Audio Engineering Society

Convention Paper 8278

Presented at the 129th Convention
2010 November 4–7 San Francisco, CA, USA

Inherent Doppler Properties of Spatial Audio

Martin J. Morrell[1], and Joshua D. Reiss[1]

[1] *Queen Mary Universty of London, London, Mile End Road, E1 4PL, United Kingdom*

Correspondence should be addressed to Martin Morrell (martin.morrell@elec.qmul.ac.uk)

ABSTRACT

The Doppler shift is a naturally occurring phenomenon that shifts the pitch of sound if the emitting objects distance to the listener is not a constant. These pitch deviations, alongside amplitude change help humans to localise a sources position, velocity and movement direction. In this paper we investigate spatial audio reproduction methods to determine if Doppler shift is present for a moving sound source. We expand spatialisation techniques to include time-variance in order to produce the Doppler shift. Recordings of several different loudspeaker layouts demonstrate the presence of Doppler with and without time-variance, comparing this to the pre-calculated theoretical values.

1. INTRODUCTION

There are various techniques for rendering spatial audio over loudspeakers. Most techniques are based on being able to localise the direction of the virtual audio source. In this paper we investigate if spatial audio techniques inherently include Doppler shift when virtual sources are moved.

Ahrens [1] and DeVries [2] have both shown that Doppler is an inherent part of wave field synthesis. They have shown that the observed Doppler shift does not match the expected Doppler shift, with the error increasing as speed increases. Ahrens later expanded the theory to include retarded time which makes the expected and observed Doppler shift error decrease.

2. DOPPLER THEORY

It is well known that sound emitted from moving sources exhibit a frequency shift on the emitted sound, causing it to be heard higher in pitch when moving towards a listener and lower in pitch when it is moving away from a listener. The Doppler shift formula may be given as [3],

$$\omega_l = \omega_s \frac{1 + \frac{V_{ls}}{c}}{1 - \frac{V_{sl}}{c}} \tag{1}$$

where ω_l and ω_s are the listener and source frequency, v_{sl} is the source to listener velocity, v_{ls} is the listeners velocity and c is the speed of sound. Velocity is positive if the source is moving away from the listener and negative if it is moving towards the listener.

2.1. Passing Doppler

If we now consider the realistic case where the sound source is passing on a line parallel to the listener. The velocity component associated with the doppler shift is the velocity associated with the trajectory towards the listener. The relationship is based on the θ, angle between the sources direction and the listener such that:

$$v_{ls} = v_s \cos(\theta) \tag{2}$$

so now we are using the velocity experienced by the listener.This makes sense since at the instantaneous moment the source is directly in front of us, the heard frequency is identical to the emitted frequency. Since we have yet to define θ the doppler shift still cannot be calculated. θ is a simple geometric problem solved by Pythagoras theorem:

$$\theta = \tan^{-1}\left(\frac{\overline{x}_{ly} - \overline{x}_{sy}}{\overline{x}_{sx} - \overline{x}_{lx}}\right) \tag{3}$$

L_x and L_y are the x and y listener coordinates and S_x and S_y are the source coordinates. Inserting this into 1 gives:

$$\omega_l = \omega_s \frac{c}{c - v_{sl}\cos\theta} \tag{4}$$

or further expanded out to:

$$\omega_l = \omega_s \frac{c}{c - v_{sl}\cos\left(\tan^{-1}\left(\frac{L_y - S_y}{S_x - L_x}\right)\right)} \tag{5}$$

Now by using this equation we can obtain the instantaneous doppler shift heard at the listener position at any point in time.

2.2. Circular Doppler

Doppler for the case of a listener anywhere in a sphere has previously been investigated and derived by Smith [3] as part of designing a Leslie speaker simulation. He derived the velocity, v_{sl} as:

$$v_l = \frac{-r_l r_s \omega_m \sin(\omega_m t)}{r_l^2 + 2r_l r_s \cos(\omega_m t) + r_s^2}\begin{pmatrix} r_l - r_s\cos(\omega_m t) \\ -r_s \sin(\omega_m t)\end{pmatrix} \tag{6}$$

where r_l and r_s are the radii from the circles centre to listener and the source respectively and ω_m is the angular velocity. The equation is now dependent on time, yielding a different velocity at every point in time, that is overlapped if the radii remain the same.

However, if we were to take the case where the listener and source are both free to move and therefore their radii change, then we derive an equation that is based not only on time, t, which is solely when the source emits its sound, as in (6), but include the time, τ, when the sound is emitted by the source. The full doppler derived equation becomes:

$$\omega_l = \omega_s \frac{c - \overline{v}_l(t).[\overline{x}_l(t) - \overline{x}_s(\tau)]/\|\overline{x}_l(t) - \overline{x}_s(\tau)]}{c - \overline{v}_s(t).[\overline{x}_l(t) - \overline{x}_s(\tau)]/\|\overline{x}_l(t) - \overline{x}_s(\tau)]} \tag{7}$$

where $\overline{x}_l$ and $\overline{x}_s$ are the listener and source vectors, $\overline{v}_l$ and $\overline{v}_s$ re the listener and source velocities. Once we alter the terms to include the radii instead of vectors, tangential velocity, v_t rather than angular velocity and take the listener to be a static position we end up with the following equation for use in our tests later on :

$$\omega_l = \frac{\omega_s c}{c + \frac{r_l v_t \sin(v_t t/r_s)}{\sqrt{r_s^2 - 2r_l r_s \cos(v_t t/r_s) + r_l^2}}} \tag{8}$$

3. INITIAL TESTING

An informal test involving turning on and off in turn 1 of 16 speakers in a line, spaced 11cm apart. The audio source used was a 1kHz sine wave taking 10 seconds to travel from one end to the other. The microphone was placed at 0.5, 1 and 2 meters from the centre of the speaker array.The authors were somewhat surprised that as shown in figure (3), although the travelling sound was heavily discretised the doppler curve was still highly visible.

Informal testing of Ambisonics was done using a source travelling from -5m to 5m on the tangent with azimuth of 0 and elevation of 0 in the horizontal plane. The recorded results of this experiment

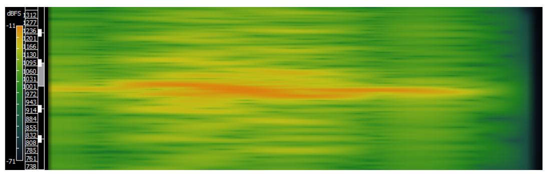

Fig. 1: Initial Doppler test using 16 discretised speakers to move a 1 kHz from left to right over 10 seconds

showed no deviation in pitch whatsoever. These preliminary experiments suggest that techniques based on amplitude may not exhibit the Doppler effect, and that a time-variant factor is required.

4. TIME-VARIANT PANNING

By altering panning functions such that they are time-variant we create a more realistic sound field representation that includes the Doppler shift of moving sound sources, thus creating a more realistic reconstruction of a real sound field. In this section we will describe the time-variant formulas that we are going to test and give the results in the following section.

4.1. Stereo Panning

Stereo panning is used to place a sound source in between two loudspeakers, this creates phantom sources which the listener perceives so that an audio scene can be described using just two speakers. By using the cosine/sine panning law constant power is kept so that the perceived auditory level remains the same as a sound moves from left to right. The spacing between standard stereo loudspeakers is 60^o, however the angle used for the panning law is 90^o, a simple linear remapping is used to convert between loudspeaker aperture placement and panning law. The cosine/sine law is given as:

$$\begin{aligned} G_L &= S_i \cos\theta \\ G_R &= S_i \sin\theta (0^o \leq \theta \leq 90^o) \end{aligned} \tag{9}$$

Where G_L and G_R are the left and right loudspeaker gains respectively and S_i is the input signal. This produces an arc between the loudspeakers on which the sound sources are placed. The perception of the sound scene can be increased and be more natural by adding in reference to distance. By doing this a distance attenuation and time-delay are needed as well as the standard panning law.

$$\begin{aligned} G_L &= S_i \cos\theta D(t) \\ G_R &= S_i \sin\theta D(t) (0^o \leq \theta \leq 90^o) \end{aligned} \tag{10}$$

where $t = d/c$, d is the distance between the virtual source and the panning arc, t is the time delay used and S_i and D is a suitable distance attenuation law such as $1/d$.

The time-variant panning for stereo can be implemented in one of two ways, shown in Figure (2):

- An audio delay line can be used for the left and right speaker respectively per sound source, driven by the source's distance as shown above. This creates $2N$ amount of delay lines.
- A Single audio delay line can be used as well as a control data delay line, where the control data delay line is a delay in the panning angle change reaching the spatialisation renderer. This is more computationally efficient as there is only $1N$ audio delay lines and an additional 1 control data delay line which is a lot more computationally efficient.

4.2. Ambisonics

Ambisonics is a method of representing a three dimensional sound scene by using the theory of spherical harmonics. The sound scene is decomposed into a finite order, M of spherical harmonics, of which contains $-M$ to M components. Ambisonics was first introduced to the audio community by Gerzon [4]. Ambisonics was originally restricted to first order and so had 4 components, 3 for the first order and 1 for the zeroth order. When using Ambisonics the highest order that is being used is stated and it is a prerequisite that all preceding orders below it are used. The use of a finite order does however limit the localisation accuracy of a sound source. When reproducing a three dimensional sound scene

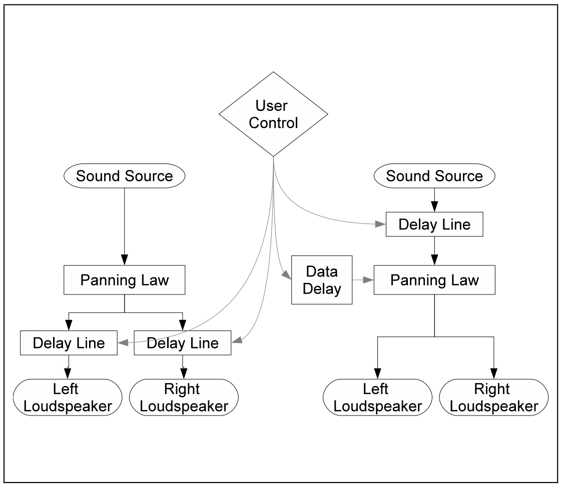

Fig. 2: Two possible ways of implementing a Stereo time-variant constant power panning law.

a reproduction array must be used, where the loudspeakers are equiangular and equidistant from the origin, the listener is located at the origin , and the minimum number of loudspeakers is determined by the order $N = (m+1)^2$.

More recent work on Ambisonics by Daniel [5] has made use of order of Ambisonics above the first. Daniel gives the 3D spherical harmonics as:

$$Y_{mn}^{\sigma(N3D)}(\theta,\phi) = \sqrt{2m+1}\tilde{P}_{mn}\begin{cases}\cos n\theta & \text{if } \sigma = +1\\ \sin n\theta & \text{if } \sigma = -1\end{cases}$$

$$\tilde{P}_m n(\sin\phi) = \sqrt{(2-\phi_{0,n})\frac{(m-n)!}{(m+n)!}P_{mn}(\sin\phi)} \tag{11}$$

where $\phi_q, q' = 1$ if $q = q'$ and 0 otherwise. P_{mn} is the Legendre function and $\tilde{P}_{mn}$ the Schmidt semi-normalised version. In the authors' experiments only horizontal reproduction will be used and the conversion to N2D is given as:

$$Y_{mn}^{(N2D)} = \sqrt{\frac{2^{2m}m!^2}{(2m+1)!}}Y_{mn}^{(N3D)} \tag{12}$$

To recompose the sound field over loudspeakers we calculate the vector of signals, S, in relation to the spherical harmonic vector, B, and the vector of loudspeaker spherical harmonics $Y_{mn}^{\sigma}(\theta_l,\phi_l)$, C, by:

$$S = pinv(C)B \tag{13}$$

We then take the sound source as a function of time by adding an auditory delay line before the Ambisonics encoding equations and also delaying the horizontal angle data by the same amount of time as shown for the Stereo time-variant panning.

4.3. Vector Base Amplitude Panning

Vector Base Amplitude Panning by Pulkki [6], is a spatialisation method for placing a sound source around a spherical loudspeaker array where the listener is at the centre of the array. The method uses a triplet of speakers at any given time to place a source in 3 dimensions keeping constant power, so $\sqrt{\sum g_1^2 + g_2^2 + g_3^2 = 1}$. The loudspeaker triplet gains are calculated by:

$$g_{123} = p^T L_{123}^{-1} \tag{14}$$

where g_{123} is the vector of loudspeaker gains, p is a vector containing the 3 dimensional coordinates of the perceived position of the sound source and L_{123} is a matrix containing the 3 dimensional coordinates of each loudspeaker within the active triplet.

Again, in this paper we are concerned with horizontal reproduction where all loudspeakers lie on a circle and so Vector Base Amplitude Panning effectively reverts to being the cosine/sine panning law between the nearest two speakers. If the sound source is directly placed at the same location as a speaker then only that loudspeaker will be active.

The time-variant implementation is carried out as in previous methods by delaying the audio signal by an amount $t = d/c$ before the spatialisation technique is applied.

4.4. Wave Field Synthesis

Wave Field Synthesis is a reproduction method for spatial audio that reconstructs the wavefront of a virtual sound source. The speaker driving function for Wave Field Synthesis is based on the Rayleigh I integral which is from the Helmhotz-Kirschoff Integral describing a homogenous area bounded by loudspeakers, monopole in the Rayleigh I integral and monopole and dipole in the Helmhotz-Kirschoff integral. The speaker driving function, Q, is given as:

$$Q = \cos(\theta_n)\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{r_0}{r_o + s_0}}\frac{\exp^{-jkr_n}}{\sqrt{r_n}} \qquad (15)$$

where θ_n is the angle between the virtual source and loudspeaker n, r_0 is the z axis (back-forth) distance between the virtual source and the loudspeaker array and s_0 the z axis between the loudspeaker and the reference line and r_n is the distance between the virtual source and loudspeaker n. In the time-domain we use:

$$A = \cos(\theta_n)\frac{1}{\sqrt{2\pi}}\sqrt{\frac{r_0}{r_0 + s_0}}\frac{1}{\sqrt{r_n}} \qquad (16)$$

and finally the time delay is implemented as:

$$t_n = r_n/c \qquad (17)$$

where c is the speed of sound in the homogenous reproduction area.

Amplitudes and time delays are calculated for each speaker based on each sound source within Wave Field Synthesis and so can lead to a large amount of computation in order to reproduce a large sound scene.

4.5. Discretized Line Array Methods

As in our original initial test of the presence of Doppler shift a line array will be used where the sound is moved across the array by use of discretized movement via the individual loudspeakers. There are two ways in which we do this:

- cosine / sine ramping between the loudspeakers on which the sound source lies. This is in effect the same as Vector Base Amplitude Panning methodology.
- Individual loudspeakers are turned on and off by a gating function: $round((P - L_1)/\Delta x) + 1$, where P is the sound sources exact position, L_1 is the left most speaker and Δx is the distance between loudspeakers used.

Using these methods allows the comparison between spatialisation techniques and synthesised movement of a sound source over closely spaced loudspeakers.

5. TESTING

In this section we will describe the tests carried out using time-variant formulas of Ambisonics and Vector Base Amplitude Panning over a horizontal loudspeaker array and the tests done on a line of 24 speakers using Wave Field Synthesis and discrete movement of sound. We will discuss the recorded results from our experiments.

5.1. Ambisonics and Vector Base Amplitude Panning

Both Ambisonics and VBAP were tested over a circular loudspeaker array featuring 12 Meyer MM4-XP loudspeakers. A DPA 4006 omni-directional microphone was used for measurements. The tests done for both spatial techniques were the same, and were:

1. 1kHz and 10kHz sine tone moving on a tangent from -5m to 5m at an azimuth and elevation of 0^o travelling at 10m/s from left to right in the horizontal plane
2. Circular motion of 1kHz and 10kHz sine tones around the array travelling at 10m/s with the microphone placed at 0.0, 0.5 and 1.0m to the left of the centre.
3. The 1kHz and 10kHz sine tones moving from 100m to 0m (on the loudspeaker array).

For the tests we used Ambisonics order 1 to 5 inclusive using all 12 loudspeakers, 1st order using 4 loudspeakers, 2nd order using 6 loudspeakers and Vector Base Amplitude Panning using 4, 6 and 12 loudspeakers.

The loudspeaker array had a radius of 1.5m and the speakers were placed equiangular starting at 0 radians with a spacing of $\pi/6$ radians.

5.1.1. Tangential Results

Figure (3) shows the 1st to 5th order Ambisonics average error results using 12 speakers, figure (4) shows a spectrogram 1st order Ambisonics using 4 speakers and figure (5) is 2nd order Ambisonics using 6 speakers. The average error results for Vector Base

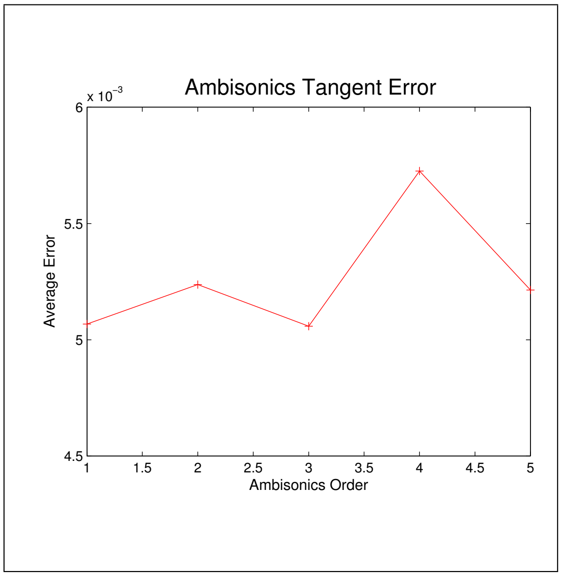

Fig. 3: The average error produced from the time-variant Ambisonics formulas for orders 1 to 5 where the sound source moving on a tangent at 0 radians with a speed of 10m/s from -5 to 5 using 12 speakers in a circle.

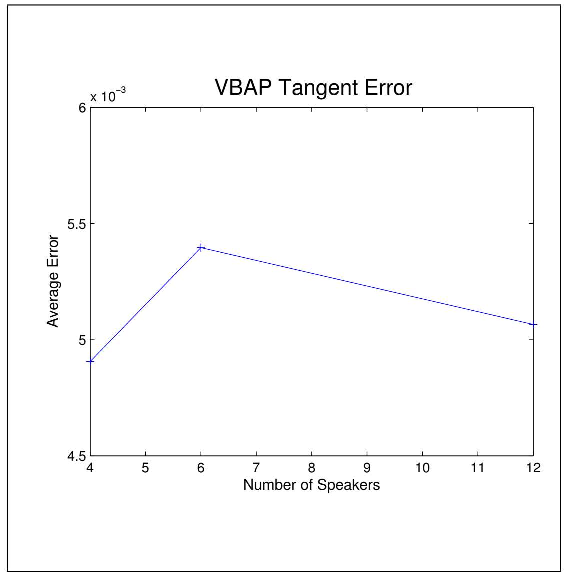

Fig. 6: The average error produced from the time-variant Vector Base Amplitude Panning where the sound source moving on a tangent at 0 radians with a speed of 10m/s from -5 to 5 using 4, 6 and 12 loudspeakers.

Amplitude Panning are shown in figure (6) for 4 , 6and 12 loudspeakers all for 1kHz. All spectograms are produced using a window of 9600 samples with 4x oversampling and 0.9375 overlap using a Hann window. The results for the average error of Doppler shift in regards to Ambisonics order does not show any general trend. In most cases of measured property against Ambisonics order the performance will improve with the increase in order, but we do not find that in the case of reproducing Doppler. For the increase in speakers for Vector Base Amplitude panning no pattern emerge either and the difference in average error between the results is less than 0.0005.

As the order of Ambisonics increases the smearing across frequency bands of the doppler shift appears to become less, especially noticeable when the sound passes directly in front of the microphone.

The results for 10kHz showed errors in the recorded results. As such figure (7) shows the results for 5th order Ambisonics and Vector Base Amplitude Panning both using 12 speakers which should give the least error for both spatialisation techniques. The recorded results show that there the sound has two simultaneous frequencies at certain points and overlapping Doppler shifts rather than the smooth shifts that are seen for the 1kHz results.

5.1.2. Circular Results

Figure (8) shows a comparison of 5th order Ambisonics and Vector Base Amplitude Panning against the theoretical Doppler shift at a distance of 1m from the centre of the array both using 12 loudspeakers and a 1kHz tone. Unfortunately for none of the tests the results showed sign of Doppler shift. Both these methods are intended for the listener to be at the centre of the array where no shift should be present and correctly no shift was recorded at the 0m position. In the Ambisonics case cancellation of the shift could be explained by the negative lobes of the reconstructed signals and the error of reconstruction at that distance would be fairly high. The positive outcome can be seen that the time-variant version performs as it should. This is in fact the same as the normal panning since this Doppler shift relies on the physical locations and therefore the loudspeaker delays in reaching the microphone position rather than the renderers time-delay lines.

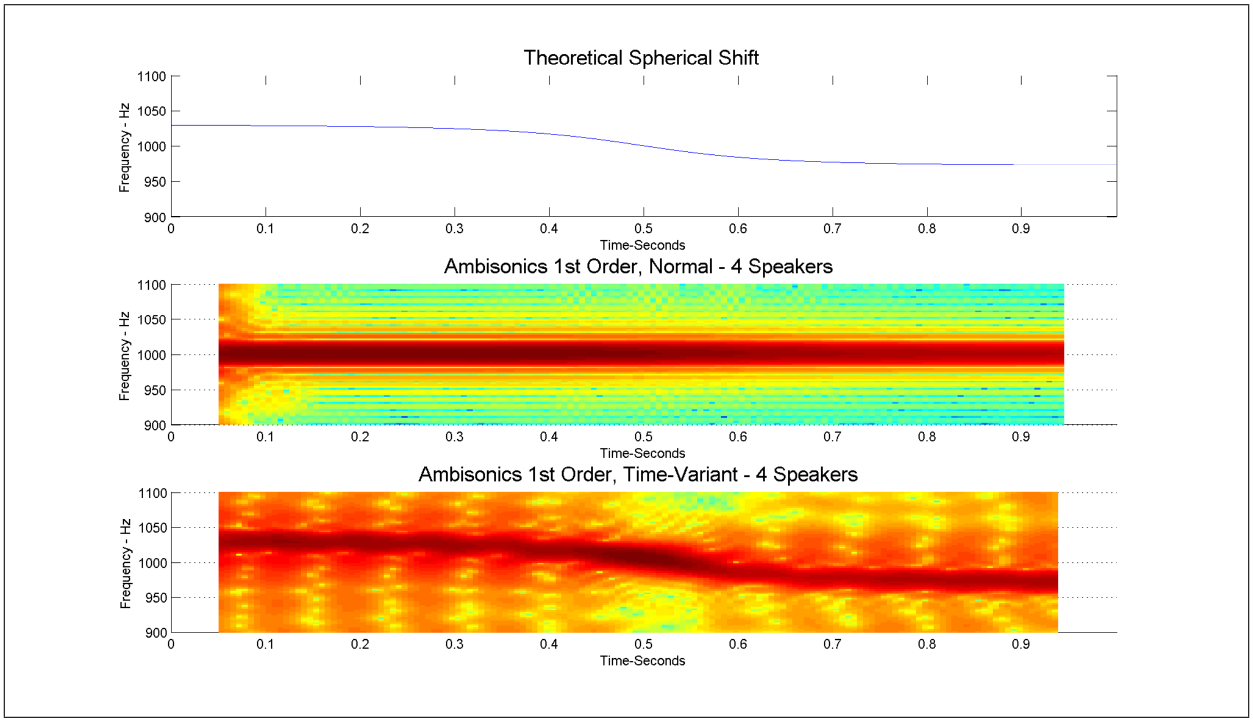

Fig. 4: 1st Order Ambisonics using 4 speakers with the sound source moving on a tangent at 0 radians with a speed of 10m/s from -5 to 5.

5.2. Back-Forth Results

From the standard Doppler formula for the sound source moving directly towards the microphone we expect a constant frequency shift. Figure (9) shows the results for 1st order ambisonics all using 12 speakers for 1kHz and figure (10) shows Vector Base Amplitude Panning using 4 speakers for 10kHz. The results for 1kHz show a constant frequency shift with 0 error, however once again for 10kHz the reproduction has multiple perceivable frequencies.

5.3. Wave Field Synthesis and Discrete Line Array Testing

We set up an array of 24 Meyer MM4-XP speakers in a line adjacent to one another spaced 0.14m apart.

The tests for wave field synthesis had the DPA 4006 omni-directional microphone placed 0.5m and 1m from the centre of the array and were:

1. 1kHz and 10kHz sine tone moving from left to right at 10m/s

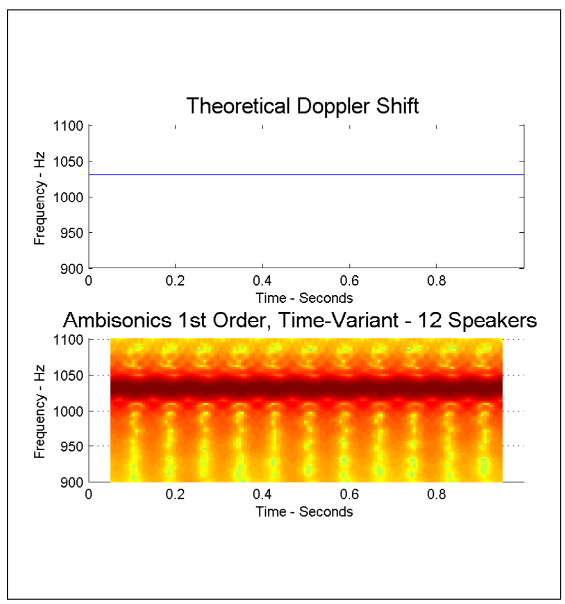

Fig. 9: 1st Order Ambisonics showing a 1kHz source moving towards the listener at 10m/s.

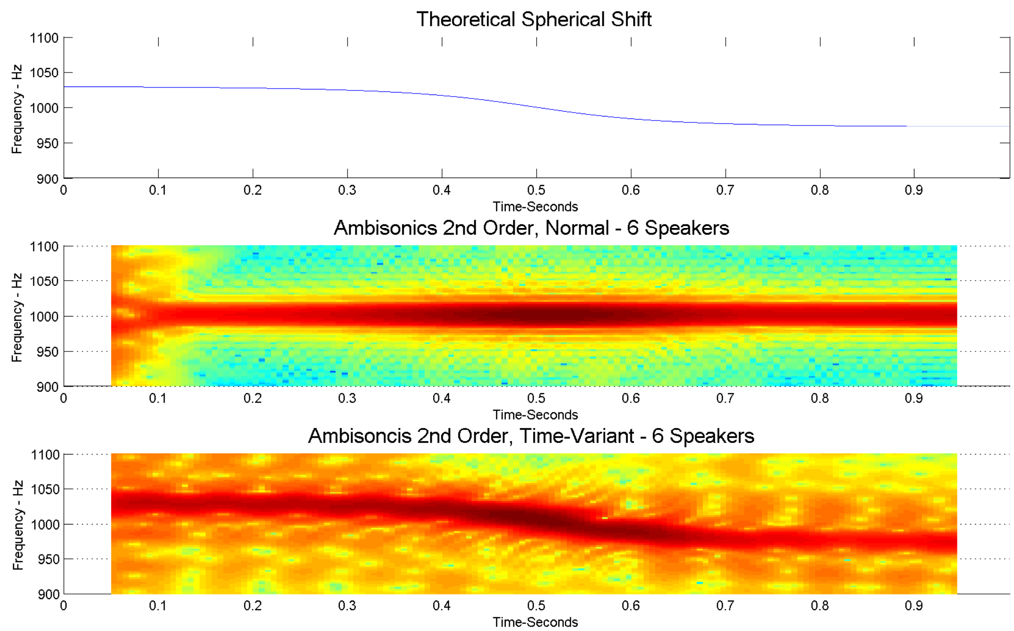

Fig. 5: 2nd Order Ambisonics using 6 speakers with the sound source moving on a tangent at 0 radians with a speed of 10m/s from -5 to 5.

2. 1kHz and 10kHz sine tone moving from 101m to 1m at 10m/s

Both tests were carried out at both distances for Wave Field Synthesis and the two discretized methods. For the discretized methods we carried out the tests using every speaker then every 2nd, 3rd 4th 5th and 6th speaker so that a full comparison of the method can be carried out.

5.4. Across Results

Figure (11) shows the results for a 1kHz measured at 1m using from every 1 to every 6 speakers to move the source, whilst figure (12) shows the result of Wave Field Synthesis for the same setup. As one might expect the results for using every speaker are the best for the discretized methods with the Doppler shift becoming less aparant with the larger gap between speakers. The ramping method had a slightly lower average error in most cases, this can be accounted for by the inclusion of amplitude panning. Perceptual differences between the methods could be significant. The result for Wave Field Synthesis showed multiple frequencies and not the exact Doppler shift we predict.

If we look at the result of 10kHz for the best case of 1kHz, that being using every loudspeaker, we can see in figure (13) that no frequency shift is detected. The frequency remains at 10kHz, albeit fluctuating slightly, but nowhere near what the Doppler shift should be.

5.5. Forward Moving Doppler

The forward moving Doppler implementation has already proven to work for Ambisonics and Vector Base Amplitude Panning. Now we look at the inherent implementation within Wave Field Synthesis. Figure (14) shows that the intended Doppler shift is produced but at the same time so is a shift lower than the source frequency.

6. CONCLUSIONS

We have shown that by including time-variant versions of spatial audio techniques that intended

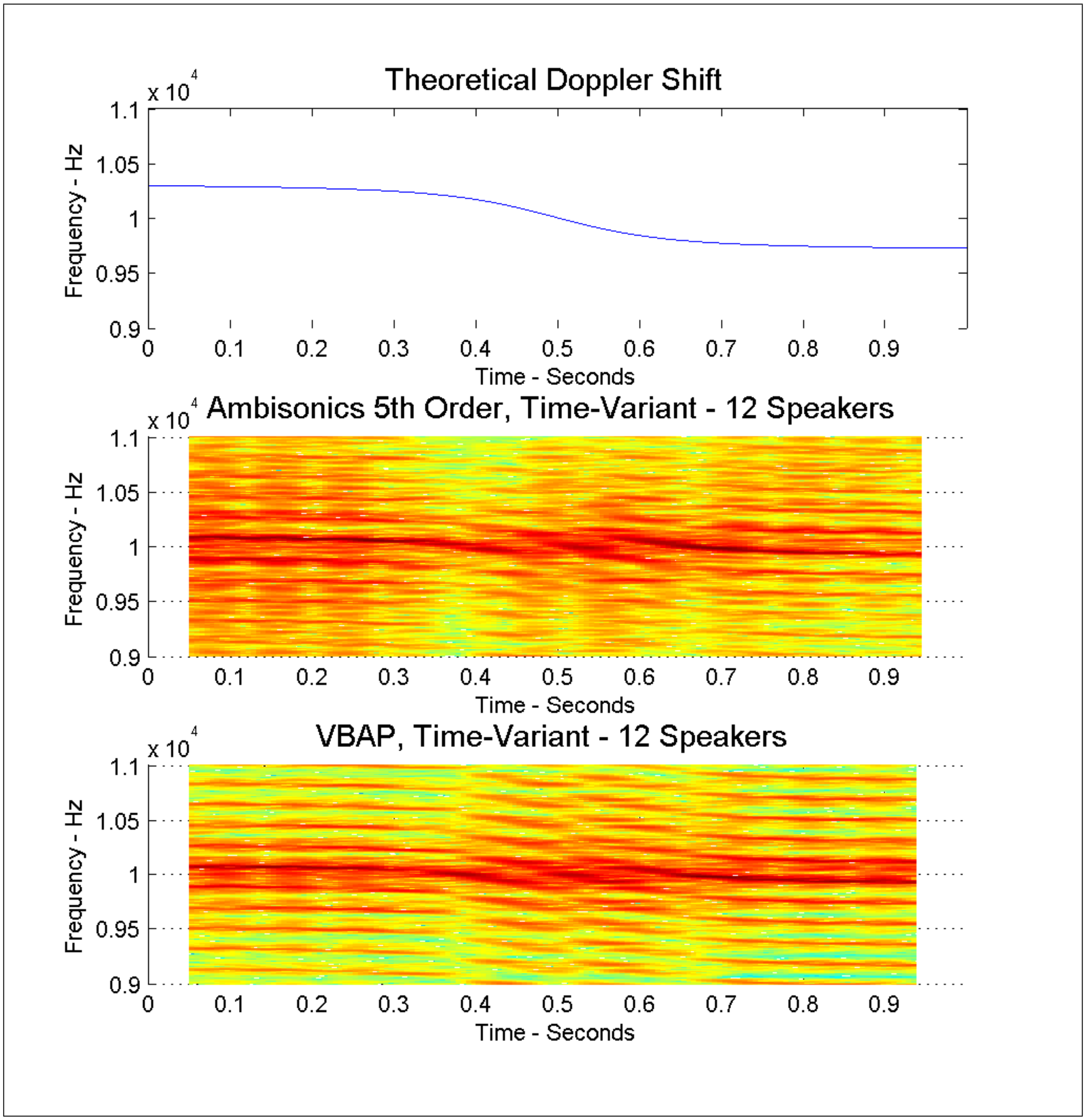

Fig. 7: Comparison of 5th Order Ambisonics and VBAP using 12 speakers with the 10kHz source moving on a tangent at 0 radians with a speed of 10m/s from -5 to 5.

Doppler shifts can become an inherent part of the spatialization. This produces a more realistic and naturally true audio reproduction of the sound scene. The circular motion Doppler test that relied on the physical properties of speakers and the microphone placement did not show the Doppler shift for

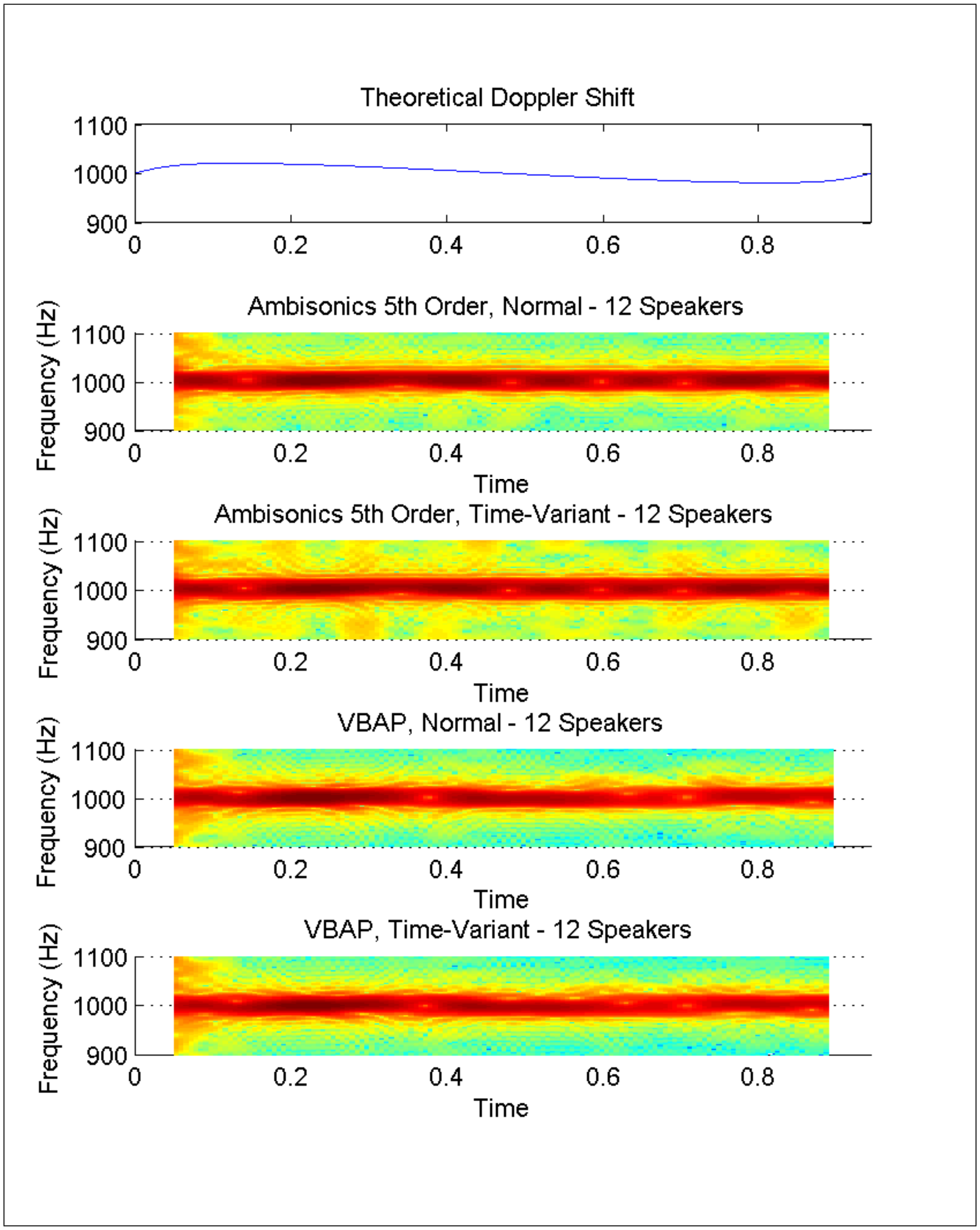

Fig. 8: Comparison of 5th Order Ambisonics and VBAP using 12 speakers with a 1kHz sine tone moving round a circle of radius 1.5m at 10m/s where the microphone is placed 1.0m from the centre.

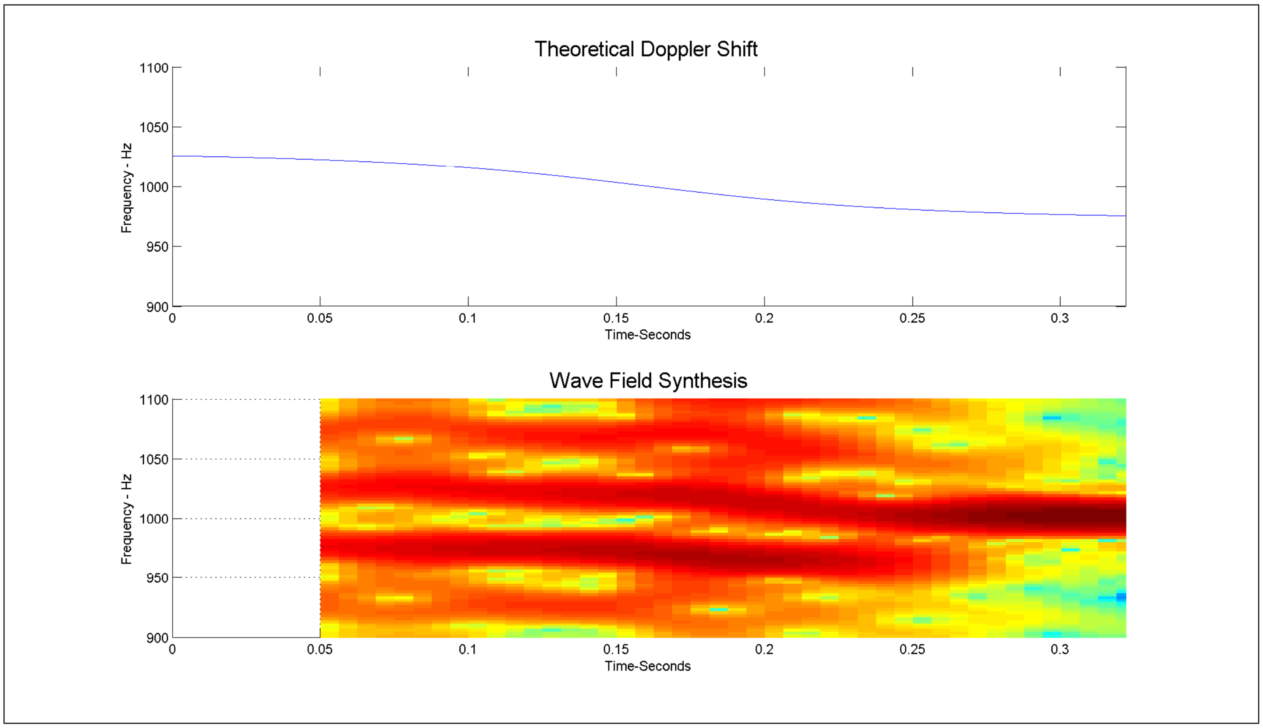

Fig. 12: 1kHz sound source moving left to right across the speaker array using Wave Field Synthesis.

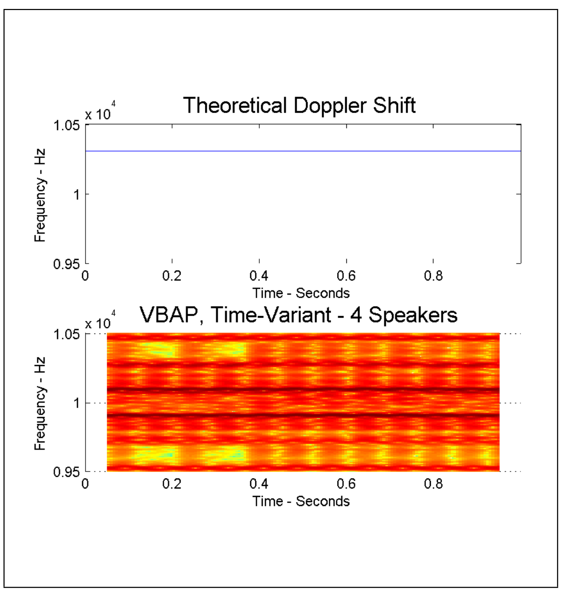

Fig. 10: Vector Base Amplitude Panning using 4 loudspeakers for a 10kHz source moving towards the listener at 10m/s.

either Ambisonics of any order or Vector Base Amplitude Panning using different numbers of speakers, however both of the techniques investigated assume and require the listener to be at the centre of the array for the spatialisation to work correctly, where there would of course be no Doppler from the test we performed. The tangential Doppler results for both the circular array and line array gave good results when including the time-variance implementation. It was surprising that the Doppler shift in Ambisonics did not become dependent on the order used, neither did the amount of speakers used for Ambisonics or Vector Base Amplitude Panning have any significant effect. For the line array however, the larger the gap between the speakers the larger the error in Doppler shift, this was most noticeable on spectrograms of the results but the average error produced was still acceptable. In all cases the time-variance created a steady shift in frequency when a sound source was directly moving towards the listener. We have shown that Doppler shift can become an inherent part of spatial audio and its implementation can

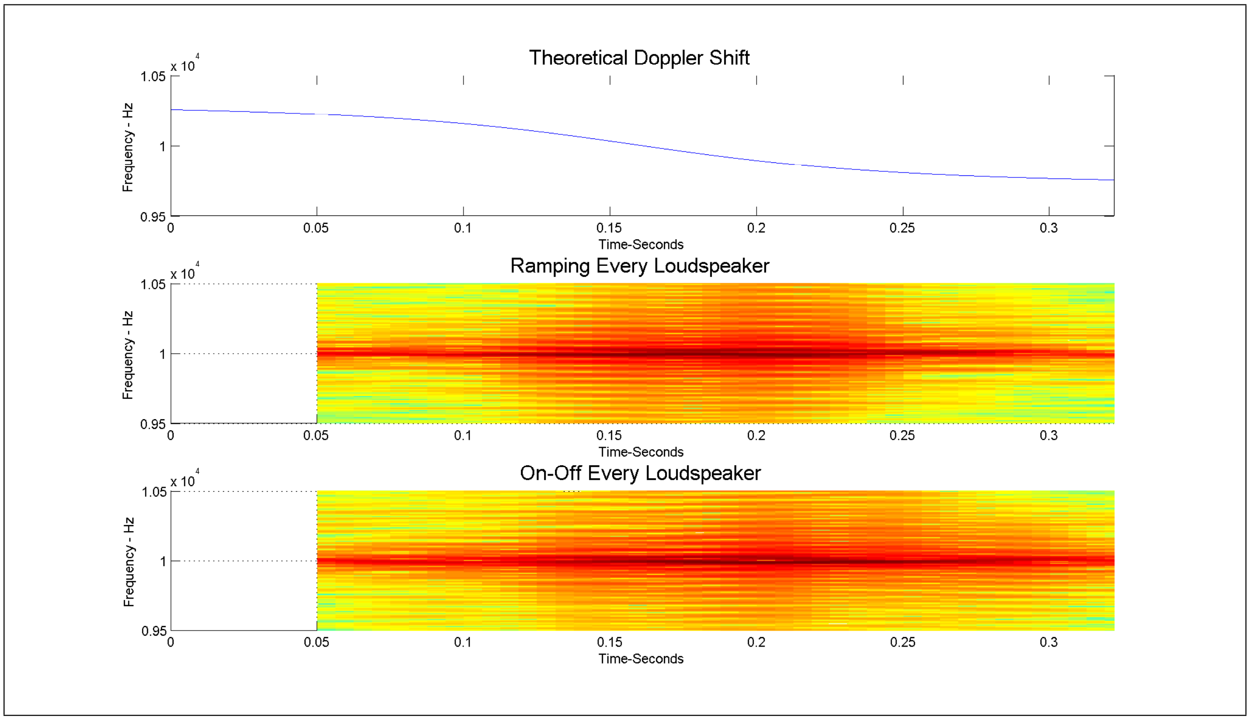

Fig. 13: 10kHz sound source moving left to right across the speaker array using every speaker.

be done quite easily to produce reliable results.

7. REFERENCES

[1] S. Ahrens, Jens; Spors, "Reproduction of moving virtual sound sources with special attention to the doppler effect," in *Audio Engineering Society Convention 124*, 5 2008.

[2] D. de Vries, "Sound reinforcement by wavefield synthesis: Adaptation of the synthesis operator to the loudspeaker directivity characteristics," *J. Audio Eng. Soc*, vol. 44, no. 12, pp. 1120–1131, 1996.

[3] J. Smith, S. Serafin, J. Abel, and D. Berers, "Doppler simulation and the leslie," *DAFX02*, September 26-28 2002.

[4] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Audio Eng. Soc*, vol. 21, no. 1, pp. 2–10, 1973.

[5] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format," in *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*, 5 2003.

[6] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc*, vol. 45, no. 6, pp. 456–466, 1997.

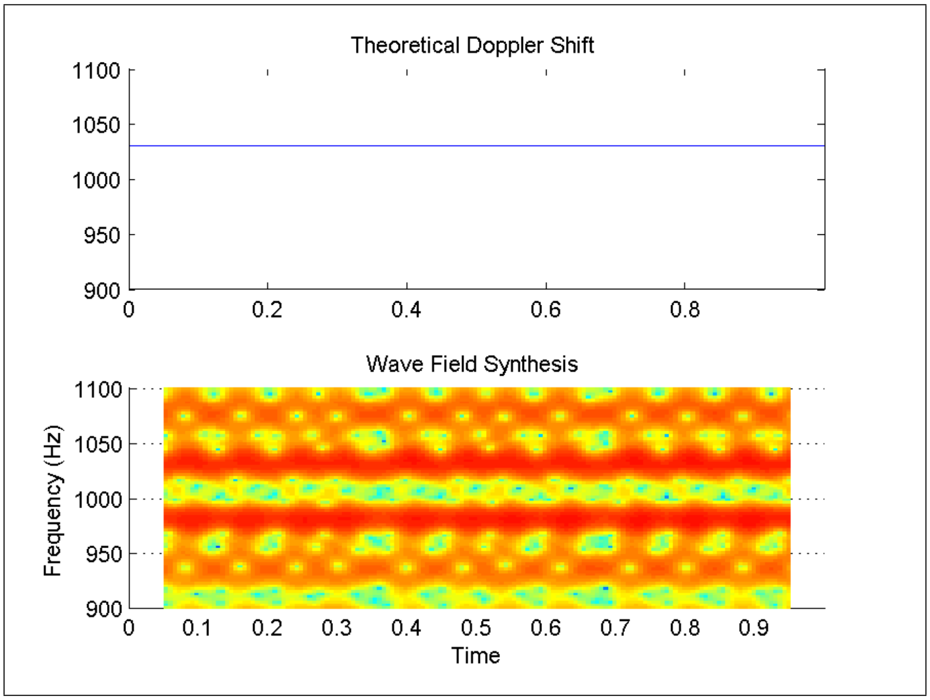

Fig. 14: 1kHz sound source moving towards the microphone at 10m/s.

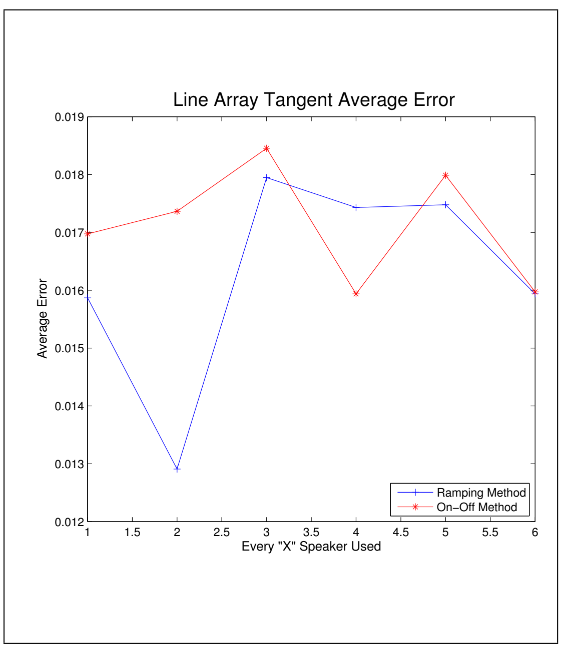

Fig. 11: 1kHz sound source moving left to right across the speaker array using Wave Field Synthesis.