INFERENCES FROM DICHOTIC PITCH FOR BINAURAL MODELING

PACS: 43.66.Pn

Bilsen, Frans; Raatgever, Johan
Applied Physics Department, Delft University of Technology,
Lorentzweg 1,
2628 CJ Delft,
The Netherlands,
Tel: 31.15.3109603
E-mail: f.a.bilsen@tnw.tudelft.nl

ABSTRACT

A dichotic pitch can be perceived due to the binaural presentation of white noise with a particular interaural phase relationship. Three perceptual aspects are characteristic: the pitch value, the timbre of the phenomenon, and the lateralized position of the pitch image. As dichotic pitch phenomena are generally considered to be natural epiphenomena of the mechanisms of binaural hearing, inferences for binaural modeling can be drawn from their behavior. Current models like Cross Correlation, Equalization Cancellation (EC), modified EC, and the Central Spectrum theory (CAP-CS) were inspected for their ability to explain the main aspects of dichotic pitch. It turns out that only CAP-CS theory is able to predict all existing data in a consistent way. Typical examples will be highlighted during the presentation.

INTRODUCTION

A dichotic pitch can be perceived when white noise is presented binaurally by headphones to a listener, provided that a particular interaural phase relationship exists between the left-ear and right-ear signals. In other words, the information at either ear independently is only able to evoke a sensation of white noise, but the stimulation of both ears together produces the sensation of pitch. Generally, a dichotic pitch is perceived somewhere in the head amidst the noisy sound filling the binaural space; roughly, its image is "lateralized" along the imaginary axis connecting the left and right ear.

In summary, each dichotic pitch phenomenon is characterised by three perceptual properties depending on the specific interaural parameter, viz. pitch value, timbre, and in-head position (lateralization) of the pitch image.

With analogue noise as basic signal, two types of interaural phase relationship have been reported in the literature, namely sharp 2ð (or ð) phase transition(s) in limited frequency range(s), or relatively large (> 3 ms) interaural time delays. Such interaural phase relations do not occur in daily life; they are non-ecological, as they are the result of signal generation in the laboratory. Such signals are the more non-ecological as the amplitude spectra of both the left and right ear channels are always made flat to avoid monaural pitch information.

This immediately might raise the question whether phenomena evoked by non-ecological signals should be seriously considered. As has been repeatedly shown in the past, dichotic

pitch phenomena have perceptual properties (pitch and timbre) that are very similar to those of daily-life signals. This implies that the human central pitch processor seems to deal with pitch information always similarly, independently from the place in the auditory system where the information originates, be it in the outer ear or in the auditory brainstem.

Pitch information, apparently, can be either monaural (monotic or diotic), or binaural (dichotic) in origin. Thus, the system is thought to be parsimonious and unique in its pitch processing strategy. Moreover, one cannot think of any reason for the human brain to have developed a separate pitch processor for non-ecological signals. Therefore, dichotic pitch phenomena are generally considered to be *natural byproducts (epiphenomena) of the binaural system*, and thus they can be used to study this system as well as the central pitch processor.

In addition to the localisation of sound sources, one of the most important and intriguing capabilities of the binaural system is its ability to single out a wanted sound (e.g. a communicating human voice) from disturbing noises that are present at the same time in the acoustic environment. Only little is known of the underlying neural processes of this so-called cocktail-party phenomenon. In view of the apparent analogy (the dichotic pitch being singled out from the disturbing head-filled noisy sound), it seems obvious to exploit dichotic pitch phenomena for the study of localisation and the cocktail-party phenomenon together with the underlying binaural-interaction system.

EXPERIMENTAL FACTS ON DICHOTIC PITCH

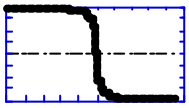
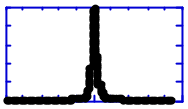
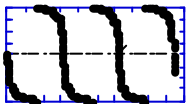
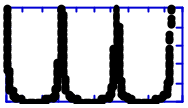
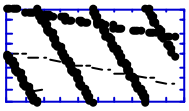
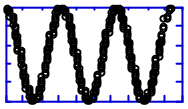
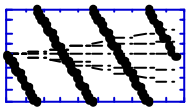
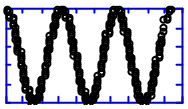
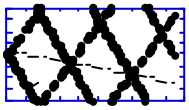
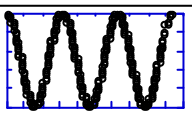
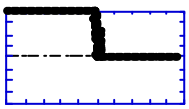
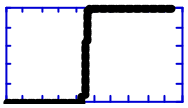
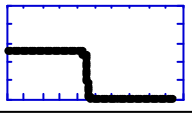
Acron	Interaural phas	Pitch	Lateralizatio	Central Spectrum	CC	EC	mEC
HP^{+}		f_c	0				
HP^{-}		f_c	$\pm\frac{1}{2f_c}$				
$MPSP^{+}$		f_0	0				
$MPSP^{-}$		f_0	± 0.8				
aFP^{+}_{-}		$\frac{1}{T_1 - T_2}$	$-T_2$				
aFP^{+}_{+}		$\frac{1}{T_1 - T_2 \pm 0.8}$	$-T_2 \pm 0.8$				
DRP^{+}		$\frac{1}{T +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

*Fig. 1. Dichotic pitch data "summarized" by Central Spectrum equations; frequency (f) in kHz and time (T, t) in ms; * no data available. CC, EC (addition), and mEC model performance is expressed by + (correct), – (incorrect) for pitch, lateralization respectively.*

The following dichotic pitches have been reported in the past: the *Huggins Pitch* (HP) for a 2ð-phase transition in a limited frequency range [1], the *(a)symmetric Fourcin Pitch* (aFP and sFP)

for two uncorrelated noises with interaural delays T_1 and T_2 [2, 3], the *Dichotic Repetition Pitch* (DRP) for only one single interaural delay T [4], the *Multiple Phase Shift Pitch* (MPSP) for a series of 2ð-phase transitions equally spaced in frequency [5], the *Binaural Edge Pitch* (BEP) for a ð-phase transition in a limited frequency range [6] and the *Binaural Coherence Edge Pitch* (BICEP) [7]. Acronyms and interaural phase configurations are summarized in Fig. 1 (columns 1 and 2).

HP, BEP and BICEP have a pure-tone character, while DRP, FP and MPSP behave like a "low" pitch (compare periodicity pitch, virtual pitch, residue pitch, repetition pitch). In addition, a dichotic pitch has a more or less well-defined binaural image separated, in general, from the (diffuse) image of the generating dichotic noise itself. As both pitch value and pitch image position (lateralization) have been shown to be correctly predicted by the Central Spectrum (CS) theory [3, 8, 9, 10, 11] (see below), existing data are “summarized” by CS equations in columns 3 and 4 of Fig. 1.

CENTRAL SPECTRUM MODEL

A successful theoretical concept to explain dichotic pitch is the Central Spectrum (CS) theory [8]. Based on cochlear frequency analysis and Jeffress' binaural cross correlation network, it calculates a *"Central Activity Pattern (CAP)"* as a function of frequency (*f*) and internal delay (t_i). The central pitch processor scans this CAP for familiar spectral patterns. For example, a sharp isolated peak will give rise to a pure-tone-like pitch. A well-modulated periodic spectral pattern at a particular internal delay will give rise to a “low” pitch comparable to repetition pitch or periodicity pitch (residue pitch, virtual pitch). In general, the pattern selected, the *"Central Spectrum (CS)",* is claimed to predict the value of the pitch; the internal delay (t_i) where the pattern is found, determines the perceived lateral position (lateralization) of the pitch image.

The CS theory was devised for qualitative understanding rather than for exact quantitative prediction of central spectra. Detailed physiological and psychophysical knowledge of the peripheral hearing organ is not built in, although peripheral filtering is included formally in the original formulation of the model. Its elegance still is its mathematical simplicity, providing insight with a minimum of calculus. Because of its success in the past, and also for didactical reasons, calculations were confined to the idealised case of infinitely sharp frequency analysis. Also temporal jitter in the cross correlation process is not included.

The selection mechanism for a Central Spectrum to be a serious candidate as predictor of pitch was not mathematically specified in the original formulation of the model [8]. Instead, the following *selection criteria* for the scanning process as described above were assumed:

(1) Resemblance with familiar (monaural) spectral patterns, for example, a single isolated spectral peak, or a series of equidistant peaks,
(2) Common internal delay ("straightness") for a series of spectral peaks,
(3) Maximum modulation depth in the spectrum selected. In the idealised formulation, this requirement will simply be fulfilled by claiming an *infinite peak-to-valley ratio* (or synonymously: an infinite level difference between peaks and valleys on a log scale).

Assuming *idealised* frequency analysis the (normalized) Central Activity Pattern (CAP) can be expressed in three alternative ways by

$$\mathrm{CAP}(f, t_i) = [H(f) + \exp\ j2\pi f t_i]^2, \quad (1a)$$

$$= 1 + \mathrm{Re}\{S_{rl}(f) \exp\ j2\pi f t_i\}, \quad (1b)$$

$$= 1 + \cos\{\varphi(f) + 2\pi f t_i\}, \quad (1c)$$

with *f* frequency and t_i internal delay. $H(f)$ represents the complex interaural transfer function with $|H(f)|^2 = 1$ for white noise as input, $S_{rl}(f)$ the cross-power spectral density, and $\varphi(f)$ the interaural phase relationship. Substituting the t_i value(s) for which the above selection criteria are fulfilled, central spectra (CS) are obtained as shown for each case in Fig. 1, column 5.

In the past, some experiments were devoted to the notion that dichotic pitch images behave like *"time images"* [3, 8, 13], as they show hardly any sensitivity to interaural intensity differences (IIDs). The CS theory is in agreement with this experimental fact, because IIDs only appear to affect the modulation depth of central spectra resulting in a decrease of the salience of the pitch, not its value nor its intracranial position (compare Eq. 1). In contrast, the image(s) belonging to the noise stimuli itself, substantially are affected by an IID.

In accordance with CS theory, pitches and their lateralizations can already be prognosed from the interaural phase patterns (column 2) by inspecting the dash-dotted lines. Being straight and going through the origin (0 phase, 0 frequency), these lines symbolise an internal delay t_i (similar to an interaural delay *T*). For example, for HP^+ and $MPSP^+$ the intersection with the phase pattern indicates the value and position of peaks in the central activity pattern (CAP) at $t_i = 0$. For aFP^+ the dash-dotted line runs parallel to the dashed line T_2 and shifted by ð, thus indicating a straight valley of zero power from noise 2 in the CAP at $t_i = T_2$, which "highlights" the central-spectrum part due to noise 1 at this internal delay. Different highlighting is obtained in the case of sFP^{++} by the additive interference of T_1 and T_2 at t_i.

Such highlighting is absent with the DRP stimulus, which therefore offers an infinite range of central spectra each with its own pitch and lateralization [10]. In other words, for each value of the internal delay, a well-modulated cosinusoidal function of frequency is found waxing and waning between 0 and 1, thus with a peak-to-valley ratio equal to infinity. This implies that no pitch at all might be expected due to mutual competition of an infinite number of candidate spectra. This might explain that some authors do not find DRP. On the other hand, the historical reports of a single faint pitch in the center of the head are reconciled with the CS model only if strong prevalence for the central position would be assumed. However, other data on dichotic pitch but also data on lateralization with conventional stimuli plead against such an assumption.

Culling et al. [9] performed calculations with three versions of the CS model: 1) with infinitely sharp frequency analysis, 2) with time and frequency weighting, and 3) with ROEX filtering, thereby taking notice of central activity patterns generated within an internal delay of ±1.5 ms. They also included a search algorithm to select central spectra with a large modulation depth. It is worth recalling that they found that the assumption of infinitely sharp frequency resolution as used in the derivation of Eq. (1) seems to have only little influence on predictions of DRP and FP. However, to predict existence regions it will be necessary, of course, to include the properties of the peripheral auditory system as good as possible.

In the following sections, it is examined to what extent also other current theories comply with these data. A summary is given in columns 6 to 8 of Fig. 1. Correct prediction is indicated with + and incorrect or non-prediction with – for pitch value and lateralization respectively (+,–).

CROSS CORRELATION

Adopting the spirit, if not the letter, of Licklider's Triplex Theory, Fourcin [2] tried to explain his original findings on aFP with the wide-band interaural cross-correlation function. Especially the need for two cross-correlation peaks (i.e. a peak pair) to obtain a strong dichotic pitch prompted him to stress the importance of neural delay (cross correlation) and "comparison patterning".

With our recent knowledge of the novel pitch called sFP, the inadequacy of the concept of cross correlation (CC) is manifest already from the simple fact that identical pitch values are predicted for the aFP and sFP cases, which is in conflict with the data [3]. Further, it is unclear how ambiguity of pitch should be predicted from one cross-correlation peak pair, the more so as the peaks have equal polarity as in the case of aFP_+^+. Also, how should one explain the experimental fact that a negative peak at T_2 predicts a pitch image position corresponding to an internal delay T_2, while the image of a single interaurally-delayed noise is predicted by a positive instead of a negative peak at the same external delay? Finally, internal delays as long as 10 ms and longer would be needed to predict low-valued FPs, which is unrealistic from an ecological or physiological point of view.

Alternatively, one might consider the possible virtues of a "Summary Cross Correlogram (SCCG)", to be defined as the result of the "addition" of peripherally-filtered cross correlation functions, very much in analogy with the Summary Auto Correlogram (SACG) as promoted in

recent studies on monaural periodicity pitch [14]. It has been shown that the SACG resembles the wide-band auto correlation function in its main features (e.g. position of first peak). Likewise, the SCCG is expected to resemble the wide-band cross correlation function This, however, is unable to explain dichotic pitch behaviour for reasons similar to those mentioned above [3].

EQUALIZATION-CANCELLATION

Durlach's EC model

Durlach's original Equalization-Cancellation (EC) model [12] is basically able to predict HP, MPSP, BEP and BICEP values [7, 9]. Also aFP^{+}_{-} is correctly predicted in addition mode. However, the model has to switch to subtraction mode for aFP^{-}_{+} [3, 9]. Further, sFP data are not predicted by the EC model, simply because equalization by interaural delay always recovers the difference between the two delays, not the averaged value.

The interaural delay needed in the cancellation process could possibly be extracted as an indicator for pitch-image position. But given this possibility, we still are faced with the problem that multiple images are not predicted. Moreover, the correct prediction of both pitch value and lateralization always calls for addition instead of subtraction in the cancellation process. Therefore, in column 7 of Fig.1, we choose to consider the EC model in its *addition mode only* (Note that this implies a deviation from the general preference for subtraction in the modelling of BMLDs). Further, it is assumed that the EC mechanism (in the absence of a signal) strives for maximum reduction of the noise.

Culling et al.'s mEC model

Culling and colleagues [9] proposed a modified Equalization-Cancellation (mEC) model performing an equalization by adjustment of internal delay (and/or level) in each frequency channel (auditory filter) *independently*. An obvious reason for its failure to predict sFP along with aFP is its unique way of operation, i.e. to generate only one optimal "*recovered spectrum*". Further, lateralization is not dealt with by the mEC model, because the possibility to extract a single equalization delay as an indicator for laterality, is essentially absent.

For the DRP stimulus, the mE-C model does not produce any recovered spectrum at all. This non-prediction might be considered at odds with the existence [10] of a continuum of pitches as predicted by the application of Eq. (1). On the other hand, it might as well be considered indicative for the extremely low salience of DRP, or the difficulty that some listeners have to perceive or match DRP.

ALTERNATIVE (COMBINATIONS OF) MODELS

Culling - Akeroyd

As Culling et al.'s mEC model does not intend to explain the lateralization of dichotic pitches, a new model, the "reconstruction-comparison" model, was designed by Akeroyd et al. [11] specifically to predict the lateralization, not the value of the pitch, of simple dichotic pitches comparable to HP. In combination with the mEC model it is expected also to predict pitch values. It is based on the idea that the binaural auditory scene is partioned into two separate objects, one for the dichotic pitch and one for the background noise, before the lateralization of each object is computed. Essentially, it determines the lateralization of the dichotic pitch from the across-frequency average of the remainder after subtraction of a reconstructed noise cross-correlogram from the dichotic-pitch cross-correlogram.

Breebaart - de Cheveigné

Recently, Breebaart et al. [15] showed that a binaural model based on contralateral inhibition optimally accounts for signal detection data. Thereby, the binaural representation of stimuli is based on the Jeffress model supplied with EI-cells instead of EE-cells. For dichotic pitch stimuli, this results in an alternative CAP where maxima are replaced by minima and vice versa. In such a model, the minima also determine lateralization. Of course, when claiming parsimony, dichotic-pitch extraction based on spectral minima cannot be reconciled with monaural-pitch extraction (conventionally) based on maxima. A way out is to combine Breebaart's contralateral inhibition model with the de Cheveigné's [16] EI-cell-based cancellation theory of pitch. As the

latter model is comparable to the calculation of –SACG, it has the disadvantage of the need for long correlation delays.

CONCLUSIONS

Psychophysical facts

Experiments in the past have shown that:

- monaural and binaural equivalence of pitch and timbre (parsimony) exists,
- dichotic pitch value is coupled with pitch image position,
- dichotic pitch image position depends on ITD, not on IID (compare "time image").
- pitch value extraction seems to precede pitch image lateralization

Inferences for binaural modeling

- Facts 1 and 2 are *not consistently* predicted by Cross Correlation (CC) solely or a Summary Cross Correlogram (SCCG), neither by Equalization and Cancellation (EC) or Modified Equalization and Cancellation (mEC).
- Facts 1 and 2 are *consistently* predicted by the Central Spectrum (CAP-CS) theory implying spectral pattern matching (or mathematically equivalent: SACG for pitch value, but not SCCG for lateralization as in Licklider's triplex theory). However, the strong prevalence for a single centralized DRP percept is not yet understood.
- It is tempting to conclude from fact 3 that, for binaural signals in general, the integrated processing of ITDs and IIDs takes place beyond the level at which dichotic pitches are generated.
- Fact 4 seems in agreement with the independent finding that, in general, perceptual grouping precedes the localisation of complex sounds.

BIBLIOGRAPHICAL REFERENCES

[1] E. M. Cramer and W. H. Huggins, J. Acoust. Soc. Am. **30** (1958), 413-417.
[2] A. J. Fourcin, in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg, A. W. Sijthoff, Leiden, 1970, pp. 319-328.
[3] F. A. Bilsen and J. Raatgever, J. Acoust. Soc. Am. **108** (2000), 272-284.
[4] F. A. Bilsen, and J. L. Goldstein, J. Acoust. Soc. Am. **55** (1974), 292-296.
[5] F. A. Bilsen, J. Acoust. Soc. Am. **59** (1976), 467-468.
[6] M. A. Klein and W. M. Hartmann, J. Acoust. Soc. Am. **70** (1981), 51-61.
[7] W. M. Hartmann and C. D. McMillon, J. Acoust. Soc. Am. **109** (2001), 294-305.
[8] J. Raatgever and F. A. Bilsen, J. Acoust. Soc. Am. **80** (1986), 429-441.
[9] J. F. Culling, A. Q. Summerfield and D. H. Marshall, J. Acoust. Soc. Am. **103** (1998), 3509-3539.
[10] F. A. Bilsen, in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart et al., Shaker Publishing, Maastricht, 2001, pp. 145-152.
[11] M. A. Akeroyd and A. Q. Summerfield. J. Acoust. Soc. Am. **108** (2000), 316-334.
[12] N. I. Durlach, in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias, Academic, New York, 1972, pp. 369-462.
[13] A. N. Grange and C. Trahiotis, J. Acoust. Soc. Am. **100** (1996), 1901-1904.
[14] W. A. Yost, R. Patterson and S. Sheft, J. Acoust. Soc. Am. **99** (1996), 1066-1078.
[15] J. Breebaart, S. van de Par and A. Kohlrausch, J. Acoust. Soc. Am. **110** (2001), 1074-1117.
[16] A. de Cheveigné. J. Acoust. Soc. Am. **103** (1998), 1261-1271.