Hondurasgate: audio analysis and authentication of twelve recordings

Report by Earshot

July 25, 2026

Investigation in brief

By identifying audio artefacts such as breaths, vocal hesitations, background noises and microphone distortions, and by performing machine-learning voice comparison for AI generated speech, Earshot concludes the following:

- Recordings number 146, 088, 114, 086, 107, and 030 are **highly likely** authentic recordings of Juan Orlando Hernández and not AI generated;
- Recordings number 168, 175, and 106 are **highly likely** authentic recordings of Nasry Asfura and not AI generated;
- Recordings number 223 and 190 are **highly likely** authentic recordings of María Antonieta Mejía and not AI generated;
- Recording number 147 is **highly likely** an authentic recording of Cosette López and not AI generated.

Contents

1. Investigation in depth 4
2. Methodology 6
2.1 Critical listening for audio artefacts 6
2.2 Machine-learning voice comparison 7
3. Juan Orlando Hernández 9
art4-FILTRACION_146.ogg 9
art5-FILTRACION_088.ogg 9
art4-FILTRACION_114.ogg 10
art5-FILTRACION_086.ogg 11
art5-FILTRACION_107.ogg 11
art4-FILTRACION_030.ogg 12
4. Nasry Asfura 14
art6-FILTRACION_168.ogg 14
art6-FILTRACION_175.ogg 14
art4-FILTRACION_106.ogg 15
5. María Antonieta Mejía 17
art4-FILTRACION_223.ogg 17
art5-FILTRACION_190.ogg 18
6. Cosette López 19
art4-FILTRACION_147.ogg 19
7. Appendix: artefact timestamps 20
Juan Orlando Hernández 20
Nasry Asfura 27
María Antonieta Mejía 32
Cosette López 35

1. Investigation in depth

Earshot analyzed 12 recordings with the alleged voices of Juan Orlando Hernández, Nasry Asfura, Cosette López, and María Antonieta Mejía.[1] Earshot concludes that all 12 recordings are highly likely authentic and not AI generated. Earshot reached its conclusions through two lines of analysis: critical listening for audio artefacts consistent with authentic speech, and machine-learning voice comparison with known recordings of the alleged speakers.

Earshot identified multiple audio artefacts across each of the 12 recordings that are consistent with natural speech recorded in real-world conditions. These artefacts belong to two categories: speech artefacts and acoustic artefacts. Earshot identified speech artefacts such as breaths and vocal hesitations heard between sentences, and identified acoustic artefacts such as background noises and microphone distortions (see paragraph 2). The presence of multiple artefacts in a single recording point to a coherent acoustic web that is consistent with speech recorded in real-world conditions. However, advances in AI generated audio are progressively incorporating these sounds into the generative process. These advances pose challenges to methodologies of authentication. To address these challenges, Earshot is continually testing the reproducibility of speech and acoustic artefacts by using voice cloning programs such as Coqui (see chapter 2).[2] This methodology allows Earshot to assess the reliability of each artefact in determining a recording's authenticity. Earshot identified several artefacts that current AI generative programs do not reproduce, such as breaths preceded by the distinct sound of the tapping of teeth or lips, yawns, and background noises with the same room resonance as the voice (see chapter 2). Together with the coherent acoustic web identified across all 12 recordings, Earshot finds these artefacts unlikely to have been reproduced through AI generated speech. See the appendix for the full list of timestamped artefacts.

Earshot supplemented the analysis of audio artefacts through a second line of analysis. Earshot performed a machine-learning voice comparison using the program 'Resemblyzer' for all the 12 recordings. Resemblyser analyses short samples of a voice and translates its complex patterns of rhythm, prosody, frequency composition and harmonics into a numerical value. This value is compared to the value of a known recording of the speaker in question. The comparison between the values of an alleged voice and a known voice produces a score that determines the similarity between the two voices and is an indication of the alleged speaker's authenticity (see paragraph 3).[3] The findings from this analysis supported the conclusions derived from critical listening across all 12 recordings.

Additional evidence of the recordings' authenticity is seen through the relatively low sample rate found across all 12 recordings. Sample rate is how many times per second a microphone measures sound pressure. Digital audio is typically recorded at 44,100 or 48,000 samples per second. These higher sample rates capture a wider range of frequencies and therefore more detail.

[1] https://hondurasgate.ch/fact-checking (Accessed July 2026).
[2] https://github.com/-ai/tts (Accessed July 2026).
[3] https://github.com/resemble-ai/resemblyzer (Accessed July 2026).

Lower sample rates are common in telephonic audio, which only needs to capture speech frequencies (0 to 8,000 Hz) rather than the full audible range (0 to 20,000 Hz). All 12 recordings contain frequencies only up to 4,500 Hz, indicating they were recorded at a relatively low sample rate. Earshot's ongoing experiments have not encountered AI generated audio with such a low sample rate. While it is possible to manually reduce any recording's sample rate, Earshot finds this likely a product of authentic telephonic conversations.

The likelihood of Earshot's conclusions was derived from a quantitative analysis of the number of audio artefacts identified in each recording and from the score obtained through the machine-learning voice comparison process.

Earshot's findings are limited by the low telephonic quality of all 12 recordings. Furthermore, ongoing advances in machine-learning algorithms is constantly improving the quality of AI generated speech and making its detectability more difficult. Further speech analysis on accent and dialect should be conducted by a Spanish-speaking linguist.

2. Methodology

2.1 Critical listening for audio artefacts

2.1.1 Speech artefacts: breaths

The presence of breaths between sentences is typically an indication that a voice is likely authentic. AI voice generation focuses predominantly on replicating the unique frequencies that give a particular voice its timbre and prosody. While it is also possible to generate sounds such as breaths, these sounds often have to be individually added to a text prompt. Moreover, Earshot's previous investigations in AI generated speech identified digital artefacts that distinguish breaths generated using AI from authentic breathing. These digital artefacts appear as rapid, successive dips in volume that resemble a tremolo effect.

Closer analysis of the breaths of a recording can help to identify 'breath profiles' consistent with authentic speech. A speaker's 'breath profile' can be established when multiple breaths contain similar patterns in frequency content, duration and dynamics. One such breath profile is found when breaths are characterised by a brief tapping of the lips or teeth preceding the intake of air. This profile was identified on several of the recordings analysed. Earshot tested the reproducibility of this breath profile in AI generated speech by using the voice cloning program Coqui. By training Coqui with samples of speech containing this unique breath profile Earshot found that the generated audio contained generic breaths instead of the breath profile of the audio used for training. The presence of this breath profile in a suspect recording is therefore an initial indication of a recording's authenticity.

2.1.2 Speech artefacts: vocal hesitations

The presence of vocal hesitations such as 'um' or 'eh' are an indication that the voice is likely authentic. Similarly to the sound of breaths, it is possible to generate sounds of vocal hesitations using machine-learning algorithms by individually adding these hesitations to a text prompt. Earshot finds it more likely that the presence of these vocal hesitations, in combination with the sounds of breaths, is a natural product of authentic speech rather than AI generated speech.

2.1.3 Acoustic artefacts: microphone distortions

Microphone distortions are typically caused by strong exhalations by the speaker directly onto the microphone, either from direct breathing or from pronouncing high-pressure 'plosive' consonants such as 'p', 't' or 'k'. Plosive consonants and exhaled air cause the audio signal to distort under the pressure exerted onto the microphone's sensitive diaphragm. By using the voice cloning program Coqui, Earshot tested if AI generated speech is capable of reproducing such microphonic distortions. Earshot trained Coqui using samples of audio containing microphonic distortions and found that AI can only generate microphonic distortions on plosive consonants. This finding indicates that microphonic distortions around plosive consonants cannot be considered alone as evidence of the authenticity of a recording. The presence of distortions that are not produced by

plosive consonants however can corroborate the overall authenticity of these acoustic artefacts, and therefore, of the recording analysed.

2.1.4 Acoustic artefacts: background noises

The presence of background noises is an indication of a rich acoustic web consistent with recordings made in real-world conditions. While these sounds can be manually edited in a recording, as well as individually prompted in the machine-learning generation process, Earshot has not encountered such cases in previous investigations and finds it more likely that these sounds are a product of authentic recordings.

Additionally, background noises containing the same room resonance as a voice are a strong indication that the recording was made in real-world conditions. Room resonance is produced by the unique architecture of the space where a recording is made. Every enclosed space has a 'modal' resonance, or a set of frequencies at which sound waves reflect and reinforce each other. This is determined by the room's dimensions, shape, and construction materials. When a sound containing those specific frequencies is introduced into the room, the room is acoustically 'excited', and those frequencies ring out longer and sound louder than others.

Continuous background noises such as air conditioning or ambient room tones help to establish a consistent noise floor across a recording and are an indication that a recording is likely authentic. Telephonic conversations are often heavily compressed such the noise floor is suppressed in order to prioritise the intelligibility of speech. When identifiable, the presence of a consistent noise floor is a strong indication that a recording was made in real-world conditions and is likely authentic and not AI generated.

2.2 Machine-learning voice comparison

Earshot supplemented its findings from critical listening by using the machine-learning voice comparison program 'Resemblyzer'. Resemblyzer works by analysing a voice recording and translating the voice's complex characteristics (fundamental frequencies, harmonics, tone, timbre, rhythm, cadence, etc.) into a single 'summary' numerical value. This summary value is compared with the summary value from a known recording of the speaker in question. This comparison is scored from 0 to 1. A score of 0 indicates the two voices do not share any characteristics and are either two different speakers, or that alleged voice has been generated with AI. Conversely, a score of 1 indicates that the two voices share exactly the same characteristics and are therefore the same speaker.[4]

Additionally, when using Reesemblyzer Earshot compares the alleged speakers with two decoy voices in order to establish a benchmark value on the 0 to 1 scoring scale. Scores above this

[4] https://github.com/resemble-ai/resemblyzer (Accessed July 2026).

benchmark value are considered an initial indication that the alleged voices match the known voices, and that therefore the alleged voices are likely authentic and not AI generated.

3. Juan Orlando Hernández

art4-FILTRACION_146.ogg

Earshot identified a total of 15 audio artefacts (6 breaths, 1 vocal hesitation, 5 microphone distortions, and 3 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1). Additionally, Earshot identified a breath profile that further suggests that the recording is authentic. 3 out of 6 breaths are characterised by a brief tapping of the lips or teeth preceding the intake of air. Earshot identified the same breath profile in known recordings of Hernández. As mentioned in paragraph 2.1.1, although breaths can be reproduced by AI, Earshot has not encountered AI generated breaths preceded by the tapping of lips or teeth. The presence of a breath profile in this recording that matches the breath profile of Hernández suggests the recording is highly likely authentic and not AI-generated.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Hernández.[5] This comparison produced an average score of 0.78. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.48. This score provides a benchmark value above which the score between the alleged voice and the known voice of Hernández can be considered a match and therefore authentic. The high score of 0.78 between the alleged voice and the known voice of Hernández, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

art5-FILTRACION_088.ogg

Earshot identified a total of 47 audio artefacts (27 breaths, 1 vocal hesitation, 4 microphone distortions, and 15 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1). Additionally, Earshot identified a yawn at 00:27.572 followed by a characteristic 'yawningly' speech, caused by the relaxed and open muscles of the jaw and mouth. Earshot used voice cloning program Coqui to test the reproducibility of the combination of a yawn and the following 'yawningly' speech. Earshot did not encounter either of these two sounds, suggesting the yawn and the following speech in this recording are highly likely products of authentic speech.

[5] https://youtu.be/ufNkSDx24EA?si=6bpMHw1ugYX2piOk (Accessed July 2026).

A closer analysis of the background noises strengthens the conclusion that the recording is highly likely authentic. 3 out of 15 background noises have the same resonance as the voice of the speaker, observed between 300 and 600 Hz. As mentioned in paragraph 2.1.4, this resonance is unlikely to have been produced through AI generated audio and suggests that the noises with this resonance are produced from the same room as the speaker in authentic, real-world conditions.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Hernández.[6] This comparison produced an average score of 0.78. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.49. This score provides a benchmark value above which the score between the alleged voice and the known voice of Hernández can be considered a match and therefore authentic. The high score of 0.78 between the alleged voice and the known voice of Hernández, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

art4-FILTRACION_114.ogg

Earshot identified a total of 33 audio artefacts (19 breaths, 2 vocal hesitations, and 12 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1). Additionally, Earshot identified a breath profile that further suggests that the recording is authentic. 6 out of 19 breaths are characterised by a brief tapping of the lips or teeth preceding the intake of air. Earshot identified the same breath profile in known recordings of Hernández. As mentioned in paragraph 2.1.1, although breaths can be reproduced by AI, Earshot has not encountered AI generated breaths preceded by the tapping of lips or teeth. The presence of a breath profile in this recording that matches the breath profile of Hernández suggests the recording is highly likely authentic and not AI-generated.

A closer analysis of the background noises strengthens the conclusion that the recording is highly likely authentic. 3 out of 12 background noises have the same resonance as the voice of the speaker, observed between 300 and 600 Hz. As mentioned in paragraph 2.1.4, this resonance is unlikely to have been produced through AI generated audio and suggests that the noises with this resonance are produced from the same room as the speaker in authentic, real-world conditions.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Hernández.[7] This comparison produced an average score of 0.72. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.46. This score provides a benchmark value above

[6] https://youtu.be/ufNkSDx24EA?si=6bpMHw1ugYX2piOk (Accessed July 2026).
[7] https://youtu.be/ufNkSDx24EA?si=6bpMHw1ugYX2piOk (Accessed July 2026).

which the score between the alleged voice and the known voice of Hernández can be considered a match and therefore authentic. The high score of 0.72 between the alleged voice and the known voice of Hernández, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

art5-FILTRACION_086.ogg

Earshot identified a total of 41 audio artefacts (22 breaths, 1 vocal hesitation, 6 microphone distortions, and 12 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1). Additionally, Earshot identified a breath profile that further suggests that the recording is authentic. 7 out of 22 breaths are characterised by a brief tapping of the lips or teeth preceding the intake of air. Earshot identified the same breath profile in known recordings of Hernández. As mentioned in paragraph 2.1.1, although breaths can be reproduced by AI, Earshot has not encountered AI generated breaths preceded by the tapping of lips or teeth. The presence of a breath profile in this recording that matches the breath profile of Hernández suggests the recording is highly likely authentic and not AI-generated.

A closer analysis of the background noises strengthens the conclusion that the recording is highly likely authentic. 4 out of 12 background noises have the same resonance as the voice of the speaker, observed between 300 and 600 Hz. As mentioned in paragraph 2.1.4, this resonance is unlikely to have been produced through AI generated audio and suggests that the noises with this resonance are produced from the same room as the speaker in authentic, real-world conditions.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Hernández.[8] This comparison produced an average score of 0.75. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.48. This score provides a benchmark value above which the score between the alleged voice and the known voice of Hernández can be considered a match and therefore authentic. The high score of 0.75 between the alleged voice and the known voice of Hernández, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

art5-FILTRACION_107.ogg

Earshot identified a total of 23 audio artefacts (11 breaths, 6 microphone distortions, and 6 background noises).

[8] https://youtu.be/ufNkSDx24EA?si=6bpMHw1ugYX2piOk (Accessed July 2026).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1). Earshot identified 1 breath that is characterised by a brief tapping of the lips or teeth preceding the intake of air. Earshot identified the same types of breaths in known recordings of Hernández. As mentioned in paragraph 2.1.1, although breaths can be reproduced by AI, Earshot has not encountered AI generated breaths preceded by the tapping of lips or teeth. The presence of a breath profile in this recording that matches the breath profile of Hernández suggests the recording is highly likely authentic and not AI-generated.

A closer analysis of the background noises strengthens the conclusion that the recording is highly likely authentic. 2 out of 6 background noises have the same resonance as the voice of the speaker, observed between 300 and 600 Hz. As mentioned in paragraph 2.1.4, this resonance is unlikely to have been produced through AI generated audio and suggests that the noises with this resonance are produced from the same room as the speaker in authentic, real-world conditions.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Hernández.[9] This comparison produced an average score of 0.77. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.49. This score provides a benchmark value above which the score between the alleged voice and the known voice of Hernández can be considered a match and therefore authentic. The high score of 0.77 between the alleged voice and the known voice of Hernández, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

art4-FILTRACION_030.ogg

Earshot identified a total of 12 audio artefacts (7 breaths, 1 vocal hesitation, 1 microphone distortion, and 3 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1). Additionally, Earshot identified a breath profile that further suggests that the recording is authentic. 2 out of 7 breaths are characterised by a brief tapping of the lips or teeth preceding the intake of air. Earshot identified the same breath profile in known recordings of Hernández. As mentioned in paragraph 2.1.1, although breaths can be reproduced by AI, Earshot has not encountered AI generated breaths preceded by the tapping of lips or teeth. The presence of a breath profile in this recording that matches the breath profile of Hernández suggests the recording is highly likely authentic and not AI-generated.

[9] https://youtu.be/ufNkSDx24EA?si=6bpMHw1ugYX2piOk (Accessed July 2026).

Earshot identified a consistent noise floor throughout the recording. As mentioned in chapter 2.1.4, this noise floor is characterised by broadband frequency content caused by the natural ambient conditions of the room where the recording was made. Unless this sound was manually inserted in the recording, the presence of a continuous noise floor is additional evidence that the recording was made in a real-world environment and is highly likely authentic and not AI generated.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Hernández.[10] This comparison produced an average score of 0.76. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.49. This score provides a benchmark value above which the score between the alleged voice and the known voice of Hernández can be considered a match and therefore authentic. The high score of 0.76 between the alleged voice and the known voice of Hernández, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

[10] https://youtu.be/ufNkSDx24EA?si=6bpMHw1ugYX2piOk (Accessed July 2026).

4. Nasry Asfura

art6-FILTRACION_168.ogg

Earshot identified a total of 48 audio artefacts (25 breaths, 4 microphone distortions, and 19 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1). Additionally, Earshot identified a breath profile that further suggests that the recording is authentic. 6 out of 25 breaths are characterised by a brief tapping of the lips or teeth preceding the intake of air. Earshot identified the same breath profile in known recordings of Asfura. As mentioned in paragraph 2.1.1, although breaths can be reproduced by AI, Earshot has not encountered AI generated breaths preceded by the tapping of lips or teeth. The presence of a breath profile in this recording that matches the breath profile of Asfura suggests the recording is highly likely authentic and not AI-generated.

Earshot identified a consistent noise floor throughout the recording. As mentioned in chapter 2.1.4, this noise floor is characterised by broadband frequency content caused by the natural ambient conditions of the room where the recording was made. Additionally, Earshot identified the continuous humming sound of an air conditioning unit. Unless these sounds were manually inserted in the recording, the presence of a continuous noise floor is additional evidence that the recording was made in a real-world environment and is highly likely authentic and not AI generated.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Asfura.[11] This comparison produced an average score of 0.77. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.54. This score provides a benchmark value above which the score between the alleged voice and the known voice of Asfura can be considered a match and therefore authentic. The high score of 0.77 between the alleged voice and the known voice of Asfura, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

art6-FILTRACION_175.ogg

Earshot identified a total of 23 audio artefacts (12 breaths, 1 microphone distortions, and 10 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

[11] https://www.youtube.com/shorts/MCx0vFayb-0 (Accessed July 2026).

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1). Additionally, Earshot identified a breath profile that further suggests that the recording is authentic. 3 out of 12 breaths are characterised by a brief tapping of the lips or teeth preceding the intake of air. Earshot identified the same breath profile in known recordings of Asfura. As mentioned in paragraph 2.1.1, although breaths can be reproduced by AI, Earshot has not encountered AI generated breaths preceded by the tapping of lips or teeth. The presence of a breath profile in this recording that matches the breath profile of Asfura suggests the recording is highly likely authentic and not AI-generated.

A closer analysis of the background noises strengthens the conclusion that the recording is highly likely authentic. Earshot identified one background noise that has the same resonance as the voice of the speaker, observed between 300 and 1000 Hz. As mentioned in paragraph 2.1.4, this resonance is unlikely to have been produced through AI generated audio and suggests that the noises with this resonance are produced from the same room as the speaker in authentic, real-world conditions.

Earshot identified a consistent noise floor throughout the recording. As mentioned in chapter 2.1.4, this noise floor is characterised by broadband frequency content caused by the natural ambient conditions of the room where the recording was made. Additionally, Earshot identified the continuous humming sound of an air conditioning unit. Unless these sounds were manually inserted in the recording, the presence of a continuous noise floor is additional evidence that the recording was made in a real-world environment and is highly likely authentic and not AI generated.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Asfura.[12] This comparison produced an average score of 0.78. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.55. This score provides a benchmark value above which the score between the alleged voice and the known voice of Asfura can be considered a match and therefore authentic. The high score of 0.78 between the alleged voice and the known voice of Asfura, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

art4-FILTRACION_106.ogg

Earshot identified a total of 11 audio artefacts (5 breaths, 2 vocal hesitations, and 4 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1).

[12] https://www.youtube.com/shorts/MCx0vFayb-0 (Accessed July 2026).

Earshot identified a consistent noise floor throughout the recording. As mentioned in chapter 2.1.4, this noise floor is characterised by broadband frequency content caused by the natural ambient conditions of the room where the recording was made. Additionally, Earshot identified the continuous humming sound of an air conditioning unit. Unless these sounds were manually inserted in the recording, the presence of a continuous noise floor is additional evidence that the recording was made in a real-world environment and is highly likely authentic and not AI generated.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Asfura.[13] This comparison produced an average score of 0.76. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.53. This score provides a benchmark value above which the score between the alleged voice and the known voice of Asfura can be considered a match and therefore authentic. The high score of 0.76 between the alleged voice and the known voice of Asfura, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

[13] https://www.youtube.com/shorts/MCx0vFayb-0 (Accessed July 2026).

5. María Antonieta Mejía

art4-FILTRACION_223.ogg

Earshot identified a total of 52 audio artefacts (36 breaths, 2 vocal hesitations, 9 microphone distortions, and 5 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1). Additionally, Earshot identified a breath profile that further suggests that the recording is authentic. 7 out of 36 breaths are characterised by a brief tapping of the lips or teeth preceding the intake of air. Earshot identified the same breath profile in known recordings of Mejía. As mentioned in paragraph 2.1.1, although breaths can be reproduced by AI, Earshot has not encountered AI generated breaths preceded by the tapping of lips or teeth. The presence of a breath profile in this recording that matches the breath profile of Mejía suggests the recording is highly likely authentic and not AI-generated.

A closer analysis of the microphone distortions strengthens the conclusion that the recording is highly likely authentic. 4 out of 9 microphone distortions are caused by the speaker pronouncing plosive consonants such as 'p' or 't', while the remaining distortions are caused by the speaker exhaling directly onto the microphone. As mentioned in paragraph 2.1.3, the presence of microphone distortions around plosive consonants cannot alone be used to assess the authenticity of a recording. However, the combined presence of microphone distortions caused by direct exhalation and by plosive consonants suggests the audio is highly likely a recording of real speech.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Mejía.[14] This comparison produced an average score of 0.76. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.58. This score provides a benchmark value above which the score between the alleged voice and the known voice of Mejía can be considered a match and therefore authentic. The high score of 0.76 between the alleged voice and the known voice of Mejía, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

[14] https://youtu.be/64H9YrI3Fds?si=qHMGKH3QMVgPIEUK (Accessed July 2026).

art5-FILTRACION_190.ogg

Earshot identified a total of 17 audio artefacts (10 breaths, 3 microphone distortions, and 4 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1). Additionally, Earshot identified a breath profile that further suggests that the recording is authentic. 6 out of 10 breaths are characterised by a brief tapping of the lips or teeth preceding the intake of air. Earshot identified the same breath profile in known recordings of Mejía. As mentioned in paragraph 2.1.1, although breaths can be reproduced by AI, Earshot has not encountered AI generated breaths preceded by the tapping of lips or teeth. The presence of a breath profile in this recording that matches the breath profile of Mejía suggests the recording is highly likely authentic and not AI-generated.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of Mejía.[15] This comparison produced an average score of 0.74. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.49. This score provides a benchmark value above which the score between the alleged voice and the known voice of Mejía can be considered a match and therefore authentic. The high score of 0.74 between the alleged voice and the known voice of Mejía, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

[15] https://youtu.be/64H9YrI3Fds?si=qHMGKH3QMVgPIEUK (Accessed July 2026).

6. Cosette López

art4-FILTRACION_147.ogg

Earshot identified a total of 8 audio artefacts (2 breaths and 6 background noises).

As mentioned in chapter 2, the presence of multiple artefacts is an indication of a coherent acoustic web consistent with recordings made in real-world conditions.

A closer analysis of the breaths did not identify the digital artefacts that Earshot has encountered in previous investigations of AI-generated speech (see paragraph 2.1.1).

A closer analysis of the background noises strengthens the conclusion that the recording is highly likely authentic. 3 out of 6 background noises have the same resonance as the voice of the speaker, observed between 100 and 800 Hz. As mentioned in paragraph 2.1.4, this resonance is unlikely to have been produced through AI generated audio and suggests that the noises with this resonance are produced from the same room as the speaker in authentic, real-world conditions.

Earshot used Resemblyzer (see chapter 3) to perform a voice comparison between the alleged voice to a known recording of López.[16] This comparison produced an average score of 0.69. In addition, Earshot performed a voice comparison between the alleged voice and two decoy voices. This comparison produced an average score of 0.46. This score provides a benchmark value above which the score between the alleged voice and the known voice of López can be considered a match and therefore authentic. The high score of 0.69 between the alleged voice and the known voice of López, relative to the decoy voices, indicates that the recording is highly likely authentic and not AI generated.

[16] https://www.youtube.com/shorts/RjzLIICCKBI (Accessed July 2026).

7. Appendix: artefact timestamps

Juan Orlando Hernández

art4-FILTRACION_146.ogg		
Background noise	00:03.827	
Breath	00:11.000	
Vocal hesitation	00:11.688	
Microphone distortion	00:14.500	
Breath	00:15.101	Breath preceded by brief tapping of lips or teeth
Microphone distortion	00:23.071	
Breath	00:24.000	Breath preceded by brief tapping of lips or teeth
Microphone distortion	00:27.448	
Breath	00:33.658	
Background noise	00:37.713	
Breath	00:37.925	Breath preceded by brief tapping of lips or teeth
Microphone distortion	00:43.500	
Background noise	00:44.660	
Breath	00:45.000	
Microphone distortion	00:47.282	

art5-FILTRACION_088.ogg		
Background noise	00:01.343	Matching room resonance as speech
Breath	00:03.770	
Breath	00:11.581	
Breath	00:14.000	

Breath	00:17.699	
Breath	00:20.571	
Breath	00:24.837	
Breath	00:27.572	Yawn followed by 'yawningly' speech
Breath	00:39.385	
Breath	00:41.663	
Background noise	00:44.102	
Background noise	00:45.635	
Breath	00:46.440	
Breath	00:53.066	
Vocal hesitation	00:56.843	
Breath	00:58.125	
Background noise	01:03.000	Matching room resonance as speech
Breath	01:05.918	
Breath	01:07.738	
Breath	01:16.577	
Background noise	01:22.097	
Background noise	01:22.889	
Background noise	01:25.818	
Breath	01:26.547	
Breath	01:30.738	
Microphone distortion	01:34.313	
Breath	01:34.746	
Microphone distortion	01:36.813	
Breath	01:39.689	

Background noise	01:43.355	
Breath	01:43.888	
Microphone distortion	01:47.480	
Breath	01:48.099	
Breath	01:52.604	
Breath	01:59.000	
Background noise	02:00.500	
Breath	02:03.000	
Breath	02:05.360	
Background noise	02:09.638	
Breath	02:11.353	
Background noise	02:16.165	Matching room resonance as speech
Background noise	02:16.364	
Background noise	02:17.000	
Microphone distortion	02:19.158	
Background noise	02:20.034	
Breath	02:20.174	
Background noise	02:25.127	

art4-FILTRACION_114.ogg		
Background noise	00:00.000	
Background noise	00:04.761	
Breath	00:21.722	Breath preceded by brief tapping of lips or teeth
Breath	00:34.000	

Background noise	00:34.405	
Background noise	00:34.847	Matching room resonance as speech
Background noise	00:41.327	
Breath	00:46.545	Breath preceded by brief tapping of lips or teeth
Vocal hesitation	00:55.408	
Background noise	00:58.367	
Background noise	01:05.561	Matching room resonance as speech
Background noise	01:06.785	
Breath	01:17.793	
Breath	01:20.731	
Breath	01:26.000	
Vocal hesitation	01:33.520	
Breath	01:46.104	
Breath	01:50.000	Breath preceded by brief tapping of lips or teeth
Background noise	01:54.457	
Breath	01:57.000	Breath preceded by brief tapping of lips or teeth
Breath	02:00.710	
Breath	02:06.000	
Breath	02:16.405	Breath preceded by brief tapping of lips or teeth
Breath	02:20.554	
Breath	02:25.589	
Breath	02:29.636	Breath preceded by brief tapping of lips or teeth
Breath	02:38.615	
Breath	02:48.242	Breath preceded by brief tapping of lips or teeth
Breath	02:50.588	
Background noise	02:56.000	Matching room resonance as speech
Background noise	02:56.300	

Breath	03:05.555	
Background noise	03:09.116	

art4-FILTRACION_114.ogg		
Background noise	00:02.227	Matching room resonance as speech
Background noise	00:03.693	
Breath	00:07.298	Breath preceded by brief tapping of lips or teeth
Breath	00:11.934	
Background noise	00:14.500	
Breath	00:14.769	Breath preceded by brief tapping of lips or teeth
Breath	00:24.500	
Breath	00:28.935	
Microphone distortion	00:35.667	
Breath	00:36.358	Breath preceded by brief tapping of lips or teeth
Breath	00:41.500	
Breath	00:47.823	
Microphone distortion	00:49.500	
Breath	00:50.287	
Background noise	00:54.441	Matching room resonance as speech
Breath	00:54.699	
Breath	01:02.000	Breath preceded by brief tapping of lips or teeth
Microphone distortion	01:05.359	
Background noise	01:12.207	
Breath	01:12.695	
Background noise	01:13.317	

Background noise	01:13.500	
Breath	01:15.821	Breath preceded by brief tapping of lips or teeth
Breath	01:18.568	Breath preceded by brief tapping of lips or teeth
Background noise	01:23.000	
Background noise	01:26.095	
Breath	01:32.420	
Breath	01:34.403	
Breath	01:37.581	
Background noise	01:44.397	Matching room resonance as speech
Background noise	01:45.209	Matching room resonance as speech
Microphone distortion	01:51.130	
Microphone distortion	01:54.811	
Breath	01:55.000	
Background noise	01:56.342	
Vocal hesitation	01:59.697	
Breath	02:00.054	
Breath	02:02.500	
Microphone distortion	02:05.116	
Breath	02:06.000	
Breath	02:11.868	Breath preceded by brief tapping of lips or teeth

art5-FILTRACION_107.ogg		
Breath	00:04.414	
Background noise	00:08.104	Matching room resonance as speech
Breath	00:10.129	Breath preceded by brief tapping of lips or teeth
Background noise	00:16.149	

Microphone distortion	00:21.000	
Background noise	00:26.766	
Background noise	00:29.249	Matching room resonance as speech
Microphone distortion	00:30.051	
Breath	00:32.000	
Microphone distortion	00:42.260	
Breath	00:42.884	
Breath	00:46.000	
Breath	00:49.528	
Breath	00:55.812	
Breath	01:01.109	
Background noise	01:02.554	
Microphone distortion	01:06.678	
Microphone distortion	01:10.500	
Microphone distortion	01:12.728	
Breath	01:20.000	
Breath	01:24.230	
Breath	01:27.565	
Background noise	01:33.207	

art4-FILTRACION_030.ogg		
Breath	00:00.157	
Microphone distortion	00:01.583	
Breath	00:04.180	Breath preceded by brief tapping of lips or teeth

Vocal hesitation	00:05.060	
Breath	00:09.075	
Breath	00:13.210	
Breath	00:15.540	Breath preceded by brief tapping of lips or teeth
Background noise	00:17.763	
Breath	00:21.192	
Background noise	00:27.215	
Background noise	00:27.767	
Breath	00:31.640	

Nasry Asfura

art6-FILTRACION_168.ogg		
Background noise	(full recording)	Consistent noise floor
Breath	00:01.70882086	Breath preceded by brief tapping of lips or teeth
Breath	00:04.10607710	
Breath	00:06.39693878	
Breath	00:09.29195011	
Breath	00:11.59176871	
Breath	00:15.54945578	
Breath	00:19.80832200	
Background noise	00:23.50000000	
Microphone distortion	00:24.92795918	
Background noise	00:25.16321995	
Background noise	00:25.50521542	

Breath	00:29.86714286	
Breath	00:34.89108844	
Breath	00:39.40024943	Breath preceded by brief tapping of lips or teeth
Breath	00:42.86800454	Breath preceded by brief tapping of lips or teeth
Background noise	00:50.54038549	
Background noise	00:50.68242630	
Breath	00:50.84133787	
Background noise	00:51.67303855	
Breath	00:53.81827664	
Breath	01:00.40600907	Breath preceded by brief tapping of lips or teeth
Breath	01:05.55362812	
Breath	01:10.34446712	
Background noise	01:11.69310658	
Background noise	01:12.23460317	
Breath	01:12.50000000	Breath preceded by brief tapping of lips or teeth
Background noise	01:12.86235828	
Background noise	01:13.38485261	
Background noise	01:13.70190476	
Breath	01:17.10052154	
Background noise	01:19.30639456	
Breath	01:21.00000000	
Background noise	01:22.63263039	
Breath	01:23.85151927	
Microphone distortion	01:24.04274376	
Breath	01:26.11757370	

Microphone distortion	01:26.50000000	
Breath	01:30.55961451	
Background noise	01:30.69136054	
Background noise	01:32.64297052	
Breath	01:32.84478458	Breath preceded by brief tapping of lips or teeth
Background noise	01:35.08306122	
Microphone distortion	01:36.06569161	
Breath	01:36.23176871	
Background noise	01:38.54764172	
Background noise	01:42.36217687	
Breath	01:42.58006803	

art6-FILTRACION_175.ogg		
Background noise	(full recording)	Consistent noise floor
Breath	00:00.000	
Background noise	00:00.176	
Breath	00:04.597	
Background noise	00:08.683	
Breath	00:10.622	
Background noise	00:10.916	
Microphone distortion	00:13.707	
Breath	00:13.843	
Breath	00:18.160	Breath preceded by brief tapping of lips or teeth
Breath	00:22.460	

Breath	00:27.000	
Breath	00:30.389	Breath preceded by brief tapping of lips or teeth
Background noise	00:33.500	
Background noise	00:33.820	
Breath	00:36.390	
Background noise	00:38.310	
Breath	00:38.585	Breath preceded by brief tapping of lips or teeth
Background noise	00:41.112	
Background noise	00:41.521	Matching room resonance as speech
Breath	00:44.338	
Breath	00:47.567	
Background noise	00:53.671	

art4-FILTRACION_106.ogg		
Background noise	(full recording)	Consistent noise floor
Breath	00:08.256	
Background noise	00:09.643	
Breath	00:15.752	
Vocal hesitation	00:17.394	
Breath	00:30.744	
Background noise	00:31.849	
Breath	00:38.080	
Background noise	00:38.944	
Breath	00:43.233	
Vocal hesitation	00:45.364	

María Antonieta Mejía

art4-FILTRACION_223.ogg		
Breath	00:02.311	Breath preceded by brief tapping of lips or teeth
Breath	00:07.000	
Breath	00:09.514	
Breath	00:12.248	Breath preceded by brief tapping of lips or teeth
Microphone disortion	00:15.046	Microphone distoriton on plosive consonant
Breath	00:18.632	
Breath	00:24.708	Breath preceded by brief tapping of lips or teeth
Breath	00:29.669	Breath preceded by brief tapping of lips or teeth
Breath	00:33.803	
Background noise	00:34.154	
Vocal hesitation	00:36.000	
Background noise	00:37.000	
Breath	00:38.357	
Microphone disortion	00:40.000	
Breath	00:44.114	
Breath	00:46.000	
Breath	00:47.792	
Breath	00:54.602	
Background noise	00:57.439	
Breath	01:01.401	
Breath	01:03.597	
Breath	01:08.084	
Breath	01:15.000	
Breath	01:18.260	

Background noise	01:20.196	
Background noise	01:22.432	
Microphone disortion	01:24.341	Microphone distoriton on plosive consonant
Breath	01:28.281	
Microphone disortion	01:31.728	Microphone distoriton on plosive consonant
Breath	01:33.000	
Microphone disortion	01:37.245	Microphone distoriton on plosive consonant
Breath	01:38.500	
Breath	01:47.421	Breath preceded by brief tapping of lips or teeth
Microphone disortion	01:53.059	
Breath	01:53.245	
Breath	01:56.317	Breath preceded by brief tapping of lips or teeth
Breath	02:02.137	
Vocal hesitation	02:02.867	
Breath	02:09.141	
Microphone disortion	02:14.115	
Breath	02:14.200	
Microphone disortion	02:18.322	
Breath	02:18.400	
Breath	02:23.664	
Breath	02:26.572	
Breath	02:28.659	
Breath	02:33.225	
Microphone disortion	02:36.539	
Breath	02:38.139	
Breath	02:42.441	

Breath	02:46.561	Breath preceded by brief tapping of lips or teeth
Breath	02:52.281	

art5-FILTRACION_190.ogg		
Breath	00:00.166	
Breath	00:03.549	
Background noise	00:06.052	
Breath	00:08.886	
Background noise	00:10.690	
Breath	00:13.260	
Breath	00:16.539	Breath preceded by brief tapping of lips or teeth
Breath	00:19.089	Breath preceded by brief tapping of lips or teeth
Background noise	00:21.469	
Breath	00:21.873	Breath preceded by brief tapping of lips or teeth
Breath	00:28.200	Breath preceded by brief tapping of lips or teeth
Microphone distortion	00:30.429	
Background noise	00:30.759	
Microphone distortion	00:32.379	
Breath	00:39.093	Breath preceded by brief tapping of lips or teeth
Breath	00:45.710	Breath preceded by brief tapping of lips or teeth
Microphone distortion	00:46.860	

Cosette López

art4-FILTRACION_147.ogg		
Breath	00:02.800	
Breath	00:04.347	
Background noise	00:11.303	
Background noise	00:11.455	Matching room resonance as speech
Background noise	00:17.500	
Background noise	00:21.937	
Background noise	00:35.839	Matching room resonance as speech
Background noise	00:38.720	Matching room resonance as speech