Audio Engineering Society

Convention Paper

Presented at the 136th Convention
2014 April 26–29 Berlin, Germany

Ambidio: Sound Stage Width Extension for Internal Laptop Loudspeakers

Tsai-Yi Wu[1], Agnieszka Roginska[2], and Ralph Glasgal[3]

[1,2] *Music Technology, New York University, New York, NY, 10012, USA*

[3] *Ambiophonics Institute, Rockleigh, New Jersey, 07647, USA*

Correspondence should be addressed to Tsai-Yi Wu (tyw238@nyu.edu)

ABSTRACT

This paper introduces a sound stage width extension method for internal loudspeakers. Ambidio is a real-time application that enhances a stereo sound file playing on a laptop in order to provide a more immersive experience over built-in laptop loudspeakers. The method, based on Ambiophonics principles, is relatively robust to a listener's head position, and requires no measured/synthesized HRTFs. The key novelty of the approach is the pre/post-processing algorithm that dynamically tracks the image spread and modifies it to fit the hardware setting in real-time. Two detailed evaluations are provided to assess the robustness of the proposed method. Experimental results show that the average perceived stage width of Ambidio is 176° using internal speakers, while keeping a relatively flat frequency response and a higher user preference rating.

1. INTRODUCTION

Mobile devices with stereo speakers such as laptops, tablets are increasingly popular. According to Google's research people spend 4.4 hours on average in front of screens each day during leisure time [14] and they mostly use their laptop for entertainment. [19] Even though audio is essential to most popular laptop entertainment, e.g. games, movies, and music, the performance of laptop internal speakers is not yet satisfactory. [9] Indeed, the sound of laptop speakers is limited by the tight space constraints, and becomes even tinnier when it goes across the keyboard. This results in a narrow and unrealistic stereo image that makes all sound perceived as coming from one direction, and appear to be more monophonic. There is a need to improve the sound quality of built-in laptop loudspeaker sound to accompany the rich graphics users normally get from the screen, for more immersive experiences.

The perceived width can be widened by applying spatial enhancement techniques. [24] Achieving spatial enhancement passively during playback is ideal for laptop entertainment since it would be compatible with existing stereo recordings. The previous general approach was to create virtual surround sound speakers by simulating the effect of Head-Related Transfer Functions (HRTFs). Also the ideal listening area of the virtual surround speakers is then very narrow in stereo loudspeaker reproduction, and the effect is sensitive to an individual's body shape. [5][15][27] This paper proposes a new approach that addresses these problems.

This present work aims to provide laptop entertainers another option to deliver an improved overall immersive experience and relatively unrestricted range of motion without a complicated setup or required additional equipment. The main algorithmic contribution leverages the combination of Ambiophonics principles [12] and dynamic Mid-Side techniques to project any stereo components playing in a laptop to a wider sound stage, on-the-fly, while keeping a static center image. The present algorithm works without any measured/synthesized HRTFs. A Mac OS X menu bar application (Ambidio) is introduced. Finally, the performance of the present algorithm is validated by comparing it with the competitive technology, SRS iWOW, and traditional stereo in both subjective and objective ways.

2. BACKGROUND

2.1. Crosstalk Cancellation

Crosstalk is an inherent problem in stereo loudspeaker playback. It occurs when a sound "falsely reaches the ear on the opposite side from each speaker."[15] Unwanted spectral coloration is caused because of constructive and destructive interference between the original signal and the crosstalk signal. Moreover, conflicting spatial cues are created and cause spatial distortion. As a result, localization fails and the stereo image collapses to the position of the loudspeakers. The solution to this problem is a crosstalk cancellation (XTC) algorithm. XTC refers to adding a crosstalk canceling vector to the opposite speaker to acoustically cancel the crosstalk signal at a listener's eardrum. With properly positioned speakers, XTC also has the ability to enhance traditional stereo recording by simulating virtual speakers located further apart than the real ones as in normal Blumlein stereo. The standard was to use generalized HRTFs to represent the angles of the two physical loudspeakers for XTC. [15] [23]

2.2. Spatial Enhancement

The idea of spatial enhancement, or stereo image widening, can be traced back to the time when stereophonic audio first spread around the world. Many early experiments were done in artificial stereophony and its optimization. See [11] for review. Atal introduced a method to produce virtual widening of the loudspeakers for phantom sources. [2] Virtual loudspeakers are located in between the real loudspeakers. This was followed by the transaural system. [8] Based on the same idea, [16] proposed a stereo widening method that adds a slightly attenuated, phase-shifted, and delayed version signal to its opposite channel. This principle is further elaborated by [1] and others to produce a more natural stereo widening effect. As the modern high-speed digital signal processing chips have gotten faster, it is possible to compute more complicated algorithms in real-time. For example, [26] proposed a real-time method that would map the original stereo recording to virtual sound sources. When spatial enhancement is done during reproduction, i.e. at the user end, it passively changes the input stereo signal to increase the overall spatial perception without any special "guideline" from the original audio producer. The method proposed in this paper belongs to this category.

2.3. Related Work

At the time of this writing, there are several commercial products that have spatial enhancement features implemented in different but also similar approaches. For example: products from Dolby Labs, QSound Labs, and Harman make use of generalized HRTFs. [20][3][6] Products from DTS (formerly SRS Lab), on the other hand, use frequency- and amplitude-dependent processing. Specifically, the side signal is filtered by a curve in the shape of average rear HRTFs. [17] Lastly, the BACCH filter introduced by Princeton University is an optimal crosstalk cancellation filter that creates a flat frequency response at the loudspeakers. Without involving generalized HRTFs, it uses actual impulse responses measured from a dummy head or listener's ear canals. [7]

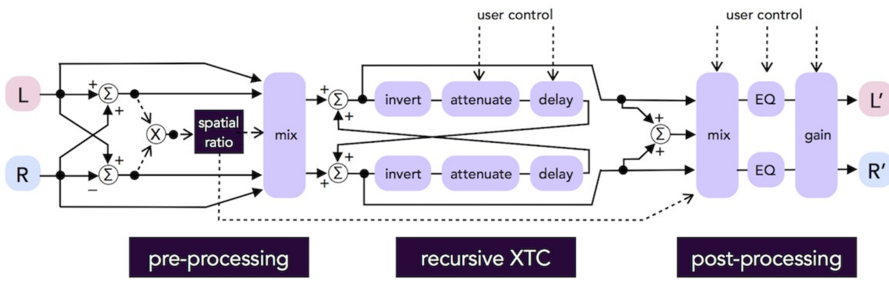

Fig. 1: Modified RACE algorithm. Signal flow (solid line), user controlled parameters (dashed line)

All of the above mentioned technologies use either analyzed or measured HRTFs to enhance immersion. As noted, HRTFs are very individualized, so none of the generalized HRTFs can compete with a listener's own. Listening with someone else's "ear" can cause spatial distortion resulting degraded sound localization. Furthermore, since the HRTFs have fixed angles, a listener needs to stay at the same fixed position as the HRTFs to get the benefit of the XTC. If a listener moves away from the sweet spot, which accordingly changes the HRTFs necessary to cancel the crosstalk, the image collapses. [4]

2.4. Ambiophonics

Introduced by Ralph Glasgal in the 1980s, Ambiophonics is an HRTF-free method taking advantage of psychoacoustic principles to project an existing stereo file up to a 180° stage by two closely spaced (10-20°) frontal speakers and a XTC filter. Ambiophonics is driven by a Recursive Ambiophonic Crosstalk Eliminator (RACE) algorithm, which cancels the crosstalk signal from the opposite speaker recursively. See [12][13] for a detailed theory of Ambiophonics.

Techniques based on Ambiophonics have several advantages. Foremost, there is no 30°speaker at each side, so there is no possibility to make the sound more like two point sources at the speakers. Also, the ideal listening area is wider and more robust with respect to the listener's head movement not only because the two loudspeakers are placed close to each other [25], but also because there is no synthesized/measured HRTFs involved. Hence, it allows listeners to listen with their own ears. Since the separation angle of laptop internal speakers is always smaller than the traditional 60° listening triangle, the laptop configuration is well suited for the application of the Ambiophonics techniques.

3. IMPLEMENTATION

The name "Ambidio" suggests the underlying theory—"Ambiophonics Audio." Ambidio is a stereo-in-stereo-out program mainly written in C++ and developed on Mac OS X 10.8. The program can be divided into four parts: First, a kernel is implemented which serves as a virtual audio device that captures the system audio output. The captured output is then routed to a separate application in the user space where the present spatial enhancement algorithm (Section 3.1) is applied. To keep flexibility, the default output can be either laptop internal speakers or external speakers. We also provided an intuitive user interface. (Section 3.2)

3.1. Crosstalk Cancellation Filter Design

As illustrated in Fig. 1, a spatial ratio is calculated from the correlation between channels. The stereo inputs are first sent to a mixing block, which balances the center image and the ambience sound based on a calculated spatial ratio and selected thresholds. By doing this, the side components are sure to be expanded efficiently by the XTC filter. The signals are then passed to the heart of the RACE algorithm—invert, attenuate, and delay. Parameters are optimized to match the physical configuration of the hardware.

After being processed by the RACE algorithm, the depth and the room ambience of a stereo stream that were once buried under the collapsed image is recovered. With such a feature, the audio content potentially appears to be farther in distance. Note that the use of artificial reverberation or even a small pan from the center would make the center image drift to the side. Therefore, the RACE-enabled output is adjusted based on the spatial ratio and a selected threshold. This keeps the center image strong and stable enough for listeners, as it is an important feature to make the content understandable. Next, an EQ is applied to eliminate the audible coloration in high frequency bands created by using non- ideal factors in respect to the size of the head and the laptop. Finally, a gain control block makes sure every signal is within the proper amplitude range and is at the same loudness as the original input signal.

The present algorithm can also be implemented in the frequency domain to improve computational performance. Unlike having infinite recursive crosstalk cancellation in time domain processing, it only reduces the crosstalk signals to -90 dB, yet still enough to make the crosstalk inaudible.

3.2. User Interface Design

Ambidio has a very simple and intuitive user interface. After launching the application, a small Ambidio icon appears on the menu bar. Clicking on the icon can bring up a drop-down menu, where users can quit the program or open the control panel. An advanced control panel is also available that allows users to optimize the effect.

4. SUBJECTIVE EVALUATION

4.1. Subjective Evaluation Design

The subjective experiment was designed following the paradigms of [22] and [10].

4.1.1. Participants and Apparatus

Forty-four graduate students and faculty members from New York University (17 females, 27 males, M$\pm$SE=27.0$\pm$5.0) voluntarily participated and were assigned to the same task. All participants had normal hearing. Seventy three percent of the subjects were music majors or worked in a music-related field. The experiment took place individually in the Spatial Auditory Research Lab of the Music Technology program at NYU. It was run on a MacBook Pro 15-inch. Participants were seated on a swivel chair. They were allowed to move freely, turn their head, adjust the height of the chair, and sit in any posture or distance they were comfortable with.

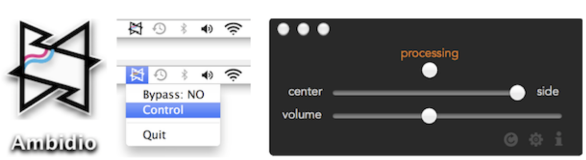

Fig. 2: User Interface of Ambidio

4.1.2. Process Methods

The evaluation included three different audio processing methods: (a) a reference without any processing; (b) the competitive technology, SRS iWOW (version 3.3); (c) the main output of this present work, Ambidio. SRS iWOW (chosen mainly because of its availability) was played using the default laptop built-in speaker profile as provided by the manufacturer. Ambidio, on the other hand, used the default value as it came out of the spatial enhancement algorithm with no extra center image adjustment.

4.1.3. Stimuli

Four 20-second sequences were selected for the experiment. To eliminate the inherent stage difference between sequences, all the sequences contained noticeable hard-panned sounds. It was designed to include different types of popular media entertainment experiences, including game, movie and music. As a result, there were a total of *12 clips* (3 methods x 4 sequences), and *176 trials* (44 subjects x 4 sequences). All clips were prerecorded by ProTools 10 and normalized to the same B-weighted sound pressure level to prevent any bias created by level differences as suggested in [22]. The normalized clips were stored as 44.1 kHz, 24 bit .wav files.

Name	Content	Resource	Video
Game	BioShock Infinite	gameplay	v
Movie	Jurassic Park (1993)	YouTube	v
Choir	Banquet Fugue	CD	
Jazz	Jazz music	downloaded	

Table 1: Description of the Selected Sequences

4.1.4. Procedure

Custom software was written to run and guide the subjects through the entire listening test. (Fig. 3) It provides interactive figures that help subjects to visualize a rating. The program randomly determines the order of the four sequences. In each trial, three clips with the same sequence were processed with different methods in a random order. Subjects were asked to judge the (a) stage width, (b) depth (perspective), and (c) presence (nearness) of each clip. Dummy variables were created to represent different levels of perceived depth/presence.

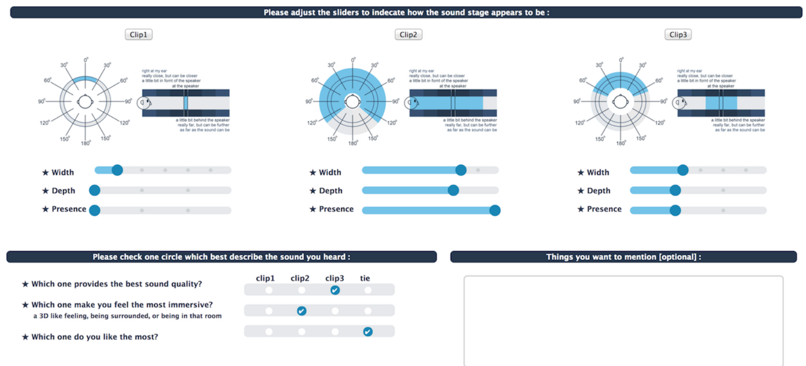

Fig. 3: The custom software used in the experiment

In addition, subjects also had to choose one clip that (d) has the best sound quality; (e) is the most immersive, and (f) is their favorite. They were allowed to listen to the clips more than once and go back and forth between trials until they felt satisfied with their answer.

4.1.5. Data Analysis

All the data were analyzed at an alpha level of .05. Levene's test was used to examine the variance's homogeneity whitin individual group. Either N-way ANOVA or Brown-Forsythe's test was used depending on the homogeneity. If the F value is larger than alpha, a post-hoc test would be performed by either Fisher's LSD or Scheffe's method depending on the sample sizes for each group.

4.2. Subjective Evaluation Result

No significant effect was found for subjects' background or the presence of the accompanying video for all rated features. There was no significant difference between the sequences unless otherwise specified. The results are reported as M±SE.

4.2.1. Perceived Width, Depth, and Presence

Considering only the trials where the reference clip is correctly rated, the perceived stage width is 176±3° for Ambidio. (N=44) The stage width boost was analyzed to eliminate the impact of the presentation order, the sequence types, and the differences within subjects. The average stage width boost of Ambidio is significantly larger than SRS iWOW. ($F(3{,}172)=5.15$, $p<.001$) In addition, Ambidio works equally well on four sequences ($F(3{,}172)=1.49$, $p=.22$), whereas significant differences were observed between sequences when the method is SRS iWOW. ($F(3{,}172)=5.15$, $p<.001$)

On the depth rating, 44% subjects rated the perceived depth as "far" (score≥2) when the process method is Ambidio, compared to 23% and 18% for SRS iWOW and the reference clip, respectively. . Ambidio provided deeper depth than the other two methods. ($p<.05$) The rated depth did not significantly differ between SRS iWOW and the hidden reference. ($p=.06$) For the presence rating, 68% participants rated the presence as "close" (score≥2) in clips processed by Ambidio, compared to 37% and

Score	Depth	Presence
0	right at the speakers	
1	a bit behind the screen	a bit in front of the screen
2	far, but can be farther	close, but can be closer
3	as far as it can be	right at my ears

Table 2: Dummy Variables for Depth and Presence

	Ambidio	SRS iWOW	reference
Average Absolute Width	176±3°	74±6°	59±0.7°
Average Stage Boost	111±4°	19±4°	-
Average Depth	1.32±0.07	0.86±0.06	0.75±0.06
Average Presence	1.80±0.07	1.23±0.07	0.82±0.06

Table 3: Perceived Rating Results (N=44, n=176)

21% in clips processed by SRS iWOW and the reference clip, respectively. The simple main effect test revealed that the perceived presence provided by Ambidio was significantly closer than the other two methods. ($p<.001$) Table 4 shows the descriptive results of each rated feature.

	Ambidio	SRS iWOW	reference	tie
Best Width	93%	2%	0%	5%
Best Depth	41%	8%	10%	41%
Best Presence	52%	12%	3%	33%

Table 4: Best Rating within Trials (N=44, n=176)

4.2.2. Sound Quality, Immersion, and Preference

Table 5 summarizes the result of the multiple-choice selections. The preference selection data were analyzed to examine their relationship to the best sound quality selection and the most immersive selection. As shown in Table 6, there were 84 trials where a subject was asked to choose between the sound quality and the immersion. 61% preferred the latter one. There were 80 trials in which a single clip was voted to have the best sound quality, to be the most immersive, and to be the preferred clip. Among 80 trials, 79 of them were Ambidio.

	Ambidio	SRS iWOW	reference	tie
Best Sound Quality	46%	30%	13%	11%
Best Immersion	94%	3%	1%	2%
Preference	73%	15%	5%	7%

Table 5: Best Rating within Trials (N=44, n=176)

5. OBJECTIVE MEASUREMENT

Quality	A	A	A	A		A	
Immersion	B	B	B	B		A	
Preference	A	B	C	tie		A	
Count		84 (48%)		12 (7%)		80 (45%)	
	Quality	Immersion	Either	-	Ambidio	SRS iWOW	reference
	30 (36%)	51 (61%)	3 (4%)	-	79 (99%)	1 (1%)	0 (0%)

Table 6: All Possible Combinations for Multiple Choices (N=44, n=176)

5.1. Objective Measurement Design

An objective experiment was conducted to examine the frequency response of Ambidio. The experiment took place in the same room as described above, and used the same apparatus as in subjective evaluation. A Neumann KU100 dummy head was used. It was set at the same height as the author's head facing down to mimic a listener watching the screen. Test signals were generated in 24-bit, 44.1 kHz: (*a*) identical pink noise in left/right channels (coherent signal for center image); (*b*) same as (a), but with the right channel at -6 dB (intermediate); (*c*) same as (a), but with the right channel 180° phase inverted. [22] A 20-second white noise signal was also generated in the same format.

Two impulse responses corresponding to each ear were recorded separately. (as shown in Fig. 4 top) The recorded data was computed by 2^{14} point FFT with Blackman-Harris window for harmonic isolation. The resulting spectrum was normalized and deconvolved using the corresponding impulse response. A $1/24^{\text{th}}$-octave filter was applied. If the algorithm introduced no coloration and eliminated all the crosstalk, the deconvolved result should be identical to the original sound file.

5.2. Objective Measurement Result

Fig. 5 shows the frequency responses of each test clip. Frequencies outside a window of 100-10kHz

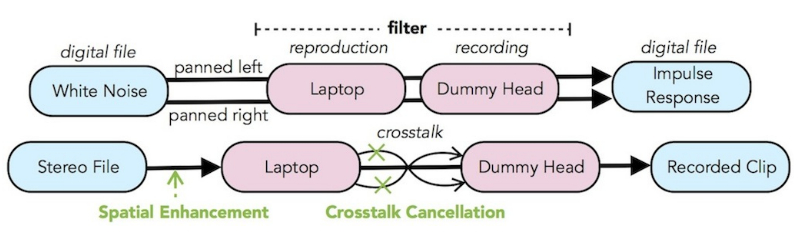

Fig. 4: Illustration of the Signal Chain in the Objective Experiment

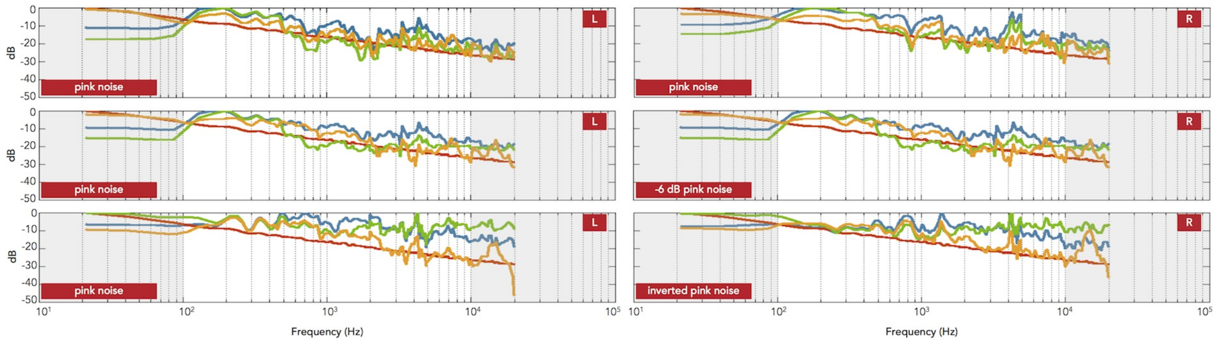

Fig. 5: The measured response. red: digital file; blue: reference; green: SRS iWOW; orange: Ambidio. The area covered in gray is too low/high for a laptop to reproduced, so they are not to be taken as shown.

may have some distortions, due to the laptop's inability to reproduce these frequencies accurately. As mentioned before, the ultimate goal is to fit the red line in Fig. 5, i.e. the original sound file. The error of the stereo reference was significant in all of the three test clips. It is also noteworthy that the three left curves for Ambidio were similar.

6. DISCUSSION AND CONCLUSIONS

The main results of the present paper show that Ambidio had an average stage width boost of 111°. (N=44)) This could be the reason that it is neither based solely on conventional widening methods, nor HRTFs. In addition, Ambidio not only performed better in depth and presence rating (Table 3), but also picked as the favorite clips among 73% of 176 trials. Also revealing from the selection result, widening the stage width could be the easiest way to create a feeing of immersion as nearly all (95%) of the clips selected as the most immersive were providing the widest soundstage width. On the other hand, the objective results show that spatial distortion appears to be reduced when crosstalk signals are minimized, leading to more successful spatial enhancement.

Immersion is an important feature for laptop entertainment, as underlined by results in the experiment described above where most subjects chose immersion rather than sound quality as the most important feature. It is noteworthy that one subject mentioned: "I already got used to listening to that kind of sound (of internal speakers)." It suggests that listeners may need time to adapt themselves to listen to sound that isn't purely coming from the front.

One of the limitations of Ambidio is that it does not support headphones since the underlying theories do not apply. This can still be done by simulating virtual speakers for headphones. Moreover, when there is no interaural differences available (both channels identical), no realistic stage width can be auralized.

With the increasing importance of media entertainment using consumer-grade hardware, passive spatial enhancement techniques have the potential to expand the sound stage during playback, thus providing a more immersive listening experience for gaming, movies, or pure music listening purposes. This paper introduces a method, Ambidio, to extend the width of the sound stage for built-in laptop loudspeakers. Ambidio relies on the traditional stereo inputs and does not need additional preprocessing to extend the perceived stage width, immersiveness and realism. A subjective experiment was conducted in which Ambidio was compared with a commercial spatial enhancement program—SRS iWOW by 44 subjects in an acoustic controlled environment using laptop internal speakers. Results from subjective tests indicate that the method presented here shows promise in significantly extending the sound stage width, increasing the sense of immersion and minimizing spectral coloration.

7. ACKNOWLEDGEMENTS

The author would like to thank Pei-Lun Hsieh, Shao-Ting Sun, and Cheng-I Wang for all the helpful conversations. Special thanks to Hao Li for his advice and comments on the manuscript.

8. REFERENCES

[1] R. M. Aarts. Phantom sources applied to stereo-base widening. *Journal of the Audio Engineering Society*, 48(3):181–189, 2000.

[2] B. S. Atal. Apparent sound source translator, 1966. US Patent 3,236,949.

[3] J. Bauck and D. H. Cooper. Generalized transaural stereo and applications. *Journal of the Audio Engineering Society*, 44(9):683–705, 1996.

[4] D. R. Begault. *3D-sound for virtual reality and multimedia*. Academic Press Professional, Boston, 1994.

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

[6] C. P. Brown. Surround sound virtualizer and method with dynamic range compression, 2011. US Patent 2011/0,243,338.

[7] E. Choueiri. Spectrally uncolored optimal croostalk cancellation for audio through loudspeakers, 2013. US Patent 2013/0,163,766.

[8] D. H. Cooper and J. L. Bauck. Prospects for transaural recording. *Journal of the Audio Engineering Society*, 37(1/2):3–19, 1989.

[9] Dolby Laboratories. Dolby survey finds entertainment features become must-have on pcs for college students, 2010.

[10] Dolby Laboratories. Pc postprocessing technologies: A competitive analysis, 2011.

[11] J. Eargle, editor. *Stereophonic Techniques: An anthology of reprinted articles on stereophonic techniques*. Audio Engineering Society, 1986.

[12] R. Glasgal. 360° localization via 4. x race processing. In *Audio Engineering Society Convention 123*, 2007.

[13] R. Glasgal. *Ambiophonics: Beyond Surround Sound to Virtual Sonic Reality*. Ambiophonics Institute, New Jersey, 2nd edition, 2009.

[14] Google. The new multi-screen world: Understanding cross-platform consumer behavior, 2012.

[15] G. S. Kendall. A 3-d sound primer: directional hearing and stereo reproduction. *Computer music journal*, 19(4):23–46, 1995.

[16] W. Kitzen and P. Boers. Applications of a digital audio-signal processor in tv sets. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84.*, volume 9, pages 527–530. IEEE, 1984.

[17] A. I. Klayman. Stereo enhancement system, 1988. US Patent 4,748,669.

[18] A. Kraemer. Two speakers are better than 5.1 [surround sound]. *Spectrum, IEEE*, 38(5):70–74, 2001.

[19] Logitech and Wakefield Research. Survey finds our close relationship with our laptops isn't without issues, 2010.

[20] D. D. Lowe and J. W. Lees. Sound imaging process, 1991. US Patent 5,046,097.

[21] R. C. Maher, E. Lindemann, and J. Barish. Old and new techniques for artificial stereophonic image enhancement. In *Audio Engineering Society Convention 101*, 1996.

[22] S. Olive. Evaluation of five commercial stereo enhancement 3d audio software plug-ins. In *Audio Engineering Society Convention 110*, 2001.

[23] F. Rumsey. *Spatial audio*. Focal Press, 2001.

[24] M. R. Schroeder. Listening with two ears. *Music perception*, pages 255–280, 1993.

[25] T. Takeuchi, P. A. Nelson, O. Kirkeby, and H. Hamada. Robustness of the performance of the" stereo dipole" to misalignment of head position. In *Audio Engineering Society Convention 102*, 1997.

[26] C. Tsakostas, A. Floros, and Y. Deliyiannis. Binaural rendering for enhanced 3d audio perception. In *Proceedings of the AudioMostly 2007 2nd Conference on Interaction with Sound*, pages 27–29, 2007.

[27] W. A. Yost. *Fundamentals of Hearing: An Introduction*. Academic Press, 4th edition, 2007.