1 Experiment on using Stereo Sound for an Audio Keyboard.

1.1 Goal of experiment

Previous studies have shown that stereo sound can be used to expedite certain audio tasks. In most cases it has been found that one can not use more than three audio channels. The goal of this experiment is to establish if one can communicate using more than 3 audio channels, by quantitatively comparing the results from 3, 4, and 5 audio channels. In our case, only the alphabet is read in different audio channels (not words), and the same letters are always read in the same position. For example, *a* will always be read in the right ear, whereas *c* will always be read in the left ear. The difference between our application of stereo sound and previous studies is that, in our case, the user always knows what will happen in each channel, making it easier to focus on the channel where the appropriate letter occurs. We hope to show that participants can communicate using 5 audio channels. In a follow-up experiment we will then optimise the interface to achieve optimal performance.

1.2 Experimental setup

The user performance on 3, 4, and 5 channels is tested. The order in which the channels are tested is randomised, in order to make a fair comparison. For example:

- Participant 1: Channel 3, 4, 5
- Participant 2: Channel 5, 3, 4

The experiment is configured as followed:

- Tutorial sessions (28 minutes).
- Nasa work load factors (18 minutes).
- Breaks (10 minutes).
- Testing (63 minutes).
- Questions (1 minute).

The speed at which the alphabet is read to the user is incrementally increased while testing a specific channel. These letter speeds are always tested in the same order, namely slow, medium, and then fast. The experimental setup for each channel is summarised as follows:

- **Tutorial** (5-15 minutes):
  - The system is explained to the user (5 minutes).
  - The user gets a chance to listen to the alphabet and to get used to the configuration (with visual assistance). (3-5 minutes).
  - The user is asked to select 2-3 letters from each audio channel (3-5 minutes).

- The first tutorial session takes approximately 15 minutes whereas the other two take approximately 7 minutes each.
- **Channel test** (7-8 minutes per speed setting:) The user selects letters displayed to him/her.
- **Nasa forms** (2 minutes): The user evaluates the cognitive demand of the task.
- **Break** (5 minutes): After all three speed settings have been tested the user gets a 5 minute break.

1.3 Results for each participant

User	Channel	Speed	Theoretical speed estimate(wpm)	Measured speed (wpm)	Total number of words	Number of correct words	Number of words in top three list	Number of erroneous word selections	Number of time-out errors	Min edit distance	Average number of clicks per character
1	3	Slow	0.96	0.93	4	3	4	1	0	0.50	2.12
1	3	Medium	1.07	1.12	8	7	8	1	0	0.12	1.90
1	3	Fast	1.33	1.52	8	7	8	0	1	0.00	2.45
1	4	Slow	0.96	1.02	10	8	10	2	0	0.70	1.94
1	4	Medium	1.07	1.23	6	5	6	0	1	0.00	2.56
1	4	Fast	1.33	1.37	5	3	5	1	1	0.75	3.42
1	5	Slow	0.96	1.03	6	5	6	0	1	0.00	2.64
1	5	Medium	1.07	1.27	7	6	6	0	1	0.00	2.25
1	5	Fast	1.33	1.46	10	6	9	2	2	1.75	3.48
2	3	Slow	0.96	1.05	8	7	8	1	0	0.38	1.84
2	3	Medium	1.07	1.09	9	7	9	1	1	0.12	2.56
2	3	Fast	1.33	1.65	7	6	7	0	1	0.00	2.24
2	4	Slow	0.96	1.12	6	3	6	2	1	1.80	3.44
2	4	Medium	1.07	1.34	4	2	4	0	2	0.00	6.70
2	4	Fast	1.33	1.15	7	7	7	0	0	0.00	2.37
2	5	Slow	0.96	1.15	11	7	9	2	2	1.89	2.80
2	5	Medium	1.07	1.27	6	5	5	0	1	0.00	2.60
2	5	Fast	1.33	1.53	7	5	7	0	2	0.00	3.52
3	3	Slow	0.96	0.96	8	8	8	0	0	0.00	1.97
3	3	Medium	1.07	1.05	5	5	5	0	0	0.00	1.92
3	3	Fast	1.33	1.30	8	8	8	0	0	0.00	1.96
3	4	Slow	0.96	0.94	6	5	6	0	1	0.00	2.94
3	4	Medium	1.07	1.31	7	5	7	0	2	0.00	3.28
3	4	Fast	1.33	1.36	8	6	8	1	1	0.14	2.71
3	5	Slow	0.96	0.96	7	7	7	0	0	0.00	1.90
3	5	Medium	1.07	1.10	6	5	6	1	0	0.17	1.89
3	5	Fast	1.33	1.40	6	4	5	1	1	0.20	3.25
4	3	Slow	0.96	0.96	6	5	6	1	0	0.17	1.97
4	3	Medium	1.07	1.21	6	6	6	0	0	0.00	1.65
4	3	Fast	1.33	1.78	8	4	8	2	2	0.50	3.37
4	4	Slow	0.96	1.14	4	4	4	0	0	0.00	1.66
4	4	Medium	1.07	1.01	4	4	4	0	0	0.00	1.89
4	4	Fast	1.33	1.21	6	4	6	1	1	1.40	3.38
4	5	Slow	0.96	0.89	6	5	6	1	0	0.17	1.99
4	5	Medium	1.07	1.10	9	7	9	1	1	1.25	2.46
4	5	Fast	1.33	1.19	8	8	8	0	0	0.00	2.38
5	3	Slow	0.96	1.00	7	7	7	0	0	0.00	1.82
5	3	Medium	1.07	1.16	7	7	7	0	0	0.00	1.77
5	3	Fast	1.33	1.20	7	6	7	1	0	0.57	2.12
5	4	Slow	0.96	1.16	5	3	5	1	1	0.25	2.73
5	4	Medium	1.07	1.01	6	6	6	0	0	0.00	2.07
5	4	Fast	1.33	1.04	6	6	6	0	0	0.00	2.81
5	5	Slow	0.96	1.33	4	4	4	0	0	0.00	1.43
5	5	Medium	1.07	1.47	6	5	6	1	0	0.83	1.53
5	5	Fast	1.33	1.26	6	3	6	1	2	0.25	4.78

1.4 Results, averaged over all users

Channel	Speed	Theoretical speed estimate(wpm)	Measured speed (wpm)	Total number of words (written in 7-8 minutes)	Average % correct words	Average % words in top three list	Average % erroneous word selections	Average % timeout errors	Min edit distance	Average number of clicks per character
3	Slow	0.96	0.98	6.60	90.91	100.00	9.09	0.00	0.21	1.94
3	Medium	1.07	1.13	7.00	91.43	100.00	5.71	2.86	0.05	1.96
3	Fast	1.33	1.49	7.60	81.58	100.00	7.89	10.53	0.21	2.43
4	Slow	0.96	1.08	6.20	74.19	100.00	16.13	9.68	0.55	2.54
4	Medium	1.07	1.18	5.40	81.48	100.00	0.00	18.52	0.00	3.30
4	Fast	1.33	1.22	6.40	81.25	100.00	9.38	9.38	0.46	2.94
5	Slow	0.96	1.07	6.80	82.35	94.12	8.82	8.82	0.41	2.15
5	Medium	1.07	1.24	6.80	82.35	94.12	8.82	8.82	0.45	2.15
5	Fast	1.33	1.37	7.40	70.27	94.59	10.81	18.92	0.44	3.48

- **User**: Participant number.
- **Channel**: The maximum number of simultaneous voices reading the alphabet.
- **Speed**: Three speeds were tested for each channel (how fast the alphabet is being read).
- **Theoretical speed estimate(wpm)**: Theoretical approximation of how many words can be written per minute. Here it was assumed that each letter is selected once in a word consisting of 5 characters, and that the alphabet was read twice for each click. This measurement is based only on how long it takes to read the alphabet (i.e., synthesised voice commands such as ”next”, and system delays to save data have been omitted).
- **Measured speed (wpm)**: The speed that was measured during the experiment. This measurement is based only on how many times the alphabet was read to make a specific word selection. This value can be slightly more than the theoretical value if the alphabet was read more than twice before a click was made, or if the user had to repeat a word (due to erroneous clicks, or due to the system that could not distinguish between similar words like *go* and *to*).
- **Total number of words**: The total number of words written.
- **Number of correct words**: The total number of words ($\leq$ total number of words) that were correct after a word selection was made by the system.
- **Number of words in top three list**: The system selects a word automatically if its probability is above 0.9. However, in some cases, especially when a timeout error occurs, the correct word was in the top three list of most probable words. This can happen especially when the word is short, e.g., when the system can not decide between *go* and *to*, and happens more frequently in four and five channels, as more sounds are played simultaneously. Thus, we also measure the total number of times the ground-truth word was one of the top three in the dictionary, when the system had to make a decision (a word selection / time out error) to quantify performance.
- **Number of erroneous word selections**: The total number of words that were wrong when the system made a decision.
- **Number of timeout errors**: The number of times the user could not get to the correct word after repeating the word twice.
- **Min edit distance**: The average number of operations needed to map the selected word onto the correct word after a word selection was made (ie, this calculation does not include timeout errors).
- **Average number of clicks per character**: The average number of clicks per character - this value is only computed for correct word selections, and should be less if more work is done on the randomisation of the alphabet and the click distribution.

2 Possible problems

- The randomised sequence of the channel 3 configuration happens to be not so random. May let channel 3 look better compared to the others. The randomisation was done automatically beforehand.
- More work needs to be done on the randomisation of channel 5 - this may let this channel look worse than the others. If we use the top three criterion this is not so much of a problem.
- People seem to struggle to hear Sad Sam (comment from all users so far) - this may let channel 4 look worse compared to the others.