Masking in Perceptual Audio Coding and Audio Quality

Perceptual audio coding is widely used in modern audio compression formats such as MP3 (MPEG-1 Audio Layer III) and AAC (Advanced Audio Coding) to reduce file size while maintaining high audio quality. One of the most fundamental techniques in these codecs is masking, a psychoacoustic phenomenon that exploits the limitations of human hearing to remove inaudible or less perceptible sounds.
This project will explain, demonstrate, and analyze two different types of psychoacoustic masking in perceptual audio coding, including frequency masking and temporal masking. Well-designed and conductive experiments will analyze how psychoacoustic masking affects audio quality and how to use them in real-world examples.

Purpose of Experiments:

The goal of this project was to analyze and apply temporal and frequency masking to audio signals and measure their effects using objective and subjective methods. Specifically, we aimed to:

- Demonstrate how stronger sounds mask weaker ones in temporal masking and frequency masking.
- Use the PEAQ (Perceptual Evaluation of Audio Quality) standard to quantify the perceptual impact of these modifications.
- Show how masking principles are applied in real-world audio compression by selectively removing inaudible components.

Summary of Experiments:

Experiment 1: PEAQ Evaluation of Audio Compression Objective:

Goal:

Evaluate the effect of lossy compression, converting WAV file to MP3 at various bitrates, on perceived audio quality.

Method:

1. A high-quality reference WAV file was encoded into MP3 at bitrates **64, 96, 160, and 320 kbps**.
2. Each MP3 file was then decoded and compared against the original using **PEAQ (ITU-R BS.1387)** to obtain an **Objective Difference Grade (ODG)** score.

Results:

```
 WAVE file: BeeMoved_Ref_48kHz16bit.wav
   Number of frames  : 1914048 (39.88 s)
   Sampling frequency: 48000
   Number of channels: 2 (16-bit integer)
 WAVE file: BeeMoved_320kbps_48kHz16bit.wav
   Number of frames  : 1914048 (39.88 s)
   Sampling frequency: 48000
   Number of channels: 2 (16-bit integer)
PEAQ Data Boundaries: 756 (0.016 s) - 1907092 (39.731 s)
Bitrate: 320 kbps | PEAQ ODG = 0.04


 WAVE file: BeeMoved_Ref_48kHz16bit.wav
   Number of frames  : 1914048 (39.88 s)
   Sampling frequency: 48000
   Number of channels: 2 (16-bit integer)
 WAVE file: BeeMoved_160kbps_48kHz16bit.wav
   Number of frames  : 1914048 (39.88 s)
   Sampling frequency: 48000
   Number of channels: 2 (16-bit integer)
PEAQ Data Boundaries: 756 (0.016 s) - 1907092 (39.731 s)
Bitrate: 160 kbps | PEAQ ODG = -0.29


 WAVE file: BeeMoved_Ref_48kHz16bit.wav
   Number of frames  : 1914048 (39.88 s)
   Sampling frequency: 48000
   Number of channels: 2 (16-bit integer)
 WAVE file: BeeMoved_96kbps_48kHz16bit.wav
   Number of frames  : 1914048 (39.88 s)
   Sampling frequency: 48000
   Number of channels: 2 (16-bit integer)
PEAQ Data Boundaries: 756 (0.016 s) - 1907092 (39.731 s)
Bitrate:  96 kbps | PEAQ ODG = -1.95
```

```
 WAVE file: BeeMoved_Ref_48kHz16bit.wav
   Number of frames  : 1914048 (39.88 s)
   Sampling frequency: 48000
   Number of channels: 2 (16-bit integer)
 WAVE file: BeeMoved_64kbps_48kHz16bit.wav
   Number of frames  : 1914048 (39.88 s)
   Sampling frequency: 48000
   Number of channels: 2 (16-bit integer)
PEAQ Data Boundaries: 756 (0.016 s) - 1907092 (39.731 s)
Bitrate:  64 kbps | PEAQ ODG = -3.44
```

- **320 kbps yielded an ODG of 0.04**, indicating **almost no audible difference from the original**.
- **160 kbps and above showed minimal perceptual degradation**, demonstrating that higher bitrates preserve more details.
- **96 kbps showed moderate degradation**, with an ODG of **-1.95**, meaning some quality loss but still acceptable in certain applications.
- **Low bitrates (64 kbps)** showed significant degradation, with **PEAQ scores at -3.44**, indicating **noticeable artifacts** and compression loss.

These results confirm the effectiveness of **perceptual coding**—by discarding masked sounds, we achieve significant data reduction while maintaining acceptable quality.

Experiment 2: Frequency Masking Analysis

Goal:

- Generate and visualize how a loud sound at a given frequency masks quieter sounds nearby

Method:

- Generate a loud tone at **1 kHz** and a quieter tone at **1.5 kHz**
- The **1 kHz tone** was played alongside a **1.5 kHz tone** at varying levels (-12 dB, -18 dB, -24dB, -30 dB).
- Spectral magnitude plots and energy difference calculations were used to analyze masking effects.
- The **difference in total energy** was measured across different tone levels to quantify masking intensity.

Results:

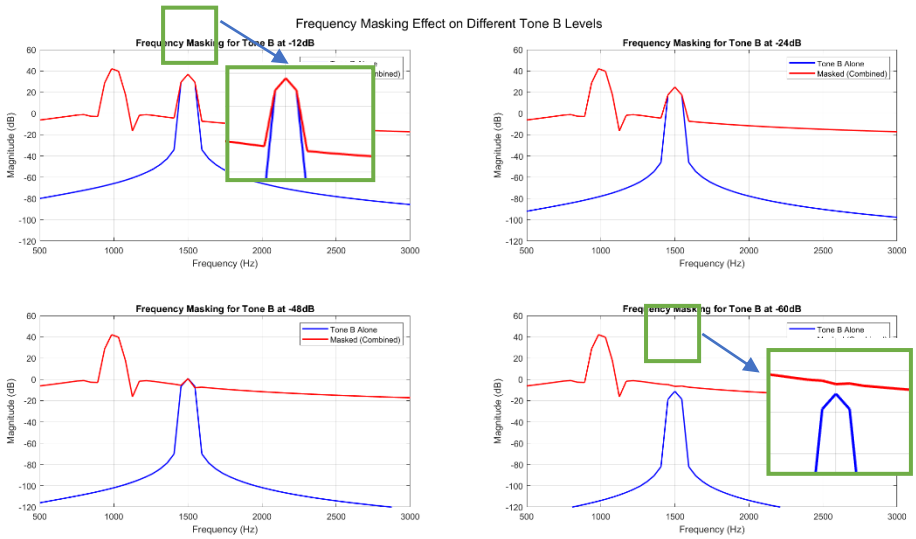

Figure 1(Frequency Masking Comparison)

1. **At higher amplitudes of Tone B (-12 dB, -24 dB)**:
   - Tone B contributes significantly to the combined spectrum.
   - There is a noticeable peak around **1.5 kHz** in the red (combined) curve.
   - The blue curve (Tone B alone) and the red curve (combined) both show distinct peaks.
2. **As Tone B decreases in amplitude (-48 dB, -60 dB)**:
   - The peak at **1.5 kHz** (Tone B) becomes **less pronounced**.
   - The **red curve (combined) barely** shows any deviation at **Tone B's location** at -60 dB.
   - This indicates that **Tone B is no longer contributing meaningfully** to the spectrum.

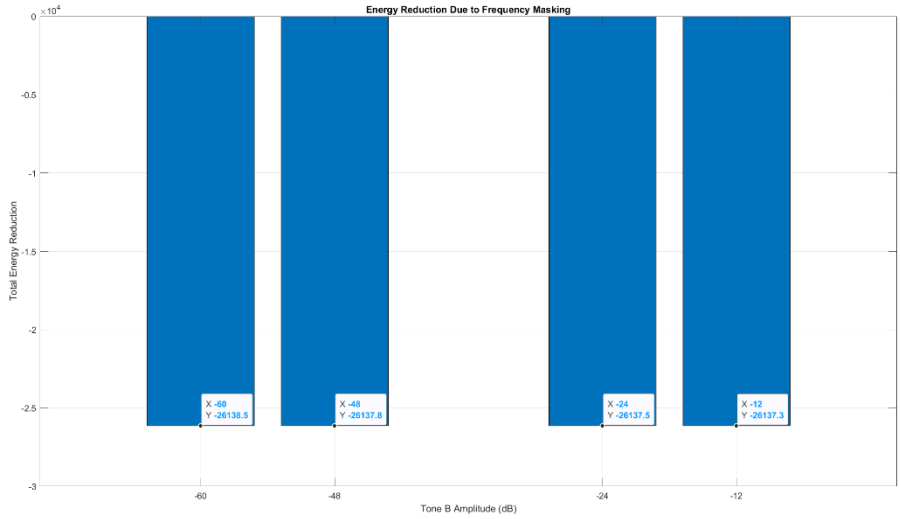

Figure 2(Energy Reduction)

3. The total energy reduction stays around the same value for all tested levels of Tone B. Even when Tone B is at **-60 dB**, where it should be fully masked, the

energy reduction is almost identical to when Tone B is at **-12 dB**.

4. When Tone B is sufficiently weak, it gets absorbed into the overall spectrum dominated by Tone A, meaning its contribution to total energy is negligible.
5. This supports the idea that **masked components in audio can be removed without significantly affecting the overall energy distribution**, which is critical for perceptual audio compression.

Experiment 3: Temporal Masking Analysis

Goal:

- Generate and visualize how a loud sound affects the perception of sounds that follow it.

Method:

1. A **strong 1 kHz tone** was played before a **weaker 2 kHz tone** with various delays (10ms, 50ms, 100ms, 300ms, 500ms).
2. Spectrograms and listening tests were conducted to observe when the **2 kHz tone** became audible.
3. A **difference spectrogram** was used to highlight the masked regions.

Results:

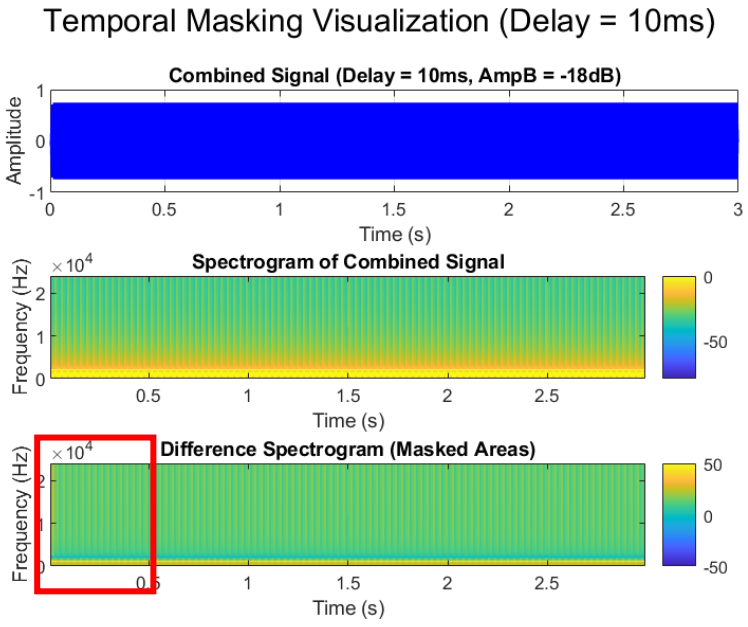

Figure 3(Temporal Masking 10ms Delay)

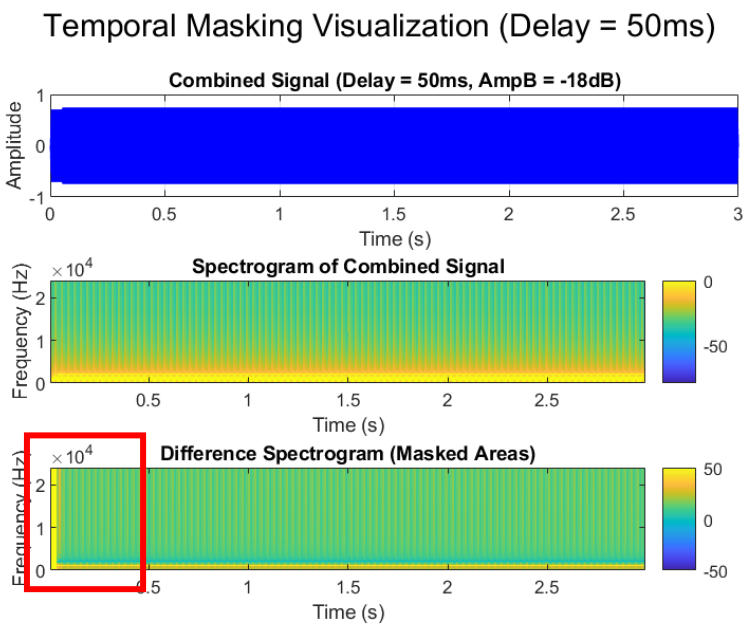

Figure 4(Temporal Masking 50ms Delay)

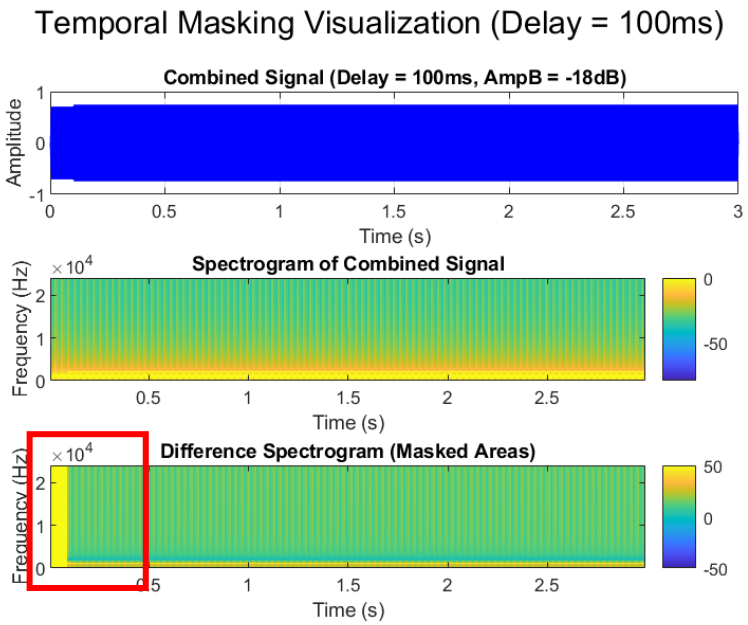

Figure 5(Temporal Masking 100ms Delay)

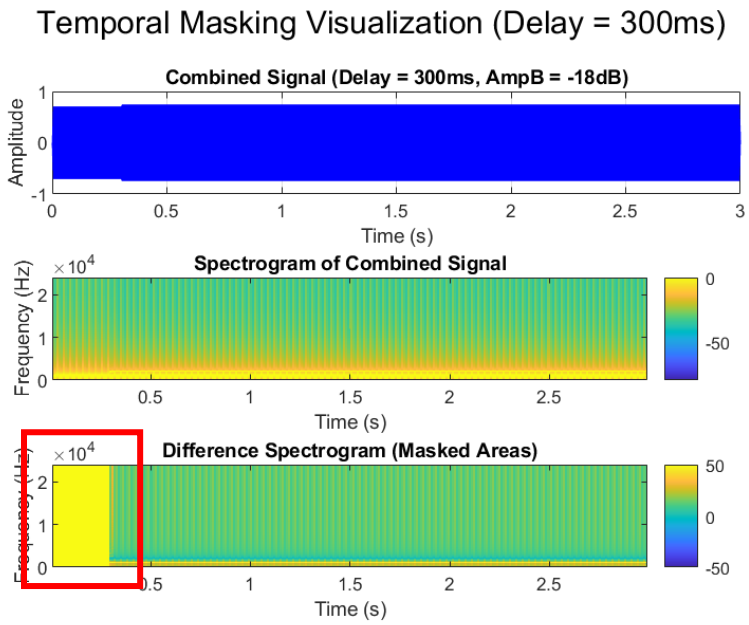

Figure 6(Temporal Masking 300ms Delay)

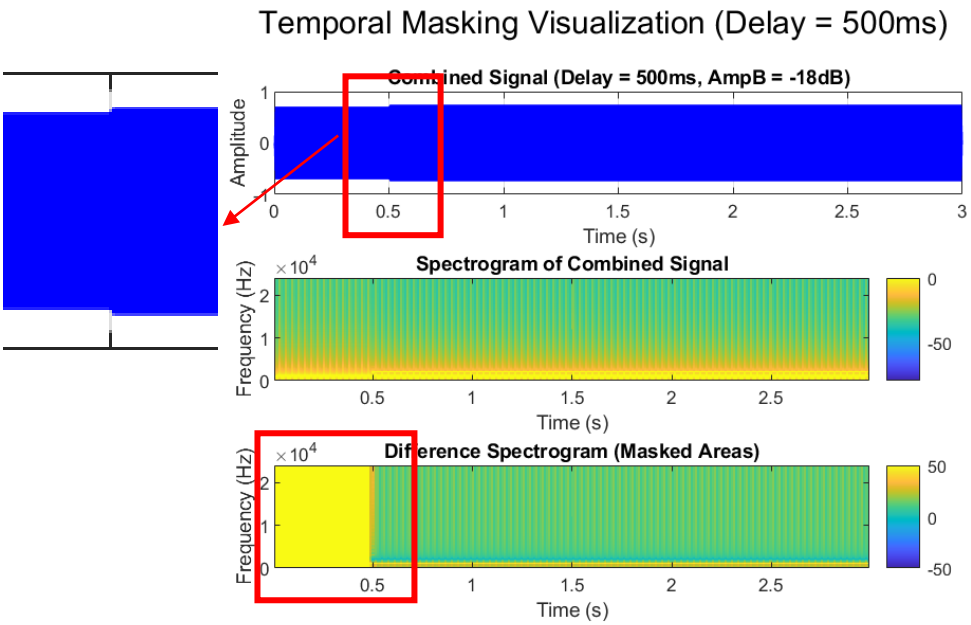

Figure 7(Temporal Masking 500ms Delay)

About Difference Spectrogram (Masked Areas):

It is used for visualizing **where and how much masking occurs**. It represents the **difference in spectral energy** between two spectrograms:

1. **Combined Signal (1 kHz tone + 2 kHz tone)**
2. **Reference Spectrogram (Expected Signal without Masking)**

The difference spectrogram is calculated as:

Difference = Spectrogram of Combined Signal – Spectrogram of 2 kHz tone Alone

This reveals which parts of the **2 kHz tone are being masked** by the **1 kHz tone** at different time delays.

- **For Short Delays (Figure 3, 4 & 5)**:
  - The **2 kHz tone is almost entirely absent** in the difference spectrogram.
  - The **masked regions appear as green areas**, meaning the energy of the 2 kHz tone has been **completely suppressed** by the masker.
  - This confirms that temporal masking is strongest **right after a loud sound**.
- **For Longer Delay (Figure 6 & 7)**:
  - The 2 kHz tone begins to **emerge in the spectrogram**.
  - The difference spectrogram shows a **partial reduction in masking**, suggesting that after a certain delay, the brain starts recognizing the weaker tone.
  - At **500ms delay**, 2 kHz tone is **clearly visible** in both the spectrogram and the difference spectrogram.
  - The masker has **minimal masking effect**, meaning the human ear can easily distinguish the 2 kHz tone. This confirms that temporal masking **weakens significantly over time**.

Experiment 4: Masking on a real-world music

Goal:

- Apply both **temporal and frequency masking** to a real music example and analyze its effect on perceived audio quality.

Method:

- Cut the music sample into snippet and apply **both** frequency and temporal masking.
- Using **PEAQ** to evaluate quality change.
- **Frequency Masking:**
  - Compute STFT and attenuate weak frequency components in each frame.
  - If a frequency component is 20 dB lower than the strongest component

in that frame, it is attenuated by 50%

- **Temporal Masking:**
  - Drum kicks are considered as maskers.
  - Used envelope extraction and peak detection to find strong drum transients.
  - Applied low-frequency removal (80–160 Hz) for 50–150ms after each drum transient

Results:

- Removing masked drum frequencies resulted in **minimal perceptual change, the overall shape of amplitude stays the same**, confirming that **masking can be used for compression without noticeable degradation**.
- **The amplitude vs. time graph (figure 8) showed that the overall energy of the damaged signal remained similar**, confirming that perceptually irrelevant sounds were removed while preserving core elements.
- PEAQ scores (figure 9) showed a slight degradation but remained within acceptable quality levels.

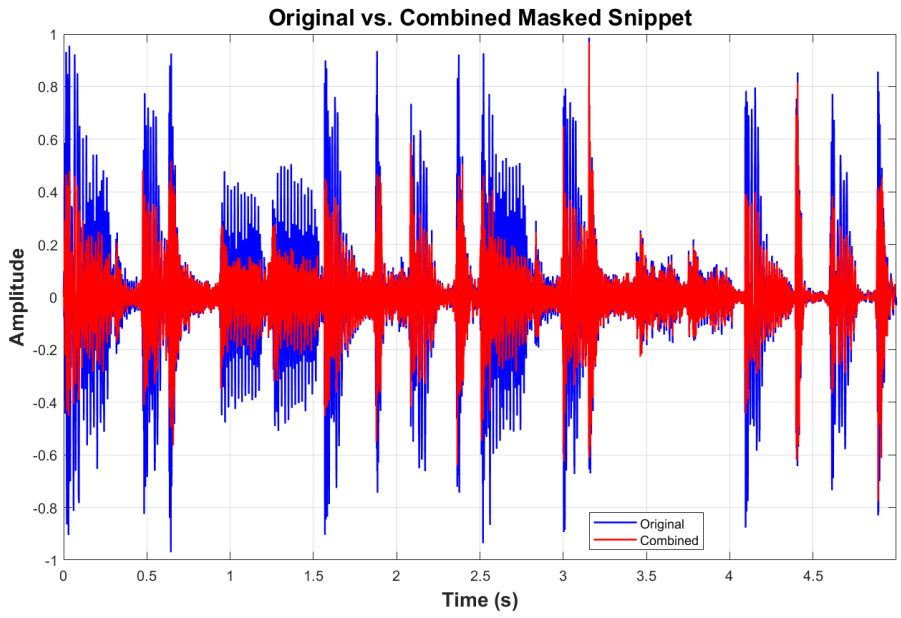

Figure 8(Amplitude plots)

```
PEAQ Objective Difference Grade (ODG): -1.44
```

Figure 9(PEAQ score)

Standard Used:

- **PEAQ (Perceptual Evaluation of Audio Quality):**
    - An international standard (ITU-R BS.1387) designed to objectively assess the perceived quality of audio signals.
    - Primarily for evaluating the degradation caused by lossy compression techniques such as MP3 and AAC.
    - Instead of relying on subjective listening tests, PEAQ **models human auditory perception** and computes an **Objective Difference Grade (ODG)** that correlates with how listeners perceive audio quality.
    - For all the other information about PEAQ, please see reference.
- **MPEG Layer 3 (MP3) and AAC**: Use both **frequency and temporal masking** for data reduction.
- **Psychoacoustic Models**: Used in perceptual coding to remove masked components.

Sources:

Technical Standards:

1. Thiede, T., Treurniet, W. C., Bitto, R., Schmidmer, C., Sporer, T., Beerends, J. G., Colomes, C., Keyhl, M., Stoll, G., Brandenburg, K., & Feiten, B. (2000). *PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality*. *Journal of the Audio Engineering Society*, 48(1/2), 3-29.
2. Câmpeanu, D., & Câmpeanu, A. (n.d.). *PEAQ – An Objective Method to Assess the Perceptual Quality of Audio Compressed Files*. "Politehnica" University of Timișoara, Department of Communications.
3. Holfelt, J. (n.d.). *PEAQ* [GitHub Repository]. Retrieved from https://github.com/jholfelt/PEAQ

Academic Papers & Books:

1. Bosi, M., & Goldberg, R. E. (2002). *Introduction to digital audio coding and standards*. Kluwer Academic Publishers.

2. Sayood, K. (2017). *Introduction to Data Compression* (Fifth edition.). Elsevier Science & Technology.
3. Painter, T., & Spanias, A. (2000). Perceptual coding of digital audio. *Proceedings of the IEEE*, *88*(4), 451–515. https://doi.org/10.1109/5.842996
4. Schuller, G. (2020). *Filter Banks and Audio Coding: Compressing Audio Signals Using Python* (1st Edition 2020). Springer Nature. https://doi.org/10.1007/978-3-030-51249-1

Other:

1. Blue Monday FM. (2014). *Bee Moved* [Song]. On *Bee Moved* [Album]. Retrieved from https://bluemondayfm.bandcamp.com/releases
2. CrashBulb. (n.d.). *Audio Sample 779199* [Sound file]. Retrieved from https://freesound.org/people/CrashBulb/sounds/779199/
3. iZotope. (n.d.). *What is Frequency Masking?* Retrieved from https://www.izotope.com/en/learn/what-is-frequency-masking.html