

The writing processes and learning strategies of initial users of speech recognition

A case study on the adaptation process of two professional writers

Mariëlle Leijten & Luuk Van Waes

This paper describes the adaptation and learning process of writers who have started using speech recognition systems for writing business texts. To gather the process data for this study we have chosen complementary research methods. First the participants were asked to fill in a questionnaire, then they received instruction in the speech recognition system, they were observed five times using the speech recognition system during their day-to-day work and filled in a logging questionnaire after each task.

The data from these sessions were used to describe the adaptation strategies during the learning process by revealing the planning, formulating and reviewing behaviour of the writers. This article focuses on (1) the effect of speech recognition on cognitive processes of writers; (2) the learning strategies of initial users; (3) the description of a research method, categorization model and notation model to answer research questions as extensively as possible.

To illustrate the possibilities of the categorization model we describe a case study in which we show the learning and writing process of two experienced dictators. Both writers have a comparable experience in professional writing, word processing and classical dictating. However, they do differ in learning style (cf. Kolb 1984). The case study shows that this difference in learning style is decisive in the adaptation process and because of this it also shows that the used method enabled us to describe these kind of writing process data adequately.

Keywords: Writing processes, learning strategies, speech recognition, research method, writing modes

1 Introduction

Until a few years ago research on speech recognition was focused on the qualitative (technical) improvement of the technology. Research on the applications of speech recognition in a business context was difficult, because of major technical shortcomings. However, continued research efforts have achieved a significant breakthrough in developing commercial speech products (Lernout&Hauspie, Philips, Dragon & IBM) that make it possible to dictate fluently to a computer with an acceptable error margin (after practice). Since the late nineties some usability studies have been conducted on the use of speech recognition systems (Halverson et al., 1999; Karat et al., 1999; Karat et al., 2000) but, so far, little is known about the cognitive processes of writers who are using speech input devices.

Previous research has shown that next to several social and individual factors, also the writing mode (such as pen and paper versus computer) influences the organization of writing processes (Hayes, 1996). The emergence of speech technology as a new input mode for writing processes has created a new writing mode, which may again influence the organization of writing processes. Writing with speech recognition systems develops a hybrid modus between dictating and writing with keyboard&mouse. Contrary to traditional dictating processes, speech recognition systems offer the possibility of immediate text feedback on the screen as with writing with keyboard&mouse. This seems to be the biggest advantage of speech recognition over classical dictating. The inability of writers to keep track of the external representation of the text produced so far is a common problem in classical dictating. To monitor the text, dictators must maintain a mental representation of the text in their memory, or if this is not sufficient, relisten to the tape (Reece & Cumming, 1996). These characteristics of classical dictating all impose an extra cognitive load on the writer. But using speech recognition and being able to see the text written so far may also have a negative effect on the cognitive load, because of incorrect text that sometimes appears on the screen due to technical shortcomings of the speech recognition software. It is our hypothesis that an incorrect representation of the text on the screen could also hinder the writer to build a coherent representation of the text.

The goal of this article is not to stress the technical aspects of the current speech technology systems, but to show how users adapt their writing to this new mode and what strategies they develop in using speech recognition systems in their day-to-day work. (In the study reported here we used L&H Voice Xpress Legal™ version 4.)¹ The scope of this paper is threefold. Firstly, we focus on the cognitive aspects that characterize the writing processes of people writing in this new writing mode, i.e. using speech technology. The other focus in this study is the learning process and how initial users adapt to a new writing mode: we would like to investigate how different users learn to write with a new writing mode. How do they solve problems? How do they use a combination of different writing modes for different writing activities? And when do they switch between different modes? How do the writers adapt their planning, formulating and reviewing behaviour? In answering these questions we will also pay attention to the way in which these kind of research questions can be answered, by describing a categorization model for speech recognition data. Finally we will focus on the methodology we have developed to observe and analyze the writing processes with the speech recognition mode. In paragraph two we will first position this study in research about writing processes, speech recognition and learning processes. Then we will give a description of the research project this study is based on. Paragraph four gives a description of the analyzing methods we used and the categorization model we developed. We also give a short description of a notation model we developed to make the data accessible and to illustrate specific observations. In the fifth paragraph the purpose of the case study is explained and in the following paragraph the results of the case study are elaborated. Conclusions from this case study are drawn in paragraph six and in paragraph seven we will discuss these conclusions in the broader perspective of the research project, related research and further research.

2 Related research: writing processes

Writing media have been the topic of several studies on writing processes (Flower & Hayes, 1980; Haas, 1996; Van Waes & Schellens, 2003). These studies show that the use of a specific writing medium often creates complementary perspectives to describe some of the cognitive processes that support writing in more detail.² By observing the writing process in the speech recognition mode, we also hope to gain insight into some writing processes in general. Or as Haas stated: *“As is often the case, new situations – in this case new technological contexts and situations for writing – bring to the fore aspects of writing that may have been there all along, but that have not been previously noticed. [...] – something with which writers have presumably been operating all along – becomes obvious in such a new situation or context: the changed technological context of composing with word processing.”* (Haas, 1996, p. 117)

The effects of the writing medium are wide ranging; the writing experience, the writing process, and the resulting texts are all influenced by the medium. Several studies have been conducted to compare writing processes in handwriting and in computer-based writing. These studies show that the writing medium has an effect on the writing process. For example, the initial planning in computer-based writing is significantly less than in handwriting. These studies also show that writers shift attention from the writing task to the medium. An aspect which is definitely to be expected in the speech recognition mode, because the interference of the writing medium is even bigger - especially in the beginning - than with word processing.

In the revised writing model of Hayes (1996) the composing medium arises in the task environment (see figure 1). In the former model (Flower & Hayes, 1980) the ‘text produced so far’ already played a significant role. This part of the task environment also influences this study because the writers are able to reread the text they have produced so far on the screen, contrary to what they normally do when dictating with classical dictating devices. The text remains unseen until the dictated text is typed

¹ Lernout & Hauspie is nowadays Scansoft.

² When examining writing processes for dictating, writing patterns occur that differ from computer writing processes.

out. With speech recognition the writer can constantly refer to the text, keep track of the progress and revise already written text segments.

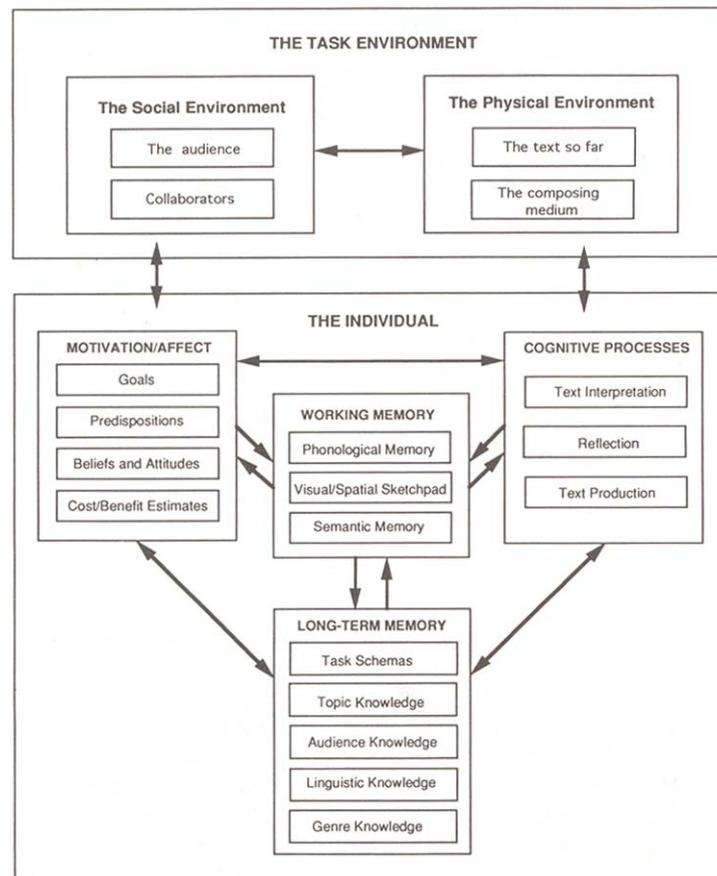


Figure 1 Writing model of Flower & Hayes (Hayes, 1996, p. 4)

Speech recognition can be described as a hybrid writing mode, because it is more or less a combination of dictating and computer writing. Traditional dictating is characterized by a high degree of linearity in the production of text. Only few revisions are made. Sentences or phrases are dictated one after the other. The only revising usually taking place is a mental revision before the text is dictated to the recorder. In contradiction to the traditional dictating mode, writers using speech technology get immediate written feedback on the computer screen. This creates the possibility to review the text in all stages of the writing process, opening the gates to non-linearity. A high degree of non-linearity is typical of computer writing processes (Eklundh, 1994, Van Waes, 1991). Most computer writers consider the paragraph, or even a sentence, as a unit that is planned, formulated, reviewed and revised in short recursive episodes. The constant feedback on the screen offers them the possibility to revise a lot, without losing the overview of the final text.

Different components of the writing process require evaluation of the text produced so far. This is a very complex task, even more if a text is not at all visible as it is in the dictating mode. But also in the computer mode text evaluation is difficult. A study of Wood et al. (2002) showed the importance of the physical availability of the text produced so far. They examined how academics use computer technology when writing academic papers. They suggested that hard copy probably offers cognitive support that may not be available in computer writing³ (Wood et al., 2002). And next to findings of Haas (1996), they show the advantages of capturing the ‘sense’ of a document. Computer technology

³ Wood et al. (2002) state that revisions are better traced in paper versions of a text. However, they do not mention the revising tools that are available in Word processors these days.

might not sufficiently support the memory and organizational demands that the writers need to have of a text. These demands, however, differ per writer. Consequently, we might assume that writers with different learning characteristics may require different ways to facilitate their writing process.

2.1 Speech recognition and writing

With the term 'speech recognition' we refer in this article to the use of dictating software. Texts are written by talking to the computer. The speech recognition software just adds a toolbar to a normal working environment on screen (see figure 2). The programmes are often called continuous speech recognizers. This means that the software has an input of fluent speech, and that it is not necessary to spell words discretely. By talking to the computer via a headset with a microphone, users see their text appear on the screen. Every action that used to be done by keyboard&mouse can be done with speech input; dictate text, navigate through programmes or texts, apply functions like making words bold, and open or close files. Speech recognizers are speaker-dependent. The user has to train the software to teach the system his or her voice. Then the software adapts dynamically to the particular voice characteristics of the user. When a person first begins to use the speech software, the recognition and performance is still poor. The performance of the speech recognition software improves as usage and training time grows (Rodman, 1999).



Figure 2 Example of speech recognition toolbar (Lernout & Hauspie Voice Xpress)

Speech recognition enables writers to talk to the computer at a normal speaking rate. This could cause the production rate of these texts to be much higher, because we can speak faster than most of us can handle the keyboard to type their texts. But the errors that are caused by speech recognition are easy to miss, because they are most of the time spelt correctly (Honeycutt, 2003). This slows down the writing and especially the revising process, because writers have to be extra careful to avoid these kind of mistakes appearing in their final text.

Whereas most research in the field of speech technology has dealt with the technical improvement of speech recognition and the role of phonology, only a few studies have focussed on the use of speech recognition as a writing tool (Halverson et al, 1999; Karat et al, 1999; Karat et al, 2000). These studies were mainly focussed on the usability of speech recognition as a writing mode. In the study of Karat et al. (1999), for instance, initial and professional users tested the usability of three speech recognition systems. The initial users were given two kinds of text creating tasks. They had to perform one task with speech recognition, the other with keyboard&mouse. Results showed that users experienced a great deal of difficulty correcting errors with speech input and subjects tended to stay in the speech modality much longer during a correction than the researchers had expected. Perhaps this can be explained by the goal of the study (usability and error correction) and the fact that the participants were observed all the time and that they had just received extensive training in speech based correction methods.

Also the use of speech recognition by the learning- and physically-disabled has received some attention. Quinlan (2002) studied the writing process and products of forty 11 to 14 years aged children. Twenty of these children experienced writing difficulties. The children composed a series of four narratives in four writing conditions, via handwriting and speech recognition, with advanced planning and without. Quinlan hypothesised that speech recognition provides cognitive benefits to children with writing difficulties and that advanced planning is supportive for the real-time planning process. Results showed that diffluent writers produced longer texts when using speech recognition. Speech recognition seems to reduce transcription-related interference enabling children to produce more fluently written – and thus longer – texts. For the fluent writers, composing with speech

recognition did not lead to improved narratives. Advance planning had a significant positive effect on text quality for both groups.

Hartley et al. (2001) focussed on differences between the writing products of keyboard&mouse and speech recognition. They compared the writing products produced in different writing technologies of three expert writers over a period of thirty years. Although the writing styles of the participants differed, the individual styles remained remarkably constant over time.

In another study Hartley et al. (2002) compared the writing products of two experienced writers, one using keyboard and the other using speech recognition. The writing products of both writers differed in sentence length, the number of long sentences and the use of the first pronoun.

In the relatively short history of speech recognition as a writing mode, no specific studies have been conducted on the effect of this new writing mode on the writing process. A study that perhaps relates the most to this subject is from Reece & Cumming (1996) on the Listening Word Processor. In their experiment participants dictated to a hidden typist. This study had the benefit that the text that appeared on the screen had no word recognition errors, as opposed to the text that appears when working with a speech recognizer. In their experiment with 30 young children (10-12 years old) they found very promising results for the LWP. The children had to write a text in three different writing modes: handwriting, dictation en LWP. Texts written in the LWP condition were generally superior to those produced by the other two methods. Reece and Cumming state that the Listening Word Processor fostered a different composing process to that seen with dictation:

“If long pauses are assumed to represent planning episodes, then there were indications that the LWP encouraged planning in a fashion not seen with dictating. We interpret this as follows: because the LWP provides the writer with a visual record of the text, the writer has no need to allocate the substantial portion of working memory to maintaining some form of ongoing representation of the composition. Instead, writers can turn their attention to a consideration of higher level aspects of the composition task.” (Reece & Cumming, 1996, p. 375)

Because of the young age of the participants of this study it is hard to say whether these results can be generalised to professional (adult) writers.

2.2 Learning

An important characteristic of this study is that the participants have to learn to produce texts in a different way. Therefore, we will pay attention to different learning styles. In this study we used Kolb's definition on learning: “Learning is the process whereby knowledge is created through the transformation of experience.” (Kolb, 1984, p. 38) In this definition we would like to emphasize two aspects of the learning process which are especially important from our perspective. The first one is the emphasis on the process of adaptation and learning as opposed to content or outcomes. The second one is that knowledge is a transformation process, being continuously created and recreated. The knowledge people have of working with speech recognition is constantly transformed by the experience they built up while working with the programme.

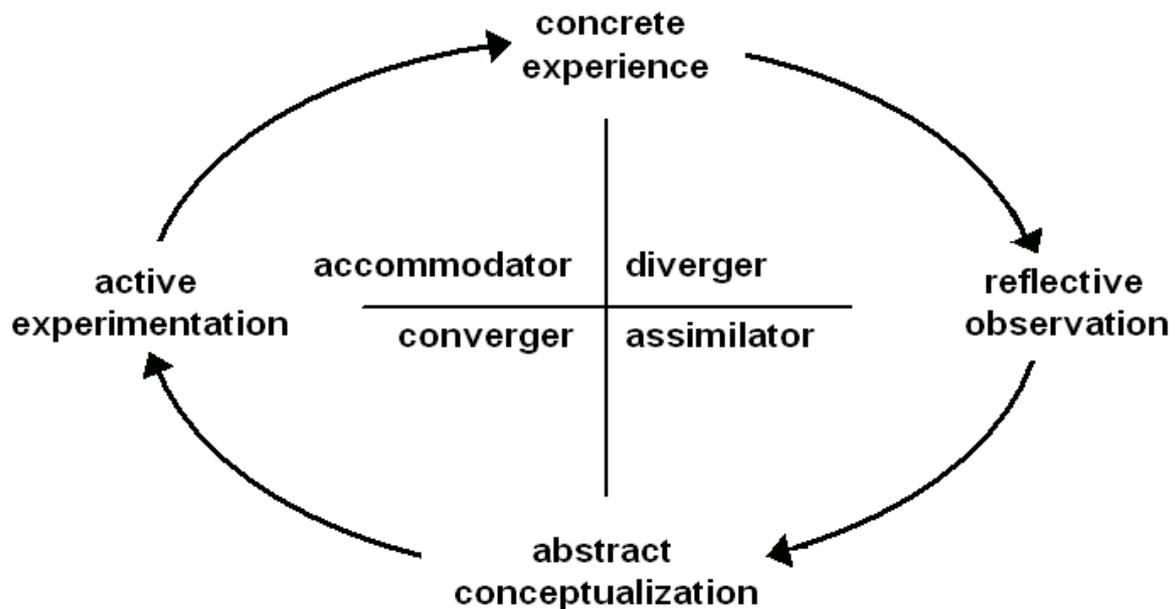


Figure 3 Learning styles (Kolb, 1984)

Kolb describes the process of learning as a four-stage cycle involving four adaptive learning modes: concrete experience, reflective observation, abstract conceptualization, and active experimentation.

In his *Learning Style Inventory* he evaluates the relative preferences an individual holds for each of these four learning modes. The *Learning Style Inventory* is an objective, self-scored instrument that reveals four statistically prevalent learning styles: diverger, converger, assimilator, and accommodator.

People with a converger or assimilator learning type prefer a higher level of abstract conceptualization. Convergents prefer to learn via the direct application of ideas and theories and have been described as somewhat unemotional with a preference to work on their own. Assimilators are good at taking in a wide range of information and reducing it to a more logical form. They tend to prefer theoretical models and deductive reasoning; this leads to a greater interest in abstract concepts and ideas than interaction with other people.

Divergers and accommodators are people that rely heavily on concrete experiences when learning. Accommodators, because they also prefer active experimentation, are described as having the ability to carry out plans and get things done. They get involved quickly in new material through a trial and error method of learning. Divergers are identified by their ability to look at a learning situation from multiple points of view. They often have a hard time to make a decision and mostly prefer to observe rather than participate (Kolb, 1984; Terrell, 2002).

For this study we have opted to observe writers in their professional environment. The texts they were writing were part of their 'normal' work; the only difference was that we asked them to use speech recognition as a writing medium. Kolb predicts that persons in technology and information science careers generally benefit from a converger or an assimilator learning style. Therefore, we would expect that for the kind of task we used in our study, especially those participants who are characterized by these learning styles, would perform the best. But of course, participants whose 'preferred' learning style is less suitable for learning to work with a new writing mode, will also be able to adapt their learning style to meet the requirements of this new mode, perhaps less efficiently. In this study we will take a closer look at the concrete strategies participants develop in learning to work with speech recognition.

In this article we focus on the feasibility of the methodology we have developed to study writing processes of speech recognition and the related learning processes of writers learning to write their business texts with the new medium. We are mainly interested in the general cognitive processes that characterize the writing process and will appear by the focus on a new writing medium: How will the

writers adapt their writing to the new medium? Will their writing strategy change over time? Will the influence of the ‘text produced so far’ change because of the immediate feedback on the screen? Will the possibility of errors in the feedback effect the writers strategy?

3 Description of the research project

In this paragraph we describe the method we have developed to reveal the elements that characterize the adaptation strategies of the writers involved in the experiment. First we will give a description of the participants, then we will explain the design and procedure of the research study and finally we will describe the materials the participants produced.

3.1 Participants

In this research project we observed the writing processes of 20 participants for about one month. We chose to observe two different groups of writers: lawyers and academics. Most of the lawyers that participated in the study were used to dictating their texts with classical dictating devices. The academics on the other hand were not familiar with dictating. This difference in writing preferences enabled us to take into account previous dictating experience in the description of the adaptation processes. For this article, however, we will focus on two participants who are both used to dictating their texts with classical dictating devices. In a case study we will describe two participants who have the same writing experience, but who are characterized by a different learning style.

3.2 Design and procedure

Before we started the experiment the participants had to fill in two questionnaires. The first questionnaire focused on information about the participants’ prior knowledge of speech technology, their attitude towards writing with a computer, their learning strategies and writing profile. It also addressed their experience in writing business texts and using a computer or a dictating system. The second questionnaire contained a Dutch translation of Kolb’s Learning Style Inventory (cf. supra), which offered us information about differences in learning styles. In a post hoc analysis we used this information to study the influence of learning styles on adaptation processes.

Before the participants started using speech recognition for the first time, they watched an introductory video about the use of the speech technology programme. The software company provided this video. The participants were then informed about the procedure of the study in more detail. They were asked to use the speech recognition system during their day-to-day work for about at least three hours a week. The participants could decide for themselves how to use the software and were not restricted to only use the speech input.

In total we observed the participants five times, after 1, 3, 6, 9 and 12 hours, while writing in their own environment.

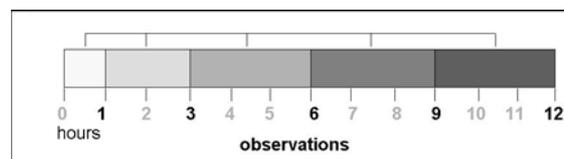


Figure 4 Timeline observation sessions

We collected the writing process data by an on-line camera (Camtasia™) and sound recorder (Quickrecord™). Because of the combination of the different input modes (keyboard, mouse and speech) we could not use existing logging programmes. We also observed the participants during each writing session and took notes about specific writing circumstances that could not be registered in another way (see figure 5). These recordings and notes enabled us to reconstruct the writing process in detail.

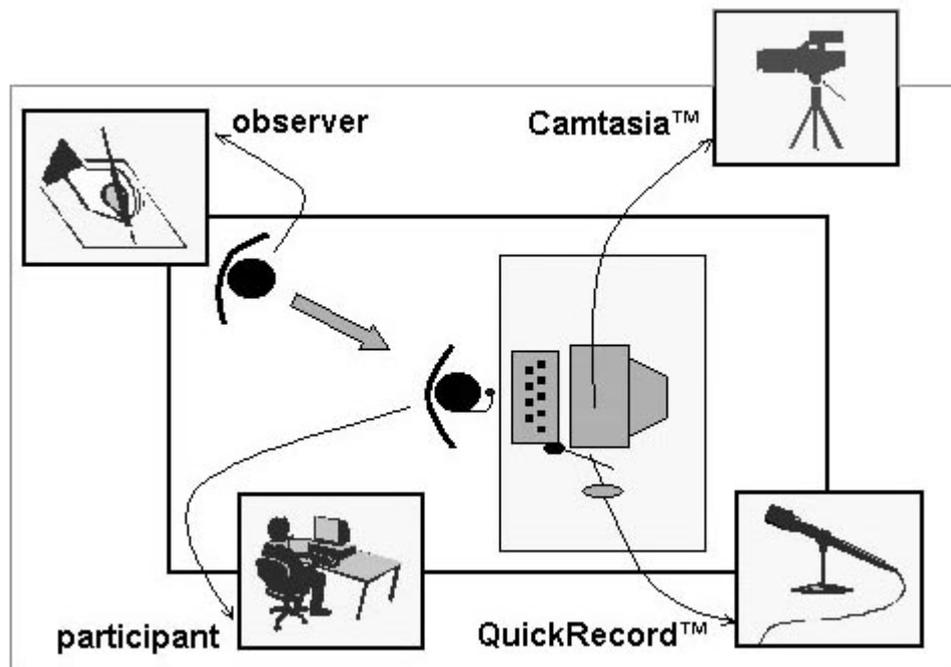


Figure 5 Observation setting

Between the observation sessions the participants were asked to fill in a logging questionnaire after each task they had performed with speech recognition software. This logging questionnaire was mainly meant to track information about what tasks were performed and which specific functions of the programme were explored. With this questionnaire we wanted to gather as much additional information as possible because the participants also worked with the programme and developed new strategies between the different observation sessions.

3.3 Materials

The different tasks the participants conducted were job-related and part of their normal writing activities, e.g. letters, e-mails or reports. During the observation sessions we collected both product and process data. The product data gave us information about the length of the texts (e.g. number of words, mean words per sentence), and the duration of the observation. The process data - collected by the on-line screencam, the sound recorder, and the observer - provided us with data about the gradual development of the writing process (e.g. mode switches, pauses and revisions).

In total we observed the writers for about 100 hours and they produced approximately 80 pages (about 24.000 words) of professional business texts.

4 Analysis

For the analysis of the data different methods were used. The analysis of the product data was aimed at contextualizing the writing activity. However, the focus of this study is on the process data. In view of this, we developed a *categorization* and a *transcription* model that takes the complexity of the hybrid writing mode into account and makes the enormous amount of process data accessible for further research.

4.1 Categorization model

The categorization model was developed to describe and categorize the different aspects of the learning and writing process. The model takes the complexity of the hybrid writing mode into account and makes the enormous amount of process data accessible for further research.

First, we will pay attention to the multimodal use of speech recognition. Then, we will describe the problem solving process of the participants using speech recognition. We use the term repair to refer to technical problems and revisions. The third category we will describe is about the revisions that are made to the text. Finally, we describe the instrumentation we have used to analyze the pauses during the writing sessions.

4.1.1 Writing modes

Speech recognition software allows writers to conduct a writing session only with the use of speech. But, will writers choose for this input as their main writing mode, or will they for example switch frequently between the different writing modes? Therefore we calculated the number of mode switches and the time writers spend in a writing mode. We were also interested in the effect that mode switches have on the cognitive writing activity. The ‘writer’ can decide to change the writing mode at any time in the writing process. Especially when problems rise in the use of the speech technology or when deciding to revise the text produced so far, a switch from the speech mode to the keyboard or mouse mode could occur.

For this study we categorized the writing modes separately. To analyze the mode switches we categorized each mode separately, but for the repairs we also combined the modes. We did not take the order of the combined modes into account.

- speech recognition
- keyboard
- mouse

4.1.2 Repairs

A problem that occurs frequently in human-computer spoken interaction is that the computer does not correctly ‘hears’ or ‘understands’, and consequently takes action in a way that is different from the message spoken by the user. This phenomenon is to compare with an ordinary ‘speaker-hearer’ situation in which problems of misunderstanding occur, because of noise in the conversation. To regain common understanding the speaker or hearer of an utterance has to repair the problem. In conversation analysis the concept of repairs is used to refer to a wide range of conversational phenomena including correction of misunderstanding and mishearing, word recovery or self-editings (cf. Schegloff et al. 1977). In repairs we can distinguish self-repairs and other-repairs. Self-repair refers to cases in which speakers repair errors or correct the appropriateness of their own speech. Other-repair on the other hand refers to the repair of disfluencies of the speaker by the hearer. In both cases the speaker is the person who initiates the error or indistinct phrasing.

For speech recognition (or other human-computer situations) we can add a third category: recognition repair (Stifelman 1993). In this case the hearer, i.e. the speech recognizer, is more or less responsible for the error (see figure 6). The user can also still be responsible for the error, for example because of a lack of knowledge about the speech recognizer (cf. *infra*).

		responsible for error ⁴	
		speaker	hearer
repairs error	speaker	self-repair (revision)	recognition repair (technical problem)
	hearer	other-repair	

Figure 6 Taxonomy of repairs (L.J. Stiefelman, 1993, p. 3)

According to a contribution model of Clark and Schaeffer, a conversation consists of a series of *presentations* and *acceptances*. A participant of a conversation who is presenting an utterance wants to know whether his utterance is accepted by the hearer. If we apply this theory to writing with speech recognition, we see that this is exactly what writers frequently (can) do when writing with speech recognition. After one or several utterances writers (can) search for evidence that their utterances have been accepted by the speech recognizer by rereading the text on the screen. In this study we noticed that most writers reread their texts produced so far to verify whether the dictated text corresponds to the text that appears on the screen.

The CA-concept of *repairs* is a valuable addition to the traditional concept of reviewing. It enabled us to incorporate a broader view of reviewing in this study, taking into account the specificity of the reviewing process in the speech recognition mode. Repairs are not limited to traditional revisions in the text (such as deletions and additions), but also the correction of misrecognitions or technical problems are taken into account. Also the distance of the repair can differ. It could either be done after one word or after several utterances. Not every problem does lead to a repair though. Sometimes problems or flaws are ignored, for whatever reason (Mazeland, 2003). For this study we adopt the concept repairs in the context of writing with speech recognition. In order to do this, we focus only on the ‘provoker’ of the error. In this study repairs refer to both ‘technical problems’ and ‘revisions’. The technical problems are – in most cases – caused by the speech recognizer (hearer) and have to be solved by the user (speaker). The revisions are both initiated and conducted by the speaker.

In this paragraph we will first describe the technical problems. We classified every speech recognition error and user strategy for repair. In paragraph 4.1.3 we will discuss the categorization of revisions separately.

4.1.2.1 Technical problems

We distinguished eight categories of technical problems:

- single misrecognition (dictated text is misrecognized and a single word appears differently than was intended);
- multiple misrecognition (dictated text is misrecognized and multiple words appear differently than intended);
- command misrecognition (a command is misrecognized and the computer performs another action. For example, instead of executing the command ‘end of the sentence’ correctly, ‘end of the sentence’ appears on the screen);
- dictated text as command;
- ∅ (hum or cough as text);
- ∅ unaccountable errors;
- no recognition (the computer does not respond);
- computer crash.

⁴ ‘Error’ has multiple meanings in this context. A text element of which the writer is not satisfied about for example is also called an ‘error’, whereas the writer only improves the text.

Intention & outcome

As mentioned earlier in the text the speech input sometimes led to something else. Therefore, we compared the intention of the writer and the actual outcome of the speech recognition system.

input	intention	outcome
1. word A	word A	word B
2. word	symbol	symbol
3. word	function key	function key
4. word	navigation	navigation
5. no input	no response	no response

The string word-word-word means for example a misrecognition. The writer intended to formulate a certain word, but in stead of that particular word another word was recognized by the speech recognizer. The string word-function key-word means that a command has been misinterpreted and has ended in the text as a text segment.

Cause

We especially focused on those problems because we think that the way in which the participants deal with those problems and try to solve them reveals important aspects of the writing and learning process. To value the technical problems in the way in which the writers solve them, it is interesting to know what caused the (technical) problem. Is it the user? Is it the speech recognizer? Or, is there no reasonable explanation for an action performed by the programme?

- user - ignorance
- user - inaccurate
- environment - noise
- software - ignorance
- software - illogical
- software - unknown
- user - unknown

Number of attempts

We counted the number of attempts to solve the problem, because when using keyboard&mouse, you can solve a problem at once. This is not really self-evident using speech technology. Repairs could either be done by keyboard&mouse or by speech input.

Writing mode

We also wanted to gather information about the mode the participants used to solve a problem. Do they hold to the same mode or do they switch writing modes to solve a problem? When writing with a classical dictating device, writers first produce the text and then make most corrections in a second phase. In this phase they probably use another mode, pen & paper or the word processor, to make these changes. Would they continue this pattern when writing in the speech recognition mode? Therefore we analyzed the three different writing modes (speech recognition, keyboard&mouse and all possible combinations between these modes (cf. infra)).

Level & distance of correction

For the distance analysis we compared the movements of the participant through the text: within the sentence, within the paragraph or outside the paragraph and the direction of the movement. Does he write a sentence and then correct all the errors or does he choose to write a paragraph and then correct all the errors in that paragraph?

Besides we made a distinction for all categories whether the movement was within or outside the computer screen.

- within sentence, within screen
- within sentence, outside screen
- within paragraph, within screen
- within paragraph, outside screen
- outside paragraph, within screen
- outside paragraph, outside screen

Direction of correction

Because we were also interested in the distance of the corrections, we have distinguished between *backward* and *forward* movements. Backward corrections start before the level of inscription and forward revisions occur in the text after the point of inscription.

Absolute time and interval

We noted the absolute time of the technical problems in the writing process and afterwards we calculated the position of the problems in the writing process by dividing the session into ten equal parts which made it possible to divide the session in equal parts.

Level of problem solving I & II

An extraordinary characteristic of speech recognition is that it is sometimes more time-consuming to correct just one character instead of a whole sentence or a sentence segment. Therefore we distinguished two levels of problem solving: a first level in which the writer actually wants to revise, and another level in which the writer actually revised. We categorized the levels: character, word, sentence segment, sentence, paragraph and punctuation.

	problem (level 1)	solution (level 2)
1.	character	character
2.	word	word
3.	sentence segment	sentence segment
4.	sentence	sentence
5.	paragraph	paragraph
6.	punctuation	punctuation

4.1.2.2 Revisions

In the repair analysis we distinguish the technical problems from revisions. Revisions are changes in the text after an evaluation of the previously written text. These changes are not initiated by technical conflicts caused by the programme but are meant to change the content, formulation or appearance of the text. Revision already has been studied over the years in different writing modes (e.g. pen and paper versus the word processor). Matsubashi defines revision as follows: *‘a revision is an episode in which the writer stops the pen’s forward movement and makes a change in the previously written text. The writer may then either resume the pen’s forward movement or make another change at another location.’* (Matsubashi, 1987, p. 208)

For writing with the word processor one could state we would have to replace *pen* with *keyboard&mouse*. Studies (Lee, 2003; Van Waes, 2003) show that when writing with a word processor, most writers revise more, and revise more locally than when writing with pen and paper. Writers also make more changes at the surface level, and they revise closer to where they are producing text at that moment. It is obvious that the shift of writing with pen and paper to a word processor caused a shift in revising behaviour. Therefore we are interested in the revising behaviour of writers using speech recognition. Would this new writing medium also cause a shift in revision behaviour?

Research about dictation as a writing method often focuses on text production. Texts written by dictating are composed more rapidly and are often longer. Dictation does not seem to affect the text quality (for a description of research on dictating see Reece & Cumming, 1996). Studies do show that dictation is more appropriate for writing simple sequential text, rather than texts with a more complex structure. This is why writers do not revise very much when dictating and when they do revise, it is at a local level (Gould, 1978).

We categorized the revisions on eight different levels: writing mode, type of revision, level of revision (I and II), goal of revision, distance of revision in written text produced so far, the direction of revision, absolute time, interval and writing phase. Below we will give a full description of the different levels.

Writing mode

We categorized the three different writing modes, viz. speech recognition, keyboard&mouse and all possible combinations between these modes (cf. infra).

Type of revision (Van Waes, 1992, some are based on Faigley & Witte, 1981)

- addition
- deletion
- substitution
- permutation
- consolidation

Level of revision I & II

We categorized the revisions on two levels: a first level in which the writer actually wants to revise, and another level in which the writer actually revised. We categorized the levels: character, word, sentence segment, sentence, paragraph and punctuation (cf. infra).

Goal of revision

- surface changes (lay-out, punctuation)
- text-based changes

Distance of revision (remoteness)

- within sentence, within screen
- within sentence, outside screen
- within paragraph, within screen
- within paragraph, outside screen
- outside paragraph, within screen
- outside paragraph, outside screen

Direction of revision

We distinguished between backward and forward revisions. *Backward* revisions start before the level of inscription and *forward* revisions occur in the text after the point of inscription.

Absolute time, interval and writing phase

We calculated the absolute time of the revision in the writing process and afterwards we calculated the position of the revision in the writing process. Furthermore, we distinguished between revisions that were made within the first writing phase (first draft) and revisions that were made in the second writing phase and later.

4.1.3 Pauses

The analysis of the pausing behaviour is the fourth perspective to look at the data of this case study. Characteristics of the pausing behaviour of the writers enable us to gain a better insight of the cognitive planning processes. Or, to cite Matsuhashi (1982, p. 270): *‘Moments of scribal inactivity during writing reflect time for the writers to engage in cognitive planning and decision-making behaviour. (...) [They] offer clues to cognitive planning processes during written discourse production.’* Of course, we are aware that cognitive processes - might - also occur during speaking and writing.

Because of the hybrid character of the speech recognition mode we want to pay attention to pauses that are related to characteristics of the writing medium. Levelt states that the flow of speech is immediately interrupted when a problem is detected (Levelt, 1983). As for dictation (Schilperoord, 1996), this could also be the case for speech input with speech recognition. Our hypothesis is that in writing processes in which speech input is used, pauses will also mark the choice of writing mode. Especially when writers are confronted with technical problems caused by the speech recognition software (cf. supra), and are forced to revise their texts technically, they seem to hesitate whether or not to switch from the speech mode to the keyboard or mouse mode. Because we observed initial users of the technology, the cognitive activity preceding this decision is probably more explicit and takes more time. We think that during the adaptation process the participants will develop new strategies for dealing with technical problems and will be better able to predict (and avoid) this kind of problems. For the pause analyses we categorized the following elements.

Duration of pauses

Absolute length of every pause longer than 1 second.⁵

Number of pauses

Counting of every pause longer than 1 second.

Temporal location of pauses

The pausing times were divided as follows: absolutely, into 5-minute segments; and relatively, into ten equal units, whereby units 1-5 and units 6-10 are also referred to respectively as Part 1 and Part 2.

Linguistic location

Three locations were selected as relevant: within the sentence, at sentence boundaries, and at paragraph boundaries.

The above categories of writing modes, technical problems, revisions and pauses enabled us to describe the writing process data. But next to this we wanted to be able to illustrate the data by examples. So far, we have been able to describe the data but we are not able to show what actually happens. For this, we had to develop a transcription model. This model will be described in the next paragraph.

4.2 Transcription model

We developed a (preliminary) *transcription model* to represent the writing process in a multi-layered linear representation. This model is partly based on the S-notation as developed by Kollberg (1998). The S-notation is a format for representing revisions during the writing process (Eklundh & Kollberg, 1992). An important goal of the S-notation, for this study, is that the notation should be non-redundant and consistent. In other words, it has to be possible to derive the written text from the notation in an unambiguous way.

⁵ When writers produce a text by dictating there occur physical pauses. These pauses normally last less than .25 seconds, so we have excluded them from our data (Schilperoord, 1996).

As in the S-notation we use the following symbols for insertions and deletions: in the S-notation 'i' is the sequential number of the revised text. As we have showed in the categorization model, in our study we focus on technical problems and revisions. Therefore the 'i' in this study stands for sequential number of corrections of (technical) problems caused by the speech recognition software and the textual revisions made by the writer.

{inserted text} ⁱ	an insertion immediately following break i
[deleted text] ⁱ	a deletion immediately following break i
ⁱ	the break with sequential number i

Because of the speech recognition mode we distinguish between spoken and typed input. To represent these different types of input we have used different fonts: sanserif for spoken input and serif for keyboard&mouse input. Other activities than the normal production of text are also divided into subscript (voice input) and superscript (keyboard&mouse). Text deletions are visually marked because they are struck out. Commands, either with voice or keyboard&mouse are printed in italics (see figure 7).

Notation system

Spoken text

Typed text

{insertion with voice}ⁱ
{~~insertion with keyboard~~}ⁱ

[~~deletion with voice~~]ⁱ
[~~deletion with keyboard~~]ⁱ

<command with voice>
<command with keyboard>

|ⁱ start repair

/50\ pause time

Mode switch

Indicated by the movement of the switch:

 speech
 keyboard
 mouse

 selection by mouse
a.b.c. spelling mode

Movement by cursor

← cursor movement to the left
↑ cursor movement one line up
→ cursor movement to the right
↓ cursor movement one line down
◀X deletion with cursor to the left
X▶ deletion with cursor to the right

Figure 7 Notation system for writing processes with speech recognition

Speech recognition allows writers to produce and edit their text with voice input. Therefore, we also distinguished between spoken and written insertion, deletions and commands. The insertions are indicated with braces ({&}), the deletions are indicated by square brackets ([&]) and the command actions are indicated by tag brackets (<&>). Appropriate symbols indicate which writing mode is used and what movements the writers make through the text.

In the S-notation all writing actions can be read in a single text layer. This is impossible for reporting a writing process with speech recognition. Therefore we had to develop a multi-layered transcription model. We also chose to give a linear representation of the text. In layer 1 the spoken or written input are transcribed, including commands. Because the spoken input does not necessarily lead to a correct output we added a special layer in the model which shows what actually appears on the computer screen. This can be seen in layer 2. The third layer provides information about the writing modes one

uses and the pausing times. To give an indication of the time in the writing process, time stamps are put in a different column (for an example of the notation model see figure 8). Figure 7 gives a complete overview of the symbols used for the transcription model.

time	Data of what the writer actually says
	Data of what appears on the computer screen
	<i>Information about commands in the different modes (speech, keyboard, mouse) (revision)</i>
p 0.00.00	Dear mrs. Baker <comma> <new line> <new line> Employees who leave the company with ten years of employment are entitled both to the Dear mrs. Baker, ,,, and least ^leave the company with ^ten years of employment are entitled both to the •
p 0.00.21	company contributions and the retirement benefit payment deductions <full stop> ehm Those employees who leave the company with less than company contributions and the retirement benefit payment deductions. then dose ^employee ^sweeter company with less than / 0.00.25-0.00.29 \
p 0.00.32	ten years of employment will receive employee paycheck contributions made to their retirement accounts <full stop> ten years of employment will receive ^ply ^paycheck contributions made to their retirement accounts. ^ / 0.00.39 0.00.41-0.00.47 \ ^ select 'and
r 0.00.48	employees who leave those who leave the employee employees who leave those s who leave the employee least' • ^ <select then dose> ^ • ^ ^ply •

Figure 8 Example of transcription model

The example above shows a short segment of a business letter (for the 'original text' see figure 9). The writer starts in the SR-mode with the opening of the letter: 'dear mrs. baker'. The punctuation 'comma' and ... are also spoken. Two line spaces are inserted with voice command. Then two sentences are dictated, with a short pause of 4 seconds after the first sentence. After a pause of 7 seconds the writer switches to a mouse-input to select the first misrecognition 'and least'. With speech input the correct text 'employees who leave' is inserted. The writer continues in speech recognition to select the second misrecognition and corrects it to 'those'. Then he positions his mouse after the word 'employee'. After this word he places an 's' with keyboard input to make the word plural and he continues to delete the word 'sweeter'. To repair it to the correct word he switches again to speech input and dictates 'who leave the'. Finally he switches to the mouse mode to select the fourth misrecognition 'apply' and changes it with speech input to 'employee'.

In figure 9 the final text of this short transcription is shown.

Dear mrs. Baker,

Employees who leave the company with ten years of employment are entitled both to the company contributions and the retirement benefit payment deductions. Those employees who leave the company with less than ten years of employment will receive employee pay check contributions made to their retirement accounts.

...

Figure 9 'Original' text of the transcription

At the moment the elaboration of this short text segment is very time-consuming, because the transcriptions cannot be generated automatically. Consequently, the notation is merely meant to illustrate striking passages and not to describe full writing sessions. The linear representation of the notation also makes it difficult to interpret in longer texts because of recursive complexity.

5 Case study

In the case study we present in this article, we will focus on the adaptation processes of two participants, *Frederik* and *Bart*. This case study compares the writing and adaptation processes of two

participants who only had to adapt to the speech software because they both had experience in dictating as a lawyer.

Frederik, associate in a large law firm, has 5 years of work experience and about 6 years of experience working with computers and classical dictating devices. He claimed not to be patient when working with a computer. He indicated that he does not study new computer programmes, but that he does put extra effort and time into learning to work with them. Frederik expected that writing with speech recognition software would be easier and faster than the traditional way of writing texts.

Bart has a bit more working experience. He has been a lawyer for 9 years, but has a comparable level of experience in working with computers and dictating devices. Contrary to Frederik, he indicated in the questionnaire to be quite patient in learning to work with (new) computer programmes. Besides, he is willing to study the possibilities of a new software tool. Doing this he mostly uses a manual. His most important reason to start writing with speech recognition is his expectation that speech recognition is less tiring and more pleasant.

So, both participants had a comparable level of experience in working with computers and dictating devices. However the reason to select them for this case study was that they were characterized by a different learning style. Some of their answers in the questionnaire reported above already gave an indication for their differences in learning style (cf. Learning Style Inventory of Kolb 1984, as described in figure 3).

Frederik is, according to Kolb's taxonomy, an *accommodator*. He had a high score on active experimentation and concrete experience. Kolb's model states that people with an accommodative orientation tend to solve problems in an intuitive trial-and-error manner. They also rely heavily on other people for information rather than trust their own analytic ability. People with this learning style are sometimes seen as impatient and 'pushy'. As Kolb predicted, Frederik's learning style is determined by his job to the extent that he has to make a lot of decisions in uncertain and unpredictable circumstances.

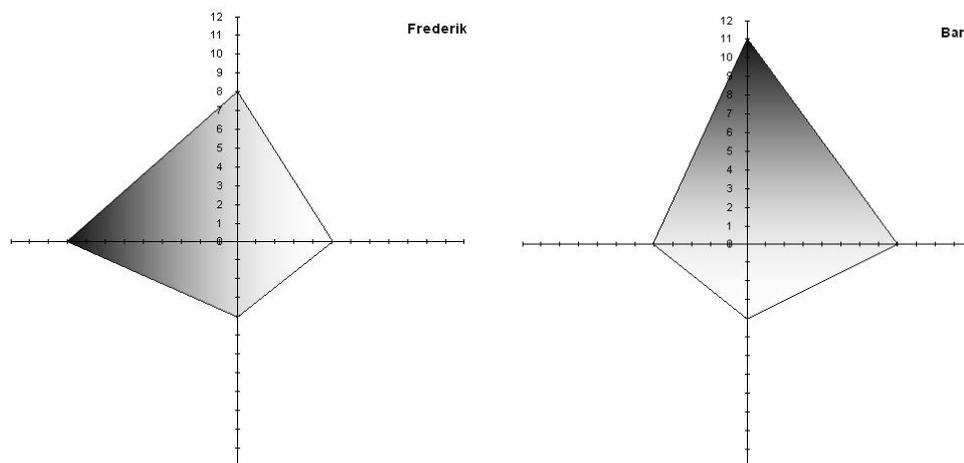


Figure 10 Graphical representation of the learning profiles of Frederik (left) and Bart (right)

Bart is a so-called *diverger*. He scored high on concrete experience and reflective observation. The emphasis in this learning style is on adaptation by observation rather than action. A person of this type performs better in situations that call for generation of alternative ideas and implications. Furthermore, they tend to be imaginative and they prefer to reflect on situations. Characteristics of his job that relate to his learning style are personal relationships and effective communication with other people.

Both participants have another learning style than Kolb predicted to be most adequate for learning to deal with (new) technology. However, since learning styles are adaptable to various tasks, we can predict that both participants will be – in some way – influenced by the task requirements of writing with a new writing medium. The active orientation of both participants will guide them to learn certain tasks and to improve active skills.

Both participants were also asked to describe their writing style for business texts (cf. design and procedure). Frederik and Bart both indicated that they hardly plan their texts in advance, but that during the writing process they often frequently pause briefly to plan the text. In the beginning of the writing process they plan more, but the further in the writing process, the less their writing processes are fragmented by pauses. They both predominantly reread on the paragraph level and then decide to revise or not. When writing a more difficult, non-routinely text, they indicate to have a different strategy. Frederik indicates that he also starts writing non-familiar texts immediately, while Bart indicates that he often takes his time to plan this kind of texts.

5.1 Writing product

In our description of the results we want to focus on two aspects (cf. infra objectives). On the one hand we want to compare the different writing strategies of the two participants and on the other hand we want to pay attention to the learning and adaptation process of the participants from observation session one to five.

During the observation sessions both participants wrote several texts in their own working environment. The different tasks the participants conducted were always functional tasks: letters, emails, conclusions, summons or reports. Because they were familiar with these kind of texts, the tasks could be performed in a structured and habitual way as Schilperoord states: *“Combined with knowledge of - in our case - legal dossiers, legal procedures and so forth, discourse schematic knowledge enables a lawyer to produce routine letters fluently and with remarkable ease. The text producer merely ‘tells’ what he already knows.”* (Schilperoord, 1996, p. 284)

The observation session differed in length because the participants were observed while working on a task they had chosen themselves. The first observation of both participants lasted about 5 minutes and the mean time of the remaining observations is 23’31” (Frederik: mean=18’05”, Bart: mean=21’18”). In this time Frederik produced an average of 17.76 words per minute and Bart 20.33 words per minute. Because of the variable length of the observation sessions, the texts also differed in length. In the first observation the participants explored the programme and wrote a very short text; during the subsequent observations their texts were between 247 and 685 words per session (Frederik: mean=355 words, Bart: mean=552,5 words). Table 1 gives a description of the five observation sessions.

Table 1: Characterization of writing product in the 5 observations

	Task	Length observation minutes : seconds	Number of words in final text	Mean number of words per minute
Frederik				
1	letter	04'08"	92	22.27
2	petition	22'17"	449	20.15
3	conclusion	21'23"	419	19.59
4	conclusion	23'33"	247	10.49
5	letter	19'46"	305	15.43
Bart				
1	e-mail	06'00"	62	10.33
2	conclusion	24'33"	586	23.86
3	petition	22'12"	520	23.42
4	report	17'25"	419	24.05
5	petition	37'00"	685	18.52

5.2 Writing modes

The mode analysis shows that both participants use the possibility of speech recognition quite differently (table 2). Frederik uses the speech input and the keyboard&mouse input almost just as much. The part of the speech mode drops gradually from the first observation session (47%) to the final session (34%). Bart on the other hand, uses the speech input almost twice as much as Frederik. For over 80% of his writing times, he uses the speech mode. This way of writing adheres to his traditional dictating habits and he hardly changes this strategy over the different observation sessions

Table 2: Mean use of writing modes in percentages (5 observation sessions)

	Frederik		Bart	
	mean (%)	sd	mean (%)	sd
speech	42,5	8,9	81,1	7,6
keyboard	44,4	13,9	13,8	7,2
mouse	13,2	9,4	5,1	2,2
total	100,0		100,0	

If we compare the use of writing modes between the first part of the writing process with the second (table 3), we notice that Bart prefers to write his texts with speech input and that this choice of writing mode stays quite homogeneous over the two writing parts. Speech is his preferred writing mode, also in the writing phase where not formulating, but revising is an important subprocess. Frederik on the contrary, uses speech input in the first part of the writing process for more than 50% of the time, but in the second part he definitely prefers writing with keyboard&mouse. The use of speech input to dictate text segments drops to a mean of 32.7 %. In the first observation session Frederik used speech for 40% of the process time in the second part of the writing process. In the last session however speech input almost completely disappeared from the writing process. He hardly uses it to finish his text (cf. *infra*). In the writing process of Bart we hardly see any changes in the use of writing modes in his writing sessions. His use of speech recognition remains quite constant (80%) in the different writing sessions and in both parts of the writing process.

Table 3: Mean part of speech mode in both parts of the writing process in percentages (5 observation sessions)

	Frederik		Bart	
	mean (%)	sd	mean (%)	sd
part 1	52,1	17,39	83,3	7,1
part 2	32,7	20,23	78,3	8,7

In table 4 we present the data of the amount of times the participants switch between the speech mode and the other writing modes. The total number of switches over the five observations is almost the same for both participants. Frederik switches 177 times and Bart 173 times. Consequently they switch one to two times per minute from speech input to one of the other writing modes (Frederik: 1.92 switches from speech per minute, Bart 1.65 per minute).

Table 4: Mean number of switches between speech mode and the other writing modes during the writing process (5 observation sessions)

	Frederik			Bart		
	number	%	sd	number	%	sd
Speech → keyboard	126	71,19	11,90	137	79,19	25,90
Speech → mouse	51	28,81	5,54	36	20,81	7,30
Total number of switches from speech	177	100,0	16,92	173	100,0	33,00

Although the total number of switches from speech to another writing mode is comparable for both participants, their writing behaviour in the five sessions differ. In the first sessions Frederik switches from writing mode more than twice as much as Bart. In the fourth session they reach the same level and switch about two times per minute between writing modes. In the last session the number of switches of Frederik drop below the number of switches of Bart. Frederik's modest use of the speech mode in the second part of the writing process is partly due to this.

Summarizing, we can say that the mode analyses confirm to a large extent the learning styles of both participants. Frederik, the accommodator, actively explores in the first sessions the possibilities of the speech recognition mode, but gives in after some trial and error. His use of speech recognition is rather restricted. Bart, on the other hand, as a diverger, uses the speech recognition system much more intensively and he explores the possibilities in several phases of the writing process. Table 1 shows that these explorations do not have a negative effect on the writing 'performance' Bart produces even more words per minute than Frederik [mean per session: Bart 20.33 (sd. 5.66) vs. Frederik 17.76 (sd. 4.72) words⁶ per minute].

5.3 Repairs

In total we observed more than 500 repairs (technical problems and revisions). Frederik interrupts the linearity of his writing process 306 times in five observation sessions and Bart has 223 recursive actions (Frederik, mean=61.20 per session vs. Bart, mean=44.60m see table 5). If we correct these data for the time differences between the observation sessions, this results in 3.45 repairs per minute for Frederik, as opposed to 2.09 repairs per minute for Bart. This probably seems much, but this number is more or less in line with findings of Levy & Ransdell (2002) in a study to 'concurrent memory loads'. They state that presenting a concurrent task too frequently (comparable to interruptions of e.g. technical problems) raises the possibility that the research participant will devote full attention to the secondary task, reducing the opportunities for proficient text generation. They determined that critical events for secondary tasks could occur about four times a minute which

⁶ This has been calculated on the number of words in the final text and therefore does not concern the total amount of words produced during the writing process.

seems to be in line with our observation on the number of times that Frederik and Bart are willing to interrupt their writing process.

Table 5: Part of speech input for repairs in percentages (5 observation sessions)

	Frederik		Bart	
	%	sd	%	sd
Speech	6.63	5.24	55.01	29.98
Keyboard&mouse	88.13	8.62	31.48	22.74
Combination	5.24	3.75	13.51	11.04
Mean number of repairs per session	61.20	33.64	44.60	26.27
Mean number of repairs per minute	3.45	0.97	2.09	0.25

The data in table 5 also show that both participants deal with repairs quite differently. As we have seen in the mode analyses, Frederik hardly uses the speech mode to solve technical problems or to revise (6.6%), while Bart tries to use speech recognition in 55% of the cases for recursive text interventions. If we add the repairs that are made with a combination of writing modes, including speech, the difference is even more extreme. Frederik does 90% of the repairs exclusively with keyboard&mouse. For Bart this percentage is 30. Bart chooses to explore the possibilities of speech recognition. During the first observation sessions Bart in 30 to 40% of the cases needs more than one attempt to solve a (technical or textual) problem. During the course of the observation session the solutions taking more than one attempt decline to less than 10%. Frederik, however, only in a few number of cases needs more than one attempt to solve a problem, because he prefers the safe approach of using keyboard&mouse to repair his text. So, he seems to be less eager to explore the possibilities of speech recognition for this purpose.

If we compare the proportion of revisions within the repairs on the one hand and the proportion of technical problems on the other hand (table 6), it is clear that the second category is the most extensive one (Frederik: 81.4% vs. Bart: 91.1%). That this pattern is typical to the adaptation process is proved by the number of technical problems in the last observation session. The number of technical problems in this session is the lowest (Frederik: 70.7% vs. Bart: 79.8%).

The analyses of the technical problems shows a different pattern for both writers. Frederik struggles more with the new technology than Bart. If we compare the technical problems per 100 words, we see that Frederik has to deal with almost one third as many technical problems than Bart (Frederik: mean=16.75 problems per 100 words vs. Bart mean=10.43)

Table 6: Repairs per minute and distribution between technical problems and revisions (5 observation sessions)

	Frederik		Bart	
	mean	sd	mean	sd
Problems	50.00	27.48	39.00	19.84
Revisions	11.20	9.42	5.60	6.88
Repairs	61.20	33.64	44.60	26.27
Technical problems relative to total number of repairs	81.39	9.61	91.09	8.02
Mean number of attempts per technical problem	1.27	0.93	1.37	0.88
Technical problems per 100 words	16.57	6.16	10.43	5.12

We can again relate this result to the learning styles of both participants. Frederik is fairly impatient and he is strongly focused on productivity. As a result, he keeps writing in his traditional way: dictating larger text segments with only short interruptions. He verifies the result of his dictations on the computer screen rather superficially. Consequently he is not able, or does not want to adjust his use of the speech recognition tool. Frederik does not train the speech recognizer consequently for a better recognition. He repairs his technical problems mostly in the second writing phase when he rereads the text produced so far (cf. *infra*; table 7), mostly without the use of speech recognition. Bart, on the other hand, uses the text that appears on the screen more consistently during the formulating phase. He reflects on the possible causes of the problems that occur (misrecognitions) and consequently tries to anticipate and to avoid these problems in the future. This strategy results in a lower problem ratio.

5.4 Revisions

It is striking that both professional writers revise very little in the five observation sessions. Partly we can explain this by the fact that they perform mainly routine tasks. Besides, the writing strategy of writers who dictate their texts is important: they preformulate mentally and therefore already partly revise their text before one word is dictated. A large part of the revisions of dictators takes place in the writers' mind. This causes the revisions from the screen to be limited and to be situated higher in the text hierarchy (Frederik: 43.21% of the revisions are above the word level vs. Bart: 73.87%; see table 5.7). Both writers partly maintain this writing style in the new writing mode, but we still see that the representation of the text starts to play an active role in the writing process. Contrary to the writing process that is typical for traditional dictators (Gould & Alforo, 1984; Schilperoord 1996) we now observe that recursivity becomes more apparent in the writing process of these writers.

Nevertheless, both writers also show a different revising behaviour. Frederik revises 3.41 words per 100 words vs. Bart: 0.93. If we assume that the mental revision is also reflected in the pausing behaviour of the participants, then we can add that factor as an explanation. Bart's pausing time is about 60% of his writing process, while for Frederik this amount is about 30% of the total time (see table 8). This distribution of pausing and writing remains constant for both writers over the five observation sessions. The difference in pausing behaviour is also in line with the learning styles of both participants: experimenting vs. reflective.

Table 7: Distribution of the revisions per level and per writing phase (5 observation sessions)

	Frederik		Bart	
	%	sd	%	sd
Level				
word	56,79	13,52	26,13	26,94
sentence	35,90	5,73	35,80	44,67
higher	7,31	11,08	18,07	31,35
Writing phase				
phase 1	42,95	30,01	75,29	43,31
phase 2	57,05	30,01	4,71 ⁷	10,52

⁷ This percentage is in total 80% because in the first observation session Bart does not revise his text.

Also the distribution of the revisions in the two writing phases differs for both participants (table 7). Frederik is a writer who – in accordance with his previous dictating experience – mainly revises in the second phase of the writing process, after he has finished a first draft of his text. Approximately 60% of his revisions are in the second phase. Bart, however, hardly revises in the second writing phase, only 5%. This kind of revising behaviour matches closely the writing profile of the professional writer who uses a word processor to write his texts. These writers also prefer to revise their texts during the writing process of the first draft, leaving only few revisions for the rereading phase.

5.5 Pauses

Many researchers analyzed pause patterns (e.g. Matsushashi, 1981, Schilperoord, 1996) to study writing processes. Pauses can be seen as indicators of cognitive processes like planning and reviewing. For the analysis of the pauses, all pauses longer than 1 second were timed using the on-line video reconstruction of the writing process. In total 1238 pauses were registered. Frederik paused 517 times in the five observations and Bart 721 times (see table 8).

Table 8: Mean number and length of pauses over 5 observations

	Frederik n=517		Bart n=721	
	mean	sd	mean	sd
number of pauses	103,50	51,00	141,70	85,00
length of pauses in seconds	3,70	1,20	6,00	1,60

Frederik pauses on average 103 times per session, while Bart pauses about 142 times. Frederik's pauses are almost equally distributed over the different writing sessions, but if we divide the writing session into two equal parts, we see that Frederik pauses more often (61.7%) in the first part of the session than in the second part (38.3%). Bart pauses are almost equally distributed over the different writing sessions and in both parts of the writing process (first part 49.3% vs. 50.7% in the second part).

We also calculated the mean length of the pauses (table 8). This analysis shows a difference between the two participants. Frederik's pauses are shorter than Bart: Frederik mean=3,70; sd. 1,20 – Bart mean 6.00; sd. 1.60 ($F(1,1237)=58,64, p<.000$). The total pausing time of Frederik is 32.3% of the total production time; Bart's pausing time takes up to 61.7% of the production time. This is almost double of the time Bart uses for planning.

From these analyses we can conclude that Bart more often pauses than Frederik and that his pauses are much longer. Frederik's writing process is a bit more fluent than Bart's. To illustrate this, we have calculated a fragmentation ratio representing the total number of pauses divided by the total writing time. Frederik's ratio is 11.13⁷ and Bart's 9.79, being highly comparable. Bart does not interrupt his

⁷ The fragmentation ratio of the second observation session of Frederik is much lower than in the other sessions, because of the

writing process as often as Frederik, but when he does interrupt his session he interrupts it for a significantly longer period of time. An explanation for this difference could be that Bart tries to repair his text with the use of speech recognition, as we have seen in the repair analysis. Because he is unfamiliar with speech recognition, this strategy will take up some of his time. That this explanation is reasonable can also be inferred from the development of the number of pauses and their length over the different writing session. For Bart the number of pauses increases over the different sessions, but their length drops. If we take a closer look at the data, we also see that in the later sessions he repairs problems and revisions in fewer attempts than in the first sessions. In conclusion, we can say that Bart is developing a successful strategy: in the beginning it took him a long time to solve a problem, further in the observation sessions, however, he becomes more successful and his pauses therefore become shorter. Frederik's number of pauses remain stable in the last three observation sessions and the length also consistently fluctuates around 4 seconds. This can be seen as evidence for the fact that Frederik is satisfied by his way of working with speech recognition: he uses it for dictating, but to repair his text he switches to the keyboard&mouse mode and reviews his text in this mode.

6 Conclusion

In this case study we have proposed an analyses of two professional writers experienced in dictating, who work for the first time with speech recognition as a dictating device. Both writers are characterized by a different learning profile (accommodator vs. diverger). In this paper we wanted to learn more about the cognitive aspects of learning to write with a new writing mode. How do writers adapt to a new writing mode? And we wanted to find out how learning styles influence the use of speech recognition for the first period of time.

In the mode analysis we have seen that Frederik explores the possibilities of the new writing mode actively in the first writing sessions, but that he opts rather fast for a selective and modest use of the speech mode. He uses speech recognition mainly to dictate his text and uses it less to solve technical problems or to revise his text. This changing pattern confirms the behaviour that is characteristic of accommodators. Initially he switches a lot between the different writing modes, but in the last sessions the number of switches is remarkably lower. Moreover, during the completion of the text Frederik 'evolves' in his use of speech recognition to a writing pattern that is comparable to the writing processes of writers using classical dictating devices. He dictates his text and then adjusts his text in the second writing phase by using keyboard&mouse. The text on the screen has a more passive, monitoring function and not a guiding function.

Bart explores more systematically and more patiently the different possibilities of the new writing mode. His writing process is very reflective and is characterized by long pauses. During these pauses Bart rereads actively the text produced so far on the screen and when necessary he repairs his text. If possible he tries to work with speech input, limiting the number of mode switches. Bart uses the speech mode during 80% of his writing time, which is twice as much as Frederik. In Bart's case speech recognition is definitely not an elevated dictating device, but more a hybrid writing tool that combines some elements of the classical dictating devices with the classical computer mode (keyboard&mouse). In the adaptation process of Bart we hardly see any quantitative movements between the writing sessions. His profile remains stable over the different sessions. Being a 'diverger' he explores in different ways the possibilities of speech recognition and therefore he develops through reflection and observation, a way of working with speech recognition satisfactory (22 words per minute). In the studies of Karat (2000) and Halverson (1999) the initial users produced 13.6 words per minute (keyboard&mouse=32.5 words per minute) and the four experienced users who worked for 30 hours with speech recognition produced 25.1 words per minute.

enormous amount of repairs he has to conduct (observation 2: number of repairs=103, mean repairs=61,2). A possible explanation could be that when a writer encounters too many problems in a text, he prefers to 'save up' all these problems to solve them at once at a later moment in the writing process. This strategy results in fewer pauses and interruptions in the first phase of the writing process.

Table 9: Characteristics of the adaptation and writing process of Frederik and Bart

	Frederik	Bart
learning style	accommodator	diverger
use of speech mode	active exploration selective use (mostly formulating)	systematic exploration dominant use (formulating and revising/repairs)
adaptation process	gradual evolution of hybrid use to a more traditional dictating process	continue adaptation and hybrid use
revision behaviour	mainly in the second writing phase	mainly in the first writing phase

These descriptions show that both participants have a different adaptation process, which evolves differently over time and seems to be driven by the learning styles of the participants. Consequently the speech mode as input for the writing process has a different effect on the writing behaviour of both writers.

7 Discussion and further research

In this paper we have described a case study that has been selected from a larger corpus of observations. In Leijten (2003) a case study has been described in a similar way. In that study the writing processes of two writers were analyzed. However, they did not differ in learning style, but differed in previous dictating experience. This study showed that the speech recognition mode does not impose a specific writing style to the writer, contrary to what previous studies show about pen and paper writers. Most studies showed that the participants had to adjust their pen and paper writing style to the new medium. Both studies confirm the potential hybrid character of speech recognition as a writing mode. They also show that the feedback on the screen about the text produced so far also can influence the writing process, but that the extent to which writers use this feedback while dictating their texts can be very different. The descriptions of the case studies show that the chosen observation instruments and analyzing methods enable us to analyze and describe the specific aspects of the writing processes.

This study showed that we can focus on several variables that are important to answer our main research questions. For the learning strategy we will focus on problems and solutions, because there you can see how the participants learn to deal with the speech recognition software. For the writing strategy we will mainly focus on pausing times and revisions. Pausing times provide rich information on the production process of writing. In a next study we want to present a more quantitative analysis of the available data of the total corpus of this research project. First we will apply the different aspects of the categorization model in this study to the complete dataset. A coordinated processing of this data must enable us to confirm the findings of the case studies and to deepen these results. Next to applying the same variables in the analyses, we can focus on different aspects, as seen in the categorization model.

We think that the categorization model we have developed for this study is an interesting way to look at complex data and creates a good framework to compare different learning and writing strategies. Different patterns occur when comparing the results. Some users of speech technology do explore and adapt to the writing mode, others maintain their familiar writing strategies (also Leijten 2003). In this study we can see that the participants adapted (or did not) their writing strategy to speech recognition. However, to get a complete picture, it would be interesting to compare the writing modes speech recognition vs. keyboard&mouse or the classical dictating devices in a more controlled way. Only then will we be able to really compare the writing processes as opposed to the mode writers were originally used to. In this study participants produced different text types, varying from a short routine letter to a complex summons. Besides, their writing session were of different length. In

this study these differences were not a problem, because we wanted to create a really high ethnographic validity. We preferred letting the people work in their own environment and let them develop their own learning and writing strategy. We assume, however, that these analyses form a solid ground to set up a controlled experimental study in which we will observe writers while writing texts on the basis of a clear task in different writing modes.

As stated above, the analyses of the data were very time-consuming. Therefore, we are developing more sophisticated observation instruments, which would make the analyses of several categories faster and more accurate⁸. These tools will make it easier to describe the data and will make it possible to analyze the data more precisely on a larger scale.

Finally, we would like to mention that the purpose of this research project is not to focus on the speech recognition as a technical device. The analyses in this study show that the focus on a new writing mode can enable us to gain a better insight into the specific aspects of cognitive processes that are responsible for the development of writing processes. This line of approach will be even more central in further research⁹. We hope to report on this in the near future.

Notes

This doctoral research is supervised by Prof. Dr. P. van den Hoven, Dr. D. Janssen en Dr. L. Van Waes.

This study is part of a NOI research project on the influence of speech recognition on the writing process (Research grant of the University of Antwerp 2000-2002).

Acknowledgement

We would like to thank both participants Frederik and Bart for being willing to learn to work with speech recognition and let us observe them, while doing so. We are grateful to Lernout & Hauspie who gave us a free license to work with VoiceXpress Legal™ during the course of this study.

Literature

- Eklundh, K. & Kollberg, P. (1996). Computer tools for tracing the writing process: from keystroke records to S-notation. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (eds), *Theories, Models and Methodology in writing research*. Amsterdam: Amsterdam University Press (pp. 526-541).
- Eklundh, K.S. (1994). Linear and Non-linear strategies in computer-based writing. *Computers and Composition*, vol. 11, 203-216.
- Eklundh, K.S. & P. Kollberg (1992). *Translating keystroke records into a general notation for the writing process*, IPLab, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm.
- Faigley, L. & S. Witte (1981). *Analyzing revision*. *College Composition and Communication*, 32/4, 400-414.
- Flower, L. & J. Hayes (1981). *A cognitive process theory of writing*. *College Composition and Communication*, 32, 365-387.
- Gould, J.D. & L. Alfaro (1984). Revising documents with text editors, hand-writing recognition systems and speech-recognition systems. *Human Factors* 26, 4, 391-406.
- Gould, J.D. (1978). How experts dictate. *Journal of Experimental Psychology: Human Perception and Performance* 4, 4, 648-661.
- Halverson, C.A., Horn, D.B., Karat, C. & Karat, J. (1999). *The beauty of errors. Patterns of error correction in desktop speech systems*. IOS Press, 1-8.
- Hartley, J. Howe, M. & McKeachie, M. (2001). Writing through time: longitudinal studies of the effects of new technology on writing. *British Journal of Educational Technology*, 32/2, 141-151.
- Hartley, J., Sotto, E. & Pennebaker, J. (2003). Speaking versus typing: A case-study of the effects of using voice-recognition software on academic correspondence. *British Journal of Educational Technology*, 34/1, 5-16.

⁸ Existing logging-tools as Trace-it (Eklundh & Kollberg 1996) or ScriptLog (Strömqvist & Malmsten, 1997) are either developed for a specific environment, or not adequately adapted to the current Windows environment. A combination with the speech mode also is not possible in the current tools.

⁹ The author of this paper is writing her doctoral thesis about the influence of different writing modes on the writing process (for further information see <http://www.ufsia.ac.be/~mleyten>).

- Hayes, J.R. & L.S. Hayes, (1980). Identifying the organization of writing processes. In L.W. Gregg & E.R. Steinberg (eds.): *Cognitive Processes in writing*. (p. 3-30) Lawrence Erlbaum Associates Inc., New Jersey.
- Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In C.M. Levy & S. Ransdell, *The science of writing: Theories, methods, individual differences, and applications*. (p. 1-27) Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.
- Honeycutt, L. (2003). Researching the use of voice recognition writing software. *Computers and Composition*, 20, 77-95.
- Karat, C., Halverson, C., Horn, D., & Karat, J. (1999). *Patterns of entry and correction in large vocabulary continuous speech recognition systems*. Paper presented at CHI 99, Pittsburgh, PA (Proceedings). ACM Press, 568-575.
- Karat, J., Horn, D.B., Halverson, C.A. & Karat, C. (2000). *Overcoming Unusability. Developing strategies in speech recognition systems*. Poster presented at CHI 2000, Den Haag, Nederland (Proceedings). ACM Press, 141-142.
- Kolb, D.A. (1984). *Experiential Learning*. Englewood Cliffs, New Jersey, Prentice Hall Inc.
- Kollberg, P. (1998). *S-Notation - a computer based method for studying and representing text composition*. Stockholm, Stockholm University.
- Lee, Young-Ju (2002). A comparison of composing processes and written products in timed-essay tests across paper-and-pencil and computer modes. *Assessing Writing*, 8/2, 135-157.
- Levelt, W.J.M. (1983). Monitoring and self-repair in speech, *Cognition*, 14, 1, 41-104.
- Levy, C.M. & Ransdell, S. (2002). Writing with concurrent memory loads. In T. Olive & C.M. Levy (eds), *Contemporary tools and techniques for studying writing*. (pp. 9-30) Dordrecht: Kluwer Academic Publishers, Studies in writing, volume 10.
- Leijten, M. (2003, to appear). How do writers adapt to speech recognition? The influence of previous dictating experience on the writing process in speech technology environments. *Recent developments in writing process research*. Volume 2: methods and applications, Studies in writing, Kluwer.
- Matsuhashi, A. (1981). Pausing and planning: the tempo of written discourse production. *Research in the teaching of English*, 15(2), 113-134.
- Matsuhashi, A. (1982). Explorations in real-time production of written discourse. In M. Nystrand (Ed.) *What writers know. The language, process, and structure of written discourse* (p. 269-290). New York: Academic Press.
- Mazeland, H. (2003). *Inleiding in de conversatieanalyse*. Uitgeverij Couthino b.v., Bussum.
- Quinlan, T.H. (2002). *Speech recognition technology and the writing processes of students with writing difficulties*. Paper presented at Sig Writing 2002, Stafford, England.
- Reece, J., & G. Cumming (1996). Evaluating speech-based composition methods: Planning, dictation, and the listening word processor. In C.M. Levy & S. Ransdell. *The science of writing: Theories, methods, individual differences, and applications*. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, 361-380.
- Rodman, R.D. (1999). *Computer speech technology*. Artech House Publishers, London.
- Schegloff E., Jefferson, G. & Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language*, 53, 361-382.
- Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production*. (Proefschrift Universiteit Utrecht). Amsterdam, Atlanta: Rodopi.
- Stifelman, L.J. (1993). *User repairs of speech recognition errors: An intonational analysis*. MIT Media Laboratory Technical Report, Massachusetts, p. 36.
- Strömqvist, S. & Karlsson, H. (2001). *ScriptLog for Windows: Users manual*. Lund: University of Lund.
- Terrell, S. (2002). The effect of learning style on doctoral course completion in a Web-based learning environment. *Internet and Higher Education*, 5, 345-352.
- Van Waes, L. (1991). *De computer en het schrijfproces. De invloed van de tekstverwerker op het pauze- en revisiegedrag van schrijvers*. Enschede, Universiteit van Twente.
- Van Waes, L., & P.J. Schellens (2003). Writing Profiles. The effect of the word processor on pausing and revision patterns of experienced writers. *Journal of pragmatics*, 35, 4, 829-853.
- Wood, E., Willoughby, T. Specht, J. & L. Porter (2002). An examination of how a cross-section of academics use computer technology when writing academic papers. *Computers & Education*, 38, 287-301.

Appendix: Notation system

Spoken text

Typed text

{insertion with voice}ⁱ
{insertion with keyboard}ⁱ

[deletion with voice]ⁱ
[deletion with keyboard]ⁱ

<command with voice>
<command with keyboard>

|ⁱ start repair

/₅₀\ pause time

Mode switch

Indicated by the movement of the switch:

🗨 speech
⌨ keyboard
🖱 mouse

🖱 selection by mouse
a.b.c. spelling mode

Movement by cursor

← cursor movement to the left
↑ cursor movement one line up
→ cursor movement to the right
↓ cursor movement one line down
⊠ deletion with cursor to the left
⊡ deletion with cursor to the right

Unintentional actions

∅ hum, cough, sniff etc.
ϕ unintentional action by computer
↵ a mistake in speaking