Bayesian models for audiovisual integration in speech perception

02458: Cognitive Modeling
Fall 2021

Anna Rifé Mata - s212487
Evangelia Athanasiou - s205726
Lluís Colomer Coll - s213237

November 28, 2021

1 Statement of Contribution

This report was based on the article: **The early maximum likelihood estimation model of audiovisual integration in speech perception**[1].

Questions	**Written by**	**Reviewed by**
Part1.Q2, Part1.Q3, Part2.Q3, Part3.Q4, Part3.Q4, Part4.Q1	Lluís	All
Part1.Q4, Part2.Q4, Part2.Q5, Part3.Q1, Part4.Q3, Part4.Q4	Evangelia	All
Part1.Q1, Part2.Q1, Part2.Q2, Part3.Q2, Part4.Q2, Part4.Q4	Anna	All

Contents

2 FLMP - a discrete Bayesian model of audiovisual integration

Question 1: The exact formula for the cost function you used to fit the FLMP to the data.

All training data have been used to fit the FLMP model by the use of the maximum likelihood method. The aim was to select the model that makes the observed data more likely. As there were only two response categories, the data follows a binomial distribution. Taking that into account, we have computed the binomial probability density function at each of the values in our data, using the corresponding number of trials and the probability of success for each trial, as presented below.

$$P(k) = \binom{N}{k} \cdot p^k q^{N-k} \tag{1}$$

where:

N : Number of trials

k : Number of successes desired

p : probability of getting a success in one trial

q : probability of getting a failure in one trial

To obtain the probability of success p for each value, 10 free parameters were used: 5 of them corresponding to the theta values for an audio stimulus (θ_A) and the other 5 for a visual stimulus (θ_V). Then, the visual and auditory probabilities were computed using the following soft-max functions.

$$p_a(\theta_a) = \frac{e_a^\theta}{e_a^\theta + 1}, \quad p_v(\theta_v) = \frac{e_v^\theta}{e_v^\theta + 1} \tag{2}$$

With the p_a and p_v, the audiovisual probabilities were computed as a discrete Bayesian model of audiovisual integration (FLMP) (Equation 3).

$$p_{av}(\theta_a, \theta_v) = \frac{p_a(\theta_a) \cdot p_v(\theta_v)}{p_a(\theta_a) \cdot p_v(\theta_v) + (1 - p_a(\theta_a)) \cdot (1 - p_v(\theta_v))} \tag{3}$$

To that end, the cost function was aimed to optimize the 10 free parameters to have the maximum likelihood between the predicted binomial probability and our observed data. However, the cost function was not maximizing the likelihood but instead minimizing the negative logarithm of the likelihood (Equation 4) as being an equivalent approach.

$$f(\theta_a, \theta_v) = -\log \prod_{a=1}^{r} \prod_{v=1}^{c} \binom{N}{k} \cdot p_{av}^k(\theta_a, \theta_v) \cdot q_{av}^{N-k}(\theta_a, \theta_v) \tag{4}$$

where p_{av}^k corresponds to a 7x5 matrix where the visual, auditory and audiovisual response probabilities were stored and q_{av}^k=1-p_{av}^k. N equals to 24, being the total number of possible responses, r=5 and c=7 (corresponding to p_{av}^k matrix size, respectively).

Question 2: A table with parameter values for each subject.

With the aim to find the 10 free parameters (θ_a,θ_v values) that give the lowest negative logarithm of the likelihood, the MATLAB `fminunc(fun,θ)` function [2] has been used. The input "fun" corresponds to the cost function described in equation 4 which basically attempts to find a local minimum given by θ values. The parameters obtained for each subject are presented below.

Parameters	θ_a	θ_v
S_1	-3.31, -2.29, -1.52, 2.46, 3.21	-2.97, 0.84, 2.91, 3.58, 3.83
S_2	-3.68, -1.83, -0.32, 2.58, 4.83	-5.29, 0.91, 3.20, 5.09, 5.87
S_3	-2.65, -1.74, -0.69, 2.62, 5.41	-2.00, 0.25, 1.54, 2.37, 2.62
S_4	-4.12, -3.11, -2.04, 2.25, 4.62	-2.52, -1.19, -0.31, 1.26, 2.12
S_5	-5.43, -3.93, -1.42, 2.92, 6.24	-3.00, 0.42, 1.73, 2.42, 3.05

Table 1: θ values obtained for each subject when fitting the FLMP model to the original data. S_1, S_2, S_3, S_4, S_5 correspond to each subject, respectively.

Question 3: A table with the negative log likelihoods for the fit to the data from each subject.

The negative log likelihood values obtained when fitting the data for each subject are presented below, which lie in the range of 32.98 to 53.38. The values were computed by equation 4, using the optimized θ values obtained (Table 1).

	S_1	S_2	S_3	S_4	S_5
Negative log likelihood	47.34	32.99	53.38	46.91	42.81

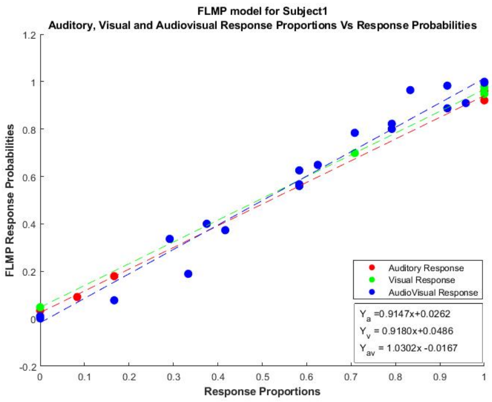
Table 2: Negative log likelihood values when fitting the data to the FLMP model for each corresponding subject.

Question 4: A scatter plot of the FLMP response probabilities vs. the response proportions for each subject. Estimate whether the model fits are good from the plots.

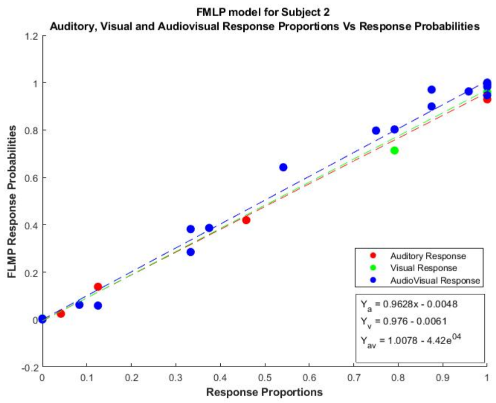
The following scatter plots in Figure 1, illustrate the Auditory, Visual and Audiovisual response proportions vs the response probabilities obtained by the FMLP model for each subject separately .

In order to linearly fit the data and estimate if the model was good, we used the `fitlm` function in MATLAB [3] which returns a linear regression model of the responses Y (Response Probabilities), for each stimulus, that fit to the data matrix X (Response Proportions). Provided that we obtained an intercept of 0-value and a slope of 1, $y = mx + n$ ($m = 1$, and $n = 0$), the model fits would exactly predict the original data.

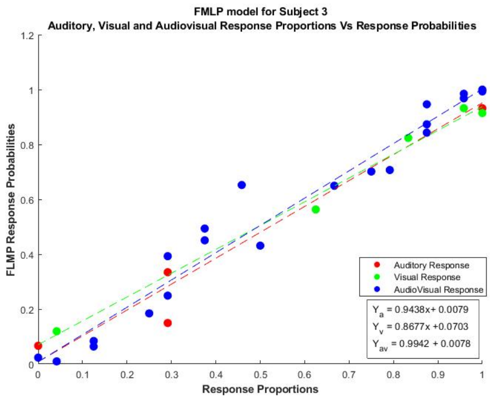
In Figure 1, for each subject, the linear regression is displayed in the bottom right part of each scatter plot. Additionally, the Root Mean Squared Error (RMSE) was also obtained for each subject and stimulus (Table 3). This statistic measurement is also known as the fit standard error and the standard error of the regression. The more the RMSE is close to 0, the better the fit.



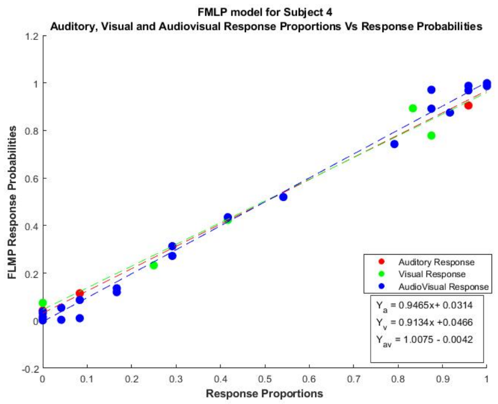
a) FLMP response probability vs response proportions for Subject 1.



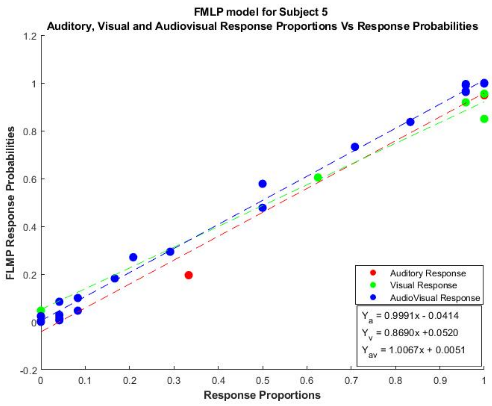
b)FLMP response probability vs response proportions for Subject 2.



c) FLMP response probability vs response proportions for Subject 3.



d) FLMP response probability vs response proportions for Subject 4.



e) FLMP response probability vs response proportions for Subject 5.

Figure 1: Scatter plots of the FLMP response probabilities vs. the response proportions for auditory, visual and audiovisual responses, including the linear regression to fit the data.

RMSE	S_1	S_2	S_3	S_4	S_5	mean(S_1...S_5)
A	0.0184	0.0310	0.0935	0.0221	0.0669	0.04638
V	0.0132	0.0381	0.0410	0.0693	0.0500	0.04214
AV	0.0533	0.0386	0.0665	0.0355	0.0280	0.04438
mean(A,V,AV)	0.0283	0.0359	0.0670	0.0423	0.0483	

Table 3: RMSE values obtained from the linear regression model for each subject (S_1, S_2, S_3, S_4, S_5) and stimulus (A:auditory, V:visual, AV: audiovisual.)

All RMSE values are smaller than 0.07, which means that the FLMP model can predict the data accurately. When considering the type of stimulus separately, the same amount of error was obtained independently of each type.

3 The MLE strong fusion model

Question 1: The exact formula for the cost function you used to fit the early MLE to the data.

As in the FLMP model, the maximum likelihood method was used to fit the MLE model that makes the observed data more likely, but this time considering a different approach to compute the response probabilities and its respective cost function.

The MLE model assumes that the perception of an stimulus, $P(S)$, causes an internal representation value x, influenced by Gaussian noise. In that way, repeated internal representations of auditory, visual and audiovisual stimulus can be assumed to be distributed following the density functions of normal distributions presented below:

$$P(S_a|x_a) = \mathcal{N}(\mu_a, \sigma_a^2) \tag{5}$$

$$P(S_v|x_v) = \mathcal{N}(\mu_v, \sigma_v^2) \tag{6}$$

$$P(S_{av|}x_{av}) = \mathcal{N}(\mu_{av}, \sigma_{av}^2) \tag{7}$$

As before, all the training data has been used, computing the binomial probability density function at each of the values in our data, and using the corresponding number of trials and the probability of success for each trial, as presented in Equation 8.

$$P(k) = \binom{N}{k} \cdot p^k q^{N-k} \tag{8}$$

where:

N : Number of trials

k : Number of successes desired

p : probability of getting a success in one trial

q : probability of getting a failure in one trial

To obtain the probability of success p for each value, 4 free parameters were used: σ_a and c_a (for the auditory stimulus) and σ_v and c_v (for the visual stimulus). c_a and c_v were used to determine the means of the five auditory and visual distributions, and consequentially p_a and p_v were computed, following the normal distribution function (Equation 10 11).

$$\mu_a = x - c_a, \quad \mu_v = x - c_v, \quad x = [1, 2, 3, 4, 5] \tag{9}$$

$$p_a(\sigma_a, c_a) = \frac{1}{\sigma_a\sqrt{2\pi}} \exp -\frac{(\mu_a - x)^2}{2\sigma_a^2} \tag{10}$$

$$p_v(\sigma_v, c_v) = \frac{1}{\sigma_v\sqrt{2\pi}} \exp -\frac{(\mu_v - x)^2}{2\sigma_v^2} \tag{11}$$

With the p_a and p_v computed, the early MLE model assumes conditional independence of visual and auditory internal representations, given the audiovisual stimulus S_{av}. Therefore, the maximum likelihood estimate of the corresponding audiovisual distribution is the normalized product of the auditory and visual probability densities. This product is also a Gaussian distribution with a mean, μ_{av}, which is a weighted sum of the means, μ_a and μ_v, of its respective distributions.

$$\mu_{av}(\sigma_a, \sigma_v, c_a, c_v) = \frac{\sigma_v^2}{\sigma_a^2 + \sigma_v^2}\mu_a + \frac{\sigma_a^2}{\sigma_a^2 + \sigma_v^2}\mu_v, \quad \sigma_{av}(\sigma_a, \sigma_v) = \sqrt{\frac{\sigma_a^2\sigma_v^2}{\sigma_a^2 + \sigma_v^2}} \tag{12}$$

Then, the probability density for audiovisual responses was computed:

$$p_{av}(\sigma_a, \sigma_v, c_a, c_v) = \frac{1}{\sigma_{av}(\sigma_a, \sigma_v)\sqrt{2\pi}} \exp -\frac{(\mu_{av}(\sigma_a, \sigma_v, c_a, c_v) - x)^2}{2\sigma_{av}^2(\sigma_a, \sigma_v)} \tag{13}$$

Taking all into account, the cost function was aimed to optimize the 4 free parameters to minimize the negative logarithm of the likelihood between the predicted binomial probability and our observed data (Equation 14).

$$f(\sigma_a, \sigma_v, c_a, c_v) = -\log \prod_{a=1}^{r} \prod_{v=1}^{c} \binom{N}{k} \cdot p_{av}^k(\sigma_a, \sigma_v, c_a, c_v) \cdot q_{av}^{N-k}(\sigma_a, \sigma_v, c_a, c_v) \tag{14}$$

where p_{av}^k corresponds to a 7x5 matrix where the visual, auditory and audiovisual response probabilities were stored and q_{av}^k=1-p_{av}^k. N equals to 24, being the total number of possible responses, r=5 and c=7 (corresponding to p_{av}^k matrix size, respectively).

Question 2: A table with parameter values for each subject.

The MATLAB `fminunc(fun,` $[\sigma_a, \sigma_v, c_a c_v]$`)` function [2] has been used. The input "fun" corresponds to the cost function described in equation 14 which basically attempts to find a local minimum given by $\sigma_a, \sigma_v, c_a c_v$ values.

Parameters	σ_a	σ_v	c_a	c_v
S_1	0.9030	0.8752	3.3117	1.8577
S_2	0.7610	0.5867	3.0206	1.9844
S_3	0.9086	1.1339	2.9025	2.0476
S_4	0.6954	1.1386	3.3580	3.1066
S_5	0.5869	0.9621	3.3245	2.0718

Table 4: $\sigma_a, \sigma_v, c_a c_v$ values obtained for each subject when fitting the early MLE model to the original data. S_1, S_2, S_3, S_4, S_5 correspond to each subject, respectively.

Question 3: A table with the negative log likelihoods for the fit to the data from each subject.

The negative log likelihood values obtained when fitting the data for each subject are presented below, which lie in the range of 53.54 - 72.80. The values were computed by equation 14, using the optimized $[\sigma_a, \sigma_v, c_a c_v]$ values obtained (Table 4).

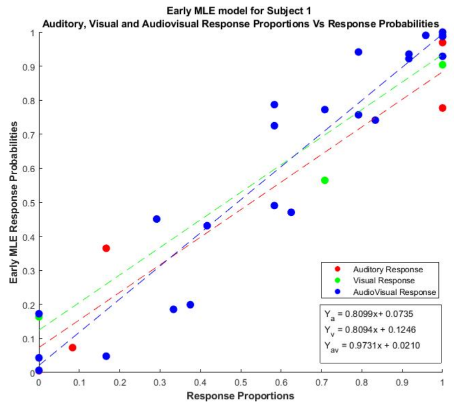
	S_1	S_2	S_3	S_4	S_5
Negative log likelihood	72.80	55.08	69.41	67.21	53.55

Table 5: Negative log likelihood values when fitting the data to the MLE model for each corresponding subject.

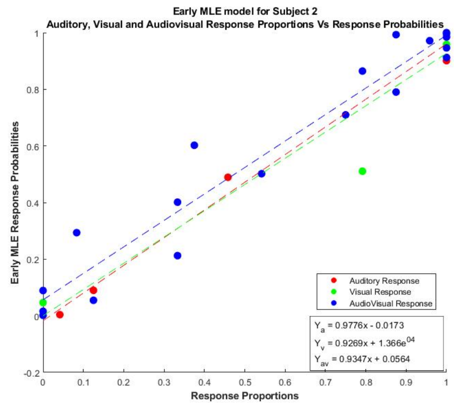
From Table 5, the minimum log value can be seen to belong to subject 2, which means that MLE model better fits for this subject.

Question 4: A scatter plot of the early MLE response probabilities vs. the response proportions for each subject. Estimate whether the model fits are good from the plots.

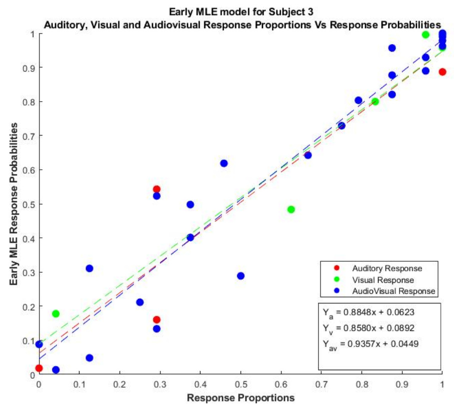
The following scatter plots in Figure 4, shows the Auditory, Visual and Audiovisual response proportions vs the response probabilities obtained by the MLE model for each subjects separately. By using the same linear regression function as in Section 2 (MATLAB `fitlm` function [3]), the linear fit for each subject was obtained.



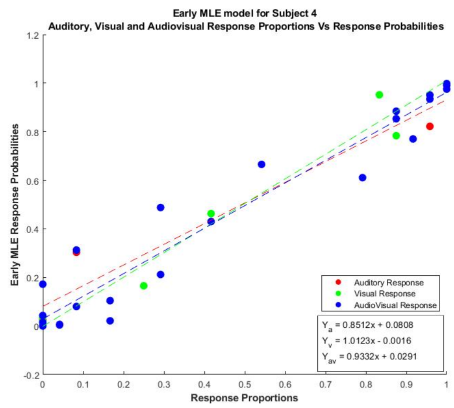
(a)Early MLE Response probabilities vs response proportions for Subject 1.



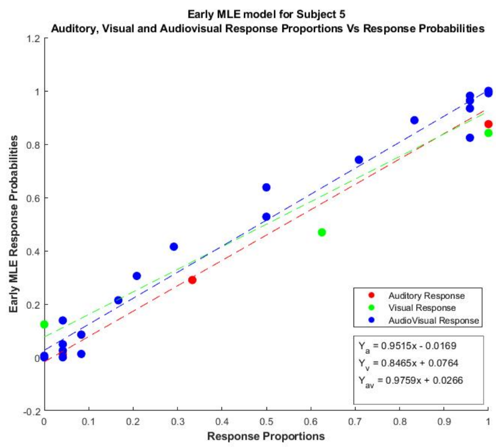
(b)Early MLE Response probabilities vs response proportions for Subject 2.



(c)Early MLE Response probabilities vs response proportions for Subject 3.



(d)Early MLE Response probabilities vs response proportions Subject 4.



(e)Early MLE Response probabilities vs response proportions for Subject 5.

Figure 4: Scatter plots of the MLE response probabilities vs. the response proportions for auditory, visual and audiovisual responses, including the linear regression to fit the data.

By looking at the linear fitting on our data and compared to the FLMP model, the early MLE present some outliers. For a better understanding, the RMSE was also computed for each subject with its respective mean values.

RMSE	S1	S2	S3	S4	S5	mean ($S_1 ... S_5$)
A	0.1322	0.0541	0.1658	0.1177	0.0518	0.1043
V	0.0966	0.1456	0.1003	0.1041	0.1183	0.1130
AV	0.1060	0.0806	0.0978	0.0956	0.0586	0.0877
mean(A,V,AV)	0.1116	0.0934	0.1213	0.1058	0.0762	

Table 6: RMSE values obtained from the linear regression model for each subject (S_1, S_2, S_3, S_4, S_5) and stimulus (A:auditory, V:visual, AV:audiovisual).

As shown in Table 6, the best model was obtained with the audiovisual stimuli, giving us the less RMSE value. However, the mean of the auditory and visual RMSE are twice the RMSE for the FLMP model. With these results in mind, the FLMP model can be conclude to fit better our data.

Question 5: A comparison of the model fits from the early MLE and the FLMP models based on the models' negative log likelihoods. Does any model fit better for any/all subjects/stimuli?

The negative log likelihoods and the RMSE were used to compare FLMP and MLE models.

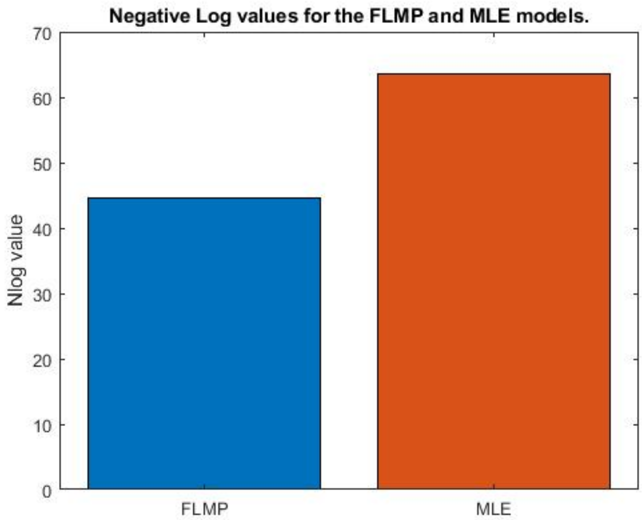


Figure 5: Mean Negative Log Likelihood values for all the subjects of the FLMP and MLE models.

Considering the mean negative log likelihood (Figure 5), the lower value was seen with the FLMP model. This leads to the conclusion that the FLMP model fits better our data than the MLE one.

On the other hand, Figure 6 shows the RMSE for the FLMP and Early MLE models. Again, the FLMP model has performed better than the MLE. Additionally, RMSE differences were observed for the different type of stimulus. For example, for the FLMP model, the Visual stimuli has slightly less RMSE compared to the other stimulus which means that it better fits the visual data. Then, the early MLE model has shown a better performance for the Audiovisual stimuli.

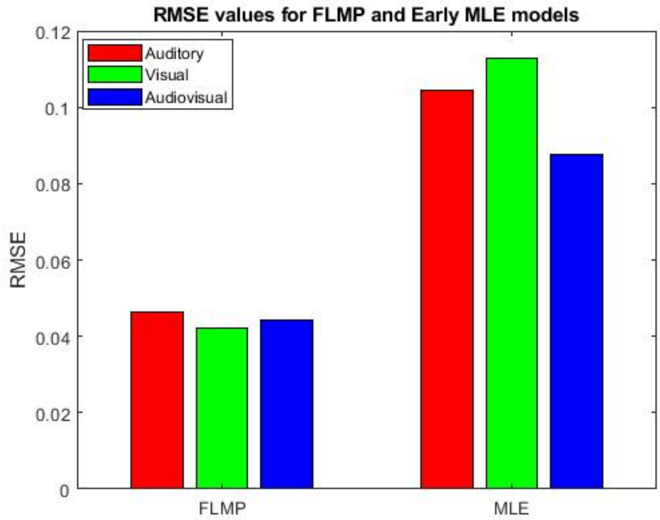


Figure 6: RMSE for the FLMP and Early MLE models for auditory, visual and audiovisual stimuli.

4 Simulation

Question 1: Create a table with the parameters that you used to simulate data for each of the two models. Select one FLMP subject and one early MLE subject for further analysis.

FLMP model parameters	θ_a S_1	θ_v S_1	θ_a S_2	θ_v S_2	θ_a S_3	θ_v S_3	θ_a S_4	θ_v S_4	θ_a S_5	θ_v S_5
	-4.6436	-4.9847	-4.4268	-5.3334	-3.7351	-5.2081	-5.2110	-5.4215	-4.5966	-1.8582
	-0.9377	-4.2514	-3.1232	-4.9028	-3.0708	-3.0127	-0.4656	-2.1795	-3.7915	-0.9705
	-0.7513	-2.7642	-2.0183	-4.5966	1.8079	-1.8768	0.7006	-0.3550	-3.5985	0.5947
	-0.6723	3.7003	-1.4575	-3.2766	2.5540	2.3608	4.1048	3.5400	2.7467	3.8820
	5.7005	5.5395	2.8576	4.6927	5.4961	4.7496	5.7721	4.4266	3.6517	5.3946

Table 7: Parameters used to simulate data for the FLMP model.

MLE model parameters	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
σ_a	0.6164	0.7221	0.9389	0.9389	0.9410
σ_v	0.5594	0.9139	0.6166	0.6166	0.8768
c_a	4.0226	3.3869	4.2833	4.2833	3.1703
c_v	2.2387	2.1338	2.1378	2.1378	2.0435

Table 8: Parameters used to simulate data for the MLE model.

The parameters for the FMLP model were five θ_a values (Auditory stimulus) and five θ_v ones (Visual stimulus), as shown in Table 7. Table 8 shows the parameters for the MLE model, which were: σ_a, σ_v, c_a and c_v.

For further analysis, one subject was selected for each model based on the lowest negative log likelihood between the simulated data and the predicted response probabilities (Table 9). The simulated subject 1 was selected for the FLMP model and simulated subject 2 for the MLE.

Model/ Log values	S1	S2	S3	S4	S5
FMLP model	18.2665	25.1392	36.2273	34.0940	36.1196
MLE model	40.6257	39.2519	40.2059	40.2059	45.5047

Table 9: Negative Log Likelihood values when fitting the simulated data to the FLMP and MLE models.

Question 2: Create a table with the simulated data for the selected FLMP subject and the early MLE subject.

FMLP/S_1 simulated data						MLE/S_2 simulated data				
0	6	9	3	24		0	0	9	18	24
0	0	0	24	24		3	7	20	24	24
0	0	0	0	15		0	0	3	11	23
0	0	0	0	19		0	0	6	15	23
0	0	1	0	23		0	3	12	23	24
9	23	24	23	24		2	9	20	23	24
17	24	24	24	24		7	15	23	24	24

Table 10: Simulated subjects based on the FLPM and MLE model.

Question 3: The negative log likelihood for the FLMP and the early MLE for the selected FLMP subject and the selected early MLE subject.

As presented in Table 9, the negative log likelihood for the selected simulated subject using the FLMP model is 18.2665. Then, the selected simulated subject for MLE has a negative log likelihood of 39.2519. Those were the lowest negative log likelihood values obtained after obtaining the model that made our simulated data more likely.

Question 4: For each subject use the models' errors to select the model that fit better. Produce a confusion matrix in which the row indicate the true model (that you used to simulate data) and the column indicate the model that fit better. The elements of the confusion matrix is the number of observations. Does this method of model selection work well based on your confusion matrix?

In order to observe how good the performance of the models was in generating the data, a confusion matrix was computed. The rows represent the true models used to simulate the data, whereas the columns show the model that fits better. For each subject, the data generated from FLMP and MLE models was fitted to both models. Thus, the model that gave the lowest negative log likelihood was considered the best fit. This procedure was repeated for every model and every subject, obtaining at the end the confusion matrix shown below.

true model\best fit	FLMP	MLE
FLMP	5	0
MLE	0	5

Table 11: Confusion Matrix between the true model and the one that fit better.

From the confusion matrix (Table 11), it can be concluded that a perfect performance was obtained. Thus, for every FLMP model, the lowest error and therefore the best fit was obtained when fitting the generated data using the FLMP model. The same happened with the MLE model. This evidences the good performance of both models when using sample data, and confirms that the model error is a good criterion for model selection.

5 Cross Validation

Question 1: Create a table with the training error for each fold for each subject for each model.

	Training Error FLMP model				
Fold n°	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
1	46.4296	31.8661	51.5977	46.4975	42.7074
2	45.9480	32.8620	51.1037	44.6689	41.3596
3	47.2967	32.9853	51.6192	46.8771	42.8089
4	44.3966	30.4632	51.6478	46.7879	40.6666
5	45.4753	30.7818	51.4525	46.6078	42.1563
6	45.4991	31.3200	51.0666	45.8713	41.0444
7	45.2950	30.8712	50.0568	44.6863	41.1695
8	46.0648	31.5309	49.4676	45.7596	41.5707
9	45.5987	30.6879	51.2491	45.1806	40.9674
10	47.2159	32.9689	52.7840	44.2498	42.7906
11	43.0754	30.9042	51.2037	46.5720	41.7666
12	45.4849	30.9240	51.0575	46.0591	41.4899
13	44.8725	32.0005	47.7924	45.2003	41.2323
14	45.5824	32.5194	50.9486	45.1161	41.0807
15	45.7774	30.9860	51.4929	45.4779	39.0120
16	45.9492	31.9387	51.8283	45.0914	37.5589
17	47.0636	32.8980	51.0888	42.7631	41.1740
18	45.3886	29.9370	50.6857	45.8947	40.6824
19	45.7068	31.4603	51.4573	45.6422	40.2497
20	45.9179	32.7755	51.8084	45.1200	41.0018
21	45.7955	32.8926	51.9515	45.0296	41.2700
22	44.7563	30.2066	51.3767	45.3202	41.2429
23	46.6482	32.8364	50.9751	44.2267	41.4379
24	45.0337	31.0753	51.5551	45.0372	40.7939
25	40.5426	26.7474	50.6441	44.9256	41.8094
26	47.2277	32.9129	51.9809	45.3147	40.9885
27	47.2840	32.9771	53.2144	42.7073	40.3651
28	47.2965	32.9832	53.2519	45.3579	42.7517
29	46.2353	32.7797	53.2672	46.6639	42.7659
30	46.8005	32.9197	52.2270	44.7989	41.5794
31	45.3524	30.8771	52.1340	45.3679	41.7372
32	43.9196	32.9086	53.2939	45.8765	42.7829
33	47.2886	32.9805	53.3583	46.5681	42.8058
34	47.3147	32.9872	53.3715	46.8410	42.8100
35	47.3205	32.9878	53.3737	46.8806	42.8119

Table 12: Training Error for each fold and subject for the FLMP model.

	Training Error MLE model				
Fold n°	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
1	72.6728	52.3883	68.9570	67.2042	53.5456
2	67.7116	53.8105	65.7961	66.3562	49.5534
3	72.6427	55.0372	67.9321	67.2112	53.5456
4	68.5086	52.5967	66.7568	67.2045	48.4155
5	68.6275	53.0006	65.3101	67.1732	53.4861
6	70.4764	52.7982	64.7104	65.1081	49.8220
7	67.5152	45.6082	66.6369	59.1659	52.5494
8	71.5165	53.4237	66.0660	66.5683	51.9615
9	69.6586	47.9811	66.2998	64.7563	49.8768
10	71.6885	54.6774	67.0612	64.7128	53.4046
11	69.3553	52.2133	67.6146	66.8435	52.4554
12	69.7496	53.2779	67.5822	66.1134	52.2821
13	70.7702	54.0239	66.1760	65.1746	51.8045
14	67.4500	55.0619	67.7774	64.9376	50.7318
15	68.2690	53.1836	63.8643	61.9819	51.6434
16	70.1358	53.9768	67.7599	65.2607	48.6296
17	67.8953	52.6578	65.3812	65.9471	51.1205
18	69.6871	53.1695	64.8252	62.4088	51.2781
19	71.0909	52.7549	67.6979	62.0163	51.6808
20	71.4680	54.9924	67.9835	63.3535	51.7367
21	71.0615	55.0751	66.4947	64.4110	51.5677
22	65.6276	52.1288	66.2265	63.8725	49.9532
23	72.6242	55.0686	68.2979	64.9112	52.2636
24	70.9774	49.9466	67.6074	65.3899	50.4367
25	70.6162	52.9766	67.6880	63.5428	50.3059
26	70.9584	54.6888	68.2396	63.9979	52.4273
27	72.5291	55.0703	68.8720	65.7719	52.2780
28	72.7789	55.0757	69.2898	66.2175	53.4522
29	71.9946	54.9605	69.1525	66.9870	53.4941
30	72.7976	55.0757	66.9551	62.4475	53.5275
31	69.5025	49.3901	67.5649	65.7223	52.5576
32	71.5328	53.6352	68.4393	66.0891	53.3262
33	72.4897	55.0383	69.1676	66.5920	53.5077
34	72.7748	55.0755	69.3653	67.0241	53.5414
35	72.8003	55.0757	69.4064	67.1647	53.5460

Table 13: Training Error for each fold and subject for the MLE model.

Question 2: Create a table with the total test error for each subject for each model.

Table 14 shows the total test errors or also called generalization errors for the FLMP model and MLE models, for each subject. This test error is presented in terms of the negative log likelihood between the true data and the predicted one from the model.

	Total test errors FLMP-model	Total test errors MLE-model
Subject 1	66.6946	89.2968
Subject 2	382.0695	73.0106
Subject 3	53.3816	80.7347
Subject 4	57.3857	79.6511
Subject 5	46.9090	53.5465

Table 14: Generalization Error when testing the model by cross-validation.

The test error values obtained from both models are within the expected range. The test errors for the FLMP lie in the interval 57.3 - 398.0, while the test errors for the early MLE were in the interval 64.8-89.2. Based on the value range we have obtained for the models, we observed that the Early MLE model has the lowest validation error. This leads to conclude that Early MLE could be a promising approach of audiovisual integration of speech when taking into account random effects in the data.

Question 3: A comparison of the model fits from the early MLE and the FLMP models based on the models' test errors. Does any model fit better for any/all subjects/stimuli?

Based on the models' test error in Table 14, the FLMP model had lowest values except for Subject 2, revealing some variability across subjects. The validation errors using the cross-validation approach were, in general, larger than without using it (Table 2 - 5), as it was expected.

Additionally, the RMSE was used to establish which model fits better based on auditory, visual and audiovisual stimuli, being correlated to the negative log likelihood.

RMSE (FLMP)	S_1	S_2	S_3	S_4	S_5	mean (S_1...S_5)
A	0.0354	0.0794	0.1218	0.0298	0.0990	0.0731
V	0.0176	0.0707	0.0701	0.1146	0.0990	0.0744
AV	0.0947	0.1477	0.1221	0.0567	0.0679	0.0978
mean(A,V,AV)	0.0492	0.0907	0.1046	0.0669	0.0886	

Table 15: RMSE values obtained from the linear regression model for each subject (S_1, S_2, S_3, S_4, S_5) and stimulus (A:auditory, V:visual, AV:audiovisual) for the FLMP model.

From the values in Table 15, it can be concluded that the performance for the FLMP model significantly varies between subjects and stimulus. Subject 4 (S_4) shows the lowest RMSE error for the auditory and audiovisual stimulus, while Subject 1 (S_1) shows the lowest error for the visual stimulus. Interestingly, when integrating visual and auditory internal representation, the RMSE of the audiovisual stimulus slightly increased.

RMSE (MLE)	S_1	S_2	S_3	S_4	S_5	mean (S_1...S_5)
A	0.1796	0.0749	0.2031	0.1409	0.0644	0.1325
V	0.1308	0.1994	0.1346	0.1335	0.1783	0.1553
AV	0.1238	0.1039	0.1124	0.1071	0.0720	0.0994
mean(A,V,AV)	0.1447	0.1260	0.1500	0.1271	0.1049	

Table 16: RMSE values obtained from the linear regression model for each subject (S_1, S_2, S_3, S_4, S_5) and stimulus (A:auditory, V:visual, AV:audiovisual) for the MLE model.

Table 16 is also showing variability between subjects and stimulus for the MLE model. Again, Subject 4 (S_4) shows the lowest RMSE error for all stimuli. In that way, MLE fit better for subject 4.

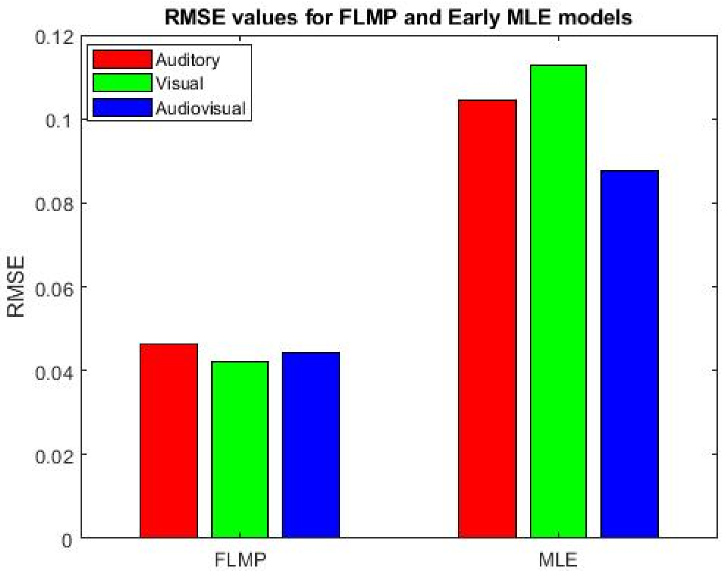
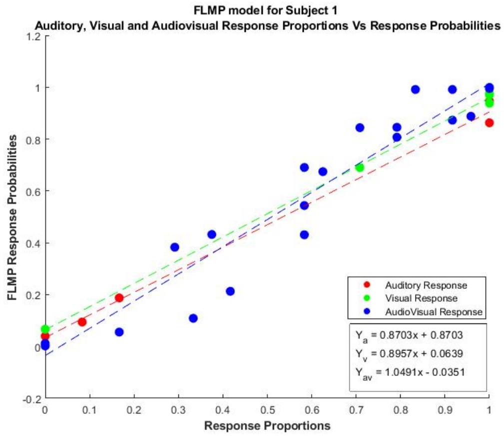


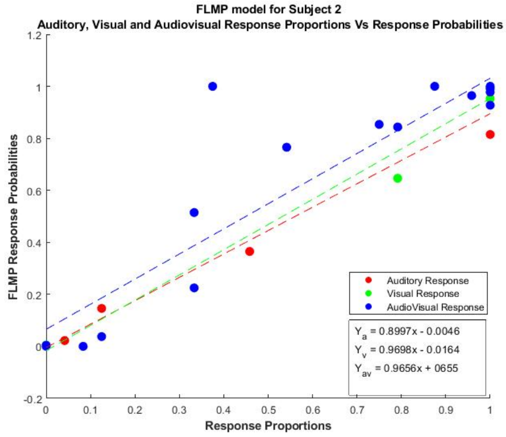
Figure 7: RMSE for the FLMP and Early MLE models for auditory, visual and audiovisual stimuli with cross validation.

Comparing the RMSE of both models, the best performance was seen with FLMP by predicting visual and auditory stimulus. When taking the audiovisual integration into account, the RMSE of both models were in the same range. Overall, the cross-validation has showed to effectively be both goodness-of-fit and model flexibility in model evaluation, and provides meaningful selection of models.

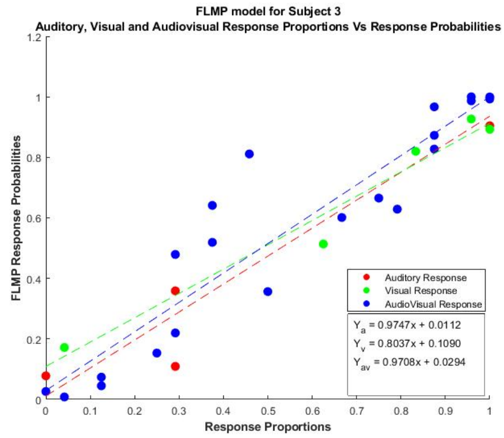
Question 4: A scatter plot of the predicted FLMP, and Early MLE response probabilities vs. the response proportions for each subject. Estimate whether the model fits are good from the plots.



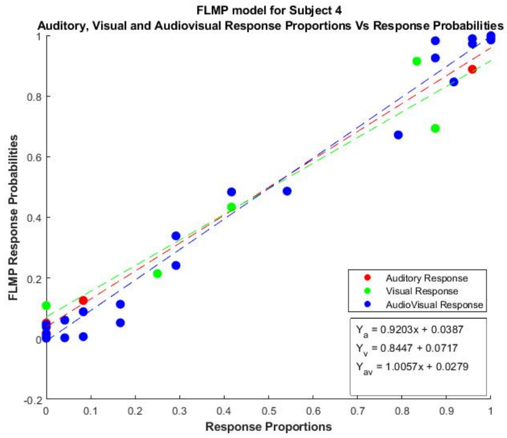
a) FLMP response probability vs response proportions for Subject 1.



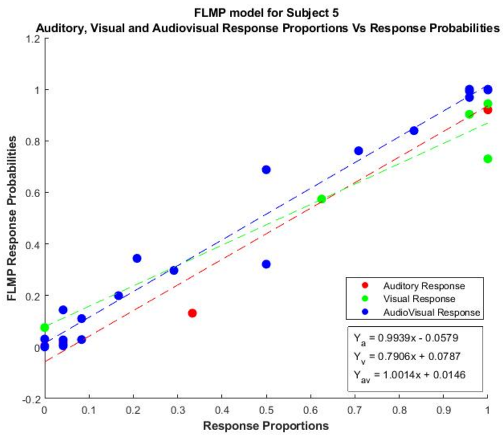
b)FLMP response probability vs response proportions for Subject 2.



c) FLMP response probability vs response proportions for Subject 3.

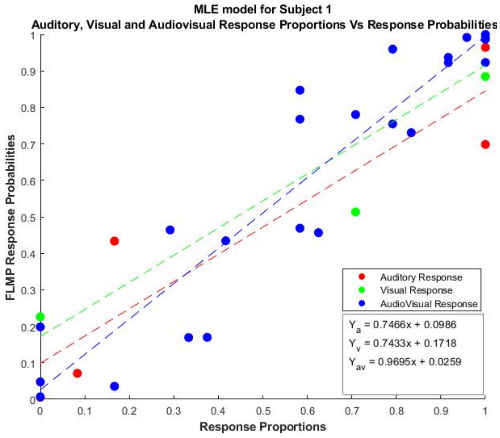


d) FLMP response probability vs response proportions for Subject 4.

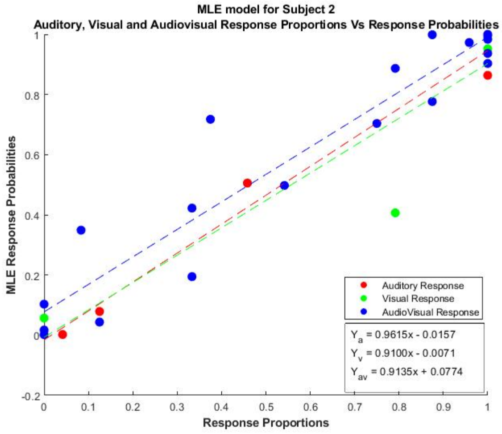


e) FLMP response probability vs response proportions for Subject 5.

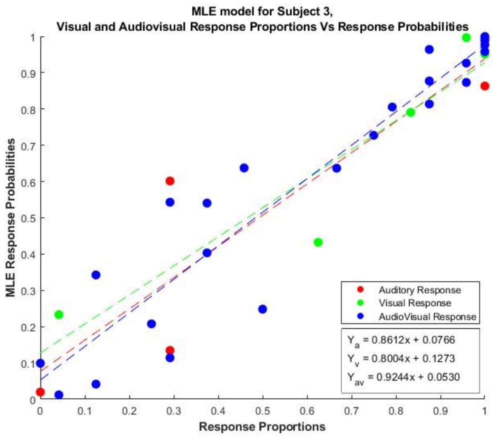
Figure 8: Scatter plots of the FLMP response probabilities vs. the response proportions for auditory, visual and audiovisual responses, after cross-validation. The linear regressions to fit the data are also included.



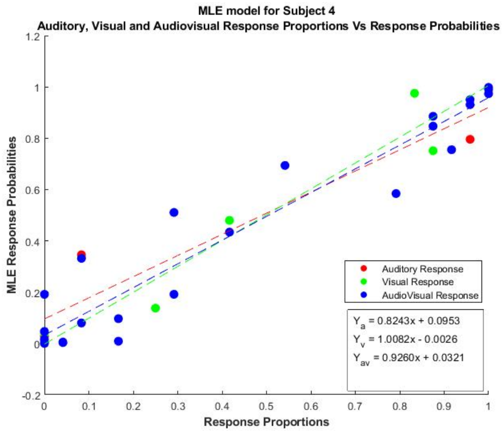
a) Early MLE response probability vs response proportions for Subject 1.



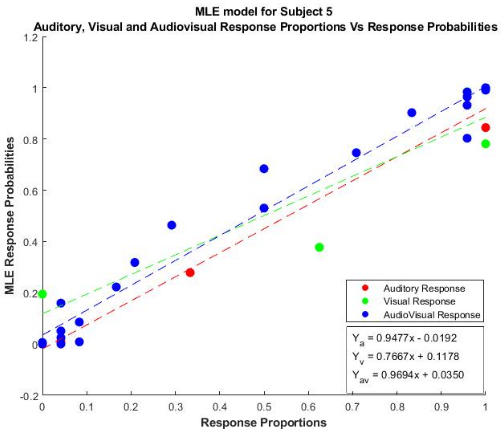
b) Early MLE response probability vs response proportions for Subject 2.



c) Early MLE response probability vs response proportions for Subject 3.



d) Early MLE response probability vs response proportions for Subject 4.



e) Early MLE response probability vs response proportions for Subject 5.

Figure 9: Scatter plots of the Early MLE response probabilities vs. the response proportions for auditory, visual and audiovisual responses after cross-validation. The linear regressions to fit the data are also included.

Figure 8 and 9 show the response proportions of the original data vs. the response probabilities of FLMP and MLE models for each subject, respectively, by using a 35-fold, cross-validation procedure. In that way, the y-axis is showing the test response probabilities.

Then, by fitting the model to the original data (using linear regression), it was not as perfect if compared it to the model fits without cross-validation. Cross-validation took into account random effects in the data, being the reason why a high error was observed when validation the testing data.

References

[1] T. S. Andersen, "The early maximum likelihood estimation model of audiovisual integration in speech perception," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2884–2891, 2015.

[2] "https://se.mathworks.com/help/optim/ug/fminunc.html."

[3] "https://se.mathworks.com/help/stats/fitlm.html."