**Source: VoiceAge Corporation**

**Title: IVAS Characterization Test Plan - Proposals for Stereo and ISM experiments**

**Agenda Item: 7.6**

**Document for: Discussion and Agreement**

1. Introduction

The present document addresses few open issues in the description of the detail experiment setup for Stereo and ISM experiments in the IVAS Characterization Test Plan [1]. Further, some corrections are proposed to remove unsupported configurations.

1. Specification of MNRU and ESDRU values

The MNRU and ESDRU values were set as follows in the IVAS Selection test [2]:

* Stereo clean speech experiment
	+ MNRU values: Q = 16, 20, 24, 28 dB
	+ ESDRU values: α = 0.1, 0.4, 0.7
* Stereo noisy speech experiment
	+ MNRU values: Q = 12, 17, 22, 27, 32 dB
	+ ESDRU values: α = 0.1, 0.3, 0.5, 0.7
* Stereo mixed content and music experiment
	+ MNRU values: Q = 12, 17, 22, 27 dB
	+ ESDRU values: α = 0.1, 0.4, 0.7
* ISM experiments (1- and 2-ISM clean speech experiments)
	+ MNRU values: Q = 15, 23, 31, 39, 47 dB
	+ ESDRU values: α = 0.1, 0.3, 0.5, 0.7

The current version 0.3.0 of the IVAS Characterization Test Plan considers 4 values for MNRUs and 4 values for ESDRUs (This is the case for all experiments if we consider the Alternative 1 for Experiments P800-2 and P800-11. For Alternative 2 only three values of ESDRUs are currently proposed). Based on the Selection test result [3], and considering the conditions defined in the Characterization Test Plan, we propose the following MNRU and ESDRU values (assuming Alternatives 1).

**Proposal:**

* Stereo experiments
	+ MNRU values: Q = 12, 17, 22, 27 dB
	+ ESDRU values: α = 0.1, 0.3, 0.5, 0.7

The following values are proposed for preliminaries:

Preliminaries for Experiment P800-1 and -2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** | **DTX** |
| 1 | c19 |  | IVAS FL | 13.2 | 5% | on |
| 2 | c26 |  | IVAS FL | 96 | 5% | on |
| 3 | c06 |  | ESDRU $α$ = 0.7 | - | - |  |
| 4 | c20 |  | IVAS FL | 16.4 | 5% | on |
| 5 | c09 |  | ESDRU $α$ = 0.1 | - | - |  |
| 6 | c22 |  | IVAS FL | 32 | 5% | on |
| 7 | c03 |  | MNRU Q = 27 dB | - | - |  |
| 8 | c01 |  | Reference | - | - |  |
| 9 | c27 |  | IVAS FL | 128 | 5% | on |
| 10 | c07 |  | ESDRU $α$ = 0.5 | - | - |  |
| 11 | c05 |  | MNRU Q = 12 dB | - | - |  |
| 12 | c24 |  | IVAS FL | 64 | 5% | on |

* ISM experiments
	+ MNRU values: Q = 18, 24, 30, 36 dB
	+ ESDRU values: α = 0.1, 0.3, 0.5, 0.7

The following values are proposed for preliminaries:

Preliminaries for Experiment P800-9, -10, and -11

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trial #** | **Label** | **Sample** | **Condition** | **Bitrate** | **FER/Profile** |
| 1 | c19 |  | IVAS FL | 13.2 | No errors |
| 2 | c26 |  | IVAS FL | 96 | No errors |
| 3 | c06 |  | ESDRU $α$ = 0.7 | - | - |
| 4 | c20 |  | IVAS FL | 16.4 | No errors |
| 5 | c09 |  | ESDRU $α$ = 0.1 | - | - |
| 6 | c22 |  | IVAS FL | 32 | No errors |
| 7 | c03 |  | MNRU Q = 36 dB | - | - |
| 8 | c01 |  | Reference | - | - |
| 9 | c27 |  | IVAS FL | 128 | No errors |
| 10 | c07 |  | ESDRU $α$ = 0.5 | - | - |
| 11 | c05 |  | MNRU Q = 18 dB | - | - |
| 12 | c24 |  | IVAS FL | 64 | No errors |

1. Unsupported bitrates in ISM experiments

The bitrate of 13.2 kbps is not supported for more than one object, and the bitrate 16.4 kbps is not supported for more than 2 objects. Consequently, it is proposed to modify the range of bitrates such that all bitrates are supported for all categories within an experiment. Note that the following proposal considers the Alternative 1 of Exp P800-11. If Alternative 2 is chosen, similarly, the bitrates of 13.2 and 16.4 kbps shall be replaced with supported bitrates.

**Proposal:**

* Bitrates of Experiment P800-9 (1 and 2 ISMs): 16.4, 24.4, 32, 48, 64, 80, 96, 128, 160 kbps. Accordingly, increase all bitrates in Tables F.9.2 and F.9.3 to the next higher bitrate.
* Bitrates of Experiments P800-10 (3 and 4 ISMs) and P800-11 (1, 2, 3 and 4 ISMs): 24.4, 32, 48, 64, 80, 96, 128, 160, 192 kbps. Accordingly, increase all bitrates in Tables F.10.2 and F.10.3, and F.11.2 and F.11.3, to the 2nd higher bitrate.
1. Balancing scenes distribution per listening panels in ISM experiments

In the current version of the Test Plan, the distribution of six spatial scenes over listening panels is not balanced for model-based categories in ISM experiments. The distribution of the scenes was copied from the Selection Test Plan where we had one P.SUPPL800 experiment dedicated to 1-ISM samples and another experiment dedicated to 2-ISM samples. Hence, the audio material in each listening panel covered all the scenes (6 scenes, 6 listening panels).

For the Characterization Test, it has been decided to combine different number of ISMs per experiment. Consequently, we have only two model-based 1-ISM categories and two model-based 2-ISM categories in Experiments P800-9 and P.800-11, but still considering 6 different spatial scenes for 1-ISM audio samples and 6 different spatial scenes for 2-ISM samples. It means that the audio material in each listening panel can comprise only 2 different spatial scenes for 1-ISM audio samples, and 2 different spatial scenes for 2-ISM samples, out of 6. If care is not taken, different listening panels might listen to significantly more or less complex audio scenes.

As an example, in the current setup, the listening panel P1 of Experiment P800-9 would listen to static model-based audio scenes only, while for the listening panel P4 all model-based audio scenes would be dynamic. Similarly for Experiment P800-11.

We thus proposal a different distribution of the spatial scenes for model-based categories, such that all listening panels listen to 1 static and 1 dynamic audio scene.

**Proposal for experiment P800-9 and -11:**

Table F.x.4: Allocation of scenes for each talker pair (category cat 1 – cat 4) and listening panel (P1-P6)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Category***  | ***Scene*** | ***Talker initial elevation*** | ***Elevation change(1*** | ***Talker initial azimuth*** | ***Azimuth change(2*** | ***Panel*** |
| ***cat 1:****M1*  | abefcd | 0°35°-90°35°0°35° | staticstatic0.3°/ frame-0.2°/ frame staticstatic | 270°180°120°0°240°180° | staticstaticstatic0.5°/ frame1°/ frame-1°/ frame | P1P2P3P4P5P6 |
| ***cat 2:****F1* | fcdabe | 35°0°35°0°35°-90° | -0.2°/ framestaticstaticstaticstatic0.3°/ frame | 300°60°120°60°300°60° | 0.5°/ frame1°/ frame-1°/ framestaticstaticstatic | P1P2P3P4P5P6 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category***  | ***Scene*** | ***Overtalk******[s]* (1** | ***1st talker elevation*** | ***2nd talker elevation*** | ***1st talker initial azimuth*** | ***1st talker azimuth change(2*** | ***2nd talker initial azimuth*** | ***2nd talker azimuth change(2*** | ***Panel*** |
| ***cat 4:****M2 + F2* | *a**b**c**d**e**f* | *-1**1**-1**1**-1**1* | *0°**35°**0°**0°**45°**30°* | *0°**35°**45°**45°**45°**30°* | *0°**10°**20°**200°**340°**120°* | *static**static**static**static**-1°/ frame**1°/ frame* | *50°**110°**170°**30°**340°**120°* | *static**static**static**-1°/ frame**-1°/ frame**-1°/ frame* | *P1**P2**P3**P4**P5**P6* |
| ***cat 5:****M3 + F3* | *d**e**f**a**b**c* | *1**-1**1**-1**1**-1* | *0°**45°**30°**0°**35°**0°* | *45°**45°**30°**0°**35°**45°* | *50°**130°**300°**30°**40°**50°* | *static**1°/ frame**1°/ frame**static**static**static* | *180°**130°**300°**230°**290°**350°* | *1°/ frame**1°/ frame**-1°/ frame**static**static**static* | *P1**P2**P3**P4**P5**P6* |

1. Defining categories 3 and 6 for Experiment P800-9 (1 and 2 ISMs)

The categories 3 and 6 for 1-ISM Experiment are currently not defined. The source sees two alternatives.

1st alternative would define categories 3 and 6 as pre-produced content, e.g. Mixed content & Music category for 1-ISM (category 3), and speech plus some background (music, effects, …) for 2-ISM (category 6).

Second alternative would define categories 3 and 6 similarly as categories 4 and 5. This would mean having two 1-ISM categories (M1, F1) and four 2-ISM categories (M2-F2, M3-F3, F2-M3, F3-M2). The reasoning behind this configuration is that it would make all male and female talkers tested equally in the sense that for 1-ISM categories sentence pairs are used while for 2-ISM categories each sample is constructed from one single female-talker sentence and one single male-talker sentence).

**Proposal - Alternative 1:**

Table F.9.5: Mixed content and Generic audio categories

|  |  |
| --- | --- |
| **Category**  | **Type** |
| cat 3 | Mixed content & Music (1 object) |
| cat 6 | 2 ISMs: Speech + background (2 objects) |

**Proposal - Alternative 2:**

Table F.9.5: Allocation of scenes for each talker pair (category cat 3 – cat 6) and listening panel (P1-P6)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Category***  | ***Scene*** | ***Overtalk******[s]* (1** | ***1st talker elevation*** | ***2nd talker elevation*** | ***1st talker initial azimuth*** | ***1st talker azimuth change(2*** | ***2nd talker initial azimuth*** | ***2nd talker azimuth change(2*** | ***Panel*** |
| ***cat 3:****M2 + F2* | *a**b**c**d**e**f* | *-1**1**-1**1**-1**1* | *0°**35°**0°**0°**45°**30°* | *0°**35°**45°**45°**45°**30°* | *0°**10°**20°**200°**340°**120°* | *static**static**static**static**-1°/ frame**1°/ frame* | *50°**110°**170°**30°**340°**120°* | *static**static**static**-1°/ frame**-1°/ frame**-1°/ frame* | *P1**P2**P3**P4**P5**P6* |
| ***cat 4:****M3 + F3* | *d**e**f**a**b**c* | *1**-1**1**-1**1**-1* | *0°**45°**30°**0°**35°**0°* | *45°**45°**30°**0°**35°**45°* | *50°**130°**300°**30°**40°**50°* | *static**1°/ frame**1°/ frame**static**static**static* | *180°**130°**300°**230°**290°**350°* | *1°/ frame**1°/ frame**-1°/ frame**static**static**static* | *P1**P2**P3**P4**P5**P6* |
| ***cat 5:****F2 + M3* | *b**c**d**e**f**a* | *1**-1**1**-1**1**-1* | *35°**0°**0°**45°**30°**0°* | *35°**45°**45°**45°**30°**0°* | *20°**30°**250°**290°**180°**10°* | *static**static**static**-1°/ frame**1°/ frame**static* | *170°**230°**340°**290°**180°**110°* | *static**static**-1°/ frame**-1°/ frame**-1°/ frame**static* | *P1**P2**P3**P4**P5**P6* |
| ***cat 6:****F3 + M2* | *e**f**a**b**c**d* | *-1**1**-1**1**-1**1* | *45°**30°**0°**35°**0°**0°* | *45°**30°**0°**35°**45°**45°* | *80°**0°**40°**50°**0°**100°* | *1°/ frame**1°/ frame**static**static**static**static* | *80°**0°**290°**350°**50°**130°* | *1°/ frame**-1°/ frame**static**static**static**1°/ frame* | *P1**P2**P3**P4**P5**P6* |

**References**

1. S4-240770 - IVAS Permanent Document IVAS-8b: Test Plan for Characterization Phase, v.0.3.0.
2. S4aA230086 - IVAS Permanent Document IVAS-8a: Test Plan for Selection, Phase v.1.1.0.
3. S4-231573 - Global Analysis Laboratory report – IVAS Selection Phase. Source: HEAD acoustics GmbH.