In [2]:
```
import pandas as pd

# Load annotations
annotations_df = pd.read_csv("annotations.csv")

# Group by filename and annotator to count annotations per annotator per file
grouped = annotations_df.groupby(['filename', 'annotator']).size().reset_index(n

# Now count how many unique annotators per file
annotator_counts = grouped.groupby('filename')['annotator'].nunique().reset_inde

# Merge back to get total annotations per file
annotation_counts = annotations_df.groupby('filename').size().reset_index(name='
merged = pd.merge(annotator_counts, annotation_counts, on='filename')

# Filter files with at least 2 annotators and multiple annotations
candidate_files = merged[(merged['num_annotators'] >= 2) & (merged['total_annota

# Show top 2 files
top_candidates = candidate_files.sort_values(by='num_annotators', ascending=Fals
print("Top candidate recordings with multiple annotators and annotations:")
print(top_candidates)
```

```
Top candidate recordings with multiple annotators and annotations:
       filename  num_annotators  total_annotations
3571  407115.mp3               3                 10
2800  352781.mp3               3                  7
```

In [4]:
```
# Filter annotations for the two selected files
selected_files = ["407115.mp3", "352781.mp3"]
selected_annotations = annotations_df[annotations_df["filename"].isin(selected_f

# Sort for readability
selected_annotations = selected_annotations.sort_values(by=["filename", "annotat

# Display the annotations grouped by file and annotator
for filename in selected_files:
    print(f"\n===== Annotations for {filename} =====")
    file_anns = selected_annotations[selected_annotations["filename"] == filenam
    for annotator in file_anns["annotator"].unique():
        print(f"\n-- Annotator: {annotator} --")
        ann = file_anns[file_anns["annotator"] == annotator][["onset", "offset",
        print(ann.to_string(index=False))
```

```
===== Annotations for 407115.mp3 =====

-- Annotator: 102834162874743039972126867430893976507845628340201871147467671756946315453713 --
    onset    offset                                  text
 0.000000 29.087982 The sound of a lively crowd outdoors.
16.856425 17.404106              A muffled tonal signal.

-- Annotator: 628837653068003366749602990709971684853876258402751854430129370711544060713 77 --
    onset    offset                                                    text
 0.037776 29.030748                       people talking in the background
 0.094440 29.030748       continuous sewing machine noise in the background
16.885809 18.132413                    horn honking once at a moderate volume
21.532240 22.401085 wheel scratching slightly sharp noise in the background

-- Annotator: 847047028104922035705084557572766621947756856102315783090111327917283162536 85 --
    onset    offset                                               text
 0.000000 22.525530 the sound of somebody sweeping in the background
 0.000000 29.030748                 people talking in the background
16.829419 17.476705                                 a honking sound
21.457510 22.266616                                 a beeping sound

===== Annotations for 352781.mp3 =====

-- Annotator: 411636573135802475484269253664194870658670826351810573916181360144468290346 --
 onset    offset                                                  text
   0.0 16.451995 Muffled voices of people having a conversation on a bus
   0.0 16.451995     A low hum of a bus driving steadily across the road

-- Annotator: 578350862087426933555573125591963038214749112007703806530322237977342586148 86 --
   onset    offset
text
0.028448 16.400499 White noise is heard in the close space, a constant, soft hiss filling the air
0.056897 16.400499                   People are talking nearby, their voices unclear and muddled.

-- Annotator: 940579719615463598462294314623876852718833849434465234293568023733472074816 51 --
   onset    offset
text
0.000000 16.400499 An engine of a heavy vehicle is running silently in the background
0.018819 16.400499                   Multiple people speaking close by evenly indoors
1.097183 16.400499                        A radio is silently playing in the distance
```

**(a) Temporal & Textual Annotation Comparison**

**Audio 1 (407115.mp3)**:

**Temporal Overlap**:

All annotators agree on a long segment (0–29s) containing multiple events.

Annotators 2 and 3 both highlight sound events between ~16s and 22s (honking, beeping, scratching).

Differences in segment precision — Annotator 1 uses broader intervals, while Annotators 2 & 3 mark finer-grained events.

**Textual Similarity & Differences**:

All three detect people talking — described as “lively crowd,” “people talking,” and “somebody sweeping” (possibly mistaken interpretation).

Annotators 2 & 3 label overlapping sound events like honks or beeps, but use slightly different language ("horn honking", "a honking sound", "a beeping sound").

Annotator 2 uniquely identifies a sewing machine and wheel scratching, indicating higher event resolution or different interpretation.

**Audio 2 (352781.mp3)**:

**Temporal Annotations (Onset/Offset)**

High agreement across annotators on the full duration of the audio (approx. 0 to 16.4s).

Small differences in onset (e.g., 0.0 vs. 0.028 or 0.056) are minimal and likely due to human precision.

All annotators treat the clip as a continuous ambient scene rather than short, isolated events.

**Similarities**:

All three annotators identify:

Human speech: Described as "muffled", "unclear", "multiple people"

Background noise: Captured as "white noise", "engine", "hum"

All are describing a confined indoor-like sound environment (e.g., inside a bus or a room)

**Differences**:

A1 explicitly identifies the context ("on a bus")

A2 generalizes to "close space" and "white noise"

A3 adds more detail, identifying a radio and engine with specific qualifiers ("silently", "evenly indoors")

In [11]:
```
metadata_df = pd.read_csv("metadata.csv")

# Show full column content (for keywords)
pd.set_option("display.max_colwidth", None)
```

```
print("Keywords for 407115.mp3:", metadata_df.loc[metadata_df["filename"] == "40
print("Keywords for 352781.mp3:", metadata_df.loc[metadata_df["filename"] == "35
```

```
Keywords for 407115.mp3: Uganda, people, sewing, Africa, textile, mall, machines,
corridor, shops, market
Keywords for 352781.mp3: engine, field-recording, people, bus, voices, talking, r
adio, vehicle, ambience
```

**b)**

**407115.mp3**

**Matches / Reliance** : "people talking", "lively crowd" → matches people, market, mall

"sewing machine noise" → directly reflects sewing, machines, textile

**Deviations / Additions**:

"horn honking", "beeping", "wheel scratching" → Not present in keywords

These are foreground events perceived by annotators but not described in metadata

**Conclusion**:

Annotations rely heavily on keywords for background context (people, sewing, market scene)

Annotators go beyond the metadata by identifying short, foreground sound events (e.g., honks)

**352781.mp3**

**Matches / Reliance**:

"engine of a heavy vehicle", "bus driving" → matches engine, bus, vehicle

"people speaking", "voices", "talking" → matches people, voices, talking

"radio playing" → directly matches radio

"white noise", "ambience" → aligns with ambience

**Deviations / Additions**:

Very few! Maybe "white noise" adds a perceptual detail, but it's covered by ambience

**Conclusion**:

This is a textbook example of alignment

Annotations match nearly all metadata keywords, showing strong reliance and task alignment

**c)**

For 407115.mp3 (Urban Environment with Crowd and Sewing Sounds)

Temporal Annotation Compliance:

Onset and Offset: The annotators followed the rule of annotating distinct sounds. For example:

People talking: Different annotators split the crowd noise into multiple events (e.g., a continuous sound, then specific people talking). The annotations' onset/offset reflects this continuous nature of the sound, as suggested by the task (e.g., "people talking" marked over a longer period).

Sewing machine: This was marked as a continuous sound with an onset at the start and an offset at the end of the sound.

Honking, beeping, and wheel scratching: These are separate events, as per the task's rule of separating events when noticeable pauses are present. Annotators made distinct annotations for each.

The temporal boundaries are consistent and correct based on the task's example. For instance, the “sewing machine noise” annotation lasts throughout the segment where the machine sound is heard, fitting the continuous sound rule.

Textual Annotation Compliance:

Specific Descriptions: The descriptions follow the task’s required level of detail:

Source and Action: Each annotation specifies the source of the sound (e.g., "sewing machine noise," "people talking").

Descriptor: Descriptions like "lively," "muffled," and "tonal" provide insight into how the sounds were perceived.

Context: Annotators provide an environmental context where relevant (e.g., “crowd outdoors,” “in the background,” “indoors”).

The annotations adhere to the one description for one sound rule (e.g., separate descriptions for "honking," "wheel scratching," and "people talking").

**Conclusion**:

The temporal and textual annotations are fully compliant with the task description. The annotators correctly followed guidelines for separating sounds, providing detailed annotations, and aligning their work with the task’s structure.

For 352781.mp3 (Bus Environment with People Talking and Engine Sounds)

Temporal Annotation Compliance:

Onset and Offset: The annotators marked the entire recording as a single region for continuous sounds like engine hum and ambient voices, which aligns with the continuous sound rule (e.g., "low hum of a bus driving steadily").

In contrast, short events like horn honking or radio playing were marked separately, showing the annotators correctly handled short, separate sound events and gaps in between.

The temporal boundaries for each sound event are appropriate, and the overlapping sounds (e.g., bus engine and talking) are managed correctly.

Textual Annotation Compliance:

Specific Descriptions: The descriptions follow the task's rule of detailed annotations:

Source and Action: Each annotation specifies the source of the sound (e.g., "engine of a heavy vehicle," "voices of people").

Descriptor: Descriptions like "hissing," "soft," and "muffled" describe how the sound was perceived.

Context: The context is well-defined, with mentions of "indoors," "in the distance," and "close by."

No multiple events are described in a single annotation, meaning each description corresponds to one sound event.

**Conclusion**:

The temporal and textual annotations for this recording are also fully compliant with the task description. Annotators successfully separated distinct sounds, followed the rules for continuous versus discrete events, and provided detailed textual descriptions.