GUIDELINES FOR LISTENING TEST

This listening test aims to rate the quality of a set of signals produced by source separation systems. Source separation aims to extract the signal of a target source from a mixture of several sound sources. The resulting signals may include several types of degradations compared to the clean target source, including distortions of the target source, remaining sounds from other sources and additional artificial noise.

During the test, you will be asked to address *four successive tasks*:

1. rate the global quality compared to the reference for each test signal,

2. rate the quality in terms of preservation of the target source in each test signal,

3. rate the quality in terms of suppression of other sources in each test signal,

4. rate the quality in terms of absence of additional artificial noise in each test signal.

For each task, the test will involve a *training phase* and an *evaluation phase*.

During the training phase, you will have to listen to all the sounds to

- train yourself address the required task and learn the range of observed quality according to that task,

- set the volume of your headphones so that it's comfortable but you can clearly hear differences between sounds (the volume can't be changed later on).

The evaluation phase involves ten experiments. In each experiment, you will have to rate the quality of eight test sounds compared to a reference sound (clean target) on a scale from 0 to 100, where larger ratings indicate better quality. You can listen to the sounds as many times as you want before deciding. You should make sure that

- the ratings between pairs of sounds are consistent, *i.e.* if one sound has better quality than another, it should be rated better,

- the ratings between different experiments are consistent, *i.e.* if two sounds from different experiments have the same quality, they should be rated equally,

- the whole rating scale is used, *i.e.* sounds with perfect quality should be rated 100 and the worst test sound over all experiments (but not necessarily the worst test sound in each experiment) should be rated 0.

The expected total duration of the test is 1 hour and 40 minutes, that is 25 minutes per task. It is recommend that you make a long break of at least 25 minutes between each task (but *don't close the GUI*). You can also make additional short breaks within each task.