On the variations of inter-aural time differences (ITDs) with frequency

V. Benichoux, M. Rébillat and R. Brette

Dpt. d'études cognitives, ENS Paris, Ecole Normale Supérieure, 29 Rue d'Ulm, 75230 Paris, France
benichoux@crans.org

Inter-aural time differences (ITDs) constitute an important localization cue for azimuth estimation, particularly below 1.5 kHz. As a first approximation, it is commonly assumed that ITDs do not depend on frequency. Nevertheless, Kuhn (JASA, 1977) shows theoretically and experimentally that due to diffraction effects around the head, ITDs depend on frequency. Low frequency ITDs should thus theoretically be 1.5 times greater than high frequency ones. To study this point, different classical tools are adapted to compute the ITD variations with frequency: onset time differences, maximum of the cross correlation, and phase differences. The reliability of each tool regarding ITD computation is assessed on the basis of head-related transfer functions (HRTFs) coming from a spherical head model. The effective frequency dependence of ITDs is finally shown by analyzing real animal HRTFs.

1 Introduction

Mammals and birds use mainly Interaural Time differences (ITDs) and Interaural Level Differences (ILDs) to localize sound sources in their environment. Even though it is commonly accepted that the ILD conveys information through location-dependent variations across the audible spectrum, the ITD is usually thought of as a *broadband* quantity, that is, it does not depend on frequency, and is often reported as a single quantity.

The view of the ITD as a single broadband quantity is a good first approximation of the high frequency limit of the ITDs, as supported by theoretical studies [3]. However, a more detailed analysis of the acoustics of the head indicates that ITDs vary significantly with frequency. As an example, [3] considered a spherical head model with rigid boundaries, and showed that in this case the ratio of the high to the low frequency ITDs is equal to 2/3 near the horizontal plane. Accordingly, given that a typical cat hears a maximum high-frequency ITD of about 350 μs, then the maximum low-frequency ITD is of about 450 μs, which is a sufficiently high difference for the cat to distinguish.

Additionaly, neurophysiological insight shows that the ITD is extracted in a frequency-dependent way. Indeed one of the first stages of the auditory processing is a form of spectral decomposition induced by the cochlea, and the ITDs are extracted downstream, by neurons that display sensitivity only in a restricted frequency band. This raises the question of the relevance of the variations of ITDs across the frequency spectrum to animal behavior.

To be able to assess the functional advantage (if any) of the frequency dependence of the ITDs, new *frequency-dependent* methods of ITD estimation need to be devised. Then one must make sure that the observed variations are not due to estimation error, which is done in the present study by quantifying the robustness of the different methods to measurement noise. Finally, using synthesized cat HRTFs, it is shown that the variations of ITDs can help disambiguate between sources originating from the front or the back of the interaural axis, and also convey proprioceptive information, i.e. information about the animal's body position.

2 Frequency-dependent ITDs

Usual definitions of the ITDs distinguish between the *interaural phase delay* and the *interaural group delay*, in the present study, ITDs are defined as the phase delays. Those have already been shown to depend on frequency, both in theoretical and experimental studies [3, 1, 5]. In his classical textbook, Blauert [5] reports frequency-dependent interaural phase delays, but the apparent noise makes it hard to conclude on a potential systematic variation of the ITD with frequency. For the purpose of the present study, a review of existing frequency-dependent methods to compute ITDs is presented, and all methods are evaluated in terms of their robustness to measurement noise.

A complete linear representation of the acoustical effects of the head, body, etc. on the incoming wavefield is given in the frequency domain by Head Related Transfer Functions (*HRTF*s, or alternatively impulse responses, *HRIR*, in the time domain), a pair of filters for every position usually lying on a sphere around the subject's head. Those filters can be either experimentally measured, or computed theoretically for simple geometrical shapes (e.g. for a sphere [1]) or more complex ones [2]. The methods described here explain how to obtain frequency-dependent estimation of ITDs offline from those filter representations, i.e. not on ongoing signals.

Three methods were considered, two based on the temporal representation (HRIRs), and one on the frequency representation (HRTFs). They were adapted from classical estimators to yield frequency-dependent results. For the time-based methods, the HRIRs are first passed through a bank of bandpass filters with variable center frequency (CF), and then the classical (broadband) method is applied to the result. The filterbank used here is a Gammatone filterbank because it is known to be a simple yet good representation of the cochlear spectral decomposition. Obviously, any other type of bandpass filter would yield similar results, provided is has approximately the same bandwidth.

2.1 Onset time differences

A natural way to estimate broadband ITDs is to compare the times of arrival of the waves at the eardrum. This can be done on HRIRs by computing the onset times of the two impulse responses. Typically, a threshold is arbitrarily picked $\alpha \in [0, 1]$, and the onset time of each impulse response is computed as the time when a fraction α of the maximum of the impulse response is reached by the sound pressure:

$$T^{Onset}(h_r) = \min_t \{h_r(t) \leq \alpha \max_s (h_r(s))\} \quad (1)$$

An ITD then follows by $ITD^{Onset} = T^{Onset}(h_r) - T^{Onset}(h_l)$. As mentioned ealier, a frequency-dependent equivalent of this estimator is devised by filtering the HRIR prior to computation with a Gammatone filterbank. Noting $h_{l,r}^{CF}(t)$ the HRIR filtered through a Gammatone filter with center frequency CF, one can then define:

$$ITD^{Onset}(CF) = T^{Onset}(h_r^{CF}) - T^{Onset}(h_l^{CF}) \quad (2)$$

2.2 Cross-correlation

Another popular estimator of ITDs is the peak of the cross-correlation function of the two HRIRs, which is known to

represent the difference in phase delays of the two filters [6]. It can be similarly adapted to compute frequency-dependent ITDs by first passing the impulse responses through a bank of Gammatone filters:

$$ITD^{Xcorr}(CF) = argmax_\tau \int_{-\infty}^{+\infty} h_L^{CF}(t) \times h_R^{CF}(\tau + t)dt \quad (3)$$

2.3 Phase differences

Finally the definition of ITD as the interaural phase delay difference suggests an immediate estimator of the frequency-dependent ITDs, that is the difference in phases between the two HRTFs converted into time delays. Those can be extracted by computing the unwrapped phase (the $\angle(.)$ operator) of the ratio of the two transfer functions:

$$ITD^{Phase}(CF) = \frac{1}{2\pi CF} \left\langle \angle \left(\frac{HRTF^r(f)}{HRTF^l(f)} \right) \right\rangle_\Gamma \quad (4)$$

In order to be able to compare this estimation to the two previously described ones, phase delays are smoothed around the same center frequencies, with a weighing equal to the frequency response of a Gammatone filter (the $< . >_\Gamma$ operator). This ensures that the same frequency components are pooled when computing the ITDs in this method, as compared to the previous ones.

3 Methods: Assessing the robustness of the estimators

To compare the relative performance of the estimators derived above when facing different levels of measurement noise, a completely noise-free HRTF dataset was needed. Fortunately, it has been shown that HRTFs were well approximated by a spherical model with rigid boundaries, which has the advantage of having an analytical solution [1]. This allowed us to simulate *surrogate* experiments where measurement error was modeled by an additive gaussian white noise ξ. All impulse responses were normalized so that the front position (0°azimuth) has an RMS value of 1, and then the signal-to-noise ratio is defined as follows:

$$NSR = \frac{RMS(\xi(t))}{RMS(h(\theta, t))} \quad (5)$$

Where θ is the azimuth of the considered HRIR, and the RMS is defined as usual as:

$$RMS(h) = \sqrt{\frac{1}{T} \int_0^T h(s)^2 ds} \quad (6)$$

ITDs were then computed using the methods described above, and compared back to the original noise-free solution. All the HRTFs were generated for 1024 frequency points, at a samplerate of 44.1kHz. This simulation was done 25 times for each of 36 evenly distributed positions on the horizontal plane to the left of the sphere, at a distance of 2 meters.

Comparing the result of those experiments to the *reference* ITD (for which $NSR = -\infty\ dB$), biases, standard deviations and confidence intervals could be derived at different frequency points, in a manner independent of the azimuth (and of the absolute mean value of the ITD). Formally, statistics reported here were computed on the signed error term E defined for a given azimuth θ, center frequency and NSR:

$$E(\theta, CF, NSR) = ITD(\theta, CF, NSR) - ITD(\theta, CF, -\infty) \quad (7)$$

4 Results: Estimation performance

4.1 Non-biasedness

A first check of the validity of our approach is to test that our estimators indeed are non biased, this means that the expectation of the estimator is equal to the theoretical value. In our framework, this means that the error term E has an expected value of zero. Reported in Figure 1 are the histograms of error expectations for the three methods, pooled over different NSR ranges. As can be observed the mean is almost always zero, and in only a few cases does the mean diverge significantly from zero.

Additionally, only at very low NSRs, well below the usual NSRs encountered with modern digital recording hardware, does one find biases that are more than a few microseconds. Most of those biases are negative, indicating that all the methods are biased towards a *smaller* absolute value for the ITD. In many of those cases, this is due to an artifact that is termed "DC failure" in this study, and will be discussed in more details in the following section.

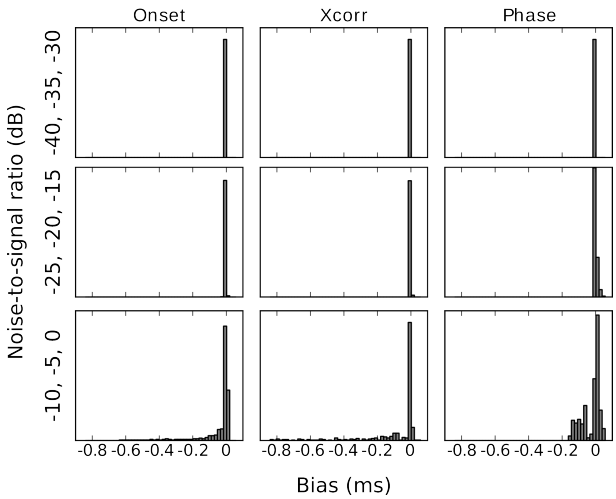

Figure 1: Histograms of biases. Biases are here defined as the mean over frequencies and positions of the error term devised in Eq. 7. Each row pools three different NSRs' biases on a single histogram. Columns are the different methods. For low NSRs, the biases have a mean of zero, but as the NSR goes up, the different methods show a bias towards smaller absolute ITDs.

4.2 Estimator dispersion

Considering the standard deviation (STD) of the error term over all positions and frequencies emphasizes some qualitative differences between the different methods, as reported in Figure 2. Yet, the first conclusion one can draw out of these simulations is that the STD is quite insensitive to NSR, indeed only at extreme noise levels (NSR bigger than -20 dB) do the STDs get bigger than 10 μs (in our case 2.5% of the maximum ITD).

The 1/f behavior of the STD of the Phase method estimator, and to a lesser extent of the Xcorr method can be easily explained. Indeed, this ITD value is obtained by dividing the average IPD over a certain window by the frequency. If the interaural phase value is non-zero for a few close-to-DC ($f = 0$) components of the spectrum, then the resulting estimation will diverge as CFgets close to zero. This effect is all

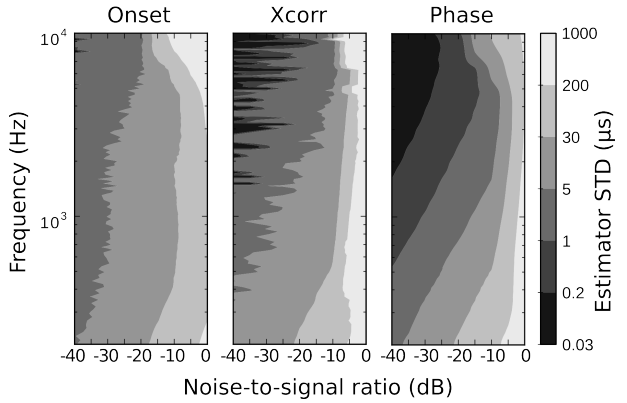

Figure 2: Estimation performances: The different panels display the standard deviation of the different estimators as a function of frequency and NSR. The first panels displays the standard deviation for the Onset method, it shows a more-or-less constant variation with NSR, even though at high or low frequencies the STD tends to be bigger. For the other two methods, the STD seems to follow a 1/f behavior.

the more problematic as signal processing hardware usually is unreliable in the very low frequency range.

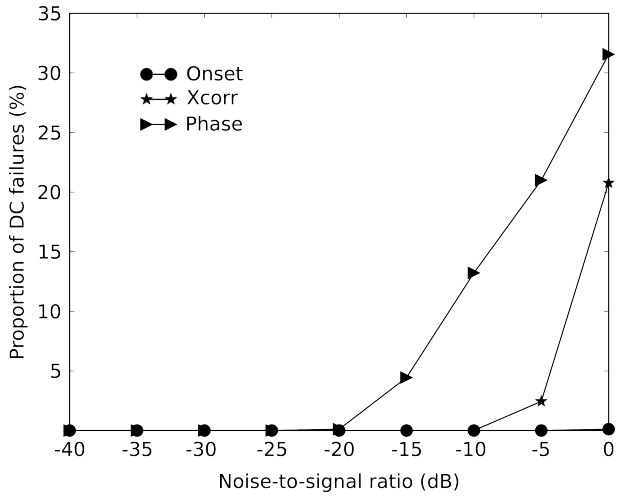

Figure 3: ITD computation failure: Indicates the proportion of abnormally high values for low frequency ITDs, termed "DC failure" (see text). These failures happend more often in phase-based methods than in the onset time differences method.

This particular sensitivity on the few first values of the complex spectrum can also lead to bigger artifacts, termed "DC failures" in the rest of the study. Those "DC failures" typically occur when the ITD magnitude is quite low, e.g. for low absolute azimuths, for which the interaural phase spectrum is dominated by that of the noise. To assess whether a computed ITD function was indeed a "DC failure", a threshold on the error term E evaluated at the lowest CF was set to 1ms. This allowed us to report failure rates as a function of NSR in the Figure 3. Indeed, for the two phase-based methods (Xcorr and Phase), this problem arises at moderate NSRs, and these methods yield abnormally high estimations for a significant proportion of the positions. This constitues a potential problem when directly using phase-derived methods to compute frequency-dependent ITDs, especially in the lowest frequencies and for low ITDs and low NSRs.

4.3 Confidence intervals and broadband variations

An additional statistic that was derived from the simulations is the 95% confidence intervals across the spectrum for all the positions. Those are a good representation of the variability of the measures, and their trustworthiness. Reported in Figure 4 are the ITD functions for four positions on the horizontal plane, for a spherical head model and alongside are plotted the confidence intervals for every measure. The point here is to show visually that the ITD variations observed across the audible spectrum are indeed bigger than the confidence intervals themselves. This constitutes a concrete

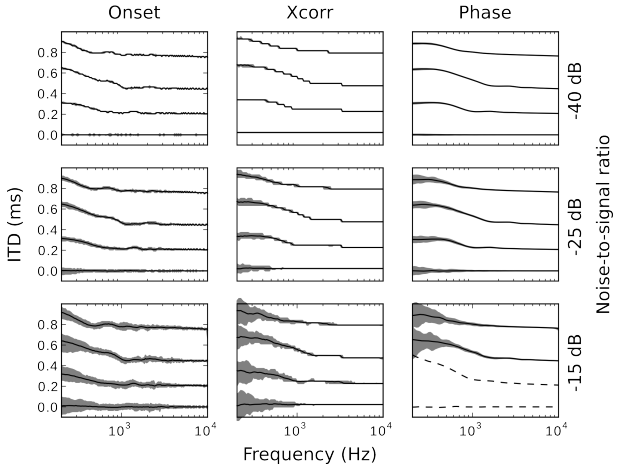

Figure 4: ITD curves for a spherical head model with a 20cm diameter. Each row is a different NSR, and each column a different estimation method. Positions shown are 0 °, 30 °, 60 °and 90 °. Gray areas indicate numerically computed 95 % confidence intervals

argument that the ITD are indeed dependent on frequency, because the variations observed are systematic, and bigger than the expected noise. Moreover, in modern experimental setups, the NSR can be as low as - 60 dB, range in which the estimator STD is expected to be too small to be noticed. Hence HRTF-derived ITD curves can indeed be trusted, provided that the signal-to-noise ratio is high enough.

5 Discussion: Investigating cat HRTFs

The cat is a widely used biological model when studying the neurophysiological basis of sound source localization. As was pointed out earlier, in such studies the ITD is often implicitly assumed to be a fixed quantity with respect to frequency, even though it has been shown that ITD-sensitive neurons' responses are frequency-dependent [4]. Hence it is of special importance to work on a more precise characterization of the ITDs for this species and moreover to try and uncover the *functional* advantage of such frequency variations, that is do those variations convey any more information than the pure, broadband quantity.

For this purpose, HRTFs were derived from 3D models (see Figure 5) of a stuffed cat using a previously published Boundary Element Method [2]. This study restricts itself to the analysis of positions on the horizontal plane, with a resolution of 5 °in azimuth. Since the filters were generated using numerical methods, they can be though of as completely free of noise and measured in absolutely anechoic conditions.

Moreover, since those HRTFs were based on 3D mesh models, it was possible to change the position of the cat head, as it was significantly slanted in the original stuffed animal.

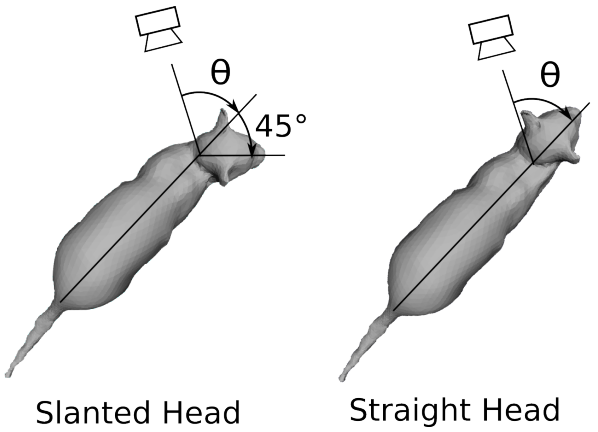

Figure 5: Cat 3D models top view: Angles are measured relative to the body axis, in the trigonometric sense (positive θ to the left of the animal). Given our conventions, the head points -45 °to the right of this axis. Keep in mind that positive ITDs come from sources to the left of the cat.

5.1 Front-back disambiguation

Reported in Figure 6 are the frequency-dependent ITDs as computed on our cat HRTFs. The gray dashed line corresponds to the upper limit of the phase locking in the cat auditory nerve, this gives an order of magnitude of the range where the cat actually processes time cues such as the ITD, i.e above $\approx 5kHz$ the cat cannot extract ITDs.

As could have been expected the ITDs for the front hemisphere are in qualitative agreement with the ones of the spherical model, displaying the same monotonously decreasing trend. Noticeably, though, the ITD curves seem to be equal up to a constant multiplicative scaling factor, i.e. the they never cross. This means that they do not convey more information than the pure broadband ITDs (in our framework, the high-frequency limit of the ITD curve).

Nonetheless, when considering ITDs on the whole horizontal plane (including the back hemidisc), one can draw different conclusions. A simple symmetry assumption implies that if the animal were a perfect sphere then the back ITDs would be exactly equal to the front ones. Multiple deviations to those assumptions hold for the cat, namely the presence of the body, and the fact that the ears do not lie on a diameter of the sphere. This implies that the ITDs for positions placed on the back of the animal should be different, and especially in their frequency variations, as shown on Figure 6. The back ITDs indeed display sharper transitions from the low-frequency to high-frequency behavior, especially for intermediate positions. Additionally, this variation occurs in a frequency range where the cat is known to process ITDs, implying that this cue could be taken advantage of to disambiguate front and back originating sound sources.

5.2 Proprioceptive information

Another striking effect on the ITD variations across the spectrum is the fact that it depends on the body posture of the animal. For sound sources originating from the front of the animal, the effect is reduced, as could be expected, because the acoustic wave does not encounter the body before reaching the ears of the animal. Notice that this would not necessarily hold if we were to consider HRTFs in a non-anechoic setup, as the body could get in the way of acoustic reflections. Hence the only effect seen here is a global shift of the ITD curves to more positive ITDs (because the head points to the right), with no significant deviation from the spherical model.

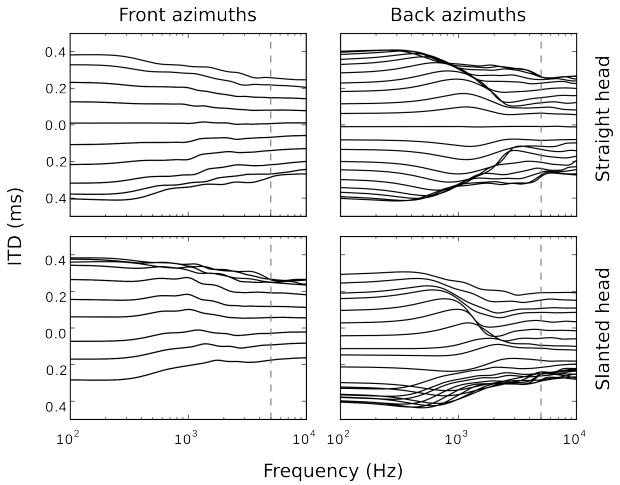

Figure 6: ITDs in the cat: effects of source direction and body posture: The vertical dashed line represents an indication of the upper limit of the ITD processing range of the cat. Bottom two panels represent the ITD values when the head of the cat is slanted, Upper panels when the head is aligned with the body. Left columns are the ITDs for positions in the front of the cat and right panels when the source is placed on the back of the animal. Notice how the body of the cat induces dramatic variations in the ITDs because of additional diffraction effects.

For sound sources originating from the back of the animal, ITD curves display a more complicated pattern. The most proeminent effect is that for sound sources that lie on the median plane (orthogonal to the interaural axis, here -45 °) the ITD is zero in the high and low frequencies, but it displays a significant variation in between, due to diffraction on the body. For left sound sources (positive ITDs), the effect is very reduced, and since the acoustic wave only *sees* the head, the ITD pattern is very much similar to that found for the spherical model, or the animal for frontal positions. As the source moves right though, the pattern seen is more complicated, and indeed significantly different from the expected one with a straight head.

6 Conclusion

This study has shown that there were multiple ways of estimating frequency-dependent ITDs. Amongst the methods presented here, some have a higher robustness to noise. The onset estimator has the undisputable advantage of showing only a moderate dependence of estimation performance with respect to frequency, i.e. it performs well in the whole audible spectrum. The natural phase estimator and the cross correlation estimator qualitatively show a 1/f behavior in the dependence of the estimation error. But the phase method has a hard time evaluating relatively small ITDs at low NSRs,

and thus might not be suited to computing the ITDs on, say, smaller mammals like the cat or the gerbil, especially at low frequencies.

Nevertheless, they all agree on the fact that the ITD is not a fixed quantity with respect to frequency. This argues for rethinking this binaural cue to take into account frequency variations.

Additionaly, those frequency variations might, much as the ILDs, convey some useful information in their frequency variations. Indeed in the example of cat HRTFs, the presence of the body for sources coming from the back impose dramatic changes in the ITD vs. freq patterns. This could enable the animal to use time cues to disambiguate sound sources coming from the back. Moreover, it seems that the body posture also has an effect on the ITDs. Whether this is an advantage (ITD variations encode an additional dimension of the stimulus) or a drawback (ITDs are not robust to animal position change) is up for discussion.

These results advocate for a reconsideration of the ITDs as a frequency-dependent quantity. Our results strongly suggest that these variations convey both proprioceptive information, and additional information about the source's localization (namely the front vs. back disambiguation). Altogether these effects should be taken into consideration when investigating the mamallian ability to localize sound sources based on binaural timing cues, design localization algorithms or rendering 3D sounds.

Acknowledgments

We would like to thank our collaboratrs, Makoto Otani for the BEM simulations, Renaud Keriven for the cat 3D models, and the Museum d'Histoire Naturelle de Paris for providing us with the stuffed cat.

This work was supported by an ERC Starting Grant (ERC StG 240132).

References

[1] R. O. Duda, W. L. Martens, *Range dependence of the response of a spherical head model*, Journal of the Acoustical society of America, **104**, 3048, (1998).

[2] M. Otani, S. Ise, *Fast calculation system specialized for head-related transfer function based on boundary element method*, Journal of the Acoustical society of America, **119**, 2589, (2006).

[3] G. F. Kuhn *Model for the interaural time differences in the azimuthal plane*, Journal of the Acoustical society of America, **62** (1), (1977).

[4] T. C. Yin, S. Kuwada, *Binaural interaction in low-frequency neurons in inferior colliculus of the cat. III. Effects of changing frequency*, Journal of neurophysiology, **50** (4), 1020–1042, (1983).

[5] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*, MIT press, (1997)

[6] S. L. Marple Jr, *Estimating group delay and phase delay via discrete-time "analytic" cross-correlation*,, IEEE Transactions on Signal Processing, **47** (9), 2604–2607, (1999).