International Symposium on Room Acoustics
Satellite Symposium of the 19th International Congress on Acoustics
Seville, 10-12 September 2007

HOW MANY POINT SOURCES IS NEEDED TO REPRESENT STRINGS IN AURALIZATION?

PACS: 43.55.Ka

Lokki, Tapio[1]
[1]Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory; P.O.Box 5400, FI-02015 TKK, Finland; Tapio.Lokki@tkk.fi

ABSTRACT
An anechoic full symphony orchestra recording is an essential stimulus for auralization studies on concert hall acoustics. Such stimulus material can be achieved by recording instruments one by one in an anechoic chamber. However, for practical reasons the recording of all strings is usually not possible and instead only one or two of each string instrument is recorded. Thus, it raises a question how strings should be represented in auralization so that string sections sound like in a large orchestra. Therefore, auralizations with different number of point sources per each section were made and a listening test was organized to find out perceptual differences. The A/B comparison paradigm was applied and subjects compared differences in perceived number of musicians, spaciousness of auralization, and the overall preference. The results suggest that strings should be modeled with one point source for each musician, but one single recording can be applied in all positions for each section. However, the found differences are quite small and it seems that reasonable auralizations can also be made by representing all strings with only five carefully selected point sources, one for each section.

INTRODUCTION
An anechoic full symphony orchestra recording is an essential stimulus for auralization studies on concert hall acoustics. To enable low noise and good quality soundtracks the instruments should be recorded one by one in an anechoic chamber. The synchronization of players can be achieved by displaying a video of the conductor to the musicians. However, usually all musicians of an orchestra are not available for such recordings and only one or two of each string instrument is recorded instead. In auralization each instrument should be represented with an individual point source, but no studies have been reported on representing string sections if only one or two recordings for each section exist. In this paper such a study is made and listening test results are reported. Different number of point sources per each section is applied in auralization which is performed with the DIVA auralization software [1,2].

PREPARING SAMPLES FOR SUBJECTIVE COMPARISON
In this section the processing of soundtracks for the listening test is explained. First the auralization method is briefly overviewed, then the modification of phases of the stimulus signals are explained, and finally the normalization of sound pressure levels of auralized soundtracks are discussed.

Room acoustics modeling and auralization
Room acoustics modeling and auralization was performed with the DIVA software [1,2]. Direct sounds and early reflections (1st and 2nd orders) were searched with the image-source method [3,4]. Static late reverberation was simulated with a feedback delay network algorithm which produces diffuse reverberation tail [5].

Each direct sound was filtered with distance dependent gain ($1/r$) and air absorption (implemented with a 2nd order IIR filter). It should be noted that sound sources were assumed to be omnidirectional, i.e., no directivity filtering was included in the modeling. In addition, the direction of the source from the listening point of view was implemented with a separated ITD and a minimum-phase HRTF (implemented with a 60 tap FIR filter). All early reflections were processed similarly in addition to material filters (4th order warped IIRs) which model the frequency dependent material absorption on boundaries. The late reverberation algorithm was fed with direct sounds filtered with diffuse field HRTFs, and it produced binaural reverberation with two uncorrelated outputs. The details of the applied signal processing are explained earlier [2].

To prepare soundtracks for the listening test auralizations with three different sound source configurations and three receiver positions were chosen in a concert hall model. Receiver positions were a normal conductor position on stage (r1), an audience seat on the main floor (r2) and another seat back of the hall (r3), see Fig. 1 top row. Source configurations were a single point source, 5 point sources (one for each string instrument), and 44 point sources (full orchestra; 12 1st violins, 10 2nd violins, 8 violas, 8 violoncellos, and 6 double basses), see Fig. 1 bottom row. In case of one position for each string section the source positions were chosen to be spatially far from each other, not as a small quintet around the conductor podium. In full orchestra case the source positions were organized in pairs as strings usually are in a symphony orchestra. Locations in a row had a small variation to prevent possible comb filtering due to equally distant sound sources from the listening position(s). In total, the different source numbers produced different number of 1st and 2nd order reflections as can be seen in Table I.

Table I. Total number of auralized direct sounds (d) and early reflections (r), presented as d + r in each case.

	receiver pos. r1	receiver pos. r2	receiver pos. r3
one sound source	1 + 29	1 + 20	1 + 10
five sound sources	5 + 90	5 + 91	5 + 66
44 sound sources	44 + 892	44 + 972	44 + 621

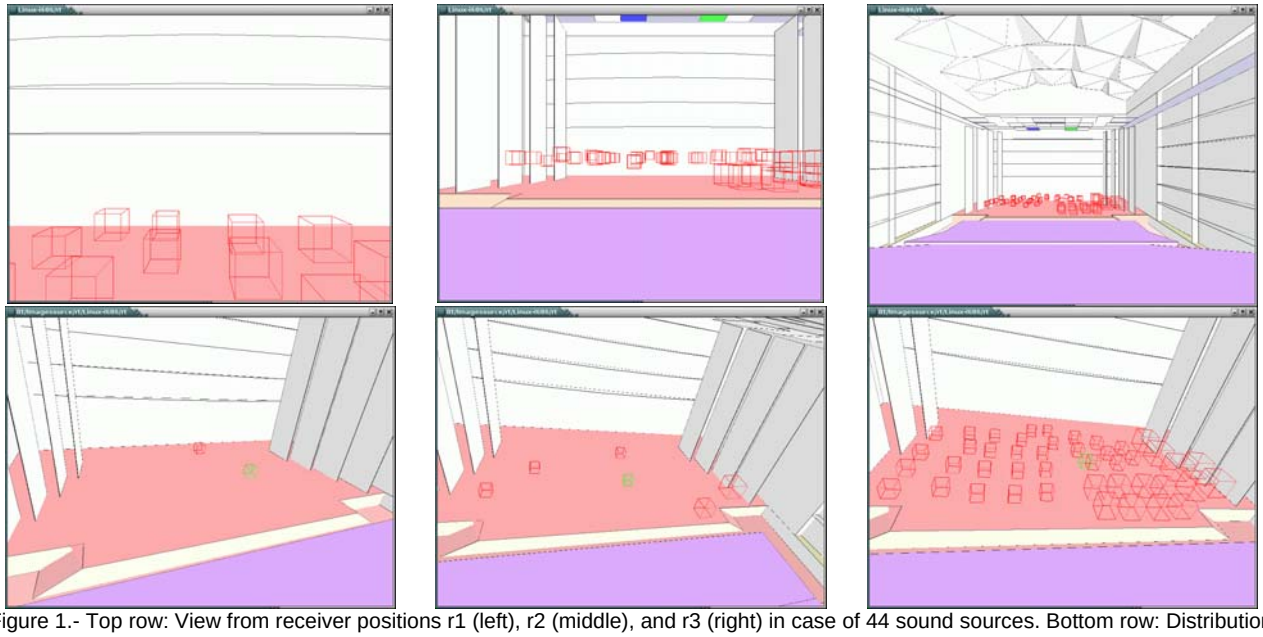

Figure 1.- Top row: View from receiver positions r1 (left), r2 (middle), and r3 (right) in case of 44 sound sources. Bottom row: Distribution of sound sources (red squares) in cases of one source (left), five sources (middle), and 44 sources (right).

Modifying the phase of signals

Only one single recording for each string instrument was available. Therefore, it was assumed that if this recording was applied to all positions inside one section the result does not sound as a large section. Thus, the phase of the signal was modified in half of the cases (see next Section) with Pitch Synchronous OverLap-Add (PSOLA) algorithm [6,7] which is often applied in speech processing. In addition to PSOLA, other ways to scramble the phases of the signals were tried, but no other well-working algorithm was found.

In the PSOLA algorithm the fundamental periods of a signal are first searched and then the signal is decomposed into a series of elementary waveforms which represent pitch periods of the signal. The reconstruction of the signal is performed with the overlap-add sum of the elementary waveforms, and the reconstruction can be done with different fundamental periods. Thus, the PSOLA process change the fundamental frequency of sound signal, but the perceived pitch is not changed. For artifact-free results, PSOLA requires the signal to be harmonic and suitable for decomposition into elementary waveforms, and indeed string sounds seemed to fill these requirements.

The PSOLA processing with different amount of modification to fundamental periods ($0.85 < gamma < 1.15$) were performed with a Matlab implementation by Blanchet [8]. After the PSOLA processing the lengths of the sound signals were not equal. For perfect synchronization the length of the signals was set equal with an algorithm called “Change Tempo without Changing Pitch” programmed by V. Johnson and D. Mazzoni in Audacity audio signal processing program (http://audacity.sourceforge.net).

Normalization of gains

The stimulus signals—20 sec. excerpt of the 3rd movement of Symphony no. 4 by Brahms—were recordings made in the Technical University of Denmark. For 44 source auralizations each source was rendered with one recording or a PSOLA processed versions of it. For 5 source auralizations the stimulus signals were multiplied with the number of musicians in each section (12, 10, 8, 8, and 6) or the PSOLA processed versions were added together respectively. Single point auralizations were rendered with sums of five stimulus signals. Although, conceptually the same number of musicians was applied in each case, the signal levels were not the same due to phase differences in PSOLA processed samples. In addition, different numbers of early reflections results that auralized samples were not on the same level in all cases. Therefore, the gains of the auralized samples were normalized on the same level for the listening test.

SUBJECTIVE LISTENING TEST PRESENTATION

The aim of the subjective test was to find out if people can perceive differences between auralizations of one, five, and 44 point sources. In addition, the need of phase modifications was tested. A paired comparison methodology, also known as A/B comparison with hidden reference, was selected to evaluate the auralizations. Since this paradigm needs a reference for comparison, and no obvious reference exists in this case, it was decided to use auralizations with 5 sources without any phase modifications as references in each listening position. The comparison was performed on the Comparison Category Rating (CCR) scale (from -3.0 to 3.0) [10] and subjects had to compare samples as a function of number of perceived musicians (PLAYERS), perceived spaciousness (SPATIAL), and preference (PREFER), see the user interface in Fig. 2.

In total 16 samples were to be rated against reference, i.e., auralizations with five sources in each listening position, see Table II. Thus, independent variables were three receiver positions (RECEIVER), three source configurations (SOURCES), and two stimulus signals (with or without PSOLA processing). The sample and the hidden reference were played synchronously so that switching between them was possible (cross-fade time was 40 ms). Each subject rated 16 pairs twice and the presentation order of samples was randomized. Before the actual listening test a brief practicing session with five pairs was completed for familiarization to samples as well as to the user interface. The listening test, implemented with the GP2 software [9], was arranged in a quiet office room and Sennheiser HD-590 headphones were applied in sound reproduction.

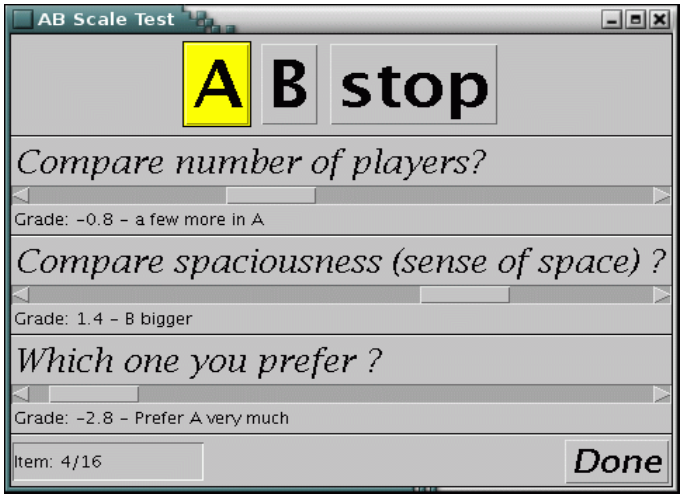

Figure 2.- Graphical user interface applied in the listening test.

RESULTS

Eight non-paid volunteers (one female and seven males, researchers at TKK/TML laboratory, ages 25-55) completed the listening test. The statistical analysis of the results was performed with the SPSS 10.0 software. First it was checked that subjects gave similar ratings on both listening rounds. Thus, dependent variables were tested with independent-samples t-test using grouping variable as repetition. No significant differences were found for PLAYERS ($p=0.962$), SPATIAL ($p=0.413$), and PREFER ($p=0.457$), thus results of both listening rounds were merged together, resulting 16 ratings for each sample pair for each question.

Second, one-sample t-test was applied to compare the mean ratings to the constant value 0 to find out which samples were perceived to differ from the references, i.e., from auralizations with five source positions. For each case (Table II), averages over subjects were tested and significant differences from the references are marked with white background in Table III. It should be noted that someone might have given negative

values and some other subject positive values for the same pair, and such a case does not produce significant difference since the average is close to zero. The means and 95% confidence intervals are presented in Fig. 3.

Table II. Tested 16 samples. The reference was always 05_no (in each listening position r1-r3), thus case 16 was exactly the same as the reference in position r1.

CASE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
REFERENCE	r1_05_no					r2_05_no					r3_05_no					r1_05_no
RECEIVER	r1	r1	r1	r1	r1	r2	r2	r2	r2	r2	r3	r3	r3	r3	r3	r1
SOURCES	44	05	44	01	01	01	44	05	44	01	01	01	44	05	44	05
PSOLA	no	yes	yes	no	yes	yes	no	yes	yes	no	yes	no	no	yes	yes	no

Table III. Probability values from one-sample t-test comparing average ratings to the constant value 0 for each case. White cells represent situations where significant differences exist on 95% level.

CASE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Receiver r1					Receiver r2					Receiver r3					Ref.
PLAYERS	.002	.146	.000	.029	.054	.208	.092	.972	.072	.044	.141	.340	.015	.014	.104	.448
SPATIAL	.042	.063	.000	.456	.954	.773	.437	.589	.039	.424	.458	.076	.000	.168	.002	.740
PREFER	.655	.964	.100	.297	.003	.000	.043	.091	.965	.016	.285	.443	.459	.002	.365	.894

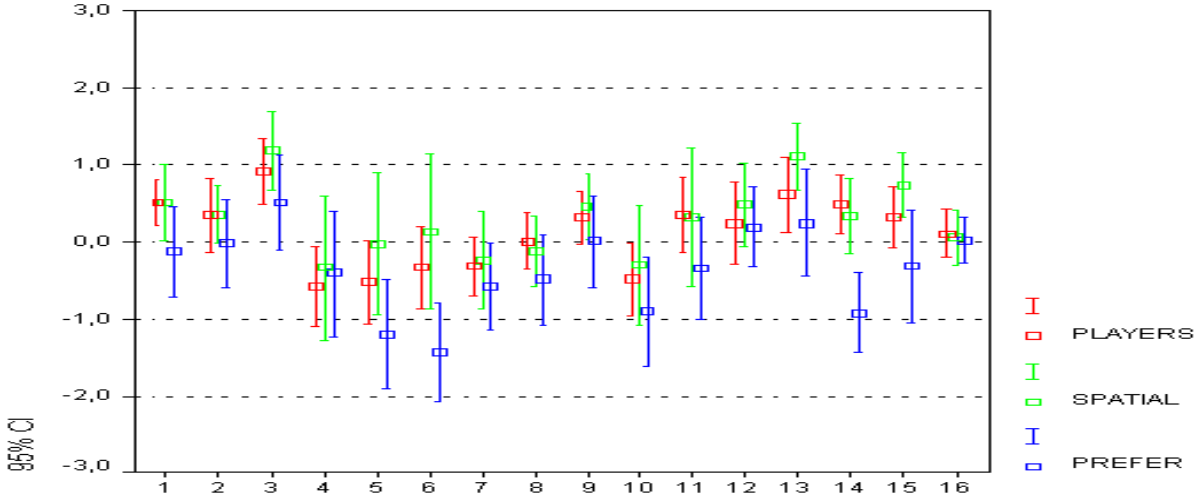

Figure 3.- Means and 95% confidence intervals of all 16 cases.

To find statistically significant differences between independent variables the analysis of variance (ANOVA) was applied (without the case 16; ref-ref). The main effects and some interesting interactions were tested and resulting ANOVA tables for all dependent variables are presented in Fig. 6. All three questions were answered so that main effects for SOURCES, RECEIVER, and SUBJECT were significant on a 95% level). The PSOLA modification to phases of signals did not produce significant perceived differences. The statistically significant interactions were found as SOURCES*RECEIVER for PLAYERS, PSOLA*RECEIVER for SPATIAL, and SOURCES*PSOLA as well as SOURCES*RECEIVER for PREFER.

In this study the biggest interest was in the number of needed sound sources in multi-source auralization, means and confidence intervals for all dependent variables are plotted again in Fig. 4, in order of 01, 05, and 44 point sources applied in auralization. In addition, dependent variable PLAYERS (perceived number of players) in a function of receiver position and number of sources is presented in Fig. 5 (left). Finally, all ratings in a function of number of sound sources are plotted in Fig. 5 (right).

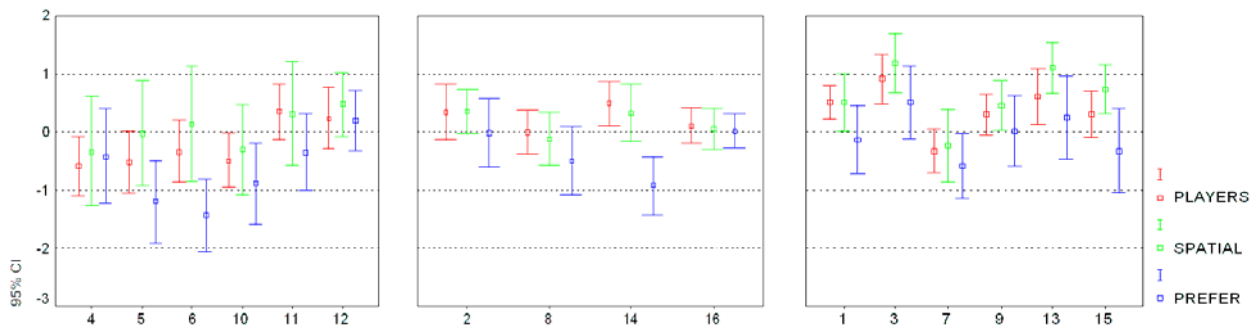

Figure 4.- Means and 95% CIs of all 16 cases ordered by number of sound sources 01 (left), 05 (middle), and 44 (right).

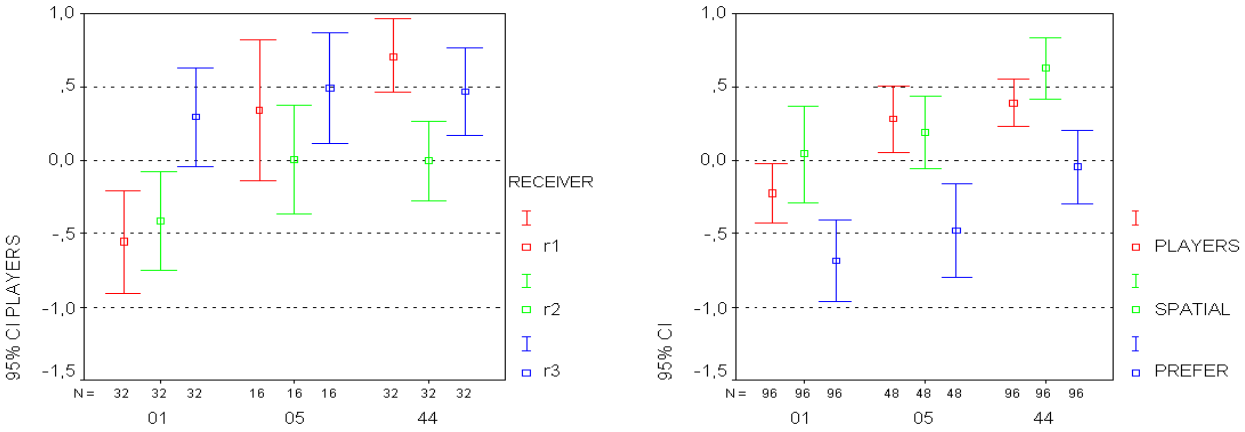

Figure 5.- Perceived number of players in a function of receiver position and number of sources (left). All ratings in a function of number of source positions (note 05 has only PSOLA ratings) (right).

DISCUSSION

Unfortunately this study did not give unambiguous answer to the question asked in the title and other studies are needed in future. However, based on the presented listening test and on the results the following remarks can be made:

- The biggest difference was on SUBJECTS (largest F-statistics for SPATIAL and PREFER), meaning a large variation in ratings between subjects. Many of them also gave verbal feedback that comparison was quite hard and differences between samples were really marginal. This was supported also with a finding that only two subjects found twice the ref-ref pair. On the other hand, this finding could also be interpreted so that other subjects were unreliable.
- In the conductor position (r1) 44 sources seemed to give an impression of larger orchestra than five sources. In the audience positions (r2 and r3) the effect is not so strong, indeed significant difference was found only with one case in position r3. Surprisingly, no significant differences were found between one source and five sources, although five sources were rated higher as seen in Fig. 5.
- The only significant differences on spaciousness were found with 44 sources. Even in the conductor position the single source was not found different than five sources, although the single source cases have largest 95% confidence intervals, meaning large variation in ratings. Some subjects told that they considered spaciousness as “perceived size of space” and others as “spatial distribution of sources”.
- There is large variation in preference ratings. No clear trend is seen, but none of the samples was preferred significantly better or worse than auralization with 5 sources without PSOLA processing.
- The result regarding PSOLA processing is interesting. If dry stimulus signals are added together it sounds like a loud quintet, but with PSOLA processed stimuli like a string orchestra. However, in auralization the spatial distribution of sources together with spatially processed early reflections create signals in which multiple copies of stimuli are added together with different delays, thus the PSOLA processing seems not to be needed. However, PSOLA processing slightly affects to preference ratings in few cases.

Acknowledgement

I like to thank prof. Jens Holger Rindel for giving me the anechoic recordings for stimulus material. In addition, Ms. Michelle Vigeant is thanked for discussions on multi-source auralizations. Finally, I’m grateful to prof. Lauri Savioja for discussions and for help in the implementation details of the DIVA software.

References

1. L. Savioja, J. Huopaniemi, T. Lokki, R. Väänänen, Creating Interactive Virtual Acoustic Environments. *Journal of the Audio Engineering Society (JAES)* **47** (1999) 675-705.
2. T. Lokki, *Physically-based Auralization -- Design, Implementation, and Evaluation*. Doctoral thesis, Helsinki University of Technology, (2002) report TML-A5, available online at http://lib.tkk.fi/Diss/2002/isbn9512261588/
3. J. B. Allen, D. A. Berkley, Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America* **65** (1979) 943-950
4. J. Borish, Extension of the image model to arbitrary polyhedra. *Journal of the Acoustical Society of America* **75** (1984) 1827-1836
5. R. Väänänen, V. Välimäki, J. Huopaniemi, M. Karjalainen, Efficient and Parametric Reverberator for Room Acoustics Modeling. in *Proceedings of the International Computer Music Conference* (ICMC97), Thessaloniki, Greece, September (1997) 200-203.
6. N. Schnell, G. Peeters, S. Lemouton, P. Manoury, X. Rodet, Synthesizing a choir in real-time using Pitch-Synchronous Overlap Add (PSOLA), *Proceedings of the International Computer Music Conference*, Berlin, Germany, September (2000) 102-108.
7. U. Zölzer (ed.), *DAFX – Digital Audio Effects*. John Wiley & Sons Ltd. (2002).
8. G. Blanchet, PSOLA implementation with Matlab, http://www.tsi.enst.fr/~blanchet/livres/PROGS/CHAP10/psola.m (2006).
9. J. Hynninen, N. Zacharov, Guineapig – a generic subjective test system for multichannel audio. *The 106th Audio Enc. Soc. (AES) Convention* (1999) preprint no. 4871.
10. ITU-T. *Recommendation P.800, Methods for Subjective Determination of Transmission Quality*. International Telecommunications Union, Telecommunications Standardization Sector. (1996).

Tests of Between-Subjects Effects

Dependent Variable: PLAYERS

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	3,253	1	3,253	2,262	,174
	Error	10,764	7,486	1,438[a]		
SOURCES	Hypothesis	18,506	2	9,253	13,648	,000
	Error	149,153	220	,678[b]		
PSOLA	Hypothesis	1,470	1	1,470	2,168	,142
	Error	149,153	220	,678[b]		
RECEIVER	Hypothesis	12,990	2	6,495	9,580	,000
	Error	149,153	220	,678[b]		
SUBJECT	Hypothesis	10,465	7	1,495	2,205	,035
	Error	149,153	220	,678[b]		
SOURCES * PSOLA	Hypothesis	,227	1	,227	,335	,564
	Error	149,153	220	,678[b]		
SOURCES * RECEIVER	Hypothesis	10,750	4	2,687	3,964	,004
	Error	149,153	220	,678[b]		
PSOLA * RECEIVER	Hypothesis	1,993	2	,997	1,470	,232
	Error	149,153	220	,678[b]		

a. ,930 MS(SUBJECT) + 7,004E-02 MS(Error)

b. MS(Error)

Tests of Between-Subjects Effects

Dependent Variable: SPATIAL

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	16,455	1	16,455	,953	,361
	Error	121,858	7,055	17,272[a]		
SOURCES	Hypothesis	18,711	2	9,355	9,674	,000
	Error	212,762	220	,967[b]		
PSOLA	Hypothesis	3,360	1	3,360	3,475	,064
	Error	212,762	220	,967[b]		
RECEIVER	Hypothesis	16,604	2	8,302	8,584	,000
	Error	212,762	220	,967[b]		
SUBJECT	Hypothesis	129,500	7	18,500	19,129	,000
	Error	212,762	220	,967[b]		
SOURCES * PSOLA	Hypothesis	,227	1	,227	,235	,629
	Error	212,762	220	,967[b]		
SOURCES * RECEIVER	Hypothesis	6,040	4	1,510	1,561	,186
	Error	212,762	220	,967[b]		
PSOLA * RECEIVER	Hypothesis	6,788	2	3,394	3,510	,032
	Error	212,762	220	,967[b]		

a. ,930 MS(SUBJECT) + 7,004E-02 MS(Error)

b. MS(Error)

Tests of Between-Subjects Effects

Dependent Variable: PREFER

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	31,094	1	31,094	1,849	,216
	Error	118,641	7,054	16,819[a]		
SOURCES	Hypothesis	19,693	2	9,847	10,731	,000
	Error	201,866	220	,918[b]		
PSOLA	Hypothesis	1,944	1	1,944	2,119	,147
	Error	201,866	220	,918[b]		
RECEIVER	Hypothesis	7,912	2	3,956	4,311	,015
	Error	201,866	220	,918[b]		
SUBJECT	Hypothesis	126,120	7	18,017	19,636	,000
	Error	201,866	220	,918[b]		
SOURCES * PSOLA	Hypothesis	8,535	1	8,535	9,301	,003
	Error	201,866	220	,918[b]		
SOURCES * RECEIVER	Hypothesis	14,185	4	3,546	3,865	,005
	Error	201,866	220	,918[b]		
PSOLA * RECEIVER	Hypothesis	3,053	2	1,526	1,663	,192
	Error	201,866	220	,918[b]		

a. ,930 MS(SUBJECT) + 7,004E-02 MS(Error)

b. MS(Error)

Figure 6.- ANOVA tables for all three dependent variables PLAYERS, SPATIAL, and PREFER.