Expanded Real-World Audio Attacks for Robustness Testing

To ensure the SoundSafe watermarking model is resilient to real-world conditions, the attack simulator must incorporate a comprehensive set of audio distortions beyond the initial ten attacks. Here's an expanded list of attack types, categorized for clarity:

I. Common Signal Processing Attacks (Beyond Initial List)

- **Dynamic Range Compression (DRC):**
  - Reduces the dynamic range of audio, affecting the watermark's intensity.
  - Include different types of compression (limiting, expansion).
- **Clipping:**
  - Distortion caused by signal exceeding the maximum amplitude.
  - Simulate varying levels of clipping.
- **De-Essing:**
  - Specific dynamic range compression targeting sibilant sounds.
  - Test how it affects watermark encoded in high frequencies.
- **Phasing/Flanging:**
  - Creates a swirling effect with delayed signal versions.
  - Vary delay and intensity.
- **Chorus/Reverb:**
  - Smears the watermark by creating multiple sound copies.
  - Simulate different room environments.
- **Pitch Shifting:**
  - Alters the pitch of the audio.
  - Impact watermarks dependent on specific frequency ranges
- **Time Stretching (More Advanced):**
  - More complex time stretching algorithms than the existing one
- **Equalization (EQ):**
  - Alters the frequency balance of audio.
  - Simulate a variety of equalization curves.
- **Noise Gates:**
  - Removes low-amplitude signals
  - Test for impact on watermarks placed on low amplitude portions of audio

II. Format and Encoding Attacks

- **Different Lossy Codecs:**
  - Test with AAC, Ogg Vorbis, WMA, in addition to MP3.
  - These codecs each affect audio and watermarks differently.
- **Lossless to Lossy Conversion (And Vice Versa):**
  - Convert lossless formats (WAV, FLAC) to lossy and back to introduce compression artifacts.
- **Bit Depth Conversion:**
  - Decreasing or increasing the bit depth to test for the effect of quantization
- **Dithering**
  - Technique used to reduce quantization noise

III. Environmental and Real-World Attacks

- **Background Noise:**
  - Add realistic ambient noises (traffic, speech, music) to mask the watermark.
- **Acoustic Transmission/Recording:**
  - Simulate microphone differences and room reverberation.
- **Over-the-Air Transmission:**
  - Simulate wireless transmission distortions (packet loss, interference).
- **Acoustic Interference:**
  - Simulate overlapping sounds
- **Electro-Magnetic Interference (EMI):**
  - Simulate distortion caused by EMI

IV. Malicious Attacks

- **Watermark Removal/Evasion Techniques:**
  - Research and simulate attacks that bypass watermarks.
- **Copy and Paste:**
  - Copying segments of audio to replace watermark with another audio segment
- **Signal Inversion:**
  - Inverting the phase of an audio signal
- **Stretching and Cutting:**
- Removing segments of audio and reassembling.
- **Collusion Attacks:**
  - Simulate combining multiple watermarked copies to remove the watermark.

V. Combined Attacks

- **Sequential Application of Attacks:**
  - Apply multiple attacks in a sequence to simulate compounding distortions.
- **Randomized Attack Chains:**
  - Randomly generate sequences of different attack types and intensities.

Implementation Guidelines

1. **Prioritize:** Implement the most common and impactful attacks first.
2. **Incremental Implementation:** Gradually add new attacks, testing after each addition.
3. **Parameter Variation:** Use a wide range of intensity levels for each attack.
4. **Randomization:** Randomize attack order and intensity during training.
5. **Analysis:** Track the Bit Error Rate (BER) on the validation set after adding a new attack type to assess its effect.
6. **Real-World Testing:** Validate the simulator performance with real-world recordings and devices.

Integration into DevOps Process:

1. **Attack Implementation:** Add code modules for new attacks into the attack simulator.
2. **CI/CD Pipeline Update:** Include these attacks in the training pipeline.
3. **Monitoring:** Monitor training performance after each added attack.
4. **Report and Analyze:** Track how added attacks are impacting watermarking performance.