

Pitch Perception and Grouping

HST.723 Neural Coding and
Perception of Sound

Pitch Perception. I. Pure Tones

The pitch of a pure tone is strongly related to the tone's frequency, although there are small effects of level and masking.

<1000 Hz: increased level: decreased pitch

1000-2000 Hz: little or no change

>2000 Hz: increased level: increased pitch

Difference Limens for Frequency (DLF)

The auditory system is exquisitely sensitive to changes in frequency (e.g. 2-3 Hz at 1000 Hz = 0.01 dB).

How is frequency coded - Place or timing?

Zwicker's proposal for FM detection.
(See Moore, 1997)

Place

Pros: Could in principle be used at all frequencies.

Cons: Peak of BM traveling wave shifts basally with level by $\frac{1}{2}$ octave – no similar pitch shift is seen; fails to account for poorer performance in DLFs at very high frequencies (> 4 kHz), although does a reasonable job of predicting frequency-modulation difference limens (FMDLs).

Temporal cues

Timing

Pros: Pitch estimate is basically level-invariant; may explain the absence of musical pitch above ca. 4-5 kHz.

Cons: Thought to break down totally above about 4 kHz (although some “optimal detector” models predict residual performance up to 8 or 10 kHz); harder to explain diplacusis (differences in pitch perception between the ears).

See Rose et al. (1971)

Musical pitch

Musical pitch is probably at least 2-dimensional:

- *Tone height*: monotonically related to frequency
- *Tone chroma*: related to pitch class (note name)

Circularity in pitch judgments: changes in chroma but no change in height. In circular pitch is a half-octave interval perceived as going up or down? (Deutsch, 1986)

Musical pitch of pure tones breaks down above about 5 kHz: octave matches become erratic and melodies are no longer recognized. Differences in frequency are still detected – only tone chroma is absent.

- Further evidence for the influence of temporal coding?

Pitch Perception in Complex Tones

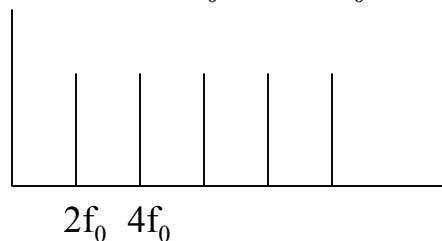
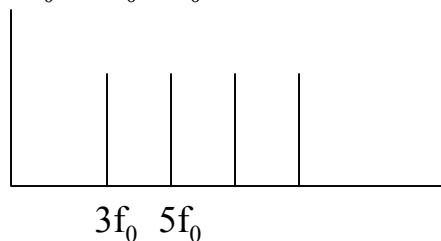
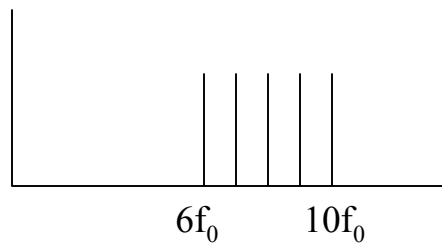
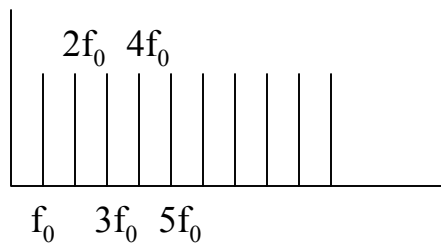
Many complex sounds are harmonic, i.e. consist of components at integer multiples of the fundamental frequency (f_0). They generally have a pitch corresponding to the f_0 .

Place Theory (Ohm, 1843; Helmholtz, 1863)

Lowest component of a complex determines its pitch. Other components important only for timbre.

BUT: Pitch does not change when the fundamental is not physically present or is masked (Seebeck, 1841; Schouten, 1940; Licklider, 1956).

What pitches do these complexes elicit?



(pitch of this depends on f_0 ;
Flanagan and Guttman)

Demonstration: ASA CD Tracks 40-42

Pitch ambiguities

A complex comprising 800, 1000, and 1200 Hz has a pitch of 200 Hz.

What about 850, 1050, and 1250 Hz?
(ASA CD track 38)

Pitch salience becomes weaker when only higher harmonics are present. (ASA CD track 43-45). Why is this?

Resolved vs. unresolved harmonics

(See Moore, 1997 and Plomp, 1964)

Mechanisms of Complex Pitch Perception

Temporal Theory (Schouten, 1940):

Pitch is extracted from the summed waveform of adjacent components. This requires that some components interact.

Pattern Recognition Theory (e.g. Goldstein, 1973):

The frequencies of individual components are determined and the “best-fitting” f_0 is selected. This requires that some components remain resolved *and* that some form of “harmonic template” exists.

Evidence against a “pure” temporal model

- Pitch sensation is strongest for low-order (resolved) harmonics (Plomp, 1964; Ritsma, 1964).
- Pitch can be elicited by only two components, one in each ear (Houtsma and Goldstein, 1972).
- Pitch can be elicited by consecutively presented harmonics (Hall and Peters, 1981).

Evidence against a “pure” pattern recognition theory

- Very high, unresolved harmonics can still produce a (weaker) pitch sensation
- Aperiodic, sinusoidally amplitude-modulated (SAM) white noise can produce a pitch sensation (Burns and Viemeister, 1976; 1981).

Autocorrelation model of pitch perception

- Based on an original proposal by Licklider (1951).
- The stimulus within each frequency channel is correlated (delayed, multiplied and averaged) with itself (through delay lines).
- This produces peaks at time intervals corresponding to multiples of the stimulus period.
- Pooling interval histograms across frequency produces an overall estimate of the “dominant” interval, which generally corresponds to the fundamental frequency.

(See Meddis and Hewitt, 1991)

Autocorrelation model

Pros:

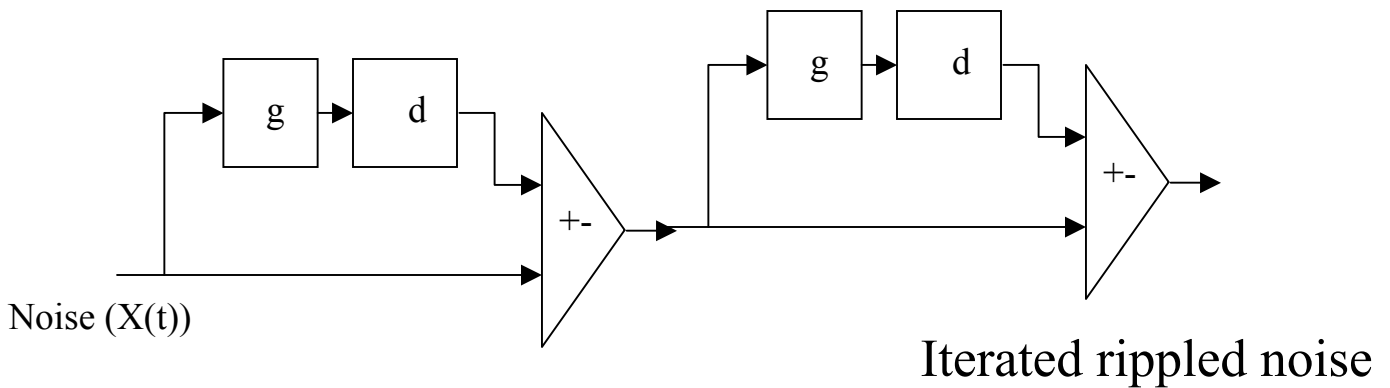
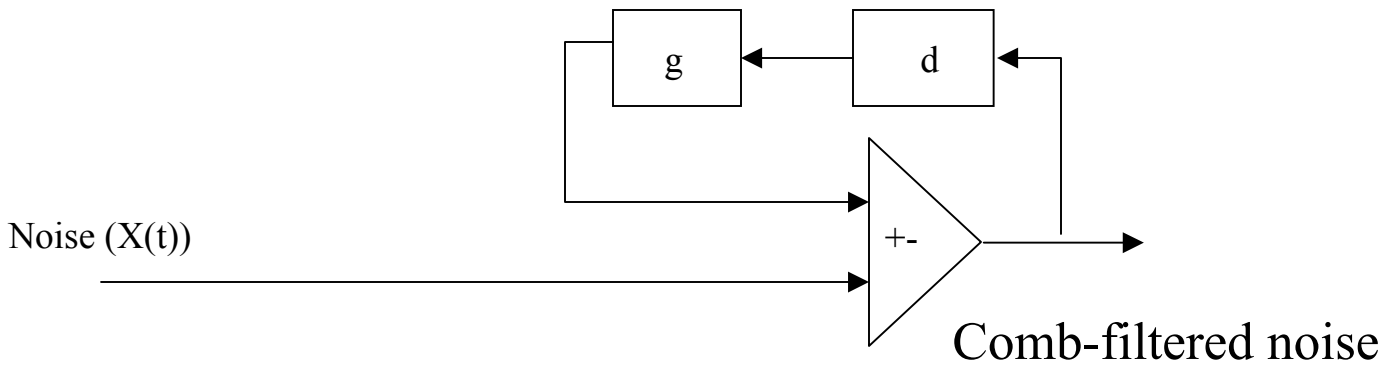
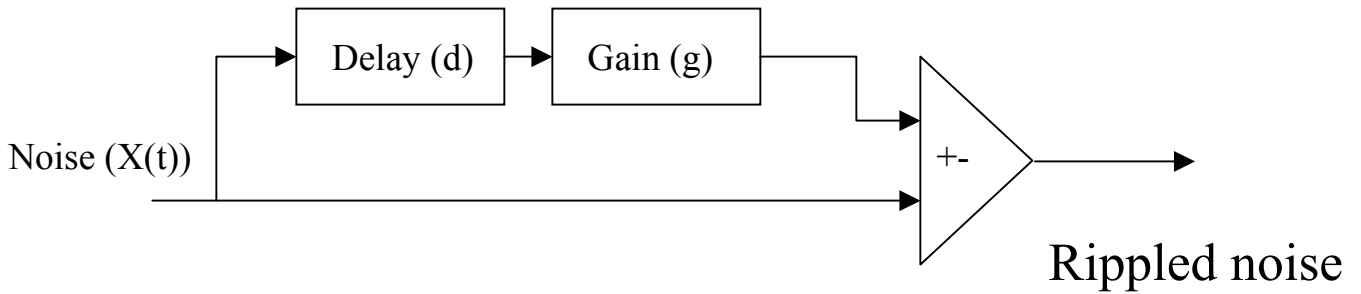
- Model can deal with both resolved and unresolved harmonics
- Predicts no effect of phase for resolved harmonics, but strong phase effects for unresolved harmonics, in line with data (Meddis & Hewitt, 1991).
- Predicts a dominance region of pitch, roughly in line with early psychophysical data, due to reduction in phase locking with frequency.

Cons:

- Deals *too well* with unresolved harmonics – predicts no difference based on resolvability, in contrast to psychophysical data (Carlyon and Shackleton, 1994).
- Dominance region based on absolute, not relative, frequency, in contrast to data.

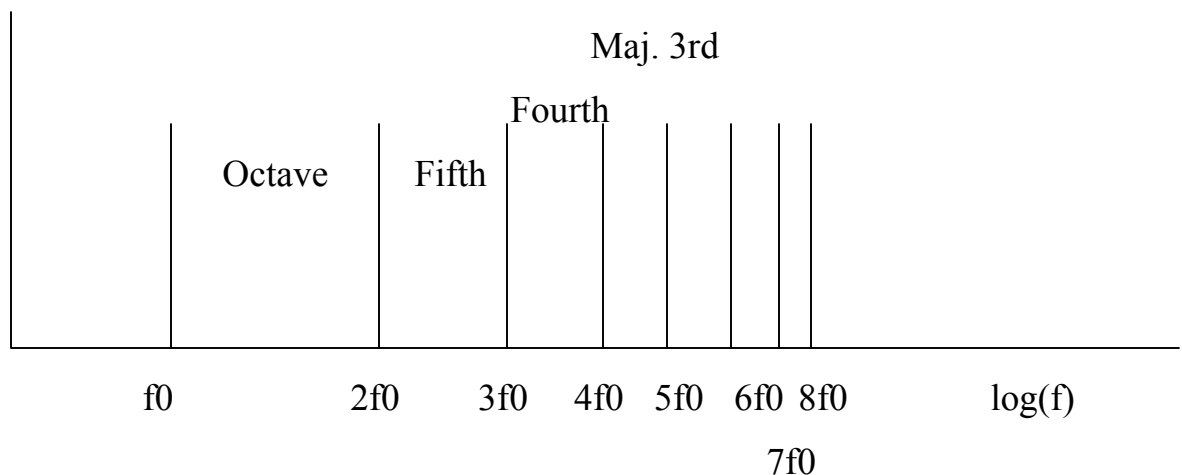
[N.B. The “template” model of Shamma and Klein (2000) involves place and timing coding, but not in the traditional sense.]

“Regular Interval Noise”



Musical intervals: Consonance and Dissonance

- In the West, the equal-tempered scale has been adopted, with the octave split into twelve equal steps on a log scale, i.e., 4 semitones higher is $2^{4/12}$ times higher in frequency.
- This is a compromise: the intervals in the harmonic series only approximate the notes of the scale.



- Perceived dissonance is in part due to beating effects between neighboring harmonics. Remaining effect of perceived consonance and dissonance may be simply cultural.

Auditory Grouping and Pitch

Simultaneous, harmonically related tones tend to form a single auditory object, which makes ecological sense.

What happens if one component is slightly out of tune?

Harmonicity can be a strong cue in binding components together, but it can be overridden by competing cues or expectations (Darwin et al., 1994).

A mistuned harmonic can be “heard out” more easily, but can still contribute to the overall pitch of the complex. This is an example of “duplex perception”.

Demonstrations of Simultaneous Grouping Principles

Demo 1: Mistuning of a single component. (Track 18)

Demo 2: Two groups of harmonics: Effect of “common fate” or inharmonicity? (Track 19)

Demo 3: Simultaneous vs. sequential streaming and effects of onset disparity. (Track 22).

Recognizing Sounds

Timbre is:

“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.”

(ANSI, 1960)

Effect of Spectrum on Timbre

Example 1:

Hemony clarion bell

Example 2:

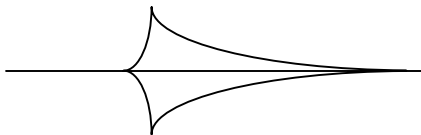
Guitar tone

Traditional view: Spectrum dominant in determining timbre; auditory system relatively insensitive to phase.

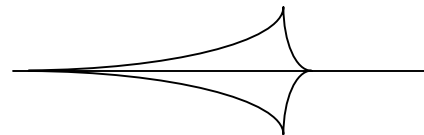
However, most sounds can be easily recognized despite very large spectral distortions, due to room acoustics or poor-quality electroacoustic transmission.

Effect of Temporal Envelope on Timbre

Piano tone



Reversed piano tone



Time

Damped and ramped sinusoids also have very different sound qualities (Patterson, 1994). Neural correlates are currently a subject of research (e.g., Lu et al., 2001).

Analysis and resynthesis has revealed the importance of dynamic spectrotemporal profile, as well as the role of inharmonic tones in musical instruments. Spectral and temporal properties of most instruments change with F0 and level, making realistic synthesis very difficult.

Sequential Stream Segregation

Pure tones

Demo 1: Stream segregation depends on frequency and repetition rate. (Track 1)

Demo 2: Segregation allows the recognition of a melody within one stream. (Track 5)

Demo 3: Segregation implies loss of rhythmic information across streams. (Track 3)

Sequential Segregation of Complex Tones

Peripheral channeling (tonotopic), or higher level “object” recognition?

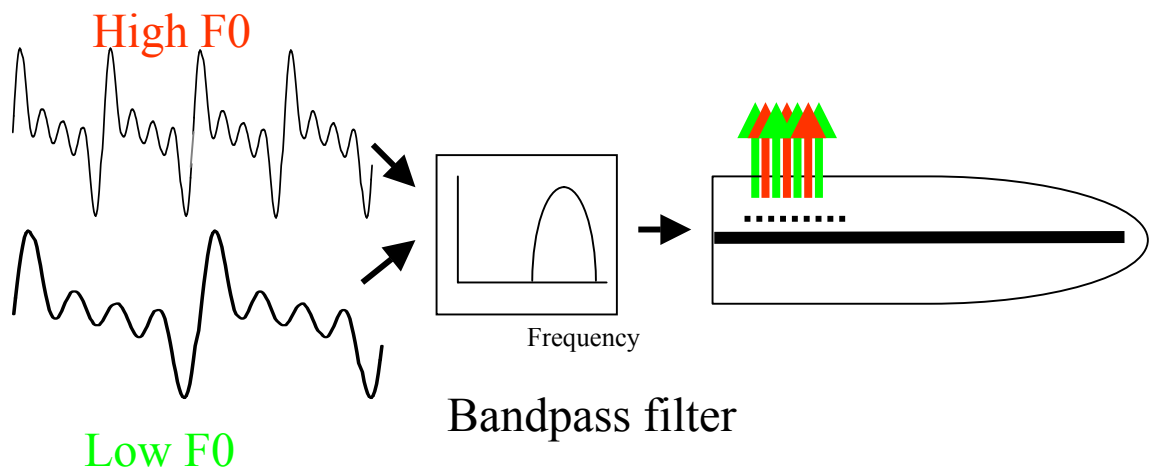
Van Noorden (1975): segregation of complex tones with the same F0 but different harmonics: Evidence for role of peripheral filtering?

Hartmann and Johnson (1991): Interleaved melodies. Varied timbre and lateralization:

Results suggest dominance of peripheral channeling. However, higher level processes are clearly involved (repetition rate, build-up effects, etc.)

Periodicity vs. Spectrum

Vliegen and Oxenham (1999) measured streaming using unresolved harmonic complexes with different F0s.



No difference in streaming ability found for resolved or unresolved harmonics:

Differences in spectrum are sufficient but not *necessary* for streaming.

Results not compatible with peripheral channeling hypothesis.