Bit-rate saving in multichannel sound

Bit-Rate Saving in Multichannel Sound: Using a Band-Limited Channel to Transmit the Center Signal*

S. L. J. D. E. VAN DE PAR** AND W. R. TH. TEN KATE

Philips Research Laboratories, 5600 JA Eindhoven, The Netherlands

AND

A. KOHLRAUSCH AND A. J. M. HOUTSMA

Institute for Perception Research, 5600 MB Eindhoven, The Netherlands

A method is proposed to achieve full-frequency-range three-channel (left, right, and center) sound reproduction in systems that have only two full-range sound channels and some band-limited commentary channels. The low-frequency part of the center signal, which matches the bandwidth of the commentary channels, is added to the (multilingual) speech signals in each of the commentary channels. The remaining high-frequency part is added in the left and right channels as in conventional mixdowns. Sound reproduction of this signal by a conventional two-channel receiver remains unaltered. The low-frequency part of the center signal is mixed to the left and right signals together with the speech once the user has selected a commentary channel. Three-channel reproduction is obtained by routing the selected commentary channel to a central loudspeaker. Listening tests revealed that sound reproduction according to the proposed scheme could not be distinguished from original three-channel reproduction. This scheme can be applied to proposed standards such as D2MAC and MPEG2.

0 INTRODUCTION

Besides high-definition pictures, future high-definition television (HDTV) will also feature multichannel sound [1]. Current proposals for sound transmission typically allow full-range audio channels (32–48-kHz sampling rate) as well as channels of limited bandwidth (16–24-kHz sampling rate). The latter are thought to be used for commentaries in multilingual broadcasting. Combinations of these are allowed as long as the total bit rate is below a certain limit. A typical configuration for loudspeakers would be three frontal loudspeakers (left, center, right) and two surround loudspeakers [1]. This study addresses only aspects of reproduction through the three frontal loudspeakers.

It is known that the addition of a third center loudspeaker to conventional stereo is very beneficial for sound reproduction in combination with pictures [1], [2]. The center loudspeaker adds considerably to the directional stability of sounds which have to be heard in front of the listener. To obtain the advantage of directional stability, a three-channel mix is required as the source signal.

In this paper a scheme is proposed which provides an appropriate signal for the three frontal loudspeakers by using two full-bandwidth (left, right) channels and one limited-bandwidth (commentary) channel. The basic idea is to add the center signal to each of the commentary signals before transmission. At the receiver end, one of the commentary channels is selected. By routing the selected commentary channel to a center loudspeaker, not only the preferred language, but also the center signal will be heard on this center loudspeaker. In this way three-channel sound reproduction is obtained.

If conventional two-channel reproduction is desired, one of the commentary channels is selected by the user and is routed to the left and right loudspeakers. In this

* Manuscript received 1993 May 4; revised 1994 March 2.
** Now with the Institute for Perception Research.

way the center signal is mixed to the left and right loudspeakers, in addition to the selected commentary signal, leaving two-channel reproduction unaffected.

Because commentary channels are typically only half bandwidth, only the lower half of the frequency content of the center signal can be transmitted through a commentary channel. If no special measures are taken, the high frequencies of the center signal are lost. In this paper we investigated whether this is acceptable, or whether it is necessary to transmit the center-signal high frequencies by some other means, for example, by mixing them to the left and right channels at the broadcasting end. In the latter case further mixing can be applied at the receiver end.

It is important that the localization properties of the original three-channel sound be not affected by the proposed scheme because localization stability was the main reason for introducing an extra center loudspeaker. It is expected that localization is not affected by redistribution of the high-frequency part of the center signal. The reason for this is that, in stereophonic listening, the main localization cues are in the frequencies between 100 Hz and 8 kHz [3]. It was expected, however, that the high-frequency part of the center signal cannot be left out of the sound field, as timbre changes could otherwise become audible.

The aim of the present paper is to evaluate the perceptual quality of several possible transmission schemes. All these schemes result in a three-channel sound reproduction by using two full- and one half-bandwidth channels for transmission. In the next section several transmission systems are described in detail. In subsequent sections the listening tests and their results are described.

1 SYSTEM DESCRIPTION

In this section a detailed technical description of several possible transmission schemes is given. As mentioned, a three-channel full-range mix must be available at the broadcasting end. The three source signals of this mix are the left, right, and center signals, which will be denoted by L, R, and C, respectively. In addition to these signals some commentary signals are available. They are C_i, where i is the number of the commentary signal.

For the transmission only two full-range channels are available plus a certain number of band-limited commentary channels. The signals which are transmitted through these channels will be denoted by L', R', and C_i'.

At the receiver end, the signals can be mixed in several ways and routed to several loudspeakers. If three loudspeakers are used, they are placed as shown in Fig. 1. The signals which are played back at the loudspeakers are labeled L'', R'', and C''. For two-channel reproduction only L'' and R'' are present.

Since the proposed scheme requires a separate treatment of the low- and high-frequency parts for some signals, the subscripts LF and HF will be used to indicate these parts of the signal. The transition frequency between the low-frequency part and the high-frequency part is 7.5 kHz in order to avoid aliasing distortions when the low-frequency part is digitally coded with a 16-kHz sampling frequency.

1.1 Transmission Systems

We considered two alternative ways of transmitting the center signal for our test. The one possibility is

$$L' = L$$

$$R' = R$$

$$C_i' = C_i + C_{LF}, \qquad i \in [1, N] \tag{1}$$

where N is the number of commentary channels. The summation in $C_i' = C_i + C_{LF}$ refers to the addition of sample values.

The second possibility is

$$L' = L + \tfrac{1}{2}\sqrt{2}\,C_{HF}$$

$$R' = R + \tfrac{1}{2}\sqrt{2}\,C_{HF}$$

$$C_i' = C_i + C_{LF}, \qquad i \in [1, N]\,. \tag{2}$$

In this option the remaining high frequencies of the center signal C_{HF} are transmitted with a 3-dB reduction in amplitude together with the left and right signals. Because both alternatives use two full-bandwidth channels and one half-bandwidth commentary channel for the reproduction of three full-bandwidth signals (L, R, and C), $2\tfrac{1}{2}$-channel sound reproduction is effectively obtained.

1.2 Mixing Options

If the broadcasting scheme of Eqs. (1) is used, a simple $2\tfrac{1}{2}$-channel option results for sound reproduction.

1: NC (no center high frequencies)

$$L'' = L' = L$$

$$R'' = R' = R$$

$$C'' = C_i' = C_{LF} + C_i\,. \tag{3}$$

All the C_{HF} information is lost in this option, which means that, in comparison with full three-channel sound, timbre changes may be expected.

Reception of the signals given by Eqs. (2) leaves sev-

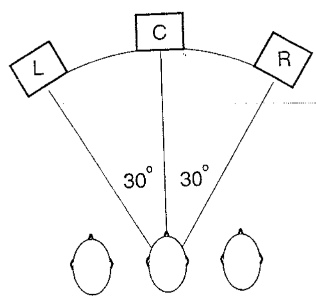


Fig. 1. Setup for listening test.

eral options to reproduce them over the three loudspeakers.

2) CTLR (center high frequencies to left and right)

$$L'' = L' = L + \tfrac{1}{2}\sqrt{2}\, C_{HF}$$

$$R'' = R' = R + \tfrac{1}{2}\sqrt{2}\, C_{HF}$$

$$C'' = C'_i = C_i + C_{LF}\,. \tag{4}$$

In this option the left and right source signals L and R are reproduced by the corresponding loudspeakers. The center source signal C is split into a low-frequency part, which is reproduced by the center loudspeaker, and a high-frequency part, which is sent through the left and right loudspeakers with a 3-dB attenuation. This attenuation is applied because it is assumed that the C_{HF} signals in both loudspeakers will add incoherently. Therefore the acoustic power spectrum of the center signal remains unchanged. This option does not need any signal processing at the receiver end.

3) OCS (optimal channel separation)

$$L'' = L'_{LF} + \sqrt{2/3}\, L'_{HF} = L_{LF} + \sqrt{2/3}\,(L + \tfrac{1}{2}\sqrt{2}\, C)_{HF}$$

$$R'' = R'_{LF} + \sqrt{2/3}\, R'_{HF} = R_{LF} + \sqrt{2/3}\,(R + \tfrac{1}{2}\sqrt{2}\, C)_{HF}$$

$$C'' = C'_i + \sqrt{1/3}\,(L'_{HF} + R'_{HF})$$

$$= C_{LF} + C_i + \sqrt{2/3}\,[C + \tfrac{1}{2}\sqrt{2}\,(L + R)]_{HF}\,. \tag{5}$$

In this option high-frequency signals are reproduced after remixing the received L'_{HF} and R'_{HF} signals. The coefficients of the remix are chosen such that optimal channel separation results. There is a 3-dB channel separation[1] in the high frequencies for all source signals. Assuming that the high-frequency parts of the reproduced signals at the place of the listener add incoherently, the acoustic power spectra of the L and R source signals are unchanged in this remix. The high-frequency part of C, however, is boosted by 1.2 dB compared with the low-frequency part.

1.3 Reproduction through Two Loudspeakers

The problem of reproducing a three-channel downmix through two loudspeakers is usually solved by adding the center signal, attenuated by 3 dB, to the left and right channels. If applied to the routing scheme of Eqs. (2), the following relationship is obtained:

$$L'' = L' + \tfrac{1}{2}\sqrt{2}\, C'_i = L + \tfrac{1}{2}\sqrt{2}\,(C_{HF} + C_{LF}) + \tfrac{1}{2}\sqrt{2}\, C_i$$

$$R'' = R' + \tfrac{1}{2}\sqrt{2}\, C'_i = R + \tfrac{1}{2}\sqrt{2}\,(C_{HF} + C_{LF}) + \tfrac{1}{2}\sqrt{2}\, C_i\,. \tag{6}$$

If a perfect reconstruction filter is used for splitting the center signal into two parts, C_{LF} and C_{HF} combine perfectly into the original center signal. In our listening tests a polyphase filter was used, which is such a perfect reconstruction filter [4].

2 LISTENING TESTS

The purpose of the proposed coding and decoding scheme is to achieve sound reproduction which is indistinguishable from full-bandwidth three-channel sound by using only two full-bandwidth channels and a half-bandwidth commentary channel. A listening test was designed to measure to what extent subjects can distinguish the various 2½-channel sound options from the three-channel sound. An appropriate testing method is the duo–trio test [5]. In such a test two stimuli A and B are presented to a subject in the order AAB or ABA. The subject has to identify which stimulus is the odd one out.[2] In this test paradigm, in contrast with the more conventional AB/BA paradigm, where the subject certainly has to be familiar with the characteristics of the stimuli A and B to be able to identify AB or BA, the subject does not need to be familiar with the stimuli.

Because, in general, stimuli can be difficult to distinguish, the subject will give a certain percentage of correct and incorrect responses. The percentage correct, $P_{correct}$, can be used as an estimate of the probability to answer correctly. From $P_{correct}$ the sensitivity d' of the subject to differences between the presented stimuli can be calculated, assuming no response bias is present. This is done by using the relationship [5].

$$P_{correct} = 1 - \Phi\left(\frac{d'}{\sqrt{2}}\right) - \Phi\left(\frac{d'}{\sqrt{6}}\right) + 2\Phi\left(\frac{d'}{\sqrt{2}}\right)\Phi\left(\frac{d'}{\sqrt{6}}\right) \tag{7}$$

in which $\Phi(k)$ is the error function,

$$\Phi(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{k} e^{-\frac{1}{2}t^2}\,dt\,. \tag{8}$$

The discrimination threshold is defined by performance for which $d' = 1$. If a subject's responses lead to a d' smaller than unity, it is said that the subject is unable to distinguish between stimuli A and B.

Full three-channel reproduction was used as one stim-

[1] Channel separation means the difference in level of a source signal (L, R, or C) in its destination loudspeaker and in an adjacent loudspeaker.

[2] This is known as a 3-interval 2-alternative forced-choice paradigm (3I2AFC).

ulus throughout all the experiments. The other stimulus was either the two-channel reproduction or one of the 2½-channel sound options. The source signal used was a one-sample pulse plus a band of noise.[3] The band was 0.5 kHz wide and at the upper side of the low-frequency part (7.0–7.5 kHz). Its energy was equal to the energy of the pulse,

$$\frac{U_{\text{noise}}}{U_{\text{pulse}}} \approx 1 . \tag{9}$$

The signal is broad band, and pulses can be localized rather easily if they are reproduced through single loudspeakers distributed throughout a room [7]. The pulse-noise signal was spatially coded into a three-channel signal by using the intensity weighting method, according to Theile [1].[4] The three-channel sound was then processed according to the proposed options [Eqs. (3), (4), and (5)].

2.1 Details of Listening Test

In Section 1.2 several mixing options were introduced. These options and a two-channel downmix of the three-channel signal were used in the listening test. The two-channel downmix used was

$$L'' = L + \tfrac{1}{2}\sqrt{2}\,C$$

$$R'' = R + \tfrac{1}{2}\sqrt{2}\,C . \tag{10}$$

The following comparisons were made:

1) Two channels versus three channels
2) NC versus three channels
3) CTLR versus three channels
4) OCS versus three channels.

The first comparison is of interest because it gives a reference of how well a subject can distinguish two-channel sound generated according to Eqs. (10) from three-channel sound. The second comparison tests whether the C_{HF} information is necessary. Comparisons 3) and 4), CTLR and OCS, test the systems that seem to offer the best chances of not being distinguishable from the original three-channel sound.

Because it is likely that the perceptibility of the difference between the original and the processed sound is dependent on the horizontal angle at which the sound sample is virtually located, measurements were made with sound placed at different angles. Three different horizontal angle positions for the phantom source were used, +25°, +15°, and +5°, and also their symmetric counterparts −25°, −15°, and −5°. (An angle of 0° would specify a pulse programmed in the center loudspeaker, an angle of +30° would designate a pulse programmed in the right loudspeaker; see Fig. 1.) From now on these placements will be referred to as placement categories 1, 2, and 3, which are sounds placed at the angles ±25°, ±15°, and ±5°, respectively.

Because there were four comparison types and three placement categories, and because each stimulus configuration was repeated 20 times, there were 240 trials. The 240 trials were grouped into two runs of 120 trials each. Each run was divided into four blocks, corresponding to comparisons 1, 2, 3, and 4, respectively. The placement categories were presented in random order, but also in a balanced way within each block. For the repeated presentation of each stimulus configuration, the sound sequences *AAB* and *ABA* were presented 10 times each. For each of the 10 sequences *AAB* or *ABA* the first interval (*A*) contained each of the two stimuli that had to be compared five times. So the first interval (*A*) represented the three-channel reference condition in 50% of the trials.

It is likely that the listening position of the subject influences the perceptibility of the differences between the presented stimuli. Therefore three listening positions were chosen, the central position and the two adjacent positions. Fig. 1 shows the setup of the listening room. The loudspeakers are placed in a circle having a radius of 3 m with the optimal listening position in the center of the circle. The seats are placed 0.85 m apart. The protocol used during the listening tests is described in the Appendix.

The answers of the subjects were recorded by response boxes and sent to a computer. After each trial, consisting of three pulses with 0.3-s intervals in between, the subjects had 2 s to reply. Feedback was given directly after the response, so the subjects could see the correct answers. When a subject did not reply, the feedback was still given after the 2-s response time. The feedback time was at least 1 s. A graphic display of the stimulus presentation is given in Fig. 2.

The experimental setup is shown in Fig. 3. The stimuli were prerecorded on analog audio tape. The experiment was under the control of a personal computer which also recorded the subject's responses. Response boxes were used containing answering buttons to be pressed by the subjects, LEDs to provide feedback, and a display showing the trial number.

Three Philips motional-feedback loudspeakers were used to produce the sound. The stimuli were produced at a comfortable listening level. The listening room had a volume of 210 m^3 and a reverberation time of about 0.4 s.

All but one of the persons that were invited to partici-

[3] After the first pilot tests it appeared that, in addition to possible localization cues, there was a much stronger timbre cue by which it was easy to distinguish between the original three-channel sound and any of the processed sounds. This timbre cue was caused by the steep filtering that was used to separate the signals in a high-frequency part and a low-frequency part [6]. In order to avoid this timbre cue the noise band was added. The timbre cue does not appear when natural sounds are processed. Music excerpts were processed, but none revealed the effect.

[4] For a given horizontal angle at which a sound sample has to be perceived, a level difference between two adjacent loudspeakers can be defined. The horizontal angle is approximately linearly dependent on this intensity difference expressed in decibels. An intensity difference of 13 dB corresponds to a sound that is audible at the position of the loudspeaker with the higher intensity. An intensity difference of 0 dB corresponds to a sound audible at the center point between two loudspeakers.

pate in the listening test were doing research on topics related to sound and a majority had some practice in doing listening tests.

3 RESULTS AND DISCUSSION OF LOCALIZATION TEST

Measured sensitivities d' of individual subjects are shown in Fig. 4, whereas Fig. 5 depicts the average d' values over all subjects for all the stimulus configurations. A dot indicates the measured sensitivity d', the columns indicate the statistical significance level: if the population of subjects would respond at chance level (50%), only 5% of the measured d' values is expected to be above this significance level. Furthermore, for each stimulus configuration the figure lists results for each phantom source location (rows) and listener position (columns).

In the case of two channels versus three channels (comparison 1), it is obvious that most subjects can distinguish between the stimuli, except for placement category 1 (which is at ±25°), where it is somewhat more difficult to distinguish between the stimuli, especially for those subjects sitting in the middle. In placement category 3 (±5°) a d' larger than unity is measured in most cases, even in the central listening position. Therefore it may be concluded that, in general, subjects distinguish between two-channel and three-channel reproduction, at least for the mixdown method used in this test.

Comparison 2 (NC high frequencies versus three channels) shows a similar picture. In placement category 1 there seems to be an even smaller perceptual difference between reference and test stimulus than in comparison 1. However, in placement category 3 there is definitely a larger difference. Category 3 appears to be the most critical in both comparisons 1 and 2. This can be explained by the fact that the center signal is dominant in this category, and it is this center signal that is subject to changes, not the left or right signals.

For comparison 3 (CTLR) the measured d' values are low. Looking at Fig. 5 it is clear that the average value of d' is smaller than 1 in all cases. This implies that, on average, subjects were not able to distinguish CTLR from three-channel reproduction.

Some of the individual d' values are larger than unity, and seven are statistically significant above chance level. Subject A seems to be able to distinguish CTLR from three-channel sound provided the listening position is central. If the remainder of the subjects is considered, only 6% of the measured d' values is significantly above the statistical chance level. These results seem to imply that one of the subjects was able to distinguish the stimuli. The rest of the subjects had 6% of their measurements above the significance limit, which is what can be expected considering the stochastic character of the measured d'. Therefore the hypothesis which says that this group cannot distinguish the stimuli is not rejected.

For comparison 4 (OCS) it appears to be slightly easier to distinguish between the stimuli, compared with comparison 3. The average d' is larger than unity in several cases, which implies that, on average, subjects were able to distinguish the stimuli to a perceptually significant degree. The $\bar{d}'$ is highest in categories 1 and 2. The reason for this is not clear. It could be that the high-

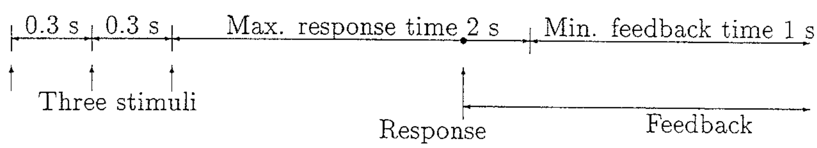


Fig. 2. Time structure of one trial in listening test.

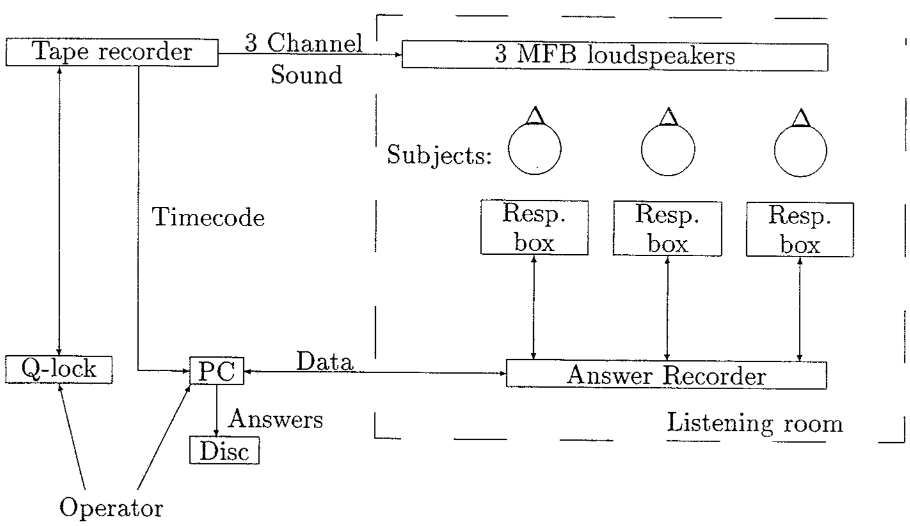


Fig. 3. Experimental setup.

frequency part of a sound intended for the left loudspeaker is distributed over the left and center loudspeakers. The high-frequency part of a sound, however, which is intended for the center loudspeaker is distributed evenly over the left and right loudspeakers. The unbalanced redistribution of high frequencies could be the cause of this effect, but this conclusion needs further investigation.

When looking at the results for comparison 4 for the individual subjects, it is clear that more d' values are above statistical significance level than in comparison 3. Even when omitting subject A, it still appears that 11% of the measurements is above the statistical significance level. This is more than expected for subjects responding at chance level.

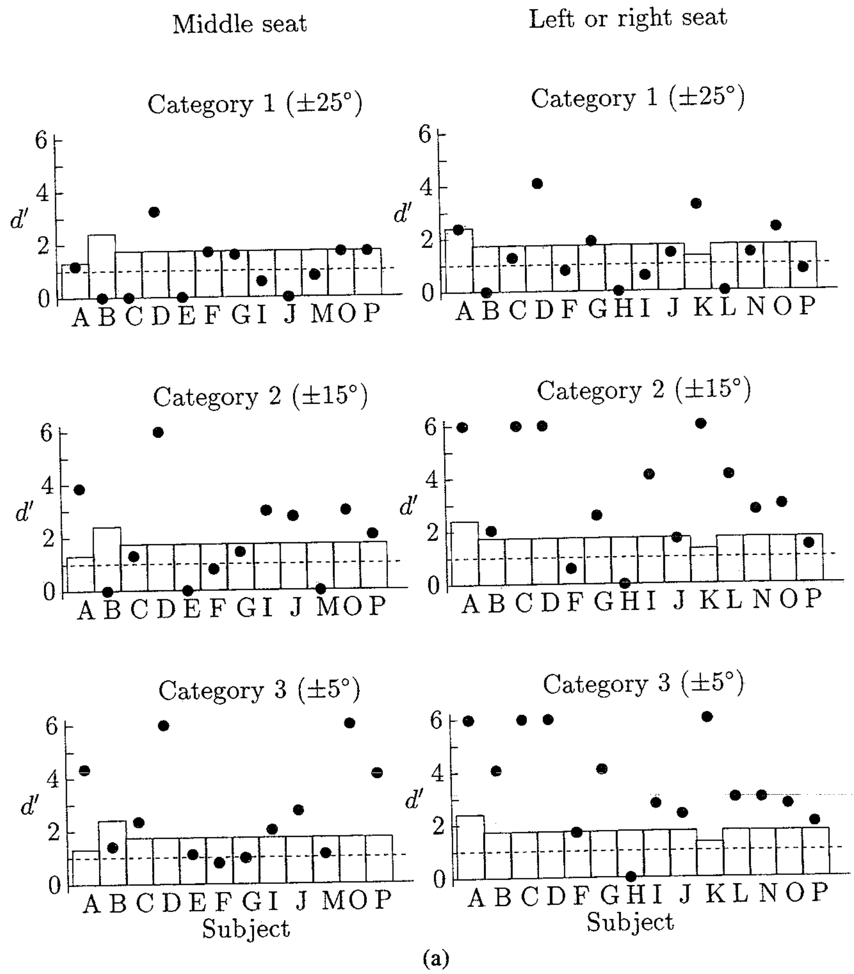
It is clear that subjects had the greatest difficulty in distinguishing between three-channel sound and CTLR sound (comparison 3). From the present results, obtained by using pulses as a source signal, it is concluded that the majority of people cannot distinguish between these two sounds. From an application viewpoint this is fortunate, because CTLR is easier to obtain than OCS since the latter needs two extra filters at the receiver end. From comparison 2 it can be concluded that the center high frequencies cannot be left out of the total sound. By comparing the results of comparisons 1 and 3, it is concluded that CTLR sound is much closer to three-channel sound than two-channel sound. Other experiments not reported here [8] confirm this conclusion.

Comparison 1, two-channel versus three-channel sound, has been repeated in another test with the pulse plus noise band programmed exactly in the center of the sound stage (0°). What was measured in this case is to what extent the phantom source, shaped by the left and right loudspeakers, is distinguishable from a real sound source in the center loudspeaker. It was found that they were easily distinguishable. Average d' values of more than 4 were found for subjects sitting in both a central and a noncentral listening position. This implies that a phantom source sounds different from a real sound source.

4 LISTENING TEST WITH REALISTIC PROGRAM SOURCES

In the experiment described clicks were used as stimuli. Human localization is very sensitive to clicks. It is therefore believed that the test results present a kind of worst-case situation. It would, however, be interesting to verify the results by using a more realistic type of program material.

Such an experiment has neen conducted within the



(a)

Fig. 4. Sensitivities d' for all subjects for all stimulus configurations. (a) Comparison 1, stereo versus 3 channels. (b) Comparison 2, NC versus 3 channels. (c) Comparison 3, CTLR versus 3 channels. (d) Comparison 4, OCS versus 3 channels.

European Eureka-95 HDTV project [9]. Subjects were asked to rate eight excerpts of about 15-s duration "on the basis of the locations and distribution of sound sources compared with those at the event itself, and the definition of these frontal sound images" [9]. The excerpts were taken from original HDTV productions. In the test the video was reproduced using an HDTV video screen, whereas the audio was reproduced as either one-channel (mono), two-channel (stereo), three-channel, or 2½-channel sound. The one- and two-channel versions were derived by downmixing from the three-channel version. The 2½-channel version was obtained from the three-channel version according to the CTLR scheme. Subjects were seated either on-axis or one or two seats away from the axis.

The results of this test showed an increased rating with the number of channels, that is, three-channel was preferred to two-channel, which in turn was preferred to one-channel reproduction. Averaged across all eight items and all listener positions, the quality increase from mono to stereo reproduction was about as large as the quality increase from two- to three-channel reproduction. Interestingly, for listeners in the center position the major improvement occurred from one- to two-channel reproduction, while off-center listeners rated the improvement from two- to three-channel reproduction as largest. This confirms the function of the center loudspeaker [1], [2]—it enlarges the listening area.

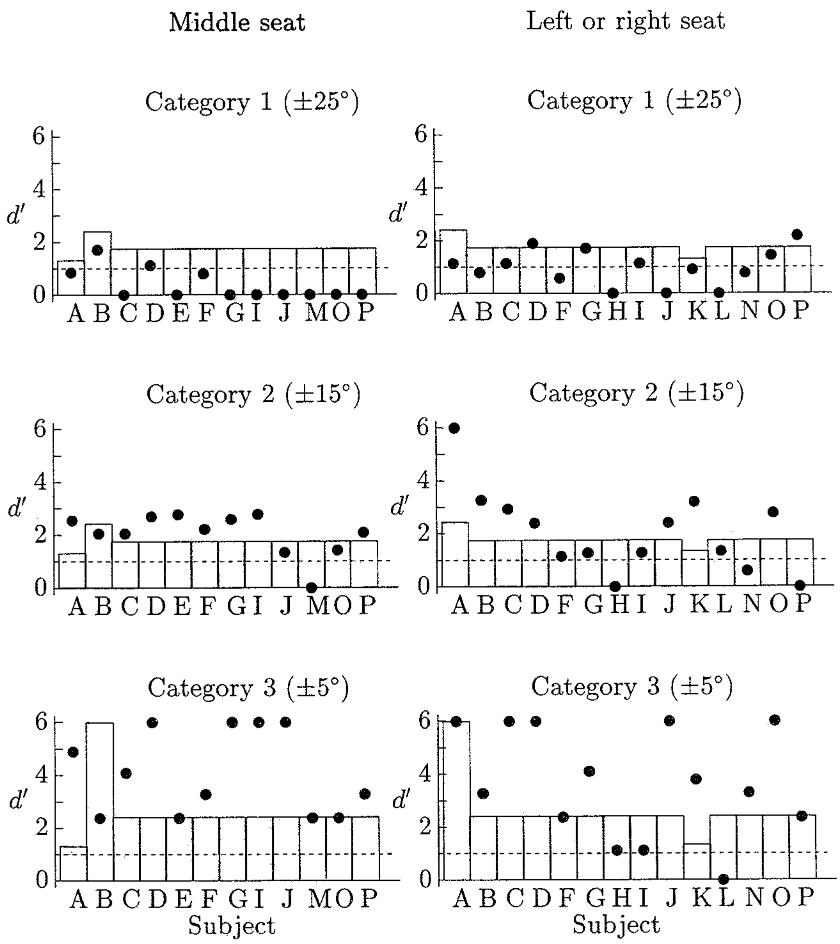
When comparing the ratings for the 2½- and three-channel systems, no significant difference between both systems could be identified. This result with realistic program material supports the conclusion from the localization experiment with clicks.

5 IMPLICATIONS FOR CURRENT STANDARDS

The perceptual equivalence of transmitting the high-frequency part of the center signal through the left and right channels with transmission through the center channel is useful for multichannel bit-rate reduction systems. The bit-stream format currently in the process of being standardized by ISO/MPEG includes a mode in which the center signal is bandwidth limited [10]. The present paper shows that if the mode is chosen, the CTLR option should be used to decode the signal.

If a transmission system offers half-bandwidth commentary channels, the capacity of the complete center channel can be saved as described in Section 1. In other words, if that system is a two-channel system augmented with commentary channels, it may be extended to a three-channel system. D2MAC is such a system.

D2MAC (MAC stands for multiplexed analogue component, D2 is the version identifier) allows for full-range audio channels (32-kHz sampling rate) as well as for



(b)

Fig. 4. continued

channels of limited bandwidth (16-kHz sampling rate). The latter are thought to be used for commentary in multilingual broadcasting. Any combination of these is allowed as long as the total bit rate does not exceed the capacity of the system. The actual configuration also depends on other factors, such as coding options and error protection levels chosen. Reproduction is thought to be by one or two loudspeakers.

D2MAC allows for different coefficients for mixing the selected commentary signal into the two-channel sound [11], [12]. Here the proposed CTLR option imposes a restriction since it requires a (net) −3-dB mixing of the center signal with the left and right signals. Furthermore, if a user prefers to have the commentary channels switched off, the user will lose the low frequencies of the center information as well. A remedy would be to provide a commentary channel which contains no commentary but only the C_{LF} signal.

6 CONCLUSIONS

CTLR is the best substitute for three-channel sound. It requires only 2½ channels. Since some half-bandwidth channels are usually available within most standards, this amounts to saving a full channel.

OCS is also an acceptable option, but less so than CTLR. This was not expected because OCS distributes the high-frequency energy more evenly over the loudspeakers. Because the OCS option requires more complexity at the receiver side, the CTLR option is even more preferred.

7 REFERENCES

[1] G. Theile, "HDTV Sound Systems: How Many Channels?," in *Proc. AES 10th Int. Conf.* (1991), pp. 147–162.

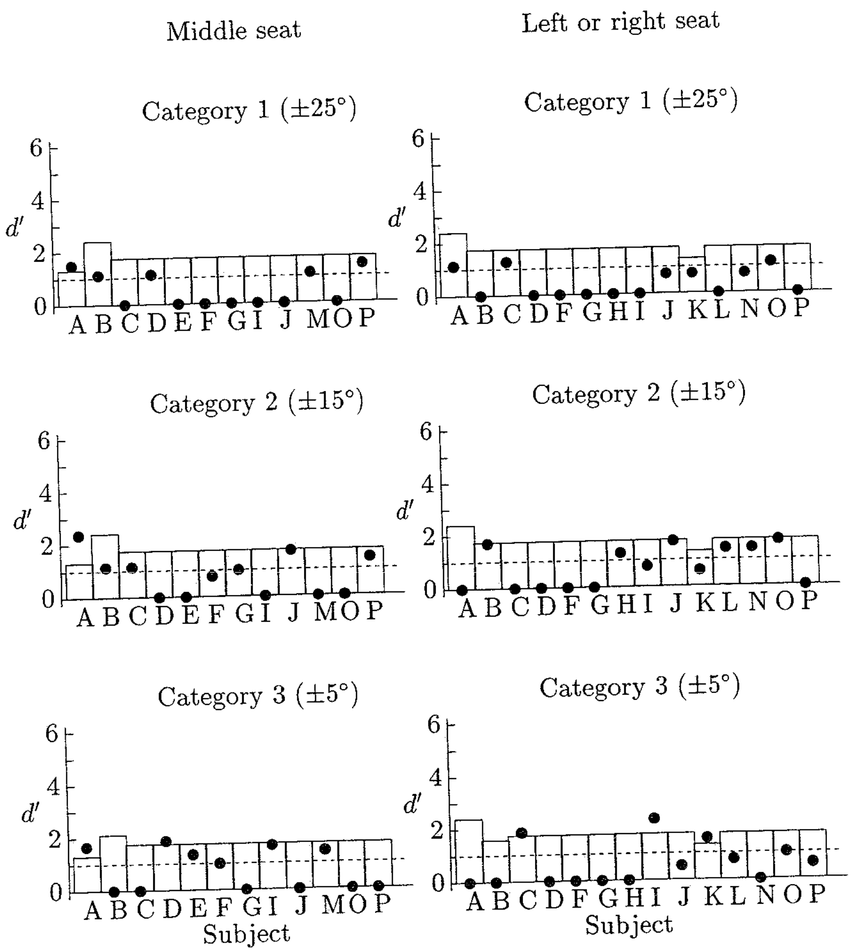
[2] K. Ohgushi, S. Komiyama, K. Kurozumi, A. Morita, J. Ujihara, and K. Tsujimoto, "Subjective Evaluation of Multi-Channel Stereophony for HDTV," *IEEE Trans. Broadcast.*, vol. BC-33, pp. 197–202 (1987 Dec.).

[3] F. K. Harvey and M. R. Schroeder, "Subjective Evaluation of Factors Affecting Two-Channel Stereophony," *J. Audio Eng. Soc.*, vol. 9, pp. 19–28 (1961 Jan.).

[4] J. H. Rothweiler, "Polyphase Quadrature Filters—A New Subband Coding Technique," in *Proc. ICASSP 83* (Boston, MA, 1983), pp. 1280–1283.

[5] R. A. Bradley, "Some Relationships among Sensory Difference Tests," *Biometrics*, no. 19, pp. 385–397 (1963).

[6] A. Kohlrausch and A. J. M. Houtsma, "Pitch Related to Spectral Edges of Broadband Signals," *Phil. Trans. R. Soc.*, no. 336, pp. 375–382 (1992).



(c)

Fig. 4. continued

[7] W. M. Wagenaars, "Localization of Sound in a Room with Reflecting Walls," *J. Audio Eng. Soc.*, vol. 38, pp. 99–110 (1990 Mar.).

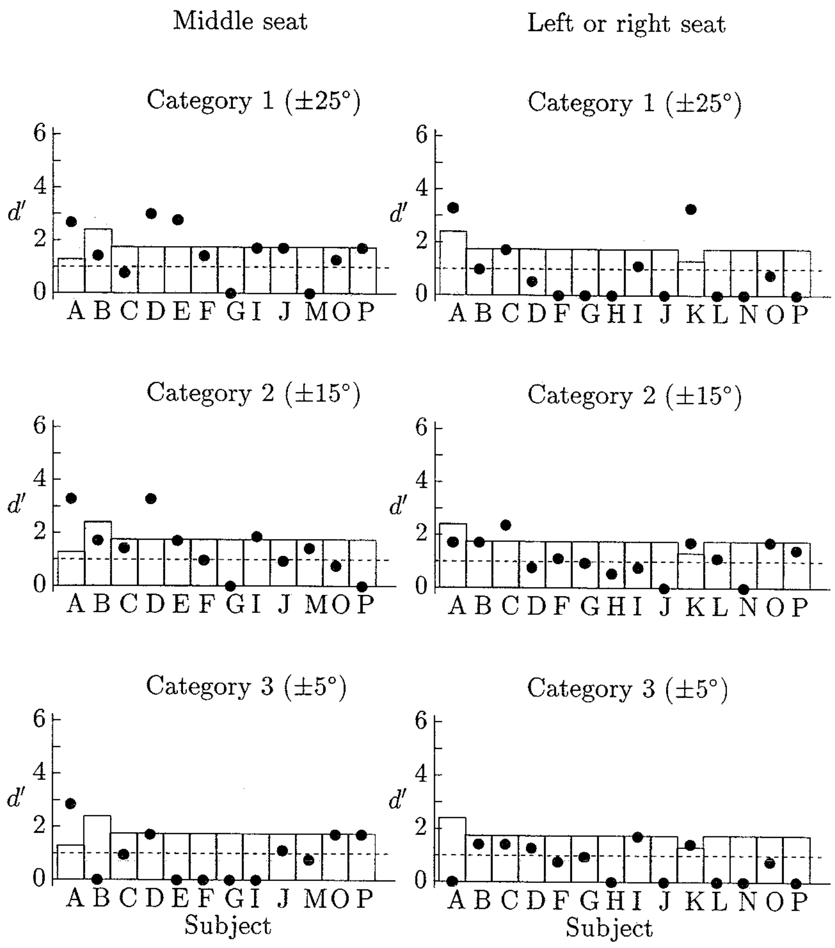
[8] S. van de Par, "On 2½-Channel Sound Compared to Three-Channel Sound," Rep. 858, Institute for Perception Research (1992).

[9] "Report on the Second Subjective Test on Multi-Channel Sound Frontal Representation," EU95-011-094 (1993 Jan.); also available as ITU-R Doc. ITU-R TG10-1/26 and ITU-R TG10-3/32.

[10] W. R. Th. ten Kate, R. C. E. Heijmans, S. P. Straetemans and P. M. Boers, "5-Channel Musicam Codec," presented at the 94th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 41, p. 406 (1993 May), preprint 3671.

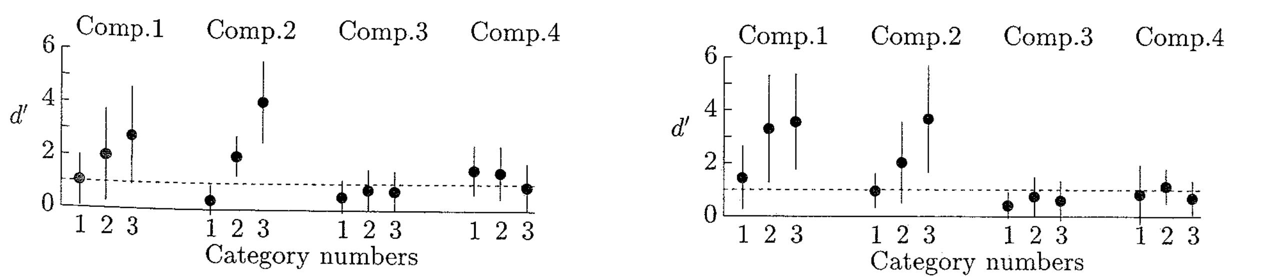
[11] "The D-MAC/Packet System for Satellite and Cable," *IBA Tech. Rev.*, pp. 30–37 (1988 Nov.).

[12] "Specification of the Systems of the MAC/Packet Family," Tech. 3258E, European Broadcasting Union (1986 Oct.).



(d)

Fig. 4. continued



(a) (b)

Fig. 5. Average sensitivities for all stimulus configurations. Comp. 1—stereo versus 3 channels; Comp. 2—NC versus 3 channels; Comp. 3—CTLR versus 3 channels; Comp. 4—OCS versus 3 channels; categories 1—±25°; 2—±15°; 3—±5°. (a) Subjects in middle seat. (b) Subjects in left or right seat.

APPENDIX

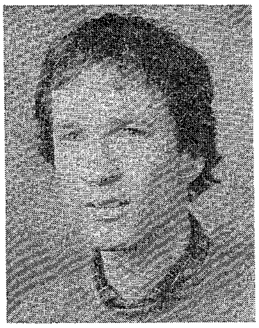
The protocol used during the listening tests is as follows:

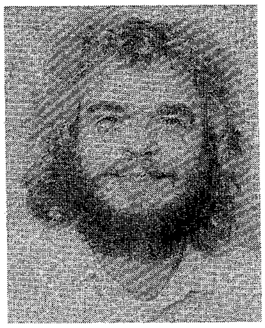
- Let the subjects take their places and give them instructions, explain details if necessary.
- Carry out training with 15 trials to let the subjects become acquainted with the experiment. Repeat if necessary.
- Do a run of 120 trials.
- Short pause and ask for comments.
- Do another run of 120 trials (same listening position).
- Show the results (percentages correct) to the subjects.

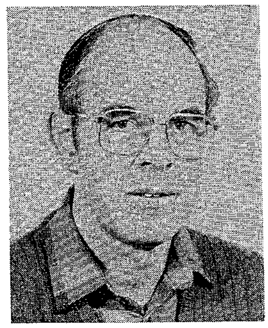
The same protocol is repeated later on with the same subjects in another listening position.

THE AUTHORS


S. L. J. D. E. van de Par


W. R. Th. ten Kate


A. Kohlrausch


A. J. M. Houtsma

Steven van de Par was born in Goirle, The Netherlands, in 1966. He became interested in physics, audio, and music, and decided to study the former and keep the others as hobbies.

During his physics study at the Technical University of Eindhoven he had the opportunity to combine his hobbies and study during his masters-thesis work at the Philips Research Laboratories Eindhoven. He is currently working on the topic of binaural perception at the Institute for Perception Research in Eindhoven.

•

Warner R. Th. ten Kate was born in Leiden, The Netherlands, in 1959. He studied electrical engineering at Delft University of Technology, graduating in 1982 cum laude, and received the 1983 prize awarded by the Delft University Fund. During the final stages of his studies his research was directed at solar cells of amorphous silicon and silicon radiation detectors. He received the Ph.D. degree in 1987.

Since 1988 Dr. ten Kate has been working in the Acoustics Group of Philips Research Laboratories. In 1985 he also began studying the French horn at the Royal Conservatory in The Hague and graduated in 1989 with distinction.

•

Armin Kohlrausch was born in Bremen (Germany) in 1953. He studied physics at the University of Goettingen and specialized in acoustics. He received his master's degree in 1980 and his Ph.D. in 1984, both under supervision of Prof. M. R. Schroeder. From 1985 until 1990 he worked at the Drittes Physikalisches Institut at the University of Goettingen, where he was responsible for research and teaching in the fields of psychoacoustics and room acoustics.

In 1991 he joined the Hearing and Speech group of the Institute for Perception Research (IPO) in Eindhoven, The Netherlands. His main scientific interest is the experimental study and modeling of auditory perception in humans.

•

Adrianus J. M. Houtsma was born in Noordwijkerhout, The Netherlands, in 1938. He received the Music Diploma A, major organ (Arnhem, 1961) and the B.A. in theology (Nijmegen, 1963). In 1965 he received the S.B. in E.E. from the University of Villanova, followed by S.M. and Ph.D. degrees from the Massachusetts Institute of Technology in 1966 and 1971. Until 1982 he was lecturer in the departments of Humanities and Electrical Engineering at M.I.T., after which he joined the research staff of the Institute of Perception Research at the Eindhoven University of Technology. There he holds the chair Psychoacoustics and its Technical Applications.

Professor Houtsma is a member of the Dutch Acoustical Society (NAG), the European Acoustical Association, the Dutch Society of Audiology, the Catgut Acoustical Society, and a fellow of the Acoustical Society of America.