Loudspeaker sound quality: comparison of assessment procedures

Vincent Koehl and Mathieu Paquier
Laboratoire d'Informatique des Systèmes Complexes, LISyC EA 3883
Université de Bretagne Occidentale
6, avenue Victor Le Gorgeu
CS 93837
29238 Brest Cedex 3, France

Introduction

A subjective evaluation of loudspeakers is a difficult and time-consuming task. AES [1] and IEC [2] have both provided recommendations for loudspeakers listening tests, but till now no standardized technique has been adopted.

[1] Audio Engineering Society, "AES20-1996: AES recommended practice for professional audio - Subjective evaluation of loudspeakers (reaffirmed 2007)", *Journal of the Audio Engineering Society 44(5)*, 382-400 (1996).

[2] International Electrotechnical Commission, "Sound system equipment - Part 13: Listening tests on loudspeakers", *IEC Publication 60268-13* (1998).

Goal of the study

A paired comparison is an easy and reliable way to assess loudspeaker sound quality. But laboratory tests are not representative of real-life comparison situations (audiophile, sound engineer…). As the test becomes more controllable and reliable, it differs from realistic conditions for comparisons.

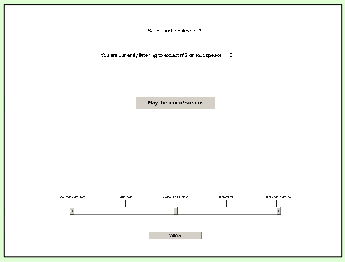
- short stimuli or noise used for comparison are not consistent with real life,
- loudness is equalized to compensate for its effect over preference.

The aim of this study is to see whether realistic comparisons are consistent with laboratory tests. For that purpose, 4 different loudspeakers were evaluated during 3 different tasks (all based on paired comparisons) to investigate the influence of the task over the preference ratings.

Assessment procedures

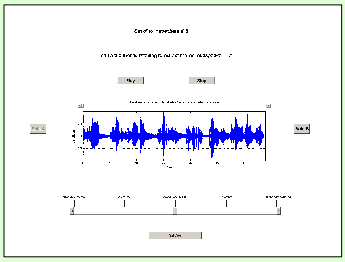
3 different assessment procedures were designed to gradually get close to real-life comparison approach.

Procedure 1



- Short excerpts (≈ 5 s)
- Consecutive presentation (A-B comparison)
- Matched loudness

Procedure 2



- Long excerpts (≈ 30 s)
- Alternate presentation (free switching)
- Matched loudness

These procedures are commonly used for loudspeaker comparison and are well described in the literature.

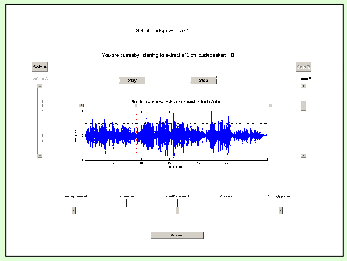
- Short stimuli imply that listeners make their comparisons on the same basis.
- Long stimuli can be listened to in a way that is closer to the real life.

For both procedures, loudness was equalized by ear by 3 expert listeners.

In real-life comparison situation, the listener is free to set the volume to his convenience.

A realistic comparison approach would be :
- Long music excerpts
- No loudness matching

Procedure 3



- Long excerpts (≈ 30 s)
- Alternate presentation (free switching)
- Reproduction level up to the listener (± 6 dB around the matched loudness)

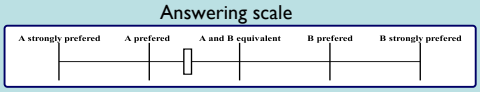
Experimental setup

48 listeners were involved in the experiment, each test consisted in 3 sessions (1 session per procedure) carried out consecutively, but in random order.

1 session = 18 trials (Paired comparisons)
- 6 pairs: $\frac{N(N-1)}{2} = 6$
- 3 excerpts:
  - Leonard Bernstein, "West Side Story", symphonic orchestra,
  - Ben Harper, "I want to be ready", male voice and acoustic guitar,
  - George Gershwin, "Rhapsody in blue", piano solo.

For each stimulus pair, the subject was asked to give his preference on a continuous scale.

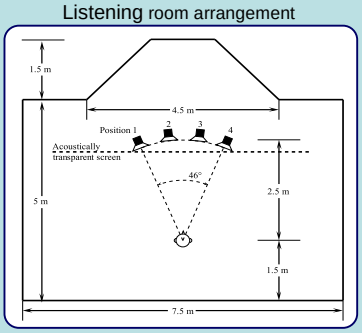
Answering scale



- Subjects could directly listen to sound radiations by loudspeakers.
- Loudspeakers were presented in monophonic reproduction.

Loudspeaker positions were exchanged throughout the experiment to compensate for the positional influence over the preference ratings:
- 24 possible combinations,
- 2 listeners per configuration.

Listening room arrangement

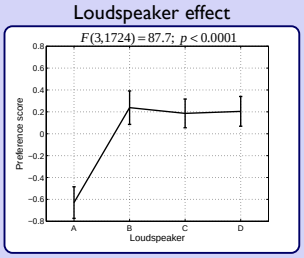


Results

- Preference was measured within [-1;1], P_{ij} stands for the preference of stimulus i over stimulus j.
- A negative preference P_{ij} means that the loudspeaker j was preferred to i.

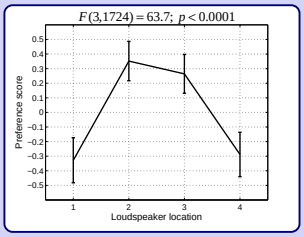
Preference score for loudspeaker i could be obtained by using : $S_i = \sum_{j \neq i} P_{ij}$

Loudspeaker effect



- Loudspeakers B, C and D were systematically preferred to A.
- No significant difference between B, C and D.

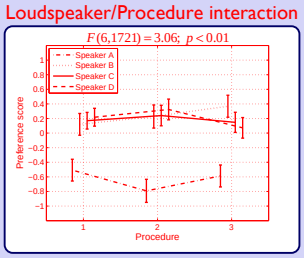
Position effect



- Loudspeakers placed in front of the listener (positions 2 and 3) obtained significantly better scores than the ones on the sides (positions 1 and 4).

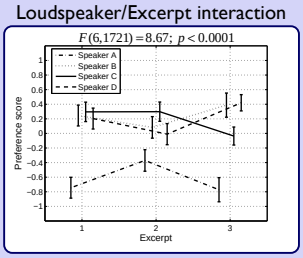
Could be due to:
- Room excitations (depending on the position),
- Loudspeaker identification (easier, when it is set laterally),
- Head movement (towards the apparent source direction).

Loudspeaker/Procedure interaction



- Whatever the procedure, loudspeakers B,C and D were significantly preferred to A.
- With procedures 1 and 2, loudspeakers B, C and D were not statistically different.
- With procedure 3, B was significantly better rated than C ($p < 0.05$) and D ($p < 0.001$).

Loudspeaker/Excerpt interaction



- Excerpt 1 was poorly discriminating: only loudspeaker A was significantly different from the other ones.
- Excerpts 2 and 3 were much more discriminating: all loudspeaker pairs were significantly different, except for the B-D pair.

Conclusion

As shown by Loudspeaker/Procedure interaction, the preference score for one loudspeaker depends on the comparison procedure. The loudspeaker scores are roughly consistent from a procedure to another one, but all procedures have not the same discriminating power.

- When the listeners were given the possibility of setting the reproduction level, their assessments tended to be more discriminating.

The different tests showed significant effects of the excerpt and the loudspeaker position, in agreement with previous studies.

Contact : vincent.koehl@univ-brest.fr or mathieu.paquier@univ-brest.fr