FLIC Preliminary Evaluation Report – Extracts

Please note that the following information has been drawn from two preliminary reports, is therefore not exhaustive (in total there were 5 different learning institutes involved, this report comprises information on only 2 of the institutes as at the time of preparation, the other courses had not been finished.) and will be added to by the final evaluation results – expected for beginning of June 2006.

Background

The FLIC system is a specialised computer-based system designed to help learners acquire an additional language in a more natural way, based on the capability of providing sophisticated delivery, and potential mixing of, instructor's and user's speech.

In this project five European centres participated in a study of the effects of the FLIC system on second language learning.

Evaluation Summary
Groups were pre-tested on objective measures of language ability and assigned to matched groups on the basis of their expertise, for either traditional or FLIC enhanced training. Following training, post tests were undertaken, and a questionnaire completed. A series of specialised tests of language (the FLICA tests) were developed and delivered to measure the effects of the training on prosody, phonology, accuracy of nonsense word repetition, and accent. Case studies were conducted in Sheffield, using a similar approach, but including an evoked potential study before and after training to identify any changes in the brain.

In this interim report we update the analyses of the FLIC evaluation study. Full data and analyses will be presented at the end of the project. Results indicate, as in our previous reports, satisfactory results from the paper-based tests, with solid improvements for both groups on objective measures of language competence (Decel) and self-evaluations of speaking, listening and vocabulary skills, with improvements in all these tests for both groups, but with greater improvement for the FLIC group in DECEL and Listening skills. The 'FLICA' (advanced listening and speaking) tests were administered at all centres, but the FLICA analyses here are based only on two of the centres, where English was the target language.

The results available so far suggest that FLIC has been successful in its aims of improving language learning in the participants, in particular with participants whose starting point was poor.

Both groups showed improvements on all aspects of prosody and phonological production, including accuracy, accent and pronunciation. The FLIC groups showed greater improvements overall on all aspects of these tests. Statistical analyses indicated a significant effect of training overall, but no significant effect of group, or interaction between group and training.

In conclusion, the results of the tests at both the individual centre and overall group level to date support the research predictions and illustrate that although both techniques are effective, the FLIC system seems to have the edge in effectiveness. If these results are supported by further analysis of the remaining German speaking centres, they have implications for the wider use of the FLIC equipment.

Group allocation

Pretesting involved the administration of Decel to establish the level of competence of each participant. Decel is an adaptive test that provides objective data on the starting level for each participant in the study, which ranged from less than 1 (instantiated here as 0.5) to 3. The critical Decel scores necessary to assign the participants to their training groups were all returned, and computer print outs of the results are available for further analysis. These scores were ranked in order of ability, and assigned alternately to the FLIC or traditional group teaching. This generated two well-matched groups, assigned randomly to either FLIC or traditional intervention. A background questionnaire was also completed. The groups were well balanced in terms of their prior experience of their target language. Eleven participants were included, 6 receiving FLIC and 5 traditional teaching. There were 3 males and 3 females in the FLIC group and 2 males and 3 females in the traditional group. Students were aged between 19 and 40, with only one in the age group 40+, all but two were native French speakers, and all but two were unemployed. This background data suggests that the groups were relatively homogeneous.

The hypotheses
- i) Both teaching methods will improve language learning overall
- ii) Specific benefits will be found in the FLIC group for speech reception and aspects of speech production
- iii) No specific benefits are predicted from the FLIC training for vocabulary.
- iv) The case studies will reveal changes in brain processing of second language speech sounds

Key issues motivating the analysis include the following:
- i) Have the measures used captured the improvements in the groups following training?
- ii) Do the users responses reflect these improvements (based on self ratings and the final questionnaire)?
- iii) Is FLIC more successful than the traditional teaching in aspects of language learning?

Data have now been collected from all five evaluation sites, and our final analyses are close to completion. Here we present the results from analysis of DECEL for all centres, and results from the analysis of the self-assessment statements. A comparison of performance at pre and post-test for DECEL and self ratings is included.

Progress towards analysis of the FLICA programs is also reported here. These sound files require intensive analysis, but have the potential to provide a more sensitive measure of progress than existing tests. We have identified the need for more sensitive analysis of naming speeds and phonological discrimination, and a computer program is being developed to handle this. A preliminary blind analysis of the sound files from participants learning English has been completed for two centres, and a second rater is working towards completing the second analysis in order to establish further inter-rater reliability. Native German speakers have been identified locally to analyse the remaining German files.

Fig 4. Improvement from pre to post test on self rated vocabulary for FLIC and traditional groups
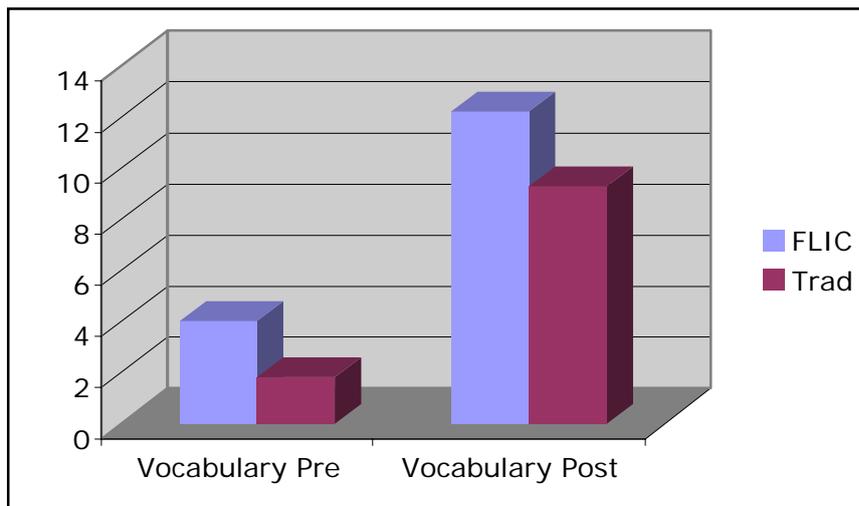

Figure 4 above also shows striking improvements in self-ratings of vocabulary skill. By contrast with only 2/4 positive responses at pre-test, participants are producing an average 10/12 positive responses. In terms of individual improvements, scores at pre-test ranged from 1-6, and at post-test from 6 to 20, indicating improvements for every participant.

### P3: Phonological Discrimination (Minimal pairs)

This is intended to assess the ability to hear the phonemes in the second language and involves presentation of two words (or nonsense words). Half of the time the stimuli are the same (though spoken by different speakers), and half of the time the stimuli are different, with one phoneme changed. The user's task is to click on 'same' or 'different' for each pair.

Figure 5. Improvements from pre-post test on paper-based tests, DECEL, Listening, speaking and vocabulary for the 2 groups
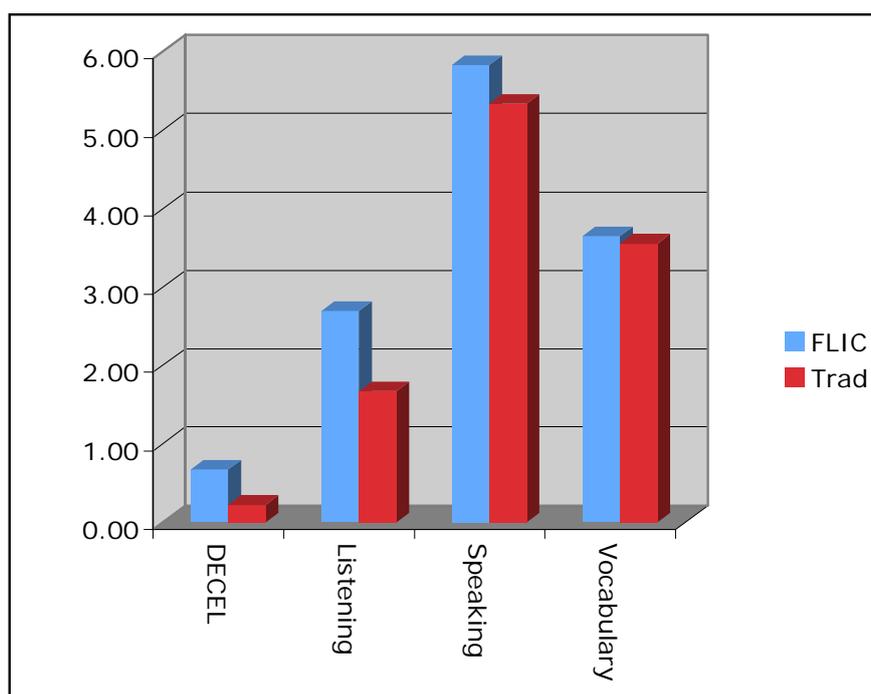
It may be seen from the figure above that there is greater improvement for the FLIC groups on all aspects of performance, with the greatest effects on listening and speaking, and objective improvement in DECEL, in line with the predictions of the study.

Fig 8. Changes from pre to post test on phonological discrimination for FLIC and traditional groups
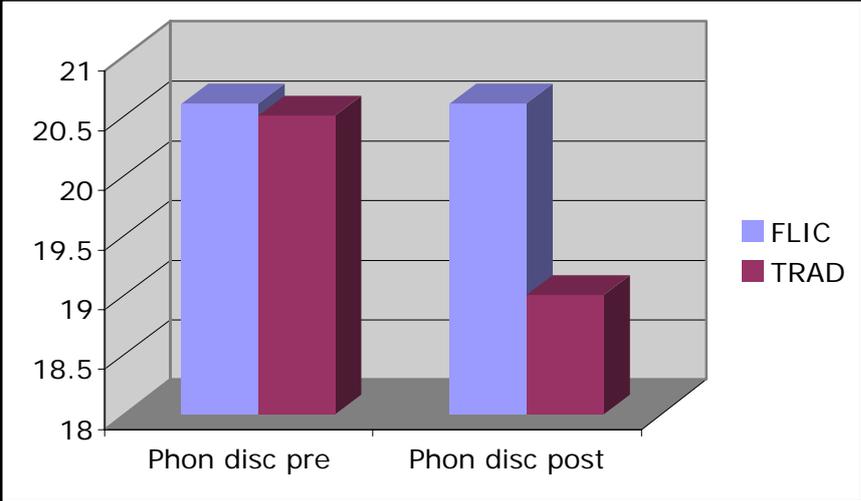
29 stimuli are presented and an element of guessing is possible, which would generate a score of around 50% correct. It may be seen from figure 6 above, that participants are scoring well above chance levels. Pairs have been chosen which tap some of the more difficult discriminations in the target language. Results were not always consistent from pre to post test, suggesting that some element of guessing is involved.

Consideration of the figures above show that for both centres the FLIC group are performing better than the traditional group, in the accuracy, the accent and the pronunciation in the phonological production task.

Figure 13. Accent score for prosodic production

Accent on Prosodic Production task by FLIC and TRAD groups (post test) at Melun and Wisoak CEL
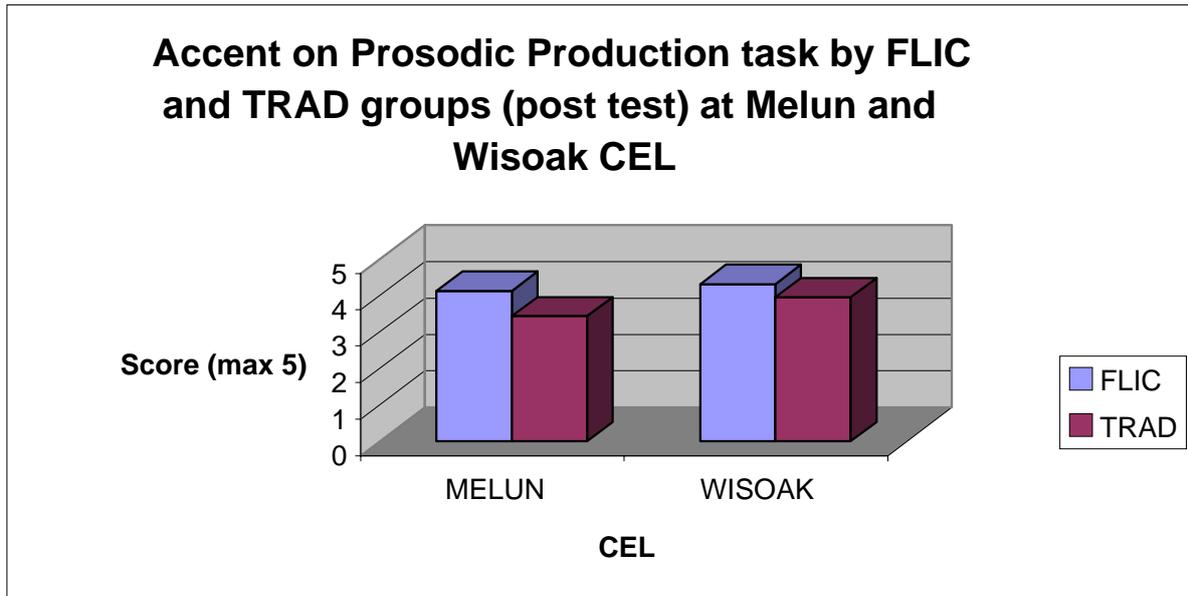
It may be seen from the figures above that for both centres the FLIC group are performing better than the traditional group, in both the accuracy of their prosody and their accent. Prosody is one of the areas in which it was predicted that the FLIC system would lead to improved performance.

Figure 23. Improvements in listening speaking and vocabulary following training
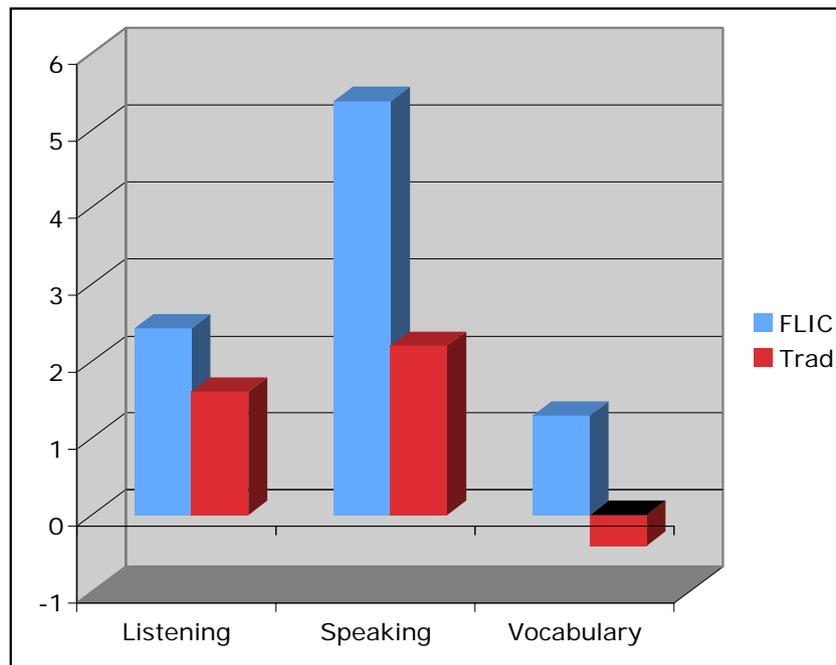
It can be clearly seen from the figure above that the FLIC intervention at this centre has a striking effect on self-ratings of speaking skills, and a solid effect on listening and vocabulary, with improvements on all tests greater for FLIC than traditional interventions. One of the main differences between this group and the Arles group is the level of performance at pre-test on Decel.

4. Evaluation of the strengths and weaknesses of the FLIC and traditional teaching

Participants were asked to write open-ended comments on how they perceived the FLIC and the traditional teaching, giving strengths and weaknesses. This has so far led to remarkably positive results. Interestingly, this group had initially been equally happy to be taught using either FLIC or traditional methods. This suggests that they had not formed strong preconceptions on whether or not FLIC would be of benefit to them. It is therefore particularly striking to note the general satisfaction with and confidence in FLIC. It is also gratifying that the traditional course itself is well rated, with high praise for the teachers and the course. Nevertheless, it is clear that participants feel the teaching has been significantly enhanced by the addition of FLIC.

Returns have not yet been translated into English, but it is notable that all 6 FLIC participants praise FLIC for its effects on concentration and pronunciation, and the richness and interest of the course. No-one noted any substantive weaknesses in the system, but several suggested that further enhancements could be made by inserting a blank page into the program for note taking.

Feedback overall on the course was very positive, with a traditional participant noting
"Une tres bonne ambiance, avec un bon groupe, tres bonne entente'

The FLIC group reported
'Encourage le concentration et donc la memorisation de la langue est plus facile' and that they found the traditional teaching *'Moins de concentration pour l'ecoute, moins de effects de prononciation, moins de rythme et de dynamisme',* identifying specific limitations, despite their overall satisfaction with the teaching.

The hypotheses under investigation

Hypothesis 1: Both teaching methods will improve language learning overall

This was clearly supported (see Table 1 and also the results of the inferential analyses in section 2).

Hypothesis 2: Specific benefits will be found in the FLIC group for speech reception and aspects of speech production

This is largely supported by the effect size data (Fig. 5); by the FLICA data (§2.4).

Hypothesis 3: No specific benefits are predicted from the FLIC training for vocabulary.

As predicted, no specific benefits were found for Vocabulary training for FLIC (§2.2). On the other hand, the zero effect size for vocabulary is a solid achievement for FLIC in that the traditional training gave considerably more vocabulary and grammar than the FLIC training.

Hypothesiss 4. Are there changes in brain organisation following FLIC training?

This issue will be addressed via analysis of the four recently completed individual case studies.

5. Summary

It should be stressed that this interim report has been compiled in a very short time, when not all of the data was data collected within two weeks of the report, and not yet fully analysed. Nonetheless, in summary the preliminary results from this first evaluation of the FLIC system are encouraging.

(i) The data for all three forms of assessment have proved analysable for both groups.

(ii) The DECEL data on objective comprehension and sentence completion indicate improvements for both groups, with a tendency for greater improvement for the FLIC group.

(iii) The self-rating data on speaking, listening and vocabulary indicate improvements for both groups, again with a tendency for greater improvement for the FLIC group.

(iv) The objective FLICA data on rapid naming, phonological discrimination and nonsense word repetition also generally show improvement following training, though the numbers of participants analysed to date means that comparisons are difficult.

(v) The final feedback data indicate very positive assessments of their experience by the participants, with intriguing suggestions that the FLIC approach may indeed help with skills of listening and speaking that the traditional approach does not reach.

Evaluations of FLIC by the two participants with the lowest level of expertise at pre-test are particularly glowing. Both these participants had previously learned their target language, English, for over 2 years, which suggests that FLIC may be effective with participants whose skills are hard to improve. If this pattern continues to be found in the remaining evaluations, it would seem that the FLIC system represents a significant enhancement on traditional teaching methods.