Application Note:

Audio performance requirements for Audio Front End of i.MX RT106A/L/S and i.MX RT105S

Document information	
Title	Audio performance requirements for Audio Front End of i.MX RT106A/L/S and i.MX RT105S
Status	
Version	1.5
Date	April 16, 2021
Author	frederic.kubelec@nxp.com
Reviewers	Laurent.pilati@nxp.com,
Document ID	
Security	0 The attached material and the information contained therein are proprietary to NXP Software and are issued only under strict confidentiality arrangements. Without a separate written authorization, it shall not be used for any purpose whatsoever, reproduced, copied in whole or in part, adapted, modified, or disseminated. It must be returned to NXP Software upon its first request.
Usage	**NXP**

Change History

Version	Status	Description	Author	Date
0.5	DRAFT	Reworked Version	Frederic Kubelec	2019-04-18
0.6	DRAFT	Review	Laurent Pilati	2019-04-23
1.1		Last Version	Frederic Kubelec	2019-04-23
1.2		Reworked	Rick Bye	2019-05-08
1.3		Added RT106L	Rick Bye	2020-02-21
1.4		Fixed broken link in 5.1.1.1	Rick Bye	2020-06-01
1.5		Added i.MX RT106S & RT105S	Rick Bye	2021-04-16

Table of content

Contents

1. Introduction

Goal of this document is to:

- Describe the performance requirements at the input of the audio front end (AFE) for the i.MX RT106A/L/S and i.MX RT105S, and provide measured example results.

2. Reference & Abbreviations

References	
[1]	ITU-T P.341 : "Transmission characteristics for wideband digital loud speaking and hands-free telephony terminals"
[2]	IEEE1329 : IEEE Standard Method for Measuring Transmission Performance of Speakerphones
[3]	VCB_AudioPerformanceMeasurement_protocol.doc

Table 1: Reference documents

Abbreviations	
AEC	Acoustic Echo Cancellation
AFE	Audio Front End
AOP	Acoustic Overload Point
AVS	Alexa Voice Service
DUT	Device Under Test
DR	Dynamic Range
ERL	Echo Return Loss
ERLE	Echo return Loss Enhancement
FE	Front End Audio
PCM	Pulse-Code Modulation
PDM	Pulse Density Modulation
SLR	Sending Loudness Rating
SNR	Signal to Noise Ratio
SPL	Sound Pressure Level
VCB	Voice Control Board

Table 2: Abbreviations

3. System definition

The following parameter definitions explain how to define a microphone input audio path, from an acoustic signal to the N-bits of digital data that will be processed by the AFE.
That audio chain can be split as following:

1. Acoustic signal acquisition by a digital microphone;
2. Filtering (decimation filter for PDM data);
3. Digital gain before the input of the front end processor.

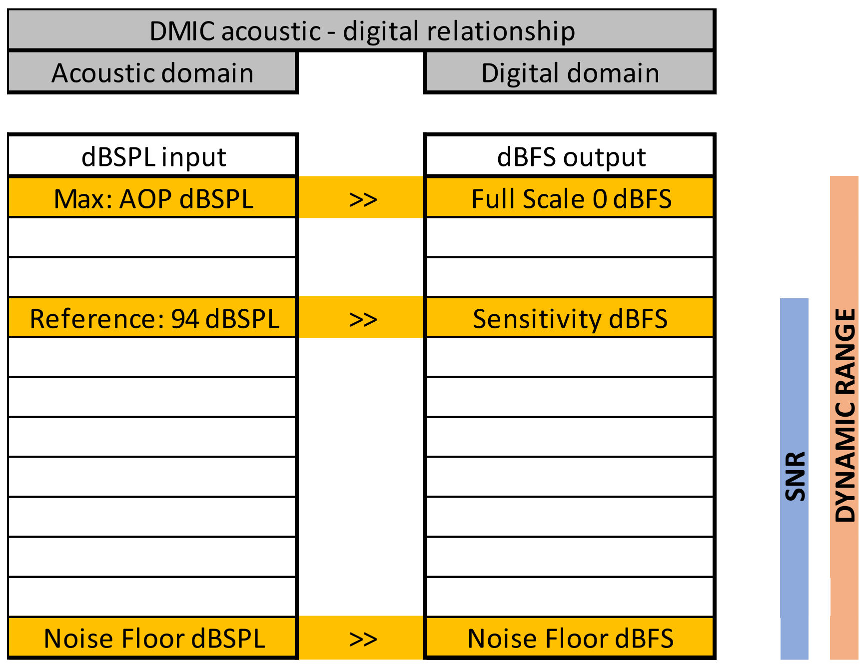


Figure 1: DMIC parameters: from dBSPL input to dBFS output (without gain)

3.1. Sensitivity (SLR)

The digital microphone's acoustic input levels in dB SPL are RMS measurements, the output of digital microphones is referenced to dB FS (full scale) which is a peak value.

The sensitivity specification should be used to match the microphone's output signal level across the dynamic range of interest to the signal level of the audio chain.

Microphone sensitivity is measured with a 1 kHz sine wave at a 94 dB SPL sound pressure level (1 pascal (Pa) pressure). Sensitivity of digital microphones depends on the maximum acoustic overload point (AOP): The sensitivity is the difference between the 94 dB SPL reference and the AOP. So, if a digital microphone's AOP is 120 dB, then its sensitivity will be −26 dB FS (94 dB – 120 dB).

Sensitivity is not the only criteria define the quality of a microphone. Other metrics such as SNR, dynamic range and THD are also important.

3.2. SNR

The challenge is to choose an AOP that keeps some headroom at maximum speaker playback volume, while keeping good SNR for far field voice recording.

The i.MX RT106A/L/S and i.MX RT105S use pulse density modulation (PDM) format digital MEMS microphones. Internally, the single-bit PDM format is converted into 16 bits PCM format at a 16 kHz sample rate.

SNR is the difference between the microphone output for a 1 kHz sine wave at 94 dB SPL and the microphone output without any acoustic input at the microphone (noise floor).
For Knowles SPH0641 microphones, SNR is 64.3 dB.

3.3. Dynamic Range

The microphone's dynamic range is the difference between the acoustic overloading point (AOP) and the noise floor.
Since the system is using 16 bits, the AOP of the microphone should be chosen so that there is no saturation at maximum speaker volume level. This will guarantee the best SNR for far field speech recording.

Note: Clipping is not allowed for Audio Front End processing.

3.4. Echo path

Acoustic Echo Cancellation (AEC) is required for the barge-in use case. A key parameter for the acoustic form factor is the Terminal Coupling Loss, which is translated in echo return loss (ERL) in the digital domain.

This is a loss in signal strength (usually positive value) between the original signal ,*S,* (playback from an integrated loudspeaker) and the echo, *E,* of that signal transmitted to the microphone path. Ideally, it should be as high as possible, leading to a small echo (Industrial Design requires a good form factor to minimize physically coupled echo).

The performance of an echo canceller is measured in echo return loss enhancement (ERLE), which is the amount of additional loss applied by the echo canceller.

Different echo sources on DUT:

- Acoustic coupling from loudspeaker to microphone: Direct sound from the loudspeaker to the microphones.
- Physical coupling (vibrations of the loudspeaker transfer to the microphone via the casing).

The most challenging barge-in scenario is when playing loud audio on the integrated loudspeaker in a far field application (meaning strong echo), while listening for a distant user's voice (meaning low speech level at microphone):

Tuning process:

- Find the maximum speaker playback level so that the digital microphone input remains below its AOP, because clipping is not allowed for Audio Front End processing;
- Define the margin to avoid reaching the AOP:
    - Based on Alexa Voice Service (AVS) 1.9 specification, maximum average speaker output level is 90 dBc while playing the AVS music pattern;
    - For this music playback level, keep at least 3 dB margin from the microphone AOP;
- Check the saturation and voice recognition performance, find a good balance between microphone gain and playback level.

4. i.MX RT106A/L/S Voice Control Development Kits

A voice control development kit (SLN-ALEXA-IOT, SLN-LOCAL-IOT or SLN-LOCAL2-IOT) development kits was used as the device under test (DUT). (All kits share an identical hardware design, the only difference is the MCU: i.MX RT106A, i.MX RT106L, or i.MX RT106S.)

4.1. Digital Microphone part number

Knowles SPH0641LM4H-1 digital microphones (DMIC) are used on the voice dev kit:

KNOWLES SPH0641LM4H-1

Standard Performance Mode

TEST CONDITIONS: f_{CLOCK} = 2.4 MHz, V_{DD}=1.8 V, unless otherwise indicated

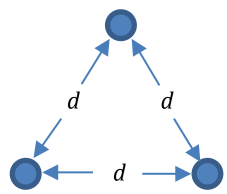
Parameter	Symbol	Conditions	Min	Typ	Max	Units
Supply Current[1,2,3]	I_{DD}	f_{CLOCK}=2.4 MHz	-	620	700	µA
Sensitivity[1]	S	94 dB SPL @ 1 kHz	-27	-26	-25	dBFS
Signal to Noise Ratio	SNR	94 dB SPL @ 1 kHz, A-weighted, f_{CLOCK}=2.4 MHz	-	64.3	-	dB(A)
Total Harmonic Distortion	THD	94 dB SPL @ 1 kHz, S = Typ	-	0.2	-	%
Acoustic Overload Point	AOP	10% THD @ 1 kHz, S = Typ	-	120	-	dB SPL
Power Supply Rejection Ratio	PSRR	200 mVpp sinewave @ 1 kHz	-	55	-	dBV/FS
Power Supply Rejection	PSR+N	100 mVpp square wave @ 217 Hz, A-weighted	-	-84	-	dBFS (A)

Table 3: Voice development kit microphone specification extract

On the i.MX RT106x voice development kits, the distance between the DMICs and loudspeaker is about 20 mm (DMIC to center of the speaker)

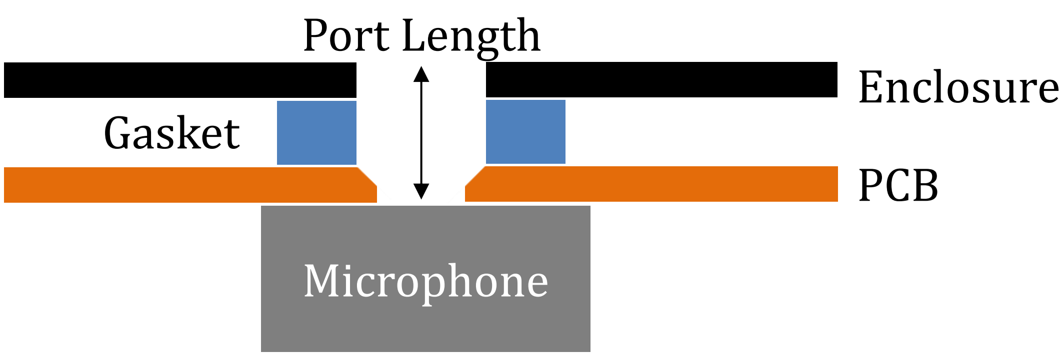
5.2 Microphone topology

For a three-microphone array, the microphones must be arranged as an equilateral triangle and can be spaced anywhere from 25 to 80 mm apart. A three-microphone array can be used for applications that require partial or a full 360-degree sound field.



5.3 Microphone Port Length

The port length is the distance between the hole of the microphone and top of the surface facing the outside world (typically the surface of the product enclosure). As the port length increases, the Helmholtz resonance gets closer to the range required for voice recognition. If this resonant frequency gets too close to the voice range, it will cause major issues.



The formula for the Helmholtz resonance frequency is the following:

$$f = \frac{c * D}{4\sqrt{\pi * V * \left(L + \sqrt{\pi} * {}^{D}/_{2}\right)}}$$

c = Speed of sound
D = Diameter of the hole/port
V = Volume of the port
L = Length of the port

For the Knowles SPH0641LM4H-1 microphones the hole diameter is 1.5 mm at the top of the PCB and 0.88 mm after the beveled edge.

5. Audio test reference signals

5.1. Sensitivity/SNR/THD

1 kHz frequency, level = -3 dB FS

5.2. Frequency response

Pink noise, level = -3 dB FS

5.3. Aliasing

Frequency sweep from 8.5 kHz to 16 kHz, amplitude = -6 dB FS

5.4. Directivity

Pink noise, level = -3 dB FS

5.5. Echo path

Pink noise, level = -3 dB FS

6. Audio recording set up

6.1. Audio laboratory

The noise floor should be below 30 dB(A), with average RT60 about 300ms.

6.2. Acoustic Material

Loudspeaker frequency response should be flat (this is the case for a monitor loudspeaker). If this is not the case, equalization is be needed to compensate the non-ideal frequency response. As an example, Genelec 8020D active loudspeakers and the Roland Octa-Capture sound card for audio playback, were used in following sections.

6.3. Speaker and DUT setup

The DUT and loudspeaker should be placed at an elevation that fits a typical use case scenario. The DUT and the loudspeaker are expected to be on the same horizontal plane. Typically, the loudspeaker stands are fixed at a height of 1 m.

6.4. Calibration

In the following sections, it is expected to play a reference signal at 74 dB SPL and 94 dB SPL at DUT.

7. Audio Measurements

This section provides the audio measurements of the i.MX RT106x voice development kit. They should be taken as typical targets for any new form factor design.

7.1. Frequency response (Front End input)

7.1.1. Test conditions

Choose the 74 dB setting. Play the pink noise pattern from the reference files and record the DUT and reference microphone.

7.1.2. Results

Voice dev kit microphone frequency response relative to reference microphone for74dB condition and standard:

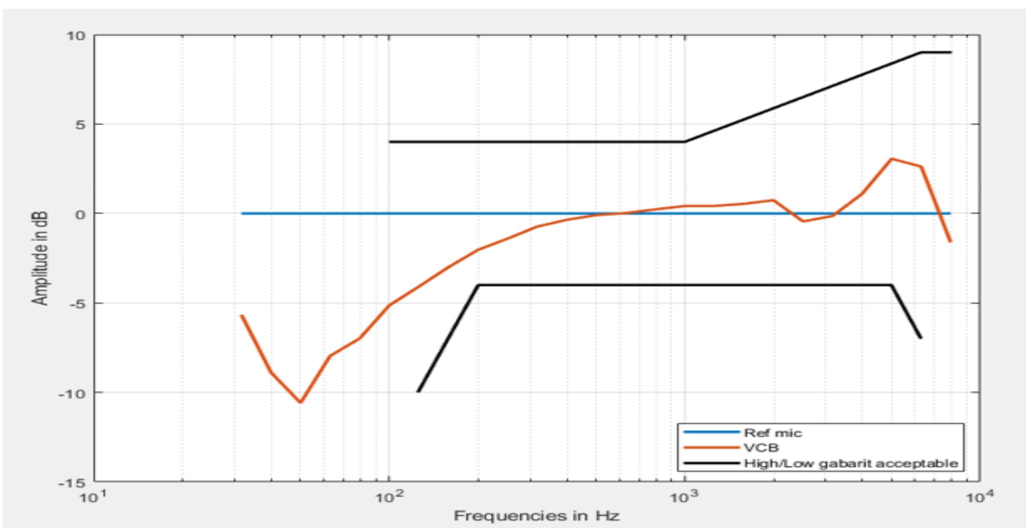


Figure 2: Compensated frequency response for voice development kit ("VCB")

Make sure the frequency response is within the ITU-T P.341 Send Frequency Mask. (Please refer to Appendix : ITU-T P.341 Mask)

7.2. Sensitivity (SLR) (Front End input)

7.2.1. Test conditions

Choose the 94 dB setting.
Play the 1 kHz sine wave on the loudspeaker and record the DUT and reference microphone.
Measured parameter: fundamental F0 level.

7.2.2. Results

Device	Measured F0 level 94 dB SPL setting
Voice dev kit	-19.7 dB FS

Table 4: Device sensitivity

Note: digital output level depends on additional gain in the audio path.

7.3. Noise floor (Front End input)

7.3.1. Test conditions

Using the reference microphone record the test room noise floor in silence conditions. Record the DUT in the same condition for a duration of 10 seconds. Compute the RMS value to measure the noise floor.

7.3.2. Results

Device	Average RMS Amplitude
DMIC voice dev kit	-78.8 dB FS
REF MIC	-83 dB FS

Table 5: Noise floor at DUT output

7.4. SNR (Front End input)

7.4.1. Test conditions

Choose the 94 dB setting.
Play the 1 kHz sine wave on the loudspeaker and record the DUT and reference microphone.

7.4.2. Results

SNR result can be cross-checked from Sensitivity and Noise results:
In 94 dB condition: SNR = -19.7 dB + 78.8 dB = 59.9 dB

Note: In order to avoid room low frequency noise, audio files were pre-filtered by a high-pass filter of 100 Hz. SNR was calculated on filtered records.

level at DUT	SNR at DMIC PATH OUT
94 dB SPL	60 dB

Table 6: SNR measurement

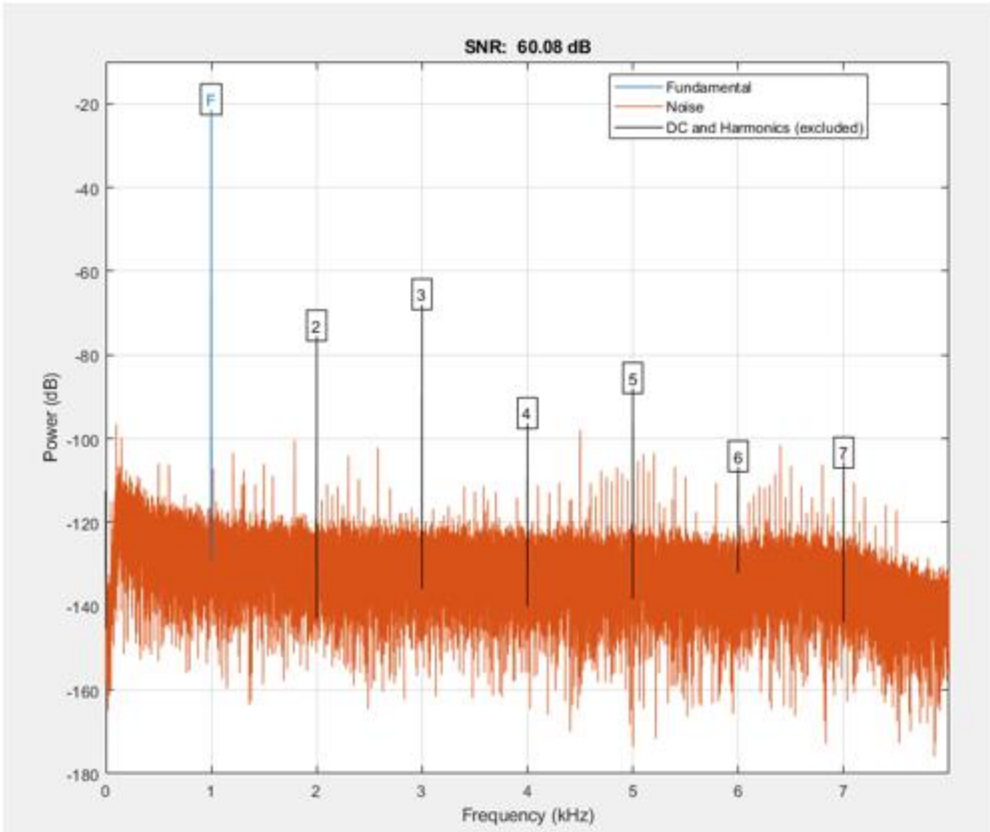


Figure 3: SNR for 94 dB SPL level and 1000 Hz input signal

7.5. Distortion (THD) (Front End input)

7.5.1. Test conditions

Choose the 94 dB or 74 dB setting.
Play the 1 kHz sine wave from the reference files on the loudspeaker and record the DUT and reference microphone.

7.5.2. Results

Note: If needed: To avoid room low frequency noise audio files were pre-filtered by a high-pass filter below 100 Hz (70 dB attenuation). THD was calculated on filtered records.

level at DUT	THD at AUDIO PATH OUT
94 dB SPL	-45.9 dB/0.5%
74 dB SPL	-47.1 dB/0.44%

Table 7: Distortion DUT / Reference Microphone

Note: THD computation can be done with MatLab `thd` function.

7.6. Directivity (Front End input)

7.6.1. Test conditions

0-degree reference to be determined;
Angle reference as described in the picture: step = +30 degrees clockwise.
Play the pink noise pattern;
Rotation step = 30 degrees between 2 successive recordings;

For each recording, user should extract the power spectrum of the following frequencies: 250 Hz, 1 kHz and 4 kHz.

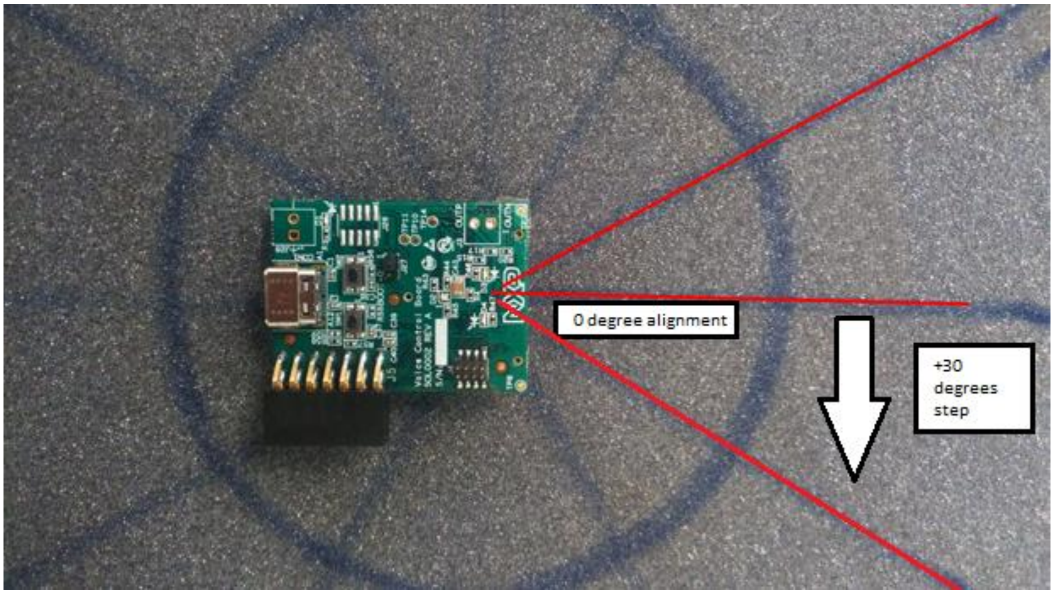


Figure 4: Angle reference

7.6.2. Results

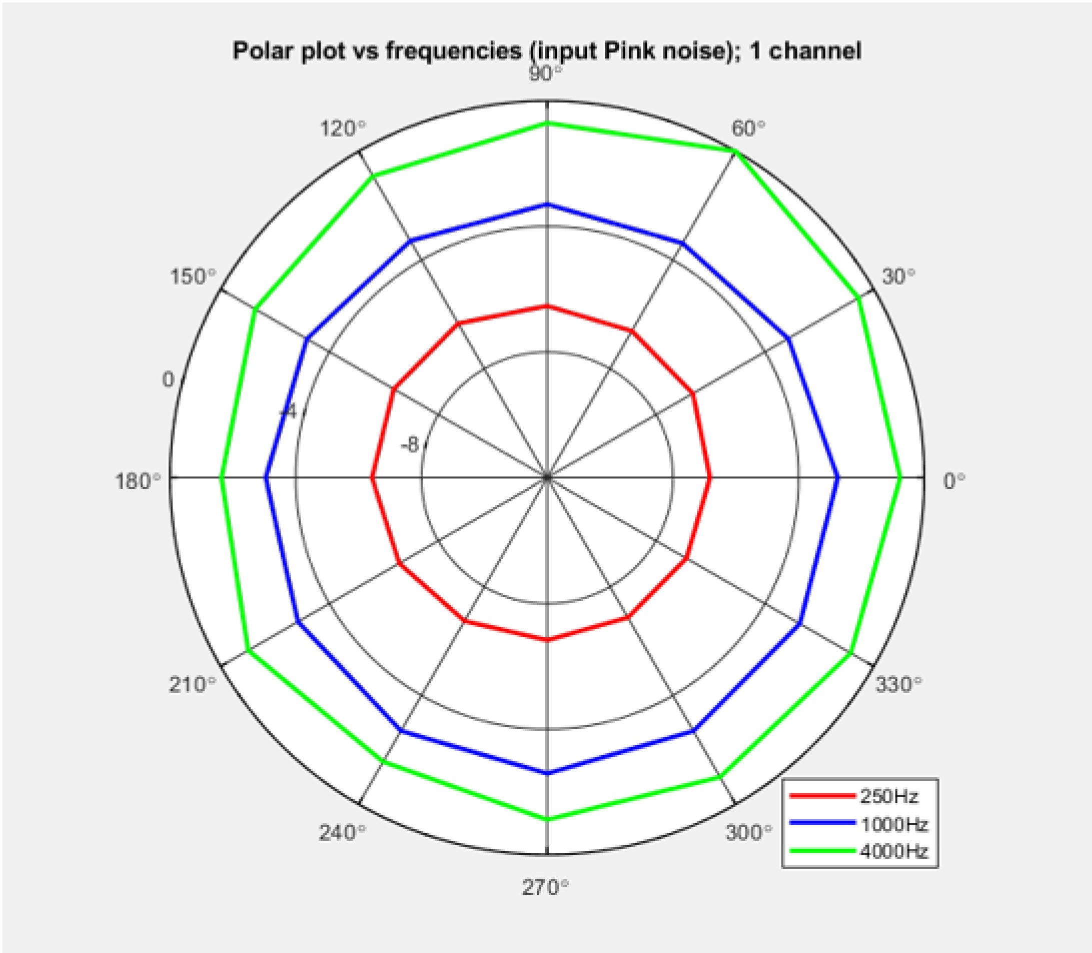


Figure 5: Directivity polar plot of voice development kit (Center MIC)

Microphones are omnidirectional without any directivity issue (ripple is below 2 dB). Attenuation of 250 Hz and 1 kHz for bottom left microphone is observed (respectively -3 dB and -6 dB) but still with ripple below 2 dB which should not impact the directivity algorithm.

7.7. Echo Path: ERL

7.7.1. Results

RMS Level echoref	RMS Level MIC3 FE input
-33.4 dB	-40.7 dB

Table 8: Echoref, FE input in Barge-In use case

Measured value on the voice development kit, MIC3 (center microphone):
Attenuation between Front End input and echoref:

- ERL = 7.3 dB

7.8. Summary table of typical metrics

Typical values at the audio front end input (id. full audio path from microphone to audio front end input):

Parameter	Typic value or range
Sensitivity (dB FS)	-20
SNR (dB):	> 60
Noise Floor (dB FS)	< -80
THD (%)	<0.5
Frequency response	Derivative of ITU-T P.341 limits
Directivity	Ripple < 3 dB
ERL	22 dB

Table 9: Typical values

8. Appendix : ITU-T P.341 Mask

Reference to Section 5.1.1.1 :

Table 6 – Sending sensitivity/frequency mask

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	4	
125	4	–10
200	4	–4
1000	4	–4
5000	(Note)	–4
6300	9	–7
8000	9	
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a logarithmic (frequency) – linear (dB) scale.		

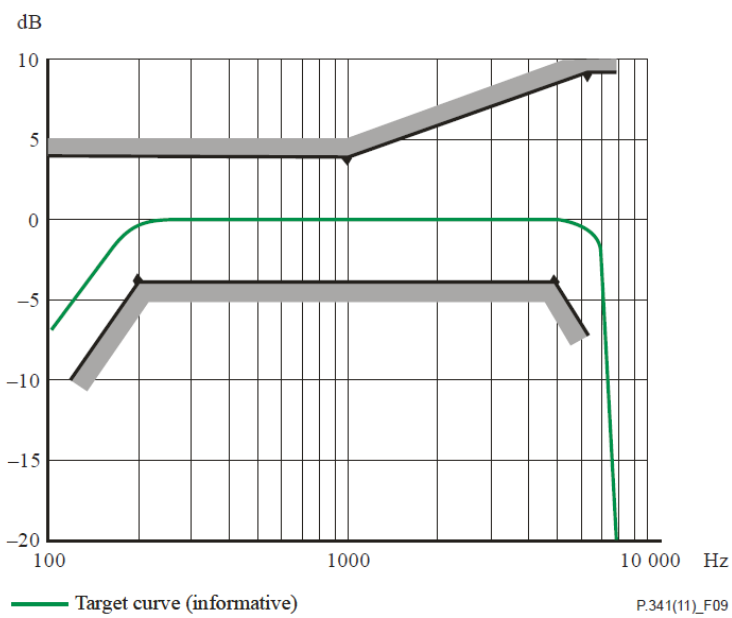


Figure 9 – Hands-free sending sensitivity/frequency mask