POST-PROCESSING OF 2-CHANNEL STEREO STUDIO RECORDINGS OF CLASSICAL MUSIC WITH ROOM SIMULATION - A PSYCHOACOUSTICAL EXPERIMENT

Alois Melka

Laboratory of Psychoacoustics
Akustika Praha
Prague, Czech Republic
a.melka@mybox.cz

Václav Syrový, Ondřej Moravec

Sound Studio of the Faculty of Music
Academy of Performing Arts
Prague, Czech Republic
syrovy@hamu.cz, moravec@hamu.cz

ABSTRACT

Thirty experienced music listeners tried to set an optimum ‘wet-to-dry signal ratio’ of the resulting mixed sound by adjusting the soft knob of the Lexicon 300 digital audio processor for eight different factory-installed room simulation effects. A pre-processed two-channel stereo studio recording of classical music was used as the input ‘dry’ signal. Results showed that experiment participants could be divided into two diverse groups, one of which preferred markedly greater values of the ‘wet’ signal than the other. The group of ‘wet’ sound advocates was composed largely of sound engineers, while the ‘dry’ sound preference came from acousticians and musicians. An approximately linear dependence of the optimal level difference of the input ‘dry’ signal and the processed ‘wet’ signal on the simulated reverberation time was found. This finding is in agreement with the conclusions of the psychoacoustical experiment carried out by Schmidt in three-dimensional synthetic sound field [9].

1. INTRODUCTION

When making a stereophonic studio recording of classical music, microphones are usually placed close to the individual sound sources, where the direct signal of the source predominates the reflected sound. At first, the sound engineer adjusts the sound levels, and, where appropriate, also the spectra of each of these almost anechoic recordings of solo instruments and/or vocals or instrumental and/or vocal groups. He then mixes them together into a two-channel ‘dry’ stereophonic recording that meets the criteria for optimum ‘spectral uniformity’, ‘sound-stage imaging’, ‘dynamics’, and ‘robustness’ of the recording [1]. In the final post-processing phase he adjusts the recording using an audio processor for ‘ambience reproduction’ [1], i.e. to imitate the acoustics of a room suitable for performing the music in question – ‘*room simulation*’ [2].

Audio processor manufacturers advise in their user manuals to begin the work with the so-called *factory presets*, i.e. their recommended parameter configurations for simulating the acoustics of various typical spaces. When an user selects a factory preset bearing a label of room type that is suitable for performing a certain piece of music (e.g. the preset ‘Church’ for church music) he should be guaranteed that the ambience reproduction of the recording – within the capabilities of two-channel stereophonic reproduction – will resemble the situation of listening to a live music performance in such room. If the user leaves the configuration of factory preset parameters unchanged his task consists only in adjusting the ‘right’ *degree of mixing* of the processed signal to the original ‘dry’ signal. In conformity with [3] we shall call this variable *OMIX* (‘overall mix’) and define it by the formula

$$OMIX = \frac{U_{proc}}{U_{inp} + U_{proc}} .100 \quad \% \qquad (1)$$

where U_{proc} is the voltage of the processed signal and U_{inp} the voltage of the input ‘dry’ signal. The output voltage of the resulting mixed signal $U_{inp} + U_{proc}$ remains constant.

In connection with this seemingly simple operation we can ask several questions that are important for the sound engineer’s work: Do optimum values of *OMIX* differ when using different presets options for the same type of music? Are experienced listeners of both live and recorded classical music unanimous in their views on optimum *OMIX* values, or is it mainly a matter of subjective preference? How large are differences in the perceived sound quality among recordings which have been processed with different room simulation effects and adjusted to an optimum *OMIX* value in each case? In order to clarify these questions we have proposed and carried out psychoacoustical experiments, the first of which is described below.

2. EXPERIMENT DESCRIPTION

As the activities of the Faculty of Music in Prague are directed primarily towards classical music, for the purpose of this experiment we used a pre-processed ‘dry’ version of a two-channel stereo recording of the third movement (Presto) of F. Benda’s E flat major Concerto for violin and string orchestra. The primary recordings were made in the recording studio of the Faculty of Music with a floor area 100 m^2, volume of 700 m^3, and reverberation time of 0,8 s. The instrument and microphone locations at the recording are indicated in Fig. 1.

For room simulation we selected *three algorithms of digital audio effects* provided by the *Lexicon 300 processor* [3]: *Random Hall*, *Random Ambience*, and *Rich Plate*. From the various options offered by the Random Hall algorithm we used the presets *Small Hall*, *Medium Hall*, *Large Hall*, and *Church*, keeping unchanged the factory-installed values of all parameters in each case. We also added two modified preset variants: *Modified Small Hall* and *Modified Church*. These differ from the original presets only in reverberation time, which was, in the first case, reduced to an extremely small value and, in the second case, increased to an extremely large value (see Table 1). From the possibilities offered by the algorithm Random Ambience, which 'can provide the missing blend and depth' according to [3], we selected the preset *Ambience*. The preset *Rich Plate* from the homonymous algorithm was selected because numerous sound engineers who have previously worked with a plate reverberator now tend to prefer its digital imitation.

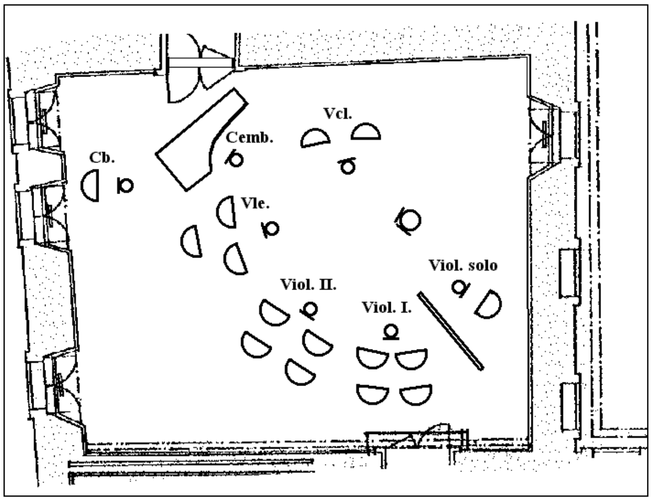

Figure 1. *Floor plan of recording arrangement*

The listening panel of 30 persons participating in this experiment consisted of 8 musicians, 9 sound engineers, and 13 acoustics researchers and university teachers.

Test sessions were arranged individually for each of the participants in the sound control room of the Faculty of Music in Prague with a floor area 35 m^2, volume of 120 m^3, and reverberation time of 0,4 s. The person concerned sat at the sound engineer's standard listening location and controlled the processor's 'soft knob'. The panel of the processor was covered to prevent the person to identify the apparatus and read the data on its display. The test administrator set the preset parameters configurations according to the experimental plan, provided the reproduction of test music and registered the *OMIX* values adjusted by the persons concerned. Each subject was instructed *to readjust the soft knob* during the roughly 3 1/2 minutes of playing time, for each of the eight variants examined, *to a position where the overall sound quality of the reproduction appeared to be most satisfactory* for him ('adjustment method' [4], [5]). To prevent the listeners from identifying the sound variants according to the sound decay after each signal interruption, fading with five seconds decrease was used. Maximum sound level at the listening position was 78 dB. The test session consisted of reading over written instruction, initial training, the test itself, and a retest which included eight items identical to those used in the test. The average duration of the test session was 69 minutes. The order of presentation of the test items was different for different subjects. The experimental design compensated time-order errors of adjustments and balanced first-order residual effects [4], [6].

3. EXPERIMENT RESULTS

Fig. 2 shows *median values, extreme values and quartile ranges of optimum OMIX adjustments* for all 19 items from A_0 to F_2, as set by the 30 participants in the experiment. Items from each part of the experiment are arranged in accordance with increasing reverberation time and/or size of the simulated room. The median values of individual optimum adjustments decrease monotonously with increasing magnitude of these variables, except for the preset variant G (Ambience).

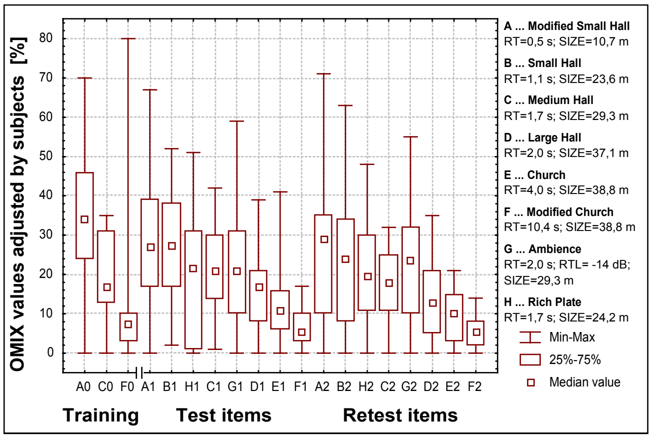

Figure 2. *Median values, extreme values and quartile ranges of individual optimum OMIX adjustments*

The results obtained from both the test and the retest are quite similar. This was confirmed by the correlation analysis, according to which the *test/retest reliability* coefficient [4] is significant for all eight preset variants. This does not necessarily mean that the *OMIX* values obtained in the test and the retest showed significant correlation for each individual. Significant correlation was identified only with respect to 13 individuals ('reliable' persons), it was not significant in respect of 17 others ('unreliable' persons). The composition of *groups of 'reliable' and 'unreliable' subjects* according to subject occupation is shown in Fig. 3.

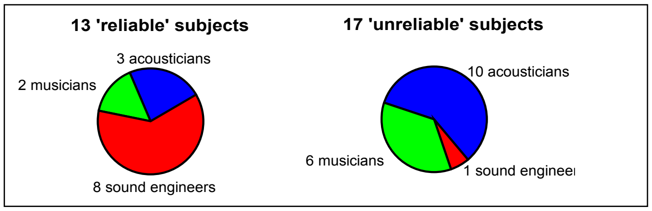

Figure 3. *Composition of 'reliable' and 'unreliable' experiment participants according to occupation*

The degree of *agreement among the subjects* was assessed for test and retest by using the Kendall concordance coefficient *W* [7]. The concordance was found significant in both cases and was better for the retest data (W = 0,446) than for the test data (W = 0,398).

We used *hierarchical cluster analysis* of the values measured in order to find out whether certain groups exist among the experiment participants with markedly different opinion as to optimum size of the *OMIX* value. In the analysis we used four different techniques ('single linkage', 'complete linkage', 'the Ward's method', and the 'k-means clustering method') [8]. An example of cluster analysis output is shown in Fig. 4.

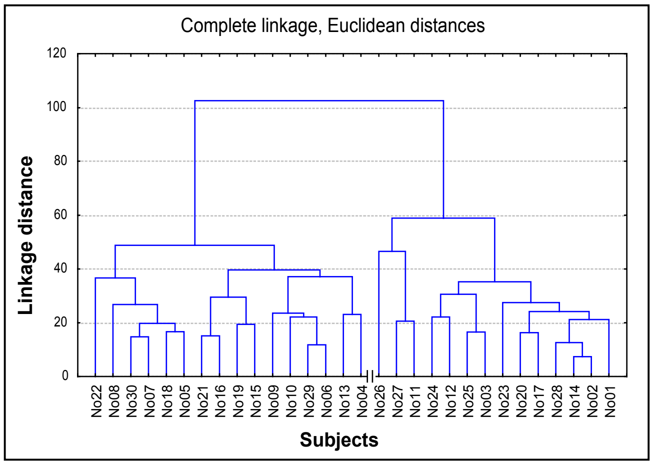

Figure 4. *Results of hierarchical clustering of 30 subjects with a 'complete linkage' joining (tree clustering) technique*

All clustering techniques led to the same conclusion. *Participants in the experiment could be divided into two groups* of approximately the same size, one of which tended to prefer markedly greater values of the added processed signal (*'wet' sound advocates*) than the other (*'dry' sound advocates*); one of the participants fluctuated between the two groups.

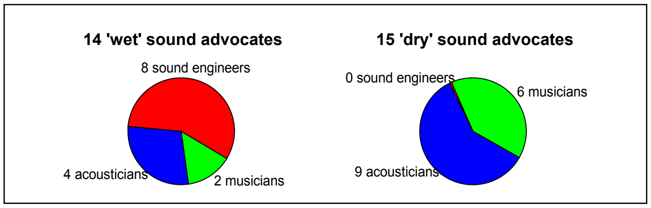

Figure 5. *Composition of advocates of 'wet' and 'dry' sound reproduction according to occupation*

The composition of the two clusters is shown in Fig. 5. Sound engineers without exception preferred greater proportions of processed sound, whereas the majority of musicians and acoustic researchers preferred a more 'dry' sound. The difference in the mean setting by the two groups is considerable. Whereas the *OMIX* median values for the 'wet' sound advocates varied between 8 % and 39 %, the range for the second group was only from 3 % to 11 % (see Table 1).

4. RESULTS ANALYSIS AND DISCUSSION

The factory-installed values of all preset parameters for eight effects observed during testing are shown in Table 1. In the upper part of this table we summarized medians of the adjusted *OMIX* values for all 30 subjects ($OMIX_{all}$), for the 'wet' sound advocates ($OMIX_{wet}$) and 'dry' sound advocates ($OMIX_{dry}$), as well as the corresponding values of

$$\Delta L = L_{inp} - L_{proc} = 20\log\frac{U_{inp}}{U_{proc}} = 20\log\frac{100-OMIX}{OMIX}\ \text{dB} \quad (2)$$

Table 1: *Medians of optimum OMIX values and corresponding ΔL values in comparison with the preset parameter values of effects under study*

EFFECT:	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
$OMIX_{all}$ [%]	29	24	18	13	10	5,5	23,5	19,5
ΔL_{all} [dB]	7,8	10,0	13,2	15,5	19,1	24,7	10,3	12,3
$OMIX_{wet}$ [%]	39	34	26	21	14	8	32	31
ΔL_{wet} [dB]	3,9	5,8	9,1	11,5	15,8	21,2	6,5	6,9
$OMIX_{dry}$ [;%]	10	8	11	5	3	3	10	11
ΔL_{dry} [dB]	19,1	21,2	18,2	25,6	30,2	30,2	19,1	18,2
RTIM [s]	0,5	1,1	1,7	2,0	4,0	10,4	2,0	1,7
RLVL [dB]	0	0	0	0	0	0	-14	0
SIZE [m]	10,7	23,6	29,3	37,1	38,8	38,8	29,3	24,2
SPRD [ms]	64	64	97	157	226	226	-	31
LINK [-]	link	link	link	link	link	link	-	link
TDCY [kHz]	5,5	5,5	4,4	3,6	2,9	2,9	-	12,8
ROLL [kHz]	1,3	1,3	1,0	0,9	4,0	4,0	3,6	14,2
BASS [-]	1,2	1,2	1,2	1,2	1,5	1,5	1,0	1,0
XOVR [kHz]	0,2	0,2	0,2	0,2	0,9	0,9	1,0	0,8
SPIN [?]	28	28	33	38	40	40	30	-
RAND [%]	-	-	-	-	-	-	-	82
DIFF [?]	65	65	68	65	70	70	75	99
WAND [ms]	6,762	6,762	8,016	10,10	10,10	10,10	7,098	-
SHAP [?]	70	70	100	120	82	82	-	16
SHLF [dB]	-6	-6	-6	-6	off	off	-	-
PDLY [ms]	0	0	0	0	36	36	0	0
DDLY [ms]	-	-	-	-	-	-	0	-
DLY1 [ms]	8	8	14	14	32	32	-	500
LVL1 [dB]	off	off	off	off	-8	-8	-	-
FBK1 [%]	-	-	-	-	-	-	-	0
DLY2 [ms]	18	18	20	20	26	26	-	500
LVL2 [dB]	off	off	off	off	-7	-7	-	-
FBK2 [%]	-	-	-	-	-	-	-	0
DLY3 [ms]	44	44	40	40	374	374	-	30
LVL3 [dB]	off	off	off	off	-14	-14	-	off
FBK3 [%]	0	0	0	0	12	12	-	-
DLY4 [ms]	34	34	48	48	462	462	-	38
LVL4 [dB]	off	off	off	off	-18	-18	-	off
FBK4 [%]	0	0	0	0	12	12	-	-
DLY5 [ms]	-	-	-	-	-	-	-	74
LVL5 [dB]	-	-	-	-	-	-	-	off
FBK5 [%]	-	-	-	-	-	-	-	0
DLY6 [ms]	-	-	-	-	-	-	-	250
LVL6 [dB]	-	-	-	-	-	-	-	off
FBK6 [%]	-	-	-	-	-	-	-	0

i.e. the optimum voltage level difference of the input ‘dry’ signal U_{inp} and the processed signal U_{proc}.

The highlighted parameter values in Tab. 1 could have influenced the perceived sound character of the corresponding effect variant during the experiment. The parameters with non-highlighted values were either switched off (‘off’) or had a zero value (‘0’) or were unrelated to the effect (‘-’).

Data in the ‘*RLVL*’ (reverberation level) row of Table 1 explain the aforementioned disturbance of the monotonous decrease of *OMIX* at the effect G (Ambience), where, in contrast to the other effects, reverberation level was reduced by 14 dB. Obviously, listeners tried to compensate for the attenuation by adjusting larger *OMIX* values in this case.

By correlating the optimum *OMIX* values of eight effects under examination with the corresponding values of all relevant parameters given in Table 1 we have found that only three parameters, namely the reverberation time (*RTIM*), simulated size of the room (*SIZE*) and degree of randomization of the reverberant sound (*SPIN*) correlate highly significantly (at the 1 % level of significance) with *OMIX*. As *RTIM* varies linearly with the setting of *SIZE* in Lexicon 300 [3], it is sufficient to examine the relation of only one of these parameters to *OMIX* or to ΔL. We chose the parameter *RTIM* and demonstrated its relation to ΔL_{all}, ΔL_{wet}, and ΔL_{dry} graphically in Fig. 6.

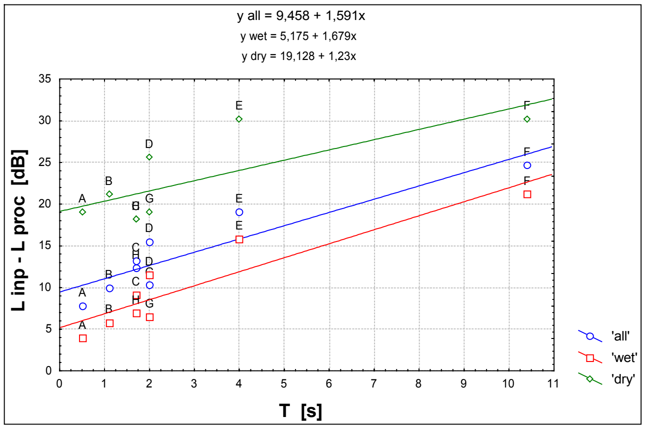

Figure 6. *Relation between the adjusted optimum value of ΔL and simulated reverberation time T for all subjects, ‘wet’ sound advocates and ‘dry’ sound advocates*

The regression lines fitted to the three sets of paired observations show that *ΔL_{opt} depends roughly linearly on T* for all the three subject groups. The slope of the regression line is 1,6 dB/s for all subjects, 1,7 dB/s for those favouring the ‘wet’ sound, and 1,2 dB/s for those favouring the ‘dry’ sound where the inaccuracy of the linear prediction was fairly large. This finding resembles the conclusions of the experimental study [9] where *W. Schmidt proved that the same degree of ‘spatial impression‘* (‘Raumeindruck’) *can be simulated* in a synthetic sound field *either by lengthening the reverberation time T* that causes an increase of ‘reverberance’ (‘Halligkeit’) *or by decreasing the quantity H* (‘Hallabstand’, ‘direct sound measure’) defined as level difference of direct sound energy W_{dir} and the energy of reverberation W_{rev}

$$H = 10\log\frac{W_{dir}}{W_{rev}} = L_{dir} - L_{rev} \quad \text{dB} \qquad (3)$$

which causes an increase of ‚spaciousness‘ (‘Räumlichkeit‘). According to [9], it is impossible to perceive the two ‘spatial impression’ components mentioned above separately. *Schmidt also found out* that *the level of reverberation above the absolute threshold of its audibility depended linearly on the subjectively equivalent value* of *T* if $H = 0$ dB.

A direct quantitative comparison between our and Schmidt’s results would not be correct for several reasons. For instance, Schmidt searched for paired values of *T* and *H* which brought about *the same magnitude of total spatial impression*, whereas we looked for values of *OMIX* for which *the overall perceived sound quality of reproduction reached an optimum*. Furthermore, while Schmidt did not simulate any early reflections at all, two single early reflections and two clusters of them were simulated in effects E, F (see Table 1).

Nevertheless, the dependence of the optimum voltage level difference ΔL_{opt} on the reverberation time *T* found in our experiment seems to be in good qualitative agreement with the findings published in [9].

The competence of sound engineers to set optimum *OMIX* values with good reproducibility as demonstrated in Fig. 3 is certainly not surprising. The reliability of *OMIX* adjusting by the other subjects might have been better following more thorough training.

A better agreement among the subjects found in the retest indicates the increased quality of subject performance with time.

The finding that musicians mostly preferred a more ‘dry’ sound character can be probably explained by the fact that they are accustomed to listening to the sound of their own instrument and the sound produced by other players in a sound field near these acoustic sources, where direct sound energy dominates.

Most subjects noted specific difficulty due to the dependence of the optimum adjustment on the changeable time structure of the music used (they tended to prefer different optimum value of *OMIX* for *tutti* and for solo violin). They were advised to try to find a ‘reasonable’ compromise setting.

5. CONCLUSIONS AND FURTHER WORK

The results of this experiment have shown that those experienced in listening to both live and reproduced classical music manage to adjust an perceptually optimal mixture of pre-processed ‘dry’ stereo music signal recorded in a studio and the same signal post-processed with various ‘room simulation‘ digital effect algorithms without very much difficulty. Audio engineers were able to do it with acceptable reliability without longer training. The reliability of adjusting was smaller in musicians and acousticians, but it increased during the test session.

Marked divergence seems to exist between the experienced listeners of classical music concerning the preferred ‘wet-to-dry signal ratio’ in the resulting mixed sound. From this point of view, our listening panel consisted of two diverse groups of listeners of equal size. Sound engineers belonged without exception to the group of ‘wet’ sound advocates, whereas most musicians and acousticians formed the group of ‘dry’ sound advocates. Different attitudes in both groups can likely be explained by different professional habits of the group members.

We found that the relation of the adjusted optimum voltage level difference of the input ‘dry’ signal and the processed signal to the simulated reverberation time was approximately linear for both the ‘wet’ sound and ‘dry’ sound advocates. This finding is in agreement with the finding of Schmidt [9] that the same degree of ‘spatial impression‘ can be reached by increasing ‘direct sound measure’ (‘Hallabstand’) and simultaneous decreasing reverberation time (or vice versa) in a simulated sound field.

Naturally, the validity of our quantitative results, which were found using a piece of early Classical chamber string music must not be generalized as applying to another sorts of music without verification.

A subsequent psychoacoustical experiment will try to find out, using the same piece of music, to what extent the listeners are able to perceive and assess sound differences between the eight studied effect variants if the processed signal is mixed with the original ‘dry’ signal in an optimum relationship.

6. ACKNOWLEDGMENTS

This work was carried out as part of the research project COST OC G6.20 ‘Digital Audio Effects – Subjective Assessment’ and was supported by the Czech Ministry of Education and Youth. We wish to thank all of the individuals who participated in this experiment for their effort.

7. REFERENCES

[1] AES20-1996: AES recommended practice for professional audio – Subjective evaluation of loudspeakers. *J. Audio Eng. Soc.* 44 (1996), 382-401.
[2] Zölzer, U., Digital Audio Signal Processing. J. Wiley&Sons, Chichester 1997.
[3] Lexicon Inc., USA: 300 Digital effects system. V 3.0 Owner‘s Manual, 1997.
[4] Guilford, J. P., Psychometric Methods. 2nd Ed. McGraw-Hill, New York 1954.
[5] Rasch, R. A. – Plomp, R., The perception of musical sounds. In: The Psychology of Music, D. Deutsch (Ed.), Academic Press, Orlando 1982.
[6] Cochran, W. G. – Cox, G. M., Experimental Designs. J. Wiley& Sons, New York 1957.
[7] Ferguson, G. A., Statistical analysis in psychology & education. 3rd Ed. McGraw-Hill, New York 1971.
[8] STATISTICA 5.0. StatSoft, Inc., Tulsa 1995.
[9] Schmidt, W., Zusammenhang zwischen Hallabstand und Nachhallzeit für den Raumeindruck (Halligkeit und Räumlichkeit bei Musik). Hochfrequenztechnik und Elektroakustik 77 (1968), 37-42.