Source: HEAD acoustics GmbH

Title: Test method for stereo capture

Document for: Agreement

# Introduction

The latest version of Pdoc ATIAS-1 [1] already contains an initial proposal for stereo capture in send direction, which already addresses two important measures of inter-channel time/level difference. However, several details regarding test method, test signal, level, analysis, etc. are still missing. For the completion of this test method, a test series was conducted.

# Test setup

The test setup of the measurement series is illustrated in Figure 1 and corresponds to the handheld hands-free UE as described in clause 2.1.4 of ATIAS-1 [1]. In order to simulate a stereo-capable UE, a stereo microphone (configurations with either X-Y or A-B setup) was positioned at a distance of 42 cm in front of the head-and-torso-simulator (HATS), which is equipped with a mouth simulator according to ITU-T P.58 [2].

To realize different angles of the talker, the microphone was mounted on a turntable and rotated around its vertical center axis from -90° to +90° in steps of 15°. The position of 0° corresponds to the default position, i.e., HATS is facing the microphone/UE.

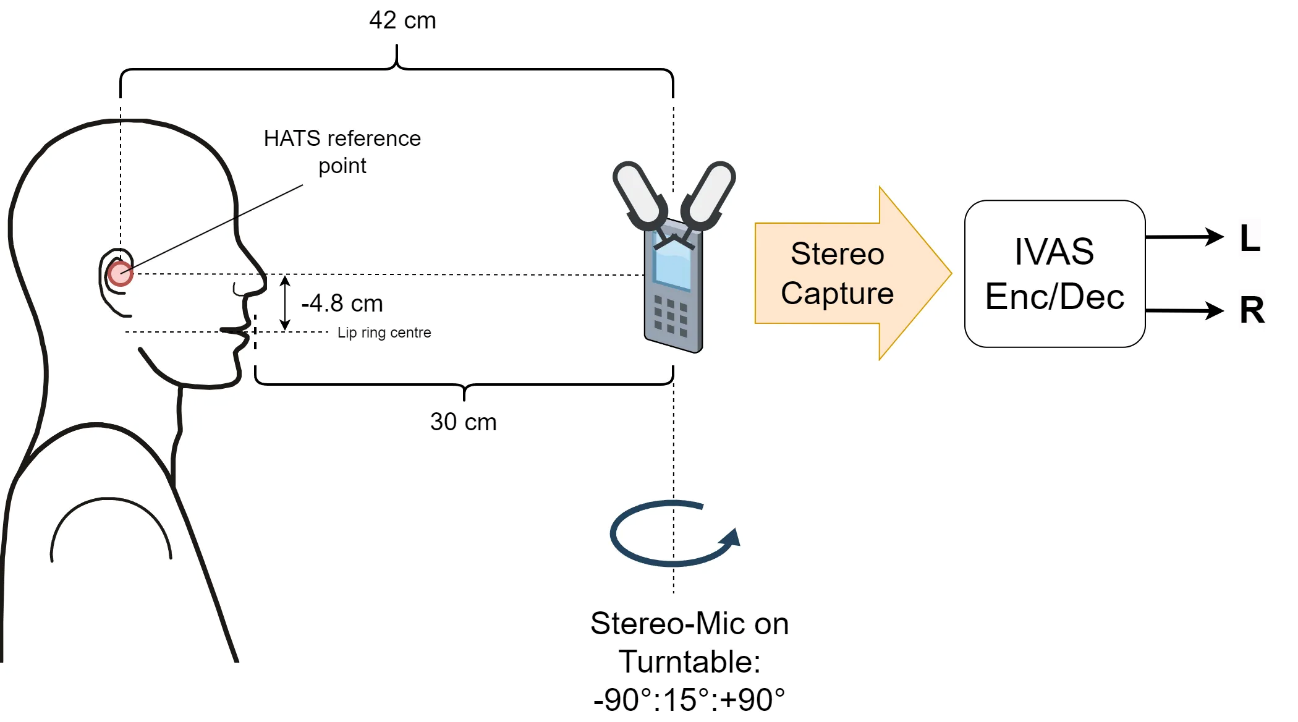


Figure 1: Test setup for stereo capture

As a test signal, the British-English single talk sequence as described in clause 7.3.2 of ITU-T P.501 [3] was calibrated to an active speech level [4] of -1.7 dBPa at the mouth reference point (MRP).

To apply a typical/reasonable sensitivity, the captured stereo signals at 0° position (left and right channel almost identical) were calibrated to -26 dBov active speech level [4]. This calibration factor was then used to calibrate the recordings of each rotation. For each microphone configuration (A-B, X-Y), the calibration factor was determined again.

The calibrated stereo signals were then passed to the IVAS encoder/decoder to simulate a transmission at all available bitrates (from 13.2 to 160 kbit/s).

# Analysis

## Inter-channel time difference (ICTD[[1]](#footnote-1))

ICTD is determined via the envelope of the segmental cross-correlation function (segment , lag ) between left and right channel signals () by means of the Hilbert transformation:

The segment duration T equals 8192 samples and an overlap of 50% was used. The ICTD is then determined as the time lag of the averaged envelope, that provides the maximum value:

## Inter-channel level difference (ICLD)

In general, ICLD is calculated as the level difference between left and right channel:

To determine and , different level calculation methods were considered:

* Active speech level (ASL) according to ITU-T P.56 [4],
* Delta-Level versus time analysis according to [5] with a time constant of 35 ms. Level aggregation is carried out by determining 15% and 85% percentiles of the resulting level difference versus time (denoted as *Perc85/15* in the following). The principle is illustrated in Figure 2.

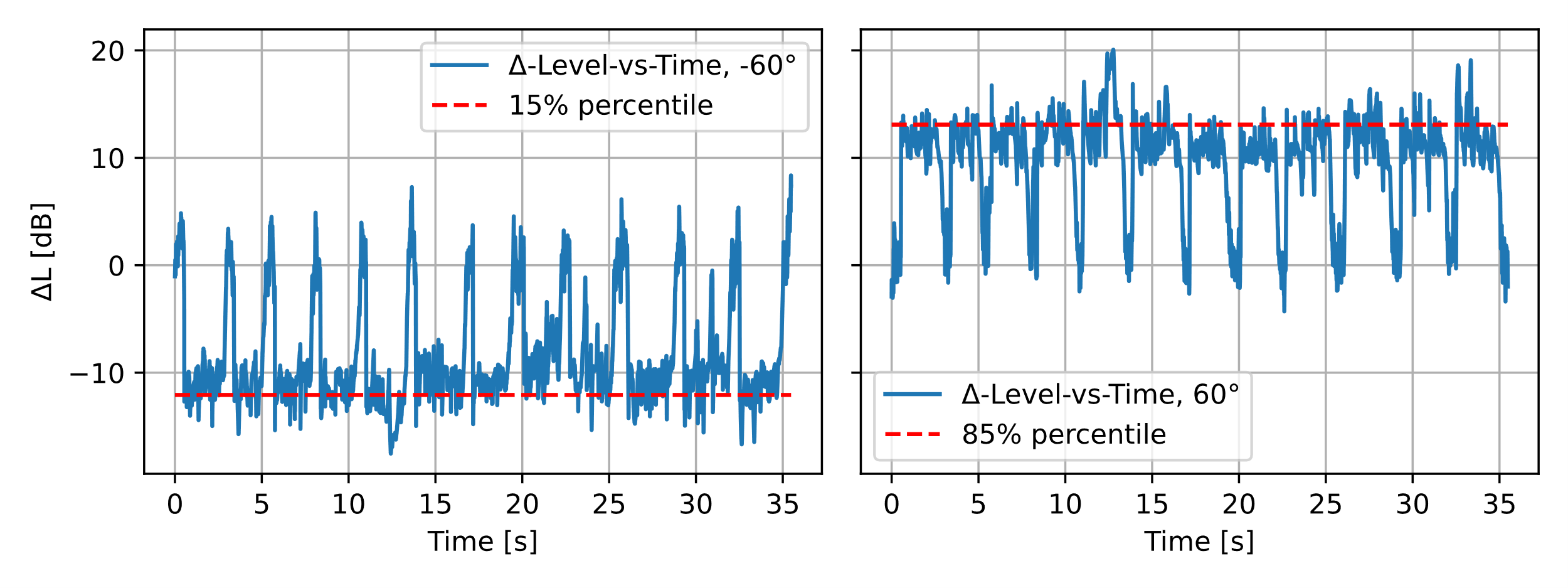


Figure 2: Examples of level-vs-time percentile analysis

# Test results

## ICTD

### Uncoded

The results for ICTD calculation and without any subsequent coding are shown in Figure 3. As expected, the A-B configuration provides a better resolution in time than X-Y. The distance between left and right microphone is approximately 11 cm, the minimum/maximum values of ~0.32 ms also seem to be reasonable.

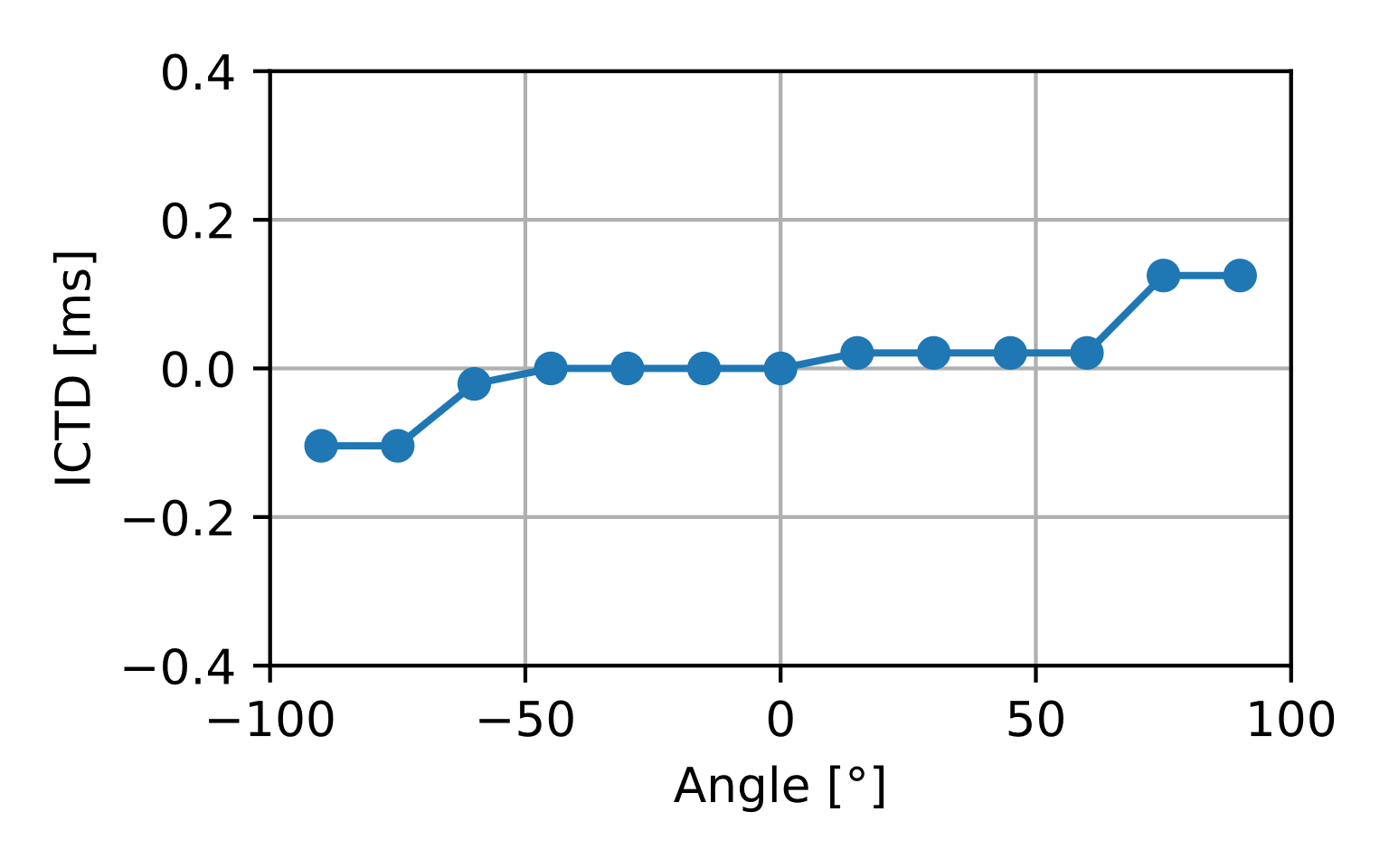
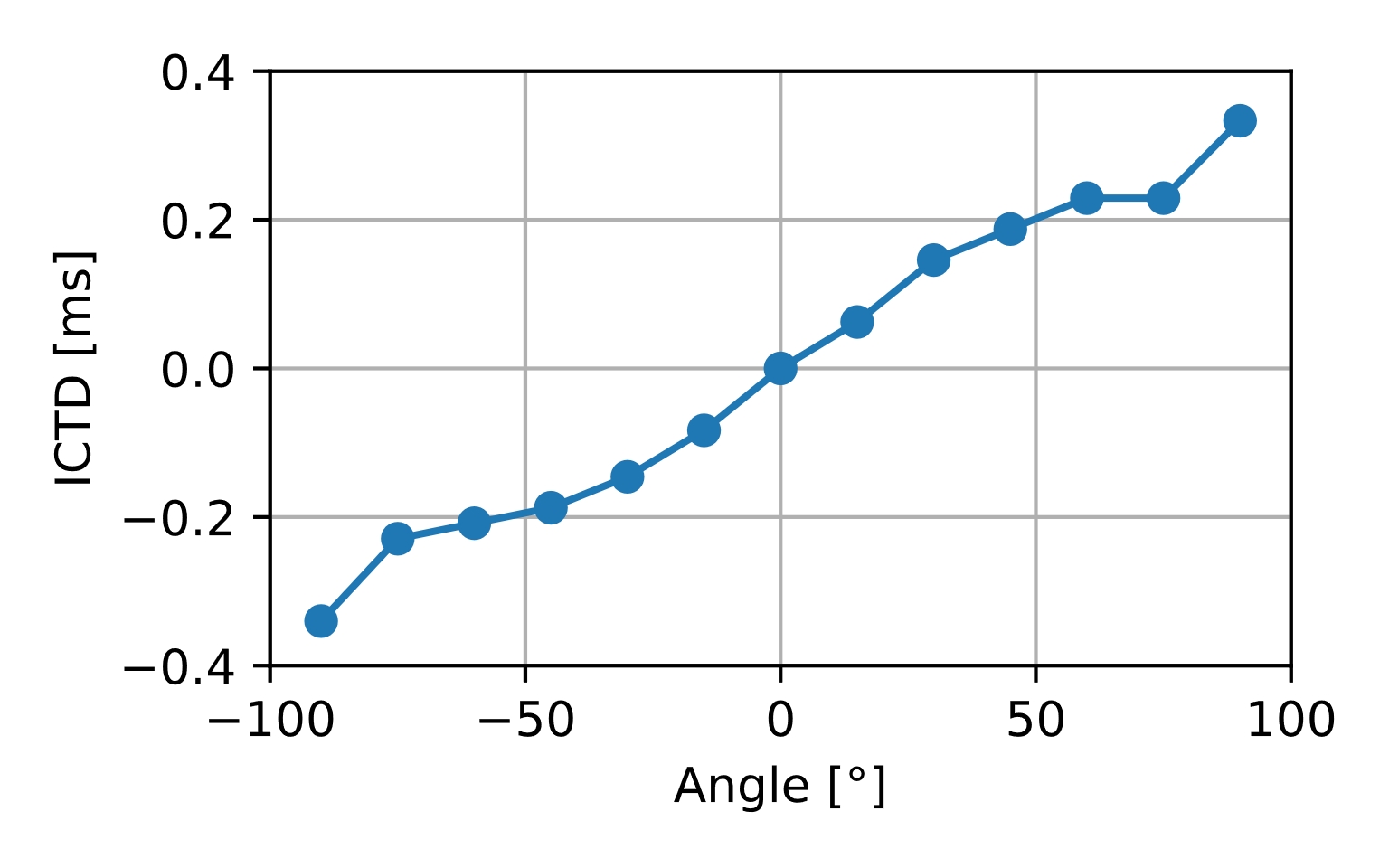


Figure 3: ICTD results versus angle, for A-B (left) and X-Y (right) configuration

### IVAS stereo mode

The calibrated stereo speech recordings were offline encoded/decoded with IVAS codec at different bitrates in stereo output format (latest floating-point version). The results for the corresponding ICTD analysis are shown in Figure 4.

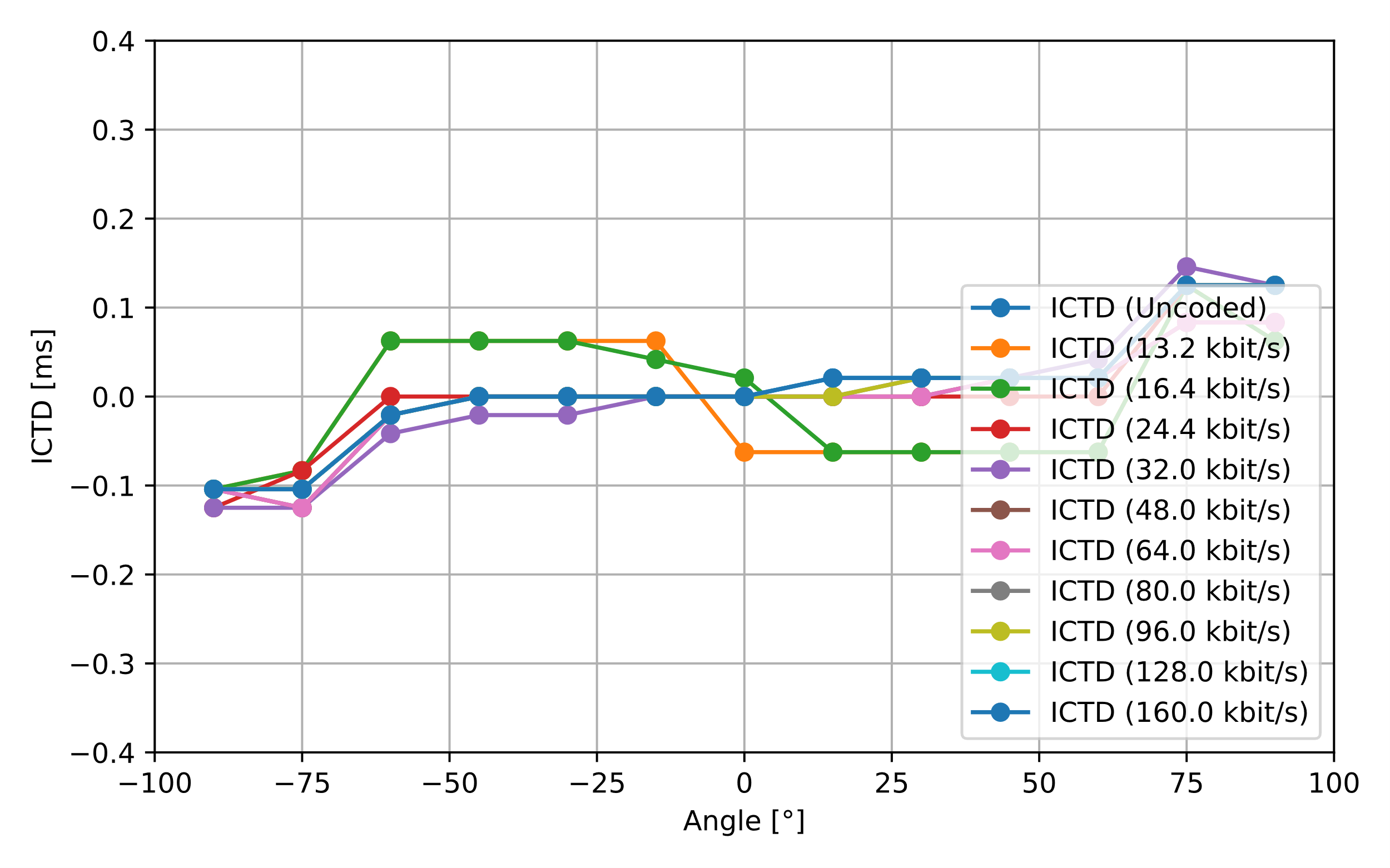
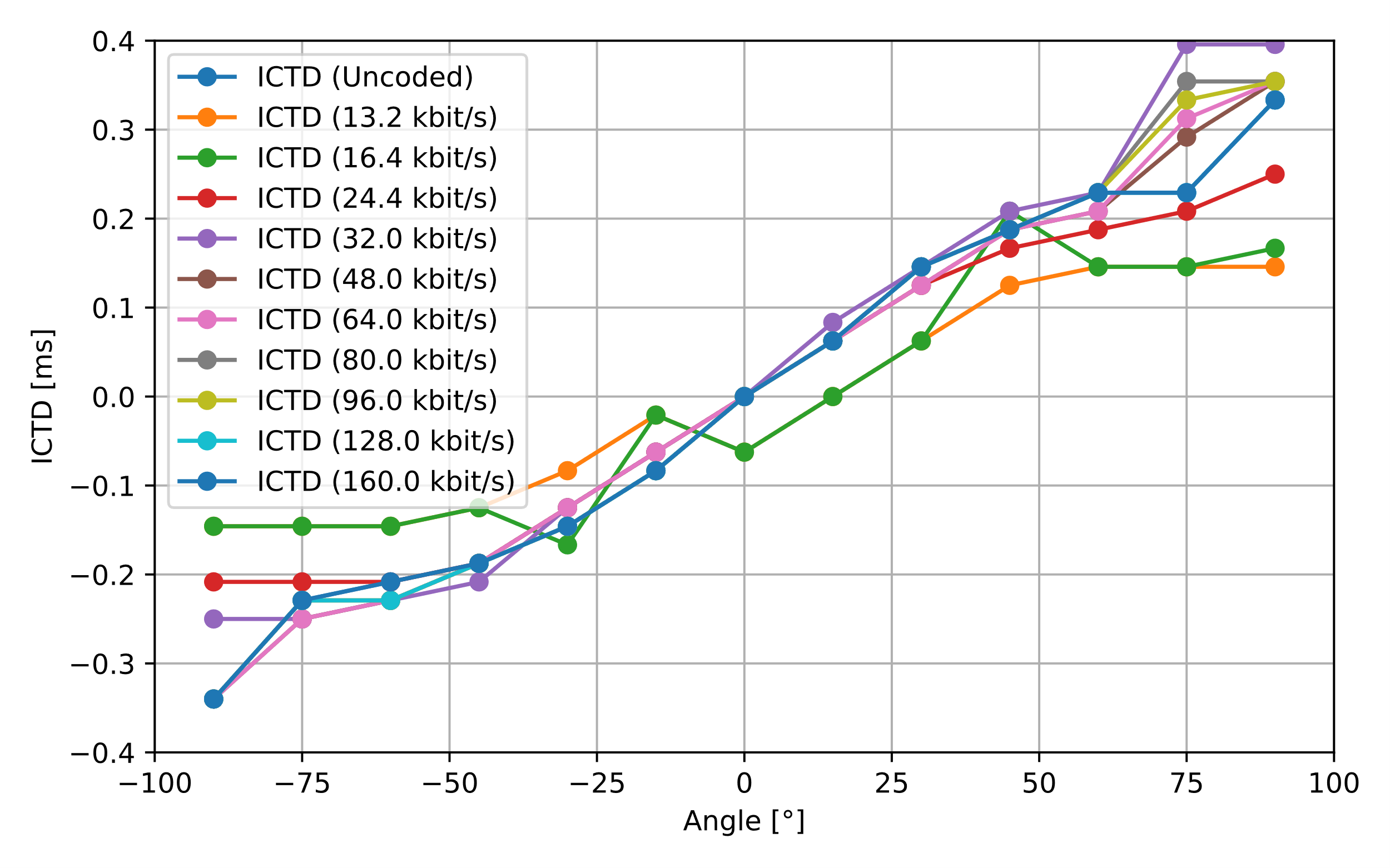


Figure 4: ICTD results with IVAS encode/decode, for A-B (left) and X-Y (right) configuration

## ICLD

### Uncoded

The results for the two different ICLD calculation methods and without any subsequent coding are shown in Figure 5. For both microphone configurations (A-B, X-Y) as well as for both calculations, the expected ICLD trend can be observed. The percentile measure obtains slightly higher values in magnitude, as it considers positive/negative peaks in the delta-level-vs-time graph. It can be observed that ICLD values up to ~12 dB can be achieved.

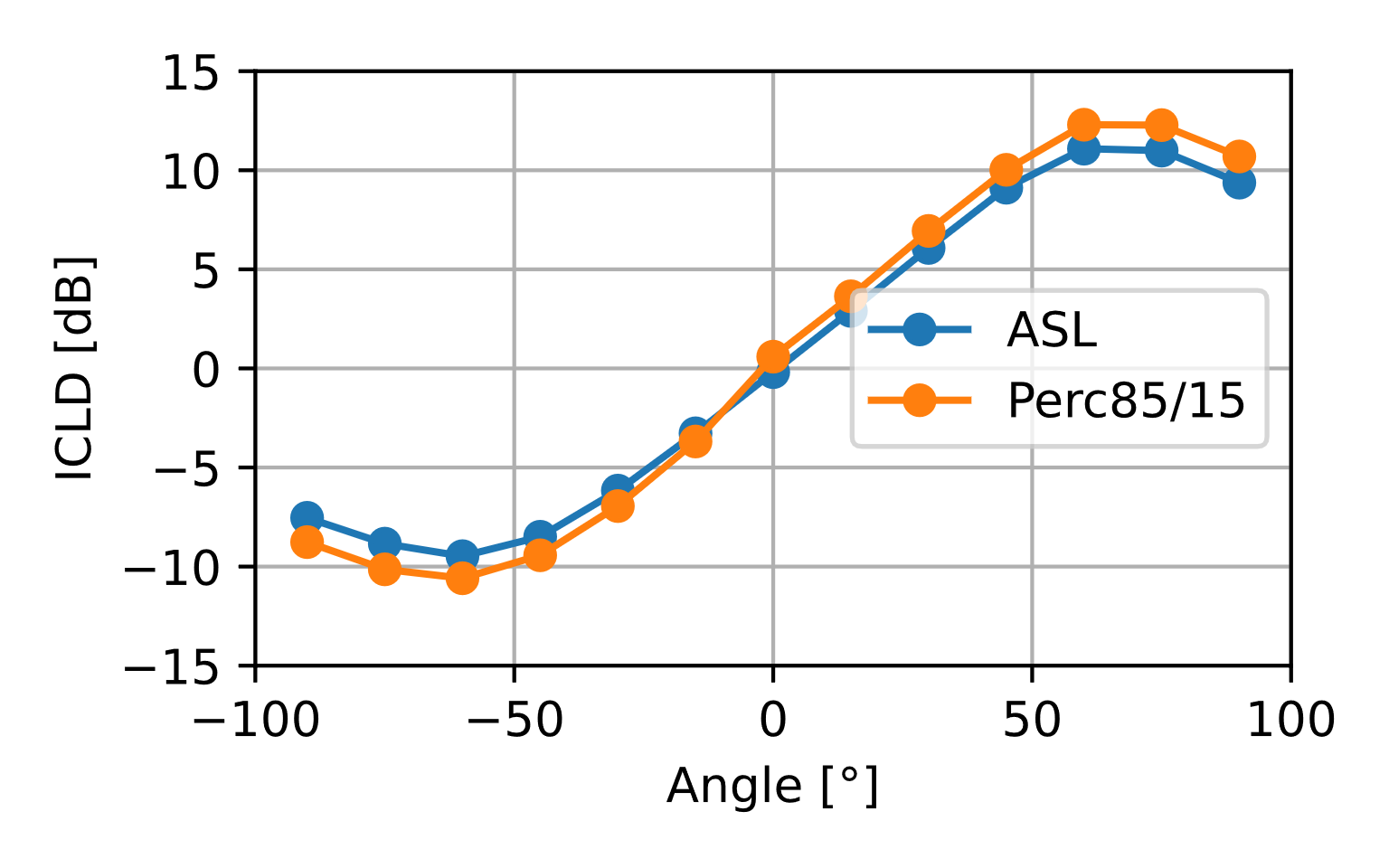
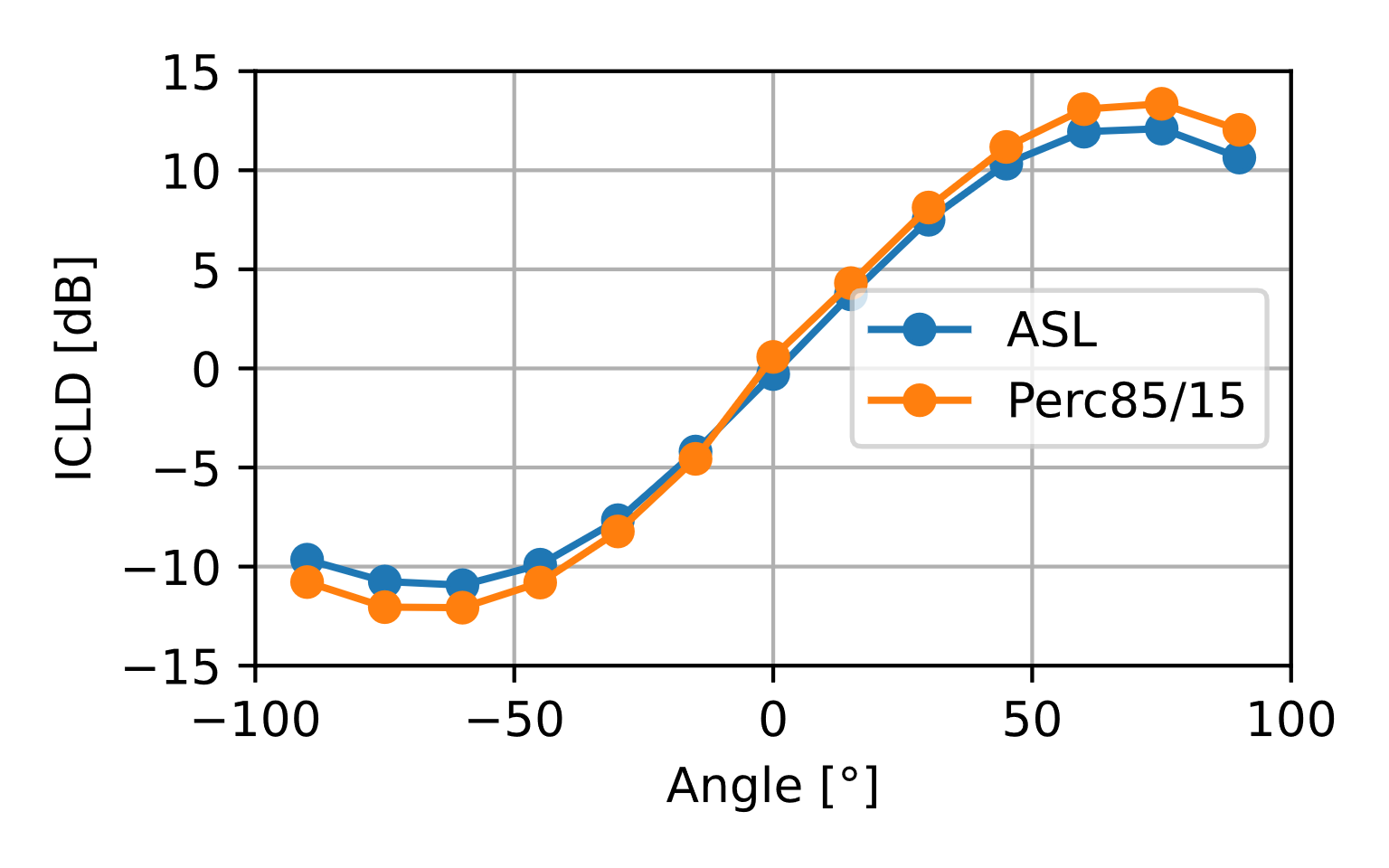


Figure 5: ICLD results versus angle, for A-B (left) and X-Y (right) configuration

### IVAS stereo mode

The calibrated stereo speech recordings were offline encoded/decoded with IVAS codec at different bitrates in stereo output format (latest floating-point version). The corresponding results for ICLD analysis are shown in Figure 6 (ASL) and Figure 7 (percentile).

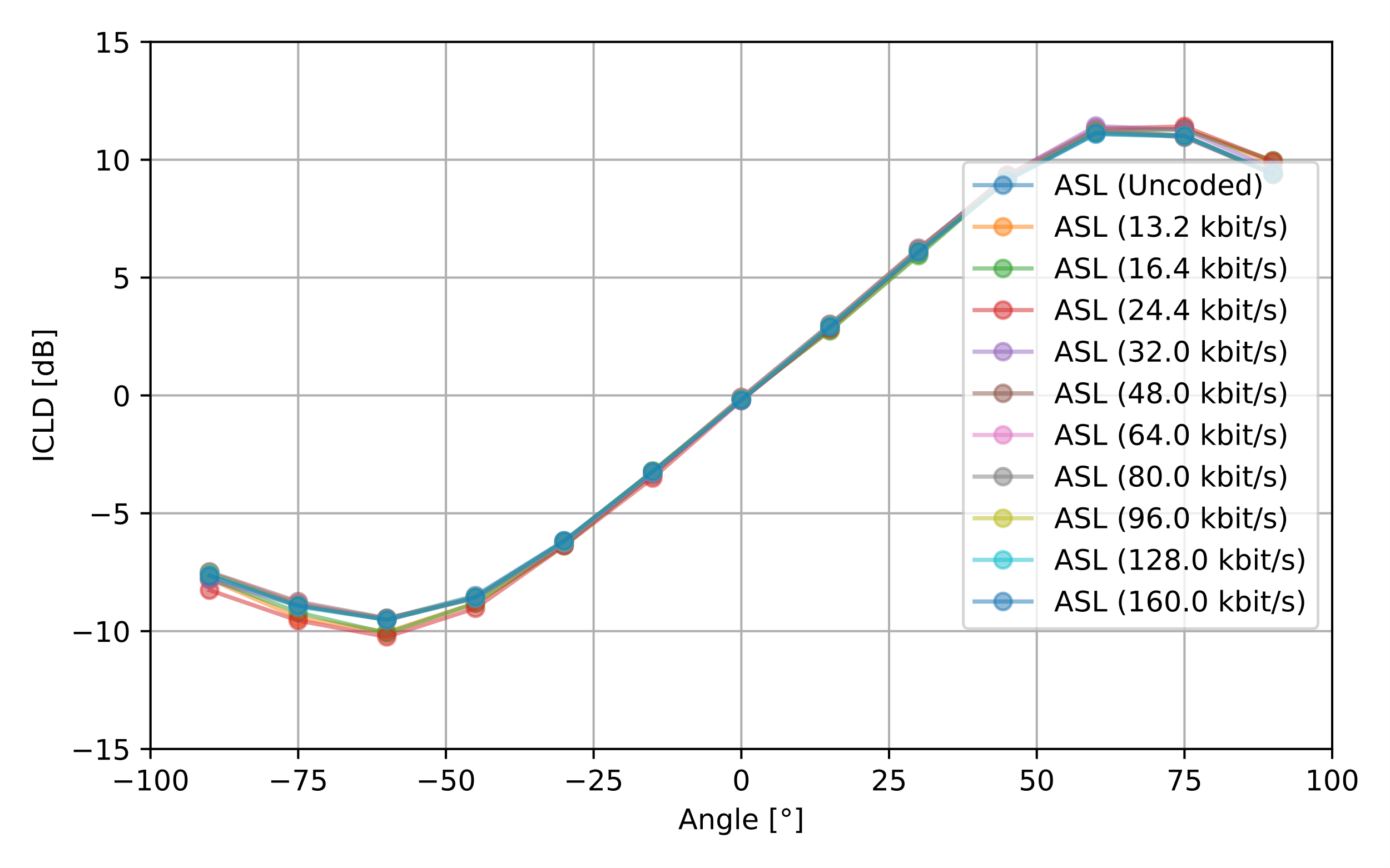
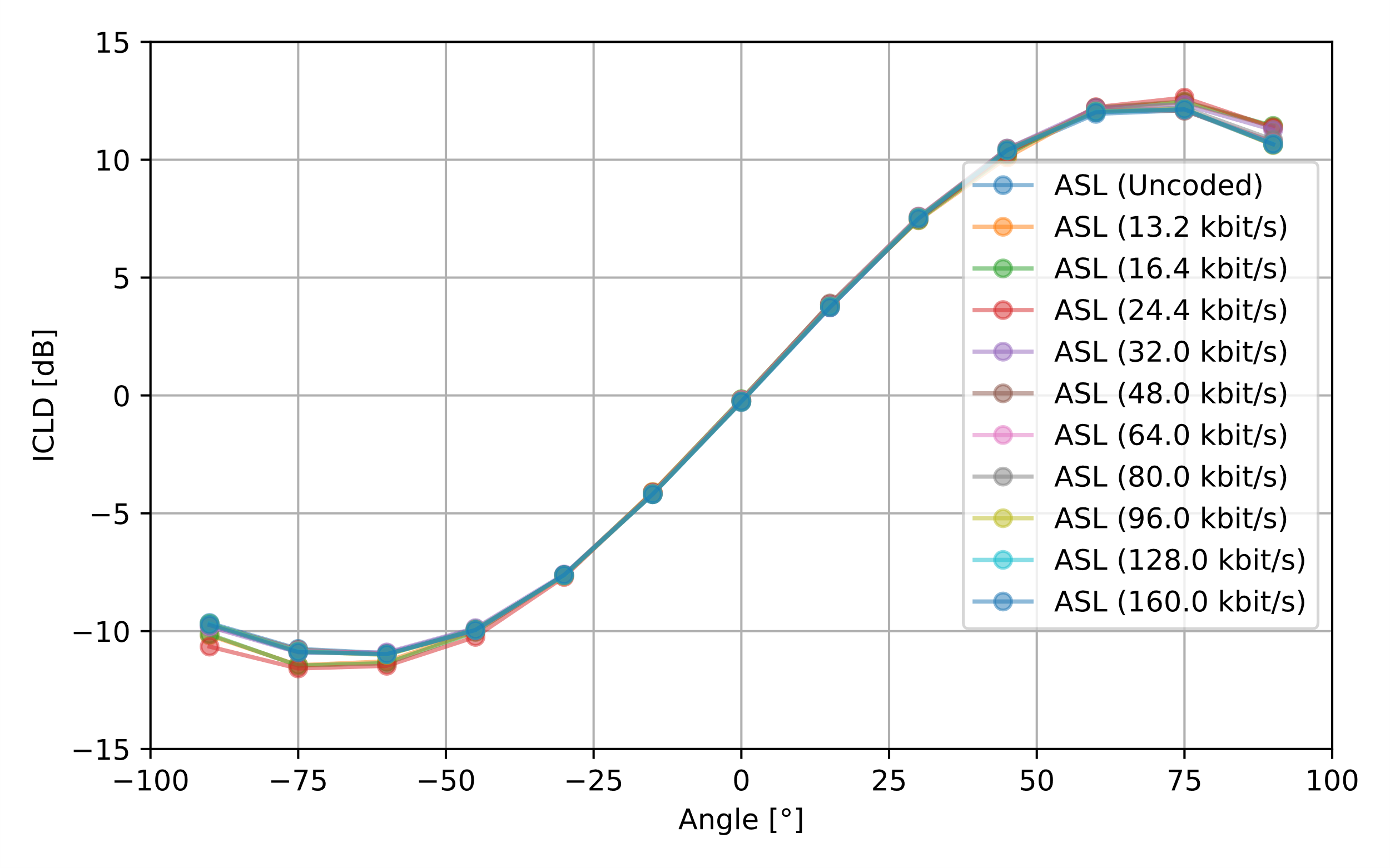


Figure 6: ICLD results (ASL) with IVAS encode/decode, for A-B (left) and X-Y (right) configuration

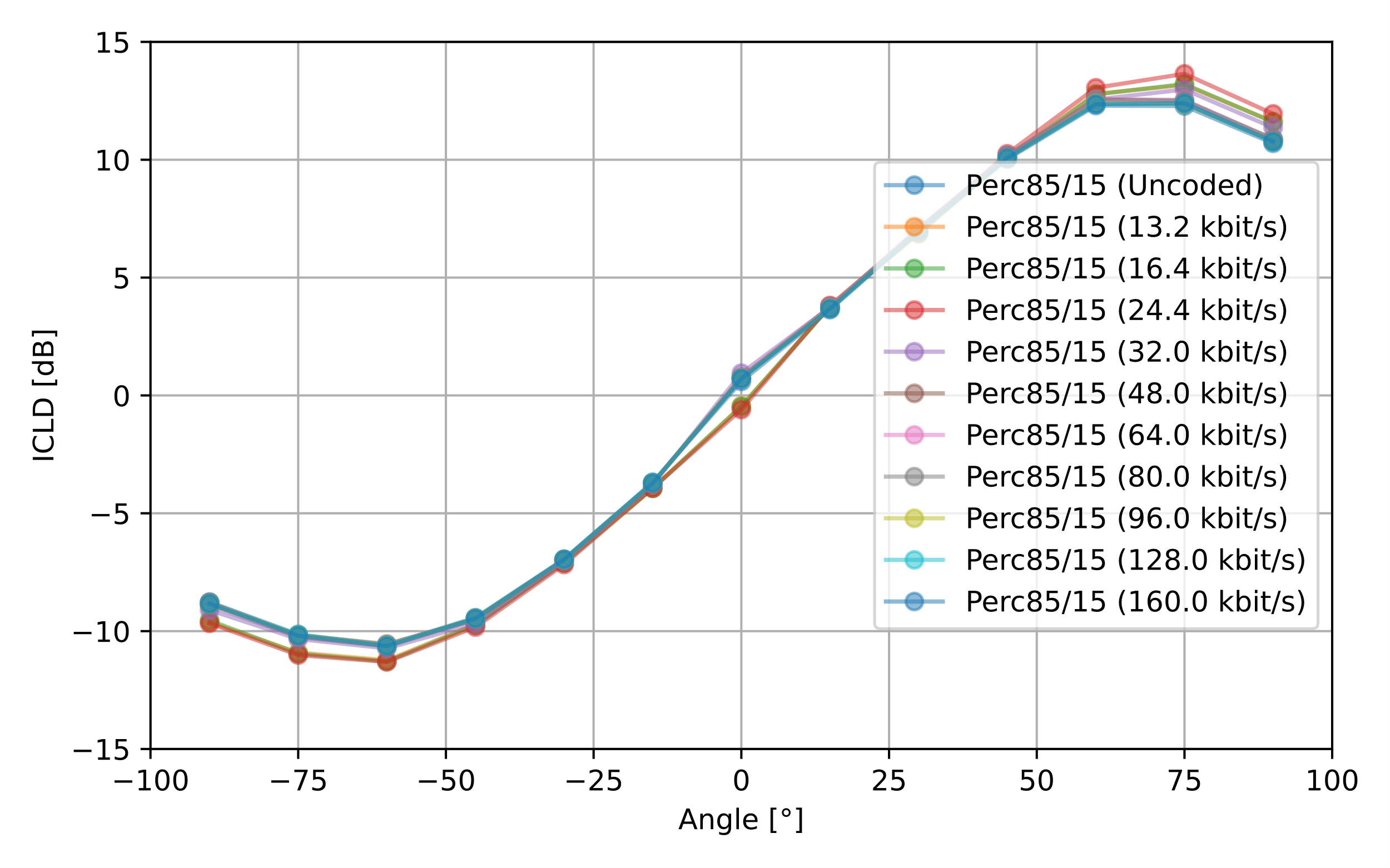
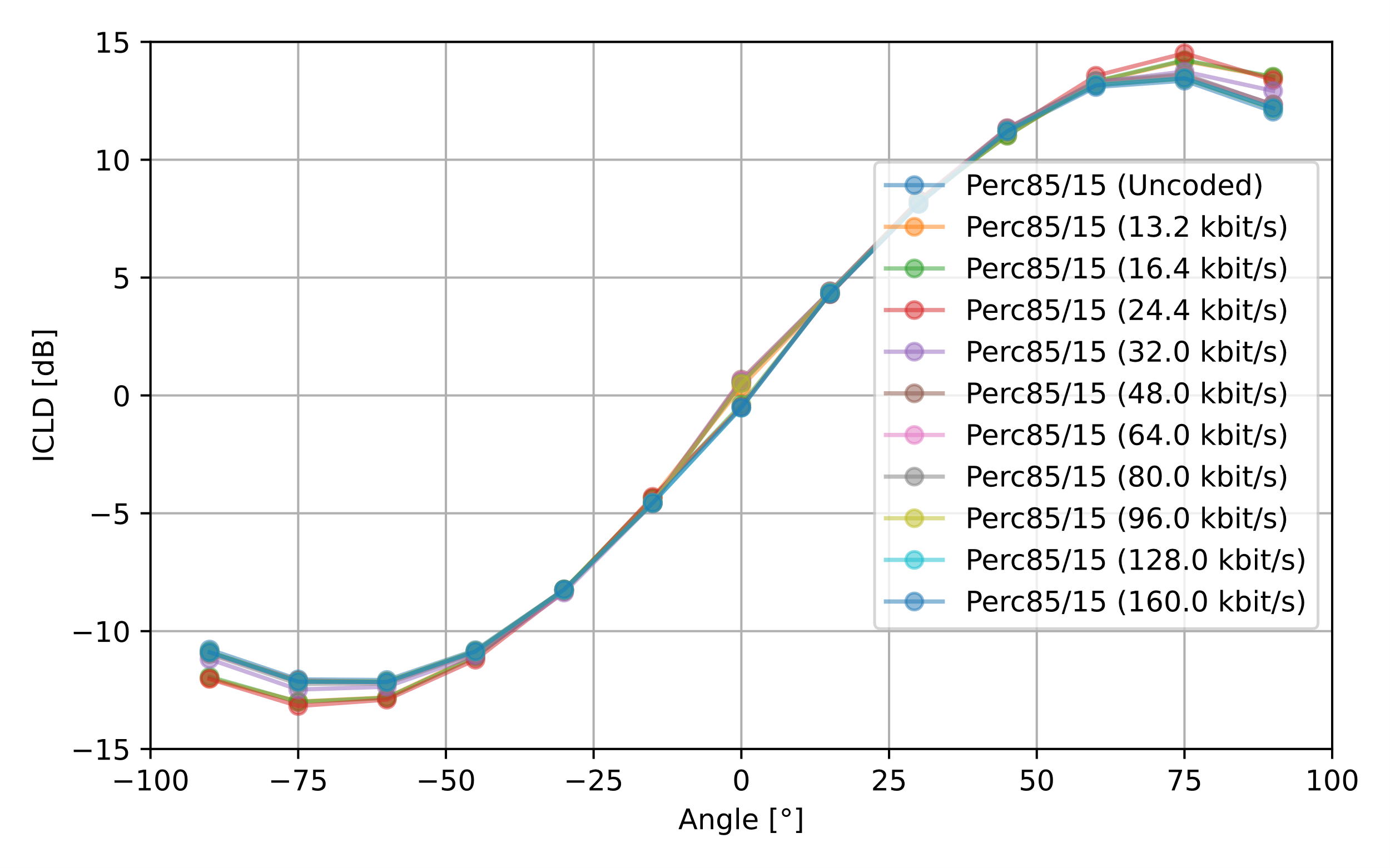


Figure 7: ICLD results (percentile) with IVAS encode/decode, for A-B (left) and X-Y (right) configuration

## Stereo image position

### Overview

ICLD and ICTD both have an impact on the perceived stereo image. Depending on the microphone configuration/hardware, subsequent signal processing and/or mixing of input channels (i.e., anything prior to the actual stereo signal output), a range of different valid ICLD and ICLD values are to be expected.

In literature, various studies have investigated the impact of ICLD and ICTD on the perceived spatial image for standard stereo playback systems (typically stereo loudspeaker setup with +/- 30°). In [6] [7], an approach independent of the playback system was introduced, which combines ICLD and ICTD to a percentage measure (+/-100%), which quantifies the expected spatial image in the stereo panorama by means of a so-called "localization curve". The minimum or maximum of the measure corresponds to a source position perceived at the left or right extreme of stereo panorama. The measure is based on the following findings derived from subjective studies and related work:

* For lower absolute ICLD values, a linear relation of 7.3% per dB is assumed.
* Higher values of ICLD increase the measure, but in a compressed way.
* For lower absolute ICTD values, a linear relation of 13% per 0.1 ms is assumed.
* Higher values of ICTD increase the measure, but in a compressed way.
* To a certain extent, ICLD and ICTD can be exchanged equivalently to pan a sound source in a stereo signal.

A first step to calculate the localization curve, ICLD and ICTD are aggregated into an overall level difference equivalent given by

According to the previous assumptions, the spatial image function is divided into a linear part, a compressed range, and a clipped maximum. It can be stated as

The linear part and the full localization curve are both illustrated in Figure 8 as a function of ICLD or ICTD.

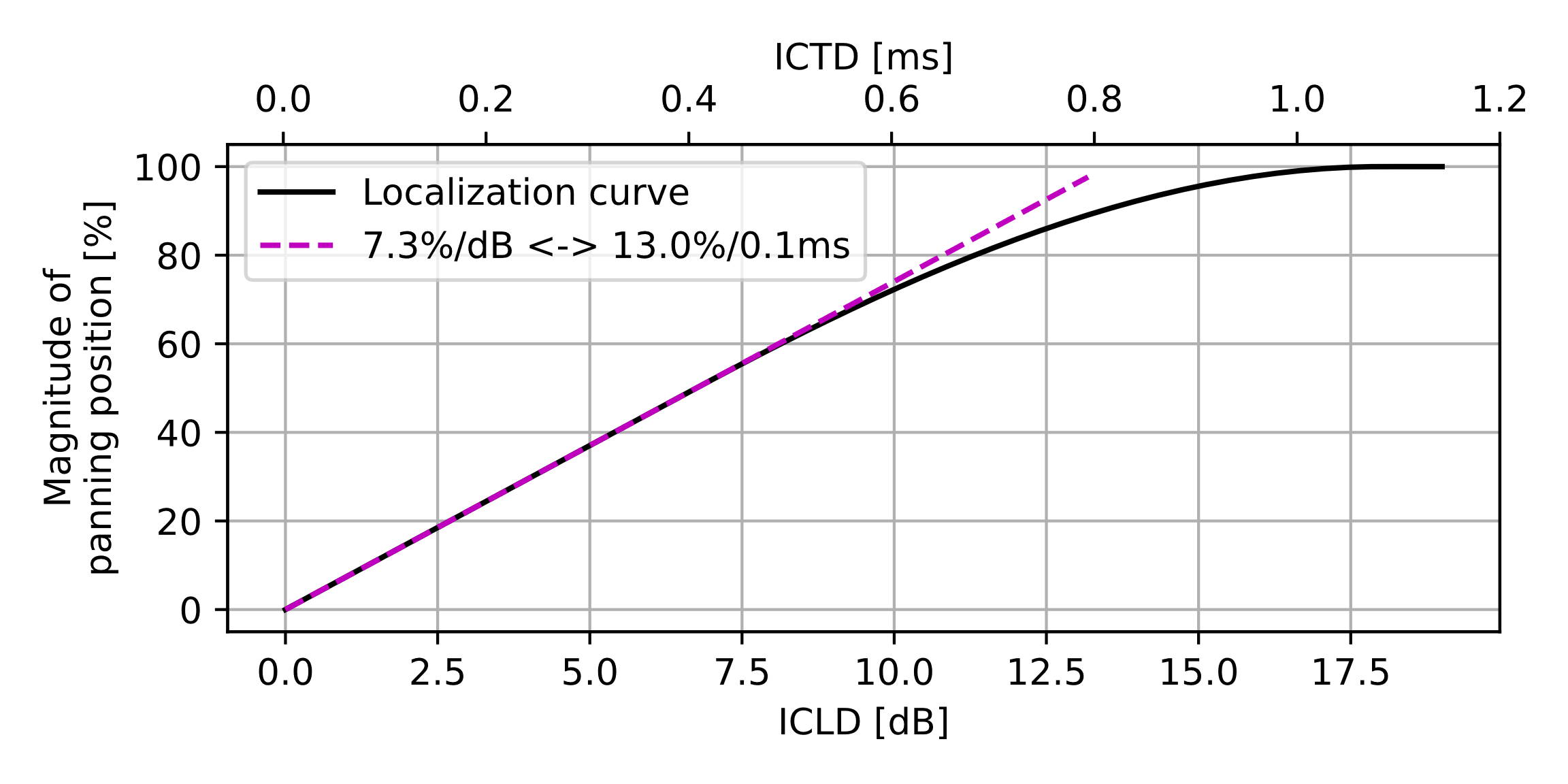


Figure 8: Estimated magnitude of panning position according to [6] & [7].

### Results for uncoded

The results for measure and without any subsequent coding are shown for both microphone configurations in Figure 9. In addition, the individual contributions of ICTD and ICLD to the overall measure are also provided. It can be noticed that the stereo image for the current test setup is quite similar for both configurations. Even though the A-B configuration provides a much wider range of ICTD, the absolute values of time differences are too low to contribute significantly to the measure. At the same time, values for ICLD are already rather high (>> 6.75 dB – compression range in the localization curve, see Figure 8). For A-B and X-Y configuration, ICLD is the dominant contribution to the overall result, which reaches in both cases about 75%.

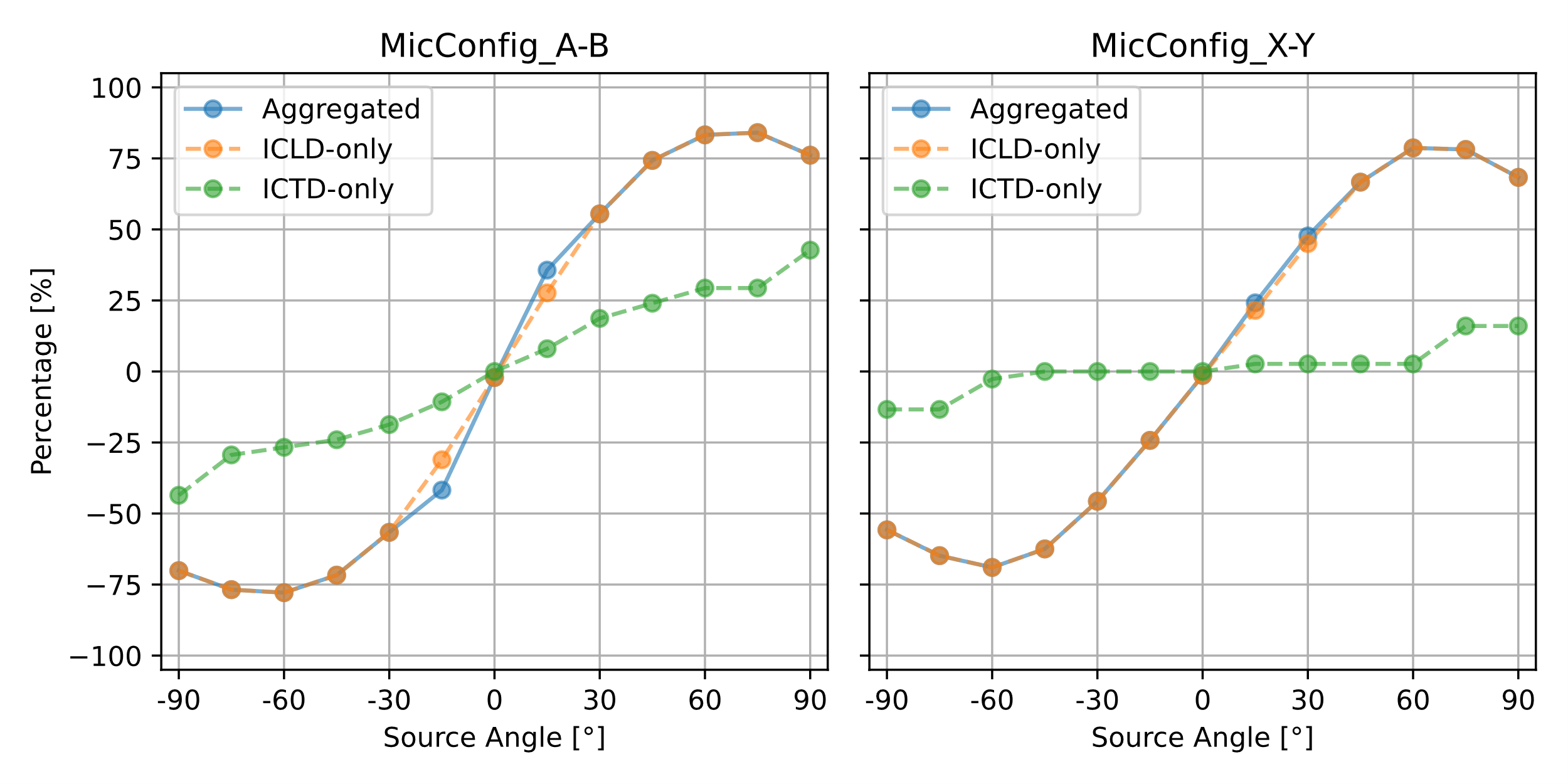


Figure 9: Stereo image for uncoded signals and individual contributions of ICTD and ICLD.

### Results for IVAS stereo mode

The corresponding measure for each signal encoded/decoded with IVAS in stereo mode is shown in Figure 10 for all bitrates.

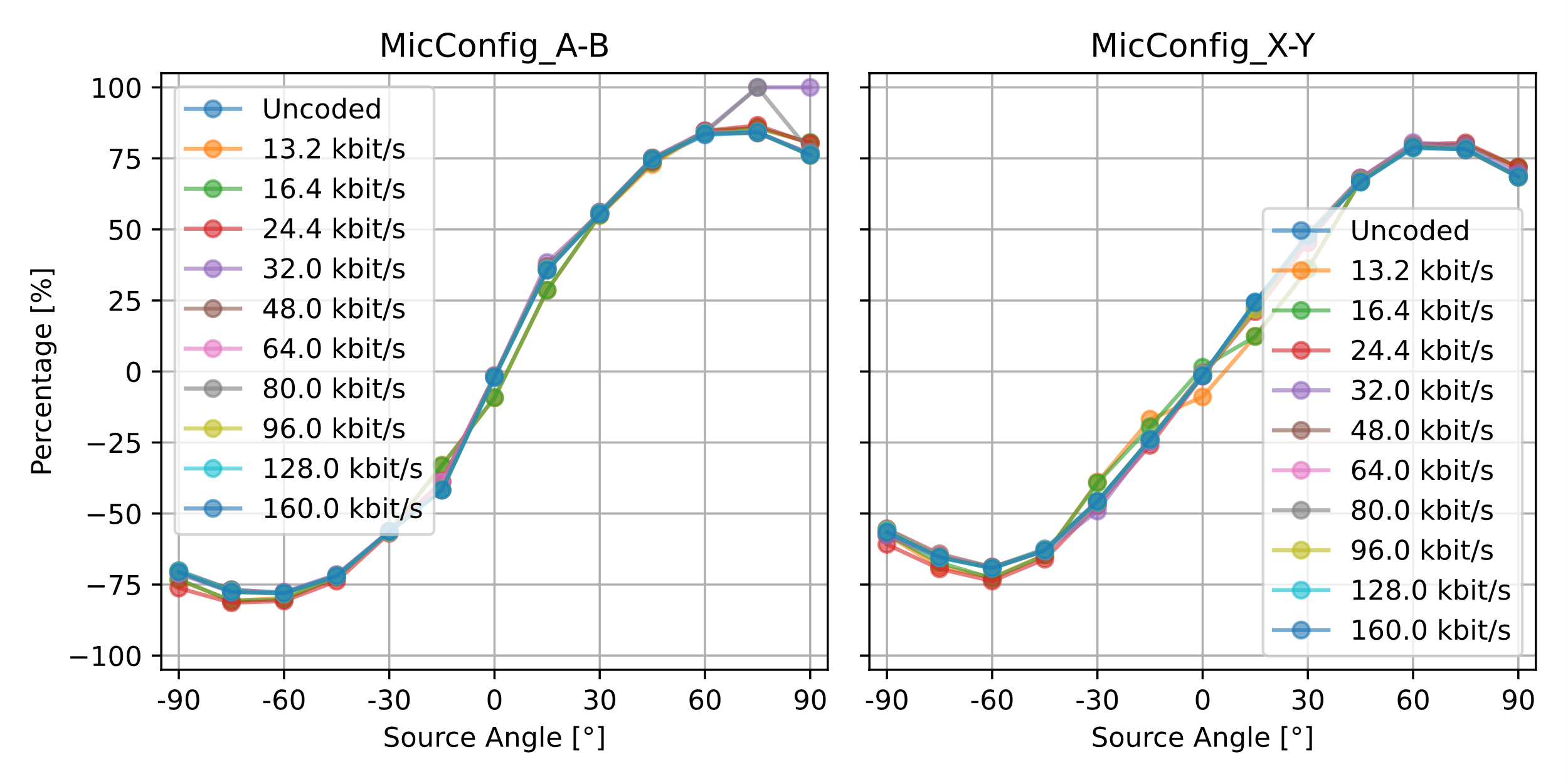


Figure 10: Stereo image for coded signals

# Conclusion

The document presented initial results of a test series for stereo capture. Measurement setup, test signal and analysis methods for ICLD and ICTD were introduced. In a first step, data for two different microphone configurations (A-B, X‑Y) was collected and analyzed to determine the expected range for both measures.

Surprisingly, the "traditional" ASL measure shows very similar performance as the percentile analysis of the delta-level-vs-time analysis, which should address temporal fluctuations in level.

In a second analysis step, IVAS encoder/decoder was applied in stereo mode at all available bitrates. For ICTD and all considered ICLD methods, the IVAS codec seems to maintain these important spatial properties of the stereo signal, even at lower bitrates.

Beside the robustness of the IVAS codec, the investigation illustrated the magnitude of ICTD and ICLD that can be expected for UEs with a form factor typical for handheld hands-free. However, it was noted that both ICTD and ICLD have an impact on the perceived stereo image. Therefore, in order to derive performance requirements which takes into account both dimensions, an aggregate measure is used, which estimates the perceived stereo image.

Annex A provides a full description of the test method, which is proposed to replace the existing clause 4.8 in Pdoc ATIAS-1.

# References

|  |  |
| --- | --- |
| [1] | 3GPP S4-231970, „PDoc ATIAS-1 v0.6.0,“ Audio SWG. |
| [2] | Recommendation ITU-T P.58, „Head and torso simulator for telephonometry,“ 03/2023. |
| [3] | ITU-T Recommendation P.501, „Test signals for use in telephony and other speech-based applications,“ 06/2022. |
| [4] | ITU-T Recommendation P.56, „Objective measurement of active speech level,“ 12/2011. |
| [5] | IEC 61672-1:2013, „Electroacoustics - Sound level meters - Part 1: Specifications,“ 2013. |
| [6] | H. Wittek, „Lokalisationskurven online – Der "Image Assistant",“ 11/2000. |
| [7] | H. Wittek und G. Theile, „The recording angle – based on localisation curves,“ in *Program of the 112th AES convention*, Munich, 2002. |
| [8] | Recommendation ITU-T P.581, „Use of head and torso simulator for hands-free and handset terminal testing,“ 07/2022. |

# Annex A: Description of test method for stereo capture

## Test setup

The test setup for handheld hands-free UE according to clause [2.1.4 in ATIAS-1] is used with a distance D = 42 cm between HATS reference point (HRP) and center of the UE. In addition to the frontal incident (αN/2 = 0°), *N=*13 talker positions between [α0 = -90° and αN-1 = +90°] shall be evaluated in steps of [15°], as indicated in Figure X and Table Y. In case the setup is realized with a turntable, the device shall be rotated around the vertical center of the UE.

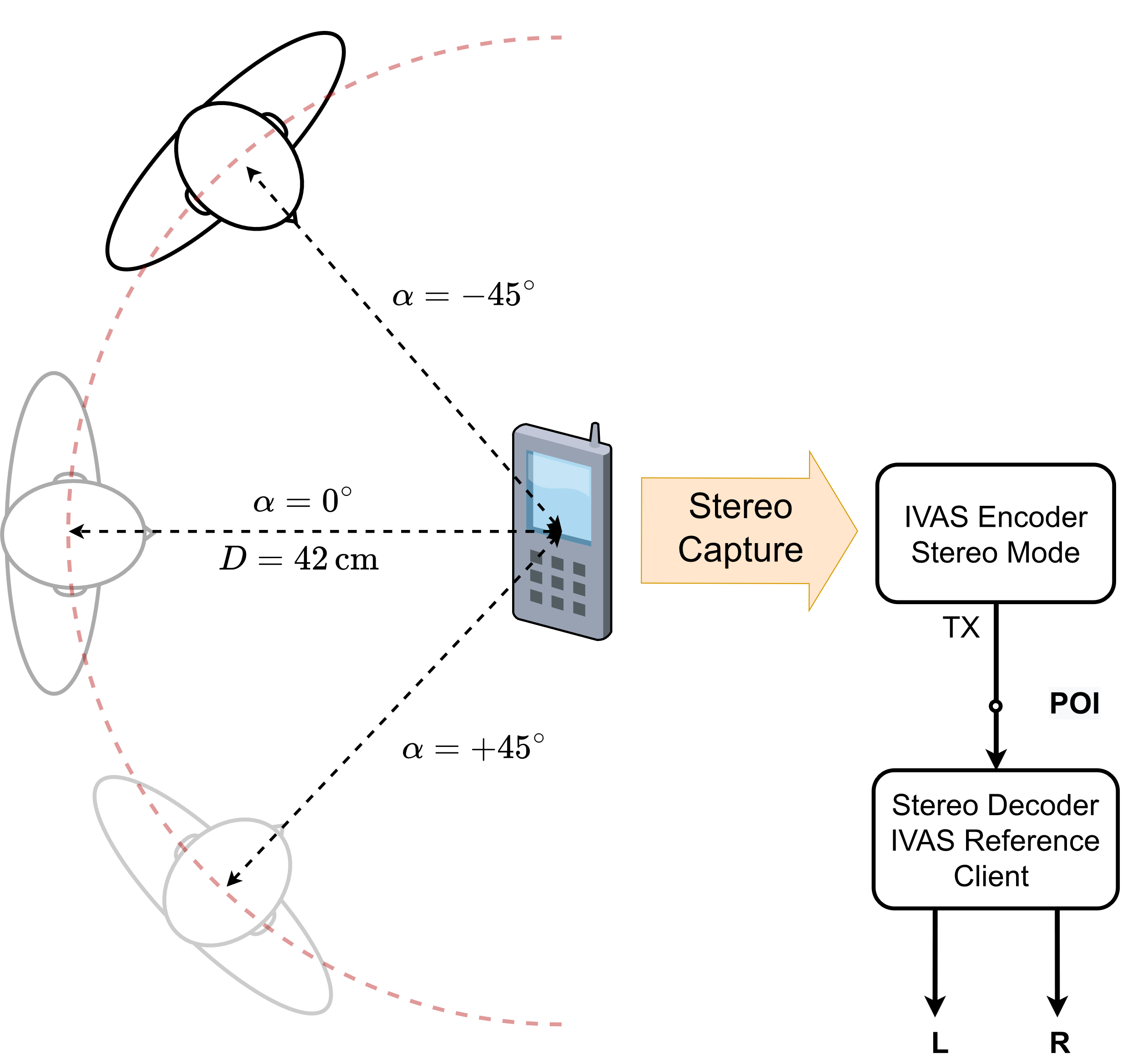


Figure X: Test setup for stereo capture

Table Y: Talker positions used for stereo capture test

|  |  |
| --- | --- |
| αi | Angle [°] |
| α0 | -90 |
| α1 | -75 |
| α2 | -60 |
| α3 | -45 |
| α4 | -30 |
| α5 | -15 |
| α6 | 0 |
| α7 | 15 |
| α8 | 30 |
| α9 | 45 |
| α10 | 60 |
| α11 | 75 |
| α12 | 90 |

## Requirements

The stereo panorama shall be consistent for source positions αi between [α2 = -60° and α10 = + 60°] according to Table Y. The estimated source positions in the stereo panorama shall be monotonically increasing, i.e., (for 2 < *i* < 11).

The stereo panorama shall provide a minimum and symmetric width between [α2 and α10], i.e., and shall be larger or equal [50%…75%].

The absolute values of the estimated source positions outside the consistent stereo panorama (α0, α1, α11., α12) shall be larger or equal than the edges of the consistent stereo panorama, i.e., and .

For frontal incidence position (α6 = 0°), the sound source shall be in the center of the stereo panorama, i.e.,

For all source positions of Table Y, the absolute value of ICTD shall be less than [1.5] ms and the absolute value of ICLD shall be less than [20…25] dB.

## Measurement

1) The orientation between HATS and UE is set to source position α0.

2) The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [xx], calibrated to an active speech level according to Recommendation ITU-T P.56 [xx] of -21.3 dBPa at HFRP (nominal speech level of -4.7 dBPa at MRP increased by 3 dB – see Recommendation ITU-T P.581 [xx]).

3) If necessary, the UE is configured for IVAS encoding in stereo mode at [TBD] kbit/s. The IVAS reference client is configured to stereo output format.

4) The left and right channel signals () are recorded by the test equipment.

5) The inter-channel time difference (ICTD) is determined as follows:

a) The envelope of the segmental cross-correlation function between and is calculated by means of the Hilbert transformation:

b) Each segment has a duration T of 8192 samples and an overlap of 50% is used. The ICTD for source position α0 is then determined as the time lag of the averaged envelope, that provides the maximum value:

6) The inter-channel level difference (ICLD) is determined as follows:

a) The active speech level according to Recommendation P.56 [xx] is calculated separately for and , resulting in and .

b) ICLD for source position α0 is calculated as:

7) The equivalent level difference for source position α0 is calculated as:

8) The estimated source position in the stereo panorama is calculated as:

9) The measurement is repeated for all source positions α1 to αN-1 to obtain to

1. The terms ILD/ITD are typically used in the context of binaural recordings and directly correspond to head-related level and time differences. For stereo, the same definition might not apply, as we are mainly interested in "technically maintaining" a given time/level between two channels. For this reason, these measures are denoted differently in the present document as ICLD/ICTD. [↑](#footnote-ref-1)