How to Train a Dragon: A comparative investigation into attitudes and efficiency of voice recognition software for formative and summative feedback

Abstract

Using a mixed methodology we investigated the suitability of the use of voice recognition software for formative and summative feedback when grading different types of Higher Education (HE) assessments. In particular we were interested in reducing the time spent grading papers, time spent sitting in a sedentary position and whether the software allowed the user to provide inadequate, satisfactory or improved feedback. The preliminary findings indicate a high satisfaction rate among users in terms of ‘ease of use’ and a reduction in time spent ‘sat’ at a computer, however, some participants reported issues with regards to technical specifications and limitations relating to marking the actual scripts.

Keywords

feedback, voice recognition software, assessment
Introduction

In recent years, voice recognition software, such as Dragon Naturally Speaking, has gained popularity and there has been an increase in users who face barriers to producing electronic text via mechanical production (Honeycutt, 2003). There exists a body of research that investigates the functionality of such programs for users with learning difficulties, such as dyslexia (De La Paz, 1999), and those with physical disabilities such as work-related upper extremity disorders (Honeycutt, 2008). Voice recognition has also been in-built into new consumer technologies, such as Xbox 360 Kinect (Microsoft, 2012), iPhone 5s (Apple Inc., 2011) and a selection of new Smart Televisions and tablets. It is likely, therefore that this technology will become a normalised form of input for young digital natives of the future. Bill Gates stated that, ‘speech is not just the future of Windows, but the future of computing itself’ (Gross and Judge, 1998, online). Given the movement towards this form of input in the consumer arena, it seems prudent that educational institutions should investigate the practical applications of such technology in terms of efficiency and accuracy for administrative tasks. In particular this method of recording data may provide benefits in terms of the timely return of assessment feedback and the physical well-being of lecturing staff as negates the necessity of desk-based, sedentary input.

The use and impact of traditional desk-based mechanical user interfaces (i.e. keyboard and mouse) to record student feedback on digitally produced feedback sheets had been raised as discussion points in both Curriculum Management Committee and Learning, Teaching and Assessment meetings at University Centre Peterborough (UCP). Many lecturers were concerned about the impact of spending long hours sat at a desk; in particular, concerns were raised about the physical impact of a sedentary lifestyle, as well as a discussion surrounding ‘quality of life’ at key marking periods such as Christmas and the New Year.

Existing research by Blatter and Bongers (2002) states that working with a computer in excess of six hours per day is associated with work-related upper limb disorders (WRULDs). The authors conclude, ‘recent increases in the number of employees working with computers and in the durations of computer and mouse use coincide with a huge prevalence increase of work related disorders of neck and upper limbs (WRULDs) and sick leave, which poses financial burdens on companies’ (Blatter and Bongers, 2002, p.1).

The use of computer technology amongst academic staff across all departments is likely to increase in the future. Hacifazlioglu, Sacli and Yengin (2007) conducted a questionnaire study across two faculties (Arts and Engineering) from a sample of 44 lecturers’ attitudes towards computer usage and reported that 70.4% of respondents used computers in excess of six hours per day, highlighting the potential risk facing lecturing staff from WRULDs.

Given the increasing issues of high capacity computer use, utilising voice recognition software as a form of input for assessment feedback may offer a solution to the health and safety hazards faced by lecturers. However, whilst Nuance Software, the creator of the Dragon range of software, claims high levels of accuracy and speed, the reality of the functionality is debated. The body of evidence relating to functionality and practicality is limited and Honeycutt (2003) suggests that the future focus of research should focus on Voice Recognition Technology (VRT) rather than other forms of dictation software. Additionally Honeycutt states that in order to understand the impact of VRT on the quality of writing, further case studies and formal experiments with variations in gender and age are required. This is also true of how the use of VRT may affect the writing process itself – for example; how bodies of text are created and the cognitive processes involved.

This study was designed to address one of the main suggestions for future research from Honeycutt’s (2003) comprehensive literature review on the use of VRT for educational purposes, ‘Whether dictation can compare to silently written forms of communication in terms of quality and the impact of VRT on the writing process’ (2003, p. 92). In order to investigate this, the researchers designed a qualitative study that collected and analysed data from a number of volunteer lecturers that were trained in the use of Dragon Naturally Speaking.
Aims of Project

- To enable lecturers to provide feedback in a timely manner.
- To enhance the quality of summative and formative feedback in order to improve student success and achievement.
- To utilise new learning technologies to develop more effective summative feedback mechanisms at UCP.

Methodology

Firstly, lecturers were trained in the use of Dragon Naturally Speaking software. This took place over a period of three months and staff were given access to specific hardware and software in order to familiarise themselves, and to be comfortable with using the equipment. Once the staff indicated they felt ‘at ease’ with, and were aware of, the study requirements, we asked them to use Dragon Naturally Speaking to record verbal feedback onto the digitally produced marking sheets. Additionally, we requested that they recorded their thoughts and feelings about using the software as opposed to the traditional method of typing feedback; some indication was given as guidance, of the sorts of responses they may record. For example, physical comfort and well-being, time spent recording, environment and ease of use, this data was extremely useful as it illustrated how the participants felt from the start to the finish in their experience with the software. Although the sample was small (three males and three females) there was some difference in the commentary with regards confidence levels and ease of use. The male participants tended to record confidence with using both the hardware (laptops) and software earlier than the female participants. Additionally their comments indicated a willingness to explore additional features that they had not been trained in, whereas the female participants were happy to ‘master’ what they knew before they ventured into ‘unknown’ menus. This data was used to inform both the questionnaire and the interviews discussed below.

Open-ended question questionnaires and unstructured interviews were used to record the views and experiences of the lecturers. The open-ended questionnaires were sent electronically to lecturers for ease of use. The questions were more detailed and subject specific, focussing on the training experience and ease of use. A sentence completion model was used for the preliminary questions, with some completely open-ended questions in the closing section. Again here, as with the written comments that the participants completed during their training, there was a slight difference between male and female answers. For example, the three male participants answered that they would feel confident installing the software at home and would not need any help; whereas the three female participants answered that they would prefer someone more experienced to install the software for them. In addition, the staff were interviewed using an unstructured interview format. This allowed for a more flexible, ‘in-depth’ exploration of some of the concerns and experiences of the staff. In particular it allowed the participants to discuss openly the issues they experienced with ease of use, and to elaborate on the cost benefit analysis of speaking versus typing. It also highlighted (albeit on a small scale) the problems faced by some users when trying to vocalise feedback that is usually typed or written as a ‘stream of consciousness’ process, a problem alluded to by Honeycutt (2003) when discussing the cognitive processes involved in text production.

Results

Three clear themes emerged from both the interviews and the questionnaire results; these are listed and discussed below.

- **Technology** - All participants were critical of the amount of memory needed to run the software. As a result, this affected the users’ hardware and many found it easier to revert to typing comments to avoid screen freeze and the loss of other data. In the unstructured interviews this aspect was deemed by many to be the most frustrating aspect of using the software and participants expressed concern that the current hardware available for use at UCP would not be ‘up to scratch’, and that the current system would struggle to accommodate the system requirements.

- **Time and location** - Some participants commented on the time it took to feel confident enough to use the software. They expressed a desire to have more structured training sessions in which they were given the opportunity to work in groups or pairs in order to ‘bounce ideas’ and help one
another. Other comments included remarks about sensitivity of the recording in relation to accents, peripheral noise (i.e. children playing on the street, pets, doors opening and closing), and, if they wanted to use the software at work, finding a suitable ‘quiet’ space or empty classroom. With one participant this also led to a discussion about carrying a laptop around as well as the student work. With a static desk-based, mechanically produced feedback system, the work stays in one place and the lecturer does not have to be responsible for carrying extra weight. This could be addressed if students’ work was marked from an electronically stored site, such as Turnitin.

- **Confidence** - Most participants indicated that they would not feel confident if asked to explain, train, or disseminate information about the software to their peers. In addition, and despite having a period time to familiarise themselves with the software, most indicated that they did not feel confident that they had ‘mastered’ the software, or that they were fully aware of all the features they could be utilising. Some participants felt the need to be seated and in a sedentary position as the microphone lead was too short. Others (in particular the three female respondents) expressed concern over more practical matters, such as, the sharing of documents that had to be double-marked or both markers working together and trying to add comments via voice recognition software. They also felt less confident about using this software to comment on videos and art pieces.

**Summary**

The results do not align well with the aims and predicted outcomes of the initial proposal. The themes and issues identified above have done little to indicate whether the use of Dragon software will improve either lecturers’ well-being, or the quality of feedback to the students.

**Conclusions and recommendations**

The results were inconclusive in relation to the primary aims. This was disappointing as it illustrated that the research design did not match the objectives. The small sample size coupled with the participants’ workload and lack of time to actively engage in using the software also had a negative impact. However, some illuminating data was produced that we hope will inform future research into this area. The study is best described as a pilot and we recommend that future researchers revise the data collection methods to incorporate focus groups and extended training sessions, and that the study is much larger in terms of the participant pool and amount of assessments that are marked electronically.

Feedback from the interviews included suggestions that a ‘comment bank’ is initially produced by the participants so they can be confident using the software to ‘draw on and add in’ ready-made ‘generic’ commentary. For example, it was suggested by one female participant that statements about referencing or sentence construction and/or structure, would be helpful. This could save time and may help participants appreciate the difference made by utilising the software.

Conversely, the research team felt, this could affect student feedback negatively if those individuals perceive the generic comments to be an indication of a lack of a personal touch on behalf of the lecturer (as discussed at the Anglia Ruskin Teaching and Learning Conference). Arguably, there is a long way to go before voice recognition software takes the place of mechanically produced feedback. However, the authors have no doubt that it will eventually replace the more traditional methods and is therefore worthy of further investigation for use in Higher Education establishments.

**References**


How to Train a Dragon: A comparative investigation into attitudes and efficiency of voice dictation software for formative and summative feedback


WELL THE TRIALS ARE GOING WELL, BUT THE BIG BREAKTHROUGH WILL BE WHEN WE GET HIM HOUSE TRAINED.
THE JOYS OF PLAYBACK WITH DRAGON...

COULDN'T HAVE SAID IT BETTER MYSELF!

COMPUTERS ARE A PAIN IN THE A**!!

JUST ANOTHER LECTURER WHO'S BEEN OVERDOING IT AT THE KEYBOARD...

Conference Cartoons by CaricaturesbyLukeWarm.co.uk