A Subjective Evaluation of the Minimum Channel Separation for Reproducing Binaural Signals over Loudspeakers*

YESENIA LACOUTURE PARODI,** *AES Associate Member,* **AND PER RUBAK,** *AES Member*
(Yesenia.Lacouture@audiolabs-erlangen.de) (pr@es.aau.dk)

Aalborg University, DK-9220, Aalborg, Denmark

Editor's Note: This *Journal* paper is a fully reviewed submission, which was awarded the Co-Student Technical Papers Award at the 128th Convention of the Audio Engineering Society in London, UK, 2010 May 22–25. The student, Yesenia Lacouture Parodi, presented the paper and was honored at the Convention. The Editor hopes this will encourage future student submissions.

To evaluate the performance of crosstalk cancellation systems, channel separation is often used as a parameter. However, a systematic evaluation of the minimum audible channel separation has not been found in the literature known to the authors. A set of subjective experiments carried out to evaluate the minimum amount of absolute channel separation needed such that the binaural signals with crosstalk are perceived to be equal to the binaural signals reproduced without crosstalk is described. A three-alternative forced-choice discrimination experiment with a simple adaptive algorithm using a weighed up–down method was carried out. The minimum audible channel separation was evaluated for listeners placed at symmetrical and asymmetrical positions with respect to the simulated loudspeakers. Eight different stimuli placed at two different locations were evaluated. Span angles of 12° and 60° were also simulated. Results indicate that in order to avoid lateralization effects, the channel separation should be below –15 dB for most stimuli and around –20 dB for broad-band noise.

0 INTRODUCTION

Binaural technology is based on the assumption that the sound pressures at the ears control the perception of the sound. Thus by employing an appropriate filter it is possible to simulate a virtual environment where the listener perceives an acoustic source located at a position where no physical source exists. Fig. 1 illustrates a binaural recording reproduced through headphones.

To reproduce binaurally recorded signals through loudspeakers it is necessary to invert the acoustic paths from the loudspeakers to the ears. This is not only in order to equalize the loudspeaker response, but also to counteract the crosstalk (see Fig. 2), that is, the signals that should be heard in the left ear are also heard in the right ear and vice versa. This process is known as crosstalk cancellation, and there exist a number of different methods to calculate the optimal inverse filters [1]–[4].

To evaluate the effective performance of a crosstalk cancellation system, channel separation is usually used as a parameter. Channel separation is defined as the magnitude ratio of the cross terms to the direct signal.[1] In other words, it is a measure of how much of the crosstalk is leaked into the desired signal. In addition, to assess the sweet spot size of a binaural reproduction system through loudspeakers a limit for the channel separation should be defined. In [5] and [6] it is suggested to set the maximum acceptable level of crosstalk relative to the desired signal to –12 dB and –10 dB, respectively. Those values were based on personal experiences. Supported by some studies of the psychoacoustic effects of early reflections and small rooms, it is stated in [7] that the maximum acceptable level of crosstalk should be –15 dB. However, no systematic evaluation of the audibility of the crosstalk and its variations with frequency has been found in the literature known to the authors.

This paper presents a set of subjective experiments carried out with the purpose of measuring the minimum

*Presented at the 128th Convention of the Audio Engineering Society, London, UK, 2010 May 22–25 under the title "A Subjective Evaluation of the Minimum Audible Channel Separation in Binaural Reproduction System through Loudspeakers"; revised 2011 March 17, May 9, and May 24.

**Now with the International Audio Laboratories Erlangen (AudioLabs), University Erlangen–Nuremberg & Fraunhofer IIS, 91058 Erlangen, Germany.

[1]Notice that this ratio has usually a negative value.

audible channel separation. We thus define the minimum audible channel separation as the maximum level of crosstalk at which the binaural signals with crosstalk are perceived to be located equal to the binaural signals reproduced without crosstalk.

Two subjective experiments were carried out using headphones. The first experiment simulated the listener as being located symmetrically with respect to the loudspeakers and pointing toward the middle between the two loudspeakers. We refer to this location as the nominal center position. The second experiment simulated the listener placed 50 mm to the left/right away from the nominal center position. Two different loudspeaker span angles were also simulated (12° and 60°).

In total 64 stimuli were assessed, including speech, broad-band noise, and narrow-band noises distributed in one-octave bands. All stimuli were placed at two different virtual locations (±40° and ±90°).

The minimum audible channel separation was evaluated using a three-alternative forced-choice (3AFC) discrimination experiment. A simple adaptive up–down method was implemented, with the rule 1-down, 2-up, which converges to 66.6% of the psychometric function [8].

To simulate the crosstalk, we added to the raw binaural signals the cross terms multiplied by a gain factor with a delay corresponding to the path differences between loudspeakers. With the gain factor we simulated the channel separation and with the delays the loudspeaker span angle as well as the listener location with respect to the loudspeakers.

This paper is organized as follows. In Section 1 we introduce the concept of channel separation and describe the simplified model used to carry out the subjective experiments. In Section 2 we describe the psychometric method used to evaluate the minimum audible channel separation. Sections 3 and 4 present the setup and results of experiments 1 and 2, respectively. A general discussion and conclusions drawn from this study are summarized in Sections 5 and 6.

1 SIMULATION OF CHANNEL SEPARATION

Fig. 2 depicts the acoustic paths from the loudspeakers to the ears with a two-channel crosstalk cancellation system. The functions H_{ij} represent the transfer functions between the ith loudspeaker and the jth ear (input voltage to ear sound pressure). Throughout this paper subscripts i and j take the values of 1 and 2. The signals d_j are the desired binaural signals to be reproduced (see Fig. 1) and signals v_j are the signals reproduced at the ears. The matrix $\boldsymbol{C}$ contains the crosstalk cancellation filters.

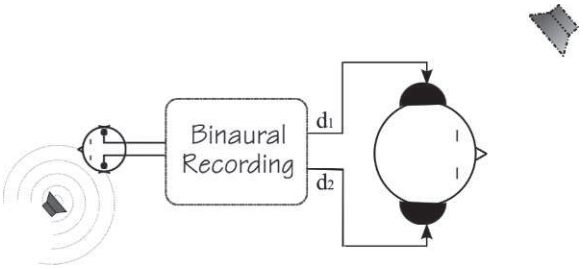

Fig. 1. Schematic of binaural recording with sound source recorded at the ears of a dummy head and reproduced through headphones. d_i—so-called binaural signals.

If perfect crosstalk cancellation is achieved, the desired binaural signals d_j are equal to the signals v_j reproduced at the ears. In other words, $\boldsymbol{HC} = \boldsymbol{I}$, where $\boldsymbol{H}$ is a matrix containing the transfer functions H_{ij} and $\boldsymbol{I}$ is the identity matrix. However, this equation system does not have a direct solution and the filters $\boldsymbol{C}$ are approximations to the required solution. Therefore in a real situation we have that $\boldsymbol{HC} = \boldsymbol{R}$, where $\boldsymbol{R}$ is a 2×2 matrix with the diagonal elements representing the direct signals, and the off-diagonal elements the crosstalk.

The channel separation (CHSP) of the aforementioned system is thus defined as the magnitude ratio of the cross terms to the direct signal [5], [6],

$$\mathrm{CHSP}_i = 20 \log_{10}\left(\frac{R_{ij}}{R_{ii}}\right) \quad [\mathrm{dB}] \tag{1}$$

where R_{ii} and R_{ij}, respectively, are the diagonal and off-diagonal elements of the matrix $\boldsymbol{R}$. Note that this ratio is frequency dependent.

The minimum audible channel separation can be defined as the maximum relative level of crosstalk R_{ij}/R_{ii} at which the signals at the ears v_i are still perceived to be located equal to the desired binaural signals d_i.

To be able to evaluate the minimum audible channel separation, we need to have complete control over the amount of crosstalk that leaks into the desired signals. This can be simulated with headphones by adding the contralateral signals to the ipsilateral signals. Thus in a

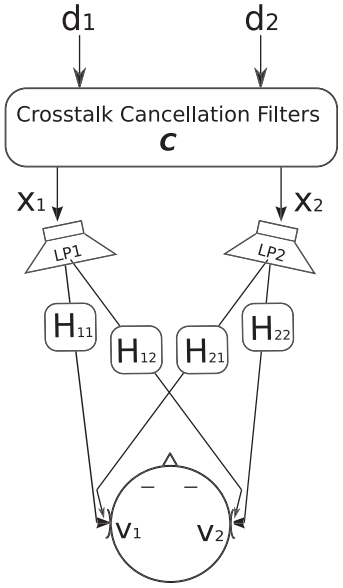

Fig. 2. Acoustic paths from loudspeakers to ears in two-channel crosstalk cancellation system.

simplified fashion, we can model the amount of crosstalk as a gain factor and a delay, as shown in Fig. 3.

Here $G_i = 10^{\mathrm{CHSP}_i/20}$ is the gain factor, in which the CHSP level is given in dB. The delays τ_{ij} correspond to the delays between channels. With this model we are able to simulate a two-channel crosstalk cancellation system in which the accuracy of the crosstalk cancellation filters is varied. Note that this is a rather simplified approach, in which the transfer functions H_{ij} from the loudspeakers to the ears are not taken into account and thus a perfect equalization is assumed. That is, after applying the ideal crosstalk canceler (see Fig. 2), the functions H_{ii} are modeled as frequency-independent delays and the magnitude of the crosstalk H_{ij} is changed in a frequency-independent fashion as well.

In an ideal case, when the listener is looking toward the center point between the two loudspeakers as illustrated in Fig. 2, the channel separation CHSP_i is symmetrical as well as the delays τ_{ij}, that is, $G_1 = G_2$ and $\tau_{12} = \tau_{21}$. When the listener is not located in a symmetrical position there is a channel separation difference between the ears and the delays are not symmetrical either.

With the proposed model we can set the gains G_i and the delays τ_{ij} in such a way that they correspond to the delay and gain factor differences of the desired span angle and listener placement with respect to the virtual loudspeakers. In this manner different span angles and listener positions can be simulated.

Even though the proposed model is a rather simplified approach, we believe it gives a pretty good insight into a listening situation where there is crosstalk. To carry out such a test with loudspeakers, it would be necessary to have complete control over the crosstalk introduced in the signals and the position of the heads of the listeners, as well as individualized HRTFs and filters for each listener. This is clearly impractical, if not infeasible. With our model the real listening situation with loudspeakers is simulated as accurately as possible by using magnitude differences of the channel separation and delays taken from several measurements of different loudspeaker configurations [9]. The delay variations were estimated by means of interaural cross correlations, and a more detailed description can be found in [10]. As mentioned before, the main simplification of our model is thus the assumption that the loudspeakers are ideally equalized.

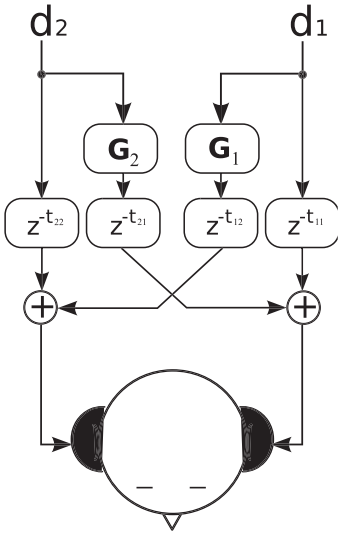

Fig. 3. Simplified model for headphone reproduction of simulated crosstalk corresponding to listening situation illustrated in Fig. 2. Amount of crosstalk is controlled by gain factors G_i; span angle and listener location are controlled with delays τ_{ij}; d_i are desired binaural signals (see Fig. 1).

Models are not meant to be 100% accurate, but to give an approximation of the reality with some controllable variables. The proposed model is thus an approximation of the psychophysical characteristics of the crosstalk in the binaural signals. Further experiments have shown that the results presented in this manuscript are consistent with real listening situations and are thus meaningful. The results of those experiments will be published in the near future.

2 METHODS

2.1 Stimuli and Binaural Synthesis

Eight different signals were assessed, including band-pass noise, one-octave band noise and speech. It is known that in a real situation the channel separation varies with frequency. Thus narrow-band noise distributed in one-octave bands was chosen to evaluate the variations with frequency of the audibility of crosstalk. The speech signal was selected as a reference for familiar and critical sounds. Table 1 summarizes the stimuli used.

Binaural synthesis was used to create the binaural signals, that is, each of the signals presented in Table 1 was convolved with head-related transfer functions

Table 1. Summary of stimuli assessed.*

Stimulus	Frequency Band [Hz]	Duration [s]
Band-pass noise	200–8000	1.2
One-octave band noise	251.1†	1.2
	501.1	1.2
	1000	1.2
	1995.3	1.2
	3981.1	1.2
	7943.3	1.2
Speech (female voice saying the numbers "seven eight")	200–4000	1.4

*All stimuli were convolved with HRTFs corresponding to ±40° and ±90°.

†Center frequencies of band-pass filters according to IEC 1260–1995 and ANSI S1.11–2004.

(HRTF) corresponding to sources located at ±40° and ±90° on the horizontal plane, where the negative angles correspond to the right side of the listener. The sign (±) of the source location was set randomly among stimuli and subjects. The HRTFs were obtained from a database containing artificial-head HRTFs measured with a 2° resolution [11]. These angles were chosen given that sound sources at those locations are externalized easily when reproduced through headphones [12].

Two different loudspeaker span angles were simulated, 12° and 60°. In addition, the experiment was divided into two different scenarios: one with the listener placed at the nominal center position, and the other with the listener displaced 50 mm laterally to the left or right of the nominal center position. The direction of the displacement was set randomly among stimuli and subjects, under the assumption that the results are symmetrical with respect to lateral displacements. To simulate span angles and listener positions, the delays τ_{ij} and the gains G_i were set according to several channel separation measurements carried out in the acoustic laboratories at Aalborg University [9], [10]. Table 2 shows a summary of the differences in channel separation between the ears and the differences in delays between the channels used for each span angle and scenario.

With these two scenarios, two span angles, and two source positions we obtain a total of 64 different stimuli. All the stimuli were set to equal loudness by adjusting their average sound pressure levels numerically to the 60-phon curve defined in ISO 226. Before starting the test the subjects were allowed to adjust the reproduction volume of the headphones through the graphical user interface if they found it necessary, and during the main test all stimuli were reproduced at the same level they had chosen. For those subjects who found a volume adjustment necessary, the observed deviations from the original levels were smaller than 2 dB.

2.2 Subjects

Thirty-two subjects with normal hearing participated in these experiments (12 females and 20 males aged between 22 and 37 years). Most subjects had some experience with listening test and discrimination procedures.

Assuming that the thresholds are normally distributed, to ensure that the 95% confidence interval is not larger than ±3 dB around the means,[2] the minimum sample size should be 7 (see [13, Eq. (3)]). The subjects were divided randomly into four groups, and each group evaluated different stimuli, that is, 16 stimuli per group, which results in eight thresholds per stimulus. The order of presentation of the stimuli followed a Latin square design in order to account for carryover effects [14].

Table 2. Summary of channel separation (CHSP) differences between ears, delays applied to crosstalk channels, and delay differences between direct signals used in the experiments to simulate each span angle and scenario.

	Center Position		50 mm Displacement	
Span Angle	12°	60°	12°	60°
$\lvert CHSP_1 - CHSP_2 \rvert$ [dB]	0	0	1.99	2.07
τ_{12} [μs]	33	166	43.8	174.8
τ_{21} [μs]	33	166	22.7	156.9
$\lvert \tau_{11} - \tau_{12} \rvert$ [μs]	0	0	8	44.3

2.3 Psychophysical Method

Thresholds were determined using an adaptive three-alternative forced-choice (3AFC) procedure. This method was chosen given that it is not too demanding for the subjects and it is rather easy to understand, thus reducing potential errors due to misunderstandings or tiredness. In each trial three test stimuli were presented with 0.1-second intervals between the stimuli. The reference stimulus was the raw binaural signal without crosstalk. All possible combinations of reference stimuli and signals with crosstalk were reproduced randomly. Each stimulus had a duration of between 1.2 and 1.4 seconds.

For each trial the order of the stimulus presentation was randomized and the subjects were asked to discriminate which of the three signals was the different one. It was observed in pilot experiments that the main audible difference between stimuli was the location. Subjects were trained and instructed to focus on the placement of the sound before starting the test.

To measure the thresholds we used the simple adaptive testing algorithm with the weighted up–down method proposed by Kaernbach [8]. The advantage of this method is that it converges to any desired point of the psychometric function and is rather simple to implement. Note that with this method it is not possible to draw a complete psychometric function because most of the observations are placed very close to the target level [15].

For the *n*AFC methods the threshold is often defined as the signal level at which the probability of correct responses is halfway between perfect performance (100% correct answers) and probability of guessing (33.3% correct answers in the case of 3AFC) [16]. With the rule 1-down, 2-up Kaernbach's method converges to 66.6% of the psychometric function, which will correspond to the threshold of the 3AFC. This means that for each correct answer the channel separation level goes down one step and for each incorrect answer the channel separation level goes up two steps.

There exist a number of ways to estimate the threshold from an adaptive up–down method. In [15] it is suggested that the mid-run estimates are rather robust, relatively efficient, and result in low bias. A run is defined as a series of steps in one direction. A min-run estimate is

[2]This is assuming a standard deviation of 4 dB observed in a pilot experiment.

calculated by taking the mean between the lower and upper turnarounds. However, we consider the medians to be more robust than the means with respect to luck and lack of attention. Thus we defined the threshold as the mean between the medians of the lower and upper turnarounds. Fig. 4 shows an example of typical data obtained from an adaptive staircase algorithm and the calculated threshold using the proposed estimate.

In order to ensure a faster convergence to the target level, we reduced the step size along the test [15]. The initial step size was set to 3 dB. After the third run it was reduced to 2 dB, and after the sixth run it was reduced to 1 dB. The algorithm was stopped after 12 runs.

2.4 Procedures and Experimental Design

Part of these experiments were conducted at the Sound and Music Innovation Technology (SMIT) laboratories at the National Chiao-Tung University in Taiwan and the rest was conducted at the acoustic laboratories at Aalborg University. For the part conducted at the SMIT laboratories we used a pair of dynamic Sennheiser HD595 headphones, and for the part conducted at Aalborg University we used a pair of open-cup Beyerdynamic DT990 headphones.

The impulse responses of both headphone pairs were measured with a dummy head. Several measurements of the headphones were carried out in order to account for fitting errors. Using the average impulse response of each headphone, 32-tap IIR inverse filters were calculated by means of the modified Yule–Walker approximation [17]. Through further measurements it was ensured that the equalization error did not exceed ± 1 dB in the frequency range between 200 and 8000 kHz. It was also ensured that the equalized impulse response did not present group delays, peaks, and bandwidths that could potentially introduce audible distortion into the stimuli [18], [19].

The experiments were carried out in sound-isolated listening rooms. Subjects had access to a graphical user interface in which for each sequence, as the stimuli were reproduced, the screen displayed buttons with the labels 1, 2, and 3, corresponding to each sound played. Subjects were instructed to click on the button associated with the sound they perceived to be different.

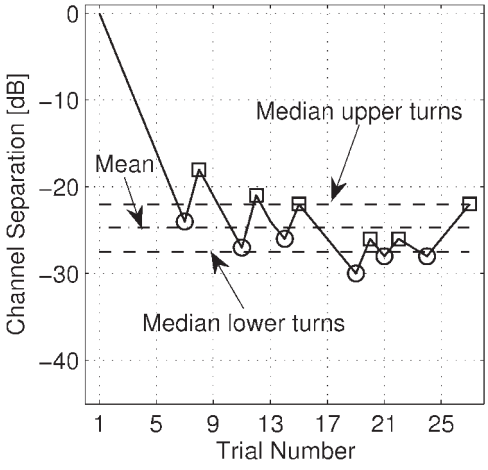

Fig. 4. Example of typical data obtained with adaptive staircase algorithm. □ upper turns; ○ lower turns. Rule applied is 1-down, 2-up.

At the beginning of the experiment, the subjects had a short training session in which they were introduced to the procedures of the test. During this training session the differences were made clearly audible (such as, $G_i = 1$), and feedback was provided to the subjects. This was done to familiarize the subjects with the differences and to improve concentration. After the training session the feedback was removed and they had the opportunity to repeat each sequence once to account for lack of attention. Subjects were encouraged to repeat the sequence only when they had heard a difference but had forgotten which one of the three was the different one. They were also instructed to focus on the position of the source.

The adaptive staircase algorithm, the graphical user interface, the playback, and all data collection were implemented in MATLAB.

The experiments were divided into four sessions per day. Each session consisted of two blocks each corresponding to different stimuli. There was a short break of about 2 minutes between blocks and a longer break of about 10 minutes between sessions. After the test, the subjects were asked to describe how the positions of the stimuli differed when they could hear a difference.

3 EXPERIMENT 1—LISTENER AT THE NOMINAL CENTER POSITION

In this experiment the minimum audible channel separation was measured for a listener placed at the center position. This means that we assumed $G_1 = G_2$, the delays lays $\tau_{ii} = 0$, and the delay $\tau_{12} = \tau_{21}$. Here the delays τ_{ij} were set to simulate span angles of 12° and 60°. These values are estimates based on channel separation measurements carried out at the acoustic laboratories at Aalborg University [9], [10] (see Table 2).

3.1 Results

In order to analyze the effects of all the possible factors (stimuli, span angle, and source position) and their interaction, we carried out a multivariate analysis of variance (MANOVA) of the channel separation thresholds.[3] When the location of the listener was simulated at the nominal center position, the span angles did not show significant effects in the thresholds. On the other hand, there was a significant effect of the source location in the CHSP thresholds ($F_{(1251)} = 25.43$, $p < 0.001$). The effect of the stimuli was also found to be significant at the 0.001 level ($F_{(7245)} = 18.06$, $p < 0.001$).

Fig. 5 shows the mean thresholds obtained for each

[3]To check the validity of the MANOVA, the error distributions were analyzed and found to be approximately normally distributed.

stimulus at the two different simulated source positions: 40° and 90°. Since no significant differences were observed with respect to the span angles, the data are sorted by source location. The mean thresholds for each source position are offset horizontally for visual clarity. The error bars correspond to the 95% confidence intervals.

We can observe a very clear trend in Fig. 5: the mean thresholds for the speech (Sp), the band-pass noise (Bp) and the narrow-band noise centered at 251 Hz (Nn_{251}) and 501 Hz (Nn_{501}) are rather homogeneous and lie close to –20 dB. In contrast, the thresholds obtained with the narrow-band noise centered at 1 kHz (Nn_{1000}) and 2 kHz (Nn_{1995}) are larger than for the rest of the stimuli and lie around –15 dB. At the higher frequency bands (Nn_{3981} and Nn_{7943}) the mean thresholds go down to approximately –25 dB when the sound source is at 90°.

To support the observed tendency, we carried a pairwise comparison between stimuli. We found the mean thresholds for the narrow-band noise centered at 1 and 2 kHz to be significantly different from the thresholds observed with the other stimuli ($p < 0.001$). The narrowband noise centered at 8 kHz is also significantly different from the other narrow-band noise and the speech signal at the 0.05 level at most (maximum p value observed $p = 0.021$). The latter observation can be a consequence of the strong dependence of the interaural level differences (ILD) on frequency [20]. Above 1.6 kHz the ILD increases systematically with frequency. When adding crosstalk to the binaural signals, we are directly affecting their natural ILD. This change could be more audible at higher frequencies where a larger ILD is expected, and this could be the reason why lower thresholds are observed with the narrow-band noise centered at 4 and 8 kHz.

Regarding the differences observed in the middle frequencies (Nn_{1000} and Nn_{1995}), one assumption for the observed pattern is that in that frequency band our accuracy to discriminate angle differences decreases substantially [21]. This is due to the different mechanisms used by humans to localize sound sources. In the low-frequency region our auditory system makes use of phase and time differences to localize sounds, whereas in the high-frequency region it uses mainly the level differences. In [21] it is suggested that in the middle-frequency region neither the phase nor the level differences are effective enough for localization. This could explain why the minimum audible channel separation is considerably larger in this region.

There is also a clear difference between the mean thresholds obtained with the narrow-band noise centered at 4 kHz. The threshold obtained with the stimulus placed at 90° is significantly smaller (–24 dB) than the threshold obtained with the stimulus placed at 40° (–15.3 dB). It is well known that above 2 kHz the ILD is remarkably larger for sound sources placed at 90° than for sound sources placed at 40° because of the head shadowing effect. Thus when adding crosstalk to the narrow-band noise centered at 4 kHz the ILD for the source placed at 90° is reduced significantly. We can also observe—to a lesser extent—a similar pattern with the narrow-band noise centered at 8 kHz. However, we can notice that this effect does not occur so dramatically with the band-pass noise and the speech signal which also contain those frequency bands. This suggests, that the changes in ILD at low frequencies—which do not vary much with the source location—and the additional low-frequency cues present in the signals, somehow mask the changes in the ILD at higher frequencies.

Furthermore there is a general tendency of larger thresholds when the source is placed at 40° than when it is

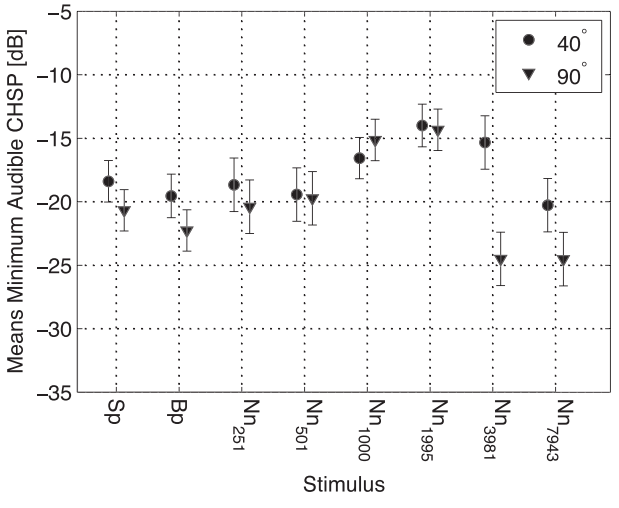

Fig. 5. Mean minimum audible channel separation (CHSP) as a function of stimulus and source location. Crosstalk is simulated for subject placed at nominal center position. Thresholds for stimuli placed at 40° (●) and 90° (▼) are offset horizontally for visual clarity. Error bars indicate 95% confidence interval.

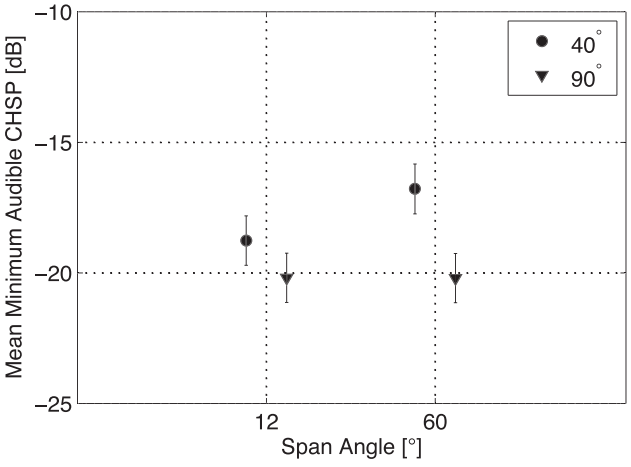

Fig. 6. Total average of minimum audible CHSP level as a function of span angle and source position. Mean thresholds for each source position are offset horizontally for visual clarity. Error bars indicate 95% confidence interval.

at 90°. This tendency is clearly shown in Fig. 6, where the average of the thresholds is plotted as a function of span angle and source position. Yet we can notice that the differences are more pronounced with the 60° span angle. The mean difference between source locations with a 60° span angle was found to be statistically significant at the 0.05 level ($F_{(1251)} = 4.3$, $p = 0.039$), whereas no significant difference was found with the 12° span angle.

4 EXPERIMENT 2—LISTENER AT A LATERALLY DISPLACED POSITION

In this experiment the minimum channel separation was measured for crosstalk corresponding to a listener displaced laterally from the nominal center position. The delays and gain differences between channels were simulated for head positions 50 mm to the left/right of the center position. The gain differences between G_1 and G_2 and the delays τ_{ij} were modeled based on measurements of the channel separation of different loudspeaker arrangements [9], [10]. These values were set such that they corresponded to the gain differences and delays obtained with the 12° and 60° span angle configurations placed on the horizontal plane (see Table 2).

4.1 Results

In this scenario we observed a particular pattern in the CHSP thresholds: some subjects consistently showed thresholds below –40 dB for some of the stimuli when the simulated span angle was 60°. The stimuli that showed this pattern more markedly were the band-pass noise and the narrow-band noise with the center frequency at 501 Hz (Nn_{501}). This lead us to believe that with such a large "signal-to-noise ratio" (that is, channel separation), what these subjects were discriminating was not the channel separation but the delays added to the signal with crosstalk. Based on this argument and in order to reduce the standard deviations of the data, we decided to divide the results for this scenario into two groups. The first group, which we will arbitrarily refer to as "delay-sensitive listeners," contains the thresholds obtained below –35 dB. The second group, which we will refer to as "normal listeners," contains the thresholds obtained above –35 dB. Note that this segregation is of thresholds and not of subjects as such. That is, subjects who were able to discriminate crosstalk levels below –35 dB with some of the stimuli, also showed "normal" thresholds with the other stimuli they evaluated and are thus included in the normal listeners group. However, we would like to stress that the subjects who were able to discriminate crosstalk levels below –35 dB did that consistently with most of the stimuli that simulated a 60° span angle. For this reason we decided to name the groups as mentioned.

Fig. 7 shows the CHSP thresholds obtained with the delay-sensitive listeners. We cannot observe any dependence on the source position. Yet, as mentioned before, we can see that the stimuli showing the largest groups of delay-sensitive listeners are the narrow-band noise centered at 501 Hz and the band-pass noise, with five subjects each. Note that no thresholds below –35 dB were observed with the narrow-band noise centered at 2 and 8 kHz, and thus they are not plotted in the figure.

It is well known that at frequencies below 1.6 kHz the interaural time difference (ITD) is the main mechanism that the human auditory system uses to localize a sound source. Since the delay differences between channels were kept constant during this experiment for the stimuli with crosstalk, this supports the hypothesis that this particular group of subjects were discriminating the ITD differences between the signal with crosstalk and the signal without crosstalk and not the channel separation differences.

In [6] the sweet spot of different loudspeaker configurations was evaluated as a function of changes in ITD. The results presented in that study show that the 60° span angle configuration is not robust to lateral displacements when looking at the temporal changes of the binaural signals. That analysis was done assuming a minimum audible ITD of 10 μs. However, in [22] a large variance was observed between subjects, when the just noticeable differences in ITD were evaluated. This could explain why some subjects were able to hear the delay differences whereas others were not.

Another possible explanation of this phenomenon is that the delay-sensitive listeners could have been able to discriminate the stimuli with crosstalk because of coloration introduced by a comb-filter effect. When superimposing the direct signal and a delayed version of itself we are creating what is known as comb-filter effect. Coloration due to comb-filter distortion can be audible if the delay is below 30 ms, depending on the signal content and the level differences [23]. In [24] it is shown that the audibility of coloration tends to decrease for delays below

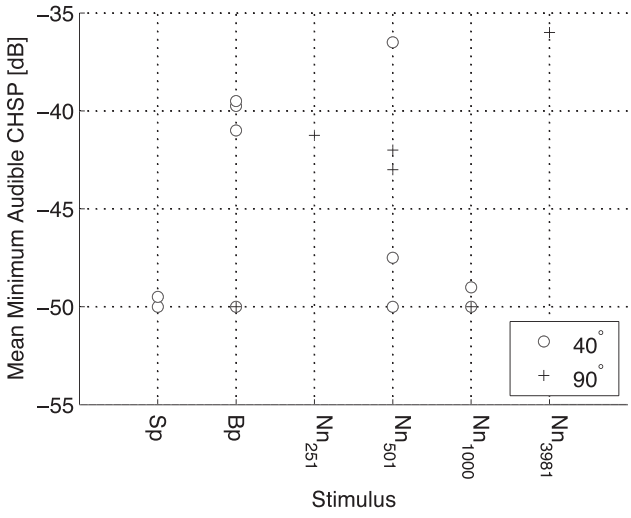

Fig. 7. Scatter plot of CHSP thresholds obtained with delay-sensitive listeners as a function of different stimuli and source locations. Thresholds correspond to simulated 60° span angle. Thresholds for stimuli placed at 40° (○) and 90° (+) are plotted.

2 ms. As shown in Table 2, the delays introduced in the experiment are on the order of microseconds, which suggests that coloration was likely below the audibility threshold, and that it was rather a slight lateralization effect that was perceived. In addition, when interviewed none of the subjects reported to have perceived coloration or pitch differences between the stimuli.

A MANOVA of the CHSP thresholds obtained with the normal listeners was carried out. When excluding the delay-sensitive listeners from the results, no significant differences were observed between span angles. Similarly to the results obtained in experiment 1, the effect of the source position on the CHSP thresholds was found to be significant ($F_{(7224)} = 19.13$, $p < 0.001$) as well as the effect of the stimuli ($F_{(1230)} = 14.52$; $p < 0.001$). The interaction between source position and stimuli was shown to be also significant at the 0.05 level ($F_{(7224)} = 2.34$; $p = 0.025$).

Fig. 8 shows the mean CHSP thresholds obtained with the normal listeners for the different stimuli placed at 40° and 90°. The mean thresholds obtained with these two source locations are offset horizontally for visual clarity.

In contrast to the pattern observed in experiment 1, when carrying out a pairwise comparison between stimuli, the narrow-band noise with center frequencies below 4 kHz showed no significant differences. Yet the band-pass noise was found to be significantly different from these stimuli and the speech signal ($p < 0.001$). The band-pass noise with crosstalk contains not only conflicting ILDs but also conflicting ITDs. In addition, in this experiment the channel separation had different levels for each ear, which corresponds to a listener displaced to the left or right from the nominal center position (see Table 2). These differences in channel separation increase also the changes in ILD. This might result in a large energence of conflicting cues (ILDs and ITD) for the band-pass noise, making discrimination of crosstalk easier for these stimuli than for the narrow-band stimuli at low frequencies and the speech signal.

Similar to the results obtained with experiment 1, the mean thresholds for the narrow-band noise centered at 4 and 8 kHz lie below –20 dB and are also significantly lower than the thresholds obtained with the narrow-band noise centered at lower frequencies.

There is also a clear difference between the mean CHSP thresholds obtained when the source was placed at 40° and when the source was placed at 90°. Similar to the trend observed with the CHSP thresholds obtained with the listener placed at the nominal center position, this difference is more pronounced with high-frequency narrow-band noise (4 and 8 kHz). In this case, however, the location of the sound source shows also a significant difference compared to the speech signal.

Fig. 9 shows the average CHSP thresholds as a function of span angle and source position. Then again, the mean differences between source positions are statistically significant for the 60° span angle.

5 GENERAL REMARKS

During the experiments the subjects were asked to give their impressions about the stimuli. They were specifically asked what they thought the difference between the locations of the stimuli was when they could hear a difference. Most subjects described the difference as one sound being placed on one of their sides whereas the other sound was either inside the head or above. Some subjects described the differences as a change in distance.

By adding crosstalk to the direct signal and delays between channels, we are indirectly varying the interaural differences of the binaural signals. As described before, the delay differences between channels were kept constant in both experiments and only the level of

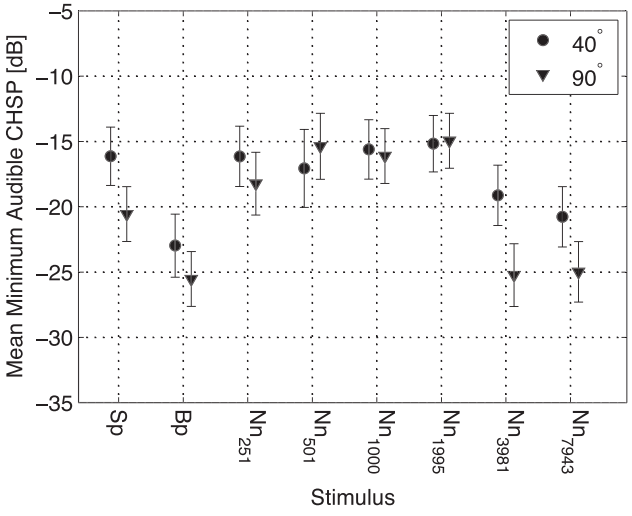

Fig. 8. Means of minimum audible CHSP obtained with normal listeners as a function of different stimuli and source locations. Thresholds for stimuli placed at 40° (●) and 90° (▼) are offset horizontally for visual clarity. Error bars indicate 95% confidence interval.

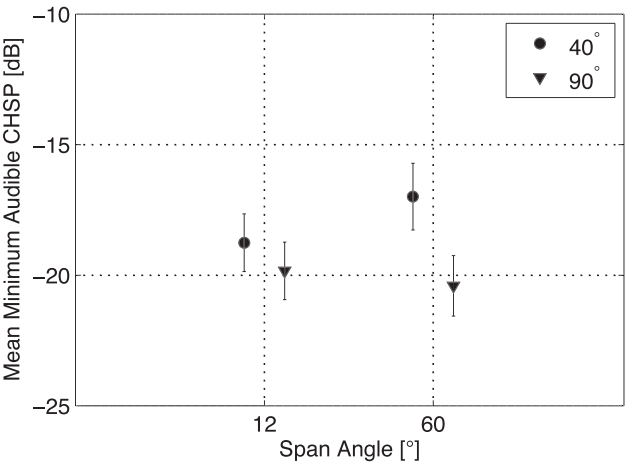

Fig. 9. Total average of minimum audible CHSP obtained with normal listeners as a function of span angle and source position. Mean thresholds for each source position are offset horizontally for visual clarity. Error bars indicate 95% confidence interval.

crosstalk was varied. We could thus argue that we were mainly varying the ILD of the binaural signals. There are of course changes in the interaural phase differences (IPDs), given that the phase does not only depend on the delay differences between the signals but also on their relative amplitude, such as, the level of crosstalk. However, for the stimuli used during the experiments those variations are negligible compared to the frequency-independent delays introduced between the channels (see Table 2). Since the delay difference between channels was kept constant in both experiments, this lead us to believe that the dominant discrimination cue was mainly changes in ILD. When simulating the 12° span angle, the delay differences between the channels were not large enough to be audible, but in the case of the 60° span angle and displaced listening position, some subjects were able to discriminate the phase differences between the binaural signals without crosstalk and the binaural signals with crosstalk.

Some experiments have shown that when changing the ILD or ITD of binaural signals, a lateralization effect usually occurs, which causes the virtual image to move to the center of the head, especially when changes in ILD do not correlate with changes in ITD [20]. In the case of the delay-sensitive listener, when the ILD difference was no longer audible, this lateralization effect was still present due to the audible differences in ITD.

6 DISCUSSION AND CONCLUSIONS

A subjective evaluation of the minimum audible channel separation for binaural reproduction systems through loudspeakers is presented. Using a simplified model, the crosstalk was simulated by adding to the direct signals the cross terms attenuated by a gain factor and a delay. Two listening experiments were carried out through headphones, in which the minimum audible channel separation was evaluated for listeners placed at the nominal center position and listeners displaced laterally 50 mm to the left/right of the center position. The minimum audible channel separation was measured using a 3AFC discrimination procedure and the simple adaptive algorithm with a weighted up–down method [8]. Even though the model presented in this paper is a simplification of the listening situation when using loudspeakers, it tries to approximate that situation as accurately as possible, and it gives a good insight into the psychophysical characteristics of binaural signals when crosstalk is introduced.

Results from the first experiment showed that the changes in ILD caused by the crosstalk are more audible at higher frequencies than at lower frequencies. In the middle frequencies, on the other hand, discrimination of the crosstalk becomes more difficult due to the decrease in accuracy of the human localization mechanism [21]. We also believe that when a broad-band stimulus such as the band-pass noise is used, the differences in ILD at high frequencies are somehow masked by the binaural cues present at lower frequencies, making discrimination of crosstalk slightly more difficult with broadband noise when the listener is placed at the nominal center position.

In the second experiment, when simulating a 60° span angle, some subjects were able to discriminate the ITD differences between binaural signals without crosstalk and binaural signals with crosstalk. This is in agreement with results presented in [10], where it is shown that the 60° span angle is not robust to lateral displacements when evaluating the ITD changes. On the other hand, audibility of coloration due to combfilter distortions can also be a possible explanation of the thresholds obtained with the delay-sensitive listeners. Even though none of the subjects reported to have perceived coloration or pitch differences during the test, we consider that this possibility should not be completely disregarded.

Disregarding the group of subjects that could discriminate the differences in ITD, the results obtained with the listener displaced laterally from the nominal center position follow similar trends as the results obtained with the listeners placed at the nominal center position. However, in this case the band-pass noise showed lower thresholds compared to the thresholds obtained in the first experiment.

In both experiments the sound sources placed at 90° showed a general tendency toward smaller thresholds than the sound sources placed at 40°. These differences are more pronounced at high frequencies and with the 60° span angle. The ILD at 90° is significantly larger than the ILD at 40° because of the natural head-shadowing effect. This is especially observed at high frequencies. Thus it is hypothesized that the crosstalk in a sound source placed at 90° is easier to discriminate than when the source is at 40° because of the significant changes in ILD at high frequencies. In addition, with the 60° span angle the delay differences between the channels is larger than with the 12° span angle, which could act as an additional cue used to discriminate the signal with crosstalk.

In summary, we could observe that the minimum audible channel separation lies below −15 dB for most stimuli evaluated. Furthermore, in the case of broadband signals such as the band-pass noise employed in the experiment, the minimum audible channel separation lies around −20 dB when the listener is at the nominal center position and −25 dB when the head is laterally displaced. Previously it has been suggested to set the maximum channel separation to −12 or −10 dB [5], [6]. However, the results obtained from this study suggest that those limits could be rather relaxed and the −15-dB limit suggested in [7] is more suitable. In most applications of binaural reproduction systems the reproduced signals are either speech or broad-band signals in general. Therefore according to the results presented in this study, the channel separation limits should be set around −20 dB instead.

Most subjects described the stimuli with crosstalk as being closer to, above, or inside their heads, which is in agreement with lateralization experiments. It is not

expected that the virtual images will fall inside the head when reproduced through a crosstalk cancellation system with insufficient channel separation, but rather that they will be placed at the location of the loudspeakers. Such an effect can have unfortunate consequences in binaural reproduction systems through loudspeakers. If, for example, a virtual environment is simulated and some of the images happen to be wrongly placed at the loudspeaker position because of insufficient channel separation at some key frequencies, the whole virtual experience can be degraded to a large extent. Hence if a proper virtual reproduction is desired, care should be taken when designing the crosstalk cancellation filters and sufficient channel separation should be allowed in the target frequency band.

7 ACKNOWLEDGMENT

Part of this work was carried out at the Sound and Music Innovation Technology (SMIT) laboratories at the National Chiao-Tung University, Taiwan. The authors would like to thank for the valuable input received from Mingsian Bai and the help with the experiments provided by his students. The authors would also like to thank the two anonymous reviewers for their valuable comments and enriching discussions.

8 REFERENCES

[1] M. R. Bai and C. C. Lee, "Development and Implementation of Cross-Talk Cancellation System in Spatial Audio Reproduction Based on Subband Filtering," *J. Sound Vibr*., vol. 290, pp. 1269–1289 (2005 Aug.).

[2] O. Kirkeby and P. A. Nelson, "Digital Filter Design for Inversion Problems in Sound Reproduction," *J. Audio Eng. Soc*., vol. 47, pp. 583–595 (1999 July/Aug.).

[3] P. A. Nelson, F. Orduña-Bustamante, and H. Hamada, "Inverse Filter Design and Equalization Zones in Multichannel Sound Reproduction," *IEEE Trans. Speech Audio Process*., vol. 3, pp. 185–192 (1995 May).

[4] D. B. Ward, "Joint Least Squares Optimization for Robust Acoustic Crosstalk Cancellation," *IEEE Trans. Speech Audio Process*., vol. 8, pp. 211–215 (2000 Feb.).

[5] M. R. Bai and C. C. Lee, "Objective and Subjective Analysis of Effects of Listening Angle on Crosstalk Cancellation in Spatial Sound Reproduction," *J. Acoust. Soc. Am*., vol. 120, pp. 1976–1989 (2006 Oct.).

[6] J. Rose, P. Nelson, B. Rafaely, and T. Takeuchi, "Sweet Spot Size of Virtual Acoustic Imaging Systems at Asymmetric Listener Locations," *J. Acoust. Soc. Am*., vol. 112, pp. 1992–2002 (2002 Nov.).

[7] A. Mouchtaris, P. Reveliotis, and C. Kyriakakis, "Inverse Filter Design for Immersive Audio Rendering over Loudspeakers," *IEEE Trans. Multimedia,* vol. 2, no. 2, pp. 77–87 (2000).

[8] C. Kaernbach, "Simple Adaptive Testing with the Weighted Up–Down Method," *Perception & Psychophys*., vol. 49, pp. 227–229 (1991).

[9] Y. Lacouture Parodi and P. Rubak, "Objective Evaluation of the Sweet Spot Size in Spatial Sound Reproduction Using Elevated Loudspeakers," *J. Acoust. Soc. Am*., vol. 128, pp. 1045–1055 (2010 Sept.).

[10] Y. Lacouture Parodi and P. Rubak, "Sweet Spot Size in Virtual Sound Reproduction: A Temporal Analysis," in *Principles and Applications of Spatial Hearing* (World Scientific, Singapore, 2011).

[11] B. P. Bovbjerg, F. Christensen, P. Minnaar, and X. Chen, "Measuring the Head-Related Transfer Functions of an Artificial Head with a High-Directional Resolution," presented at the 109th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 48, p. 1115 (2000 Nov.), preprint 5264.

[12] S. M. Kim and W. Choi, "On the Externalization of Virtual Sound Images in Headphone Reproduction: A Wiener Filter Approach," *J. Acoust. Soc. Am*., vol. 117, pp. 3657–3665 (2005 Jun.).

[13] J. Eng, "Sample Size Estimation: How Many Individuals Should Be Studied?," *Radiology*, vol. 227, pp. 309–313 (2003).

[14] G. E. P. Box, J. S. Hunter, and W. G. Hunter, *Statistics for Experimenters: Design, Discovery, and Innovation,* 2nd ed. (Wiley-Interscience, New York, 2005).

[15] H. Levitt, "Transformed Up–Down Methods in Psychoacoustics," *J. Acoust. Soc. Am*., vol. 49, pp. 467–477 (1971).

[16] C. Kaernbach, "Adaptive Threshold Estimation with Unforced-Choice Tasks," *Perception & Psychophys*., vol. 63, pp. 1377–1388 (2001).

[17] B. Friedlander and B. Porat, "The Modified Yule-Walker Method of Arma Spectral Estimation," *IEEE Trans. Aerosp. Electron. Sys.,* vol. AES-20 pp. 158–173 (1984).

[18] H Banno, K Takeda, and F Itakura, "The Effect of Group Delay Spectrum on Timbre," *Acoust. Sci. Technol*., vol. 23, pp. 1–9 (2002).

[19] H. Møller, P. Minnaar, S. K. Olesen, F. Christensen, and J. Plogsties, "On the Audibility of All-Pass Phase in Electroacoustical Transfer Functions," *Audio Eng. Soc*., vol. 55, pp. 115–133 (2007 Mar.).

[20] J. Blauert, *Spatial Hearing,* 3rd ed. (Hirzel, Stuttgart, Germany, 2001).

[21] A. W. Mills, "On The Minimum Audible Angle," *J. Acoust. Soc. Am*., vol. 30, pp. 237–246 (1958).

[22] P. F. Hoffmann and H. Møller, "Audibility of Differences in Adjacent Head-Related Transfer Functions," *Acta Acustica/Acustica*, vol. 94, pp. 945–954 (2008).

[23] H. Haas, "Influence of a Single Echo on the Audibility of Speech," *J. Audio Eng. Soc*., vol. 20, pp. 146–159 (1972 Mar.).

[24] A. M. Salomons, "Coloration and Binaural Decoloration of Sound Due to Reflections," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands (1995).

THE AUTHORS

Y. Lacouture Parodi

P. Rubak

Yesenia Lacouture Parodi was born in Colombia in 1980 and studied electronic engineering at the Pontificia Universidad Javeriana in Bogota, Colombia. She received a master's degree in acoustics in 2007 and a Ph.D. degree in 2010, both from Aalborg University, Aalborg, Denmark. Her doctoral work comprises a systematic study of binaural reproduction systems through loudspeakers, with special focus on stereo dipoles. She was visiting researcher at the laboratory for Sound and Music Innovation Technology (SMIT) at the National Chiao-Tung University, Hsin-Chu, Taiwan, from 2009 August to December. She has recently joined the International Audio Laboratories Erlangen (AudioLabs) in Germany, a joint institution of Fraunhofer IIS and University of Erlangen–Nurnberg. Her research interests include binaural techniques, perception of spatial sound, audio signal processing and immersive environments.

Dr. Lacouture Parodi is a member of the Audio Engineering Society and IEEE.

●

Per Rubak obtained an M.Sc. degree in electronic engineering from the Technical University of Denmark in 1968.

From 1968 to 1971 he was with Brüel & Kjær, Copenhagen, where his main work was the development of calibration equipment for microphones. At present he is an associate professor at the Institute of Electronic Systems, Aalborg University, Denmark. He was manager for PANACOUSTIC A/S during 1991 and 1992, where he was responsible for the development of an active hearing protector. From 2001 to 2002 he was a professor at the University College of Aarhus, Denmark, and in 2002 he returned to Aalborg University. Since 1987 the Danish National Testing Board has drawn on his expertise in the field of electroacoustic measurements on telecom equipment. During the 1980s he developed a new and innovative telephone handset (low acoustic impedance, providing a better low-frequency response), in cooperation with B&O and Jutlands Telephone Company. His main interests are room acoustics, digital room equalization, digital reverberators, perceptual effects of sound fields, electroacoustics, 3D-sound systems, active control of sound, and signal processing of audio signals.

Mr. Rubak was President for the Danish Acoustical Society from 1991 to 1999. He has published 36 papers, approximately half within the framework of the Audio Engineering Society.