XLYOPHONE SPECTRUM REVIEWER EXPLAINED

Hey y'all, In this document we will be explaining exactly what we expect from you in the role of Reviewer in the Xylophone Spectrum project and how you should evaluate. Read the whole document carefully and make sure you always have it open with you.

What is the Reviewer role and for what purpose?

There are 3 layers in total in this project. Understanding these 3 different layers is very important to understand the roles in general.

L-1: Attempt level, at this level audio recording is taken for the task and this recording is loaded into the task, selections are made and approved. The approved task moves to L0 level.

L0 : Review level, at this level it is checked whether the task sent by the attempter thinking that it is "completely correct" is actually completely correct. **If there is a fixeable error, it is corrected, approved and sent to L10 level. If there is an unfixable error, it is sent back to queue (SBQ) and sent back to L-1.** In case of SBQ, all information in the task is reset to zero.

L10: Senior Review level, where the tasks called "fixed and approved" at the L0 level are checked to make sure that they are ready to send to the customer. **The Senior Reviewer's evaluation and scoring is for the L0 Reviewer.**

What are the situations that are Fixeable and Unfixeable when reviewing?

When you claim a task to review, you will see an interface called "Output" which pulls all the data from the Audio Collector. It is very important to read this "Output" section correctly.

Click here to open audio 📋
Click here to open video 📋

S   📋 Collapse ⌃

```
1  {
2      "audioUrl": "https://scale-cds-public-us-west-
   2.s3.amazonaws.com/66193e0f5491a13716ce22b7/JTaf8S3MQcMMKx8",
3      "videoUrl": "https://audio-
   collector.scale.com/api/recordings/s3Redirect/67527b199129e6bd81b02f82_6873e0ef6f033cb86a
4      "metadata": {
5          "uploadedAt": "2025-07-13T16:38:12.184Z",
6          "transcription": "What would you do if you got a voicemail, uhm not a message but, a
   recording of you speaking, and you were talking about a place that you've never, been to
   and you've never seen, what would be your reaction?",
7          "category": "indoor",
8          "subcategory": "office",
9          "noiseLevel": 2,
10         "recordingDevice": "lenovo laptop",
11         "description": "Im in the living room, room echo, fridge, fish tank filter, daughter
   watching cartoons.",
12         "demographics": {
13             "accent": "Zulu",
14             "gender": "Female",
15             "nativeLanguages": [
16                 "English (en-ZA)",
17                 "Zulu"
18             ]
19         },
20         "distribution": {
21             "recordingLocation": "Indoor Personal Spaces",
22             "category": "Role-Play & Sound Effects",
23             "subcategory": "Absurd & Surreal Scenarios",
24             "language": "English"
25         },
26         "markedAsUsed": true
27     }
28  }
```

The sections marked in green in this section are "Fixable" sections. **TASK should not be SBQd even if there is any error in the areas marked in green.** [Cause we can fix them, no need to reset!]

⚠️ **SBQ** ⚠️ = In case of irreversible and unchangeable errors, the task is rejected and reset. It is used as a last resort.

21. RecordingLocation ⛔

It is the only part of the Output section that cannot be changed. The incorrect RECORDING LOCATION CATEGORY selected *here deserves to be SBQ as it is an irreversible error.*

7. Category ✅

```
7          "category": "indoor",
```

You can fix at the end of the task, no need to deduct points for that mistake, just warning.

8. Subcategory ✅

```
8          "subcategory": "office",
```

You can fix at the end of the task, no need to deduct points for that mistake, just warning.

9. Noise Level ✅

```
9          "noiseLevel": 2,
```

The Noise Level selected in the Output section is not important, **ignore it,** what is important is the Noise Level selected in the TASK UI. You should make the evaluation according to this:

Noise Level - IMPORTANT *

Please re-listen to your recording and select the most accurate noise level

- ○ 0 - Silent (no discernible background noise)
- ○ 1 - Low (faint, barely noticeable background noise)
- ● 2 - Moderate (clearly noticeable, but doesn't overpower the main audio)
- ○ 3: High (background noise is prominent and competes with the main audio; audio is difficult to understand, but more than 90% of the words are decipherable)
- ○ 4: Very High (background noise overpowers or distorts the main audio; audio becomes very difficult to understand, but the meaning behind the prompt can be derived and more than 70% of the words are decipherable)

22. Category ✅

```
22          "category": "Role-Play & Sound Effects",
```

You can fix it at the end of the task, and you need to deduct points for the mistake.

23. Subcategory ✅

```
23          "subcategory": "Absurd & Surreal Scenarios",
```

You can fix it at the end of the task, and you need to deduct points for the mistake.

⚠️**If the main category is incorrect, feel free to change it.**

⚠️ **If the subcategory is close to a 50/50 split, do not deduct or change points. Only correct very obviously incorrect subcategory choices and deduct points.**

How to review a task step by step ⏩

In this section, we will show you step by step how you can review in the most optimal way by adjusting the time. This may vary according to the person; you can differentiate the order of the steps here according to yourself.
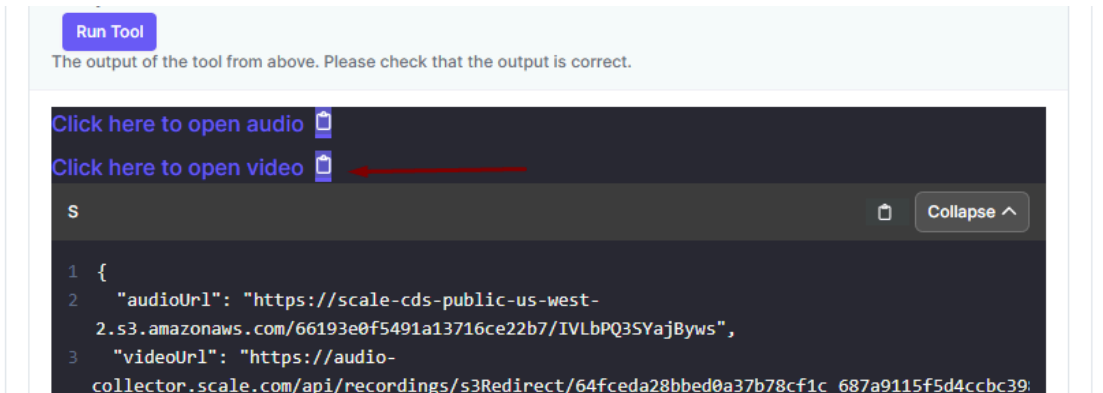
⚠️🎧⚠️: PLEASE PLEASE PLEASE, USE HEADPHONES WHILE DOING REVIEWS.

1. **CLAIM THE TASK AND CHECK VIDEO/CATEGORY ALIGNMENT**

- First of all, let's go to Line 20/21 in the Output section and check which category is selected in the RecordingLocation section.

```
16          "English (en-US)"
17        ]
18      },
19      "distribution": {
20        "recordingLocation": "Outdoor Public Environments",
21        "category": "High-Quality Conversation",
22        "subcategory": "Narrative & Story-Sharing (Realistic Contexts)",
23        "tone": "Neutral",
24        "language": "English"
25      }
26    }
```

  - **Indoor Personal:** Between 4 walls, indoor and personal spaces. Please check the details one by one in the document.
  - **Indoor Public:** Spaces, usually between 4 walls, impersonal, where different people are or can be present at the moment. Please check the details one by one in the document.
  - **Outdoor Public:** Not between 4 walls, where you can see the sky, impersonal and potential spaces where different people are or can be in the moment. Please check the details one by one in the document.
  - **Transportation Env:** Inside or private stops/waiting areas for all types of transportation.Please check the details one by one in the document.
- When you check the selected topic, go to the "Click here to open video" part of the Output.

Run Tool

The output of the tool from above. Please check that the output is correct.

Click here to open audio
Click here to open video

S   Collapse ^

```
1  {
2      "audioUrl": "https://scale-cds-public-us-west-
   2.s3.amazonaws.com/66193e0f5491a13716ce22b7/IVLbPQ3SYajByws",
3      "videoUrl": "https://audio-
   collector.scale.com/api/recordings/s3Redirect/64fceda28bbed0a37b78cf1c_687a9115f5d4ccbc39
```

- The video should speak for itself, so make sure the video is a direct match to the chosen location.
  - **Indoor Personal:** DONT SBQ, EVEN IF IT'S A DARK VIDEO OR VIDEO DOESN'T EVEN EXIST. REMEMBER THAT WARNING:

REVIEWER UPDATE
Read the instructions carefully before continuing.

IF THE TASK IS AN INDOOR PERSONAL SPACES TASK, THE VIDEO IS NOT REQUIRED. YOU DON'T NEED TO SBQ AN INDOOR PERSONAL SPACES TASK IF IT DOESN'T HAVE A VIDEO.

Also, if a task is recorded OUTSIDE, it counts as Outdoor Public Environments. Even if it's in someone's backyard. Anything outside counts for this.

"TASK HAS TO BE IN ENGLISH and make sure the contributor chooses the right accent as well."

  - **Indoor Public:** If the video and location category is not matching > SBQ and 2/5.
  - **Outdoor Public:** If the video and location category is not matching > SBQ and 2/5.
  - **Transportation Env:** If the video and location category is not matching > SBQ and 2/5.
- If RecordingLocation and video is not matching, don't check the rest, 2 points + SEND BACK TO QUEUE.

2. **CHECK THE AUDIO**

    a. **Audio Length:**

SBQ | Approve with warning | Perfect

    i. **If the audio is less than 5 seconds** > 2/5 and SBQ.

  ii. **If audio is between 5-12 seconds and the rest is perfect,** <mark>deduct 1 point and give 4/5.</mark> **If the rest is not perfect, don't deduct extra points.**

    1. Fix the task and if you:
      a. **Approve:** Use that form > https://forms.gle/tKLv6TLkbCrs2vyf7
      b. **SBQ:** No need to fill the form

  iii. **If the audio is between 12-15 seconds,** just warn them without deducting any points.

    1. Also make sure to use that form: https://forms.gle/tKLv6TLkbCrs2vyf7

  iv. **If the audio is longer than 15 seconds,** <mark>it's a good task.</mark>

 b. **Audio Noise Level**

Noise Level - IMPORTANT *

Please re-listen to your recording and select the most accurate noise level

- ○ 0 - Silent (no discernible background noise)
- ○ 1 - Low (faint, barely noticeable background noise)
- ● 2 - Moderate (clearly noticeable, but doesn't overpower the main audio)
- ○ 3: High (background noise is prominent and competes with the main audio; audio is difficult to understand, but more than 90% of the words are decipherable)
- ○ 4: Very High (background noise overpowers or distorts the main audio; audio becomes very difficult to understand, but the meaning behind the prompt can be derived and more than 70% of the words are decipherable)

  i. Listen to the audio recording for one full round just to determine the noise level, do not focus on anything else.

  ii. Then, if the Noise Level you set is different from the selected Noise Level, select the correct level and score points.

    1. **NL 0 >** I'm listening with %75 and I literally can't hear anything or little tiny minor one time thing. Rest is clean.

    2. **NL 1 >** I listen to the audio recording at 75% and in the background of the recording **I hear a very faint background noise that is not continuous or even if it is continuous, I make an extra effort to hear this continuous noise, it is really hard to notice.**

    3. **NL 2 >** I listen to the audio recording at 75% and I can hear the background sound very clearly. I turn it down to 50% and the background sound is still very clearly audible. **It's also continuous and I don't have to make excessive efforts to hear the sound.**

    4. **NL 3 >** Even when listening to the audio at 50%, there is a lot of background noise, making it difficult to hear a few words. **I had to guess some words, not all of them, but one or two of them are really hard to understand.**

    5. **NL 4 >** I don't hear 3 out of 10 words in the audio recording, I can't even understand them. It's hard to even guess.

 c. **Audio and Transcript accuracy check:**

      i.   **Check the words:** Listen the audio recording one round just to make sure that the words match 100%.

      ii.   **Check punctuation and spacing:** Listen to the audio recording for one round just to make sure that punctuation and pauses are shown correctly.

FILL THIS PART:

Answer the questions below

Reviewer Instructions: https://docs.google.com/document/d/e/2PACX-1vTccNNUa_cRFi-L-RObJHZ2kHjc7dmTKLVO1ilO7FfUzSd16jvgpBmyNtfQDxHQeStzXLjC0Gf6l-2v/pub

Does the transcript exactly match the recording audio? ☐

Does the prompt follow the location specifications, conversation type and category? ☐

Is the recording noise level accurate? ☐

Save and Continue

3. **CHECK THE ENVIRONMENT DESCRIPTION:**
   a. Make sure that the Env Description is compatible with both audio and video and has enough detail. Correct or add to it if necessary.

4. **CHECK TOPIC CATEGORY AND SUBCATEGORY:**

```
19      "distribution": {
20        "recordingLocation": "Outdoor Public Environments",
21        "category": "High-Quality Conversation",
22        "subcategory": "Narrative & Story-Sharing (Realistic Contexts)",
23        "tone": "Neutral"
```

   a. Read the prompt and go back to the "Output" section > Line 22/23.
   b. Make sure that Category and Subcategory is correct.

         i.   **IF Category is wrong:**

             1. **Fix > 2/5 and approve.**

         ii.   **IF Subcategory is wrong:**

             1. **If it's 50/50 or 60/40 fits to two different options:**

                 a. Fix > don't deduct points > approve > give feedback and explain.

             2. **If it's 75/25 and chosen one is %25 one:**

                 a. Fix > 2/5 > approve > explain the reason.

         iii.   You can make the fix at the end of the task, if you won't even make a fix, pls make sure to choose the correct Category and Subcategory. Always choose that part, never skip:

Is this prompt usable?

Unusable prompts are

- too simple (e.g. Hello!)

- mention recent events (last night's football game)

[   ]

Do you believe the audio recording took place in this location?

Note: anything other than indoor_quiet-room should have some background noise indicating as such

[   ]

Is the below category/subcategory correct?

If not, select the correct choice

Choice Path:

[Search choices...   ]

> High-Quality Conversation

> Knowledge-Based Conversation

> Role-Play & Sound Effects

SBQ RUBRIC ⬅️BACK

Rate	Criteria
2	● The CB is not speaking English ● The prompt sounds scripted or unnatural ● The prompt asks for 'recent' or time-dependent events ● The environment/location doesn't match (check video) ● The prompt is shorter than 5 seconds ● There is a song playing in the background with very clear lyrics/vocals (Copyright) ● There is a clear human voice in the background, either from a TV or from a 3rd party. The spoken words are very clear and intelligible
1	● The task is spam

FIX AND APPROVE RUBRIC ✅

Rate	Criteria
5	● The noise level of the actual audio is correct ● The transcript is perfect or requires one minor edit ● The category is perfect

	● The environment description is perfect
4	● The transcript requires two minor edits ● Recording is between 5-14 seconds
3	● The noise level of the actual audio is off by 1 point ● The transcript 3 or more of minor edits (punctuation, ellipses, etc. Words DO NOT COUNT as a minor edit) ● The environment description is missing 1-2 significant details.
2	● The noise level of the actual audio is off by more than 2+ points ● The transcript has a wrong word or is missing a word ● The transcript is missing 1+ filler words ● The category is incorrect ● The environment description is inaccurate/missing 3+ details ● The location in UI is wrong