PHANTOM SOURCE WIDENING WITH DETERMINISTIC FREQUENCY DEPENDENT TIME DELAYS

online revision: error in table 1 corrected, iem.at, iaem.at, Zotter, Oct. 2012.

Franz Zotter, Matthias Frank, Georgios Marentakis, Alois Sontacchi[*]

Institute of Electronic Music and Acoustics
University of Music and Performing Arts, Graz, Austria
{zotter,frank,marentakis,sontacchi}@iem.at

ABSTRACT

We present a novel method to adjust the perceived width of a phantom source by varying the deterministic inter channel time difference ($ICTD$) in a pair of signals over frequency. In contrast to given literature that focuses on random phase over frequency, our paper considers a deterministic approach that is open to a more systematic evaluation. Two allpass structures are described, finite impulse response (FIR) and infinite impulse response (IIR), for phase-based phantom source widening and evaluated in a formal listening test. Varying $ICTD$ over frequency essentially alters the inter-aural cross correlation coefficient at the ears of a listener and in this way provides a robust way to control the auditory source width. The subjective evaluation results fully support our observations for both noise and speech signals.

1. INTRODUCTION

Two loudspeakers emitting the same sound simultaneously will create the illusion of a phantom source, a sound localized in between. A long lasting problem in the field is the manipulation of the perceived phantom source width. Up until now, the main ways to manipulate it were the so-called pseudo stereo or decorrelation approaches.

Purely phase-based decorrelation techniques were comprehensively studied in the work of Kendall [1], which gives an overview of already convincing perceivable effects, using random phase values for different frequencies. Improvements and variations were described in [2, 3]. The control of a random phase is nevertheless a challenging task and evaluation results verify the difficulty inherent in reducing the variability in the perceptual outcome of such a process.

Pseudo stereo can be achieved by various kinds of implementations, which essentially introduce fluctuations in the frequency dependent level and phase differences of a pair of playback signals. Schröder [4] and Orban [5, 6] describe the Lauridsen network which produces a pair of spectrally complementing comb filters. Gerzon [7] discusses that previous work on Lauridsen networks suffers from phasing, a serious side-effect. He proposed various alternative strategies to reduce this effect: a) frequency varying amplitude panning, b) a modified Lauridsen network with a delayed channel, and c) unitary feedback delay networks.

Of the techniques proposed by Gerzon, frequency varying amplitude panning is very close to observations made independently by Blauert and Lindemann. In their work, they mention [8] that panning of individual frequency bands in specific directions, or the rapid interchange of the panning directions of the frequencies within a signal, reduces the inter-aural cross correlation coefficient (IACC) and increases the phantom source width, i.e., auditory source width (ASW). They also verify the inverse relation between IACC and ASW for headphone listening by combining uncorrelated noise sources. The IACC is the maximum of the normalized inter-aural cross correlation function within a maximum time shift of $\pm 1\mathrm{ms}$ [9]. Despite the evaluation using headphones, their work is fundamental when creating a pseudo-stereo or decorrelated signal pair out of a single monophonic sound.

The main focus of this work is to revise phase-based approaches that neither suffer from phasing nor require special skill in designing random variables by taking into account the observations by Blauert and Lindemann [8] and Gerzon [7]. To this end, we establish the relationship between the phase of an allpass filter and its frequency dependent group delay. Using the group delay differences in a pair of allpass filters enables the deterministic control of $ICTD$. Suitable implementations are provided for both FIR and IIR filter designs. Eventually, the target phase is determined by an accurately reproducible frequency dependent inter channel time difference ($ICTD$). This approach is equivalent to a frequency dependent time delay panning in stereophony. The perceived phantom source widening using the presented algorithms is evaluated and verified by a listening experiment and related to the IACC, using a setup illustrated in Fig. 1.

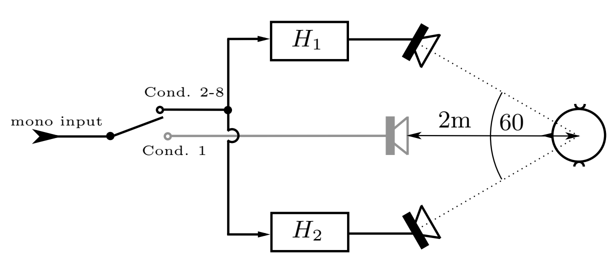

Figure 1: The setup used in the evaluation experiment.

[*] Contributions: Franz Zotter developed two allpass structures yielding frequency dependent $ICTD$ for decorrelation. Georgios Marentakis proposed to investigate ASW control through frequency dependent panning. Matthias Frank and Georgios Marentakis designed the evaluation study executed by Matthias Frank. Georgios Marentakis and Matthias Frank executed the statistical analysis. Alois Sontacchi contributed his expertise on the field of signal decorrelation, perceptual evaluation and in the presentation of research.

2. ALLPASS STRUCTURES FOR DETERMINISTIC ICTD

Decorrelation [1] designs are usually done by means of a random allpass phase $\phi[k]$ in the discrete Fourier transform (DFT) domain. Introducing random phase in a signal cannot be done without constraints imposed to maintain acceptable sound quality.

Assuming an arbitrarily designed phase pair ϕ_1, ϕ_2, the out of phase signals are easily avoided by forbidding these values in the design, i.e. $\phi_2 - \phi_1 \neq (2l+1)\pi, l \in \mathbb{Z}$.

Another issue is to ensure that the magnitude response stays unity between the frequency bins k, cf. [1], or similarly, that the impulse response does not suffer from cyclic time-domain aliasing. The difference of the random phase in successive bins is therefore subjected to a limitation $|\phi[k] - \phi[k-1]| < \Delta\phi_{\max}$. After this limitation, the achievable effect is mainly controlled by this limit and the DFT-length N. In [3] it is even proposed to make this limit depending on the Bark frequency to improve some perceptual qualities.

This phase limitation is quite interesting for the deterministic design as there is an underlying meaning that is useful. The (cyclic) group delay in one frequency bin can be estimated using the first backward difference of the phase $\tau[k] = \frac{N}{2\pi\, f_s}\Delta\phi[k]$, with the signal sampling frequency f_s. Hence, the limitation of the phase change over successive bins is observed to limit the group delay, implicitly. But it also means that the group delay $\tau[k]$ could be designed as a random variable to construct the phase

$$\phi[k] = \frac{2\pi\, f_s}{N} \sum_{k'=0}^{k} \tau[k']. \tag{1}$$

Nevertheless, the limitation of frequency dependent random group delays might not be enough as a design parameter for interesting effects influencing the phantom source image. For instance, the group delay should change often enough over frequency to avoid angular shifts of the phantom source. On the other hand, too frequent changes might cause impulse responses of inconvenient length. Therefore, a cosine function is proposed to design the group delay.

2.1. Deterministic FIR allpass design

The presented study favors a deterministic FIR allpass design that is reproducible and has a controllable variation of the $ICTD$ over frequency.

As the simplest choice of its deterministic behavior, a positive and negative cosine contour with adjustable frequency period Δf and peak value $\hat{\tau}$ can be chosen as the group delay of the transfer functions H_1, H_2

$$\tau_{1,2}(\omega) = \mp\hat{\tau}\cos(\omega/\Delta f), \tag{2}$$

with the angular frequency variable $\omega = 2\pi f$. This yields, with $\tau_2 - \tau_1$, an adjustable inter channel time delay

$$ICTD(\omega) = 2\,\hat{\tau}\,\cos(\omega/\Delta f). \tag{3}$$

By the negative integral of the group delay over ω, the sinusoidal phase $\phi(\omega) = \pm\hat{\tau}\Delta f\ \sin(\omega/\Delta f)$ of the allpass decorrelation filters are obtained, with $j = \sqrt{-1}$,

$$H_{1,2}(\omega) = e^{\pm\, j\hat{\tau}\Delta f\, \sin(\omega/\Delta f)}. \tag{4}$$

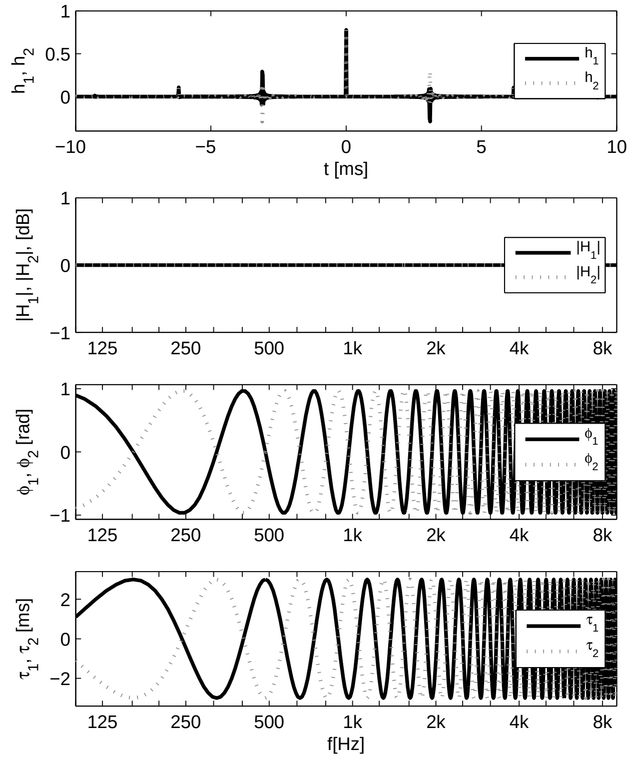

Figure 2: Impulse, magnitude, phase, and group delay responses of the deterministic FIR allpass pair.

It is not surprising that the structure of the corresponding impulse response is similar to an FM-spectrum with its Bessel functions $J_m(\mu)$ of various orders m and the modulation depth μ. For the two allpass functions $h_1(t)$ and $h_2(t)$ with opposing time delays, the following impulse responses are obtained

$$h_{1,2}(t) = \sum_{m=-\infty}^{\infty} J_{\pm m}(\hat{\tau}\Delta f)\;\delta\left(t - \frac{m}{\Delta f}\right). \tag{5}$$

The corresponding inter channel time delay is adjustable with regard to the frequency period $\Delta f/2$ in which the sign of the $ICTD$ alternates its magnitude between $-2\,\hat{\tau} \leq ICTD \leq 2\,\hat{\tau}$. The FIR implementation is made causal by introducing a suitable time delay. Its length is limited as Bessel functions vanish with large $|m|$ and small $\hat{\tau}\Delta f$. Fig. 2 shows the impulse, magnitude, phase, and group delay responses of the system. In order to avoid opposite inter channel phase, the phase argument of one channel ϕ needs to be restricted to $\pm\pi/2$, i.e.

$$\hat{\tau}\Delta f < \frac{\pi}{2}. \tag{6}$$

2.2. IIR allpass design

As an alternative IIR implementation of the phantom source widening effect, an implementation with third-octave structure can be defined. Let us assume two cascade chains $l = 1, 2$ containing

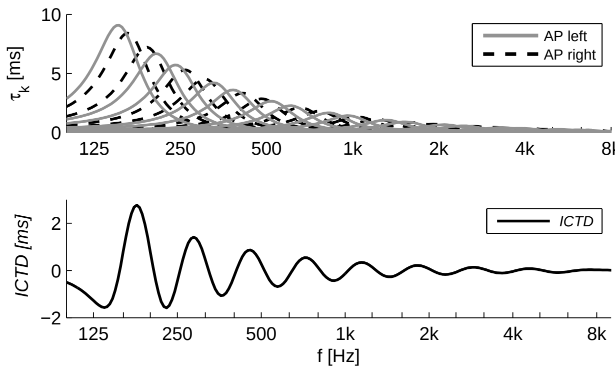

Figure 3: Group delay responses of the 2nd order IIR allpass chain elements for the channel 1 and 2 and the achieved $ICTD$.

third-octave 2nd order allpass filters

$$H_l(s) = \prod_{k=-8}^{8} H_{k,l}(s), \qquad l = 1, 2. \tag{7}$$

The allpass k in the cascade l is, for simplicity, described in the Laplace/Fourier domain as a complex continuous-time filter

$$H_{k,l}(s) = \left.\frac{s - j\omega_{k,l} - \sigma_{k,l}}{s - j\omega_{k,l} + \sigma_{k,l}}\right|_{s=j\omega} = e^{j2\arctan\frac{\omega_{k,l}-\omega}{\sigma_{k,l}}}. \tag{8}$$

We may use a reference angular frequency $\omega_0 = 2\pi$ 1kHz in order to define the center frequencies of the filters $\omega_k = 2^{k/3}\,\omega_0$ and their bandwidths $\sigma_k = \frac{\partial}{\partial k}\omega_k = \omega_k/Q$ with $Q = 1/\ln 2^{1/3}$. Two identical cascades do not create any $ICTD$ yet, but only require slight modifications to do so.

A frequency varying $ICTD$ is obtained when the center frequencies of both chains are shifted alternatingly towards lower and higher frequencies by some factors ϵ^{-1} or ϵ, respectively. Defined accordingly, the allpass parameters of the chain l are

$$\omega_{k,l} = \epsilon^{(-1)^{k+l}}\, 2^{k/3}\,\omega_0, \tag{9}$$
$$\sigma_{k,l} = \omega_{k,l}/Q. \tag{10}$$

For good results, ϵ should be bounded $1 \leq \epsilon < 2^{1/6}$ to avoid frequencies with opposite inter channel phase.

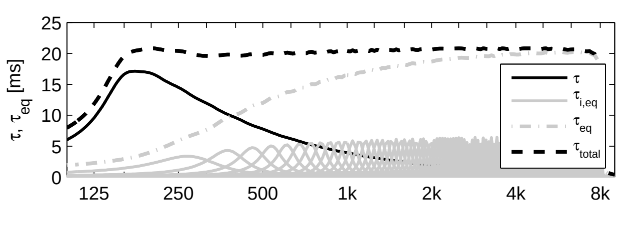

Figure 4: Group delay response of the allpass chain and its equalization with 150 2nd order IIR allpasses, and the compensated group delay curve.

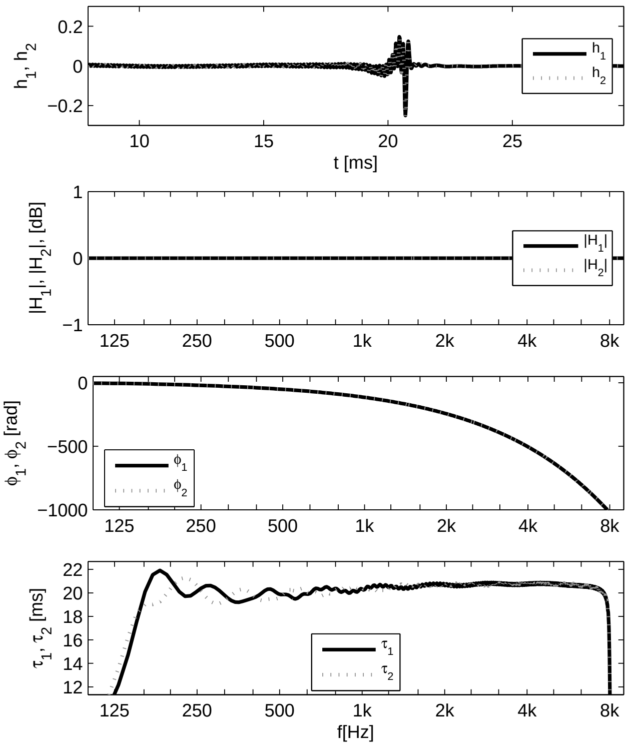

Figure 5: Impulse, magnitude, phase, and group delay responses of the deterministic IIR allpass chain pair out of third-octave 2nd order allpasses and their 150 element group delay compensation.

A closer insight into the proposed structure is obtained by regarding the group delay. The negative derivative of the phase is easy to calculate for one element k, and, in sum, yields the group delay introduced by the cascade l

$$\tau_l(\omega) = \sum_{k=-8}^{8} \frac{2\,\sigma_{k,l}^2}{\sigma_{k,l}^2 + (\omega_{k,l} - \omega)^2}. \tag{11}$$

The $ICTD$ is the difference between the group delays of both cascades

$$ICTD(\omega) = \tau_2(\omega) - \tau_1(\omega). \tag{12}$$

Essentially, this achieves the desired effect, see bottom diagram in Fig. 3, but the group delay distortion of a third-octave allpass cascade is clearly audible and slightly annoying for speech and transients. An equalizer for the group delay has been designed to avoid this in the proposed implementation, see Fig. 4. This is a 150 element allpass chain, designed to compensate for the group delay curve

$$\tau_{eq}(\omega) = \tau_c - \frac{\tau_1(\omega) + \tau_2(\omega)}{2} \tag{13}$$

up to a constant offset τ_c. For more details on the design of allpass chains of 2nd order elements with specific overall dispersion behavior see [10]. Fig. 3 shows the group delay responses of all the IIR cascade elements, its sum, the dispersion compensation allpasses and the overall obtained $ICTD$.

		Description	IACC	$\mathrm{IACC_{E3}}$
	C1	real source	0.886	0.875
	C2	phantom source	0.824	0.815
	C3	IIR allpass	0.741	0.725
1.5ms	C4	FIR allpass, $\hat{\tau}$ = 3ms, Δf = 200Hz	0.715	0.690
0.5ms	C5	FIR allpass, $\hat{\tau}$ = 1ms, Δf = 1200Hz	0.528	0.559
1.5ms	C6	FIR allpass, $\hat{\tau}$ = 3ms, Δf = 400Hz	0.598	0.554
3.0ms	C7	FIR allpass, $\hat{\tau}$ = 6ms, Δf = 200Hz	0.571	0.521
1.5ms	C8	FIR allpass, $\hat{\tau}$ = 3ms, Δf = 600Hz	0.394	0.468

Values in the table erroneously indicate the maximum ICTD observable the original stimuli and not the tau parameter, which is only half as big.
(Zotter, Oct. 2012)

Table 1: Description of the tested conditions.

	C1	C2	C3	C4	C5	C6	C7	C8
C1	-/-	91/84	91/100	95/84	100/100	100/100	95/100	100/95
C2	9/16	-/-	82/91	86/68	95/100	100/95	100/100	100/100
C3	9/0	18/9	-/-	75/48	91/77	91/93	100/100	100/100
C4	5/16	14/32	25/52	-/-	68/66	91/93	86/98	100/100
C5	0/0	5/0	9/23	32/34	-/-	77/86	86/89	100/95
C6	0/0	0/5	9/7	9/7	23/14	-/-	48/50	91/75
C7	5/0	0/0	0/0	14/2	14/11	52/50	-/-	82/70
C8	0/5	0/0	0/0	0/0	0/5	9/25	18/30	-/-

Table 2: Relative frequency dominance matrix containing the average percent of times participants responded that a column is wider than a row for both noise/speech stimuli; numbers are rounded integers of the relative votes in %.

In Fig. 5 the resulting impulse, magnitude, phase, and group delay responses are given. In this implementation, the parts of the impulse responses lying below 200Hz and above 7kHz were left untreated and come without delay for simplicity of illustration.

3. EVALUATION

We created eight conditions (described in Table 1) that correspond to the discussed techniques and evaluated them in a formal listening test. The first two conditions were control conditions, included to verify that the manipulations yield a larger source width relative to an untreated phantom source (C2) and a mono source (C1). The IIR condition C3 does not have many degrees of freedom thus it was evaluated in a single condition. C4-C8 are more versatile and were tested with various parameter settings for Δf and $\hat{\tau}$.
The IACC for our eight conditions was computed from dummy head recordings using a B&K 4128C in the real listening test setup. For completeness, a recent measure proposed in [11], the $\mathrm{IACC_{E3}}$ was also computed. $\mathrm{IACC_{E3}}$ is the mean of the IACC computed in 3 octave bands (500Hz, 1KHz and 2KHz) for the first 80ms.

3.1. Method

Participants were seated according to Figure 1 and were facing forward, i.e. the participants were facing the centre of the sound event. They were presented with all possible pairwise comparisons of the eight experimental conditions and indicated which sound in a pair (A, B, or none) is wider by pressing the corresponding buttons on a keyboard. They could listen and switch between the sounds in a pair at will.

The test was divided into two parts with different stimuli: 5s of pink noise in the first and speech in the second part. For the second part, a mono sample of 22s male English speech was used from the EBU SQAM CD [12]. Stimuli were presented at 65dB(A) for noise and 65 $\mathrm{L_{eq}(A)}$ for speech. Before each part, a short training phase was conducted to familiarize the participants with the comparison task and the stimuli. Each of the (8 choose 2 =) 28 paired comparisons was rated twice by one participant, yielding a number of 56 comparisons for each part; a total number of 112 comparisons per participant for both parts. The order of presentation and the location of each stimulus within each pair (A, B) was randomized.

3.2. Setup

Three Genelec 8020 loudspeakers were placed at -30°, 0°, and +30°, 2m from the participants' head, see Figure 1. The size of the playback room was approx. $11\mathrm{m} \times 11\mathrm{m} \times 5\mathrm{m}$. The average reverberation time $\mathrm{RT_{60}}$ was 470 ms. Although with reference to ITU-R BS.1116-1 [13] the room is large, it still is within the recommended reverberation time limits. In addition, participants were seated within the effective critical distance of the setup, which was calculated to be 2.76m. 11 participants (4 female, 7 male) participated in the listening test (age range: 23-32 years, median: 27 years). All participants were part of a trained listening panel [14, 15] and familiar with the evaluation of source width, as they had participated in another source width experiment [16].
We tested the following hypotheses: 1. The decreased IACC or $\mathrm{IACC_{E3}}$ created by the variation of the $ICTD$ over frequency affects the perceived ASW. 2. The effect of the different conditions is consistent for different input signal types as long as the $ICTD$ variability is distributed over the signal's bandwidth. 3. The perception of the ASW increases with an increase in the magnitude of the $ICTD$. 4. The perception of the ASW is affected by Δf.

3.3. Results

Participants responded consistently throughout the experimental trials. Only 11% of the repetitions for noise and 14% for speech contained a different response. Most of this variation was observed when they compared pairs C3/C4, and C6/C7. Table 2 shows the relative frequency dominance matrix averaged across all participants and repetitions. A first indication that the different conditions yielded different ASW impressions can be obtained by examining the percent of the responses in which participants judged condition C_{i+1} to be wider than C_i, $i = 1 \ldots 7$. For noise, all comparisons yield percent discrimination significantly larger than chance ($p < 0.05$) apart from comparison C6/C7 and C4/C5. Results for speech stimuli were the same with the additional exception of condition comparison C3/C4.

A more complete picture can be obtained by creating phychophysical scales that take into account the participants' responses in all possible pairwise comparisons between the eight conditions. ASW scales were constructed based on the dataset in Table 2 according to Thurstone Case V [17] and the BTL model [18] and are presented in Table 3. Conditions were ranked in the same way by both models however the differences between the scale values of each condition varied. We attribute this discrepancy to the fact that our experimental conditions were easily distinguishable yielding ceiling effects that are handled differently in the two models. Although we concentrate our analysis on the Thurstone scales, our conclusions are supported by both models.

	Thurstone		BTL	
	Noise	Speech	Noise	Speech
C1	0	0	0	0
C2	0.19	0.14	0.0018	0.0043
C3	0.33	0.38	0.0063	0.0256
C4	0.48	0.30	0.0191	0.0183
C5	0.61	0.50	0.0398	0.0692
C6	0.78	0.87	0.2648	0.3972
C7	0.82	0.89	0.1473	0.8785
C8	1	1	1	1

Table 3: Scale values normalized within [0,1] based on Table 2, using the Thurstone and BTL model.

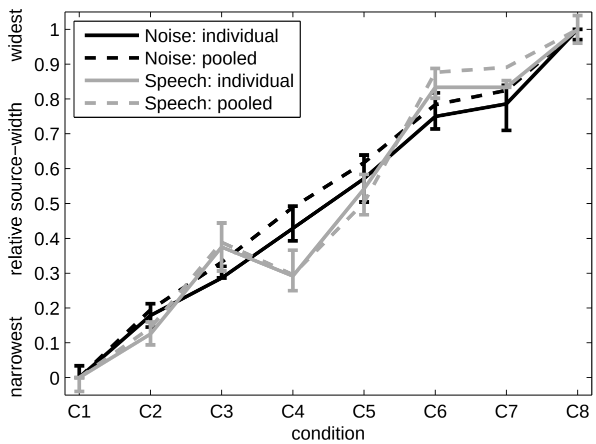

Figure 6: Thurstone scales for noise/speech: median and 95% CI using participants' individual scales compared to scales using answers pooled from all participants, Tab. 2. Scales were normalized within [0,1].

Thurstone scales for each subject were calculated based on each individual's frequency dominance matrix. It is worth noting that the median of the scales from the individuals' responses and the scales of the pooled responses constructed in the previous paragraph were almost identical (see Figure 6). A 2-way (Stimulus x Condition) ANOVA was then performed on the scale values of each individual as a function of the condition and stimulus type used in the experiment. Analysis of variance yielded no effect of stimulus type and only a rather weak interaction between stimulus and condition, F(7,70) = 2.08, $p = 0.057$. There was a significant main effect of condition, F(7,70) = 263.616, $p < 0.001$. Post-hoc comparisons using Bonferroni confidence interval adjustment showed that all conditions yielded significantly different source width perceptions ($p < 0.001$) with the exception of C3/C4 and C6/C7. In the latter case no difference was observed for both noise and speech. In the former, a significant difference was observed for noise but none for speech, attributing partially to the marginally significant interaction. The significant difference between conditions C4 and C7 indicates that ASW perception depends on the magnitude of $ICTDs$. One could try to predict the effect of Δf from the fact that C4 yielded significantly lower ASW impressions than both C6 and C8, implying an inverse relationship of ASW to Δf. However, this is not confirmed when comparing C6 and C7.

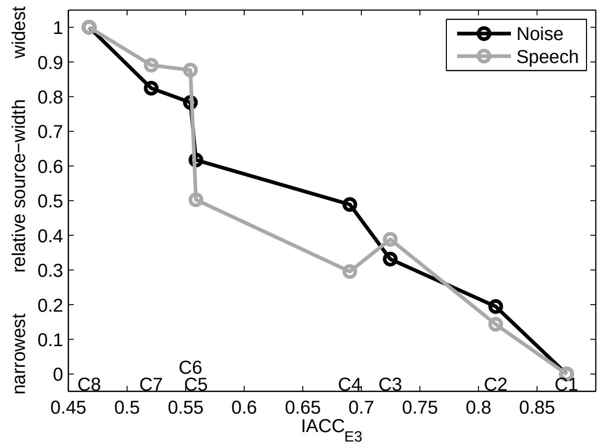

Figure 7: Thurstone scales for answers pooled from all participants as a function of the $\mathrm{IACC_{E3}}$. Scales were normalized within [0,1], $\mathrm{IACC_{E3}}$ of the conditions are indicated.

3.4. Discussion

Our hypotheses were in general confirmed by the experimental results. Hypothesis 1 stating that the decreased IACC or $\mathrm{IACC_{E3}}$ as a consequence of adjusting $ICTDs$ would yield different ASW impressions was confirmed by the significant main effect of condition observed in our experiment. The Thurstone scales from the pooled participants' responses correlated with IACC at $r = 0.95$ for noise and $r = 0.92$ for speech. Similarly, $\mathrm{IACC_{E3}}$ correlated at $r = 0.98$ and $r = 0.97$ with the noise and speech stimuli respectively, yielding a nearly linear relation as is clearly shown in Figure 7.

Hypothesis 2 stating that the manipulation of the $ICTD$ would apply to diverse stimuli types was confirmed for the noise and speech used in our experiment. The ASW scales already presented in Table 3 for noise and speech were highly correlated to each other ($r = 0.96$). However, further studies are required to fully confirm this hypothesis for other signal types.

Hypothesis 3 stating that the ASW impression would depend on the magnitude of the $ICTD$ was confirmed by the significant difference between conditions C4 and C7.

Hypothesis 4 was also confirmed, however, a proper identification of the effect of Δf will have to be discussed in future studies.

4. CONCLUSION

We have presented methods for phantom source widening using deterministic frequency dependent time delays. The methods have been implemented as deterministic FIR and IIR allpass structures. The presented algorithms are useful new audio effects for spatial sound imaging with well controllable behavior. However, clearly, the comprehensive evidence of all perceived aspects and comparison to earlier ideas about pseudo-stereo are outside the scope of this paper and remain subject to future studies. It is worth noting that in confirmation of the authors' subjective impression from listening to the presented algorithms, participants did not report annoyance or timbral deficiencies in the listening experiments. This however remains to be established in a more formal way in future studies.

Importantly, we show that frequency dependent $ICTD$ panning can be successfully used to widen a phantom source. Although our evaluation results are on $ICTD$ panning, informal experimentation shows that the same technique can also be used with frequency dependent $ICLD$ panning. Frequency dependent panning works essentially because it adjusts the inter-aural cross correlation coefficient of the received signal and therefore provides a robust way to control apparent sound source width. Several issues remain unresolved however. In particular, it would be interesting to identify when exactly image splitting starts to appear and the role of different panning envelopes and individual frequencies. Such undertakings will likely increase our knowledge about the mechanisms that leads to robust ASW representation in the brain and assist us in the creation of better ways to control ASW.

5. ACKNOWLEDGMENTS

This project is partly funded by the Austrian Research Promotion Agency (FFG), the Styrian Government and the Styrian Business Promotion Agency (SFG) under the COMET programme. Thanks to all participants for their participation and to the project partners for assistance.

6. REFERENCES

[1] G. S. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music J.*, vol. 19, no. 4, 1995.

[2] G. Potard and I. Burnett, "Decorrelation techniques for the rendering of apparent sound source width in 3d audio displays," in *Proc. DAFx-07*, Napoli, 2007.

[3] M. Bouéri and C. Kyriakakis, "Audio signal decorrelation based on a critical band approach," in *Preprint 117th Conv. Audio Eng. Soc.*, San Francisco, 2004.

[4] M. R. Schröder, "An artificial stereophonic effect obtained from a single audio signal," *J. Audio Eng. Soc.*, vol. 6, no. 2, 1958.

[5] R. Orban, "A rational technique for synthesizing pseudo-stereo from monophonic sources," *J. Audio Eng. Soc.*, vol. 18, no. 2, 1970.

[6] R. Orban, "Letters to the editor: Further thoughts on 'a rational technique for synthesizing pseudo-stereo from monophonic sources'," *J. Audio Eng. Soc.*, vol. 18, no. 4, 1970.

[7] M. A. Gerzon, "Signal processing for simulating realistic stereo images," in *Preprint 3423, 93rd Conv. Audio Eng. Soc.*, San Francisco, 1992.

[8] J. Blauert and W. Lindemann, "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," *The Journal of the Acoustical Society of America*, vol. 79, pp. 806–813, 1986.

[9] ITU, "ISO 3382:1997 acoustics - measurement of the reverberation time of rooms with reference to other acoustical parameters," Tech. Rep., 1997.

[10] J. S. Abel and J. O. Smith, "Robust design of very high-order allpass dispersion filters," in *Proc. DAFx-06*, Montreal, 2006.

[11] T. Hidaka, L. L. Beranek, and T. Okano, "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls," *J. Acoustical Soc. Am.*, vol. 98, 2, pp. 988–1007, 1995.

[12] EBU, "Tech 3253 - sound quality assessment material (sqam)," http://tech.ebu.ch/publications/sqamcd.

[13] ITU, "ITU-R BS.1116-1 methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," Tech. Rep., 1994-1997.

[14] A. Sontacchi, H. Pomberger, and R. Höldrich, "Recruiting and evaluation process of an expert listening panel," in *Fortschritte der Akustik, NAG/DAGA*, Rotterdam, 2009.

[15] M. Frank, A. Sontacchi, and R. Höldrich, "Training and guidance tool for listening panels," in *Fortschritte der Akustik, DAGA*, Berlin, 2010.

[16] M. Frank, G. Marentakis, and A. Sontacchi, "A simple technical measure for the perceived source width," in *Fortschritte der Akustik, DAGA*, Düsseldorf, 2011.

[17] L. L. Thurstone, "A law of comparative judgment," *Psychological Review*, vol. 101, no. 2, pp. 266–270, 1994.

[18] F. Wickelmaier and C. Schmid, "A matlab function to estimate choice model data from paired-comparison data," *Behavior Research Methods, Instruments and Computers*, vol. 36, no. 1, pp. 29–40, 2004.

[19] C. Faller, "Parametric coding of spatial audio," in *Proc. DAFx-04*, Napoli, 2004.