# LABORATORY EVALUATION OF SMARTPHONE AUDIO ZOOM SYSTEMS

**Dan Zhao, Arthur Drouadene, Harry Maisonneuve, Philippe Guelen, Justin Bacle**
dzhao@dxomark.com
jbacle@dxomark.com

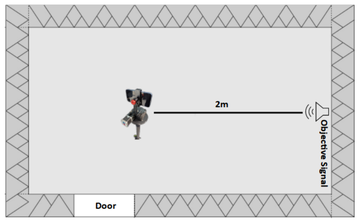
**http://corp.dxomark.com**
**24-26 Quai Alphonse le Gallo**
**92100 Boulogne-Billancourt, France**

## Abstract

When taking a video of a person speaking, you have the option of using the camera to zoom in on the speaker's face. But is your phone's microphone capable of zooming in on the speaker's voice so that it stands out above the background and competing noises?
This poster proposes a novel rating protocol for evaluating smartphone audio zoom systems using objective and perceptual testing. Audio zoom is a recently developed feature that isolates sound sources in videos based on the smartphone camera's focal point and zoom level [ Avendano and Solbach (2015) Vincent et al. (2006)]. The protocol considers the ability of a system to focus on the target sound as the main criterion for evaluation, while also discussing other important audio quality factors.

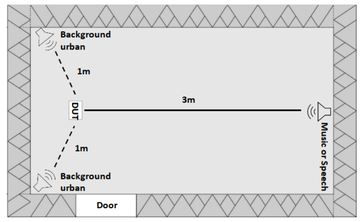
## Evaluation Protocol

Our evaluation protocol consists of two parts: objective directivity testing and perceptual listening tests. Both types of recordings are performed in a semi-anechoic chamber, with the smartphones' main camera in landscape orientation.




### Objective Directivity

- 5-second logarithmic swept-sine (20-20000Hz)
- 10° step full revolution




### Perceptual Recording

- Use cases: Speech & Music
- Attributes: Timbre, Spatial, Volume, Artifact, Background

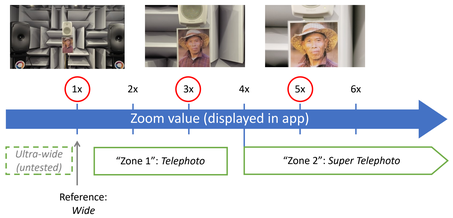
## References

Avendano, C. and L. Solbach (2015). Audio zoom. *US Patent Submitted 20*, 110–129.

Recommendation, I. (2003). Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm. *ITU-T recommendation*, 835.

Vincent, E., R. Gribonval, and C. Févotte (2006). Performance measurement in blind audio source separation. *IEEE transactions on audio, speech, and language processing 14*(4), 1462–1469.

## Evaluated Zoom Ranges




- If the camera app provides a quick zoom preset value (#x), that value will be used.
- If the camera app does not have presets, the following values (circled in red) will be used.
- If there are multiple presets available within a given range, the preset closest to the circled value will be chosen.

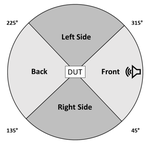
## Objective Directivity Analysis

### Directivity Graphs




A 10°-resolution directivity graph showing the ability of the device to zoom in on the content(blue: wide, red: tele, brown: super tele). Comparisons between wide, tele and super-tele in the frequency band (800 Hz – 2000 Hz) for devices A, B, C, D.

### Objective Directivity Score




Direction zones for computing objective directivity

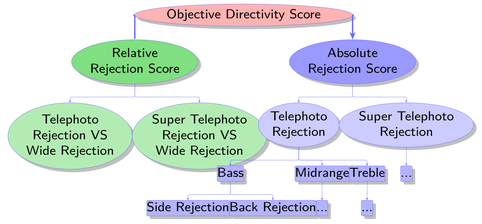
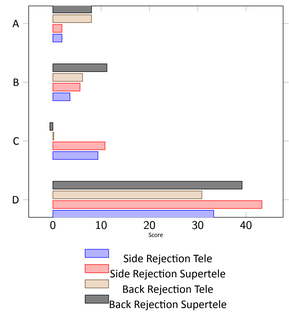



Diagram of the objective directivity algorithm

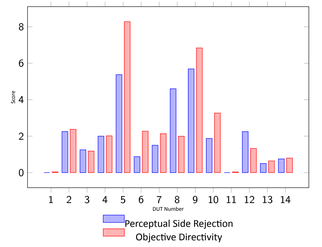



Comparison of objective side and back rejection scores in frequency band (800 Hz – 2000 Hz) for devices A, B, C, D (the higher the better)

## Perceptual Analysis

We conducted an evaluation of 14 recently released smartphones (before June 2022) that included audio zoom in their official website specifications. The evaluation focused on the audio quality of the phones and involved four expert audio engineers with qualifications in perceptual listening [Recommendation (2003)]. The engineers followed precise listening guidance for each attribute and assigned scores on a scale of 0 to 10, with higher scores indicating better performance.

Perceptual VS Objective Side Rejection




Four typical patterns of average recording loudness evolution across zoom zones

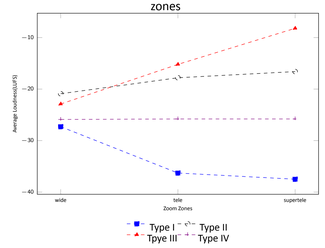
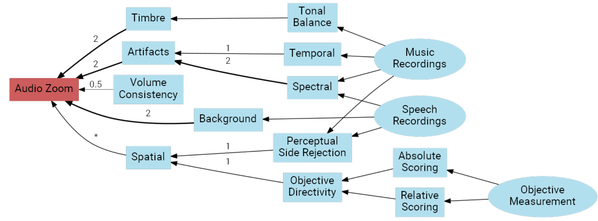
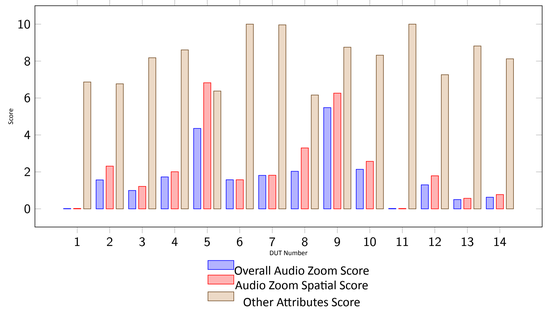



Diagram of the audio zoom scoring system




Audio zoom final score comparisons among devices




## Conclusion

- Devices offering decent audio zoom function have good or reasonable backside rejection. However, most devices' side rejection needs to be developed and improved.
- The occurrence of disturbing artifacts when zooming in can degrade the user's video recording experience. Therefore, controlling these artifacts is critical to achieving good audio zoom performance.