Source Segregation

Chris Darwin

Experimental Psychology

University of Sussex

- Ears receive mixture of sounds

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

- Stored information about sounds (eg acoustic/phonetic relations) probably concerns a single source

- Need to make single source properties (eg silence) explicit

- Single-source properties not explicit in input signal
- eg silence (Darwin & Bethel-Fox, JEP:HPP 1977)

NB experience of yodelling may alter your susceptibility to this effect

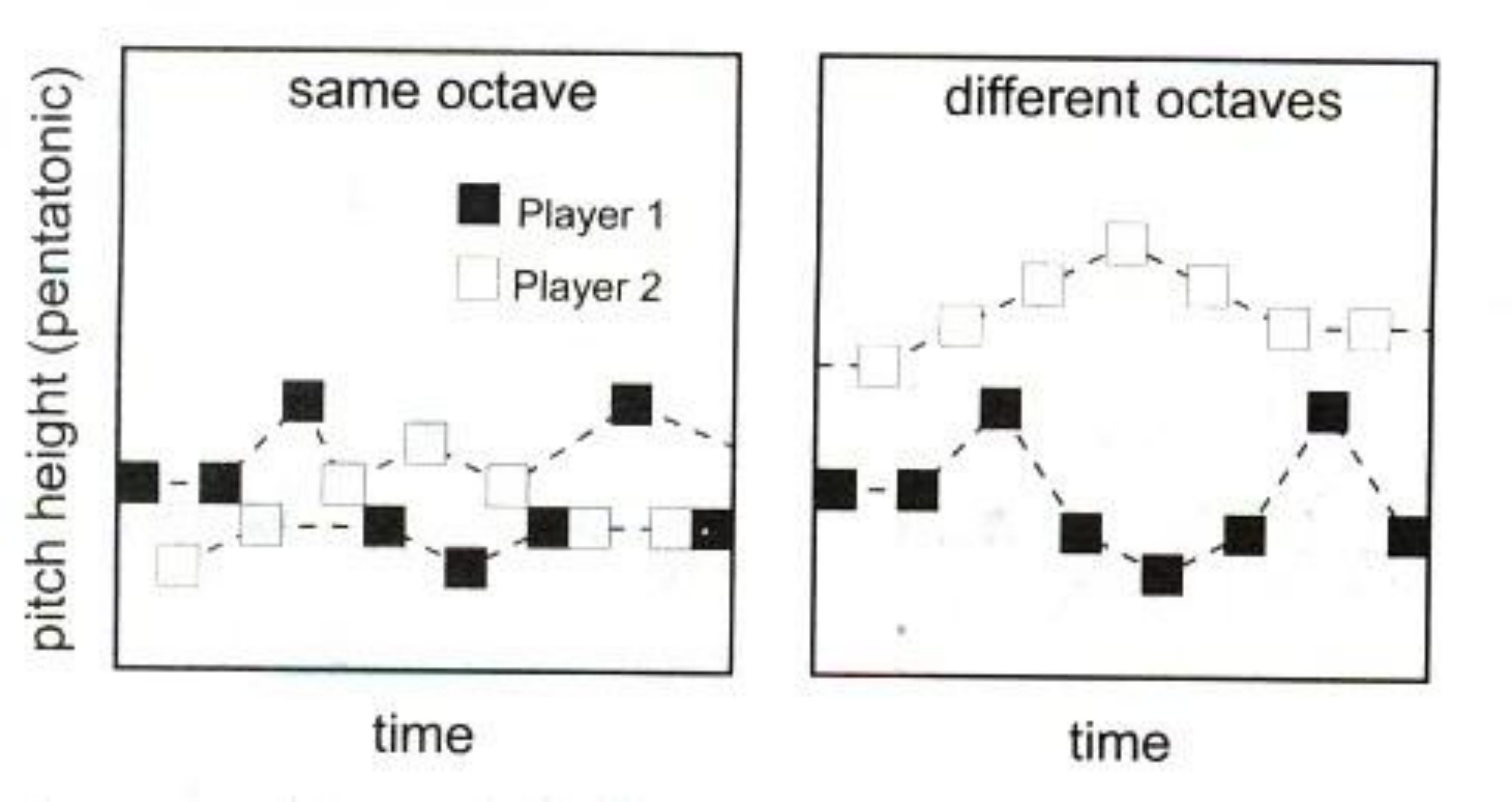
- Primitive grouping mechanisms based on general heuristics such as harmonicity and onset-time - “bottom-up” / “pure audition”

- Schema-based mechanisms based on specific knowledge (general speech constraints?) - “top-down.

- Successive segregation
  - Different frequency (or pitch)
  - Different spatial position
  - Different timbre

- Simultaneous segregation
  - Different onset-time
  - Irregular spacing in frequency
  - Location (rather unreliable)
  - Uncorrelated FM not used

Bugandan xylophone music: “Ssematimba ne Kikwabanga”



Track 7

Track 8

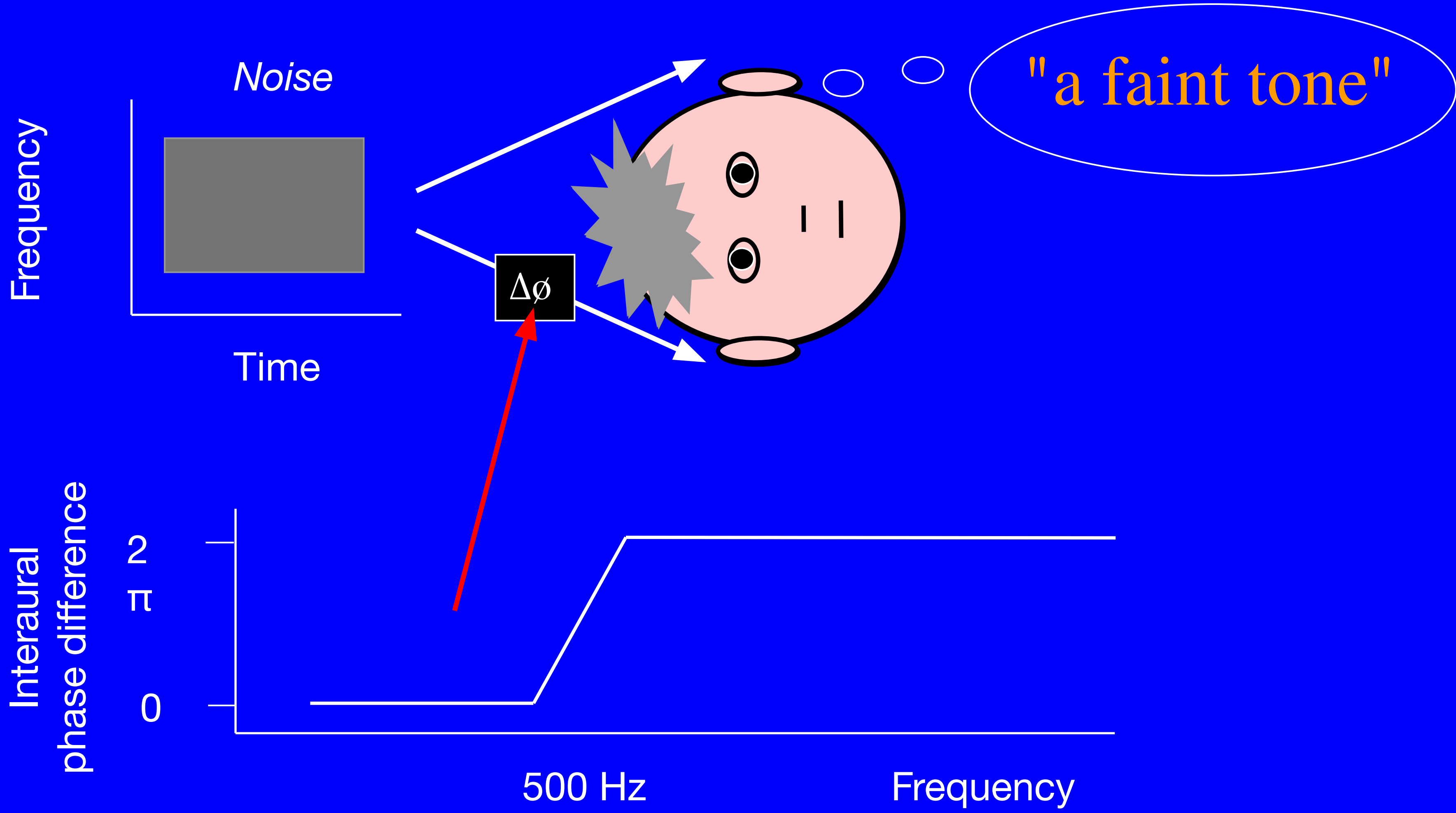
Streaming occurs for sounds

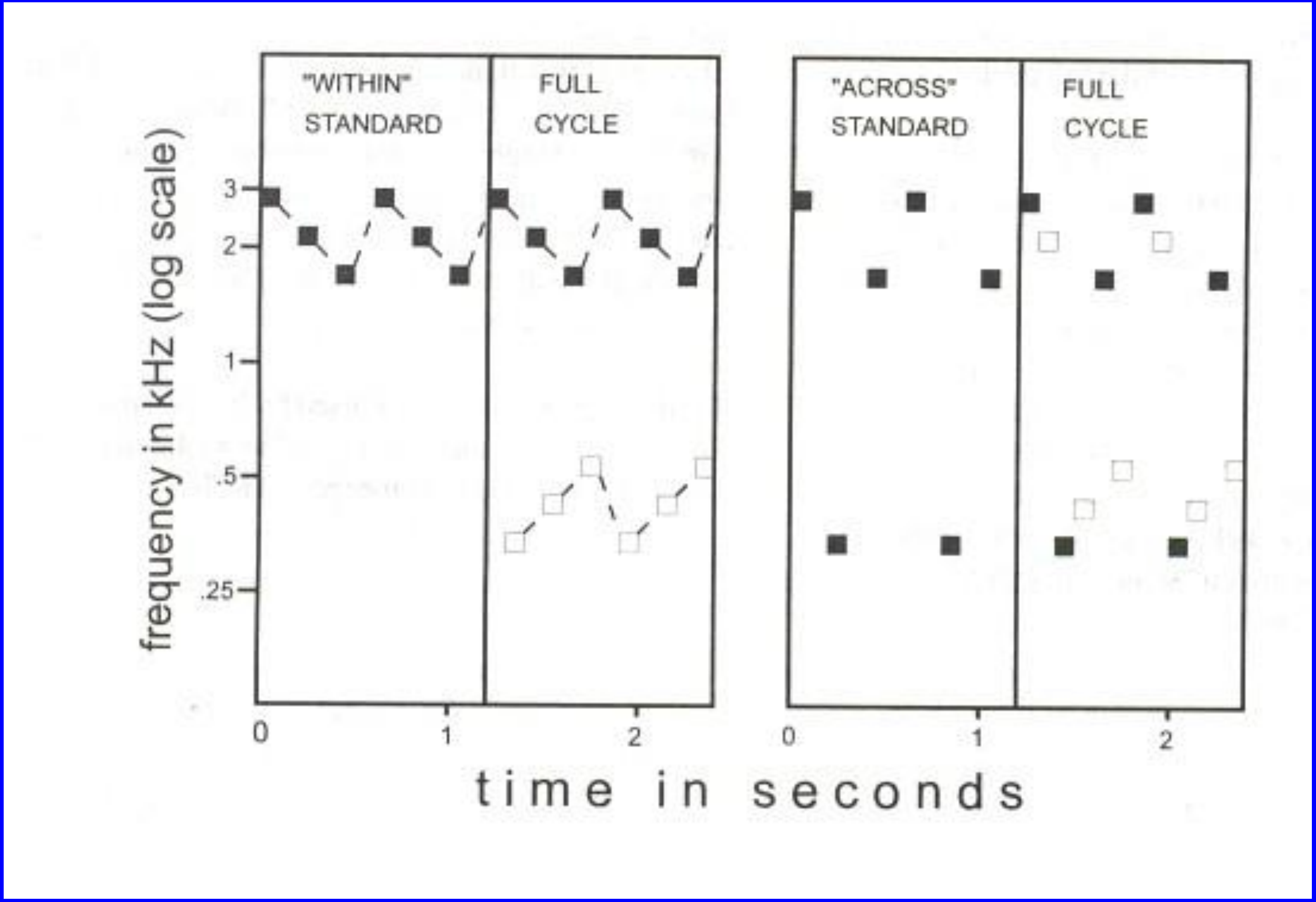
- with same auditory excitation pattern, but different periodicities

  Vliegen, J. and Oxenham, A. J. (1999). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am. 105, 339-46.

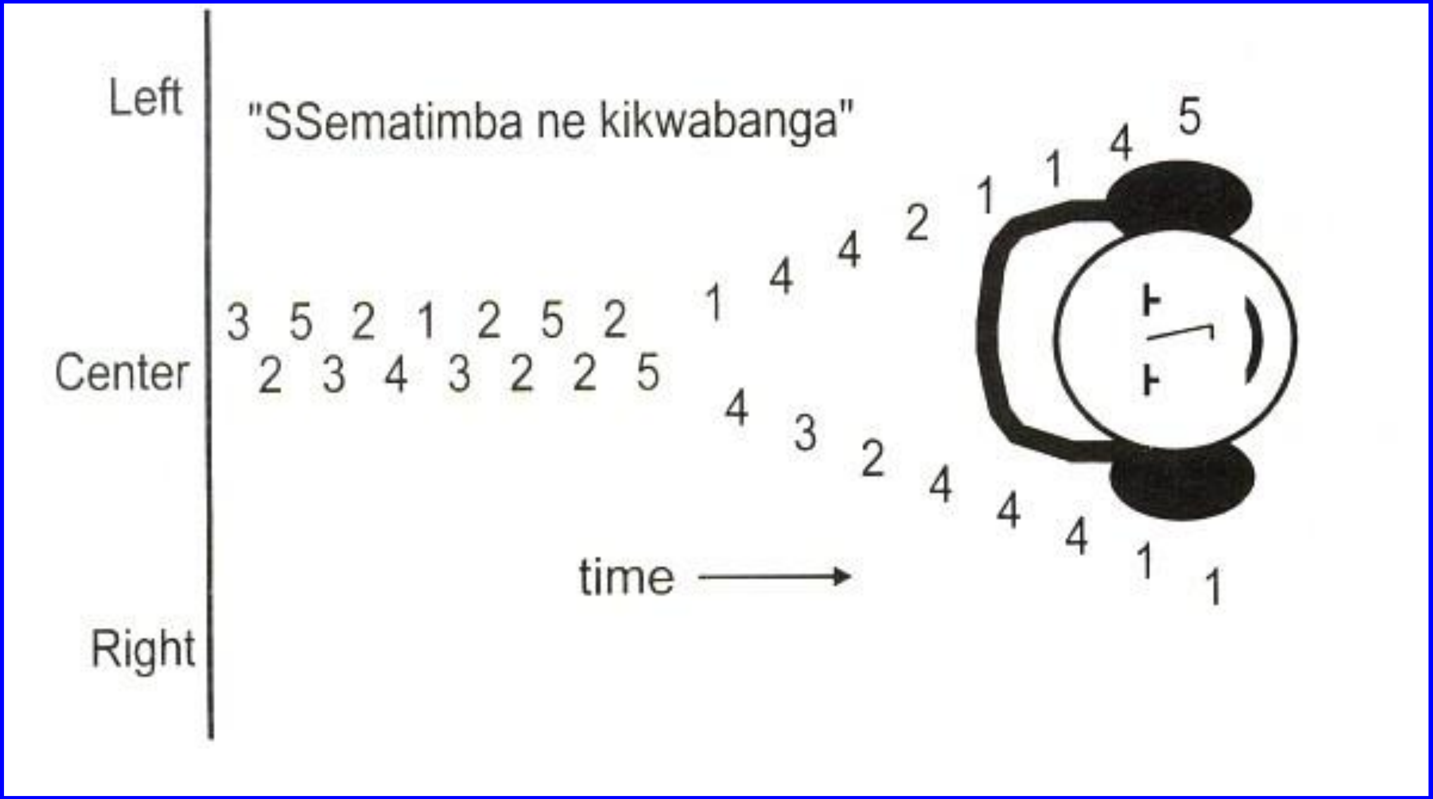
- with Huggins pitch sounds that are only defined binaurally

  Carlyon & Akeroyd


Noise
Frequency
Time
Δø
"a faint tone"
Interaural phase difference
2π
0
500 Hz
Frequency


"WITHIN" STANDARD
FULL CYCLE
"ACROSS" STANDARD
FULL CYCLE
frequency in kHz (log scale)
3
2
1
.5
.25
0
1
2
0
1
2
time in seconds

Track 2


Left
Center
Right
"SSematimba ne kikwabanga"
3 5 2 1 2 5 2
2 3 4 3 2 2 5
1 4 4 2 1 1 4 5
4 3 2 4 4 4 1 1
time

Track 41

Sach & Bailey - rhythm unmasking by ITD or spatial position ?

Rhythm A

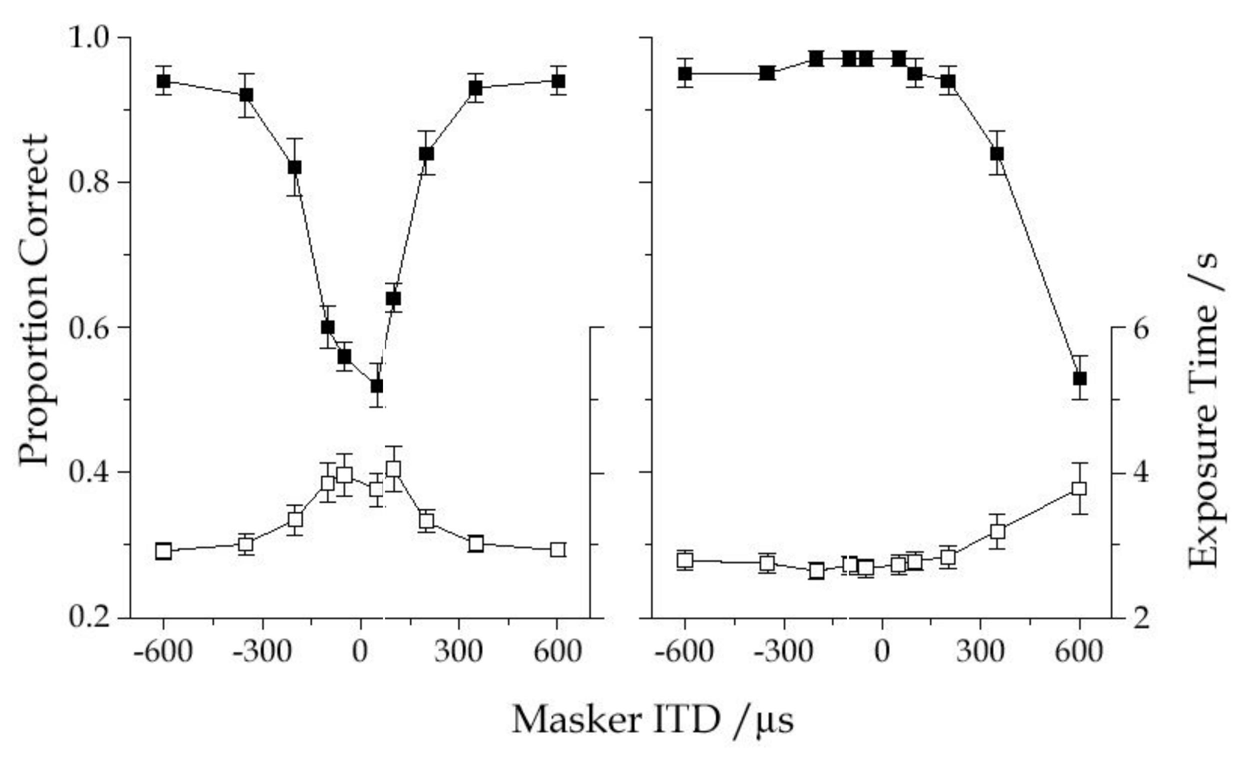
● 1 2 ● 3 4 5 6 ●

Rhythm B

● 1 2 ● 3 4 ● 5 6 7 8 ● 9 10 11 12 ●

Target • ITD=0, ILD = 0

Target • ITD=0, ILD = +4 dB



ITD sufficient but, sequential segregation by spatial position rather than by ITD alone.

Horse Morse

-LHL-LHL-LHL- --H---H---H--

-L-L-L-L-L-L-L

- Segregation takes a few seconds to build up.
- Then between-stream temporal / rhythmic judgments are very difficult

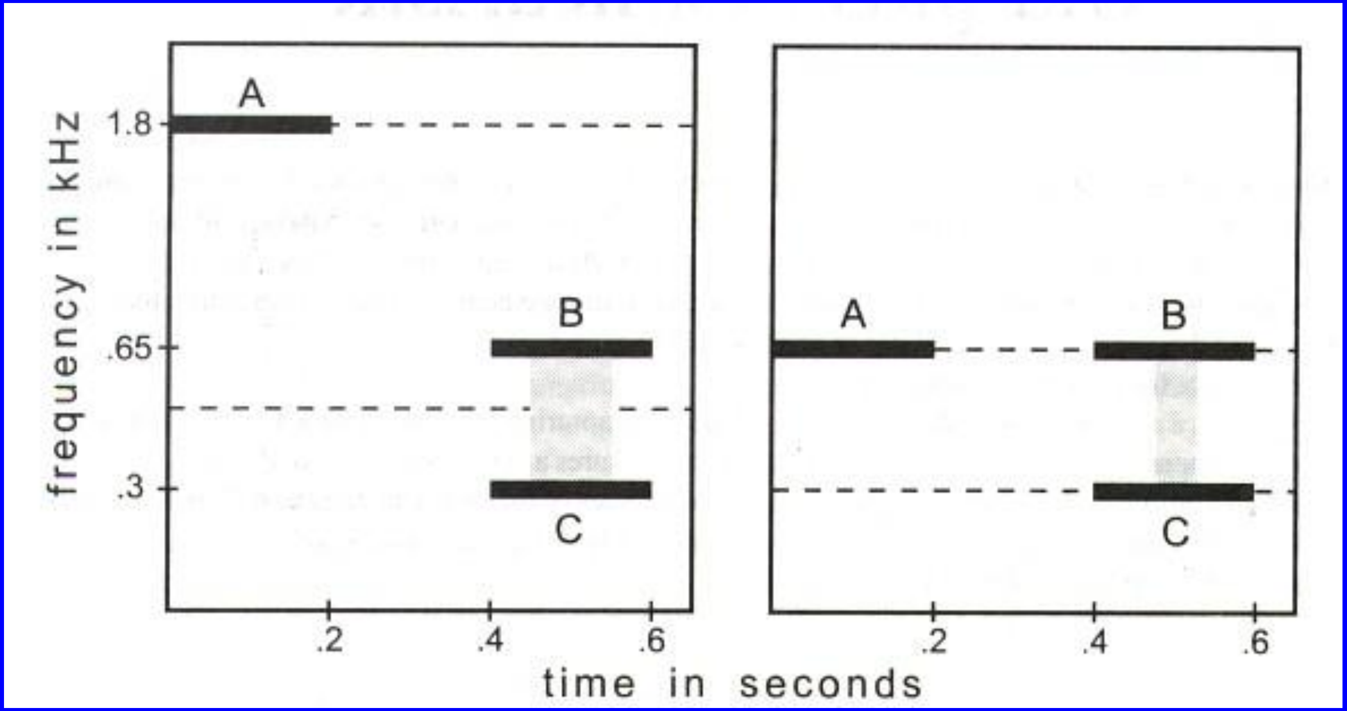
- Sequential streaming may require attention - rather than being a pre-attentive process.

Horse --> Morse

```
-LHL-LHL-LHL-   -->   --H---H---H--
            -L-L-L-L-L-L-L
```

- Horse -> Morse takes a few seconds to segregate
- These have to be seconds spent attending to the tone stream
- Does this also apply to other types of segregation?

A-B
A-BC



Freq separation of AB
Harmonicity & synchrony of BC

What is the timbre / pitch / location of a particular sound source ?

Important grouping cues

- continuity
- onset time

(Old + New)

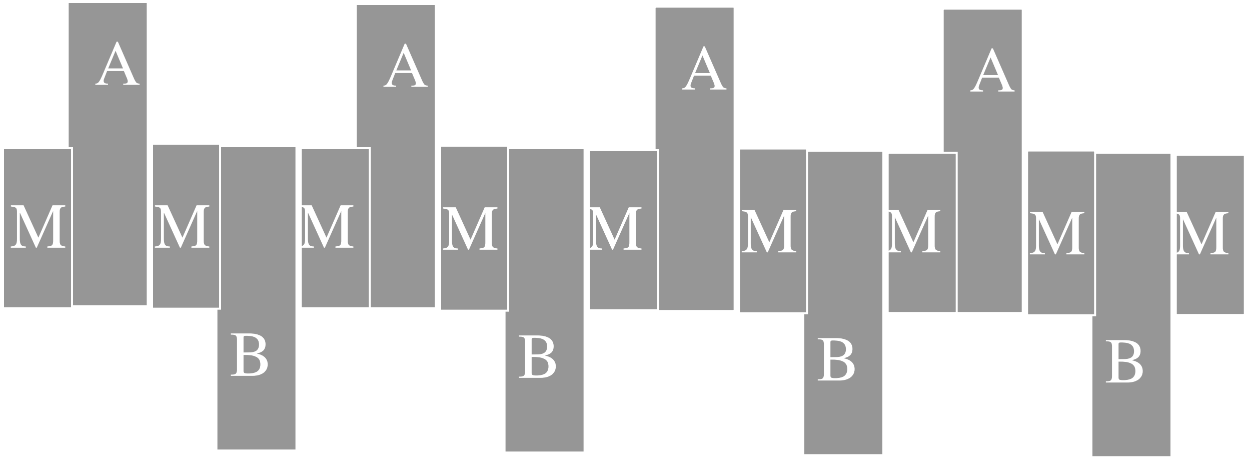
- harmonicity (or regularity of frequency spacing)

Stimulus: A followed by A+B

-> Percept of:

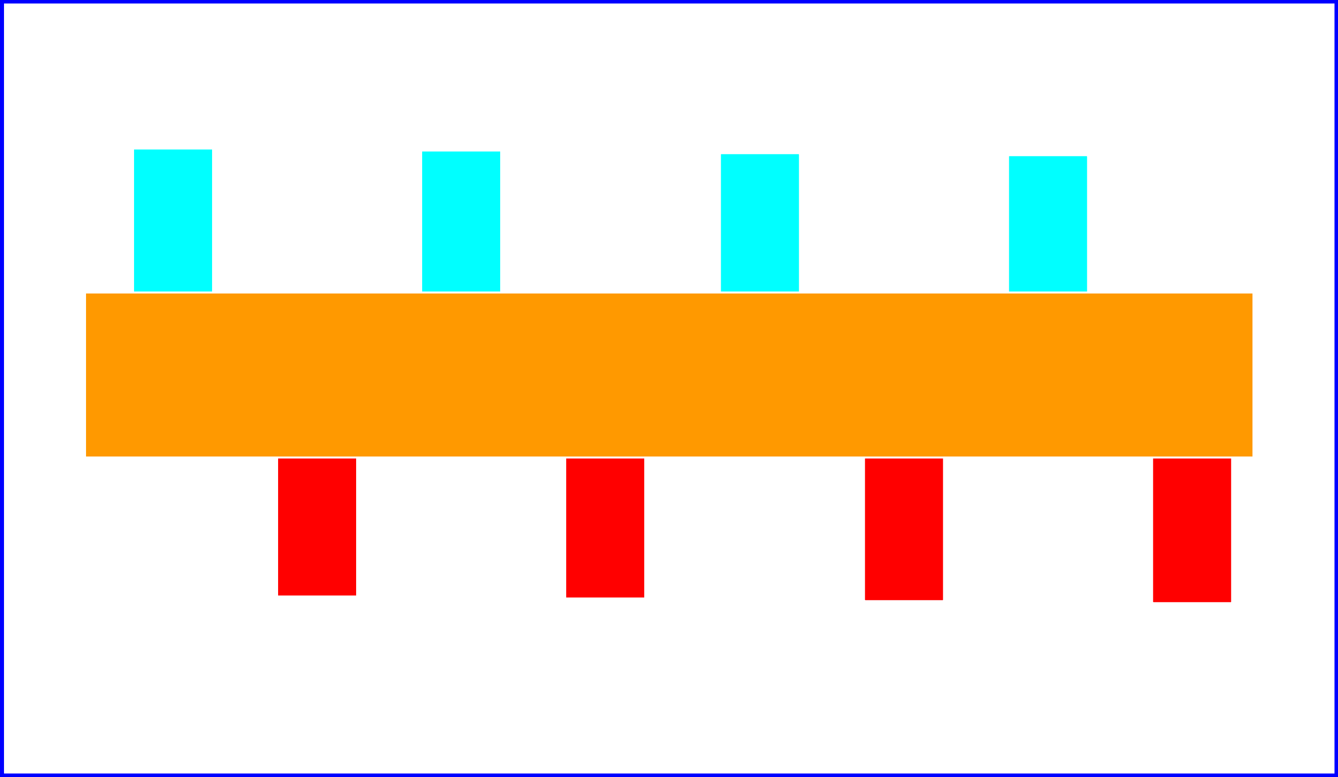
A as continuous (or repeated)

with B added as separate percept

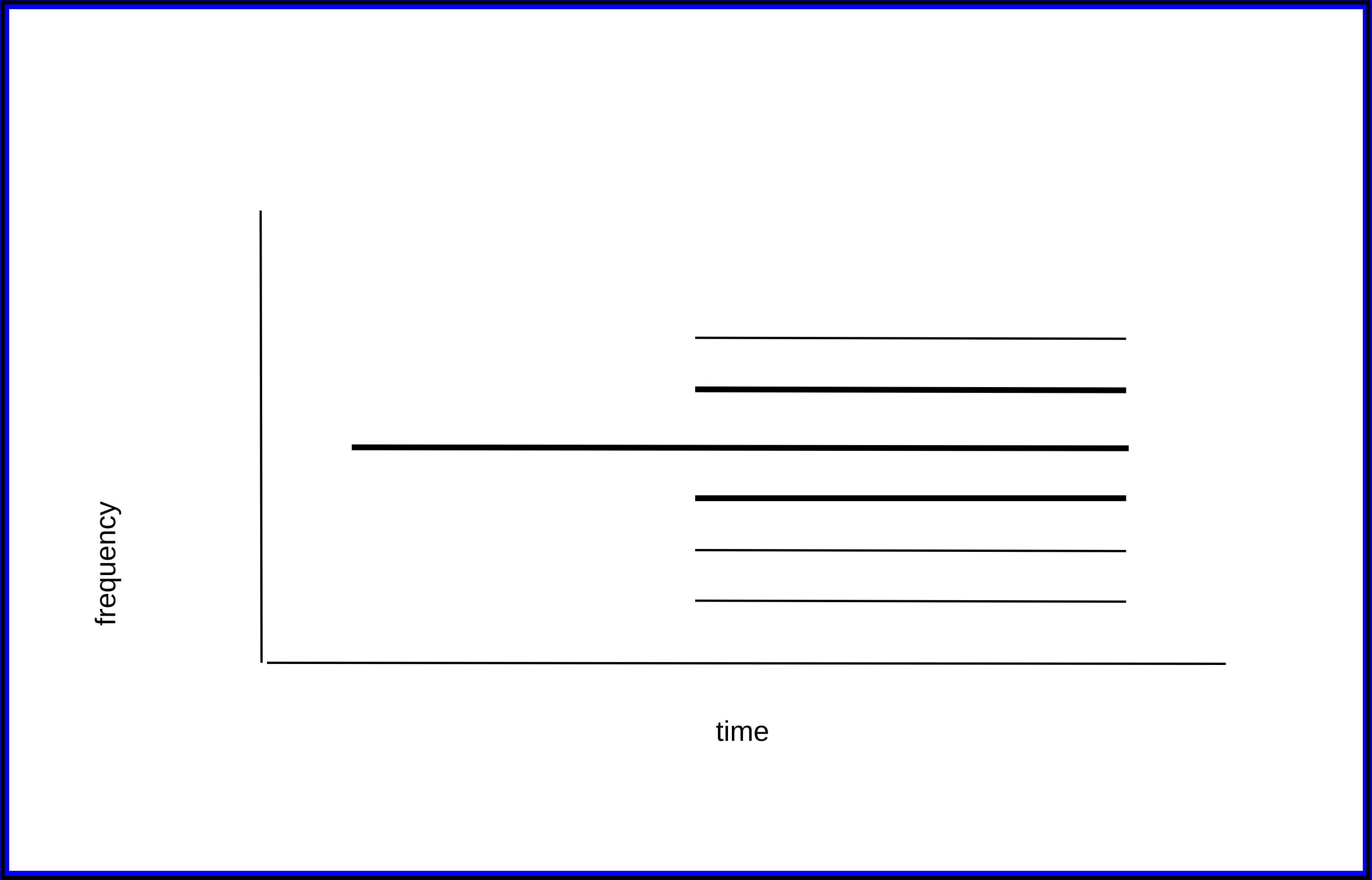


A
MAMB

B
MAMB

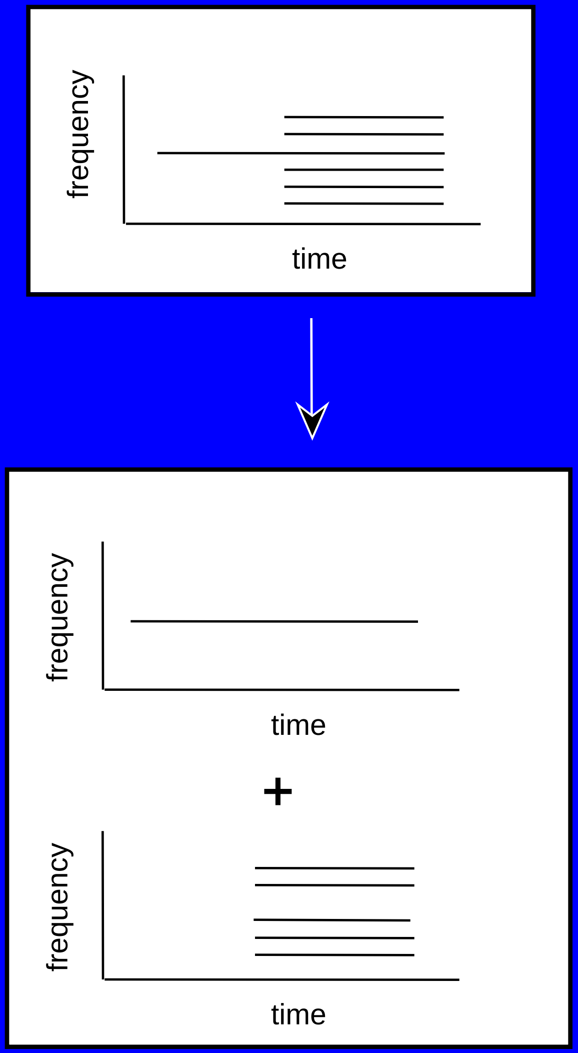


Grouping & vowel quality

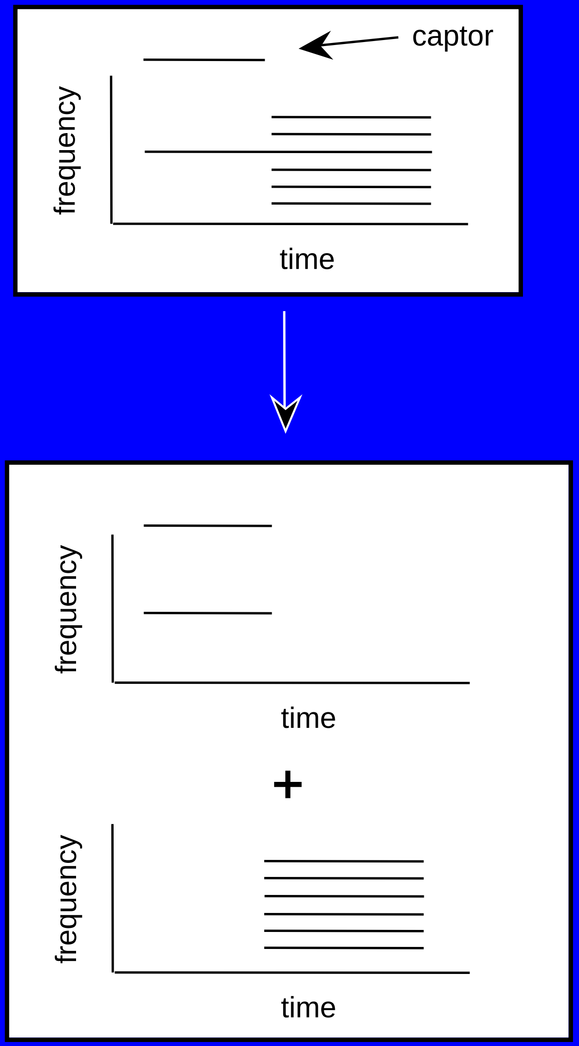


Grouping & vowel quality (2)

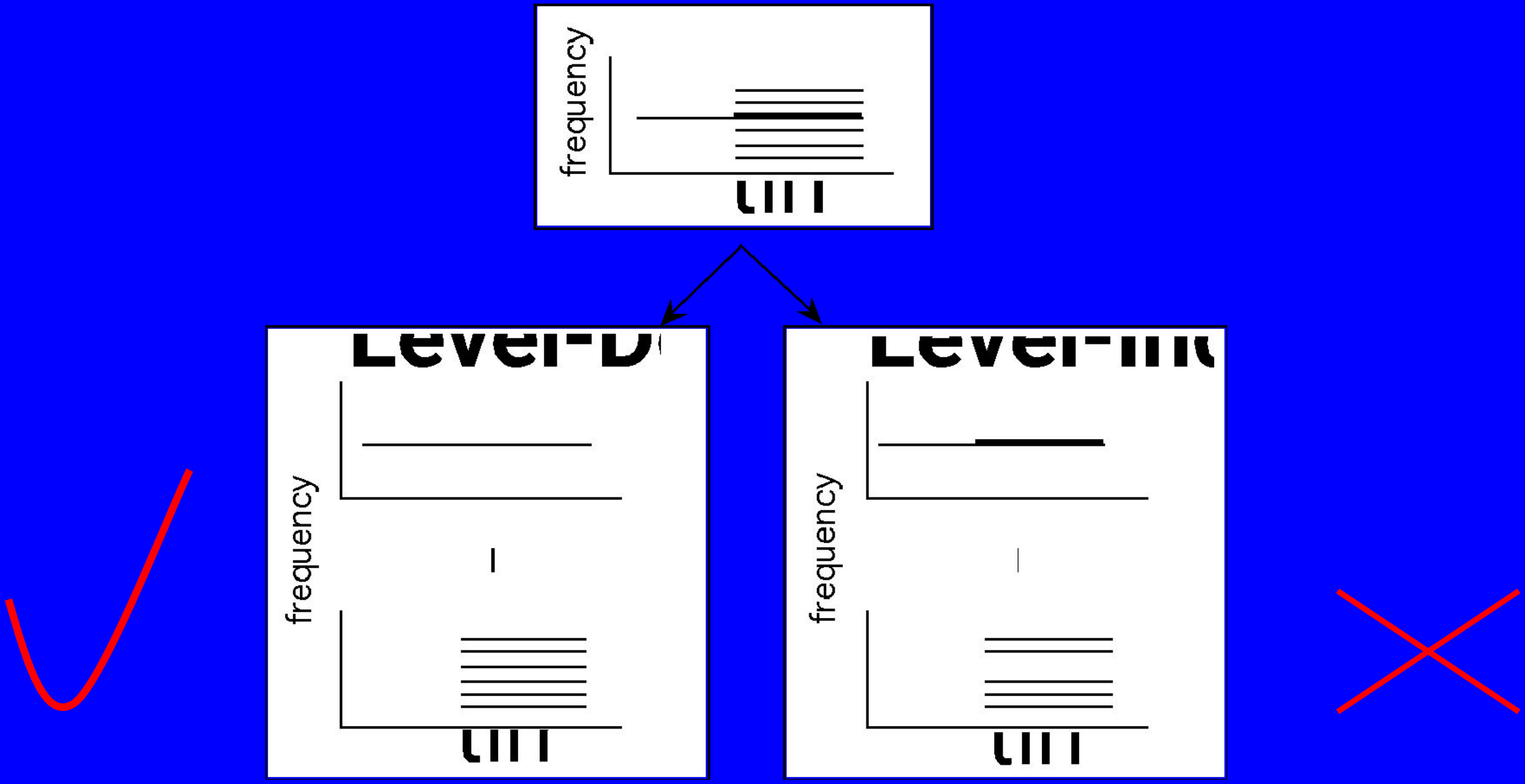
continuation removed from vowel



continuation not removed from vowel

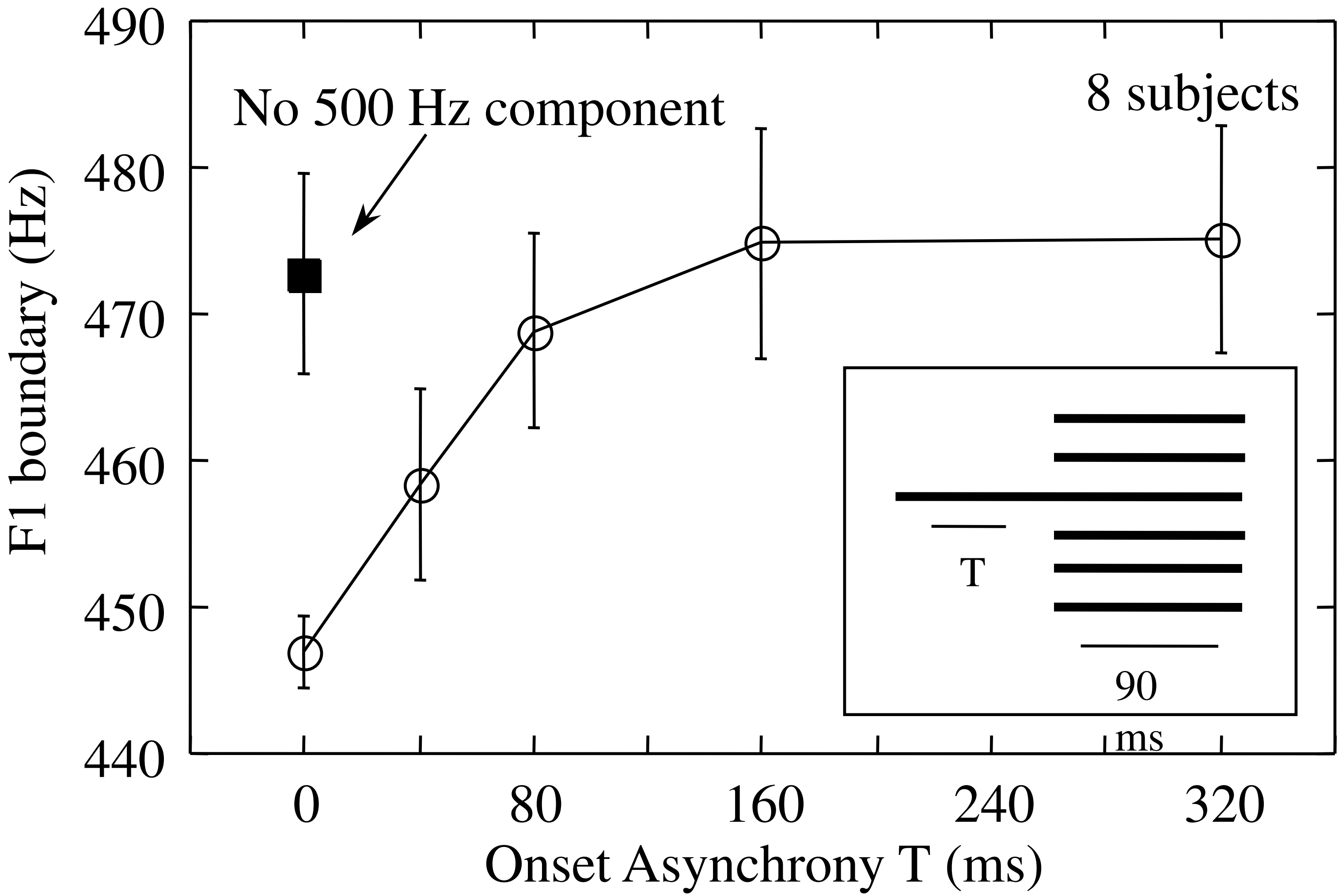


- Bregman’s Old-plus-New heuristic

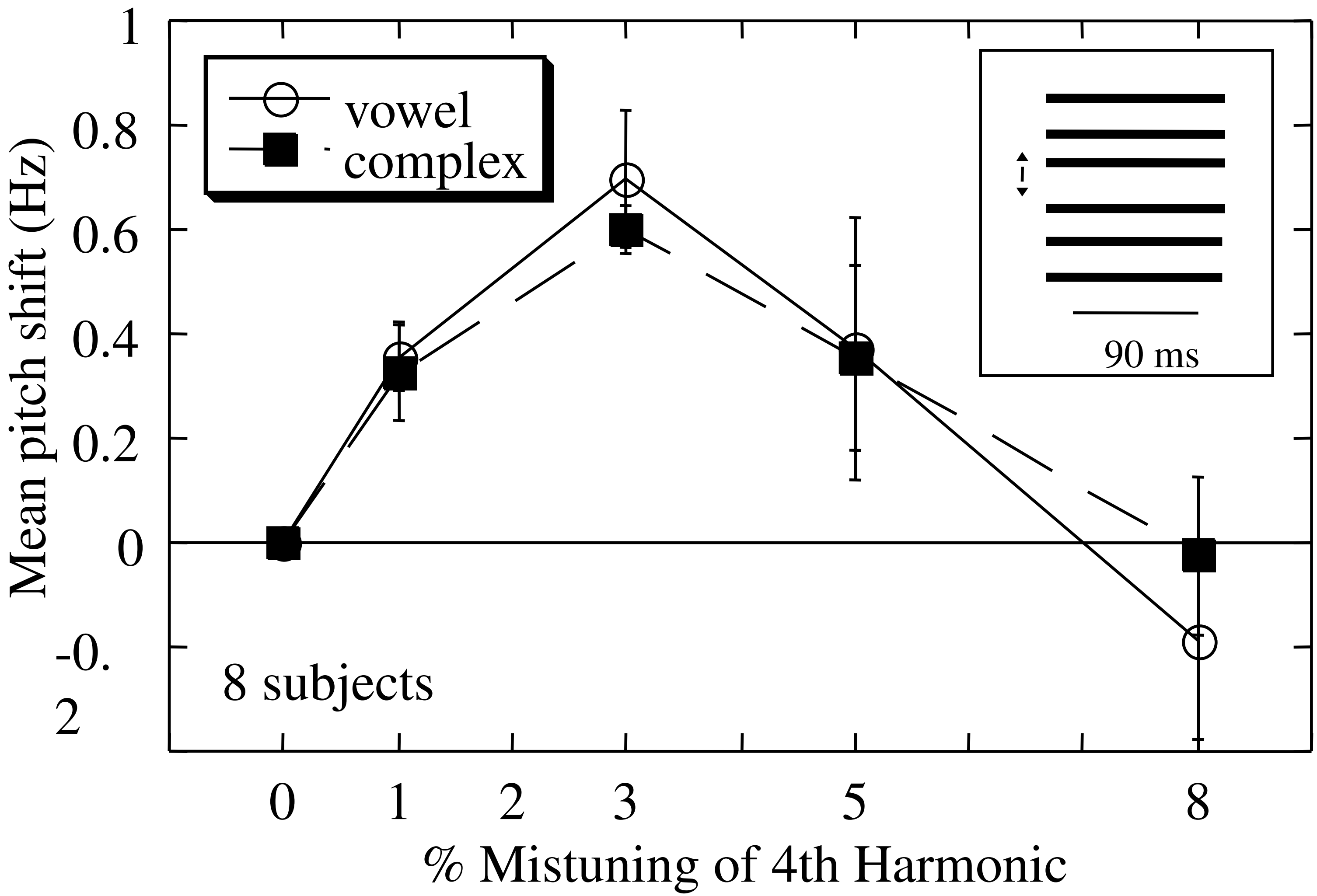


- Indicates importance of coding change.

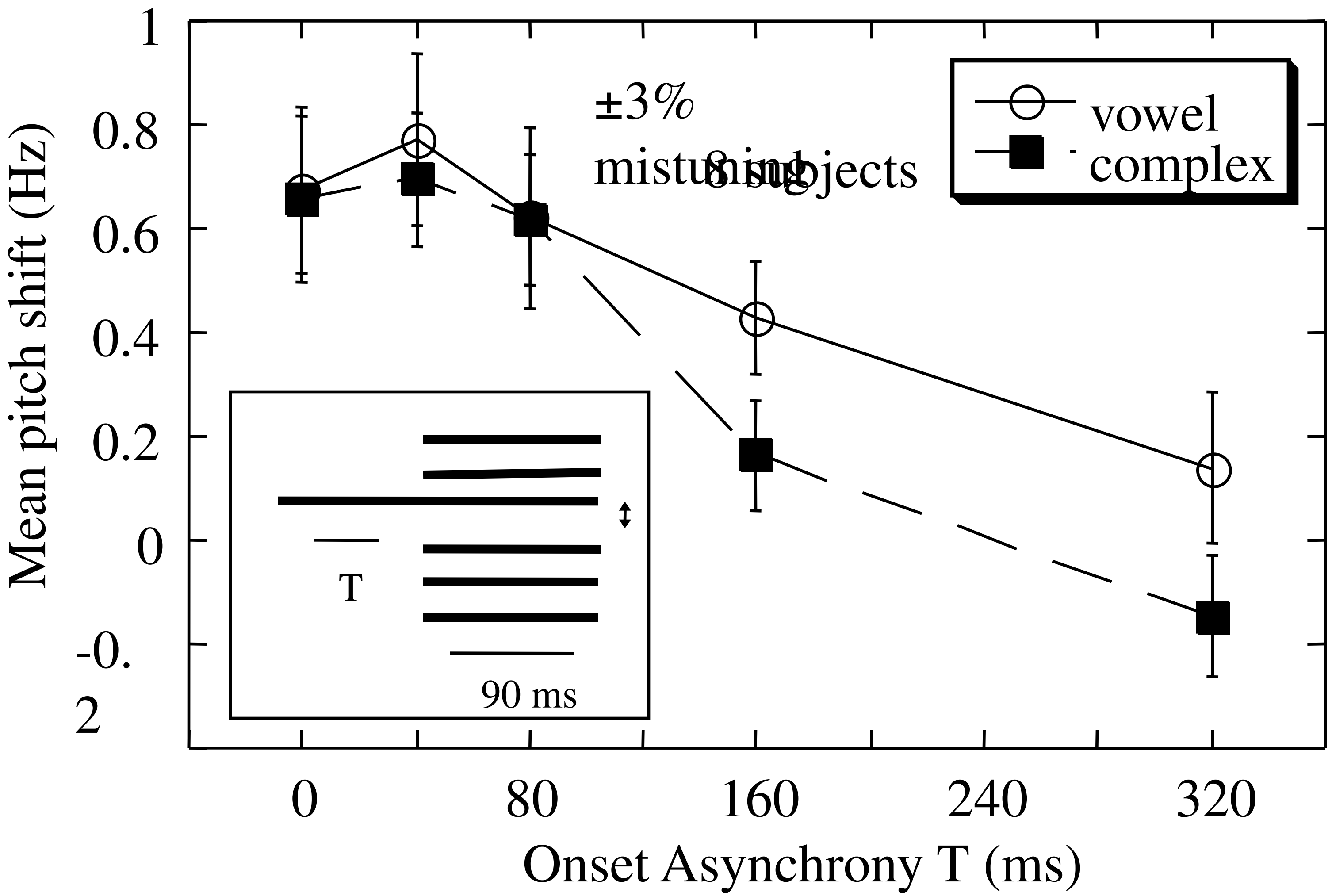
Asynchrony & vowel quality



Mistuning & pitch



Onset asynchrony & pitch



Some interesting points:

- Sequential streaming may require attention - rather than being a pre-attentive process.
- Parametric behaviour of grouping depends on what it is <u>for</u>.

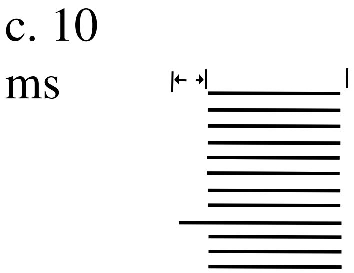
Grouping for

Effectiveness of a parameter on grouping depends on the task. Eg

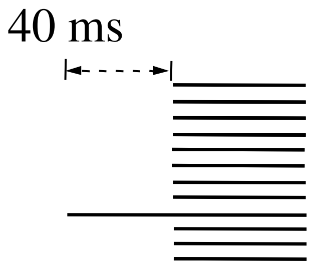
- 10-ms onset time allows a harmonic to be heard out
- 40-ms onset-time needed to remove from vowel quality
- >100-ms needed to remove it from pitch.

Minimum onset needed for:

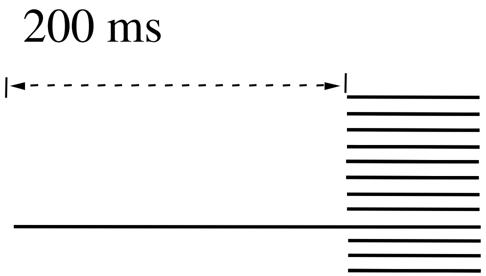
Harmonic in vowel to be heard out:



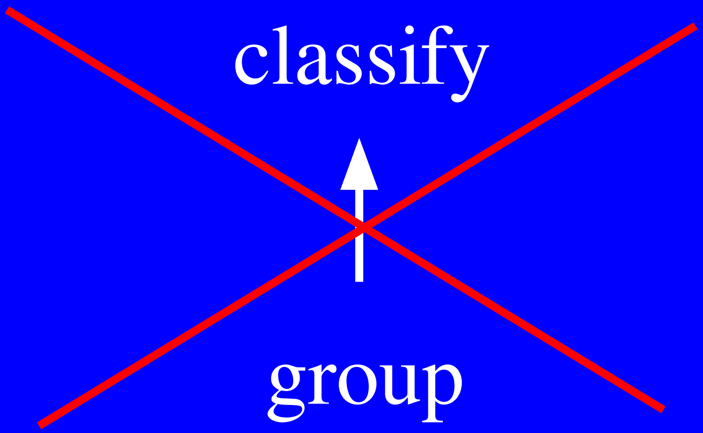
Harmonic to be removed from vowel:



Harmonic to be removed from pitch:



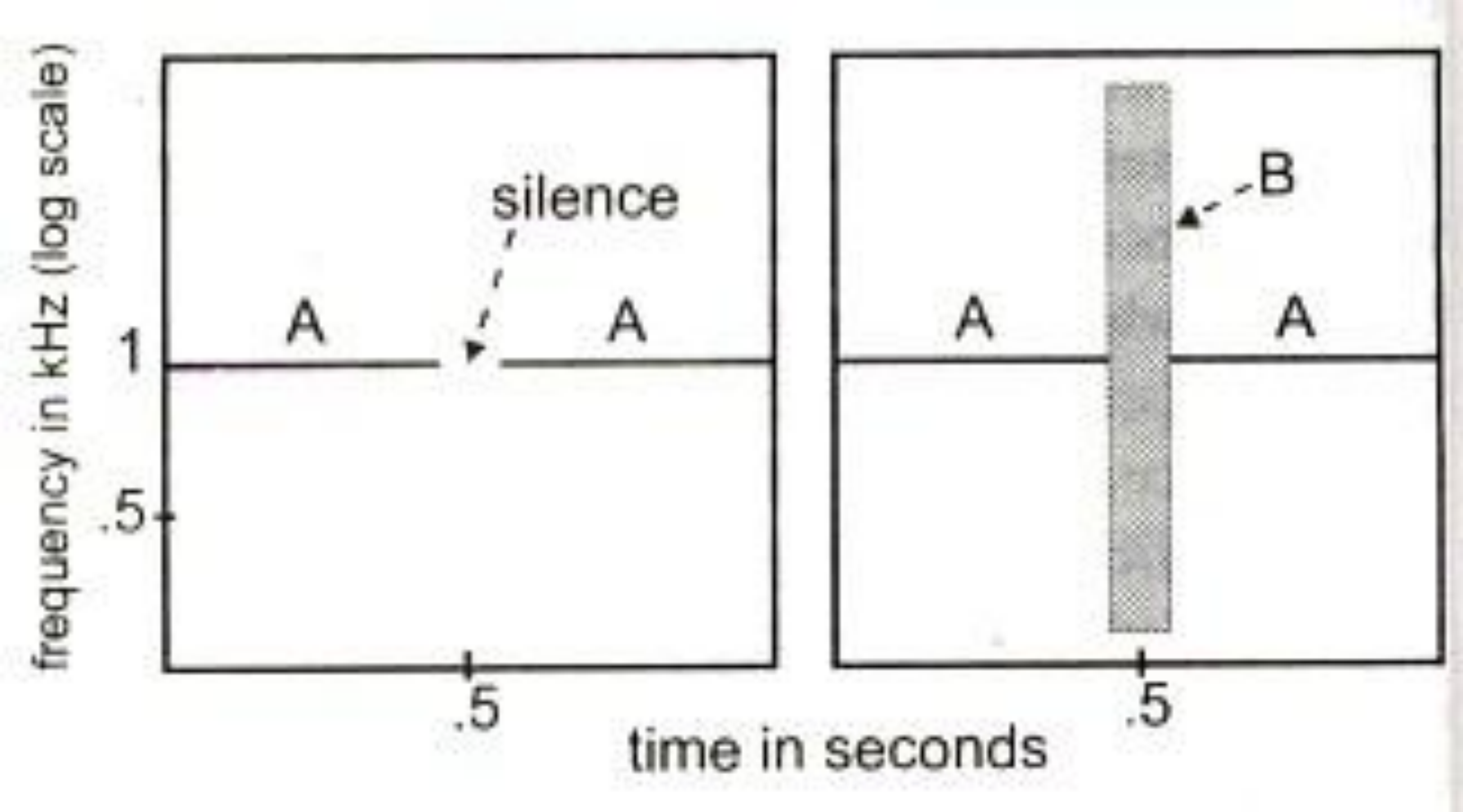
Grouping not absolute and independent of classification



classify

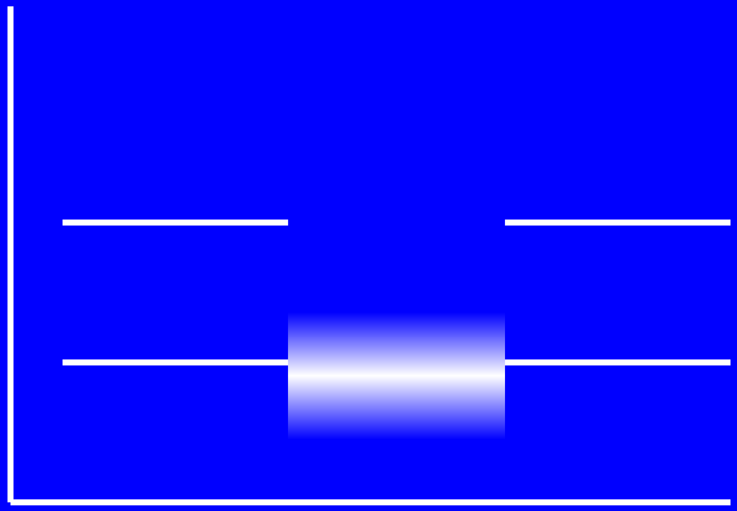
group

If B would have masked if it HAD been there, then you don't notice that it is not there.



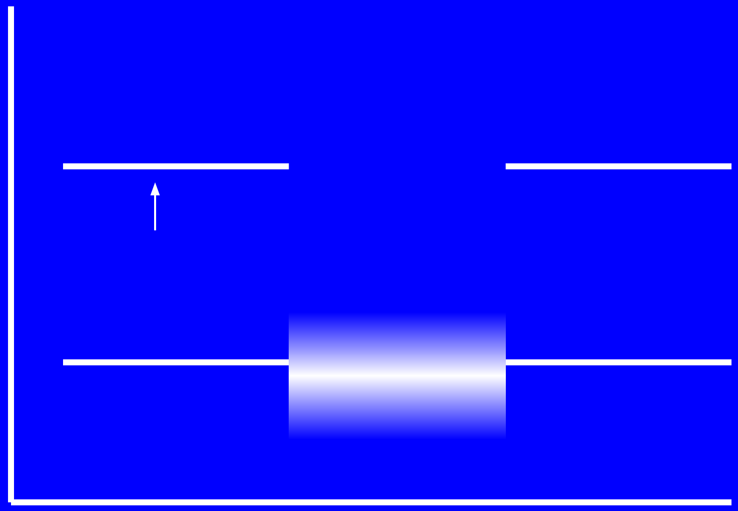
Track 28

Harmonic



1. Pulsing complex

Enharmonic



1. Pulsing high tone
2. Steady low tone

Group tones; then decide on continuity.

- Sequential streaming may require attention - rather than being a pre-attentive process.
- Parametric behaviour of grouping depends on what it is for.
- Not everything that is obvious on an auditory spectrogram can be used :
    - FM of Fo irrelevant for segregation (Carlyon, JASA 1991; Summerfield & Culling 1992)

5 Hz, 2.5% FM

frequency

Harm	Inharmonic
2500	2500
2000	2100
1500	1500
Easy	Impossible

1 2 3

Odd-one in 2 or 3 ?

Carlyon, R. P. (1991). "Discriminating between coherent and incoherent frequency modulation of complex tones," J. Acoust. Soc. Am. 89, 329-340.

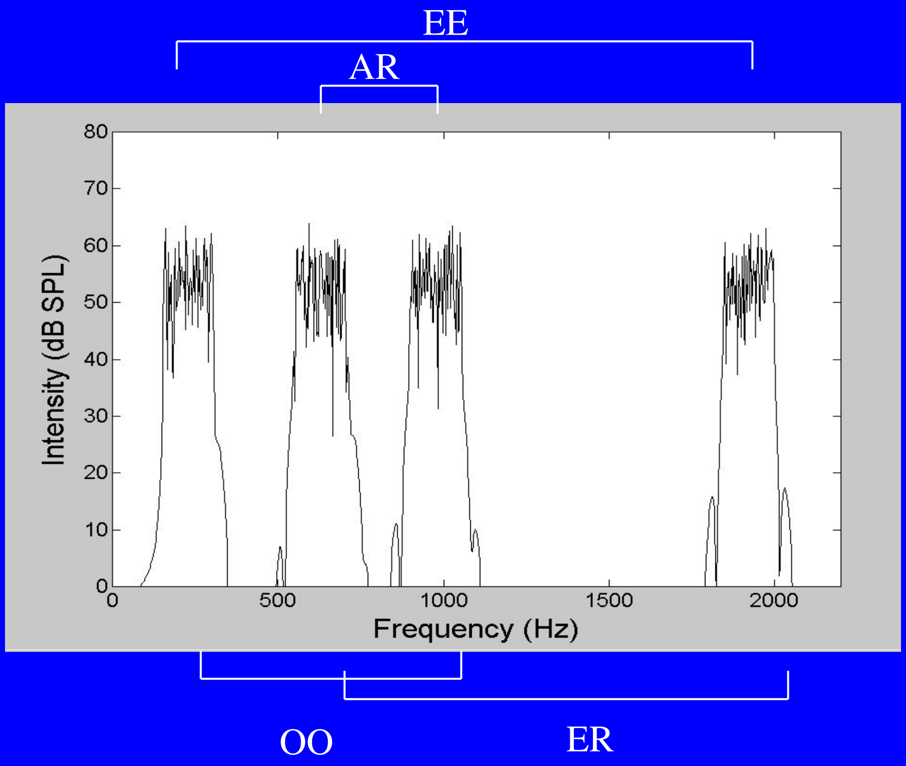
What role do localisation cues play in helping us to hear one voice in the presence of another ?

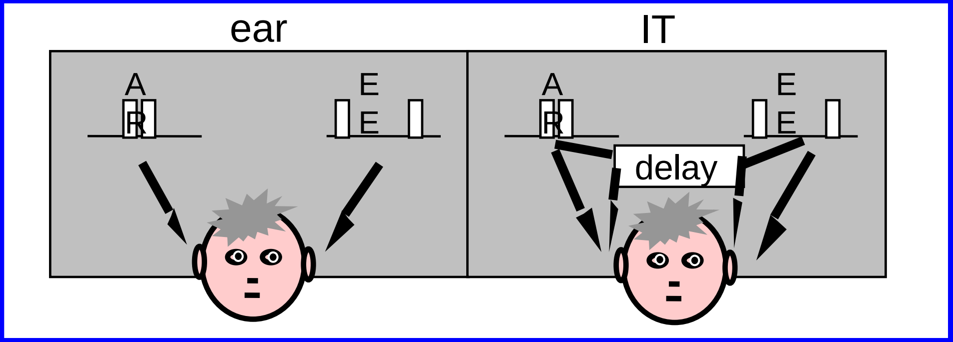
- Head shadow increases S/N at the nearer ear (Bronkhurst & Plomp, 1988).
  - … but this advantage is reduced if high frequencies inaudible (B & P, 1989)
- But do localisation cues also contribute to selectively grouping different sound sources?

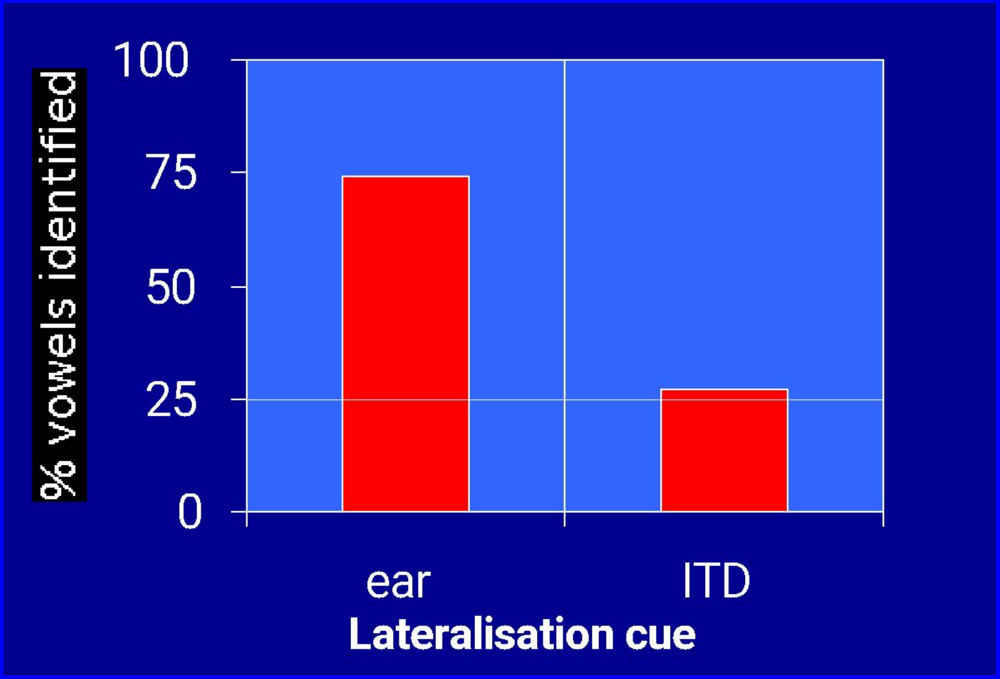
- Sequential streaming may require attention - rather than being a pre-attentive process.
- Parametric behaviour of grouping depends on what it is for.
- Not everything that is obvious on an auditory spectrogram can be used :
    - FM of Fo irrelevant for segregation (Carlyon, JASA 1991; Summerfield & Culling 1992)
- Although we can group sounds by ear, ITDs by themselves remarkably useless for simultaneous grouping. Group first then localise grouped object.

- Noise bands played to different ears group by ear, but...

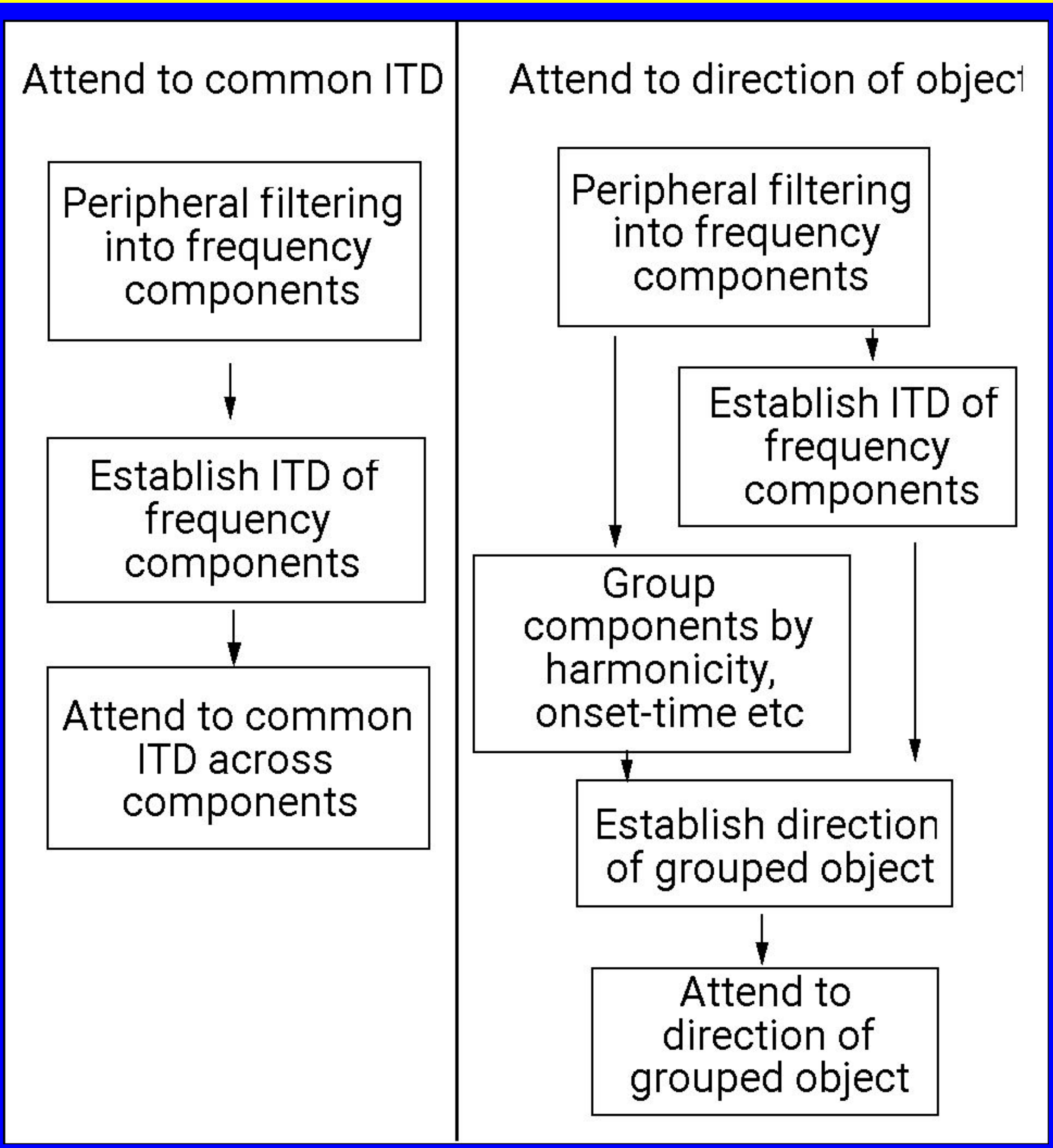
- Noise bands differing in ITD do not group by ear







Task - what vowel is on your left ? ("ee")



Attend to common ITD

Peripheral filtering into frequency components

Establish ITD of frequency components

Attend to common ITD across components

Attend to direction of object

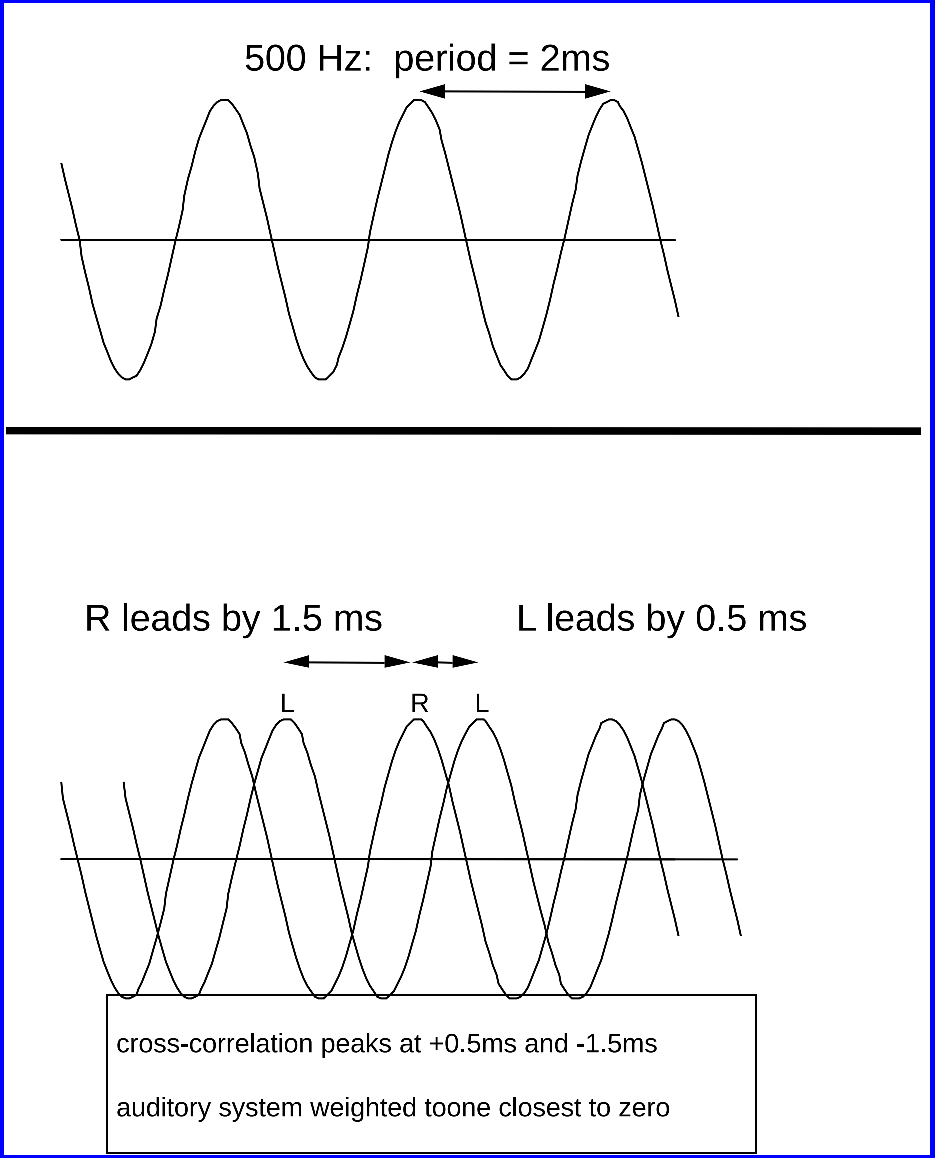
Peripheral filtering into frequency components

Establish ITD of frequency components

Group components by harmonicity, onset-time etc

Establish direction of grouped object

Attend to direction of grouped object



cross-correlation peaks at +0.5ms and -1.5ms

auditory system weighted toone closest to zero

500-Hz pure tone leading in Right ear by 1.5 ms

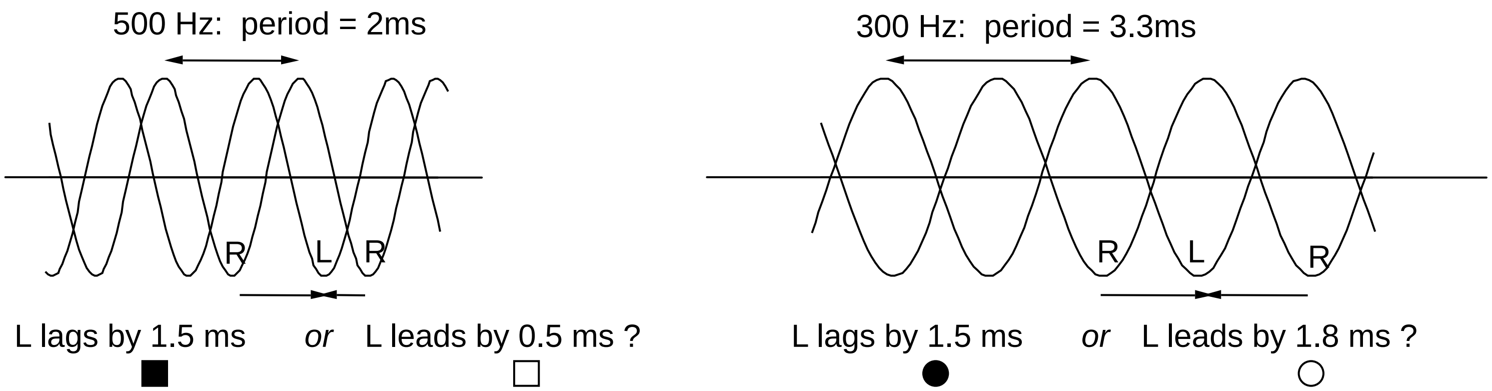
Heard on Left side

- Narrowband noise at 500 Hz with ITD of 1.5 ms (3/4 cycle) heard at lagging side.
- Increasing noise bandwidth changes location to the leading side.

Explained by across-frequency consistency of ITD.

(Jeffress, Trahiotis & Stern)

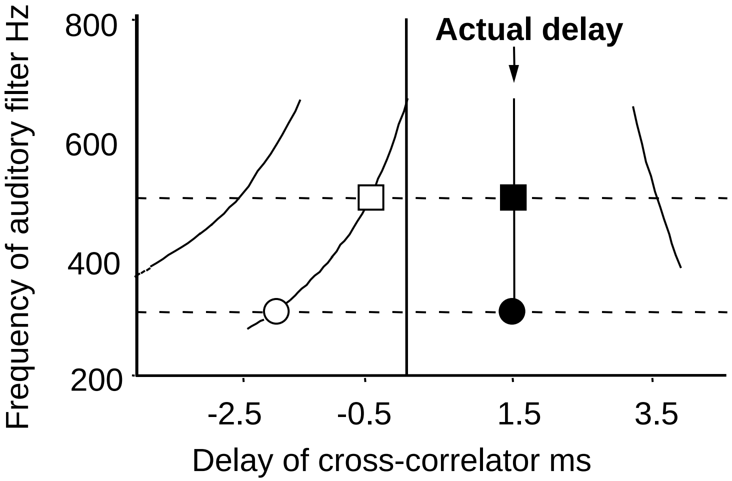
Left ear actually lags by 1.5 ms


500 Hz: period = 2ms

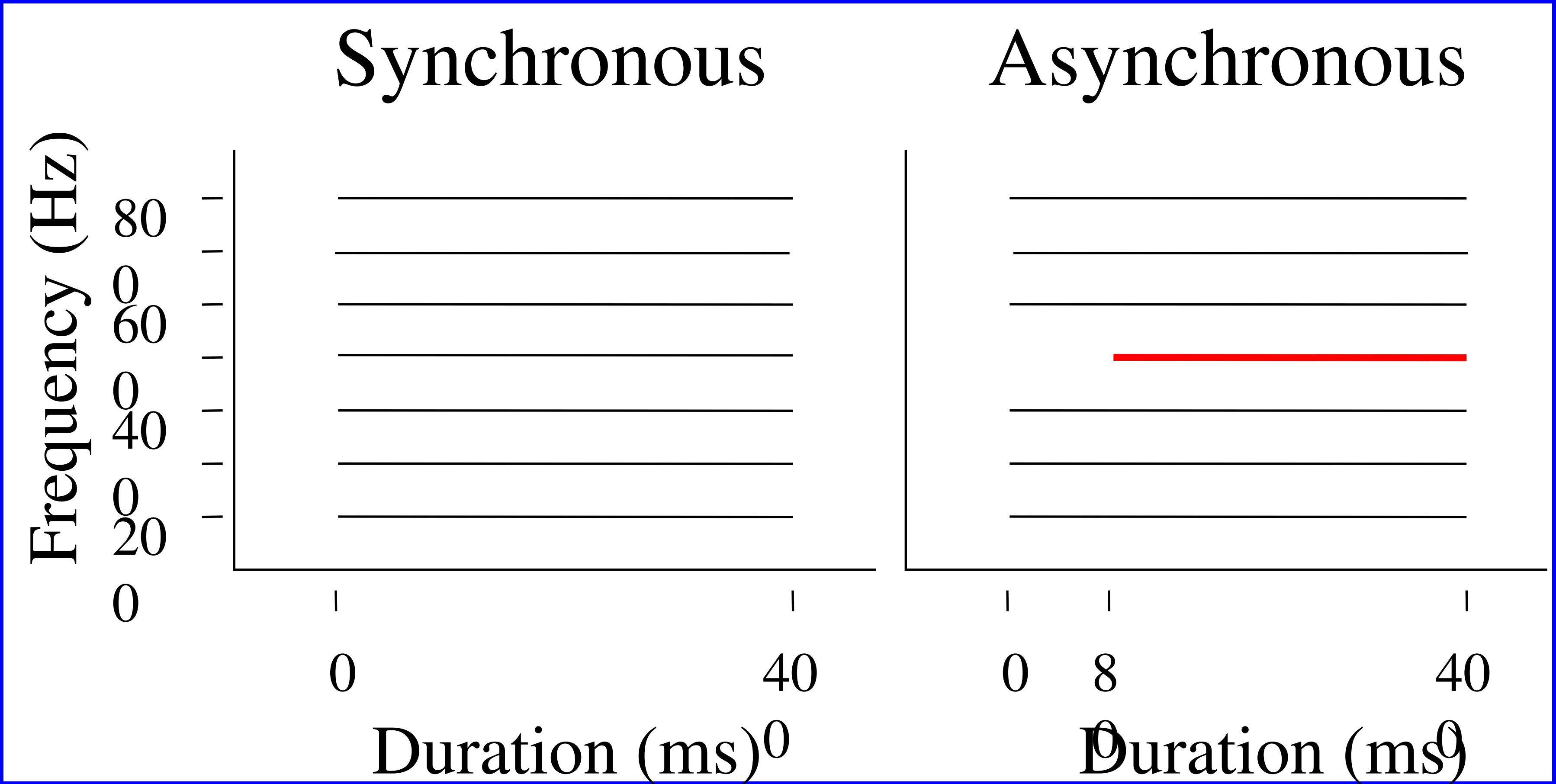
L lags by 1.5 ms ■ *or* L leads by 0.5 ms ? □

300 Hz: period = 3.3ms

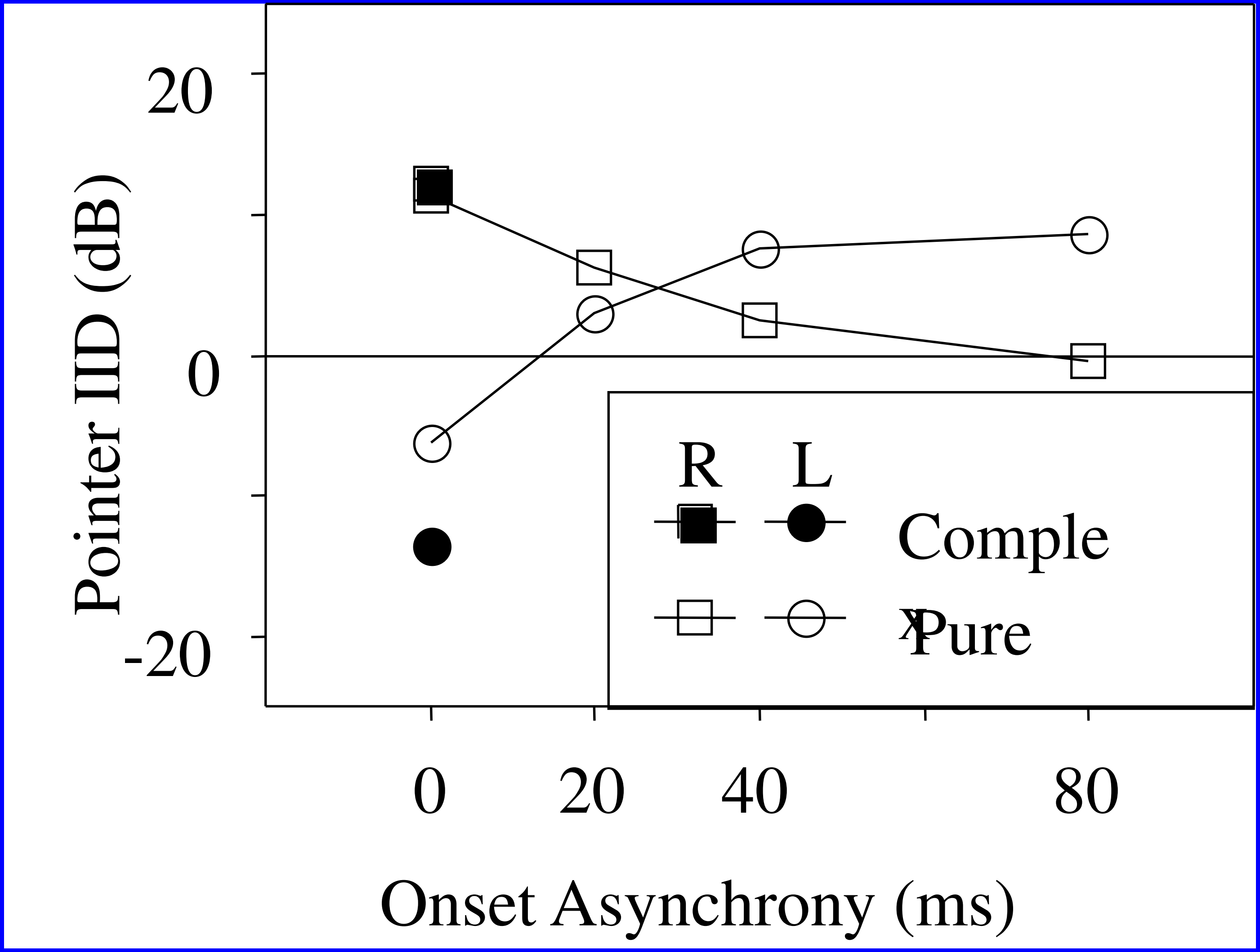
L lags by 1.5 ms ● *or* L leads by 1.8 ms ? ○

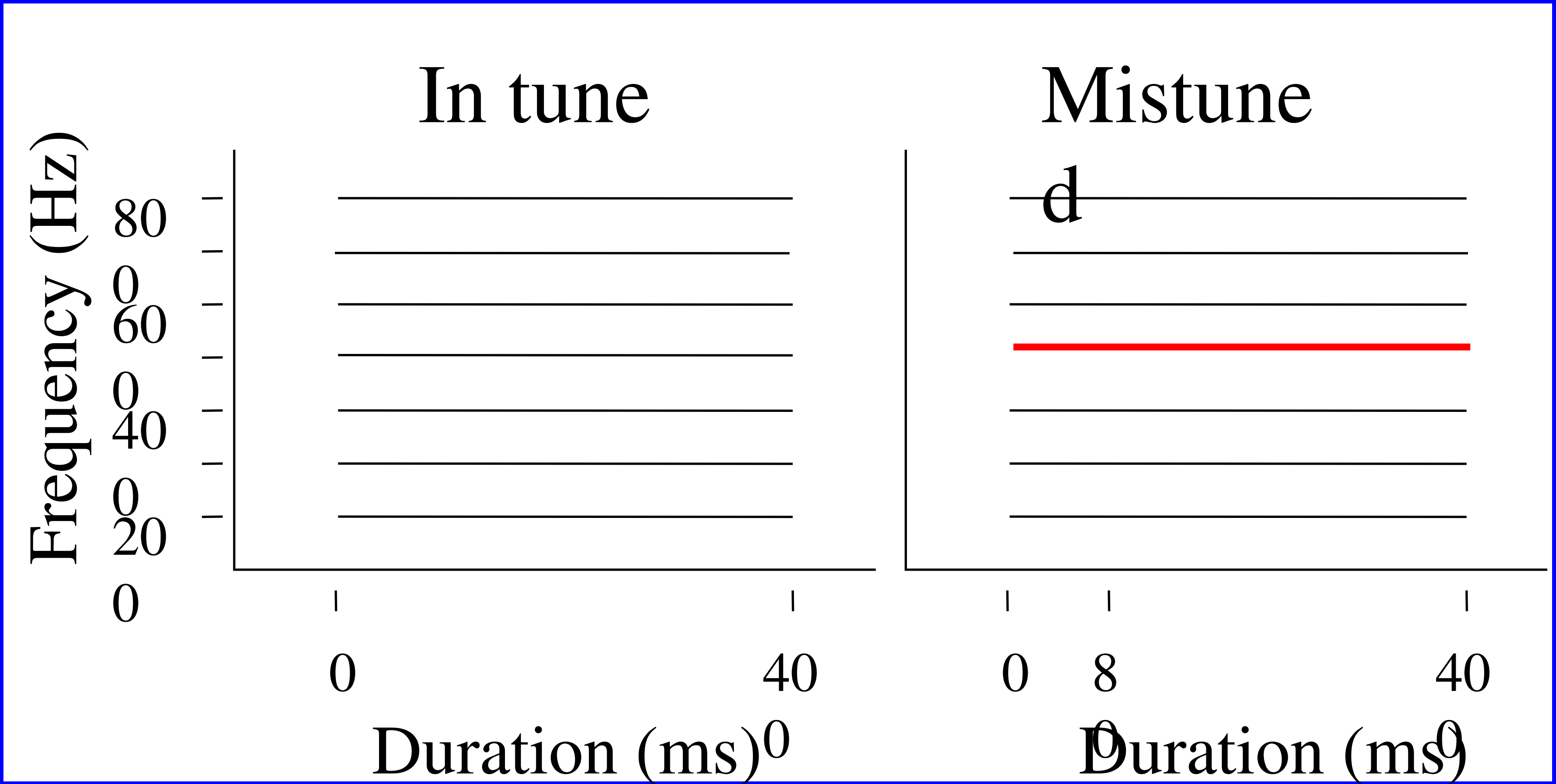


Cross-correlation peaks for noise delayed in one ear by 1.5 ms



ITD: ± 1.5 ms (3/4 cycle at 500 Hz)


Pointer IID (dB)
20
0
-20
R
L
■
●
Comple
□
○
x
Pure
0
20
40
80
Onset Asynchrony (ms)


In tune
Mistuned
Frequency (Hz)
800
600
400
200
0
0
400
Duration (ms)
0
80
400
Duration (ms)

- Primitive grouping mechanisms based on general heuristics such as harmonicity and onset-time - “bottom-up” / “pure audition”

- Schema-based mechanisms based on specific knowledge (general speech constraints?) - “top-down.

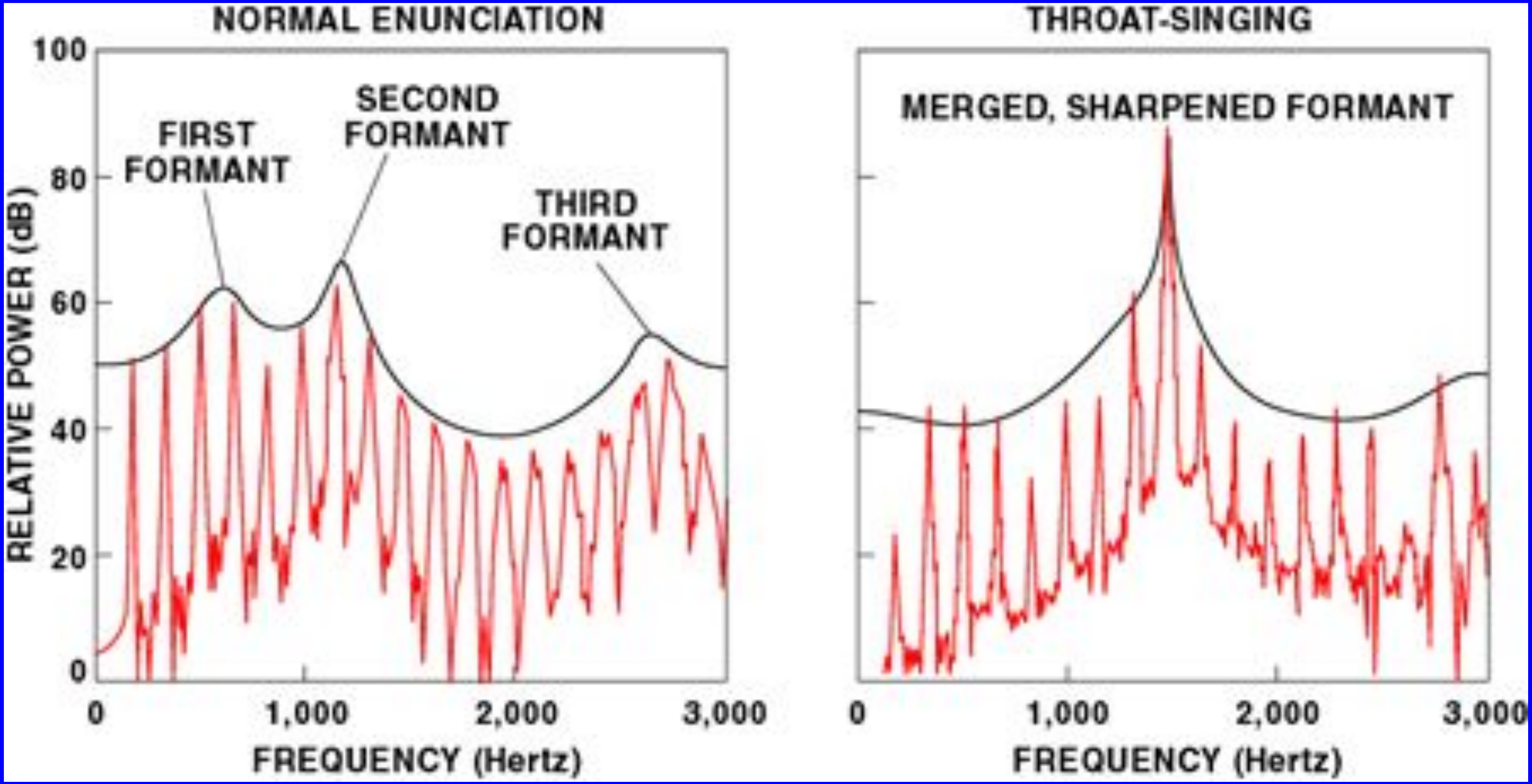
- Orchestra
    - 1° Violin section
        - Leader
            - Chord
                - Lowest note
                    - Attack
    - 2° violins…

Corresponding hierarchy of constraints ?

Multiple sources of sound:

- Vocal folds vibrating
- Aspiration
- Frication
- Burst explosion
- Clicks

Nama: Baboon's arse


NORMAL ENUNCIATION
THROAT-SINGING
100
80
60
40
20
0
RELATIVE POWER (dB)
FIRST FORMANT
SECOND FORMANT
THIRD FORMANT
MERGED, SHARPENED FORMANT
0
1,000
2,000
3,000
FREQUENCY (Hertz)
0
1,000
2,000
3,000
FREQUENCY (Hertz)

Alexei Saryglar
of
Tuva

Sygyt

Recorded at the Cedar Cultural Centre
1-29-99
Mpls, MN USA

© Steve Sklar 1999

Onset-time & continuity only bottom-up cues

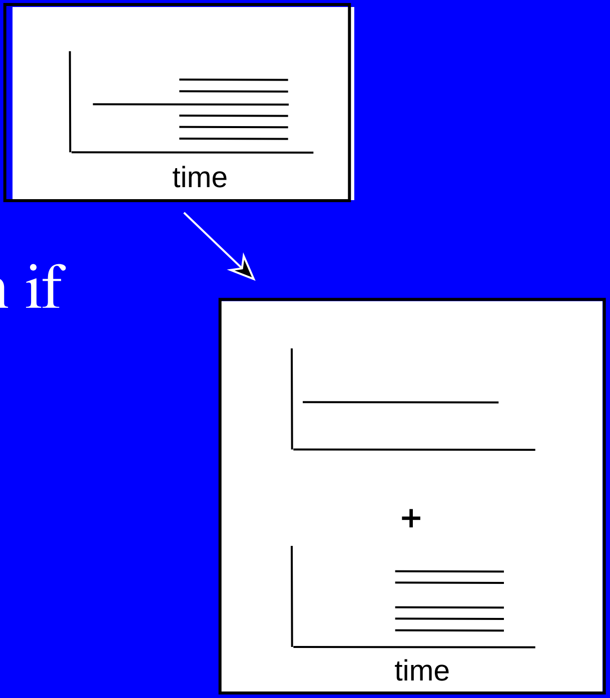
Barker & Cooke, Speech Comm 1999

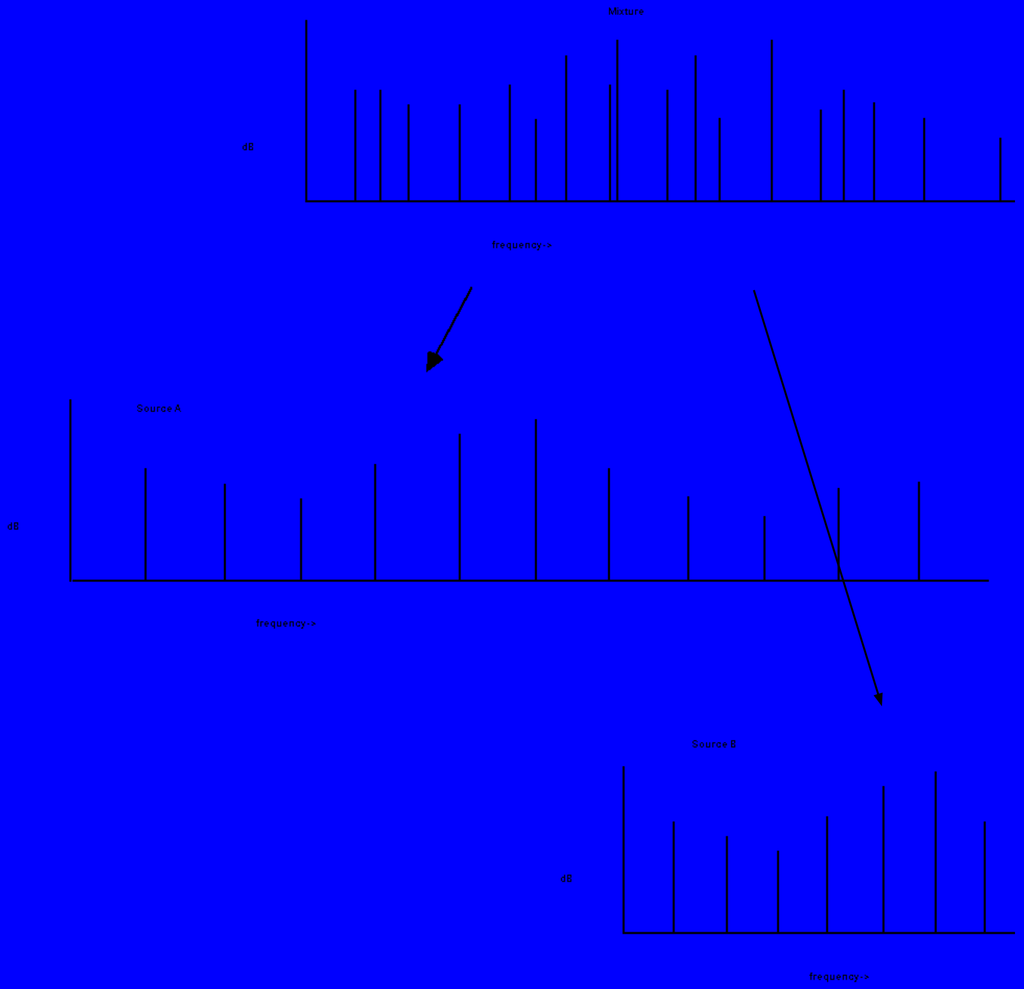
- Bottom-up processes constrain alternatives considered by top-down processes

  e.g. cafeteria model (Darwin, QJEP 1981)

Evidence:

Onset-time segregates a harmonic from a vowel, even if it produces a “worse” vowel (Darwin, JASA 1984)



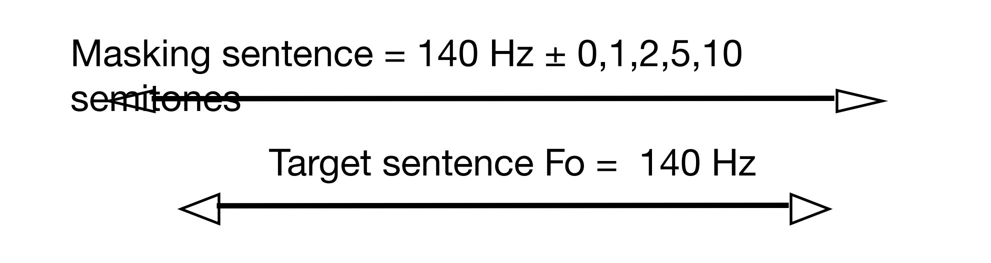


Look for:

- harmonic series
- sounds starting at the same time

Two sentences (same talker)

- only voiced consonants
- (with very few stops)

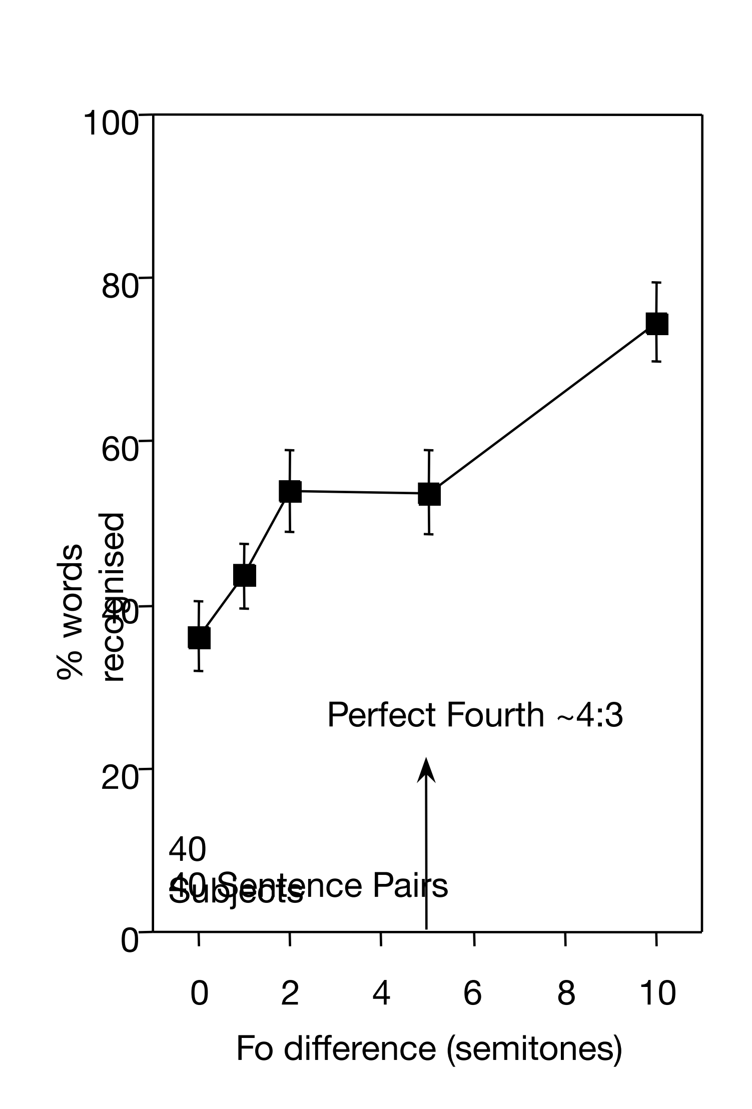


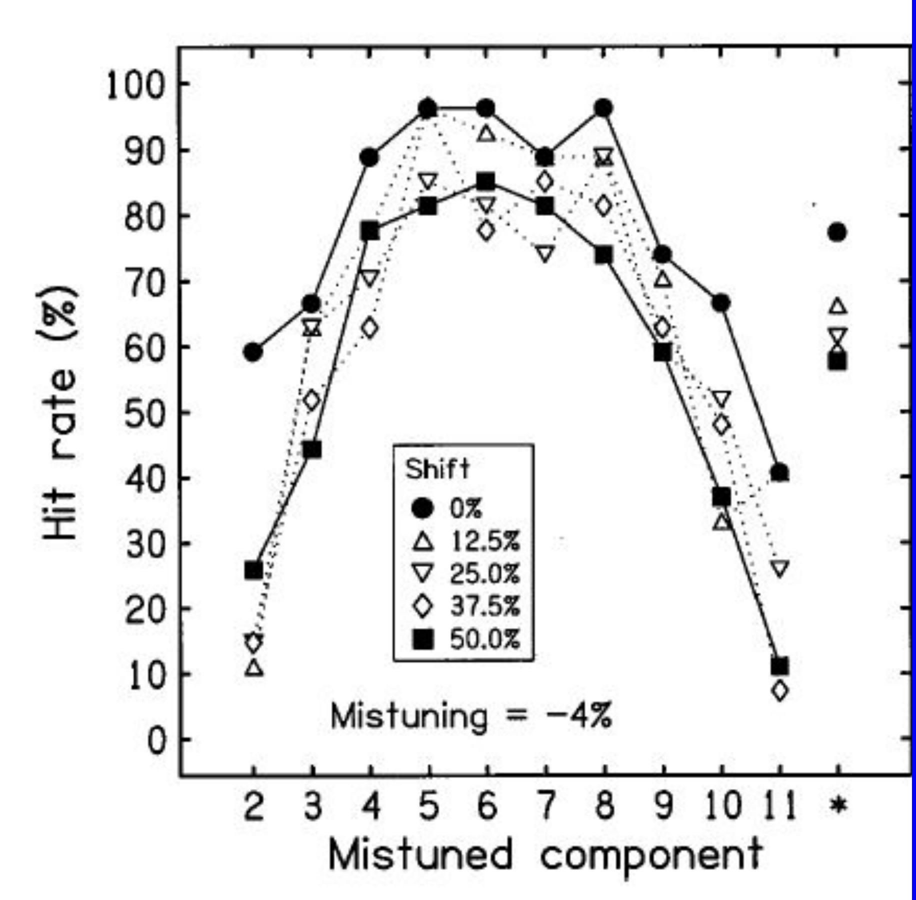
Masking sentence = 140 Hz ± 0,1,2,5,10 semitones

Target sentence Fo = 140 Hz

Task: write down target sentence

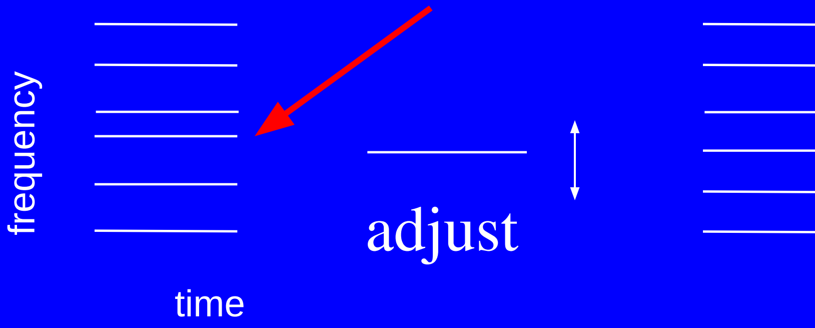
Replicates & extends Brokx & Nooteboom





Roberts and Brunstrom: Perceptual coherence of complex tones (2001)
J. Acoust. Soc. Am. 110

mistuned



adjust

Similar results for harmonic and for linearly frequency-shifted complexes

- Do grouping principles work because they provide some degree of stastistical independence in a time-frequency space?
- If so, why do the parametric values vary with the task?

Cues used by the ASA process

* The perceptual segregation of sounds in a sequence depends upon differences in their frequencies, pitches, timbres (spectral envelopes), center frequencies (of noise bands), amplitudes, and locations, and upon sudden changes of these variables. Segregation also increases as the duration of silence between sounds in the same frequency range gets longer.

* The perceptual fusion of simultaneous components to form single perceived sounds depends on their onset and offset synchrony, frequency separation, regularity of spectral spacing, binaural frequency matches, harmonic relations, parallel amplitude modulation, and parallel gliding of components. [Note to physicists: All these cases of fusion can be obtained at room temperature.]

* Different cues for stream segregation compete to control the grouping, and different cues have different strengths.

* Primitive grouping occurs even when the frequency and timing of the sequence is unpredictable.

* An increased biasing toward stream segregation builds up with longer exposure to sounds in the same frequency region.

* Stream segregation is context-dependent, involving the competition of alternative organizations,

Effects of ASA on perception

* A change in perceptual grouping can alter the perception of rhythms, melodic patterns, and overlap of sounds.

* Patterns of sounds whose members are distributed into more than one perceptual stream are much harder to perceive than those wholly contained within a single stream.

* Perceptual organization can affect perceived loudness and spatial location.

* The rules of ASA try to prevent the crossing of streams in frequency, whether the acoustic material is a sequence of discrete tones or continuously gliding tones.

* Known principles of ASA can predict the camouflage of melodies and rhythms when interfering sounds are interspersed or mixed with a to-be-recognized sequence of sounds.

* The apparent continuity of sounds through masking noise depends on ASA principles. Stimuli have included frequency glides, amplitude-varying tones, and narrow-band noises.

* A perceptual stream can alter another one by capturing some of its elements.

* The apparent spatial position of a sound can be altered if some of its energy becomes grouped with other sounds,

* Comodulation masking release (CMR) does not make the presence of the target more discriminable by simply altering the timbre of the target-masker mixture. It actually increases the subjective experience that the target is present.

* Sequential capturing can affect the perception of speech, specifically the integration of perceptually isolated components in speech-sound identification.

* The segregation of vowels increases when they have different pitches and different pitch transitions. We have looked at synthetic vowels that do or do not have harmonic relations between frequency components,

* ASA principles help explain the construction of music, e.g., rules of voice leading.

* ASA principles are used intuitively by composers to control dissonance in polyphonic music.

* The segregation of streams of visual apparent motion works in exactly the same way as auditory stream segregation.