczany jako różnica między poziomem wyjściowym między sygnałem testowym o częstotliwości 1 kHz i poziomie 94 dBSPL a szumem własnym mikrofonu w warunkach całkowitej ciszy. Typowy wskaźnik SNR dla cyfrowego mikrofonu MEMS wynosi 69 dBSPL[A].
- W niniejszej białej księdze **SNR mowy** oznacza stosunek poziomu sygnału do poziomu szumów między mową ludzką a poziomem szumów z korekcją A w sytuacji, gdy nie występuje mowa. Wskaźnik SNR mierzy się w decybelach. Przy obliczaniu teoretycznego zasięgu zbierania mowy przyjmujemy SNR mowy wynoszący 0 dB jako definicję najniższego akceptowalnego SNR. W tym kontekście szumy mogą pochodzić zarówno z hałasu otoczenia i zakłóceń elektrycznych, jak i z szumu własnego mikrofonu.
- **Poziom szumów mikrofonu**. Cyfrowy poziom szumów mikrofonu mierzy się w dBFS RMS. Najlepsze mikrofony cyfrowe mają zazwyczaj teoretyczną podłogę szumową wynoszącą -105 dBFS RMS[A].

- **Akustyczny ekwiwalent szumu mikrofonu** można obliczyć jako różnicę między poziomem testowym wynoszącym 94 dBSPL a SNR mikrofonu. Typowym przykładem może być tutaj 94 – 69 = 25 dBSPL[A].

O firmie Axis Communications

Axis wspiera rozwój inteligentnego oraz bezpiecznego świata przez poprawę bezpieczeństwa, ochrony, efektywności działania i wiedzy biznesowej. Jako firma zajmująca się technologiami sieciowymi oraz lider branży, Axis oferuje rozwiązania z zakresu dozoru wizyjnego, kontroli dostępu, systemów domofonowych i systemów audio. Ich rozszerzeniem i uzupełnieniem są inteligentne aplikacje analityczne oraz wysokiej jakości szkolenia.

Axis zatrudnia około 5000 pracowników w ponad 50 krajach oraz współpracuje z partnerami z obszaru technologii i integracji systemów na całym świecie w celu dostarczania swoich rozwiązań klientom. Firma została założona w 1984 roku i ma swoją siedzibę w Lund w Szwecji.

© 2026 — Axis Communications AB. AXIS COMMUNICATIONS, AXIS, ARTPEC i VAPIX są zarejestrowanymi znakami towarowymi firmy Axis AB w różnych jurysdykcjach. Wszystkie inne znaki towarowe są własnością odpowiednich podmiotów.Zastrzegamy sobie prawo do wprowadzania zmian bez uprzedniego powiadomienia.