Pitch perception in auditory scenes
Papers on pitch perception…

• of a single sound source
  
  LOTS - too many?

• of more than one sound source

  Almost none ± a few
Need for sound segregation

• Ears receive mixture of sounds

• We hear each sound source as having its own appropriate timbre, pitch, location

• Stored information about sounds concerns a single source
Bach: Musical Offering (strings)
Talk outline

• Low-level grouping cues in pitch perception
  – harmonic structure
    • Conjoint or disjoint allocation
  – onset-time
  – context
  – spatial cues.

• Use of a difference in harmonic structure between two sound sources to help identify, track over time, and localise simultaneous sound sources.
Excitation pattern of complex tone on basilar membrane

- Unresolved
- Resolved

Log (ish) frequency

Output of 1600 Hz filter

Output of 200 Hz filter

1/200s = 5ms
Mistuned harmonic’s contribution to pitch declines as Gaussian function of mistuning


\[ \Delta F_0 = a - k \Delta f \exp(-\Delta f^2 / 2s^2) \]

Harmonic Sieve

- Only consider frequencies that are close enough to harmonic. Useful as front-end to a Goldstein-type model of pitch perception.

Duifhuis, Willems & Sluyter *JASA* (1982).
Is “harmonic sieve” necessary with autocorrelation models?

- Autocorrelation could in principle explain mistuning effect
  - mistuned harmonic initially shifts autocorrelation peak
  - then produces its own peak
- But the numbers do not work out.
  - Meddis & Hewitt model is too tolerant of mistuning.
Disjoint allocation?

Will the 200-Hz series grab the 600-Hz component, and so make it give less of a pitch shift to the 155-Hz series?

Disjoint allocation… (2)

- No evidence of 600-Hz component being grabbed by "in-tune" 200-Hz complex
Pitch determined by conjoint allocation

• Decision on each simultaneous pitch taken independently.
Mistuning & pitch

![Graph showing the mean pitch shift (Hz) with respect to % Mistuning of 4th Harmonic for 8 subjects. The graph compares vowel and complex sounds, with error bars indicating variability. The x-axis represents % Mistuning of 4th Harmonic ranging from 0 to 8, and the y-axis represents Mean pitch shift (Hz) ranging from -0.2 to 1.0. There is a label indicating 90 ms.]
Onset asynchrony & pitch

Mean pitch shift (Hz)

±3% mistuning
8 subjects

- Onset Asynchrony T (ms)

vowel
complex

90 ms
Adaptation?

Increases effect of mistuning
Use of onset-tme

- A long onset-time removes a harmonic from pitch calculation.
- Not just due to adaptation because of regrouping experiment
- NB Onset-time much longer for pitch (100s of ms) than for hearing sound out as separate harmonic or for timbre judgements of complex (10s of ms).
Repeated context and pitch

Pitch perception is not a bacon slicer

- Pitch perception does not operate on independent time-slices
- It uses, for example, the “Old + New” heuristic to parse which components are temporally relevant
These conclusions apply to resolved harmonics

Separate simultaneous pitches only audible with unresolved harmonics when each pitch comes predominantly from a separate frequency region.

Problem for autocorrelation models

And the Old+New Heuristic breaks down. No advantage for onset-time.
Superimposing unresolved harmonics

2kHz - 3kHz

Fo = 100 Hz

Fo = 125 Hz

sum

Spatial cues?

- Contralateral mistuned component contributes almost as much as ipsilateral (k= .073 vs .061)

Cf Beerends and Houtsma *JASA* (1989).

\[
\Delta F_0 = a - k \Delta f \exp\left(-\frac{\Delta f^2}{2s^2}\right)
\]

-  
  \(s = 17.4\)
  \(k = 0.061\)

-  
  \(s = 19.9\)
  \(k = 0.073\)
Summary of constraints for pitch

- Harmonic (Gaussian s.d. 3%)
- Onset time (~100-200 ms)
- Repeated Context
- V. weak spatial constraint
- Independent estimates of different simultaneous F0s (conjoint allocation)
- N.B. These constraints different for unresolved harmonics and for e.g. timbre / vowel perception
Using pitch differences to separate different objects

- Helps intelligibility of simultaneous speech
- Helps to track one voice in presence of others
- Helps to separate objects for independent localisation
ΔFo between two sentences
(Bird & Darwin 1998; after Brokx & Nooteboom, 1982)

Two sentences (same talker)
• only voiced consonants
• (with very few stops)
Thus maximising Fo effect

Task: write down target sentence

Replicates & extends Brokx & Nooteboom
Example of using harmonicity for grouping for another property

How do we localise complex sounds when more than one sound source is present?

Maybe we group sounds first and then pool the localisation estimates of the component frequencies for that sound. More stable than grouping by location?
Localisation by ITD:
Jeffress / Trahiotis & Stern

500 Hz Narrow-band noise
Right ear leads by 1.5 ms

Hear on Left
(phase ambiguity)

500 Hz Wider-band noise
Right ear leads by 1.5 ms

Hear on Right
(consistent ITD)
Resolution of phase ambiguity by across-frequency comparison

500 Hz: period = 2 ms

300 Hz: period = 3.3 ms

Resolution of phase ambiguity by across-frequency comparison

Left ear actually lags by 1.5 ms

Cross-correlation peaks for sound delayed in one ear by 1.5 ms
Asynchrony & localisation

Asynchrony of 500 Hz tone (ms)

Pointer IID (dB)

6 Ss
Mistuning & localisation

Graph showing the relation between percent mistuning and Pointer I/D (dB) with 6 Ss.