Understanding WAV File Structure and Data Chunk Analysis

1. Introduction
WAV files store audio data in a structured format using chunks. The audio data is typically found in the 'data'chunk, but some WAV files may include additional metadata chunks such as 'LIST', which can affect how audio data is processed. In this document, we analyze two WAV files and identify why one works correctly while the other becomes corrupted after bit manipulation.

2. Comparison of Two WAV Files

Property	first.wav (Works Fine)	second.wav (Corrupts After Modification)
Chunk ID	RIFF	RIFF
File Size	222480 bytes	-1 bytes (Incorrect)
Format	WAVE	WAVE
Subchunk1 ID	fmt	fmt
Subchunk1 Size	16 bytes	16 bytes
Audio Format	1 (PCM)	1 (PCM)
Number of Channels	2	2
Sample Rate	44100 Hz	48000 Hz
Byte Rate	176400 bytes/sec	192000 bytes/sec
Block Align	4 bytes	4 bytes
Bits per Sample	16	16
Subchunk2 ID	data ✓	LIST ✗ (Extra Metadata)
Data Size	222444 bytes	26 bytes (Not Actual Data)
File Size in Disk	~217 KB	~46 MB (Much Larger)

3. Why Does second.wav Corrupt?
Issues in second.wav:
- File Size is -1 bytes (likely incorrect header value '0xFFFFFFFF').
- Subchunk2 ID is 'LIST', meaning audio data is not at the expected position.
- Some programs use LIST for metadata before data, which affects processing.

4. How to Locate the 'data' Chunk?
- Skip the first 12 bytes ("RIFF", File Size, "WAVE").
- Scan for chunks, reading 8 bytes at a time.
- If chunk ID is 'data', stop - this is where the audio starts.
- If chunk is not 'data', skip chunkSize bytes and continue.

5. Key Takeaways
- WAV files use chunks like 'fmt ', 'LIST', and 'data'.
- Always locate 'data' before modifying samples.
- PCM WAV files are easiest to work with.
- Check the file size field (if -1, header may be incorrect).