Computational Perception
15-485/785

Auditory Structure 1
Auditory Representations of Pitch and Timbre

Lyon and Shamma (1996)

- “The prime function of the auditory system is to tell the organism about its world by detecting, locating, and classifying sound sources.”

- “In a complex sound environment, deciding what sound fragments should be grouped and identified as a single sources is perhaps the most difficult and important low-level processing task.”

What do these have to do with pitch and timbre?

- “A good theory of pitch and timbre should related to properties of natural sound sources of importance to an organism, and a good computational model of pitch and timbre should related to known physiology of the auditory nervous system.”

How are pitch and timbre relevant to animal auditory systems?
Computational goals in audition

- orient to sounds
- identify sounds
- communication

What makes these hard in a natural environment

- multiple sounds
- background noise
- acoustical environment

What is sound structure?
Inner Ear

- malleus
- incus
- stapes
- tympanic cavity
- oval window
- semicircular canal
- scala media
- cochlear sac
- scala vestibuli
- helicotrema
- vestibule
- round window
- spiral lamina and basilar membrane
- Reissner's membrane
Auditory Nerve Fiber Responses: Frequency tuning

![Graph showing frequency tuning of auditory nerve fibers.](image)
Phase Locking

A
0.408 kHz
72/sec

B
0.85 kHz
179/sec

C
1.0 kHz
182/sec

D
1.5 kHz
234/sec

E
2.0 kHz
178/sec

F
2.3 kHz
83/sec

Duration of Interval (ms)

Number of Intervals
What auditory structure should be represented?

Conventional spectrograms of speech:

- upper time window: 20Hz
- lower time window: 200Hz
At a coarse temporal scale, cochleagrams and spectrograms look similar except for the scale of the frequency axis.
The plot shows the cochleagram of 'rea' from 'greasy'. This code preserves much of the fine time scale structure of each sound component, e.g. the glottal pulses. This representation allows use of correlograms.
Spike code representation of glottal pulses

 Spikegram of the "eal" part of "wealth".
Higher order auditory responses

Responses are specific to cell types.
Higher order auditory responses

Example peristimulus time histograms.
Idealized PSTHs

PST histograms to tone bursts

- Acoustic nerve
- Pauser
- On
- Chopper
- Off
- Tonic
- On-off
- Inhibitory

Number of discharges

Tone burst

Time ————————

Tone burst
Reverberation in the ideal case

- Direct sound
- First reflections
- Reverberation
Reverberation in a room

Sound travel time to wall removed.
Affect of reverberation on speech signal

Top is original, clean signal.
Bottom is signal convolved with room IRF.

“two oh six”
Initial reflections contribute to sense of room character

- zero point (first reflection) omitted
- initial part consists of discrete echoes
- later part is more continuous
- echoes can enhance intelligibility.

How?
- best concert halls have initial time-delay gap of 15-30 ms.

How to deal with reverberation?

- Use better, more directional, microphones to minimize energy from reflections
- use microphone arrays (e.g. in conference rooms) to do noise (i.e. echo) cancellation
Simulating room acoustics

a

receiver

source

b
The precedence effect

What will happen perceptually as the delay is varied?
The precedence effect

Lateralization shifts from 0 to $\sim$1ms.
Only first impulse is perceived from 1ms to $\sim$30ms
Beyond that two impulses are perceived separately (echo)
The precedence effect depends on relative sound pressure

Continuous speech (normal speed = 5 syllables/s)

Level of the primary sound ≈ 50 dB

$L_{ST} - L_{SO}$ is level difference between echo and primary auditory event.
The precedence effect depends on impulse width

Damaske’s theory that echo is detectable whenever curvature of total loudness exceeds value (i.e. an acoustic edge).

Does this experimental paradigm capture reverberation?
Using binaural inputs to reduce effects of reverberation

- Cross-correlation at one position (i.e. delay) inhibits activity at other positions.
- reduces effects of echos and reverberations.
- does not allow for frequency dependent level differences

Extension of Lindemann model: Gaik (1993)

Weightings are adjusted according to interaural transfer function.

Limitations:

- contains a number of parameters that must be set by hand (e.g. duration of cross channel inhibition)
- For ongoing signals like speech and music in reverberant environments can have problems distinguishing direct sound from reflections
- These problems compound when multiple sounds are active
Pattern generated from a binaural impulse response recorded in Grosser Musikvereinsaal in Vienna for critical band 630 (770 Hz).