Internal use only

002

TR-IT-0074

Report on first EMMI Experiments for the MIDDIM project in the context of Interpreting Telecommunications

Georges FAFIOTTE & Christian BOITET

1994.9.10

This note describes the first EMMI-based experiments conducted during our research stay at ATR, in the framework of the MIDDIM project. We first present the goals and methodological framework common to the seven phases of the two experiments conducted. Then, specific methodological points are given, first observations are accounted for, and some insights on results are sketched, though no full transcription could be performed in the very short time available.

ATR – Interpreting Telecommunications Research Laboratories
2–2 Hikari-dai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan

Table of Contents

-o-o-o-o-o-o-o-o-o-o-

Report on first EMMI Experiments for the MIDDIM project in the context of Interpreting Telecommunications

Georges FAFIOTTE & Christian BOITET

GETA, IMAG (UJF & CNRS
BP 53, F-38041 Grenoble Cedex 9

research stay at

ATR Interpreting Telecommunications
2-2 Hikari-dai, Seika-cho, Kyoto 619-02

1994.9.10

Abstract

This note describes the first EMMI-based experiments conducted during our research stay at ATR, in the framework of the MIDDIM project. We first present the goals and methodological framework common to the seven phases of the two experiments conducted. Then, specific methodological points are given, first observations are accounted for, and some insights on results are sketched, though no full transcription could be performed in the very short time available.

Introduction

This note describes the first EMMI-based experiments conducted during our research stay at ATR, in the framework of the MIDDIM project. These experiments were prepared by a first modeling of the situation, subjects and factors bound, described in a Specification Report [1].

In this Experiment Report, we first present the goals and methodological framework common to the seven phases of the experiments. Then, specific methodological points are given, first observations are accounted for, and some insights on results are sketched, though no full transcription could be performed in the very short time available.

1. Characteristics common to the experiments

1.1 Goals

The MIDDIM experiments organized here in the available time and with the settings of the multimodal generic environment currently available at ATR, were intended to use EMMI for multimodal Client-Agent interactions.

The aims of the experiments, announced in this specification document, were:

(1) *"to test what influence variations in speech recognition ability and language proficiency of the interpreter, and in domain knowledge of all participants, will have on the number of disambiguation questions asked by the (automatic) interpreter"*, and

(2) to collect spontaneous ambiguities in this multimodal interpreting situation.

As a matter of fact, we concentrated on qualitative observation of interaction modalities on the EMMI environment.

This is because our main goals were not to take measurements, nor to derive an experimental factor analysis in the classical sense, on strong or closed hypotheses.

Rather, in experimenting multimodal disambiguation with EMMI, it seems to us more important, to:

- gather, and analyze later, qualitative observations on multimodality events and effects,
- openly consider situations close to future real usage of Intepreting Telecommunication Services. For instance, the limited linguistic processing ability of any future machine interpreter justifies our choice of Wizard of Oz interpreters with limited linguistic ability. It is also important that they are efficiently trained to behave as machines.

- take a careful look at the effects of EMMI's present generic settings, and determine which improvements, if any, could help to achieve the above-mentioned goals better.

This is why, in our view, our experiments should allow

(3) to gain a better understanding:

. of user behavior in multimodal interactions, and

. of what could be an interactive multimodal simulation environment specifically customized for Interpreting Telephone situations,

following in this an exploratory then incremental prototyping scheme, well established in Software Engineering.

Such a prototyping approach first involves a Wizard of Oz Interpreter for every speech and language processing function, then provides a basis for later modular functional replacement, depending on the progress of MT/MI technology.

1.2 Methods

a. Situation

Two participants are conversing,

- the Client, seeking explanations from
- the expert Agent,

both using only their own language (two different languages),

about a question in particular domain.

In these first experiments, both participants were supposed —and intended— not to be able to understand the language spoken by the other party. This is important in our view to simulate realistic and demanding situations, in order to learn more about user behavior modeling and in environment modeling.

The Wizard of Oz Interpreter is interacting with both parties. He/she is supposed to simulate a Machine Translation Environment.

Actually, one aspect of the setting of our experiment on the EMMI environment somehow hampers the credibility of such a Machine mediation. It is that the 2 subjects well knew about the Wizard of Oz situation. They even very well knew the person 'interpreting the role' of the Machine. This Wizard Interpreter sat in a separate, but very close, cubicle. Such a setting could be easily modified in order to simulate more realistic situations.

Two Observers were on site.

One, near the Interpreter, could lightly assist him/her in managing eventual complex situations (like ordering simultaneous speech), however not playing a moderator role in the experiment.

The second, near the Client and Agent stations, only took notes.

b. Factors

Three main factors are considered, for each of the 3 experimentees:

Language Proficiency, Domain Knowledge, Auditory Aptitude,

plus a side factor: Environment Mastery.

c. Intended scope of valuation for the factors

In these first experiments, Client's and Agent's Language Proficiency (LP) in the other's language, as already reported, was intended to be as minimal as possible.

The LP of the Interpret it was intentionally required to be limited, trough an artificial lowering of the actual human proficiency, through damaged input signals to the Interpreter by means of a Vocoder system. Again this medium range ability was intended in our view to match realistic machine ability, to learn more from this realism.

Actually this 1st factor was intended here not to vary, with respect to effective intrinsic linguistic aptitudes. In fact it could well do, due to an increase of attention focusing and of acceptable cognitive load, correlatively to adaptation to the environment (cf. 4th side-factor). This drastic effect of adjustment trough self-training was anticipated, and we insisted on fully recording the very first trial and use of the participants. This expectation was clearly confirmed by the experiments.

Domain Knowledge was required to be as low as possible from 'Machine' side, in order to provoke if possible numerous natural spontaneous disambiguations, in these few experiments.

The factor was intended to stay fixed over this sub-series of experiments. This was expectedly achieved through successive experiments, with a different aspect being explained each time, so that the participants could not learn and improve their knowledge of the subject matter from one phase to the next.

Auditory Aptitude was expected to vary here for the 'Machine' Interpreter. This variation was supposed to correlate with an artificial input distortion in Client's and Agent's speeches.

The experimental situation was anyhow to provide with a variation of the Environment Mastery 4th factor, since we chose not to train users beforehand (especially the 'Machine'), in order once again to observe a wider scope of situations.

2. First experiment: Japanese-French / Technical Explanation, Algorithmics

2.1 Specific settings

a. Experimentees and other participants:

Client:	Herve Blanchon
Agent:	Kyung-Ho Loken-Kim
Wizard Interpreter:	Mutsuko Tomokiyo
1st Observer:	Christian Boitet
2nd Observer:	Georges Fafiotte

b. Specific factor variation

The two languages used were: Japanese and French.

Domain was here: Algorithmics, namely the presentation of Viterbi's algorithm, used in MT.

The factor settings were as follows.

It is recalled that in the following 'Subject/Factor Profile' Tables, x–y indicates an interval, < x > an element, < x , y > a pair, and < x–x' , y–y' > all pairs < x" ,y " > with x" in x–x' and y" in y–y".

The rating is as follows:

0	*none*	1	*poor, low*	2	*mediocre*
3	*average*	4	*fair*	5	*good*
6	*very good, expert*				

\ Factor **Subject **	**Language Proficiency** < F , J >	**Domain Knowledge**	**Auditory Aptitude** <F perc., J perc.>
H. Blanchon C	< 6 , 1 >	< 4 >	< 6 >
K.H.Loken-Kim A	< 1 , 6 >	< 5 >	< 6 >
M. Tomokiyo I (simulating a machine)	< 5 , 6 > to be compensated down to < 2 , 3 > by vocoder	< 1 >	< 1–2 , 1–2 > instead of < 2–3 , 2–3 > through vocoder distortion

The experiment allowed 3 effective sessions or phases to be held, with the following settings for the 3rd factor:

Phase	**Topic**	**Perception levels (I)** < F , J >	**Remarks**
1	Explanation of the algorithm (goals, method)	< 2 , 2 >	Distortion set to compensate for too high linguistic ability
2	Explanation of the data structures used	< 2 , 1 >	Japanese side more distorted
3	Explanation of the coding of the algorithm	< 1 , 2 >	French side more distorted

2.2 Durations, adjustments

Date: 26 / 8 / 1994

Phase 1:	14 mn.	Vocoder:	65
Phase 2:	15 mn.	Vocoder:	66 (distortion)
Phase 3:	17 mn.	Vocoder:	69 (important distortion)

2.3 Observations

a) Specific material

Several files, describing the algorithms explained, were made available on the EMMI environment, and were used.

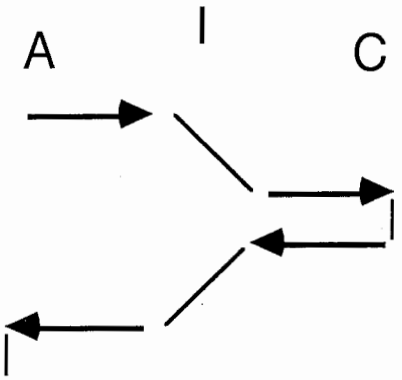
b) Condensed transcription

A model of immediate condensed transcription, with 'visual abstraction' of speech (here) and main multimodal events was originated and refined, in order to suit the experiment following-up, and

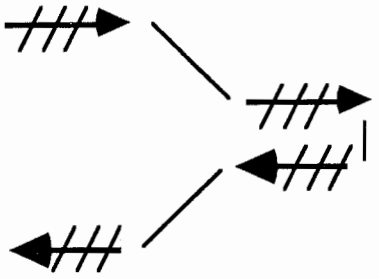
- to facilitate immediate marking out of clarification or disambiguation sub-dialogs,
- to later be used as a story-board abstract to locate and represent multimodal events,
- to help eventually for the written transcription.

Notation used is as follows, though much more flattened (enlarged here for better explanation)< It proved convenient.

Time is running here from top to bottom, but could as well be organized horizontally. Timing, main behavioral events, headlines of Clarification or Disambiguation topics are reported in ad hoc adjacent columns. Agent and Client utterances may be categorized later (Question, Answer, Neutral information, Correction of a previous utterance, etc.). Clarification or Disambiguation sub-dialogs are also oriented, and may be numbered afterwards and colored according to classes.

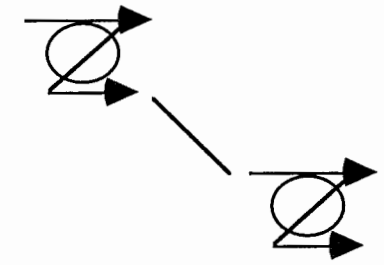


Agent speaks : 1 long utterance
(Interpreter thinks)
Interpreter speaks to Client
(Client thinks)
Client speaks : 1 long utterance
(Interpreter thinks)
Interpreter speaks to Agent
(Agent thinks)



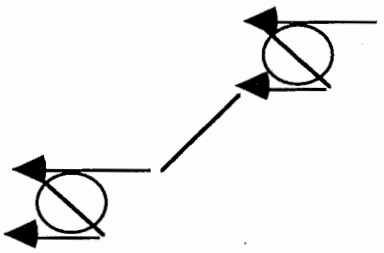
Agent speaks : 1 divided utterance
and intricatelly
Interpreter speaks to Client (divided)

Client speaks (divided)
and intricatelly
Interpreter speaks to Agent (divided)



Disambiguation sub-dialog
(Machine-Interpreter requires from Agent,
to clarify or disambiguate)

Disambiguation sub-dialog
(Client requires from Agent, through Interpreter,
to clarify or disambiguate)



Disambiguation sub-dialog
(Machine-Interpreter requires from Client,
to clarify or disambiguate)

Disambiguation sub-dialog
(Agent requires from Client, throuhg Interpreter,
to clarify or disambiguate)

c) Subject adaptation and self-training

From notes taken during the sessions, and from spontaneous account from the subjects, we noticed...

- Subject adaptation:

  Client initially spoke very fast, and rather softly. He spontaneously evolved over the sessions towards slower, louder, more distinct speaking.

  At the beginning Client and Agent did not wait for the Interpreter to finish her translation. Then they progressively did. Visual controls could help them to discipline the dialog.

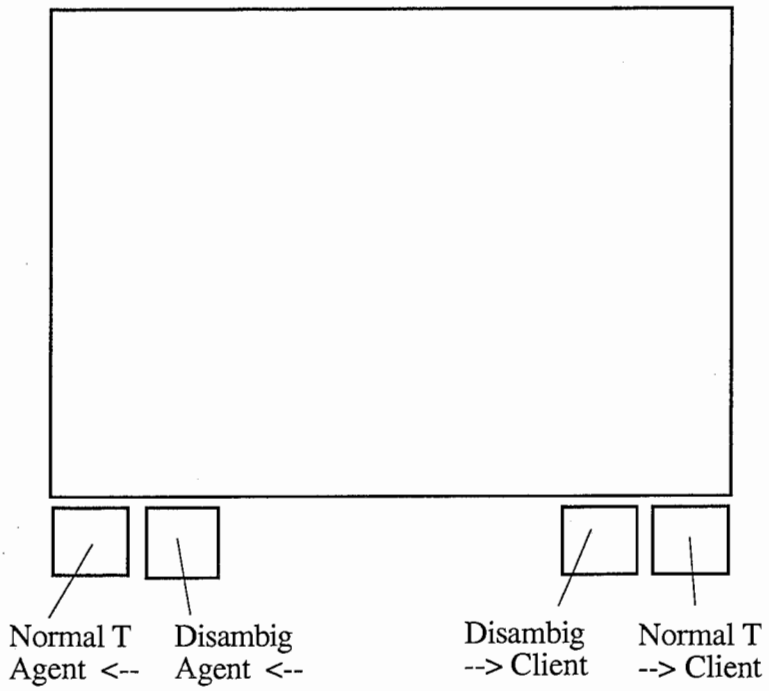
At a point, the Client did not look, for some long time, at drawings appearing to his screen, only listening to the Interpreter. The isolating effect of the headphones used probably drains off the attention after some time. A short sound could announce the arrival of a drawing.

- Interpreter's training and self-training:

The interpreter was deliberately chosen untrained neither with respect to the environment (commands...), nor to her task. She gradually took a part in monitoring the interaction with Client, and to some extent Agent. A built-in resource could help this monitoring function.

d) Environmental tuning

After phase 1, 4 post-it colored markers were added at the very bottom of Interpreter's screen.



They intended to help Interpreter in situating the nature of her speech, i.e. a 'Speaking Interpreter' state, among four states:

. yellow NTA Normal Translation towards Agent,

. yellow DA Disambiguation (or Clarification) sub-dialog with the Agent.

Interpreter then seemed (and later confirmed) to have identified better her current activity.

. blue NTC Normal Translation towards Client,

. blue DC Disambiguation (or Clarification) sub-dialog with the Client,

Meanwhile, DC and DA states or sequences on the notes were refined, on the 2nd Observer's notes, respectively in

. DCI Disamb. or Clarif. sub-dialog asked to Client by Interpreter,

. DIC Disamb. or Clarif. sub-dialog asked to Interpreter by Client,

. DAI Disamb. or Clarif. sub-dialog asked to Agent by Interpreter,

. DIA Disamb. or Clarif. sub-dialog asked to Interpreter by Agent.

Other visual aids, controls commands, could be added to build a palette of resources which would discipline and rationalize the Wizard Interpreter's task.

- The manual entire reproduction of the drawings by the human Interpreter may be a heavy task, especially because it should be done simultaneously, or concurrently, with some translating. The cognitive load is important.

An automatic drawing transfer function could ease the process. Automatic transfer of the entire drawing all at once, at a precise moment chosen and announced by the Interpreter, can be very pertinent in certain cases. In other cases, real reproduction of the rhythm and duration of the original drawing may be important and should be preserved.

Interpreter should not have to redraw, since it may be a difficult task. Drawing nicely, clearly, meaningfully, can be a skill that the Interpreter could be also trained to acquire. But Interpreter is supposed to focus on translation first. Sometimes this re-drawing could even improve the original Client's or Agent's drawing. Moreover, Interpreter's screen can become very heavily drawn (like in one session observed) thus increasing perceptive then cognitive load, unless systematically erased (therefore also on addressee's screen).

3. Second experiment: Japanese-French / Town orientation

3.1 Specific settings

a. Experimentees and other participants:

Client:	Tsuyoshi Morimoto
Agent:	Georges Fafiotte
Wizard Interpreter:	Marc Seligman
1st Observer:	Christian Boitet
2nd Observer:	Mutsuko Tomokiyo

b. Specific factor variation

The two languages used were: Japanese and French.

The factor settings were as follows.

Factor \ Subject \	Language Proficiency <F , J >	Domain Knowledge	Auditory Aptitude <F perc., J perc.>
T. Morimoto C	< 1 , 6 >	< 3 >	< 6 >
G. Fafiotte A	< 6 , 1 >	< 5 >	< 6 >
M. Seligman I (simulating a machine)	< 4 , 4 > to be compensated down to < 2 , 3 > by vocoder	< 2 >	< 1–2 , 1–2 > instead of < 2–3 , 2–3 > through vocoder distortion

It was specified that A would explain to C how to go somewhere, namely how to reach first the Grenoble Culture House then the University Campus, from Lyon-Satolas airport.

We had envisaged 3 successive sessions, with a different aspect being explained each time, so that the interpreter cannot learn and improve his knowledge of the subject matter from one phase to the next. In each phase, we decrease the auditory aptitude of the interpreter by increasing the distortion of the vocoder.

In fact the experiment allowed 4 effective sessions to be held, with the following settings for the 3rd factor:

Phase	Topic	Perception levels (I) <F , J>	Remarks
1	Explanation of how to come from Lyon-Satolas Airport to Grenoble bus terminal	< 3 , 2 >	Distortion set to compensate for too high linguistic ability
2	Explanation of how to go to the tramway station	< 2 , 1 >	Japanese side more distorted
3	Explanation of how to go to the 'Maison de la Culture'	< 1 , 1 >	Both sides more distorted
4	Explanation of how to go to the Campus	< 1 , 1 >	Both sides more distorted

3.2 Durations, adjustments

Date: 30/ 8 / 1994

Phase 1:	8 mn.	Vocoder:	65
Phase 2:	7 mn (system breakdown):	Vocoder:	66
	12 mn.		
Phase 3:	12 mn.	Vocoder:	117 (very high distortion)
Phase 4:	14 mn.	Vocoder:	69

3.3 Observations

a) Specific material and settings

Several maps in French (Lyon Airport, Grenoble Bus Terminal and Tramway Map, in French) were made available on EMMI, and used during the experimentation.

An automatic drawing transfer was made available on EMMI for the Interpreter to directly forward when he wished any drawing issued by the Agent or the Client. It eased Interpreter's task.

b) Subject adaptation and self-training

We use here observations made during the second experiment by the Wizard Interpreter, which somehow correlate with former observation derived from the first experiment.

- Wizard training

<< So a training period for the Wizard is recommended. Experiments carried out before the Wizard warms up should be treated separately. They tell you about the Wizard's training process, not about the translation process itself. They probably should not be included in the translation corpus.

The training should also explicitly cover how the Wizard should react to various problems. What do you do when you can't recognize a source expression, or can't remember a target expression? Or when you can't see something which was drawn or typed? >> [M. Seligman personal communication].

- Agent and client training

Again,

<< for participants as well as the Wizard, training effects are quite important, and again such effects should be considered when evaluating the data. At first, the speakers used normal-length utterances in natural speaking style, but soon they switched to careful, short phrases.

...I didn't directly request these changes, since I wasn't sure whether it was allowed. But the participants got the idea anyway, after about two dialogues. (To repeat, I think the Wizard should be explicitly instructed which requests are allowed, and the use and effect of the requests on the communication success rate should be investigated.) >> [idem]

c) Auditory ability

From the experiment we reinforced our first questioning on the use of the Vocoder facility, as the best means for the intended variation of factor 3. The Vocoder may be an efficient way to produce some 'machine-like' output useful in a Wizard of Oz situation. But the input fading or alteration that we wish for the Wizard Interpreter could be obtained through other methods.

<< ...the major factors in comprehension were the degree of fluency of the translator and the cooperativeness of the participants in speaking distinctly, rather than the distortion. In the early dialogues, the distortion added to the already considerable stress. But ... a fluent listener could learn even more quickly to "tune out" most reasonable levels and types of distortion. Thus I tend to think that simulated listening errors should be introduced in a more controlled way. >>> [idem]

4. Discussion

NB. Written transcriptions of the audio records are planned later on. They will allow a deeper analysis of the effect of the factors on the occurrence of ambiguities and of clarification demands, and on their multimodal solving process.

Meanwhile, partial elements can be sketched.

But further analysis only may confirm or qualify, or even balance, first elements drawn here.

4.1 Disambiguation dependency

Factor 1 - Language Proficiency:

No direct variation of this factor here.

However, we observed and it was reported by subjects tan effect on this ability to translate, resulting from a temporary cognitive over-load, especially here for the Interpreter, eventually for Agent or very inexperienced Client.

This resulted in multiplying clarification sub-dialogs. In this sense, factor 4 shows to have a side-effect on factor 1.

In our experiments, Interpreters were not trained yet on EMMI display and commands. They had to adapt through self-training over the first sessions.

Methodogically, it may be interesting of course to conduct distinct experiments with different starting Environmental Mastery (factor 4) for the Wizard Interpreter. This would allow to observe a full range of 'extensive to unnecessary' self-training situations, more precisely their differential effect on clarification and disambiguation occurrences.

Factor 2 - Domain Knowledge:

No direct variation of this factor here.

The first experiment offered a situation with a low ability in factor 2 for the Interpreter. Numerous lexical clarification seems to result. Again this situation is of interest, among other experiments to conduct, for simulating realistic MT situations in case of limited Domain or Terminology Knowledge.

Factor 3 - Auditory Aptitude (varying factor):

This factor was set up to vary along the phases of the experiments.

In the second experiment, a rapid adaptation to the distortion (reported by the Wizard Interpreter himself) limited the expected effect on speech recognition (segmentation, parsing..) for the 'Machine' simulating Interpreter.

We also observed spontaneous adjustment in Client's and Agent's speech production towards a somewhat pre-segmented slower speaking, probably compensating the increasing distortion and easing Interpreter's recognition.

Therefore, other means should be found to vary subject Auditory Aptitude, then providing with more frequent spontaneous clarification and disambiguation events.

Software distortion, simple lowered sound level, smothered input signal, could cause a more efficient 'cotton-eared' effect on subjects.

Factor 4 - Environment Mastery:

We did not intend to have variations of factor 4, initially seen as possibly important in its effect, but intentionally fixed for the sake of simplifying the experimental protocol.

As already mentioned, we better learned from these experiments how this factor can influence —more than initially expected—, users' tendency to slow down the dialog for clarification sub-dialogs, and spontaneous generation of disambiguation events.

This could be perceived in most of the Clarif.-Disambig. situations: Disambiguation provoked by Client towards Agent, Agent-provoked towards Client, Machine-provoked towards Client, Machine-provoked towards Agent.

4.2 Collecting and analyzing spontaneous ambiguities and classifications

Disambiguation situations will be collected from audio material, and classified within the MIDDIM Data Base.

Some exemplary lexical ambiguities occurred: e.g. in French "bus" also means a 'bus', but "car" means an 'inter-city' bus, not a 'car'.

4.3 Analyzing behavioral aspects in a multimodal scheme

Records of the faces of both Client and Agent are available. But video conversion of the full screen of the Interpreter was unfortunately not recorded.

This strongly scaled down the analysis of the impact of factors 4 (Environment Mastery) and 3 (Auditorial Aptitudes), in terms of observable behavioral adaptation and modality self-adjustment, by the subjects.

We did not proceed yet to further inquiry on this matter. Some elements have already been reported as observations.

4.4 On the role of the Wizard of Oz

We could confirm from these experiments that the 'Wizard of Oz' Interpreter (WOI) is of course not only a language interpreter.

He or she should as well interpret a machine interpreter (in the sense of 'playing a role of'), and actually a WOI-driven machine. This role is complex.

The WOI will have several different functions. We well agree that WOI should be trained for behavioral aspects. We again think that a medium range ability should be preserved, as far as Linguistic Aptitudes are concerned, to benefit disambiguation simulation situations of real future Machine Translation.

- WOI should behave as a simple and 'realistic' machine, in particular consider him/herself at any time in one of the following submodes or subtasks:
  - . always either translate what's being said by one participant, or
  - . enter in a sub-dialog with one participant only, for clarification or disambiguation of what this participant just said,
  - . never enter in any human-like sub-dialog with a participant, for instance referring to something which was not said, or said much before, or arguing about something about the translation, or mentioning something that obviously a machine would not know.
- WOI should act a moderator, a regulator, in the dialog of the two experimentees: prevent them to take a turn or a lead in speaking, when it is not relevant or if it is a source of disturbance at this point in the translation process.
  - . therefore he should be allowed to stop them if needed (real human interpreter do so), rather in the way a machine would do.

4.5 On further experiments and settings

- Other experiments could choose as well to focus on speech-only or speech-mainly interactions, and simulate them on EMMI.

- Visual (and very moderately audio) prompters or controls could indicate to participants their current interaction mode or the state they are supposed to be in. This would help them to discipline their interaction. Such a modification was simulated during the first experiment, using colored labeled post-its.

- The automatic drawing transfer modification, for the Interpreter was easily done on the generic EMMI environment. Other variants of this modality could be developed, to better reproduce the exact speed and duration of the drawing as issued by its participant author.

- The same modifications could well apply to written messages: automatic transfer by the Interpreter, of messages issued by other participants.

-o-o-o-o-o-o-o-o-o-o-o-

References

[1] **Fafiotte G. & Boitet C. (1994)** *Specification of first EMMI Experiments for the MIDDIM project in the context of Interpreting Telecommunications.* MIDDIM report, GETA-IMAG & ATR-ITL, Aug. 1994, 6 p.

[2] **Fais L. & Loken-Kim K.-H. (1994)** *Effects of Mode on Spontaneous English Speech in EMMI.* Technical Report, ATR-ITL, June 1994, 17 p.

[3] **Loken-Kim K.-H., Yato F. & Morimoto T. (1994)** *A Simulation Environment for Multimodal Interpreting Telecommunications.* Proc. IPSJ-AV workshop, March 1994, 5.

-o-o-o-o-o-o-o-o-o-o-o-