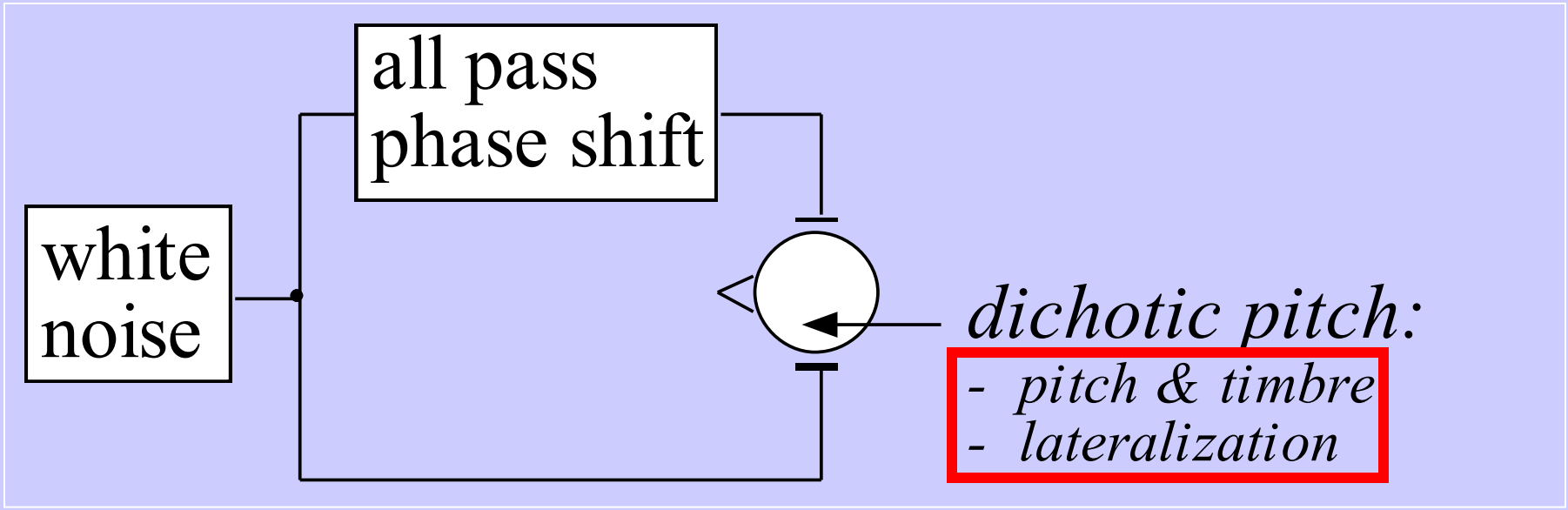
Inferences from Dichotic Pitch for Binaural Modeling

Frans Bilsen

Johan Raatgever

Definition of dichotic pitch (DP):



Psychophysical facts on dichotic pitch (DP):

1. Similarity of pitch and timbre (**parsimony**) for DP signals and natural (ecological) signals

2. **Coupling** of DP-value and -image position

3. DP-image position is **IID insensitive**

Cross Correlation (CC) hypothesis:

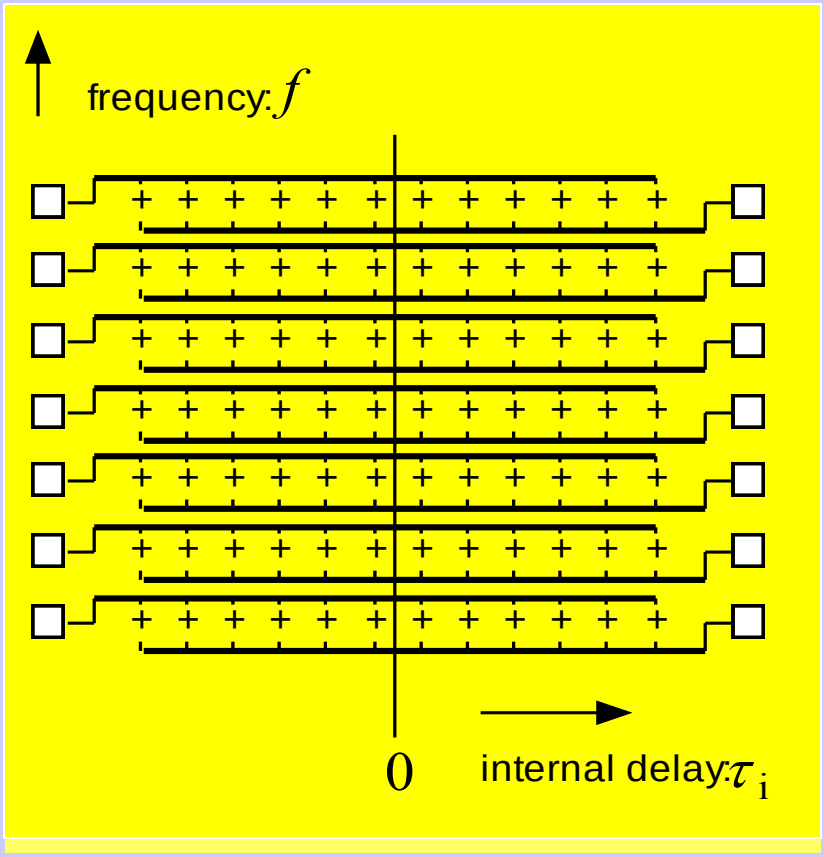
In the spirit (not the detail) of Jeffress' model and Licklider's triplex theory, Fourcin (1962, 1972) tried to explain his findings on Fourcin Pitch with the wide-band cross-correlation function.

Present extensions:

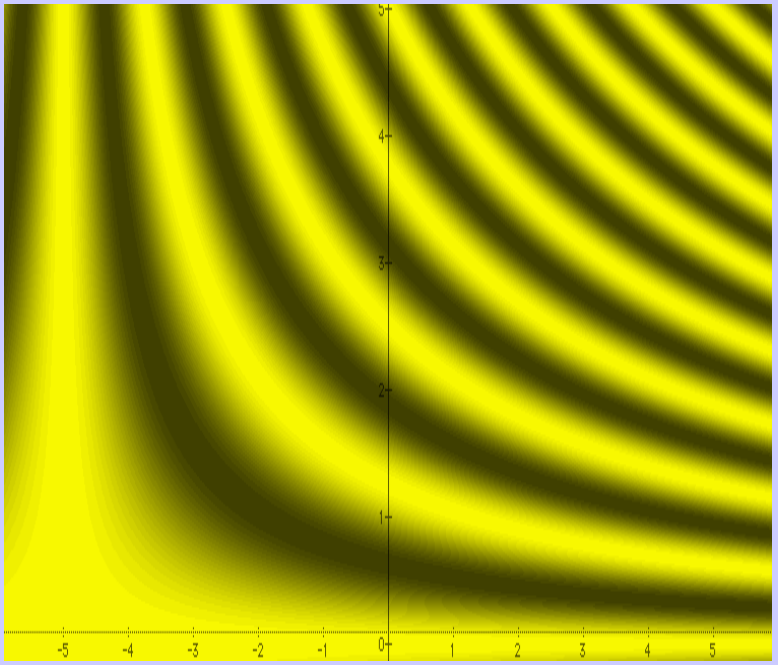
• Similarly, one might consider the possible virtues of a "*Summary Cross Correlo-Gram* (SCCG)", in analogy with the SACG (Meddis, Yost, Patterson).

• Alternatively, one might try narrow-band cross correlation

Central Activity Pattern (CAP):



N_σ



Central Spectrum (CS):

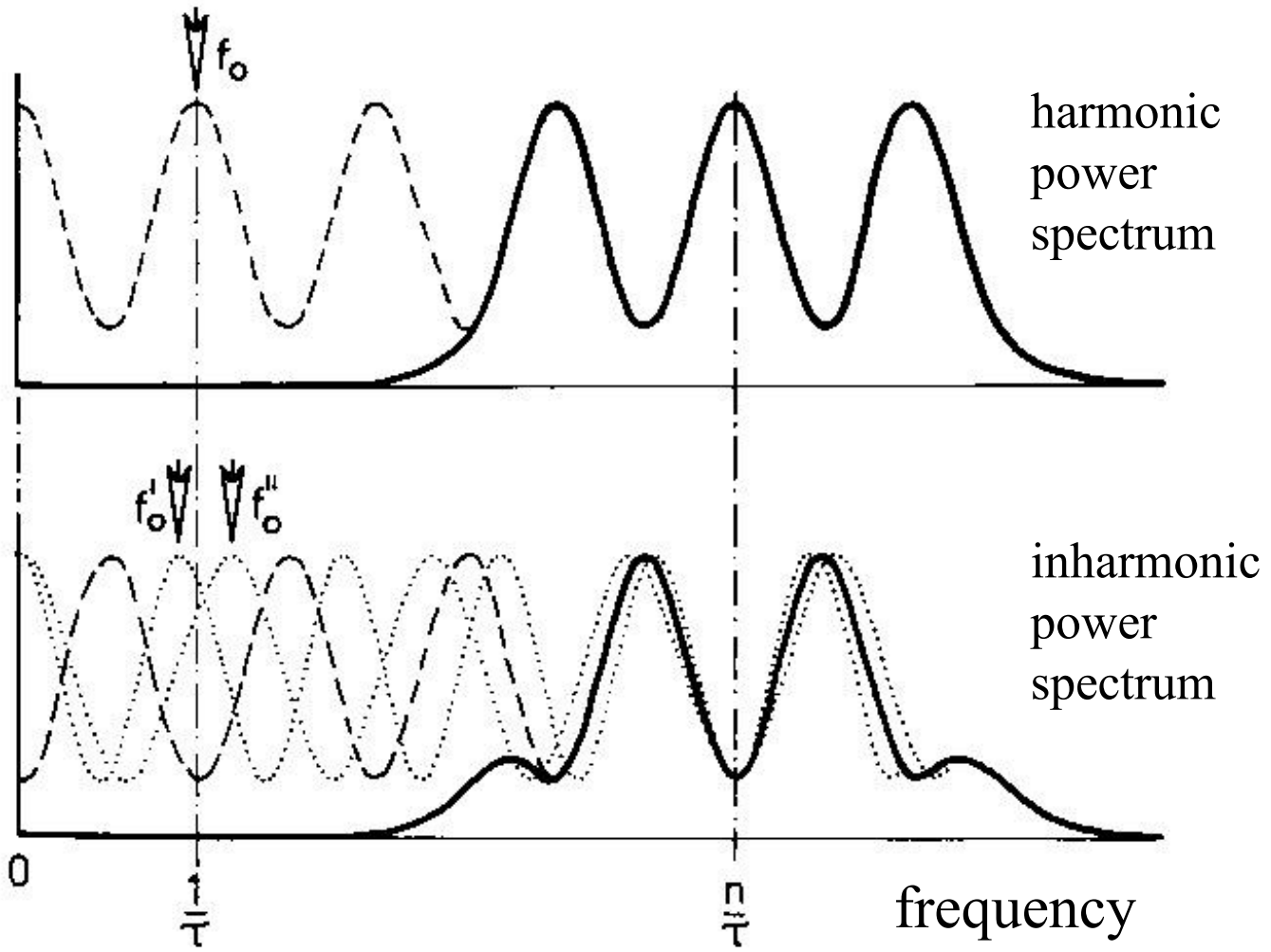
Selection from the CAP of a spectral pattern giving rise to the central sensation of pitch

Selection criteria:

- *resemblance with monaural spectra (**parsimony**)*
- *common internal delay **("straightness")***
- ***infinite** peak-to-valley ratio in the pattern selected*

Note. CAP-CS theory predicts **pitch value** on the basis of either
a) spectral pattern matching on a Central Spectrum selected, or
b) joint auto correlation applied on the time structure of resolved harmonics in a Central Spectrum (compare **SACG**)

Repetition Pitch (Low Pitch, Musical Pitch, Periodicity Pitch, Residue Pitch, Virtual Pitch) :



Equalization Cancellation (EC) model:

Equalization is performed on the left and right ear signals by adjustment of level and/or interaural delay. Then cancellation is performed by addition or subtraction (Durlach, 1972).

Present restrictions and extras:

• The *addition mode* is considered only, because the correct prediction of both pitch and lateralization always calls for addition instead of subtraction.

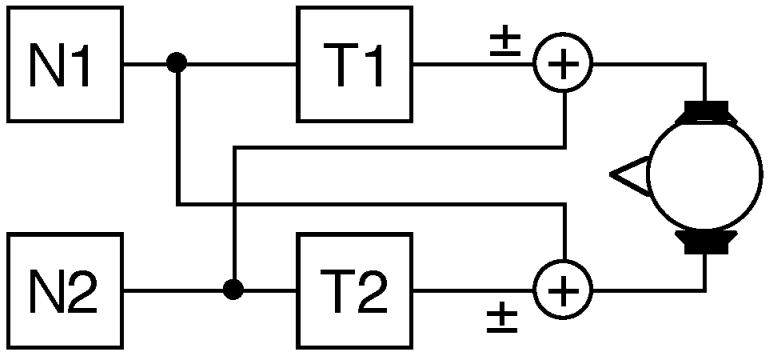
• It is assumed that the EC mechanism, in the absence of a signal, strives for maximum reduction of the noise.

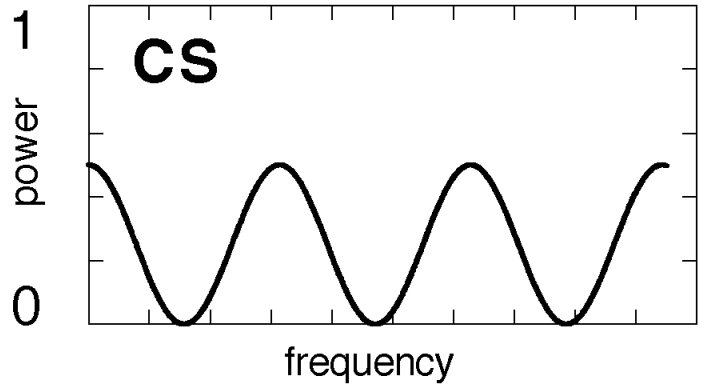
• The equalization parameter might be considered a predictor of position (lateralization)

Modified Equalization Cancellation (mEC):

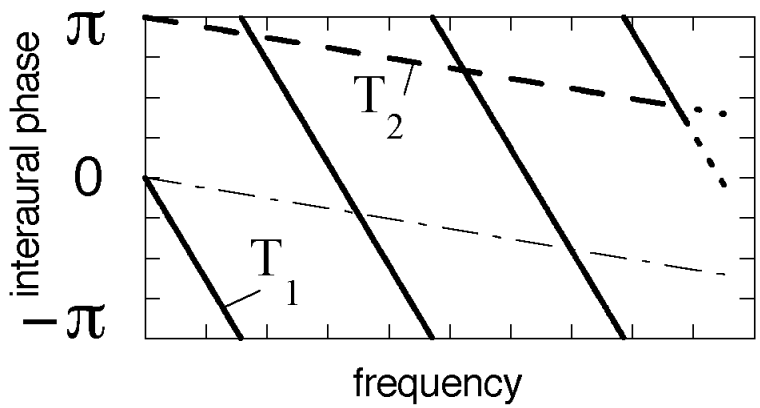
Equalization is performed by adjustment of level and/or interaural delay (up to ± 5 ms) *in each frequency channel (auditory filter) independently*. The residual energy in each filter after cancellation is plotted as a function of center frequency to generate a "*recovered spectrum*", which thus reflects the degree of interaural de-correlation present in each frequency channel (Culling et al., 1998).

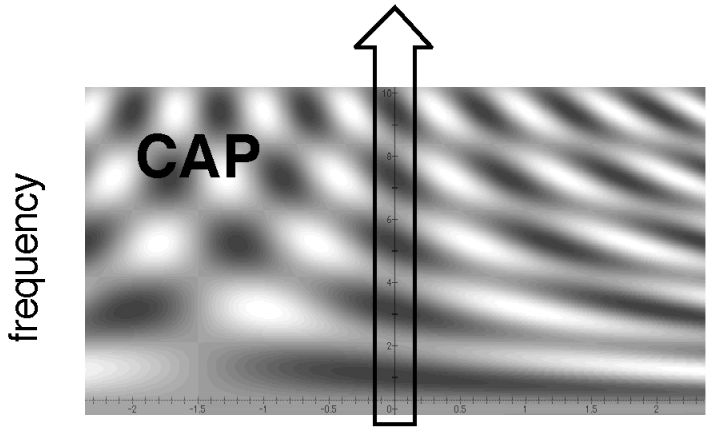
Note 1. Essentially no prediction of position (lateralization)
Note 2. One unique solution for each signal configuration

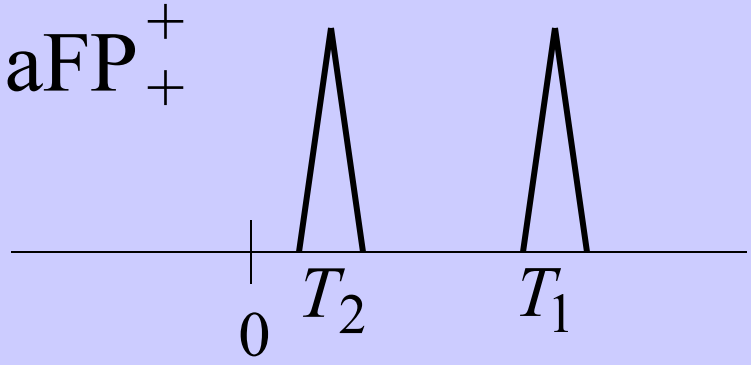

N1
T1
±
N2
T2
±

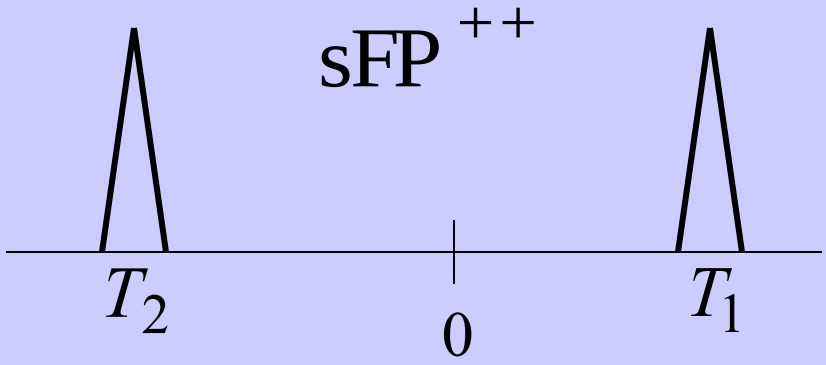

1
power
0
CS
frequency

aFP$^{+}_{-}$


π
interaural phase
0
−π
T2
T1
frequency


frequency
CAP
internal delay





pitch: $\dfrac{1}{T_1 - T_2 \pm 0.8}$

$\dfrac{2}{T_1 - T_2}$

example:

1000/(5.5-0.5±0.8)=173, 238 Hz

2000/(4.5+5.5)=200 Hz

CC:–, EC:–, mEC:+, CAP:+

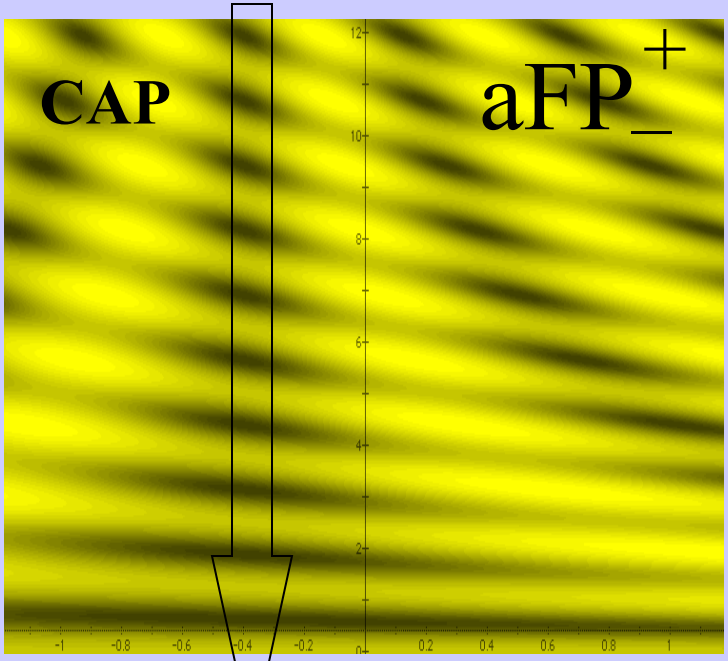
CC:–, EC:–, mEC:–, CAP:+

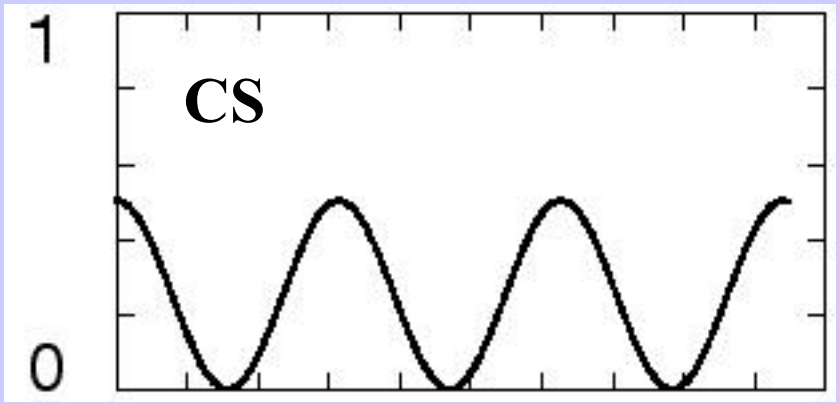
lateralization: $-T_2 \pm 0.8$

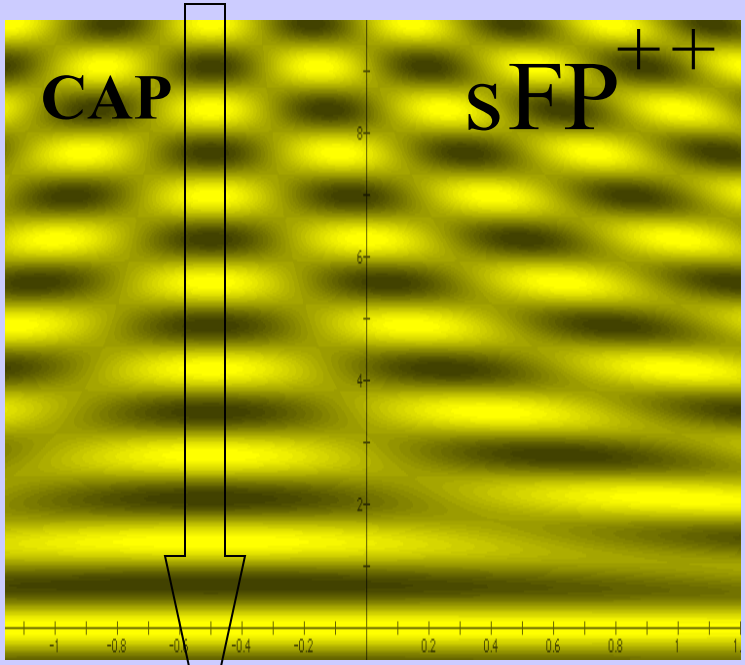
$-\dfrac{T_1 + T_2}{2}$

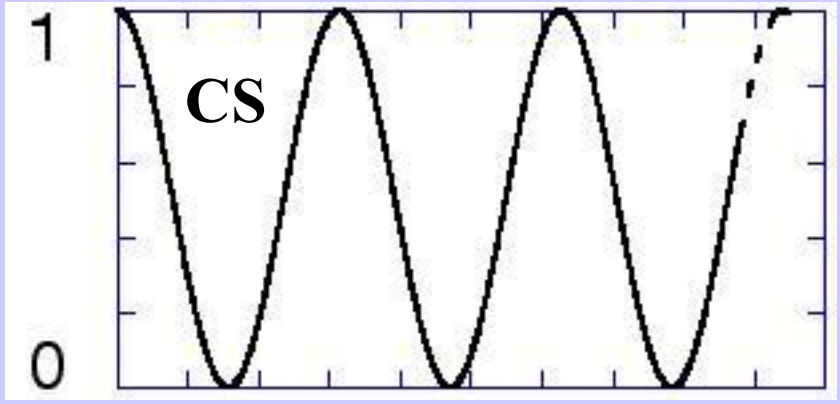
– 1.3, +0.3 ms

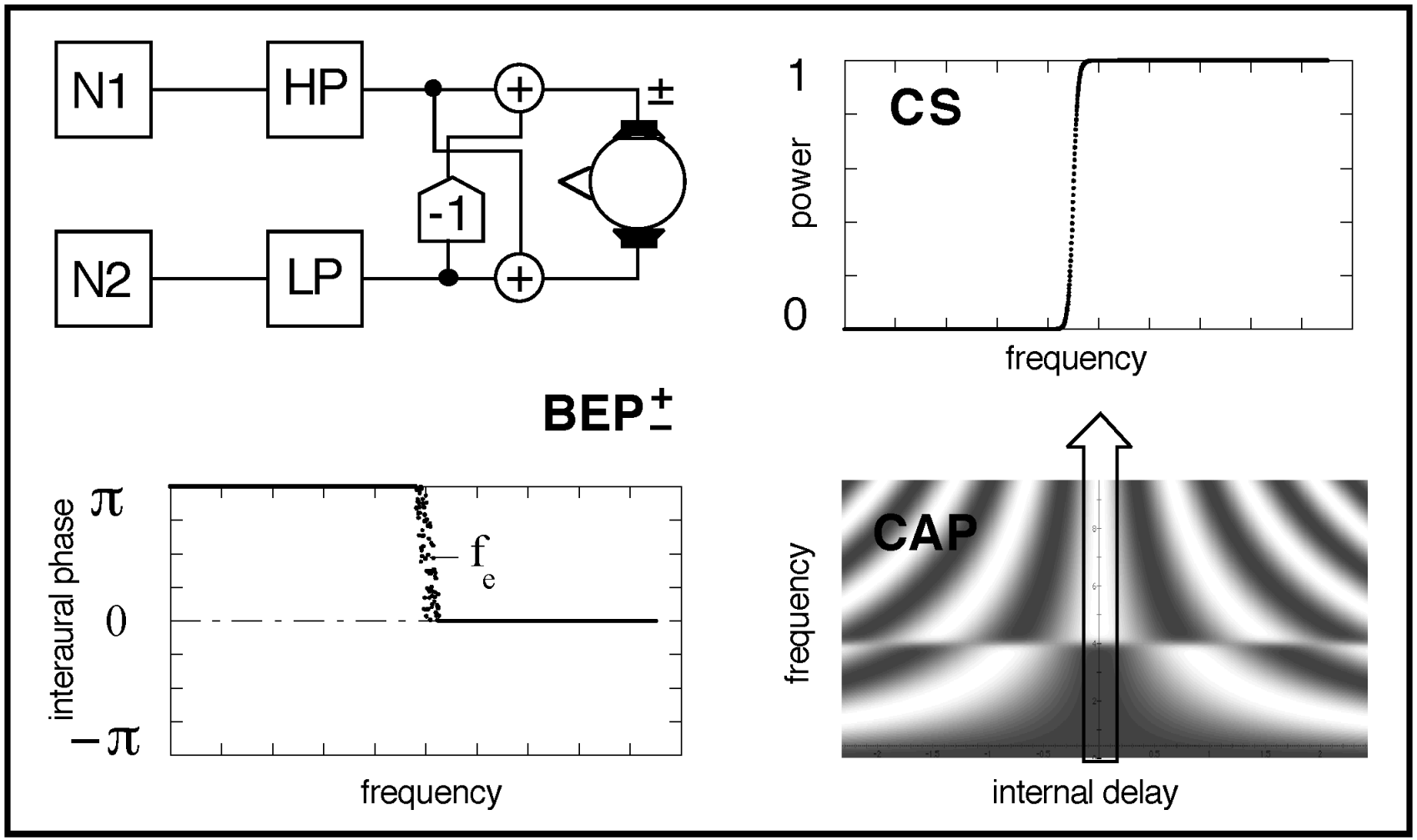
–(4.5–5.5)/2=0.5 ms


CAP
aFP_+

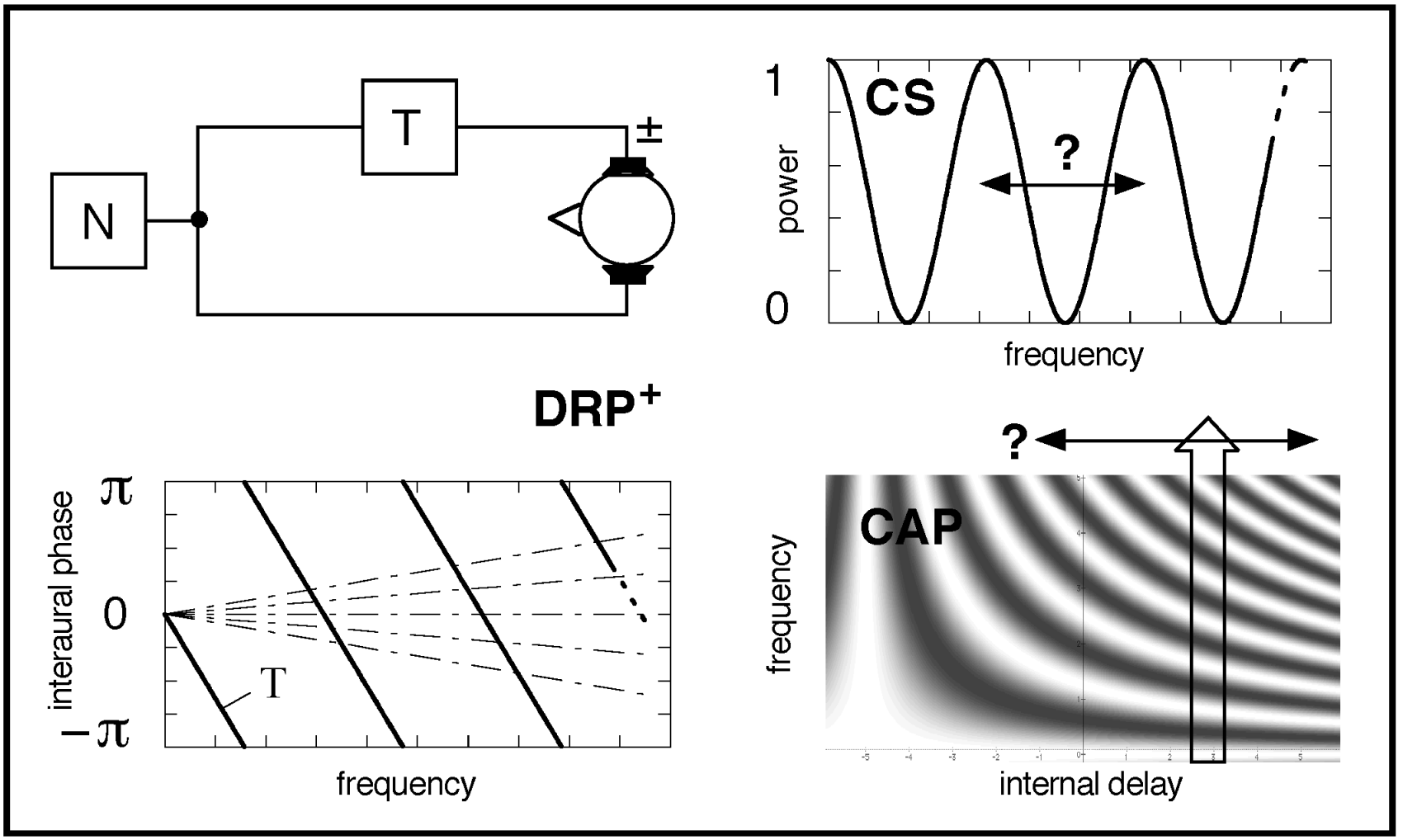

1
CS
0


CAP
sFP++

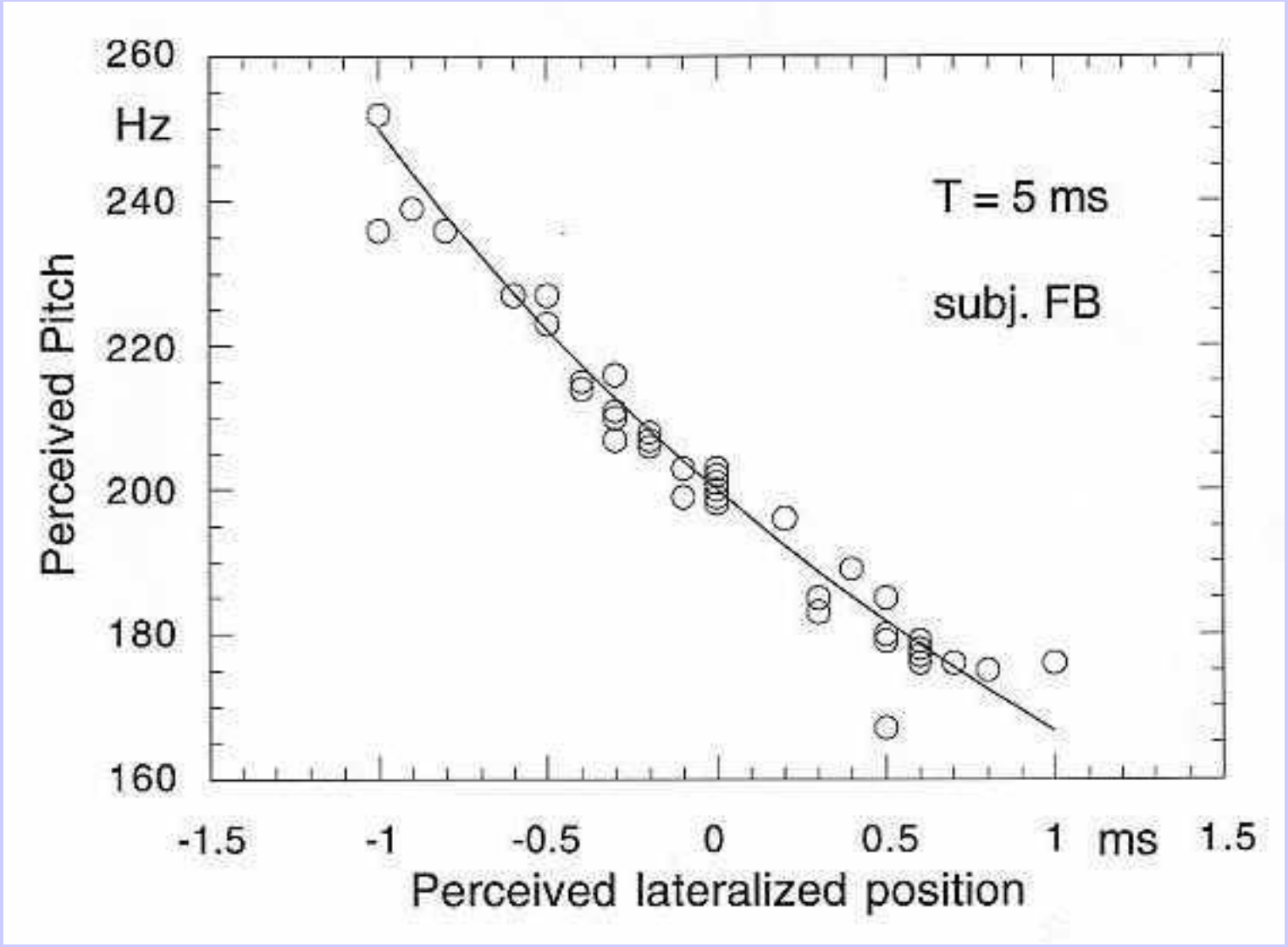

1
CS
0



CC:+, EC:±, mEC: – (vanishing peak), CAP:+



CC:–, EC:–, mEC:– (no recovered spectrum at all), CAP:+


260
Hz
240
220
200
180
160
Perceived Pitch
T = 5 ms
subj. FB
-1.5
-1
-0.5
0
0.5
1
ms
1.5
Perceived lateralized position

Overall evaluation: (pitch, lateralization)

	EC	CC	CAP	mEC	
HP	(+,+)	(+,+)	(+,+)	(+,–)	
MPSP	(–,–)	(+,–)	(+,+)	(+,–)	
aFP	(–,–)	(–,–)	(+,+)	(+,–)	
sFP	(–,–)	(–,–)	(+,+)	(–,–)	:predicts aFP only
DRP	(–,–)	(–,–)	(+,+)	(–,–)	:no CS at all
BEP	(±,±)*	(+,+)	(+,+)	(–,–)	:(vanishing) HP
BICEP	(–,–)*	(–,–)	(+,+)	(+,–)	

* Consistency in pitch and lateralization in addition mode only

Conclusions:

- The psychophysical DP data are predicted **consistently** by the CAP-CS theory only

- Pitch image position is predicted by the **internal delay** of the Central Spectrum selected, **not** by the **SCCG** (as in Licklider's triplex theory)

- Pitch value extraction seems to precede lateralization (compare Darwin c.s.)

INFERENCES FOR BINAURAL MODELING IN GENERAL:

- CC (or SCCG) is not a universal predictor of position

- ITDs and IIDs are processed separately in the "periphery"

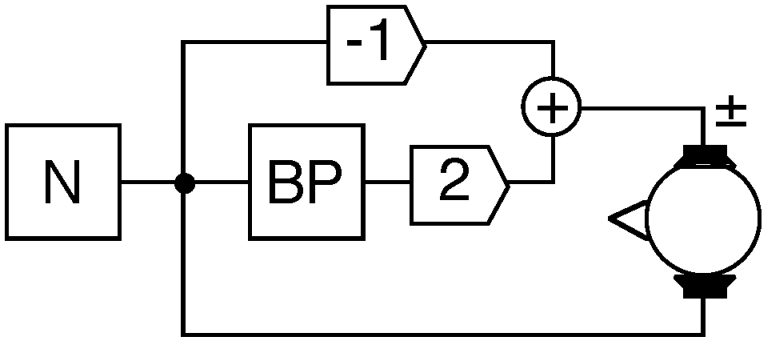
Additional note:
Combination of the EI-cell-based cancellation theory of pitch (compare: **–SACG**) (de Cheveigné) and **EI-** (instead of EE-) cell based binaural interaction for lateralization and detection (Breebaart) seems a possible though not plausible alternative to fulfil monaural and binaural pitch similarity (parsimony)

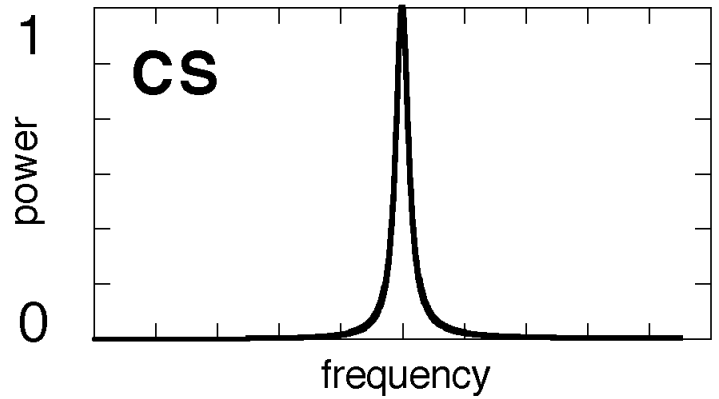
Remaining questions:

- Relations between CS straightness, DP salience and DP-image compactness have to be measured yet
- A low-frequency paradox in binaural pitch (Hartmann)?
- Why exists prevalence for a centrally localized DRP (A separate mechanism)?

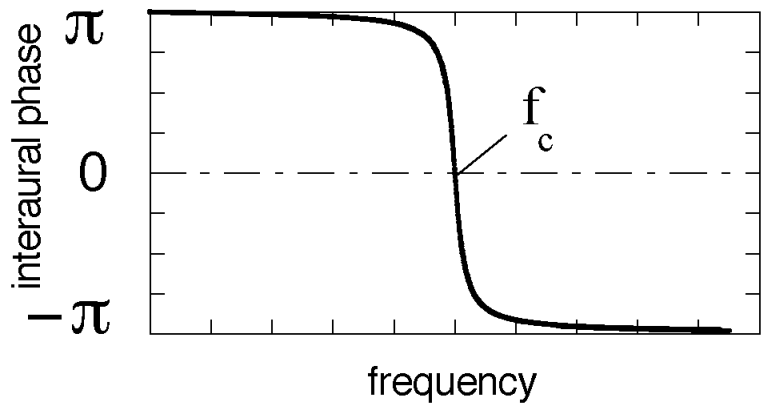
Some logic on the importance of DP phenomena:

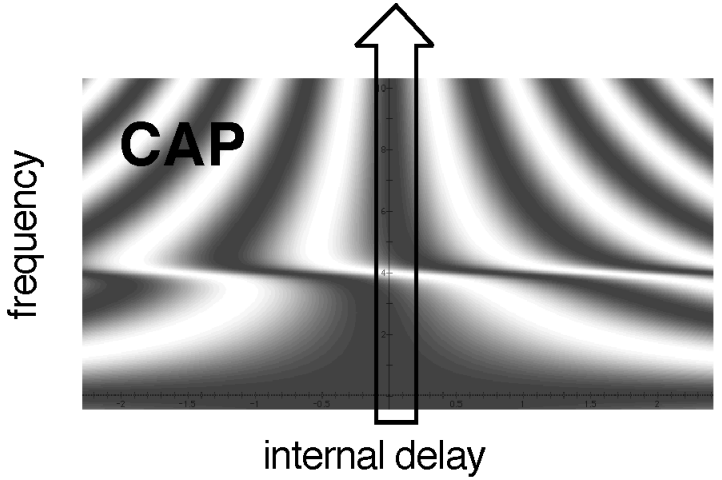
- Parsimony implies DP signals to be processed by the same central pitch processor as ecological signals
- No separate pitch processor for non-ecological signals like DP signals (teleological argument)
- Thus: DP phenomena are natural byproducts of the mechanism of binaural hearing

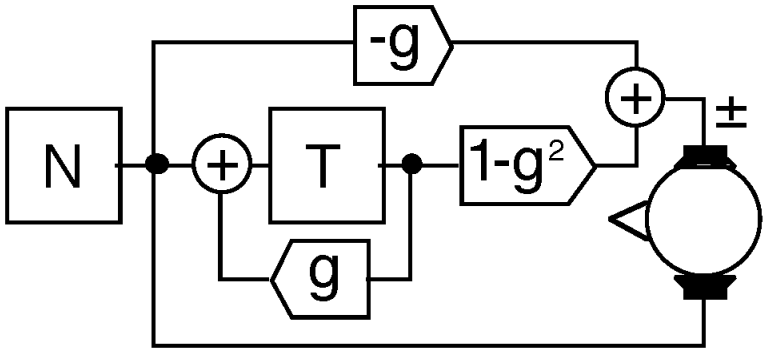

-1
N
BP
2
+
±

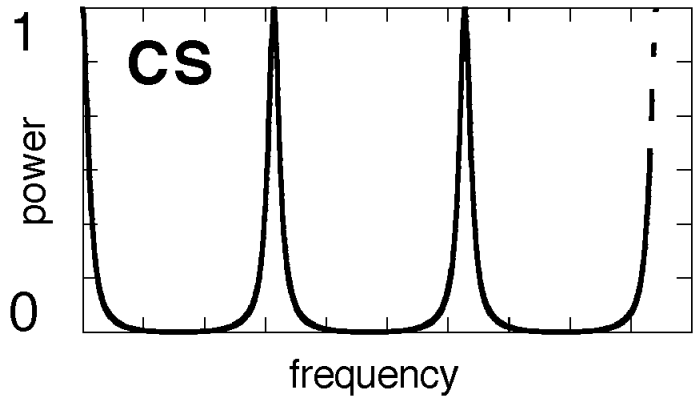

1
power
0
CS
frequency

HP+

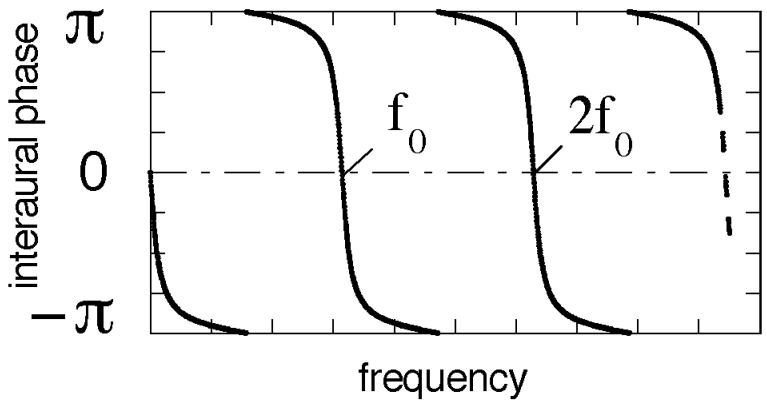

π
interaural phase
0
−π
f
c
frequency

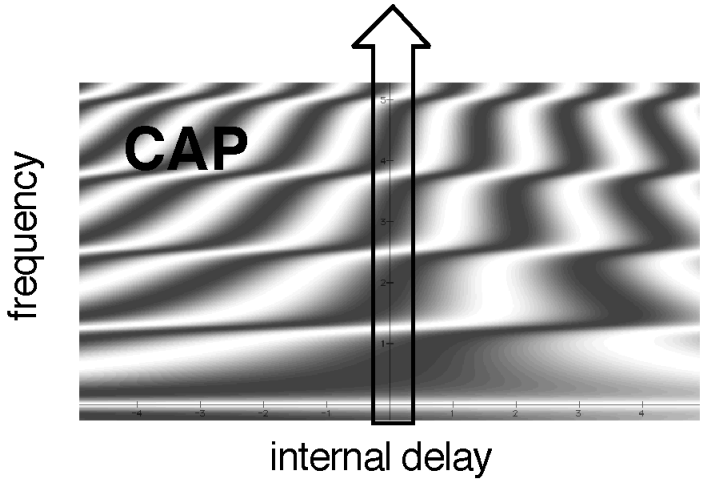

frequency
CAP
internal delay

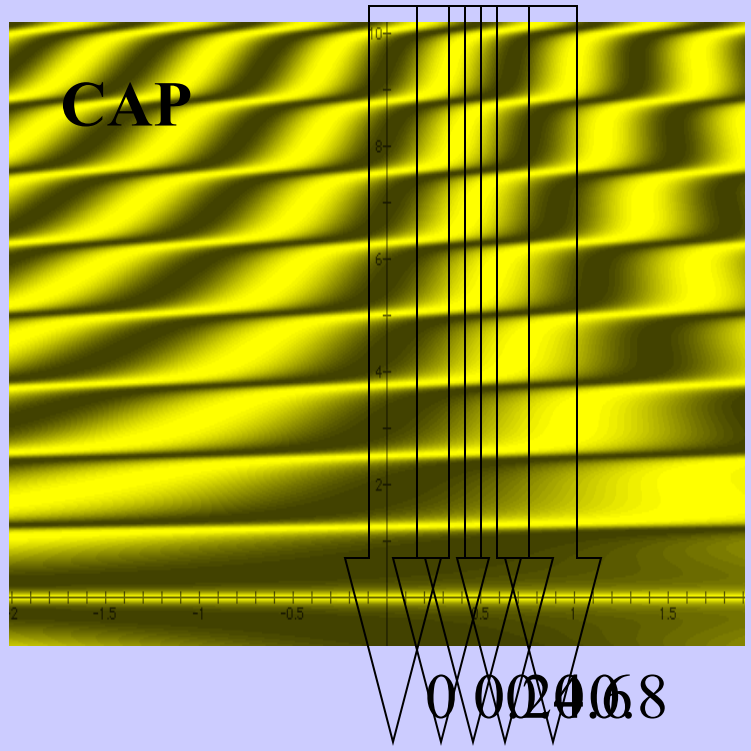

-g
N
T
1-g²
g
±

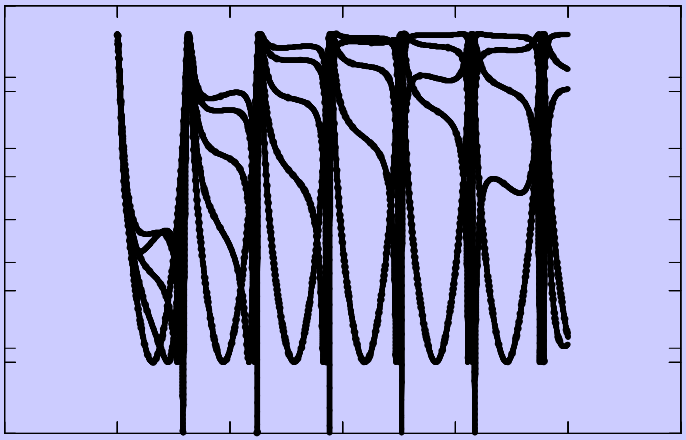

CS
1
power
0
frequency

MPSP +

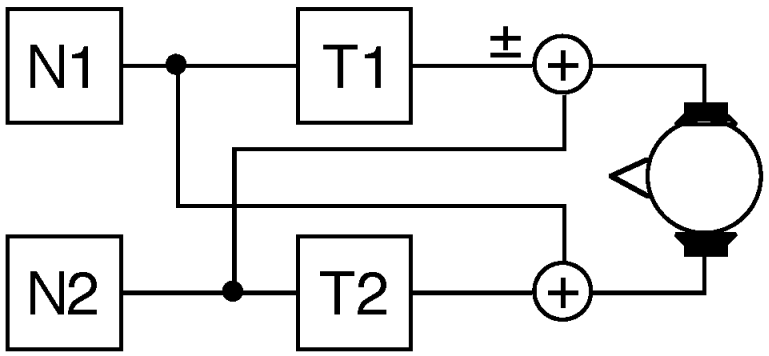

π
interaural phase
0
−π
f0
2f0
frequency

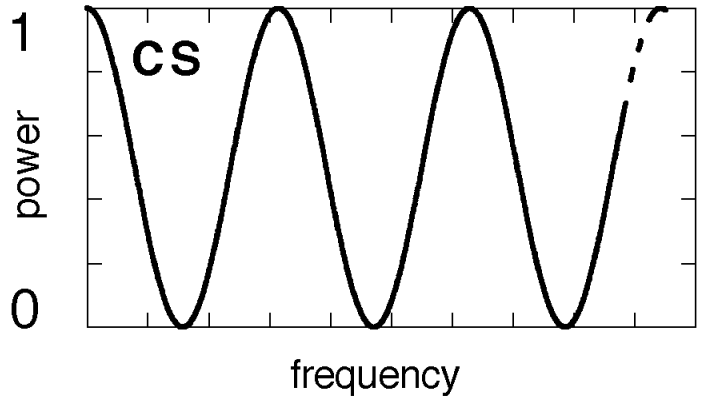

CAP
frequency
internal delay



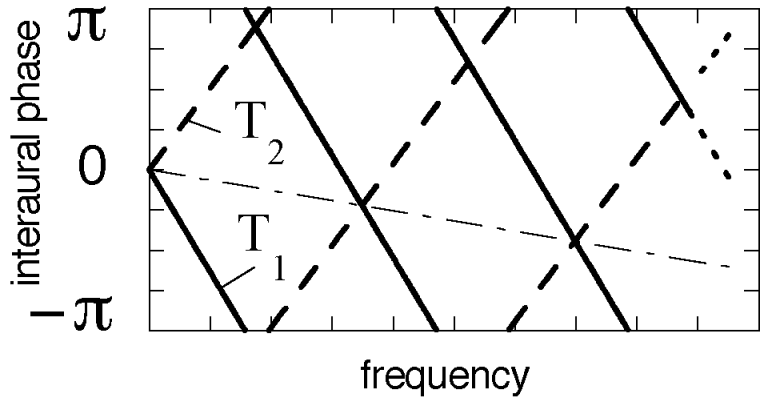


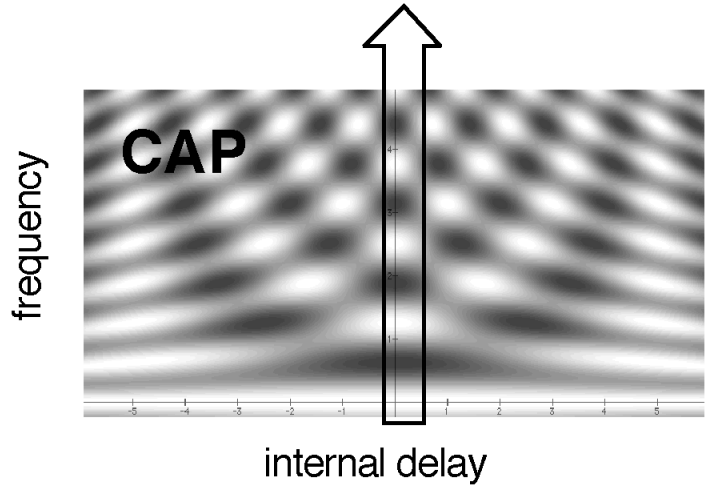
200/220-Hz MPSP interval; listening (diotically) to hypothetical central spectra at different interaural delays: 0.8, 0.6, 0.4, 0.2, 0 ms

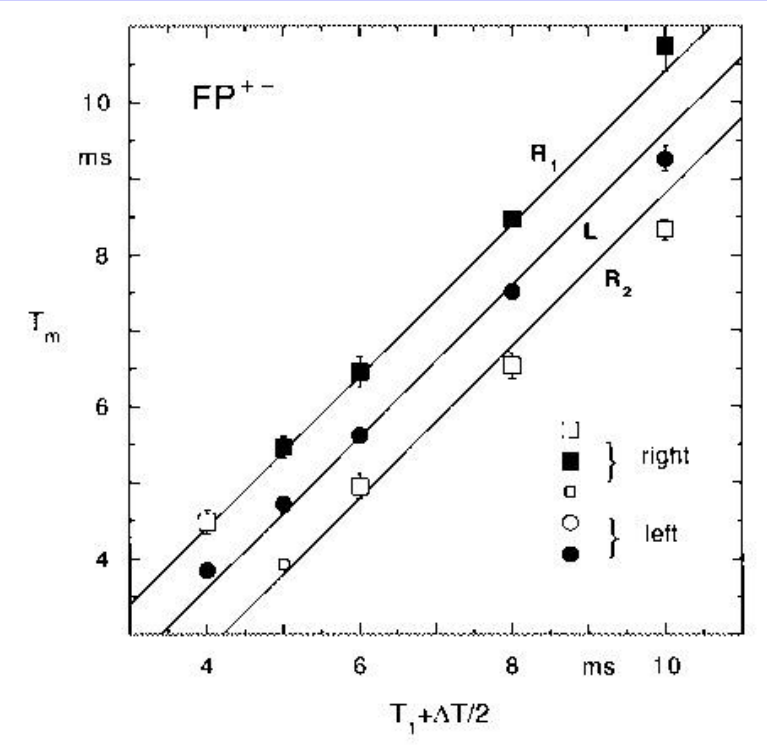

N1
T1
±
N2
T2

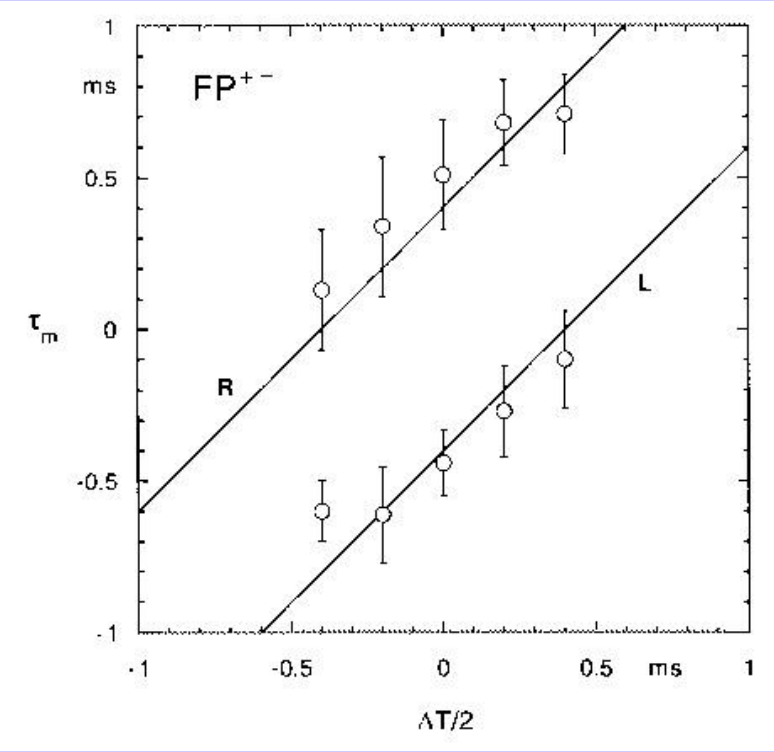

1
power
0
cs
frequency

sFP^{++}


interaural phase
π
0
−π
T2
T1
frequency

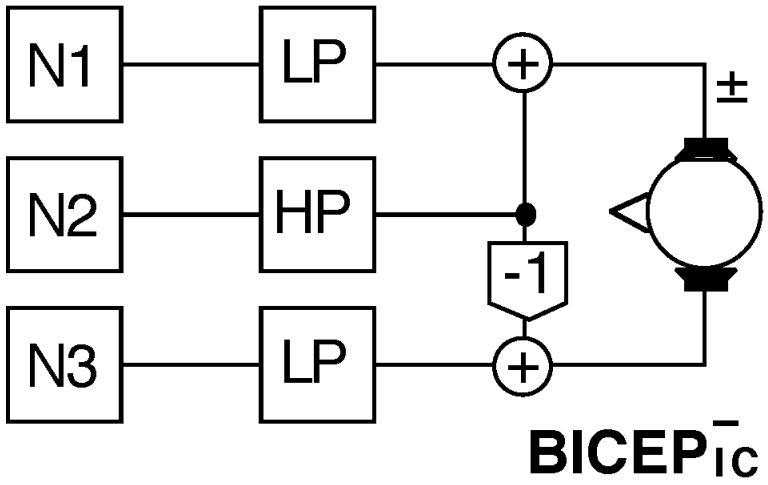

frequency
CAP
internal delay

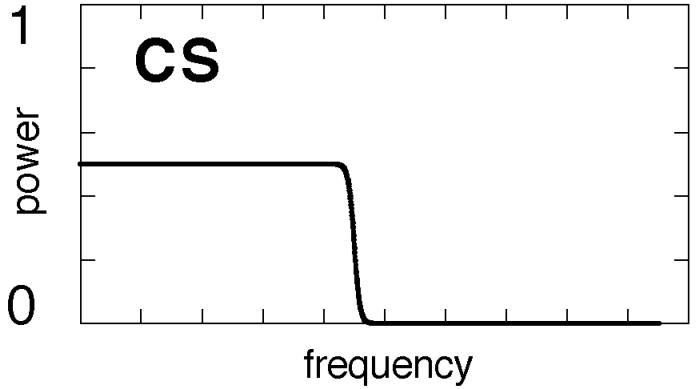


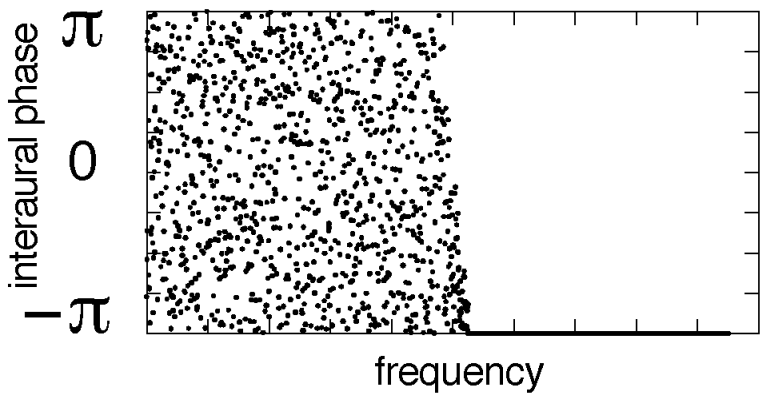


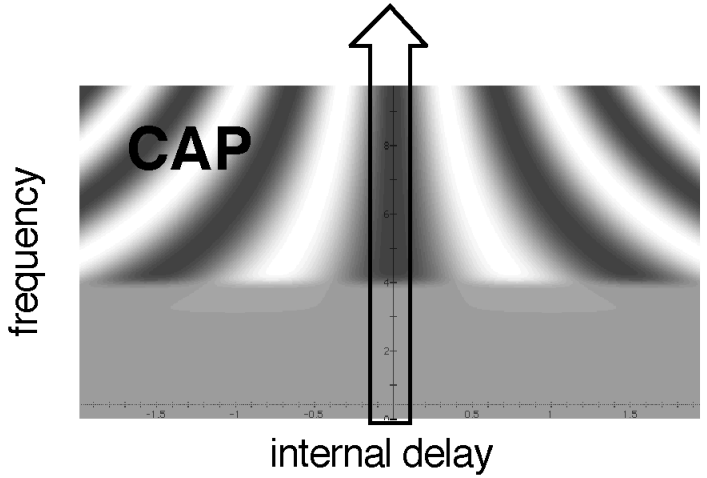
(Bilsen and Raatgever, JASA 2000)

$(\Delta T = |T_1| - |T_2|)$


N1
LP
N2
HP
N3
LP
+
+
±
-1
BICEP_IC


1
CS
power
0
frequency


interaural phase
π
0
−π
frequency


CAP
frequency
internal delay

$$\mathrm{CAP}(f, \tau_i) = [H(f) + \exp\, j2\pi f\tau_i\,]^2$$

$$= 1 + \cos\{\phi(f) + 2\pi f\tau_i\}$$

τ_i internal delay

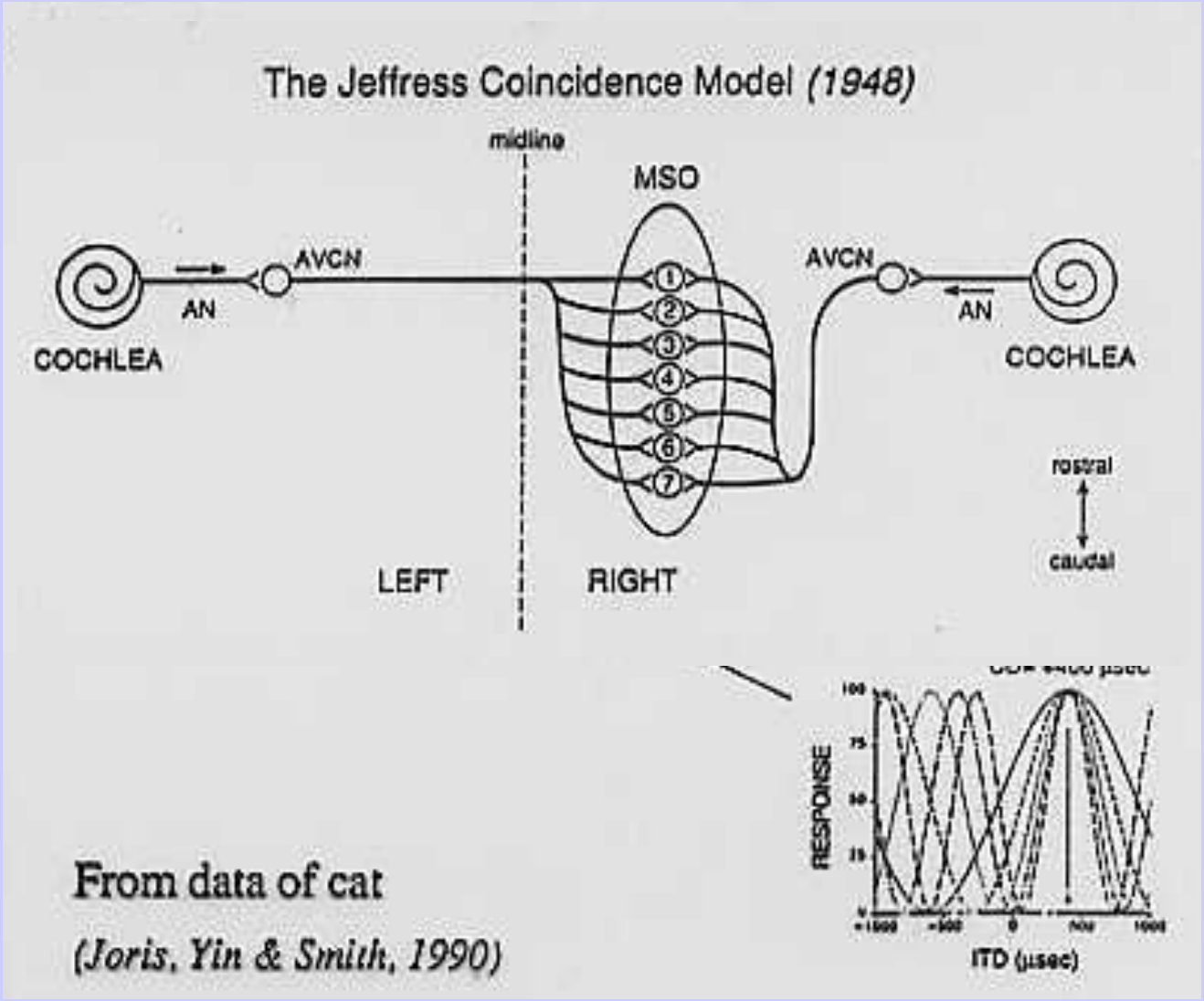
$H(f)$ complex interaural transfer function

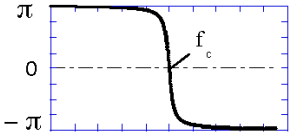
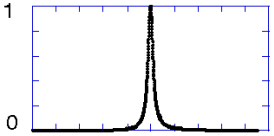
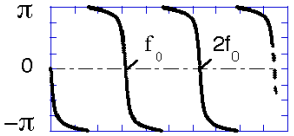
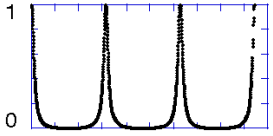
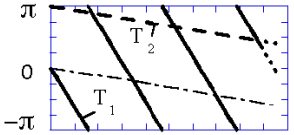
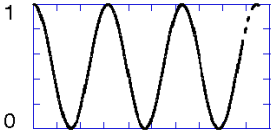
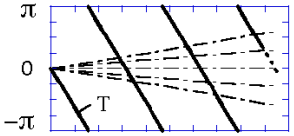
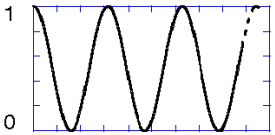
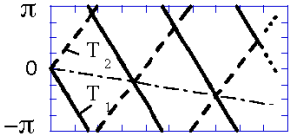
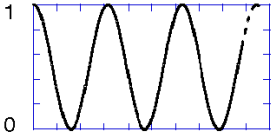
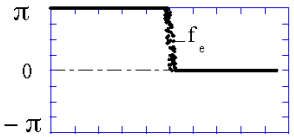
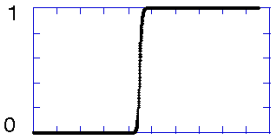
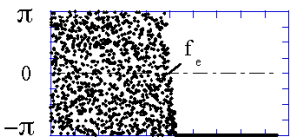
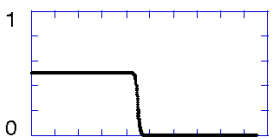
$|H(f)|^2 = 1$ white noise input

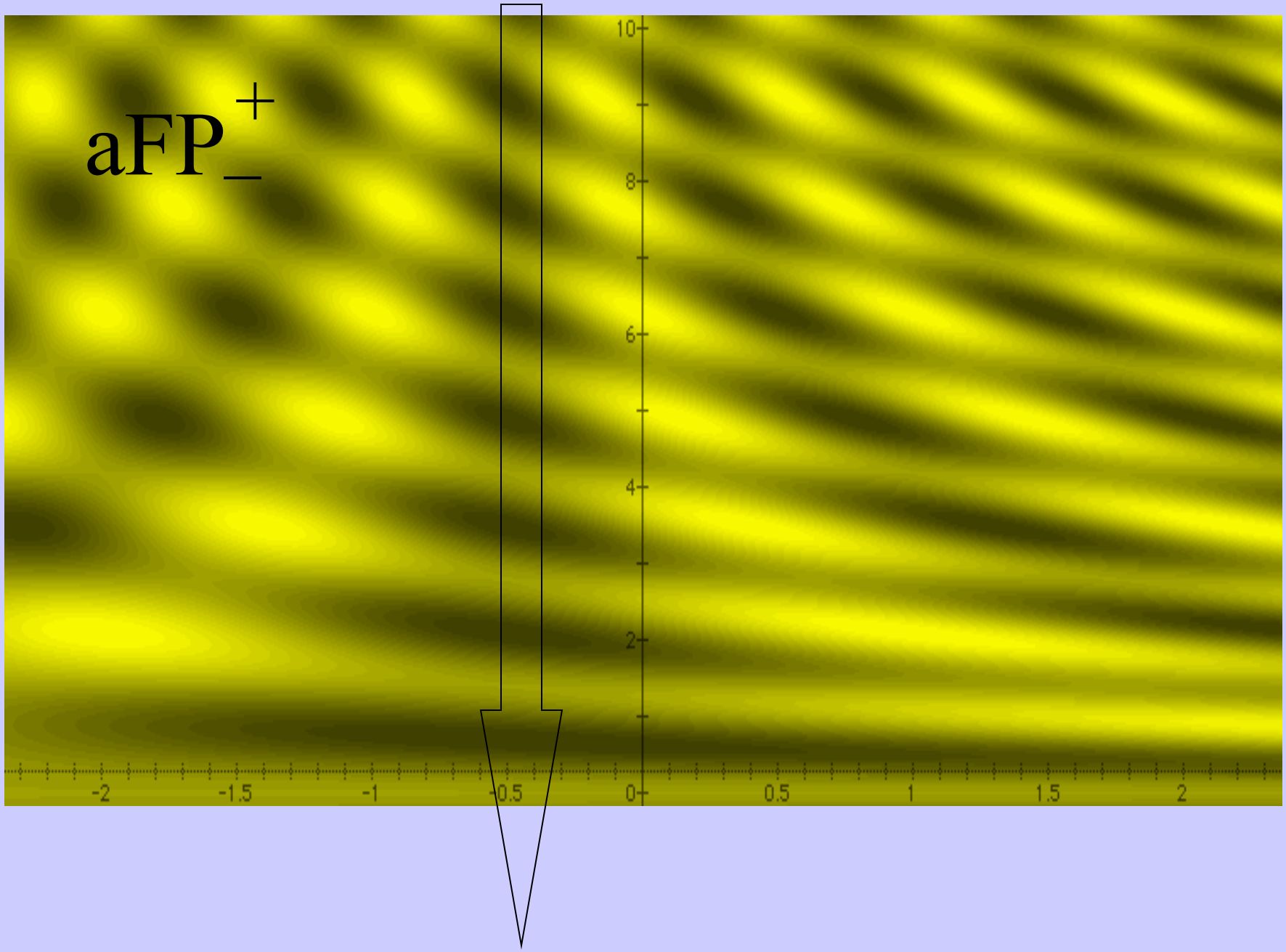
$\phi(f)$ interaural phase function

Note. After introduction of peripheral auditory properties similar pitch values are predicted (Culling et al., 1998)

Binaural Interaction:



Acron.	Interaural phase	Pitch	Lateralization	Central Spectrum	CC	EC	mEC
HP^{+}		f_c	0		+,+	+,+	+,–
$HP^{\pm}$		f_c	$\pm\frac{1}{2f_c}$		+,+	+,+	+,–
$MPSP^{+}$		f_0	0		+,–	+,+	+,–
$MPSP^{\pm}$		f_0	±0.8		+,–	–,–	+,–
$aFP^{+}_{\pm}$		$\frac{1}{T_1 - T_2}$	$-T_2$		–,–	+,+	+,–
aFP^{+}_{+}		$\frac{1}{T_1 - T_2 \pm 0.8}$	$-T_2 \pm 0.8$		–,–	–,–	+,–
DRP^{+}		$\frac{1}{T + \tau_i}$	τ_i		–,–	–,–	–,–
$DRP^{\pm}$		$\frac{1}{T + \tau_i \pm 0.8}$ *	τ_i *		–,–	–,–	–,–
sFP^{++}		$\frac{2}{T_1 - T_2}$	$-\frac{T_1 + T_2}{2}$		–,–	–,–	–,–
$sFP^{+\pm}$		$\frac{2}{T_1 - T_2 \pm 0.8 \pm 1.6}$	$-\frac{T_1 + T_2 \pm 0.8}{2}$		–,–	–,–	–,–
$BEP^{\pm}_{+}$		$f_e \pm \Delta$	$0, \pm\frac{1}{2f_e}$ *		+,+	**±,±**	–,–
$BEP^{+}_{\pm}$		$f_e \mp \Delta$	$0, \pm\frac{1}{2f_e}$ *		+,+	**±,±**	–,–
$BICEP^{\pm}_{ic}$		$f_e - \Delta$	0 *		–,–	+,+	+,–
$BICEP^{ic}_{+}$		$f_e + \Delta$	$\pm\frac{1}{2f_e}$ *		–,–	–,–	+,–


aFP_+
10
8
6
4
2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2