Voice recognition software as a real tool during rehabilitation processes

TERESA GANDARA-SAMBADE, MARIA CARMEN SOUTO-SUAREZ, JOSÉ A. OROSA
Department of Energy and M.P
University of A Coruña
Paseo de Ronda, 51, 15011. A Coruña.
SPAIN
jaorosa@udc.es

Abstract: - In the past few years, an increase in the number of illnesses due to incorrect ergonomic working conditions have been observed. In particular, in places of work like computer stations a clear relationship was observed between computer operating habits and pain in the back, arms and hands. In this sense, the illnesses ascribed to these bad working conditions are related with tendon inflammation, medically called tendonitis. Solutions for this illness are inexpensive like, for example, changing to more ergonomic furniture and the massage treatment but it is a temporal solution and the problem will reappear as soon as the treatment ends. In consequence, a realistic solution for this problem narrows down to a different interface between the operator and the computer. Among all the options defined over the past few years, the speech recognition system is one of the most important due to its low cost and its fast training. Dragon Naturally Speaking and Microsoft Word software resources will be tested in a real case study to analyze which are the better innovations that can be employed in everyday working conditions that result in this illness. Results showed its unsuitability to be employed in real life, and the lack of information and training of students in graduate and postgraduate studies about this useful tool. In particular, if we think that in the near future voice recognition will reach a status so as to be applied in a higher percentage of its expected applications, this tool could be of greater interest.

Key-Words: - voice recognition, tendonitis, software, education, rehabilitation, tool

1 Introduction

Present research work will show one possible solution for tendonitis due to incorrect ergonomic working conditions. In the initial stage, we must understand the behavior of tendons. Training of tendons and ligaments take place along with muscle contractions, but their development is more slow and may take several months. Training increases the amount of collagen fibers and, with this, tendons become thicker and stronger.

Therefore, in the strength-training programs it is necessary to start with a long period of work in which we work with very low load and high repetitions. This type of work will enhance the development of tendons and ligaments without stimulating excessive muscle development. In this way, we get stronger joints and tendons to address a subsequent strength program.

We can now understand how injuries occur in tendons, as was commented earlier. Tendons recovery very slowly if injured [1]. This limited capacity for growth and regeneration is due to poor feeding and blood flow with a situation that we also found in cartilage and scarring. For this reason, they also weaken rapidly with inactivity and aging. In fact, between 20 and 30 years of age, this situation is more pronounced, and tendons lose traction and flexibility, and are exposed to breakage more easily. Despite this, it is on crossing 30 years of age when men begin to frequently use their tendons in professional work.

The problem appears when a worker notes the first symptoms and does not experience sufficient pain to stop working, and cannot find any way to reduce the workload to their arms. Under this situation, the worker continues working till the pain is strong enough to enable him to employ a computer.

Over the past few years, an increment in the number of illnesses due to incorrect ergonomic working conditions was observed. In particular, in computer-related work, like in an office, where a clear relationship between computer working habits and pain in the back, arms and hands was

demonstrated. Illnesses related with these bad working conditions cause tendon inflammation, which is called tendonitis. Solutions for this illness are inexpensive like, for example, changing the table, chair, mouse and computer to more ergonomic furniture. In particular, most medical recommendations to prevent this kind of illness are, for example, that the keyboard and mouse must be between 3 and 8 cm below the horizontal elbow, trying to keep the wrist straight. In other cases, solutions are treatment with massage and some muscular relaxants but, despite this, it is a temporal solution and the problem will reappear as soon as the massage treatment ends, because if the situation is allowed to pass, the pain will be greater.

Finally, other solutions like proposing that each worker must get up from his seat every 30–60 min, relax the neck and back, stop using the traditional mouse and instead use a graphics tablet or a trackball, are some solutions usually proposed, but these only defer the problems. As a consequence, like with other kinds of illnesses [2], a realistic solution for this problem lies in a totally different interface between the person and the computer [3, 4]. Among all the various options, the speech recognition system is among the most important due to its low cost, and its configuration time with training processes is very fast.

2. Materials and methods

In accordance with statistical results of previous research work [5], the main software resources that were initially proposed to be employed in this case study are Dragon Naturally Speaking or Microsoft Word Speech Recognition Software [6, 7, 8, 9]. This last feature is available in versions of Microsoft Office Simplified Chinese, Traditional Chinese, English (US) and Japanese. Speech recognition is installed in all Office programs by using this feature in Microsoft Word, or performing a custom installation.

When a speaker starts using voice recognition in an Office program, he must follow the next procedure to train the voice recognition. First of all, the software must be installed and, once it is installed, it will be available on the Tools menu in all Office programs.

When the speaker wishes to train his voice recognition, he only has to employ the language bar and to configure the microphone. This training process takes approximately 15 min which is considered enough to improve accuracy. More options are typically employed like, for example, to

load and delete words as pronounced by the speaker, as we can see in Figures 1 and 2. This option will allow a worker to develop a library of terms in accordance with the specific vocabulary that each speaker tends to use in his daily work.
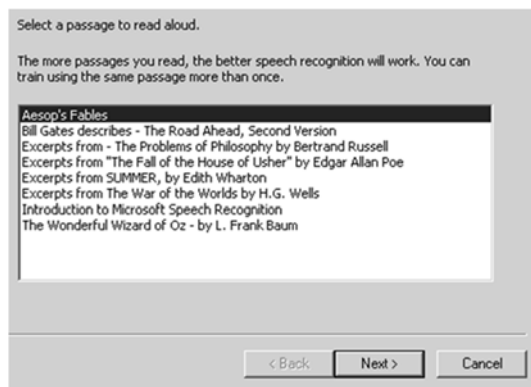

Figure 1. Speech recognition training session.

Figure 2. Add/delete words menu.

Figure 3. Speech input settings.

Teresa Gandara-Sambade,
Maria Carmen Souto-Suarez, José A. Orosa

Another important option that we can see in the speech tool of Microsoft Word is the speech input settings tool, shown in Figure 3. In particular, one of the more important settings is the assign mode keys option, as we can see in Figure 4.
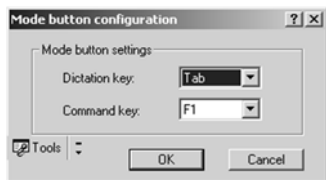

Figure 4. Mode button menu.

In Figure 4, we can see that to activate dictation, we only have to press the "Tab" button. Furthermore, we have to press the "F1" key to activate commands. The selection of these two last buttons is based on the fact that these buttons are hardly ever employed in the usual application of the keyboard. Furthermore, with only these two buttons, initial dictation can be developed. Despite this, when there is any kind of mistake in the dictation process, the last sentence can be undone with the command "Ctrl+Z", as it is specified in most Microsoft applications by default.

Another similar software resource is Dragon Naturally Speaking [10]. This is a voice recognition software that can recognize single words, but in experiments and in theory, shows a real improvement with sentences reaching accuracy rates of 99%. Like in the previous software resources, this needs a training period from some passages and correction of errors. This software, at the same time, supplies capitalization, apostrophes and spacing between words, and the most important and typical commands are also incorporated.

Figure 5. Speech recognition commands.

There are more software resources that can be employed for these same applications, like e-speaking that are based on Microsoft [11]. Despite this, as we can see in Figure 5, this voice recognition software employs typical commands related with the main applications.

Main commands employed in most of this software are shown in Table 1.

cancel	computer	maximize	powerpoint	spacebar
close	edit copy	minimize	open word	start
close ?	edit find	next	print	start
menu	edit paste	window	release	dictation
close	edit undo	open	item	start email
application	file print	open ?	replace	start excel
close	find next	file	replace all	start
document	import	open	resume	internet
close	commands	email	save	start
window	listen to	open	search and	notepad
	me	excel	replace	start
		open	select	powerpoint
		internet	select all	start word
		open		stop
		notepad		stop
		open		listening

Table 1. The commands currently defined in Voice Recognition software resources.

Owing to the feasibility of Microsoft voice speech recognition being employed in all office software, it was selected for text dictation and, at the same time, e-speaking was selected for commands dictation. This last software was selected despite the fact that it is based on the same Microsoft voice recognition technology, and is specially designed in commands recognition. Despite this, it can be employed for common dictation as well. Both software resources will be tested in a real case study to analyze which are the better tools to improve the working conditions under the effects of this illness.

Like in previous works [12], a study of voice recognition software was conducted comparing the two methods; typing on a keyboard and dictating text. In the initial stage, different parameters that allowed us to improve the actual resources were defined and analyzed.For example, one of the main parameters that must be tested is the training time. It is the minimum number of words that must be pronounced by the speaker to train the software. Despite this, we must remember that only the software influences this timing and other parameters, like the kind and configuration of microphone employed, and the ambience where this is recorded can influence the final result.

Another parameter that can improve the final results is to employ the self-learning tool from one text developed by this method. With this option, the

software learns the particular pronunciation of a word by one speaker from a previously dictated text.

A few more parameters that must be considered are the language. Sometimes dictation is not done by native speakers and, in consequence, the software must be adjusted to the particular accent of the speaker, as it was analyzed in previous works [13].

We must taking into consideration that in some works the reading levels of students of different ages was analyzed with these same kinds of software resources [14, 12].

The main results of those studies showed that a student could input text more quickly using voice dictation compared with keyboard entry, and that the accuracy was acceptable. Furthermore, these studies showed that voice recognition can be improved with practice.

In our real case study, we have proposed to develop a more in-depth study to develop the tests and comparing results of dictation, keyboard and previous works to obtain the best comparison that we can.

The first method consists of dictation by non-native English speakers, employing Microsoft Word and, on the other hand, its results were compared with the readings. During each process, the number of words with errors was counted. At the same time, the percentage of errors per unit of time was calculated as a new parameter that shows the quality of each of the different processes. Results have been shown in Figures 6 and 7.

During training, the keyboard buttons were defined letting us start and stop the dictation process. In consequence, to develop this dictation task, and thinking about the application via the keyboard in daily working tasks, selection of two keyboard buttons to start and stop dictation was done.

In accordance with the previous section, buttons like "Tab" and "F1", among others, have been selected for this test. The reason why we have selected these buttons is that they are not usually employed when working and, in consequence, cannot cause any damage to the work.

On the other hand, another interesting application of voice recognition for computer commands was tested after the training. The main commands correspond to the operative system, and the ones related with each specific applications that can be uploaded.

3. Results

In the initial stages, a text employed in previous research works [12, 15] was selected to be dictated by workers as is shown in Table 2. This text had been employed in this research work to define the reading improvement of young students, and now its results will be compared with a non-native English speaker's dictation results.

The reason why we have selected this parameter is because it shows an intermediate level that is better than writing with keyboard and dictation. In this sense, the number of recognition errors from the first reading and dictation to the last one was obtained, as we can see in Figure 6.

Pat shut the door of the warm kitchen behind her. Outside the air was like ice. It took her breath away for a moment.
All around her the white garden sparkled and shone in the pale winter sunshine.
Pat blew on her hands to warm them. Then she walked across the frosty grass to the pigeon cage that stood by the back fence.
She stood in front of the cage, watching the pigeons.
There were twelve pigeons. Some were grey with long black or white feathers in their wings. The others were brown, speckled with white.
Their turned their heads from side to side and looked at Pat, first with one dark eye, then with the other.
The pale winter sun shone through the wire HYPHEN netting and over the nesting HYPHEN boxes. Some of the birds were sitting on their nests.
The others walked up and down the cage, cooing and bobbing their heads.
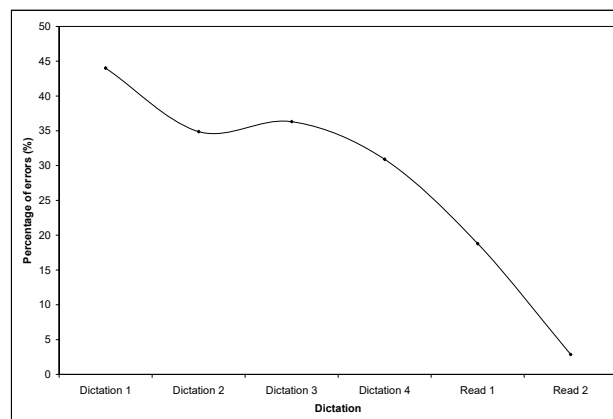
Table 2. Main text employed in dictation.

Figure 6. Percentage of errors during dictation and reading a text.

The main commands typically defined in the database of all these software resources, shown in Table 1, were dictated into a computer consecutively to obtain the accuracy of voice recognition. Furthermore, this dictation test of all commands was done 20 times, and the number of times that the commands were not recognized was

counted. In this case, the percentage of errors for each different command is reflected in Figure 7.
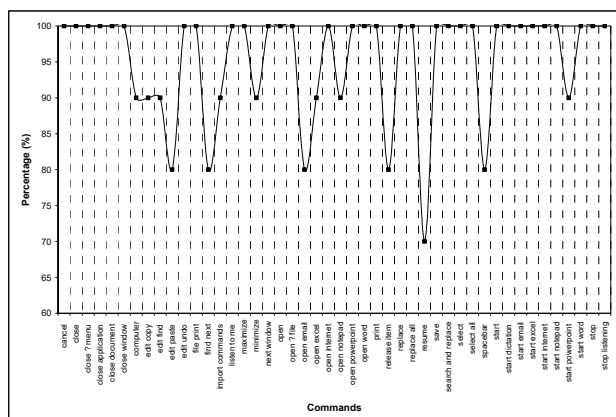

Figure 7. Percentage of errors of voice recognition software in commands dictation.

4. Discussion

The present work shows one tendonitis illness solution in two possible ways; the first is to employ the voice recognition software to dictate texts, which implies a great improvement in voice recognition of groups of words and sentences. On the other hand, the other test was designed to employ the voice recognition software to recognize some kind of fundamental commands that a computer can easily execute, being a very useful tool for persons with some kind of handicap.

As we can see in Table 1, the main commands employed in these software resources are similar to the common functions of Microsoft Office with options like copy, paste and save. These kinds of commands are usually employed with a high frequency in the typical tasks of workers.

If we analyze Figure 7, we can conclude that these commands are easily recognized by a computer with good percentages of recognition, after 20 attempts that goes from 70–100%.

In this sense, commands like "resume" showed the worst recognition percentage by the software. Another intermediate level of recognition was obtained with commands like edit paste, find next, open e-mail, release item and space bar. Finally, the most easily recognizable commands like close, open and start were always recognized by the computer (100%).

The reason why some commands are easier to be recognized can be related with the fact that the command employs only one word. Furthermore, these words are typically employed in basic English and, in consequence, are pronounced in a more

adequate way by non-native English speakers and are easily recognized by the computer.

However, when dictation is developed with long texts, the computer needs to recognize the combination of each word sound in a reduced time. In consequence, a very low percentage of recognition was obtained, as we can see in Figure 6. Despite this, as was obtained in previous research works [12], dictation recognition experiments were an improvement when the same text was dictated several times. In consequence, the higher percentage of errors reached 45% of words in the first reading, and the minimum number of mistakes with dictation was reached with 30% of errors. These values are much higher than when the text was read to students. In this last situation, it has reached values of 15 and 2% from the first to the last reading, respectively.

From these results, we can conclude that dictation of long texts have a higher percentage of errors than simple commands. In particular, commands that are simpler are composed of few words, and in particular, words more frequently employed. In consequence, these kinds of software resources present a higher usability as command recognition than for text recognition in daily work conditions of engineers and officers.

An interesting result obtained was that to improve text recognition software, speakers must be trained in the pronunciation of words. To do it, the pronunciation of the common vocabulary from each kind of word must be improved, particularly, if the speaker is not a native English speaker. As a useful tool for this problem, the adequate pronunciation of each sentence, once it was written by the keyboard, could be listened to by a speaker by using his own Microsoft speech option reading tool.

Another interesting result obtained from this practice was that selection of two keyboard buttons to start and stop dictation is a very useful methodology than to do it through dictation. In particular, buttons like "Tab" and command buttons like "F1" have been selected. The reason why the practice showed the need to select these buttons aims at the fact that they are not usually employed when text is being typed. In consequence, as a future research about voice recognition software, it is necessary to find some buttons that must never be employed in daily work to start and stop dictation.

Another important result obtained was that when we try to dictate a text and some errors of voice recognition appear, the quickest way to go back and

try the dictation of words again, is to employ the keyboard command "Ctrl+Z", as it is specified in most of Microsoft Word applications by default. This implies pressing two buttons at the same time and this can be a problem for speakers with handicaps. In consequence, an extremely urgent future work is to propose new keyboards or mouse buttons that let us control dictation in an easier way.

Finally, mean results showed that these kinds of applications can be employed in real life by computer workers in their daily tasks. Furthermore, these kinds of applications are the only way to continue working after this kind of illness appears. Owing to the usability of this methodology of future engineers, officers and, in general, all kinds of computer workers [16, 17, 18], this methodology must be learnt in graduate and postgraduate studies as a complement to programming languages subjects, and be improved upon, based on new technologies [14, 19, 20].

5. Conclusions

The present work shows one possible solution for tendonitis illness in two ways; the first is to employ the voice recognition software to dictate texts and to recognize some kind of fundamental commands that a computer can easily execute.

From this study, we can conclude that commands are easily recognized by a computer, and that the reason why some commands are easier recognized than others can be related with the fact that the command employs only one word that is typically employed in basic English and, in consequence, is pronounced in a more adequate way by non-native English speakers and to be easily recognized by a computer.

Another conclusion is that when dictation is developed with longer texts, a very low percentage of recognition was obtained due to the computer's need to recognize the combination of each word sound in a reduced time. Despite this, as was obtained in previous research works [12], dictation recognition experiments are an improvement when the same text is dictated several times.

Some interesting improvements obtained during the testing periods of this technology were obtained. The first proposal is that speakers can be trained in the pronunciation of some words. To do this, the pronunciation of the common vocabulary from each kind of word must be improved by one's own speech option reading tool of a text once it has been previously written.

The second proposal is the selection of two keyboard buttons to start and stop dictation due to it being a very useful methodology, rather than doing it by dictation. In consequence, it is proposed, as a future research work, to find some buttons that must never be employed in daily work to start and stop dictation. Furthermore, an extremely urgent future work is to propose new keyboards or mouse buttons that let us undo, in an easy way, an incorrectly recognized text.

Finally, mean results showed that these kinds of applications can be employed in real life by computer workers like engineers and scientists in their daily tasks. Furthermore, due to these kinds of applications, this is the only way to continue working after this kind of illness appears, and this methodology must be learnt in graduate and postgraduate studies as a complement to programming languages subjects.

6. *References:*

[1] M. O'Sullivan, S. Richards, W. L. Walter. Iliopsoas Tendonitis. A Complication after total hip arthroplasty. *J. Arthroplasty* Vol.22, No.2, 2007, pp.166-170.

[2] G. Guido. What the study of voice recognition in normal subjects and brain-damaged patients tells us about models of familiar people recognition. *Neuropsychologia*. Vol.49, No.9, 2011, pp. 2273-2282.

[3] Z. Wei, V.G. Duffy, R. Linn, A. Luximon. Voice recognition based human-computer interface design. *Comput. Ind. Eng.* Vol.37, No.1-2, 1999, pp.305-308.

[4] J. M. Górriz, J. Ramírez, E.W. Lang, C.G. Puntonet, I. Turias. Improved likelihood ratio test based voice activity detector applied to speech recognition. *Speech Communication*. Vol.52, No.7-8, 2010, pp.664-677.

[5] R. G. Zick, J. Olsen. Voice recognition software versus a traditional transcription service for physician charting in the ED. Am. J. Emerg. Med. Vol.19, No.4, 2001, pp.295-298.

[6] H. Lee. Researching the use of voice recognition writing software. Computers and Composition. Vol.20, No.1, 2003, pp.77-95.

[7] M. Leijten, L. V. Waes. Writing with speech recognition: The adaptation process of professional writers with and without dictating experience. *Interact. Comput.* Vol.17, 2005, pp.736-772.

[8] R. P. Ramachandran, K. R. Farrell, R. Ramachandran, R. J. Mammone. Speaker

recognition-general classifier approaches and data fusion methods. *Pattern Recogn.* Vol.35, 2002, pp.2801-2821.

[9] Micorsoft speech recognition. http://support.microsoft.com/kb/306901 (Accessed March 2012)

[10] Dragon Naturally Speaking. http://spain.nuance.com/naturallyspeaking/ (accessed March 2012)

[11] E-speaking software. http://www.e-speaking.com/index.htm (accessed March 2012)

[12] T. Batt, S. Wilson. A study of Voice – Recognition Software as a Tool for Teacher Response. *Computers and Composition.* Vol.25, 2008, pp.165-181.

[13] D. Coniam. Voice recognition software accuracy with second language speakers of English. *System.* Vol.27, No.1, 1999, pp.49-64.

[14] R. Haeb-Umbach, P. Beyerlein, D. Geller. Speech recognition algorithms for voice control interfaces. *Philips Journal of Research.* Vol.49, No.4, 1995, pp.381-397.

[15] T. Batt, S. Wilson. A Study of Voice-Recognition Software as a Tool for Teacher Response. *Computers and Composition.* Vol.25, No.2, 2008, pp.165-181.

[16] A. Rogowski Industrially oriented voice control system. *Robot Cim-int Manuf.* Available at:
http://www.sciencedirect.com/science/article/pii/S0736584511001189.

[17] A. Andics, J.M. McQueen, K.M. Petersson. Neural mechanisms for voice recognition. *NeuroImage.* Vol. 52, No.4, 2010, pp.1528-1540.

[18] T. Ayres, B. Nolan. Voice activated command and control with speech recognition over WiFi. Sci. *Comput. Program. Vol.* 59, No.1-2, 2006, pp.109-126.

[19] E. A. O'Hare, M. F. McTear. Speech recognition in the secondary school classroom: an exploratory study. *Comput. Educ.* Vol. 33, 1999, pp. 27-45.

[20] J. Takahashi, N. Sugamura, T. Hirokawa, S. Sagayama, S. Furui. Interactive voice technology development for telecommunications applications. *Speech Communication. Vol.*17, No.3-4, 1995, pp.287-301.