Theories & models of musical pitch
The search for the missing fundamental: theories & models of musical pitch

• Brief review of basics of sound & vibration
• Brief review of pitch phenomena
• Distortion theories (nonlinear processes produce F0 in the cochlea)
• Spectral pattern theories
  – Pattern-recognition/pattern-completion
  – Fletcher: frequency separation
  – The need for harmonic templates (Goldstein)
    Terhardt’s Virtual pitch: adding up the subharmonics
  – Musical pitch equivalence classes
  – Pitch classes and neural nets: Cohen & Grossberg
  – Learning pitch classes with connectionist nets: Bharucha
• Temporal theories
  – Residues: Beatings of unresolved harmonics (Schouten, 1940’s)
  – Problems with residues and envelopes
  – Temporal autocorrelation models (Licklider, 1951)
  – Interspike interval models (Moore, 1980)
  – Correlogram demonstration (Slaney & Lyon, Apple demo video)
  – Population-interval models (Meddis & Hewitt, Cariani & Delgutte)
• Problems & prospects
Vibrations create compressions and expansions of air

Sound waves are alternating local changes in pressure
These changes propagate through space as “longitudinal” waves

Condensation phase (compression):
pressure increases

Rarefaction (expansion) phase:
power pressure decreases
Waveforms

Microphones convert sound pressures to electrical voltages. Waveforms plot pressure as a function of time, i.e. a “time-series” of amplitudes. Waveforms are complete descriptions of sounds.

Audio CD’s sample sounds at 44,100 samples/sec.

Oscilloscope demonstration.
Most physical systems have multiple modes of vibrations that create resonances that favor particular sets of frequencies.

Vibrating strings or vibrating columns of air in enclosures exhibit harmonic resonance patterns.

Material structures that are struck (bells, xylophones, percussive instruments) have resonances that depend partly on their shape and therefore can produce frequencies that are not harmonically related.

More later on what this means for pitch and sound quality.
Frequency spectra

Joseph Fourier (1768-1830) showed that any waveform can be represented as the sum of many sinusoids (Fourier spectrum). George Ohm (1789-1854) and Hermann von Helmholtz (1821-1894) postulated that the ear analyzes sound analogously, first breaking sounds into their partials.

Each sinusoid of a particular frequency (frequency component, partial) has 2 parameters:

- 1) its magnitude (amplitude of the sinusoid)
- 2) its phase (relative starting time)

A sound with only one frequency component is called a pure tone. A sound with more than one is called a complex tone.

A complex tone whose partials are all part of the same harmonic series is called a harmonic complex. Such a tone is periodic -- its waveform repeats with a period equal to 1/fundamental frequency (i.e. the fundamental period). If any of the partials are not part of a harmonic series, then the sound is inharmonic.
Fundamentals and harmonics

- Periodic sounds (30-20kHz) produce pitch sensations.
- Periodic sounds consist of repeating time patterns.
- The fundamental period (F0) is the duration of the repeated pattern.
- The fundamental frequency is the repetition frequency of the pattern.
- In the Fourier domain, the frequency components of a periodic sound are all members of a harmonic series ($n = 1\times F0, 2\times F0, 3\times F0...$).
- The fundamental frequency is therefore the greatest common divisor of all of the component frequencies.
- The fundamental is also therefore a subharmonic of all component frequencies.
Demonstrations

- **Oscilloscope demonstrations**
  - Armenian flute
  - Human voice

- **Spectrum analyzer demonstrations**
  - Flute
  - Violins
  - Human voice

- **Real time spectrogram demonstrations** *(iTunes/SpectroGraph plugin)*
  - Armenian flute
  - Violins
  - Vocal music
Harmonic series

A harmonic series consists of integer multiples of a fundamental frequency, e.g. if the fundamental is 100 Hz, then the harmonic series is: 100, 200, 300, 400, 500, 600 Hz, .... etc.
The 100 Hz fundamental is the first harmonic, 200 Hz is the second harmonic. The fundamental is often denoted by F0.

The fundamental frequency is therefore the greatest common divisor of all the frequencies of the partials.

Harmonics above the fundamental constitute the overtone series.

Subharmonics are integer divisions of the fundamental:
e.g. for F0= 100 Hz, subharmonics are at 50, 33, 25, 20, 16.6 Hz etc. Subharmonics are also called undertones.

The fundamental period is 1/F0, e.g. for F0=100 Hz, it is 1/100 sec or 10 msec. Periods of the subharmonics are integer multiples of the fundamental period.
Synthetic vowels

Waveforms

Power Spectra

Autocorrelations

Formant-related
Vowel quality
Timbre

Pitch periods, $1/F_0$

[ae]
F0 = 100 Hz

[ae]
F0 = 125 Hz

[er]
F0 = 100 Hz

[er]
F0 = 125 Hz
Pitch: basic properties

• Highly precise percepts
  – Musical half step: 6% change F0
  – Minimum JND's: 0.2% at 1 kHz (20 usec time difference, comparable to ITD jnd)

• Highly robust percepts
  – Robust quality Salience is maintained at high stimulus intensities
  – Level invariant (pitch shifts < few % over 40 dB range)
  – Phase invariant (largely independent of phase spectrum, f < 2 kHz)

• Strong perceptual equivalence classes
  – Octave similarities are universally shared
  – Musical tonality (octaves, intervals, melodies) 30 Hz - 4 kHz

• Perceptual organization ("scene analysis")
  – Fusion: Common F0 is a powerful factor for grouping of frequency components

• Two mechanisms? Temporal (interval-based) & place (rate-based)
  – Temporal: predominates for periodicities < 4 kHz (level-independent, tonal)
  – Place: predominates for frequencies > 4 kHz (level-dependent, atonal)
Periodic sounds produce distinct pitches

Many different sounds produce the same pitches

Strong
• Pure tones
• Harmonic complexes
• Iterated noise

Weaker
• High harmonics
• Narrowband noise

Very weak
• AM noise
• Repeated noise

Schematic diagram representing eight signals with the same low pitch.
Pitch as a perceptual emergent

**Missing F0**

**Line spectra**

**Autocorrelation (positive part)**

Pure tone 200 Hz
Frequency ranges of (tonal) musical instruments
Duplex time-place representations
"Pitch is not simply frequency"

Musical tonality: octaves, intervals, melodies

Strong phase-locking (temporal information)

temporal representation
level-invariant, precise

place representation
level-dependent, coarse

Frequency (kHz)
Pitch height and pitch chroma

Please see Figure 1, 2, and 7 in Roger N. Shepard. "Geometrical Approximations to the Structure of Musical Pitch." *Psychological Review* 89 no. 4 (1982): 305-322.
Duplex time-place representations

**temporal representation**
- level-invariant
  - strong (low fc, low n)
  - weak (high fc, high n; F0 < 100 Hz)

**place-based representation**
- level-dependent
  - coarse

Similarity to place pattern vs. Similarity to interval pattern

cf. Terhardt's spectral and virtual pitch
A "two-mechanism" perspective (popular with some psychophysicists)

Unresolved harmonics
- Weak temporal mechanism
- Phase-dependent; first-order intervals

Resolved harmonics
- Strong spectral pattern mechanism
- Phase-independent
- Rate-place? interval-place?

Place-based representations
- Level-dependent
- Coarse

Dominance region
- \( f, F_0 \)
- \( n = 5-10 \)
Some possible auditory representations

Local
Rate-place

Masking phenomena
Loudness

Central spectrum

CF

Synchonry-place
Phase-place
Interval-place

Pure tone pitch JNDs: Goldstein

Central spectrum

CF

Population interval

Complex tone pitch

Population-interval

Interval

1/Fo

Stages of integration

All-at-once
General theories of pitch

1. Distortion theories
   – reintroduce F0 as a cochlear distortion component (Helmholtz)
   – sound delivery equipment can reintroduce F0 through distortion
   – however, masking F0 region does not mask the low pitch (Licklider)
   – low pitch thresholds and growth of salience with level not consistent with distortion processes (Plomp, Small)
   – binaurally-created pitches exist

2. Spectral pattern theories
   – Operate in frequency domain
   – Recognize harmonic relations on resolved components

3. Temporal pattern theories
   – Operate in time domain
   – Analyze interspike interval dists.
Psychological perspectives on pitch

Analytical: break sounds into frequencies (peceptual atoms), then analyze patterns (templates) (British empiricism; machine perception)

Relational: extract invariant relations from patterns (Gestaltists, Gibsonians, temporal models)

Nativist/rationalist: mechanisms for pitch are given by innate knowledge and/or computational mechanisms differences re: how recently evolved these are

Associationist: mechanisms for pitch (e.g. templates) must be acquired through experience (ontogeny, culture)

Interactionist: (Piaget) interaction between native faculties and structure of experience (self-organizing systems)
Spectral pattern theories

- Not the lowest harmonic
- Not simple harmonic spacings
- Not waveform envelope or peak-picking (pitch shift exps by Schouten & de Boer)
- Must do a real harmonic analysis of spectral fine structure to find common denominator, which is the fundamental frequency
- Terhardt: find common subharmonics
- Wightman: autocorrelation of spectra
- Goldstein, Houtsma: match spectral excitation pattern to harmonic templates
- SPINET: Use lateral inhibition/center-surround then fixed neural net to generate equivalence classes
- Barucha: adaptive connectionist networks for forming harmonic associations
Spectral pattern analysis vs. temporal pattern analysis

Note: Some models, such as Goldstein's, use interspike interval information to first form a Central Spectrum which is then analyzed using harmonic spectral templates.

There are thus dichotomies 1) between use of time and place information as the basis of the central representation, and 2) use of spectral vs. autocorrelation-like central representations.
Traveling waves along the cochlea. A traveling wave is shown at a given instant along the cochlea, which has been uncoiled for clarity. The graphs profile the amplitude of the traveling wave along the basilar membrane for different frequencies, and show that the position where the traveling wave reaches its maximum amplitude varies directly with the frequency of stimulation. (Figures adapted from Dallos, 1992 and von Bekesy, 1960)
"Virtual" pitch: F0-pitch as pattern completion
From masking patterns to "auditory filters" as a model of hearing

Power spectrum
Filter metaphor

Notion of one central spectrum that subserves perception of pitch, timbre, and loudness

2.2. **Excitation pattern** Using the filter shapes and bandwidths derived from masking experiments we can produce the excitation pattern produced by a sound. The excitation pattern shows how much energy comes through each filter in a bank of auditory filters. It is analogous to the pattern of vibration on the basilar membrane. For a 1000 Hz pure tone the excitation pattern for a normal and for a SNHL listener look like this: The excitation pattern to a complex tone is simply the sum of the patterns to the sine waves that make up the complex tone (since the model is a linear one). We can hear out a tone at a particular frequency in a mixture if there is a clear peak in the excitation pattern at that frequency. Since people suffering from SNHL have broader auditory filters their excitation patterns do not have such clear peaks. Sounds mask each other more, and so they have difficulty hearing sounds (such as speech) in noise. --Chris Darwin, U. Sussex, http://www.biols.susx.ac.uk/home/Chris_Darwin/Perception/Lecture_Notes/Hearing3/hearing3.html
Shapes of perceptually-derived "auditory filters" (Moore)
Resolution of harmonics

Threshold Shift (dB)

0 1000 2000 3000 4000 Hz
Goldstein’s harmonic templates


Terhardt's method of common subharmonics

Spectral vs. virtual pitch: duplex model

Virtual pitch computation:
1. Identify frequency components
2. Find common subharmonics
3. Strongest common subharmonic after F0 weighting is the virtual pitch

Terhardt's model has been extended by Parncutt to cover pitch multiplicity and fundamental bass of chords
Terhardt references

Neural tuning as a function of CF
Broad tuning and rate saturation at moderate levels in low-CF auditory nerve fibers confounds rate-based resolution of harmonics.

Low SR auditory nerve fiber

![Graph showing the frequency response of low SR auditory nerve fibers under different stimulus levels. The graph displays the number of spikes per second (Spikes/s) in response to various frequencies (Frequency in Hz) at different sound pressures (dB). The baseline frequency (B.F.) is 1700 Hz. The graph includes markers indicating the level of stimulus (e.g., 25 dB, 35 dB, 45 dB, 55 dB, 65 dB, 75 dB, 85 dB, 95 dB) and whether the stimulus is synchronized (Spon.) or not (No sync.).]
Spectral pattern theories - pros & cons

Do make use of frequency tuning properties of elements in the auditory system
No clear neural evidence of narrow (< 1/3 octave) frequency channels in low-BF regions (< 2 kHz)

Operate on perceptually-resolved harmonics
Do not explain low pitches of unresolved harmonics

Require templates or harmonic pattern analyzers
Little or no neural evidence for required analyzers
Problems w. templates: relative nature of pitch

Do not explain well existence region for F0
Learning theories don't account for F0 ranges or for phylogenetic ubiquity of periodicity pitch
Existence region for missing fundamentals of AM tones
ANFs

Tuning curves

Stimulus waveform

voice pitch
Fundamental period 1/F0

Cochlea

Rate threshold level (dB SPL)

Characteristic frequency (kHz)

Stimulus

Peristimulus time histograms
(100 presentations @ 60 dB SPL)
Temporal pattern theories

Σ First-order intervals (renewal density)

- Schouten’s temporal theory (1940’s) depended on interactions between unresolved (high) harmonics. It was displaced by discovery of dominance region and binaural combination pitches in the 1960’s. The idea persists, however in the form of spectral mechanisms for resolved harmonics and temporal ones for unresolved harmonics.


Σ All-order intervals (temporal autocorrelation)


Interval-based theories of pitch

First-order intervals
(renewal density)


All-order intervals
(temporal autocorrelation)

Licklider’s (1951) duplex model of pitch perception

Licklider’s binaural triplex model

J.C.R. Licklider
“Three Auditory Theories”

Fig. 2. – Schematic diagram of overall analyzer.
Fig. 2. – Schematic diagram of overall analyzer.
Delay lines

Basic schema of neuronal autocorrelator. A is the input neuron, B1, B2, B3, .... is a delay chain.
Autocorrelation and interspike intervals

**Autocorrelation functions**

\[ \text{Corr}(\tau) = \sum_t S(t) S(t - \tau) \]

- **Fundamental period** $1/F_0$
- **Time lag** $\tau$

Shift
Multiply
Sum the products
for each delay $\tau$
to compute
autocorrelation
function

**Autocorrelations of spike trains** = **Histograms of all-order intervals**

00000100000000100000000000000001000000000010000000
00000100000000100000000000000000000000000000000000000
Correlograms: interval-place displays (Slaney & Lyon)

Correlograms

Auditory nerve
Temporal coding in the auditory nerve

Work with Bertrand Delgutte
Cariani & Delgutte (1996ab)

Dial-anesthetized cats.
100 presentations/fiber
60 dB SPL

Population-interval distributions are compiled by summing together intervals from all auditory nerve fibers.

The most common intervals present in the auditory nerve are invariably related to the pitches heard at the fundamentals of harmonic complexes.
The population-interval distribution of the auditory nerve
Stimulus Autocorrelation

Pitch = 1/F0

All-order interval histograms

Characteristic frequency (kHz)

Population-interval histogram

Autocorrelation functions

Fundamental period = 1/F0

Corr(τ) = \sum_{t} S(t) S(t - τ)

Shift
Multiply
Sum the products for each delay τ to compute autocorrelation function

Autocorrelations of spike trains = Histograms of all-order intervals
Percept-driven search for neural codes:
1. Use stimuli that produce equivalent percepts
2. Look for commonalities in neural response
3. Eliminate those aspects that are not invariant

Metameric stimuli (same percept, different power spectra)

Stimulus A (AM tone, fm = 200 Hz)

Stimulus B (Click train, F0 = 200 Hz)

Neural response pattern

Neural codes, representations: those aspects of the neural response that play a functional role in subserving the perceptual distinction

Candidate "neural codes" or representations

Common aspects of neural response that covary with the percept of interest

Intensity

Location

Duration

Other parameters for which percept is invariant

Candidate "neural codes" or representations
Pitch equivalence

Six stimuli that produce a low pitch at 160 Hz

<table>
<thead>
<tr>
<th>Waveform</th>
<th>Power spectrum</th>
<th>Autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure tone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>160 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM tone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_m: 160 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_c: 640 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harms 6-12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_0: 160 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Click train</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_0: 160 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM noise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_m: 160 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM tone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_m: 160 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_c: 6400 Hz</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Population interval distribution

Pitch equivalence

Pitch frequency

Time (ms)

Frequency (Hz)

Lag (ms)

Interspike interval (ms)

Pitch period

Six stimuli that produce a low pitch at 160 Hz

- Pure tone 160 Hz
- AM tone F_m:160 Hz F_c:640 Hz
- Harms 6-12 F_0: 160 Hz
- Click train F_0: 160 Hz
- AM noise F_m: 160 Hz
- AM tone F_m: 160 Hz F_c: 6400 Hz
Level-Invariance Pitch Equivalence

Neural Data

E  Pure Tone  Fc=640 Hz, Fm=160 Hz

F  AM Tone  F=160 Hz

G  Click Train  F0=160 Hz

H  AM Noise  Fm=160 Hz

Pitch Equivalence
**A**

**AM Tone**
- $F_c = 640 \text{ Hz}$
- $F_m = 160 \text{ Hz}$

**Center frequency**
- $F_c - F_m$
- $F_c$
- $F_c + F_m$

**Frequency channels**

- **Half-wave rectification + LP filtering**

- **Channel autocorrelations**

**B**

- $>> F_c$

**C**

**AM Tone**
- $F_c = 6400 \text{ Hz}$
- $F_m = 160 \text{ Hz}$

- $\sim F_c$

**Time (ms)**

- **Time (ms)**
- **Time (ms)**
- **Interval (ms)**
The running population-interval distribution

Population interval histograms (cross sections)

Interval (ms)

# intervals

0 5 10 15

A

B
Pitch height and pitch chroma

Please see Figure 1, 2, and 7 in Roger N. Shepard. "Geometrical Approximations to the Structure of Musical Pitch." *Psychological Review* 89 no. 4 (1982): 305-322.
Pitch shift of inharmonic complex tones

$F_m = 125 \text{ Hz}$
$F_c = 750 \text{ Hz}$

$n = 6$

$n = 5.86$

$n = 5.5$

$n = 5$

Population interval distributions

Stimulus waveform

Stimulus autocorrelation

Interval (ms)
Pitch shift of inharmonic complex tones
Phase-invariant nature of all-order interval code

AM Tone
Fc=640 Hz, Fm=200 Hz

B
Fc=640 Hz, Fm=200 Hz

QFM Tone
F_{c}=640 \text{ Hz}, F_{m}=160-320 \text{ Hz}

D
F_{c}=640 \text{ Hz}, F_{m}=160-320 \text{ Hz}

de Boer's rule
1/Fm
AM Tone

F_c=640 Hz, F_m=200 Hz

QFM Tone

F_c=640 Hz, F_m=200 Hz

E AM  F_c=640 Hz, F_m=320 Hz
PST=225-275

F QFM  F_c=640 Hz, F_m=320 Hz
PST=225-275

G AM  F_c=640 Hz, F_m=256 Hz
PST=335-385

H QFM  F_c=640 Hz, F_m=256 Hz
PST=335-385
Cochlear nucleus IV: Pitch shift

**Variable-Fc AM tone**  

**F_m = 125 Hz**  

**F_c = 500-750 Hz**  

Pitch ~ de Boer's rule

---

Pooled ANF (n=47)

- 500
- 750
- 500

Interval (ms)

- 0
- 100
- 200
- 300
- 400
- 500

Peristimulus time (ms)

---

Chop-S (PVCN)

- 45-15-8 CF: 4417 Hz
- Thr: -18, SR: 39 s/s

Pauser (DCN)

- 45-15-4 CF: 408 Hz
- Thr: 21.3, SR: 159

---

Primarylike (AVCN)

- 45-17-4 CF: 408 Hz
- Thr: 21.3, SR: 159

---

- 1/F_m
- de Boer's rule (pitches)

- sub-harmonics of Fc
- 1/Fc
Dominance region for pitch (harmonics 3-5 or partials 500-1500 Hz)

- \(F_{0_{3-5}} = 80 \text{ Hz}\)
- \(F_{0_{3-5}} = 160 \text{ Hz}\)
- \(F_{0_{3-5}} = 240 \text{ Hz}\)
- \(F_{0_{3-5}} = 320 \text{ Hz}\)
- \(F_{0_{3-5}} = 480 \text{ Hz}\)

Harmonics 3-5 alone

Harmonics 6-12 alone

Harmonics 3-5 and 6-12 together

Interval (ms)
Summary

Population-interval representation of pitch at the level of the auditory nerve

Pitch of the "missing fundamental"

Pitch Equivalence

Level invariance

Pitch shift of inharmonic AM tones

Phase invariance

Dominance region

Pitch salience

Summary
Temporal theories - pros & cons

Make use of spike-timing properties of elements in early processing (to midbrain at least)
Interval-information is precise & robust & level-insensitive
No strong neurally-grounded theory of how this information is used

Unified model: account for pitches of perceptually-resolved & unresolved harmonics in an elegant way (dominant periodicity)
Explain well existence region for F0 (albeit with limits on max interval durations)
Do not explain low pitches of unresolved harmonics

Interval analyzers require precise delays & short coincidence windows
Little or no neural evidence for required analyzers
Physiological and functional representations

Rate-place profiles

Interval-place Spatiotemporal pattern (Place & time)

Global interval

Central spectrum Frequency-domain representation

Central autocorrelation Time-domain representation
Different representations can support analogous strategies for pitch extraction, recognition, and comparison

- **Central spectrum**
  - Frequency-domain representation
  - Explicit identification of individual harmonics & deduction of F0 via common subharmonics or pattern recognition

- **Central autocorrelation**
  - Time-domain representation
  - Template-based global recognition of pitch-related patterns (neural networks, harmonic templates, interval sieves)

- **Relative pitch comparison mechanisms**
  - (matching, octaves, musical intervals)
  - Spectral patterns
  - Temporal patterns
Reading/assignment for next meeting

- Tuesday. Feb. 24
- Pitch mechanisms, continued
- Perception of timbre

**Reading:**

Moore Chapter 8, pp. 269-273
Chapter in Deutsch by Risset & Wessel on Timbre
(first sections up to p. 113-118)
Handel, chapter on Identification
Pitch classes and perceptual similarity

Build up harmonic associations from repeated exposure to harmonic complex tones

Harmonic similarity relations are direct consequences of the inherent structure of interval codes
Figure 4. Similarities between population-interval representations associated with different fundamental frequencies. Simulated population-interval distributions for pure tones (left) and complex tones (right) consisting of harmonics 1-6.
Octave similarity

Pitch height and pitch chroma

Please see Figure 1, 2, and 7, in Roger N. Shepard. "Geometrical Approximations to the Structure of Musical Pitch." *Psychological Review* 89 no. 4 (1982): 305-322.

Musical tonal relations
