Unit 4: Timbre, Music Analysis and Processing

Course Notes

MA375TGR - Mathematics of Music
Semester VII - Institutional Elective

Faculty: Dr. Venugopal K.

Contents

1 Introduction to Timbre **4**
- 1.1 What is Timbre? 4
- 1.2 Why Timbre Matters 4
- 1.3 The Challenge of Defining Timbre 4
- 1.4 Modern Understanding of Timbre 5
- 1.5 Mathematical Representation 5
- 1.6 Questions and Answers 5

2 Harmonics and Spectral Content **6**
- 2.1 Partials, Fundamentals, and Overtones 6
- 2.2 Harmonic versus Inharmonic Spectra 6
- 2.3 Spectral Envelope 7
- 2.4 Brightness and Spectral Centroid 7
- 2.5 Examples of Spectral Characteristics 7
- 2.6 Questions and Answers 8

3 Fourier Analysis and Sound Decomposition **9**
- 3.1 Historical Context 9
- 3.2 The Fundamental Question 9
- 3.3 Fourier Series 9
- 3.4 Calculating Fourier Coefficients 9
- 3.5 Complex Form of Fourier Series 9
- 3.6 From Series to Transform: The Fourier Transform 10
- 3.7 Physical Interpretation 10
- 3.8 The Role of Phase in Perception 10
- 3.9 Example: Square Wave 11
- 3.10 Questions and Answers 11

4 Sub-harmonics and Combination Tones 13
4.1 Sub-harmonics: Definition and Occurrence . . . 13
4.2 Where Sub-harmonics Occur . . . 13
4.3 Difference from Virtual Pitch . . . 14
4.4 Combination Tones . . . 14
4.5 Historical Discovery . . . 14
4.6 Mathematical Explanation . . . 14
4.7 Mathematical Derivation of Combination Tones . . . 15
4.7.1 Setup and Assumptions . . . 15
4.7.2 Step 1: Apply the Nonlinear Transformation . . . 15
4.7.3 Step 2: Expand the Square . . . 15
4.7.4 Step 3: Apply Trigonometric Identities . . . 15
4.7.5 Step 4: Complete Expansion . . . 16
4.7.6 Step 5: Identify Frequency Components . . . 16
4.7.7 Physical Interpretation . . . 16
4.7.8 Extension to Cubic Nonlinearity . . . 18
4.8 Practical Significance . . . 18
4.9 Example Calculation . . . 18
4.10 Questions and Answers . . . 19

5 Limitations of Fourier Analysis: Noise and Beyond 19
5.1 What Fourier Analysis Tells Us . . . 19
5.2 The Time-Frequency Trade-off . . . 20
5.3 Types of Noise . . . 20
5.4 Why Fourier Analysis Fails for Noise . . . 21
5.5 Musical Significance of Noise . . . 21
5.6 Generating Different Colors of Noise . . . 21
5.7 Alternative Analysis Methods . . . 22
5.8 Questions and Answers . . . 22

6 Introduction to Sound Synthesis 24
6.1 What is Sound Synthesis? . . . 24
6.2 Why Synthesize Sound? . . . 24
6.3 Basic Synthesis Components . . . 25
6.4 The Importance of Envelopes . . . 25
6.5 Envelope Applications . . . 26
6.6 Low Frequency Oscillators (LFOs) . . . 26
6.7 Basic Waveforms and Their Spectra . . . 26
6.8 Questions and Answers . . . 27

7 Synthesis Methods 28
7.1 Additive Synthesis . . . 28
7.2 Subtractive Synthesis . . . 29
7.3 Frequency Modulation (FM) Synthesis . . . 30
7.4 Granular Synthesis . . . 31
7.5 Physical Modeling Synthesis . . . 32
7.6 Questions and Answers . . . 33

8 Spectral Envelopes and Timbre **36**
8.1 What is a Spectral Envelope? 36
8.2 Importance of Spectral Envelope 36
8.3 Source-Filter Model 37
8.4 Formants 37
8.5 Static vs. Dynamic Spectra 38
8.6 Amplitude Envelope vs. Spectral Envelope 38
8.7 Practical Applications 38
8.8 Example: Piano vs. Harpsichord 39
8.9 Questions and Answers 39

9 Summary and Applications **41**
9.1 Key Concepts Recap 41
9.2 Practical Applications 42
9.3 Connections to Other Units 42

A Quick Reference Tables **43**
A.1 Common Waveform Spectra 43
A.2 Synthesis Method Comparison 43
A.3 Noise Types 43
A.4 Important Formulas 43

B Sample Problems and Solutions **44**

C Conclusion **45**

1 Introduction to Timbre

1.1 What is Timbre?

Timbre (pronounced *TAM-ber*, from French) is perhaps the most complex and least understood dimension of musical sound. While pitch, loudness, and duration can be quantified relatively straightforwardly, timbre encompasses everything about a sound that is not its pitch, loudness, or duration.

Definition 1.1 (Timbre)**.** Timbre is the quality or ”color” of a sound that distinguishes different types of sound sources, even when they have the same pitch and loudness. It is determined primarily by the spectral content of the sound and how that spectrum evolves over time.

Remark 1.1. The American Standards Association (1960) defines timbre as ”that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar.”

1.2 Why Timbre Matters

Timbre is what allows us to distinguish:

- A violin from a flute, even when playing the same note
- Different vowel sounds in speech
- The same note played loudly versus softly on certain instruments
- The character or emotional quality of a sound (bright, dark, harsh, mellow)

Unlike pitch and rhythm, which have been extensively studied and formalized in music theory for centuries, timbre remained somewhat mysterious until modern acoustics and signal processing provided tools to analyze and understand it mathematically.

1.3 The Challenge of Defining Timbre

The difficulty in defining timbre stems from several factors:

1. **Multidimensional Complexity**: Timbre is not a single parameter but involves multiple perceptual dimensions including brightness, roughness, attack characteristics, and temporal evolution.
2. **Residual Definition**: Traditionally, timbre has been defined negatively—as what remains after accounting for pitch, loudness, and duration. This provides little positive understanding.
3. **Context Dependence**: Our perception of timbre can be influenced by musical context, cultural background, and individual experience.
4. **Temporal Dynamics**: Unlike pitch (which can be relatively static), timbre is inherently dynamic, changing throughout the duration of a note.

1.4 Modern Understanding of Timbre

Contemporary research has identified two primary perceptual structures that determine timbre:

1. **Spectral Energy Distribution**: The relative strength of different frequency components (partials) in the sound
2. **Temporal Evolution**: How the spectral energy distribution changes over time

This leads to a more positive definition: *Timbre consists primarily of the static and dynamic properties of a sound's spectrum.*

1.5 Mathematical Representation

A complete mathematical description of timbre requires capturing both frequency content and temporal evolution. If $s(t)$ represents a sound signal, its timbre characteristics can be analyzed through:

$$\begin{aligned} S(f,t) &= \text{Time-varying spectrum} && (1) \\ E(f) &= \text{Average spectral envelope} && (2) \\ A(t) &= \text{Amplitude envelope} && (3) \\ \phi(f,t) &= \text{Phase relationships} && (4) \end{aligned}$$

1.6 Questions and Answers

Question 1 (Short Answer): Define timbre and explain why it is difficult to define precisely.

Answer: Timbre is the quality or "color" of sound that distinguishes different sound sources having the same pitch and loudness. It is difficult to define precisely because: (1) it is multidimensional, involving brightness, roughness, attack, and temporal characteristics; (2) it has traditionally been defined residually as "everything except pitch, loudness, and duration"; (3) it depends on both spectral content and temporal evolution; and (4) perception varies with context and individual experience.

Question 2 (Short Answer): What are the two primary perceptual structures that determine timbre?

Answer: The two primary perceptual structures are:

1. **Spectral Energy Distribution**: The pattern of frequencies and their relative amplitudes present in the sound
2. **Temporal Evolution**: How the spectrum changes over time during the course of a note

Question 3 (Medium Answer): Explain how we can distinguish between a violin and a flute playing the same note at the same loudness using the concept of timbre.

Answer: When a violin and flute play the same note (e.g., A440) at the same loudness, they differ in timbre due to:

Spectral Differences:

- The violin has a richer harmonic spectrum with stronger high-frequency partials
- The flute has fewer strong partials, dominated by the fundamental and a few lower harmonics
- The relative amplitudes of harmonics differ significantly between the two

Temporal Differences:

- The violin has a characteristic attack with bow-string interaction, including scraping sounds
- The flute has a breathy attack from the air jet forming
- The decay characteristics differ—violin notes can be sustained with varying bow pressure, while flute notes decay more gradually
- The temporal envelope of each harmonic evolves differently

Our auditory system analyzes these spectral and temporal patterns and uses prior experience to identify the instruments, even though pitch and loudness are identical.

2 Harmonics and Spectral Content

2.1 Partials, Fundamentals, and Overtones

Understanding timbre requires understanding how complex sounds can be decomposed into simpler components.

Definition 2.1 (Partial)**.** A partial is a sinusoidal component of a complex tone. Partials are also called components and carry a partial characterization of the complete sound.

Definition 2.2 (Fundamental)**.** The fundamental is the lowest-frequency partial in a tone. It generally determines the perceived pitch of the sound.

Definition 2.3 (Overtone)**.** Overtones are all partials in a tone that are higher in frequency than the fundamental. The n-th overtone is the $(n+1)$-th partial.

Definition 2.4 (Harmonic)**.** A harmonic is a partial whose frequency is an integer multiple of the fundamental frequency. If the fundamental has frequency f_0, the n-th harmonic has frequency nf_0.

2.2 Harmonic versus Inharmonic Spectra

Musical sounds can be classified based on their spectral structure:

Harmonic Sounds: Most pitched musical instruments (strings, winds, brass, voice) produce harmonic spectra where:

$$f_n = n \cdot f_0, \quad n = 1, 2, 3, 4, \ldots \tag{5}$$

Inharmonic Sounds: Percussion instruments (drums, gongs, bells) generally have inharmonic spectra where:

$$f_n \neq n \cdot f_0 \tag{6}$$

The frequencies do not form simple integer ratios, leading to a less clear pitch sensation.

2.3 Spectral Envelope

Definition 2.5 (Spectral Envelope)**.** The spectral envelope is a smooth curve that connects the peaks of the individual partials in a frequency spectrum, representing the overall shape of the spectrum's energy distribution.

The spectral envelope is crucial for determining timbre. Two sounds with similar spectral envelopes will sound similar in timbre, even if their fundamental frequencies differ.

Mathematical representation:

$$E(f) = \text{envelope}\{|S(f)|\} \tag{7}$$

where $S(f)$ is the Fourier transform of the signal and the envelope function smooths over individual peaks.

2.4 Brightness and Spectral Centroid

One quantifiable aspect of timbre is brightness, related to the distribution of spectral energy:

Definition 2.6 (Spectral Centroid)**.** The spectral centroid is the "center of mass" of the spectrum, calculated as:

$$C = \frac{\sum_{n=1}^{N} f_n \cdot A_n}{\sum_{n=1}^{N} A_n} \tag{8}$$

where f_n is the frequency of the n-th partial and A_n is its amplitude.

- Higher spectral centroid → brighter, sharper timbre
- Lower spectral centroid → darker, mellower timbre

2.5 Examples of Spectral Characteristics

Example 2.1 (Flute Spectrum)**.** A flute playing A440 Hz typically has:

- Strong fundamental (440 Hz)
- Weak second harmonic (880 Hz)
- Moderate third harmonic (1320 Hz)
- Very weak higher harmonics
- Low spectral centroid → mellow, pure timbre

Example 2.2 (Trumpet Spectrum)**.** A trumpet playing A440 Hz typically has:

- Moderate fundamental (440 Hz)
- Strong harmonics up to 6-8th partial
- Significant energy in high harmonics (2000-4000 Hz)
- High spectral centroid → bright, brilliant timbre

2.6 Questions and Answers

Question 4 (Short Answer): What is the difference between a partial and a harmonic?

Answer: A **partial** is any sinusoidal component of a complex sound, regardless of its frequency relationship to other components. A **harmonic** is a special type of partial whose frequency is an integer multiple of the fundamental frequency. All harmonics are partials, but not all partials are harmonics (e.g., in bells or drums, partials are inharmonic).

Question 5 (Short Answer): Define the spectral centroid and explain its perceptual significance.

Answer: The spectral centroid is the weighted average frequency of a spectrum, calculated as:

$$C = \frac{\sum f_n A_n}{\sum A_n}$$

It represents the "center of gravity" of the spectrum. Perceptually, it correlates with brightness: higher centroid values correspond to brighter, sharper timbres, while lower values correspond to darker, mellower timbres.

Question 6 (Medium Answer): Explain the difference between harmonic and inharmonic spectra. Give examples of instruments for each category and explain why this distinction matters.

Answer:

Harmonic Spectra: In harmonic spectra, all partial frequencies are integer multiples of the fundamental: $f_n = n \cdot f_0$.

Examples: violin, flute, clarinet, trumpet, human voice

These instruments produce a clear sense of pitch because the harmonics reinforce a common fundamental.

Inharmonic Spectra: In inharmonic spectra, partial frequencies do not form simple integer ratios with the fundamental.

Examples: drums, cymbals, bells, gongs, marimba

These instruments often have ambiguous or multiple perceived pitches.

Why It Matters:

- Harmonic instruments can play melodies and harmonies clearly
- Inharmonic instruments are better for rhythm and texture
- The mathematics of tuning and scale construction applies primarily to harmonic instruments
- Different synthesis and analysis techniques are needed for each type

Mathematical Basis: For a string or air column, the physics naturally produces harmonics because the standing wave patterns must fit integer numbers of half-wavelengths. For 2D vibrating systems (drums) or complex 3D structures (bells), the physics produces inharmonic partials because the boundary conditions and mode shapes are more complex.

3 Fourier Analysis and Sound Decomposition

3.1 Historical Context

The mathematical decomposition of complex sounds into simpler components has its roots in the work of several mathematicians and scientists of the 17th and 18th centuries, including Marin Mersenne, Daniel Bernoulli, Jean-le-Rond d'Alembert, Leonhard Euler, and the Bach family. However, the systematic theory was developed by Jean Baptiste Joseph Fourier (1768-1850).

3.2 The Fundamental Question

How can a string vibrate with multiple frequencies simultaneously? This question occupied mathematicians and musicians for centuries. The answer lies in the principle of superposition and Fourier's theory of harmonic analysis.

3.3 Fourier Series

Definition 3.1 (Fourier Series)**.** Any periodic function $f(t)$ with period T can be decomposed into a (possibly infinite) sum of sines and cosines:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{2\pi nt}{T}\right) + b_n \sin\left(\frac{2\pi nt}{T}\right) \right] \tag{9}$$

where a_0, a_n, b_n are the Fourier coefficients.

The fundamental frequency is $f_0 = 1/T$, and the series contains harmonics at frequencies $f_n = n \cdot f_0$.

3.4 Calculating Fourier Coefficients

The coefficients are determined by integration over one period:

$$a_0 = \frac{2}{T} \int_0^T f(t)\, dt \tag{10}$$

$$a_n = \frac{2}{T} \int_0^T f(t) \cos\left(\frac{2\pi nt}{T}\right) dt \tag{11}$$

$$b_n = \frac{2}{T} \int_0^T f(t) \sin\left(\frac{2\pi nt}{T}\right) dt \tag{12}$$

The coefficient $a_0/2$ represents the DC component (average value), while a_n and b_n represent the amplitudes of the n-th harmonic's cosine and sine components, respectively.

3.5 Complex Form of Fourier Series

Using Euler's formula, we can write the Fourier series in complex form:

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi nt/T} \tag{13}$$

where the complex coefficients are:

$$c_n = \frac{1}{T} \int_0^T f(t) e^{-i2\pi nt/T} dt \tag{14}$$

The magnitude $|c_n|$ represents the amplitude and $\arg(c_n)$ represents the phase of the n-th harmonic.

3.6 From Series to Transform: The Fourier Transform

For non-periodic signals, we extend from discrete harmonics to a continuous frequency spectrum:

Definition 3.2 (Fourier Transform)**.** The Fourier transform of a signal $s(t)$ is:

$$S(f) = \int_{-\infty}^{\infty} s(t) e^{-i2\pi ft} dt \tag{15}$$

The inverse transform recovers the signal:

$$s(t) = \int_{-\infty}^{\infty} S(f) e^{i2\pi ft} df \tag{16}$$

Theorem 3.1 (Fourier Inversion Theorem)**.** If $s(t)$ is a reasonably well-behaved function, then the Fourier transform and its inverse are true inverses, meaning applying both operations returns the original function.

3.7 Physical Interpretation

The Fourier transform tells us:

- **Magnitude Spectrum**: $|S(f)|$ shows the amplitude of each frequency component
- **Phase Spectrum**: $\arg[S(f)]$ shows the phase relationships between components
- **Power Spectrum**: $|S(f)|^2$ shows the energy distribution across frequencies

3.8 The Role of Phase in Perception

An important finding in psychoacoustics:

Remark 3.1 (Phase and Timbre Perception)**.** For steady-state tones, phase relationships between harmonics are generally not perceivable. However, phase is crucial for:

- Attack transients (initial portion of sounds)
- Temporal localization of sound events
- Spatial perception

This is why spectral analysis (amplitude vs. frequency) is often more useful for understanding timbre than waveform analysis (amplitude vs. time).

3.9 Example: Square Wave

Example 3.1 (Fourier Series of Square Wave)**.** A square wave with period T and amplitude ± 1 has the Fourier series:

$$f(t) = \frac{4}{\pi} \sum_{n=1,3,5,\ldots}^{\infty} \frac{1}{n} \sin\left(\frac{2\pi n t}{T}\right) \tag{17}$$

Key observations:

- Contains only odd harmonics (1st, 3rd, 5th, ...)
- Amplitudes decrease as $1/n$
- No even harmonics (2nd, 4th, 6th, ...) present
- This spectrum creates the characteristic "hollow" sound of square waves

3.10 Questions and Answers

Question 7 (Short Answer): State the Fourier theorem for periodic functions in words.

Answer: The Fourier theorem states that any periodic function can be decomposed into a sum (possibly infinite) of sine and cosine waves whose frequencies are integer multiples of the fundamental frequency. These components are called harmonics, and their amplitudes and phases completely determine the original function.

Question 8 (Medium Answer): Calculate the Fourier coefficients for a sawtooth wave defined as $f(t) = t$ for $t \in [0, 2\pi)$ with period 2π.

Answer: For $f(t) = t$ on $[0, 2\pi)$ with period $T = 2\pi$:

Step 1: Calculate a_0

$$a_0 = \frac{2}{2\pi} \int_0^{2\pi} t\, dt = \frac{1}{\pi} \left[\frac{t^2}{2}\right]_0^{2\pi} = \frac{1}{\pi} \cdot 2\pi^2 = 2\pi$$

Step 2: Calculate a_n

$$a_n = \frac{1}{\pi} \int_0^{2\pi} t \cos(nt)\, dt$$

Using integration by parts twice:

$$a_n = \frac{1}{\pi} \left[\frac{t \sin(nt)}{n} + \frac{\cos(nt)}{n^2}\right]_0^{2\pi} = 0$$

(Since $\sin(2\pi n) = 0$ and $\cos(2\pi n) = \cos(0) = 1$)

Step 3: Calculate b_n

$$b_n = \frac{1}{\pi} \int_0^{2\pi} t \sin(nt)\, dt$$

Using integration by parts:

$$b_n = \frac{1}{\pi} \left[-\frac{t \cos(nt)}{n} + \frac{\sin(nt)}{n^2}\right]_0^{2\pi} = -\frac{2\pi}{n\pi} = -\frac{2}{n}$$

Result:

$$f(t) = \frac{2\pi}{2} - 2\sum_{n=1}^{\infty} \frac{\sin(nt)}{n} = \pi - 2\sum_{n=1}^{\infty} \frac{\sin(nt)}{n}$$

The sawtooth contains all harmonics with amplitudes decreasing as $1/n$, giving it a bright, buzzy timbre.

Question 9 (Long Answer): Explain the physical basis for why sine waves are considered "pure" sounds and how the human ear acts as a Fourier analyzer.

Answer:

Mathematical Basis:

Sine waves are pure sounds because they are the fundamental solutions to the differential equation of simple harmonic motion:

$$\frac{d^2x}{dt^2} + \omega^2 x = 0$$

Any object subject to a restoring force proportional to displacement ($F = -kx$) will vibrate sinusoidally. The solution is:

$$x(t) = A\sin(\omega t + \phi)$$

where $\omega = \sqrt{k/m}$ depends on the stiffness k and mass m.

How the Ear Works as a Fourier Analyzer:

The human ear performs real-time Fourier analysis through the structure of the cochlea:

1. Basilar Membrane:

- The basilar membrane inside the cochlea is approximately 35mm long
- It is stiff and narrow at the base (near the oval window)
- It is flexible and wide at the apex
- Different locations along the membrane resonate at different frequencies

2. Frequency Mapping:

- High frequencies (20,000 Hz) cause maximum vibration at the base
- Low frequencies (20 Hz) cause maximum vibration at the apex
- The mapping is approximately logarithmic

3. Simple Harmonic Oscillators: Each point on the basilar membrane acts as a damped simple harmonic oscillator with natural frequency ω_0 depending on its position. When sound enters the ear, each point responds according to:

$$\frac{d^2x}{dt^2} + 2\gamma\frac{dx}{dt} + \omega_0^2 x = F(t)$$

The steady-state response to a sinusoidal input at frequency ω is maximum when $\omega \approx \omega_0$ (resonance).

4. Frequency Decomposition: When a complex sound enters the ear:

- Each frequency component causes maximum vibration at a specific location
- Hair cells at that location fire nerve impulses
- The brain interprets the spatial pattern of firing as different frequencies
- This is essentially a biological Fourier transform!

Mathematical Parallel: The basilar membrane's response is analogous to a bank of bandpass filters:

$$H(\omega) = \frac{1}{-\omega^2 + 2i\gamma\omega + \omega_0^2}$$

A complex sound $s(t) = \sum A_n \sin(\omega_n t + \phi_n)$ causes each filter centered at $\omega_0 \approx \omega_n$ to respond with amplitude proportional to A_n.

Conclusion: The ear is structurally designed to decompose sounds into frequency components, which is exactly what Fourier analysis does mathematically. This explains why sine waves are perceptually "pure"—they excite only a single location on the basilar membrane, while complex sounds excite multiple locations simultaneously.

4 Sub-harmonics and Combination Tones

4.1 Sub-harmonics: Definition and Occurrence

Definition 4.1 (Sub-harmonic)**.** A sub-harmonic is a frequency component that is a fractional division of the fundamental frequency:

$$f_{\text{sub}} = \frac{f_0}{n}, \quad n = 2, 3, 4, \ldots \tag{18}$$

Unlike harmonics (which are integer multiples *above* the fundamental), sub-harmonics are integer divisions *below* the fundamental.

4.2 Where Sub-harmonics Occur

Sub-harmonics are less common than harmonics but appear in several contexts:

1. **Nonlinear Systems**:
   - Overdriven electronic circuits
   - Certain organ pipes at very high air pressure
   - Vocal fry in speech (also called pulse register)
2. **Period-Doubling Bifurcations**: When a system parameter is varied, the vibration period can suddenly double, creating a sub-harmonic at $f_0/2$.
3. **Parametric Oscillations**: Systems where parameters vary periodically can produce sub-harmonics.
4. **Nonlinear Resonance**: Strong driving can cause a resonator to respond at fractional frequencies.

4.3 Difference from Virtual Pitch

Important Distinction:

- **Sub-harmonics**: Actual physical frequency components present in the sound wave at frequencies below the fundamental
- **Virtual/Missing Fundamental**: A perceived pitch corresponding to a frequency not physically present, inferred by the brain from the harmonic series

4.4 Combination Tones

When two or more tones are played simultaneously, the nonlinearity of the ear (and sometimes the instrument) can create additional perceived frequencies.

Definition 4.2 (Combination Tones)**.** Combination tones are frequencies generated by nonlinear interactions between two or more primary tones. For primaries f_1 and f_2 (with $f_2 > f_1$), the main combination tones are:

$$\text{Difference tone:} \quad f_d = f_2 - f_1 \tag{19}$$

$$\text{Summation tone:} \quad f_s = f_1 + f_2 \tag{20}$$

$$\text{Cubic difference tone:} \quad f_c = 2f_1 - f_2 \tag{21}$$

4.5 Historical Discovery

Combination tones were discovered by the violinist Giuseppe Tartini (1692-1770), who noticed that when two notes were played simultaneously on a violin, a third, lower pitch could be heard. These are sometimes called "Tartini tones."

4.6 Mathematical Explanation

Consider a nonlinear system with input-output relation:

$$y = ax + bx^2 + cx^3 + \dots \tag{22}$$

If the input is $x = A\sin(2\pi f_1 t) + B\sin(2\pi f_2 t)$, the output will contain:
From the x^2 term:

- DC term
- Harmonics $2f_1, 2f_2$
- **Sum and difference:** $f_1 + f_2, |f_1 - f_2|$

From the x^3 term:

- Harmonics $3f_1, 3f_2$
- **Cubic difference:** $2f_1 - f_2, 2f_2 - f_1$
- Other combinations: $2f_1 + f_2, f_1 + 2f_2$

4.7 Mathematical Derivation of Combination Tones

We now provide a complete step-by-step derivation showing how the quadratic nonlinearity generates sum and difference tones. This derivation is fundamental to understanding the mathematical origin of combination tones.

4.7.1 Setup and Assumptions

Consider a nonlinear distortion system with the transfer characteristic:

$$y = ax + bx^2 \tag{23}$$

We focus on the quadratic term bx^2 as it is the simplest nonlinearity that produces intermodulation products.

Input Signal: Two pure tones at frequencies f_1 and f_2:

$$x(t) = A_1 \cos(2\pi f_1 t) + A_2 \cos(2\pi f_2 t) \tag{24}$$

where A_1, A_2 are the amplitudes of the two input tones.

4.7.2 Step 1: Apply the Nonlinear Transformation

The output is:

$$y = ax + bx^2 \tag{25}$$

The linear term ax simply reproduces the input frequencies (scaled by a), so we focus on the nonlinear term:

$$y_{\text{nonlinear}} = b[A_1 \cos(2\pi f_1 t) + A_2 \cos(2\pi f_2 t)]^2 \tag{26}$$

4.7.3 Step 2: Expand the Square

$$\begin{aligned} y_{\text{nonlinear}} &= b[A_1 \cos(2\pi f_1 t) + A_2 \cos(2\pi f_2 t)]^2 \\ &= b[A_1^2 \cos^2(2\pi f_1 t) + A_2^2 \cos^2(2\pi f_2 t) \\ &\quad + 2A_1 A_2 \cos(2\pi f_1 t)\cos(2\pi f_2 t)] \end{aligned} \tag{27}$$

4.7.4 Step 3: Apply Trigonometric Identities

We now apply two key trigonometric identities to simplify each term.

Identity 1 (Power-reduction formula):

$$\cos^2(\theta) = \frac{1}{2}[1 + \cos(2\theta)] \tag{28}$$

Applying this to the first two terms:

$$A_1^2 \cos^2(2\pi f_1 t) = \frac{A_1^2}{2}[1 + \cos(4\pi f_1 t)] \tag{29}$$

$$A_2^2 \cos^2(2\pi f_2 t) = \frac{A_2^2}{2}[1 + \cos(4\pi f_2 t)] \tag{30}$$

Identity 2 (Product-to-sum formula):

$$\cos(\alpha)\cos(\beta) = \frac{1}{2}[\cos(\alpha - \beta) + \cos(\alpha + \beta)] \tag{31}$$

Applying this to the cross-term:

$$\begin{aligned} 2A_1A_2\cos(2\pi f_1 t)\cos(2\pi f_2 t) &= 2A_1A_2 \cdot \frac{1}{2}[\cos(2\pi(f_1 - f_2)t) \\ &\quad + \cos(2\pi(f_1 + f_2)t)] \\ &= A_1A_2[\cos(2\pi(f_1 - f_2)t) + \cos(2\pi(f_1 + f_2)t)] \end{aligned} \tag{32}$$

4.7.5 Step 4: Complete Expansion

Combining all terms:

$$\begin{aligned} y_{\text{nonlinear}} = b\Bigg[&\frac{A_1^2}{2} + \frac{A_2^2}{2} + \frac{A_1^2}{2}\cos(4\pi f_1 t) + \frac{A_2^2}{2}\cos(4\pi f_2 t) \\ &+ A_1A_2\cos(2\pi(f_1 - f_2)t) + A_1A_2\cos(2\pi(f_1 + f_2)t)\Bigg] \end{aligned} \tag{33}$$

4.7.6 Step 5: Identify Frequency Components

The output contains the following frequency components:

1. **DC component (zero frequency):**

$$\frac{b}{2}(A_1^2 + A_2^2) \tag{34}$$

2. **Second harmonics:**

$$\frac{bA_1^2}{2}\cos(4\pi f_1 t) \quad \text{at frequency } 2f_1 \tag{35}$$

$$\frac{bA_2^2}{2}\cos(4\pi f_2 t) \quad \text{at frequency } 2f_2 \tag{36}$$

3. **Difference tone:**

$$bA_1A_2\cos(2\pi(f_1 - f_2)t) \quad \text{at frequency } |f_1 - f_2| \tag{37}$$

4. **Summation tone:**

$$bA_1A_2\cos(2\pi(f_1 + f_2)t) \quad \text{at frequency } f_1 + f_2 \tag{38}$$

4.7.7 Physical Interpretation

Key Result: The nonlinear term bx^2 creates *new* frequencies that were not present in the original input:

- The **difference frequency** $|f_1 - f_2|$ (the famous Tartini tone)

- The **summation frequency** $f_1 + f_2$
- Second harmonics $2f_1$ and $2f_2$

Amplitude Relationships:

- The difference and sum tones have amplitude proportional to A_1A_2 (product of input amplitudes)
- The second harmonics have amplitudes proportional to A_1^2 and A_2^2
- The nonlinearity coefficient b scales all the distortion products

Musical Implications:

1. **Difference tones are more audible** than summation tones because:
   - They fall in a frequency range where the ear is more sensitive
   - They are less likely to be masked by the primary tones
   - Lower frequencies generally require less energy to be perceived
2. **Amplitude dependence:** Since combination tones depend on A_1A_2, they are:
   - More prominent when both tones are loud
   - Effectively absent at low volumes
   - Characteristic of loud instruments (organs, brass)
3. **Consonance implications:** For consonant intervals with simple frequency ratios:
   - The difference tone often reinforces the harmonic series
   - Example: Perfect fifth ($f_2/f_1 = 3/2$) gives $f_d = f_2 - f_1 = f_1/2$
   - This reinforcement contributes to the perception of consonance

Example 4.1 (Calculation for Perfect Fourth)**.** Consider two tones at 600 Hz and 800 Hz (approximately a perfect fourth with ratio 4:3).

Given: $f_1 = 600$ Hz, $f_2 = 800$ Hz

Difference tone:

$$f_d = |f_2 - f_1| = |800 - 600| = 200 \text{ Hz}$$

This is approximately G3, one octave below the fundamental implied by the harmonic series.

Summation tone:

$$f_s = f_1 + f_2 = 600 + 800 = 1400 \text{ Hz}$$

This is approximately F6.

Physical interpretation: The difference tone at 200 Hz is two octaves below 800 Hz and reinforces the harmonic structure. The prominence of these combination tones in organs explains why organists can perceive bass notes from difference tones even when the fundamental is weak or absent.

4.7.8 Extension to Cubic Nonlinearity

For completeness, a cubic nonlinearity cx^3 produces additional combination tones. Using the expansion:

$$\cos^3(\theta) = \frac{3\cos(\theta) + \cos(3\theta)}{4} \tag{39}$$

and similar product identities, the cubic term generates:

- Third harmonics: $3f_1, 3f_2$
- **Cubic difference tones:** $2f_1 - f_2, 2f_2 - f_1$
- Various other combinations

The cubic difference tone $2f_1 - f_2$ is often more prominent than the quadratic difference tone $f_1 - f_2$ in biological systems, as the auditory system exhibits stronger cubic than quadratic nonlinearity.

4.8 Practical Significance

In Music Performance:

- String players use difference tones to check intonation of double stops
- Organ builders consider combination tones when voicing stops
- The richness of certain intervals partly derives from combination tones

In Hearing:

- Combination tones are mostly generated in the cochlea
- They contribute to the perception of beats and roughness
- They can affect the perception of consonance and dissonance

4.9 Example Calculation

Example 4.2 (Combination Tones from Musical Interval)**.** Consider a perfect fifth: $f_1 = 440$ Hz (A4) and $f_2 = 660$ Hz (E5).

Difference tone:

$$f_d = 660 - 440 = 220 \text{ Hz (A3, an octave below A4)}$$

Summation tone:

$$f_s = 660 + 440 = 1100 \text{ Hz (C\#6, approximately)}$$

Cubic difference tone:

$$f_c = 2(440) - 660 = 220 \text{ Hz (A3)}$$

Notice that both the difference tone and cubic difference tone reinforce A3, which is part of the harmonic series of A4, contributing to the consonant quality of the fifth.

4.10 Questions and Answers

Question 10 (Short Answer): What is a sub-harmonic and how does it differ from a harmonic?

Answer: A sub-harmonic is a frequency component that is an integer fraction of the fundamental frequency ($f_{\text{sub}} = f_0/n$), meaning it is below the fundamental. In contrast, a harmonic is an integer multiple of the fundamental ($f_n = nf_0$), meaning it is above the fundamental. Sub-harmonics are less common and typically occur in nonlinear systems or under special conditions.

Question 11 (Medium Answer): Calculate all primary combination tones for a major third interval between C4 (261.63 Hz) and E4 (329.63 Hz). Identify what notes these combination tones correspond to.

Answer: Given: $f_1 = 261.63$ Hz (C4), $f_2 = 329.63$ Hz (E4)

Difference tone:

$$f_d = 329.63 - 261.63 = 68.00 \text{ Hz} \approx C2$$

Summation tone:

$$f_s = 329.63 + 261.63 = 591.26 \text{ Hz} \approx D5$$

Cubic difference tone:

$$f_c = 2(261.63) - 329.63 = 193.63 \text{ Hz} \approx G3$$

Identification:

- 68 Hz $\approx$ C2 (two octaves below C4)
- 591 Hz $\approx$ D5 (between C5 at 523 Hz and E5 at 659 Hz)
- 194 Hz $\approx$ G3 (close to 196 Hz)

Interestingly, the cubic difference tone (G3) forms a perfect fifth below C4, reinforcing harmonic relationships in the major third interval.

5 Limitations of Fourier Analysis: Noise and Beyond

5.1 What Fourier Analysis Tells Us

Fourier analysis is a powerful tool that reveals:

- The frequency content of periodic signals (Fourier series)
- The frequency spectrum of arbitrary signals (Fourier transform)
- The relationship between time and frequency domains

However, Fourier analysis has significant limitations when applied to certain types of sounds.

5.2 The Time-Frequency Trade-off

A fundamental limitation of Fourier analysis stems from the Heisenberg-Gabor uncertainty principle:

Theorem 5.1 (Uncertainty Principle)**.** For any signal, there is a fundamental trade-off between time resolution and frequency resolution:

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi} \tag{40}$$

where Δt is the uncertainty in time and Δf is the uncertainty in frequency.

Implications:

- To know frequency precisely requires analyzing over a long time window
- To know when something happens precisely requires sacrificing frequency resolution
- Cannot simultaneously have perfect time and frequency resolution

5.3 Types of Noise

Definition 5.1 (Noise)**.** In signal processing, noise refers to random or aperiodic signals. Unlike tones, noise has continuous (non-discrete) spectra.

Common Types of Noise:

1. **White Noise**:
   - Equal power at all frequencies
   - Power spectrum: $P(f) = \text{constant}$
   - Sounds like ”static” or ”hiss”
   - $1/f^0$ spectrum
2. **Pink Noise (1/f Noise)**:
   - Power decreases with frequency
   - Power spectrum: $P(f) \propto 1/f$
   - Equal power per octave
   - Sounds more ”natural” than white noise
   - Common in natural phenomena and music
3. **Brown Noise (Red Noise)**:
   - Power decreases faster with frequency
   - Power spectrum: $P(f) \propto 1/f^2$
   - Sounds ”darker” and ”rumbling”
   - Models Brownian motion
4. **Blue Noise**:
   - Power increases with frequency
   - Power spectrum: $P(f) \propto f$
   - Sounds very bright and hissy

5.4 Why Fourier Analysis Fails for Noise

For random, non-periodic signals:

1. **No Discrete Spectrum**:
   - Noise has continuous spectrum, not discrete peaks
   - Fourier series doesn't apply (not periodic)
   - Must use Fourier transform, but interpretation is different
2. **Statistical Description Needed**:
   - Must describe average properties (power spectral density)
   - Individual Fourier coefficients are meaningless
   - Need to analyze many samples and average
3. **Non-Stationary Signals**:
   - Most real sounds change over time
   - Standard Fourier transform assumes infinite duration
   - Statistical properties vary with time

5.5 Musical Significance of Noise

Noise plays crucial roles in music:

- **Percussion Sounds**: Drums, cymbals contain significant noise components
- **Attack Transients**: The "th" in "think" or bow scratch on violin
- **Breath Sounds**: Flutes, recorders, and voice contain filtered noise
- **Texture**: Electronic music uses noise for sonic variety
- **Realism**: Adding noise to synthetic sounds makes them more natural

5.6 Generating Different Colors of Noise

White Noise Generation: Simply use random numbers: $x[n] = \text{random}(-1, 1)$

Pink Noise Generation: Filter white noise with a -3 dB/octave filter:

$$H(f) = \frac{1}{\sqrt{f}}$$

Fractal Noise: Sum multiple octaves of noise:

$$x(t) = \sum_{k=0}^{N} \frac{1}{2^k} \cdot \text{noise}(2^k t) \tag{41}$$

This creates self-similar structure across time scales.

5.7 Alternative Analysis Methods

Due to Fourier analysis limitations, alternative methods have been developed:

1. **Short-Time Fourier Transform (STFT)**:
    - Apply Fourier transform to short, windowed segments
    - Trades frequency resolution for time localization
    - Creates spectrogram showing time-varying spectrum
2. **Wavelet Analysis**:
    - Uses wavelets instead of sine waves as basis functions
    - Better time-frequency localization for transients
    - Variable time-frequency resolution
3. **Cepstral Analysis**:
    - "Spectrum of a spectrum"
    - Useful for separating source and filter in speech
    - Used in audio coding and voice recognition

5.8 Questions and Answers

Question 12 (Short Answer): State the Heisenberg-Gabor uncertainty principle for signal analysis and explain its practical implication.

Answer: The uncertainty principle states: $\Delta t \cdot \Delta f \geq 1/(4\pi)$, meaning the product of time uncertainty and frequency uncertainty cannot be arbitrarily small. Practically, this means we cannot simultaneously know exactly when a frequency occurs and what that exact frequency is—improving time resolution worsens frequency resolution and vice versa. This is why analyzing short, transient sounds with precise timing necessarily sacrifices frequency accuracy.

Question 13 (Short Answer): What is pink noise and why is it perceptually important?

Answer: Pink noise (1/f noise) has a power spectrum where power decreases proportionally to frequency: $P(f) \propto 1/f$. This means it has equal power per octave. Pink noise is perceptually important because:

- It sounds more "natural" than white noise to human ears
- Many natural phenomena exhibit 1/f characteristics
- Musical pieces often show 1/f statistical structure
- It is commonly used in audio testing and calibration

Question 14 (Long Answer): Explain why standard Fourier analysis is inadequate for analyzing percussion instruments like drums and cymbals. What are the main challenges, and what alternative approaches exist?

Answer:

Challenges with Percussion Analysis:

1. Inharmonic Spectra:

- Percussion instruments have partials that are not integer multiples of a fundamental
- Fourier series assumes periodicity, but percussion sounds are not truly periodic
- The concept of "fundamental frequency" is ambiguous or meaningless
- Example: A circular drum has modes at approximately $1.00f, 1.59f, 2.14f, 2.30f, 2.65f, \ldots$ (not integer ratios)

2. Rapid Decay and Non-Stationarity:

- Percussion sounds decay quickly (typically < 1 second)
- Different partials decay at different rates
- High frequencies often decay faster than low frequencies
- Standard Fourier transform assumes infinite or very long duration
- The spectrum is time-varying, but standard Fourier gives time-averaged spectrum

3. Significant Noise Components:

- Impact sounds (drum stick hitting skin) contain broadband noise
- Cymbals and hi-hats have large noise components
- Noise has continuous spectrum, not discrete peaks
- Fourier analysis is designed for tonal, not noise, components

4. Complex Initial Transients:

- The attack contains crucial timbral information
- Transients require good time resolution
- Fourier analysis requires long windows for good frequency resolution
- Uncertainty principle: cannot have both precise time and frequency

Mathematical Issues:
For a drum membrane, the wave equation in 2D:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \nabla^2 u$$

with circular boundary conditions leads to Bessel function solutions with eigenfrequencies:

$$f_{mn} = \frac{c}{2\pi R} \alpha_{mn}$$

where α_{mn} are zeros of Bessel functions, forming inharmonic ratios.
Alternative Approaches:
1. Short-Time Fourier Transform (STFT):

$$\text{STFT}(t, f) = \int_{-\infty}^{\infty} s(\tau) w(\tau - t) e^{-i2\pi f \tau} d\tau \tag{42}$$

where $w(t)$ is a window function centered at time t. This creates a spectrogram showing how the spectrum evolves.

2. Wavelet Transform:

$$W(a,b) = \int_{-\infty}^{\infty} s(t)\psi^*\left(\frac{t-b}{a}\right) dt \tag{43}$$

Uses wavelets ψ that provide better time-frequency localization for transients. Variable resolution: good time resolution at high frequencies, good frequency resolution at low frequencies.

3. Modal Analysis: Decompose into damped sinusoids:

$$s(t) = \sum_n A_n e^{-\gamma_n t} \sin(2\pi f_n t + \phi_n) \tag{44}$$

Explicitly models inharmonic partials and their individual decay rates.

4. Physical Modeling: Directly simulate the physics:

- Solve wave equation numerically (finite element/difference methods)
- Model the actual geometry and material properties
- Captures inharmonicity and complex decay naturally

Conclusion: Standard Fourier analysis fails for percussion because it assumes periodicity, stationarity, and infinite duration—none of which apply to percussion instruments. The time-frequency uncertainty principle fundamentally limits our ability to simultaneously capture the rapid transients and the inharmonic frequency content. Modern alternatives like STFT, wavelets, and modal analysis provide better tools by explicitly addressing these limitations through time-varying representations and better-matched basis functions.

6 Introduction to Sound Synthesis

6.1 What is Sound Synthesis?

Definition 6.1 (Sound Synthesis)**.** Sound synthesis is the process of generating audio signals using electronic or computational means, rather than by acoustic means (e.g., vibrating strings or air columns).

If Fourier analysis is the process of decomposing sounds into components, synthesis is essentially the reverse: creating sounds from basic building blocks.

6.2 Why Synthesize Sound?

Historical Motivations:

- Create sounds not possible with acoustic instruments
- Electronic instruments (1960s-1970s)
- Film and game sound effects

- Cost and portability advantages

Modern Motivations:

- Precise control over all sound parameters
- Reproducibility and automation
- Creative exploration of new timbres
- Understanding sound through reconstruction
- Virtual instruments in music production

6.3 Basic Synthesis Components

Key Elements in Most Synthesis Systems:

1. **Oscillators**: Generate basic waveforms (sine, square, sawtooth, triangle)
2. **Filters**: Shape the spectrum (low-pass, high-pass, band-pass)
3. **Envelopes**: Control how parameters change over time
4. **Modulators**: Add variation and expressiveness (LFO, vibrato, tremolo)
5. **Effects**: Add realism and space (reverb, delay, chorus)

6.4 The Importance of Envelopes

One of the most important discoveries in synthesis was the critical role of the attack portion of a sound.

Definition 6.2 (ADSR Envelope)**.** The ADSR envelope describes how the amplitude (or other parameter) changes over time:

- **Attack**: Time to reach peak amplitude from note onset
- **Decay**: Time to fall from peak to sustain level
- **Sustain**: Level maintained while note is held
- **Release**: Time to fade to silence after note is released

Remark 6.1 (Attack Transients are Crucial)**.** Research has shown that the attack portion (first 50-100 ms) of a note is the most vital for instrument identification. Removing the attack makes it very difficult to identify the instrument, even if the steady-state portion is preserved. This is why envelope generators in synthesizers often have separate controls for attack time.

6.5 Envelope Applications

Envelopes control not just amplitude, but also:

- **Filter Cutoff**: To simulate brightness changes (e.g., bell tones start bright and decay to darker tone)
- **Pitch**: For pitch bends or slides at note onset
- **Modulation Depth**: To add expression
- **Pan Position**: For spatial movement

6.6 Low Frequency Oscillators (LFOs)

Definition 6.3 (LFO)**.** A Low Frequency Oscillator produces periodic modulation in the range of approximately 0.1-20 Hz, below the audio range but perceptible as variation.

Common LFO Applications:

- **Vibrato**: Periodic pitch modulation (typically 5-7 Hz)
- **Tremolo**: Periodic amplitude modulation
- **Timbral Variation**: Modulating filter parameters
- **Pan Modulation**: Creating stereo movement

LFO parameters:

- Rate/Frequency: How fast the modulation occurs
- Depth/Amount: How much modulation is applied
- Waveform: Sine (smooth), triangle, square (abrupt), sawtooth, random
- Attack Time: Gradual fade-in of modulation

6.7 Basic Waveforms and Their Spectra

Common Oscillator Waveforms:

1. **Sine Wave**:
   - Spectrum: Single frequency component
   - Sound: Pure, flute-like
   - Use: Basic building block, sub-bass, meditation music
2. **Square Wave**:
   - Spectrum: Odd harmonics only $(1, 3, 5, 7, \ldots)$ with amplitudes $1/n$
   - Sound: Hollow, clarinet-like, "woody"
   - Use: Leads, basses, chiptune music

3. **Sawtooth Wave**:
   - Spectrum: All harmonics $(1, 2, 3, 4, \ldots)$ with amplitudes $1/n$
   - Sound: Bright, buzzy, brassy
   - Use: String pads, brass, aggressive leads
4. **Triangle Wave**:
   - Spectrum: Odd harmonics with amplitudes $1/n^2$
   - Sound: Softer than square, like a muted woodwind
   - Use: Mellow leads, flute-like sounds
5. **Pulse Wave**:
   - Spectrum: Depends on pulse width (duty cycle)
   - Sound: From thin/nasal to square wave
   - Use: Bass sounds, rhythmic elements

6.8 Questions and Answers

Question 15 (Short Answer): What is an ADSR envelope and why is the Attack phase particularly important?

Answer: ADSR stands for Attack-Decay-Sustain-Release, describing how amplitude (or other parameters) evolve over time:

- Attack: Time to reach peak from onset
- Decay: Time to fall to sustain level
- Sustain: Level while note is held
- Release: Time to fade to silence

The Attack phase is particularly important because research shows that the first 50-100 ms contains crucial information for instrument identification. The attack transients (initial irregularities, noise, and spectral evolution) are what our brains use to distinguish a piano from a guitar or a trumpet from a flute, even if they play the same note.

Question 16 (Medium Answer): Compare the spectral content and timbral characteristics of square waves and sawtooth waves. Include mathematical expressions for their Fourier series.

Answer:

Square Wave:

Fourier series:

$$f_{\text{square}}(t) = \frac{4}{\pi} \sum_{n=1,3,5,\ldots}^{\infty} \frac{1}{n} \sin(2\pi n f t)$$

Characteristics:

- Contains only odd harmonics (1st, 3rd, 5th, 7th, ...)
- Harmonic amplitudes decrease as $1/n$

- No even harmonics present
- Spectrum is relatively sparse
- Sound: Hollow, woody, clarinet-like quality
- Perceptually "softer" than sawtooth despite same fundamental amplitude

Sawtooth Wave:
Fourier series:

$$f_{\text{saw}}(t) = \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(2\pi n f t)$$

Characteristics:

- Contains all harmonics (1st, 2nd, 3rd, 4th, ...)
- Harmonic amplitudes also decrease as $1/n$
- Even harmonics present, filling in the gaps
- Spectrum is much denser
- Sound: Bright, buzzy, brassy, "aggressive"
- Perceptually much brighter due to presence of all harmonics

Key Differences:

- Sawtooth has twice as many spectral components as square wave
- Sawtooth sounds approximately twice as bright (higher spectral centroid)
- Square wave has a more "hollow" quality due to missing even harmonics
- Musically, sawtooth is used for aggressive sounds (brass, strings), while square wave is used for softer, woodwind-like sounds

7 Synthesis Methods

7.1 Additive Synthesis

Definition 7.1 (Additive Synthesis)**.** Additive synthesis creates sounds by summing individual sinusoidal components (partials), each with controllable frequency, amplitude, and phase.

Mathematical Formulation:

$$s(t) = \sum_{n=1}^{N} A_n(t) \sin[2\pi f_n t + \phi_n(t)] \tag{45}$$

where:

- $A_n(t)$ = time-varying amplitude of n-th partial

- f_n = frequency of n-th partial
- $\phi_n(t)$ = time-varying phase

Advantages:

- Complete control over every aspect of timbre
- Can recreate any periodic sound (by Fourier theory)
- Individual control over each partial's envelope
- Can create inharmonic spectra easily

Disadvantages:

- Computationally expensive (many oscillators needed)
- Requires detailed analysis of target sounds
- Many parameters to control ($3N$ parameters for N partials)
- Difficult to create realistic dynamic timbres

Historical Example: The Hammond organ (1930s) uses additive synthesis with drawbars controlling the amplitude of different harmonics, though it's a simplified version with fixed harmonic relationships.

7.2 Subtractive Synthesis

Definition 7.2 (Subtractive Synthesis)**.** Subtractive synthesis starts with harmonically rich waveforms (sawtooth, pulse) and uses filters to remove frequency content, sculpting the desired timbre.

Basic Signal Flow:

$$\text{Oscillator} \rightarrow \text{Filter} \rightarrow \text{Amplifier}$$

Each stage has an envelope:

$$\text{Env}_{\text{pitch}} \rightarrow \text{Env}_{\text{filter}} \rightarrow \text{Env}_{\text{amplitude}}$$

Filter Types:

- **Low-pass**: Removes high frequencies, creates darker sound
- **High-pass**: Removes low frequencies, creates thinner sound
- **Band-pass**: Keeps middle frequencies, creates nasal quality
- **Notch/Band-reject**: Removes middle frequencies

Filter Parameters:

- **Cutoff Frequency (f_c)**: Where filtering begins to take effect

- **Resonance (Q)**: Emphasis at cutoff frequency
- **Slope**: How steeply filter attenuates (dB/octave)

Mathematical Model (Simple Low-Pass): Transfer function:

$$H(f) = \frac{1}{1 + i2\pi f/f_c} \tag{46}$$

With resonance:

$$H(f) = \frac{1}{1 + iQ(f/f_c - f_c/f)} \tag{47}$$

Advantages:

- Computationally efficient (one oscillator + filter)
- Intuitive controls (brightness, resonance)
- Excellent for analog-style sounds
- Real-time parameter modulation is natural

Disadvantages:

- Cannot create arbitrary spectra
- Limited to removing content, not adding it
- Cannot produce clean sine waves or sparse spectra

7.3 Frequency Modulation (FM) Synthesis

Definition 7.3 (FM Synthesis)**.** FM synthesis creates complex spectra by using one oscillator (modulator) to modulate the frequency of another oscillator (carrier).

Basic FM Equation:

$$s(t) = A\sin[2\pi f_c t + I\sin(2\pi f_m t)] \tag{48}$$

where:

- f_c = carrier frequency
- f_m = modulator frequency
- I = modulation index (controls brightness)
- A = overall amplitude

Resulting Spectrum: Using Bessel functions, the spectrum contains components at:

$$f_n = f_c + nf_m, \quad n = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{49}$$

with amplitudes:

$$A_n = A \cdot J_n(I) \tag{50}$$

where J_n is the Bessel function of the first kind, order n.

Modulation Index Effects:

- $I = 0$: Pure sine wave (no modulation)
- I small (< 1): Few sidebands, soft timbre
- I large (> 5): Many sidebands, bright, complex timbre
- Increasing I increases brightness (more harmonics)

Harmonic Ratios: When $f_c : f_m$ are integer ratios, the spectrum is harmonic:

- $f_c : f_m = 1 : 1 \rightarrow$ Odd and even harmonics
- $f_c : f_m = 2 : 1 \rightarrow$ Even harmonics emphasized
- $f_c : f_m = 1 : 2 \rightarrow$ Subharmonic effects

Advantages:

- Very efficient (only 2 oscillators for rich spectra)
- Wide range of timbres from simple setup
- Modulation index controls brightness naturally
- Excellent for metallic, bell-like, and electric piano sounds

Disadvantages:

- Non-intuitive relationship between parameters and timbre
- Difficult to predict exact spectral result
- Limited to certain classes of timbres
- Complex enveloping requires many operators

7.4 Granular Synthesis

Definition 7.4 (Granular Synthesis)**.** Granular synthesis creates sound textures by combining thousands of short sound particles called grains, typically 10-100 ms in duration.

Mathematical Model:

$$s(t) = \sum_{i=1}^{N} g_i(t - t_i) \cdot w_i(t - t_i) \tag{51}$$

where:

- $g_i(t)$ = grain i (short waveform)
- $w_i(t)$ = window function (envelope for grain)
- t_i = onset time of grain i

Grain Parameters:

- Duration: 10-100 ms typically
- Density: Grains per second
- Pitch: Frequency content of each grain
- Amplitude: Volume of each grain
- Spatial position: Pan/location

Grain Types:

- **Sinusoidal**: Pure tones
- **Sampled**: Short segments from recordings
- **Synthesized**: Generated waveforms

Applications:

- Time-stretching without pitch change
- Pitch-shifting without time change
- Creating evolving textures and clouds
- Extreme sound transformations
- Ambient and experimental music

7.5 Physical Modeling Synthesis

Definition 7.5 (Physical Modeling)**.** Physical modeling synthesis simulates the physics of sound-producing mechanisms using mathematical models of vibrating systems.

Approaches:

1. **Modal Synthesis**: Model as sum of damped resonators:

$$s(t) = \sum_n A_n e^{-\gamma_n t} \sin(2\pi f_n t + \phi_n) \tag{52}$$

2. **Waveguide Synthesis**: Model traveling waves in strings/tubes:

$$y(t) = \alpha \cdot y(t - \tau) + \text{input}(t) \tag{53}$$

where τ is the delay time related to string length.

3. **Finite Element/Difference Methods**: Numerically solve wave equation:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \nabla^2 u \tag{54}$$

Advantages:

- Physically accurate and realistic

- Parameters correspond to physical properties (tension, length)
- Naturally handles complex interactions
- Automatic generation of realistic timbres

Disadvantages:

- Computationally expensive
- Requires detailed physical understanding
- Limited to sounds that can be physically modeled
- Some parameters difficult to control musically

7.6 Questions and Answers

Question 17 (Short Answer): Compare and contrast additive and subtractive synthesis approaches.

Answer:

Additive Synthesis:

- Builds sound by summing sine waves
- Complete spectral control
- Computationally expensive (many oscillators)
- "Bottom-up" approach

Subtractive Synthesis:

- Starts with rich waveform, removes content with filters
- More limited spectral control
- Computationally efficient
- "Top-down" approach

Both can create musical sounds, but additive offers more precision while subtractive offers more efficiency and intuitive real-time control.

Question 18 (Long Answer): Explain FM synthesis in detail. Include the mathematical formulation, discuss the role of the modulation index, and explain how harmonic vs. inharmonic spectra are generated. Provide a specific example calculation.

Answer:

FM Synthesis Fundamentals:

FM (Frequency Modulation) synthesis was developed by John Chowning at Stanford in the late 1960s and revolutionized digital synthesis, particularly in the Yamaha DX7 synthesizer.

Mathematical Formulation:

The basic FM equation is:

$$s(t) = A \sin[2\pi f_c t + I \sin(2\pi f_m t)] \tag{55}$$

where:

- f_c = carrier frequency (determines perceived pitch)
- f_m = modulator frequency (determines spectral spacing)
- I = modulation index (determines brightness/complexity)
- A = amplitude

The modulator creates instantaneous frequency deviations:

$$f_{\text{inst}}(t) = f_c + \Delta f \cos(2\pi f_m t) \tag{56}$$

where the peak deviation is $\Delta f = I \cdot f_m$.

Spectral Analysis:

Using the Bessel function expansion, the spectrum contains components at frequencies:

$$f_n = f_c + n f_m, \quad n = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{57}$$

with amplitudes given by Bessel functions of the first kind:

$$A_n = A \cdot J_n(I) \tag{58}$$

Role of Modulation Index:

The modulation index I is the single most important parameter:

- $I \approx 0$: Only the carrier frequency is present (sine wave)
- $I < 1$: A few sidebands appear, soft timbre
- $I = 2$: Moderate number of sidebands, rich but controlled
- $I = 5$: Many sidebands, bright metallic timbre
- $I > 10$: Very complex spectrum, harsh noise-like quality

The approximate bandwidth is:

$$BW \approx 2(I + 1) f_m \tag{59}$$

This is known as Carson's rule. As I increases, more frequency components appear and the timbre becomes brighter.

Harmonic vs. Inharmonic Spectra:

The relationship between f_c and f_m determines whether the spectrum is harmonic:

Harmonic Spectra ($f_c : f_m$ = integers):

Example: $f_c = 440$ Hz, $f_m = 440$ Hz (ratio 1:1)
Spectrum: $440 + n(440) = 440, 880, 1320, 1760, \ldots$ Hz
All components are integer multiples of 440 Hz → Harmonic series!
Perceived as clear pitch at 440 Hz.
Example: $f_c = 880$ Hz, $f_m = 440$ Hz (ratio 2:1)
Spectrum: $880 + n(440) = 440, 880, 1320, 1760, 2200, \ldots$ Hz
Even-numbered components emphasized, fundamental at 440 Hz.

Inharmonic Spectra ($f_c : f_m$ = non-integers):

Example: $f_c = 440$ Hz, $f_m = 333$ Hz (ratio ≈ 1.32:1)

Spectrum: $440 + n(333) = 107, 440, 773, 1106, 1439, \ldots$ Hz
Not integer multiples $\rightarrow$ Inharmonic!
Perceived as bell-like, metallic, or ambiguous pitch.
Detailed Example Calculation:
Let's calculate the spectrum for: - $f_c = 1000$ Hz - $f_m = 500$ Hz (ratio 2:1) - $I = 2$
Step 1: Determine frequency components:

$$\begin{aligned} f_n &= 1000 + 500n \\ f_{-4} &= -1000 \text{ Hz (ignored—negative freq)} \\ f_{-3} &= -500 \text{ Hz (ignored)} \\ f_{-2} &= 0 \text{ Hz (DC)} \\ f_{-1} &= 500 \text{ Hz} \\ f_0 &= 1000 \text{ Hz (carrier)} \\ f_1 &= 1500 \text{ Hz} \\ f_2 &= 2000 \text{ Hz} \\ f_3 &= 2500 \text{ Hz} \\ f_4 &= 3000 \text{ Hz} \end{aligned}$$

Step 2: Calculate amplitudes using Bessel functions $J_n(2)$:

$$\begin{aligned} J_0(2) &\approx 0.224 \\ J_1(2) &\approx 0.577 \\ J_2(2) &\approx 0.353 \\ J_3(2) &\approx 0.129 \\ J_4(2) &\approx 0.034 \end{aligned}$$

Step 3: Construct spectrum (relative amplitudes):

Frequency (Hz)	**Amplitude**
500	0.577
1000	0.224
1500	0.577
2000	0.353
2500	0.129
3000	0.034

Observations:

- Harmonic series: 500, 1000, 1500, 2000, 2500, 3000 Hz (all multiples of 500 Hz)
- Fundamental is at 500 Hz, not the carrier (1000 Hz)!
- Sidebands (500, 1500 Hz) are stronger than carrier (1000 Hz)
- Creates a clear pitch at 500 Hz with a rich, complex timbre

Dynamic Modulation Index:
In practice, I is often controlled by an envelope:

$$I(t) = I_{\max} \cdot \text{ADSR}(t)$$

This creates dynamic timbral evolution:

- Attack: I increases rapidly $\rightarrow$ bright, complex attack
- Decay: I decreases $\rightarrow$ darkens to sustain level
- Release: I drops to zero $\rightarrow$ returns to pure tone

This mimics the behavior of acoustic instruments where the attack is bright and the sustain is mellower.

Conclusion:

FM synthesis is remarkably efficient—with just two oscillators and one control parameter (I), it can create spectra ranging from simple sine waves to complex, bell-like timbres. The key insights are:

1. Modulation index controls spectral richness/brightness
2. Integer frequency ratios produce harmonic spectra (clear pitch)
3. Non-integer ratios produce inharmonic spectra (bells, gongs)
4. Dynamic modulation index creates evolving timbres

This explains why FM synthesis became so popular—it offers great sonic variety with computational efficiency and relatively simple parameter control.

8 Spectral Envelopes and Timbre

8.1 What is a Spectral Envelope?

Definition 8.1 (Spectral Envelope)**.** The spectral envelope is a smooth curve that represents the overall shape of a sound's frequency spectrum, connecting the peaks of individual partials and ignoring fine details.

Mathematically, if $S(f)$ is the magnitude spectrum with discrete peaks at partial frequencies f_n, the spectral envelope $E(f)$ is a smooth interpolation:

$$E(f) = \text{smooth}(\{|S(f_n)|\}) \tag{60}$$

8.2 Importance of Spectral Envelope

The spectral envelope is crucial for timbre perception because:

1. **Timbre Identity**: Sounds with similar spectral envelopes sound similar, even at different pitches
2. **Formants in Speech**: Vowel identity is determined by spectral envelope peaks (formants), not the fundamental frequency
3. **Instrument Families**: Instruments in the same family (e.g., violin, viola, cello) share similar spectral envelope shapes
4. **Brightness**: The envelope's center of mass determines perceived brightness

8.3 Source-Filter Model

A powerful framework for understanding spectral envelopes:

Definition 8.2 (Source-Filter Model)**.** Many sounds can be modeled as a source spectrum $S(f)$ passing through a filter $H(f)$:

$$\text{Output}(f) = S(f) \cdot H(f) \tag{61}$$

In the time domain (convolution):

$$\text{output}(t) = \text{source}(t) * h(t) \tag{62}$$

Examples:

- **Voice**: Vocal cord vibration (source) filtered by vocal tract (filter)
- **Violin**: Bow-string interaction (source) filtered by body resonance (filter)
- **Wind Instruments**: Reed/air jet (source) filtered by bore resonances (filter)

The source provides the harmonic series; the filter shapes the spectral envelope.

8.4 Formants

Definition 8.3 (Formant)**.** A formant is a peak in the spectral envelope, typically corresponding to a resonance in the sound-producing system.

In Human Voice: The vocal tract acts as a resonator with typically 3-5 prominent formants:

- F1: First formant (lowest, 300-800 Hz)
- F2: Second formant (800-2500 Hz)
- F3: Third formant (2000-3500 Hz)

Different vowels have characteristic F1-F2 patterns:

Vowel	F1 (Hz)	F2 (Hz)
/i/ (bee)	300	2300
/a/ (father)	750	1200
/u/ (boot)	300	800
/e/ (bed)	550	1900

Key Insight: When you whisper (no vocal cord vibration), you can still distinguish vowels because the formant structure is preserved—this is purely the filter (vocal tract shape), without the source (vocal cords).

8.5 Static vs. Dynamic Spectra

Static Spectrum: A snapshot at one instant in time, showing amplitude vs. frequency.

Dynamic Spectrum: Shows how the spectrum evolves over time—the complete description includes three dimensions:

- Time
- Frequency
- Amplitude

Representation Methods:

1. **Waterfall Plot**: Multiple static spectra stacked in time order
2. **Spectrogram (Sonogram)**: Time on x-axis, frequency on y-axis, amplitude shown by color/brightness
3. **3D Surface Plot**: Time and frequency as horizontal axes, amplitude as height

8.6 Amplitude Envelope vs. Spectral Envelope

Important distinction:

- **Amplitude Envelope** $A(t)$: How overall loudness changes over time (ADSR) Time vs. amplitude (1D over time)
- **Spectral Envelope** $E(f)$: Distribution of energy across frequencies Frequency vs. amplitude (1D over frequency)
- **Time-Varying Spectral Envelope** $E(f,t)$: How the spectrum shape evolves over time Requires 2D representation (frequency and time)

8.7 Practical Applications

In Synthesis:

- Subtractive synthesis: Filter controls spectral envelope
- FM synthesis: Modulation index affects spectral envelope
- Additive synthesis: Directly specify partial amplitudes = spectral envelope

In Analysis:

- Speech recognition: Identify formant patterns
- Music transcription: Separate instruments by spectral envelope
- Audio effects: Vocoders transfer spectral envelope between sounds

In Vocoding: A vocoder analyzes the spectral envelope of one sound and applies it to another:

$$\text{Output} = \text{Envelope(Voice)} \times \text{Source(Synth)}$$

This creates "talking synthesizer" effects.

8.8 Example: Piano vs. Harpsichord

Both are struck string instruments with harmonic spectra, but different spectral envelopes:

Piano:

- Struck by felt hammers → softer attack
- Soundboard resonances create broad spectral envelope
- Lower harmonics relatively strong
- Smooth decay with dampers

Harpsichord:

- Plucked by plectra → sharper attack
- Smaller soundboard, less resonance
- Higher harmonics relatively stronger
- Rapid initial decay, then slower

Despite both having harmonic spectra at the same pitch, the different spectral envelopes make them instantly distinguishable.

8.9 Questions and Answers

Question 19 (Short Answer): Define spectral envelope and explain why it is important for timbre.

Answer: The spectral envelope is a smooth curve representing the overall shape of a frequency spectrum, connecting the peaks of individual partials while ignoring fine details. It is important for timbre because:

- It determines the perceived "color" or quality of a sound
- Sounds with similar spectral envelopes sound similar regardless of pitch
- It encodes information about the sound-producing mechanism (resonances, filtering)
- In speech, formants (envelope peaks) determine vowel identity
- It is more perceptually relevant than fine spectral details

Question 20 (Medium Answer): Explain the source-filter model. Give two musical examples and describe how the source and filter contribute to the final timbre.

Answer:

The source-filter model represents sound production as:

$$\text{Output}(f) = \text{Source}(f) \times \text{Filter}(f)$$

Concept:

- **Source**: Generates a spectrum (usually harmonic)

- **Filter**: Shapes the spectrum by emphasizing some frequencies and attenuating others
- The filter creates the spectral envelope

Example 1: Human Voice
Source: Vocal cords vibrating in air stream

- Produces periodic pulses (glottal pulses)
- Rich harmonic spectrum with -12 dB/octave rolloff
- Fundamental frequency determines pitch

Filter: Vocal tract (throat, mouth, lips)

- Acts as resonant tube with formants (resonance peaks)
- Formant frequencies depend on shape (tongue position, mouth opening)
- Creates spectral envelope with 3-5 prominent peaks

Result:

- Pitch determined by source (vocal cord frequency)
- Vowel identity determined by filter (formant pattern)
- Same pitch can produce different vowels by changing filter only

Example 2: Violin
Source: Bow-string interaction

- Stick-slip motion creates sawtooth-like waveform
- Rich harmonic spectrum
- Fundamental determined by string length, tension, mass

Filter: Violin body resonances

- Bridge transmits vibrations to top plate
- Air and wood resonances (body modes, f-hole resonances)
- Characteristic formants around 270 Hz, 500 Hz, and higher
- Different for each violin (Stradivarius vs. factory violin)

Result:

- Pitch determined by source (string vibration)
- Tone quality (warm vs. bright) determined by filter
- Quality of instrument largely determined by filter characteristics
- Playing position (sul ponticello vs. sul tasto) affects source spectrum

Significance: The source-filter model explains why:

- We can change pitch without changing timbre (vary source, keep filter)
- We can change timbre without changing pitch (vary filter, keep source)
- Different vowels at same pitch are possible
- Instrument quality relates to filter characteristics

9 Summary and Applications

9.1 Key Concepts Recap

Timbre Understanding:

- Timbre is multidimensional (brightness, attack, spectral evolution)
- Determined by spectral content and temporal dynamics
- Attack transients crucial for instrument identification

Fourier Analysis:

- Periodic sounds: Fourier series (discrete harmonics)
- General sounds: Fourier transform (continuous spectrum)
- Magnitude spectrum more important than phase for steady sounds
- Limitations: Time-frequency uncertainty, poor for noise and transients

Synthesis Methods:

- Additive: Sum sinusoids (flexible but expensive)
- Subtractive: Filter rich waveforms (efficient, intuitive)
- FM: Frequency modulation (efficient, complex spectra)
- Granular: Micro-sound particles (textures, time-stretching)
- Physical modeling: Simulate physics (realistic but complex)

Spectral Envelopes:

- Overall spectrum shape determines timbre character
- Source-filter model: Excitation + resonance
- Formants: Envelope peaks from resonances
- Time-varying envelopes capture dynamic timbre

9.2 Practical Applications

Music Production:

- Sound design: Understanding synthesis methods
- Mixing: EQ shapes spectral envelopes
- Effects: Filters, vocoders manipulate spectra
- Sampling: Understanding formants for pitch-shifting

Audio Engineering:

- Codec design: Perceptual coding based on spectral masking
- Room acoustics: Resonances affect spectral envelope
- Noise reduction: Distinguish signal from noise spectra

Computer Music:

- Algorithmic composition: Generating timbral variation
- Real-time synthesis: Efficient methods (FM, subtractive)
- Audio effects: Spectral processing

Research:

- Timbre perception studies
- Automatic instrument recognition
- Source separation (cocktail party problem)
- Music information retrieval

9.3 Connections to Other Units

Unit 1 (Sound Fundamentals):

- Harmonics and waveforms $\rightarrow$ basis for timbre
- Frequency perception $\rightarrow$ understanding pitch in complex spectra

Unit 2 (Tuning):

- Harmonic series $\rightarrow$ natural basis for scales
- Combination tones $\rightarrow$ affect perception of intervals

Unit 3 (Rhythm):

- Temporal envelopes complement rhythmic structure
- Attack characteristics affect rhythmic perception

Unit 5 (Perception):

- Critical bands $\rightarrow$ spectral masking
- Consonance/dissonance $\rightarrow$ spectral interaction
- Psychoacoustics $\rightarrow$ how we perceive timbre

A Quick Reference Tables

A.1 Common Waveform Spectra

Waveform	Harmonic Content	Timbre Description
Sine	Fundamental only	Pure, mellow, flute-like
Square	Odd harmonics (1,3,5,...) at $1/n$ amplitude	Hollow, woody, clarinet-like
Sawtooth	All harmonics (1,2,3,...) at $1/n$ amplitude	Bright, buzzy, brassy
Triangle	Odd harmonics at $1/n^2$ amplitude	Soft, flute-like, mellow
Pulse (50%)	Same as square wave	Hollow, narrow
Pulse (narrow)	All harmonics, relatively equal	Very bright, thin, nasal

Table 1: Waveforms and Their Spectral/Timbral Characteristics

A.2 Synthesis Method Comparison

Method	Efficiency	Control	Flexibility	Realism
Additive	Low	High	High	Low-Medium
Subtractive	High	Medium	Medium	Medium
FM	High	Low	High	Medium
Granular	Medium	Medium	High	Medium-High
Physical Modeling	Low	Medium	Medium	High

Table 2: Comparison of Synthesis Methods

A.3 Noise Types

Type	Spectrum	dB/octave	Description
White	$P(f) = C$	0	Static, hiss
Pink	$P(f) \propto 1/f$	-3	Natural, balanced
Brown/Red	$P(f) \propto 1/f^2$	-6	Dark, rumbling
Blue	$P(f) \propto f$	+3	Bright, harsh
Violet	$P(f) \propto f^2$	+6	Very bright

Table 3: Types of Noise and Their Characteristics

A.4 Important Formulas

Fourier Series (Real Form):

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}[a_n \cos(n\omega t) + b_n \sin(n\omega t)]$$

Fourier Transform:

$$S(f) = \int_{-\infty}^{\infty} s(t)e^{-i2\pi ft}dt$$

Spectral Centroid:

$$C = \frac{\sum f_n A_n}{\sum A_n}$$

FM Synthesis:

$$s(t) = A\sin[2\pi f_c t + I\sin(2\pi f_m t)]$$

Combination Tones:

$$\begin{aligned} f_{\text{diff}} &= f_2 - f_1 \\ f_{\text{sum}} &= f_1 + f_2 \\ f_{\text{cubic}} &= 2f_1 - f_2 \end{aligned}$$

Source-Filter Model:

$$\text{Output}(f) = \text{Source}(f) \cdot \text{Filter}(f)$$

B Sample Problems and Solutions

Problem 1: A sound has the following spectrum: 440 Hz (amplitude 1.0), 880 Hz (0.5), 1320 Hz (0.3), 1760 Hz (0.2), 2200 Hz (0.1). Calculate the spectral centroid.

Solution:

$$C = \frac{440(1.0) + 880(0.5) + 1320(0.3) + 1760(0.2) + 2200(0.1)}{1.0 + 0.5 + 0.3 + 0.2 + 0.1}$$

$$C = \frac{440 + 440 + 396 + 352 + 220}{2.1} = \frac{1848}{2.1} = 880 \text{ Hz}$$

The spectral centroid is 880 Hz, which is the second harmonic. This indicates a moderately bright timbre.

Problem 2: For FM synthesis with $f_c = 600$ Hz, $f_m = 200$ Hz, and $I = 3$, calculate the frequencies of the first 5 sidebands on each side of the carrier.

Solution: Sidebands occur at $f_n = f_c + nf_m$:

$$\begin{aligned} f_{-5} &= 600 - 5(200) = -400 \text{ Hz (ignored)} \\ f_{-4} &= 600 - 4(200) = -200 \text{ Hz (ignored)} \\ f_{-3} &= 600 - 3(200) = 0 \text{ Hz (DC)} \\ f_{-2} &= 600 - 2(200) = 200 \text{ Hz} \\ f_{-1} &= 600 - 1(200) = 400 \text{ Hz} \\ f_0 &= 600 \text{ Hz (carrier)} \\ f_1 &= 600 + 1(200) = 800 \text{ Hz} \\ f_2 &= 600 + 2(200) = 1000 \text{ Hz} \\ f_3 &= 600 + 3(200) = 1200 \text{ Hz} \\ f_4 &= 600 + 4(200) = 1400 \text{ Hz} \\ f_5 &= 600 + 5(200) = 1600 \text{ Hz} \end{aligned}$$

Since $f_c : f_m = 3 : 1$ (integer ratio), the spectrum is harmonic with fundamental at 200 Hz.

C Conclusion

Unit 4 has explored the mathematical foundations of timbre, one of the most complex and fascinating aspects of musical sound. We have seen how Fourier analysis provides a powerful framework for understanding how complex sounds can be decomposed into simpler components, and how this decomposition relates to our perception of timbre.

Key insights include:

- Timbre is multidimensional, encompassing spectral content, temporal evolution, and attack characteristics
- Fourier analysis reveals the frequency content of sounds but has limitations for noise and transient signals
- The spectral envelope is more perceptually relevant than fine spectral details
- The source-filter model provides an elegant framework for understanding many sound-producing mechanisms
- Different synthesis methods offer various trade-offs between efficiency, control, and realism
- Attack transients are crucial for instrument identification
- Sub-harmonics and combination tones arise from nonlinear processes
- Time-varying spectra capture the dynamic nature of real musical sounds

The mathematical tools developed in this unit—Fourier series and transforms, spectral analysis, synthesis algorithms—are fundamental not only to understanding music but also to modern audio technology, from music production and synthesis to audio compression and speech recognition.

The study of timbre demonstrates how mathematics can illuminate artistic phenomena, revealing the hidden structures that give music its richness and emotional power. By understanding the mathematics of timbre, we gain both analytical tools for understanding existing sounds and creative tools for generating new ones.

References and Further Reading

1. Benson, D. J. (2006). *Music: A Mathematical Offering.* Cambridge University Press.

2. Loy, G. (2006). *Musimathics: The Mathematical Foundations of Music* (Vols. 1 & 2). MIT Press.

3. Fauvel, J., Flood, R., & Wilson, R. (Eds.). (2003). *Music and Mathematics: From Pythagoras to Fractals.* Oxford University Press.

4. Sethares, W. A. (2005). *Tuning, Timbre, Spectrum, Scale* (2nd ed.). Springer.

5. Roads, C. (1996). *The Computer Music Tutorial.* MIT Press.

6. Grey, J. M. (1977). *Multidimensional perceptual scaling of musical timbres.* Journal of the Acoustical Society of America, 61(5), 1270-1277.

7. Chowning, J. (1973). *The synthesis of complex audio spectra by means of frequency modulation.* Journal of the Audio Engineering Society, 21(7), 526-534.

8. Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and Models* (2nd ed.). Springer.