CRESCENDO

powered by SpeechMagic™

Leveling the paperwork mountain

A speech recognition for healthcare white paper

Head Office
Crescendo Systems Corporation
1600 Montgolfier
Laval, Canada H7T 0A2
Tel: (450) 973 8029
Fax: (450) 973 4449

Toronto Office
Crescendo Systems Corporation
1 Yonge Street, Suite 1801
Toronto, Ontario
Canada M5E 1W7
Tel: (416) 214-4272
Fax: (416) 369-0515

UK Office
Crescendo Systems Limited
Unit 5 Sunbury International Business Centre,
Brooklands Close
Sunbury-On-Thames
United Kingdom TW16 7DX
Tel: +1 44 (0)870 770 1717
Fax: +1 44 (0)870 770 1716

TABLE OF CONTENTS

1. Executive Summary

Thanks to speech recognition, healthcare facilities can benefit from a huge productivity gain on the part of the Transcriptionists, allowing outsourcing to be reduced or medical Transcriptionists to be redeployed. This frees physicians from their workstations and accelerates the delivery of critical medical data whenever and wherever patient care is needed.

But workflow automation is another keyword. Medical data must be accurate, consistent, complete and timely, not just transcribed faster. It needs to be easily and securely exchangeable between the different IT systems from different vendors, whether it is dictation, transcription, electronic chart, PACS, RIS, HIS, CIS etc. The digital dictation system plays a pivotal role in this puzzle. Hence the need to not only analyze how a dictation system integrates its speech recognition engine but also how efficiently it fits within the overall IT infrastructure picture.

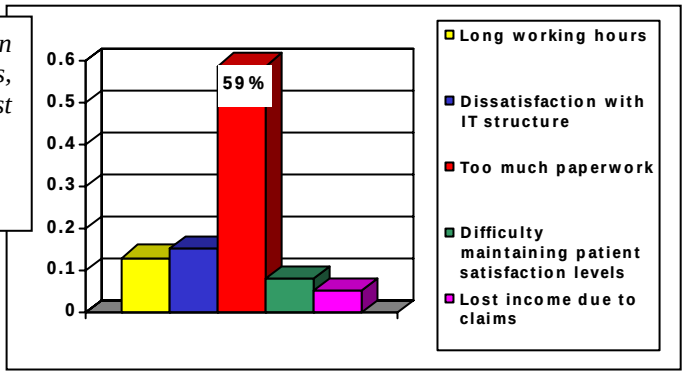
2. Challenges facing healthcare today

2.1. The paperwork mountain issue

Written medical reports are the life-blood of healthcare, resulting in huge amounts of paperwork for physicians and healthcare facilities. A recent HIMSS survey reveals paperwork as being the number one source of dissatisfaction for nurses and physicians, taking precedence over other issues such as long working hours or lost income due to claims.

"When doctors and nurses at your organization express dissatisfaction with internal services, which ONE of the following is their most frequently expressed concern?"

(Source: HIMSS 2003 Survey)



2.2. The cost of documentation

Document creation has become time-consuming and cost-intensive, with long turnaround times preventing immediate patient treatment and overloaded medical Transcriptionists struggling to keep up with growing dictation volumes. According to the Consensus Workgroup on Health Information Capture and Report Generation, report generation costs in the US are currently estimated to be well over $50B annually. New federal regulatory mandates only tend to accelerate this trend by imposing additional administrative and paperwork burdens – according to AHA, "compliance with the Health Insurance Portability and Accountability Act's privacy regulations alone are expected to cost hospitals between $4 billion and $22 billion." (White Paper - The State of Hospitals' Financial Health).

Finally, PricewaterhouseCoopers survey of hospitals and health systems reveals the following care/paperwork ratios that speak for themselves.

Care Provided	Equals Paperwork *
1 hour of Emergency Department care	1 hour of paperwork
1 hour of Surgery & Inpatient acute care	36 minutes of paperwork
1 hour of Skilled Nursing Care	30 minutes of paperwork
1 hour of Home Healthcare	48 minutes of paperwork

**(Source: PricewaterhouseCoopers survey of hospitals and health systems)*

2.3. The quality of documentation

According to the US Institute of Medicine, over 100,000 Americans die every year from medical errors, mainly due to poor medical documentation (i.e.: illegible entries, misinterpretations or lack of uniform capture). No wonder the present healthcare agenda, urging healthcare facilities around the country to improve the reliability of data and, ultimately, the quality of patient care.

2.4. The need for document automation

Paperwork and backlogs are a stress factor for physicians and patients alike. Affording physicians and hospitals greater flexibility in the recording and transcription of medical dictations and largely automating the process helps reduce waiting times for patients and improves working conditions for physicians. But this is not the only challenge facing healthcare facilities today. To comply with strict confidentiality standards and cost reduction agendas, healthcare facilities are in need of documentation that is accurate, consistent, complete, timely, exchangeable between different IT systems and accessible whenever and wherever patient care is needed.

▶ Healthcare facilities have to find a way to keep records current at lower costs while efficiently managing an ever-increasing amount of information and documentation.

3. Speech Recognition to the rescue

Most people have come to associate speech recognition with unfriendly single-use solutions and their related user-adaptation hassles. Don't we all remember a friend or colleague from the 1990's flaunting his new speech recognition gadget while spending hours trying to get a single sentence right?

Did you know?

We typically speak 5 to 7 times faster than we can type.

However, the technology has had time (and the R&D budgets) to evolve since then. Off-the-shelf, consumer software can only be made available at a low price by providing a fairly basic speech engine for generic use. To be beneficial for professional document creation, speech-recognition systems must be able to interpret what the speaker means, rather than just successfully recognize words. It also has to be carefully optimized for the specific requirements of the institution it is used in. Professional speech recognition means targeted products developed for clearly defined user groups and able to seamlessly integrate into the departmental network.*

The following gives an overview of Intelligent Speech Interpretation (ISI), the technology developed by Philips Speech Recognition Systems. Beyond the literal translation of words, ISI extracts the meaning so that high-quality documents can be produced with the minimal of human intervention.

3.1. The Human Factor

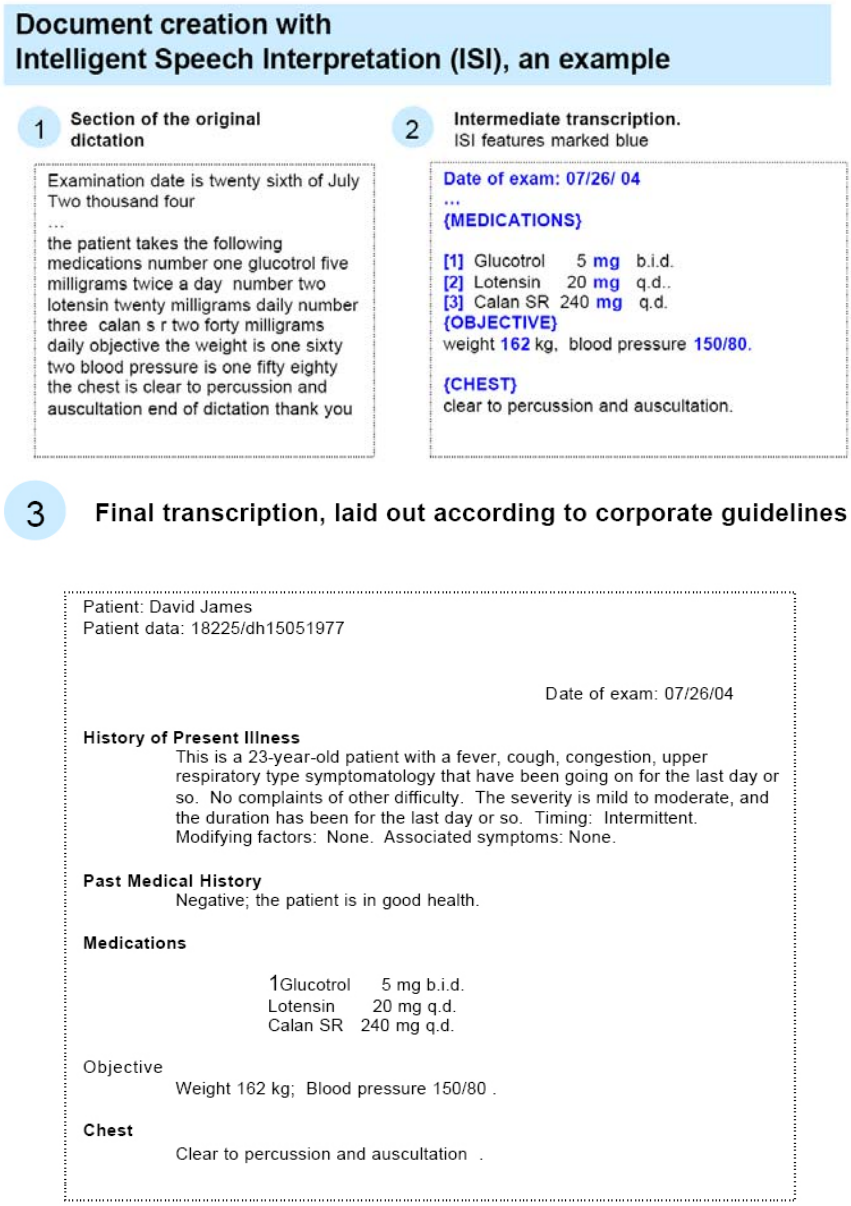
Transcriptionists do much more than simply type what was dictated. For a start, they leave out the 'um's and 'eh's, ignore dialogue that is not part of the dictation, implement corrections that are dictated as part of the text, fill the information into forms, and even rephrase sentences. They format and organize text, adding section headings, numbering lists and standard blocks of content. In short, they ensure that the final document communicates what was meant, rather than just what was said. Intelligent Speech Interpretation technology, as developed by Philips, emulates the capabilities of good medical Transcriptionists, to increase the productivity of transcription staff and free resources for more critical tasks. Crucially, the technology is just as useful to physicians who prefer to look after the reporting process themselves, as to those who delegate transcription and editing to someone else.

3.2. Situational intelligence

The initial challenges in doing this are acoustics due to background noise as well as differences in dialect, variations in pitch and speed, and how distinct or slurred the pronunciation is. By filtering out acoustic events, which have no relevance for the current report and comparing with known variations in speaker characteristics, the system can compensate for many of these deviations, and normalize the speech for further processing.

**Source: "Recognizing the Meaning" White Paper - Philips Speech Recognition Systems*

But speech recognition is not only about recognizing voice characteristics. Next, the system must recognize what the speaker said. As with other challenges in speech recognition, the ConText – the probability model of words and word sequences - of the dictation is the key to generating high quality and consistent results. This starts with vocabulary. Awareness of what people are likely to say not only helps recognize what they do say, it also helps identify what doesn't belong, for example, "PET" (photon emission tomography) is more likely in a radiologist's report than "pet" (an animal kept at home). This awareness is also about knowing the probability of a particular word, given the words used before: the probability of "PET" being followed by "scan" is much higher than it being followed by "food". Intelligent Speech Interpretation thereby offers dedicated dictionaries related to the physician's speciality that maximizes the recognition of complex profession-related terminology.



Source: "Recognizing the Meaning" White Paper - Philips Speech Recognition Systems

3.3. Word interpretation

Recognizing what was said provides a solid basis for correcting phrase and sentence structures. But spontaneous dictation often results in missing articles, verbs and punctuation, as well as redundant or repeated words and self-corrections. A clear understanding of the context helps interpret the words to identify and correct such matters.

Examples of word interpretations and situational intelligence

	Dictated Text	Recognized Text
Redundant phrases	End of dictation. Thank you.	Does not appear in final document.
Redundant phrases	Send copy of report to	Does not appear in final document.
Section headings	Condition on/at/upon/of/ discharge Next is condition...	{CONDITION ON DISCHARGE}
Dates	May five two thousand two fifth of May two thousand two...	May 5, 2002
Automatic punctuation	No chills fevers night sweats weight loss...	No chills, fevers, night sweats, ...
Silences/pauses	There hasn't been (------pause----) much change...	There has not been much change.
Non-speech dictation	There hasn't been (paper rustling) much change...	There has not been much change.
Hesitations	There hasn't been (AAHHMMMMM) much...	There has not been much...
Contraction	There hasn't been	There has not been

Source: "Recognizing the Meaning" White Paper - Philips Speech Recognition Systems

3.4. Rules and analyzes

Whether it is to dictate the content of a report or to execute commands for the word processing application, the system works internally with phonetic representations of words, and rules for the structures of phrases, sentences and documents. The developers entered basic representations and rules, along with suitable vocabulary. The system then added more detail by statistically examining large numbers of existing texts. When transcribing a dictation, the system compares the words on hand with these statistics to imply the word, phrase, sentence or document section, and adjust the output accordingly.

By working closely with a number of medical system manufacturers, the researchers from Philips have been able to correlate dictations with both the machine-recognized texts and manually corrected final

reports. This has enabled them to improve both the initial recognition rate and the quality of the reports the system delivers. In particular, it helped them discover unexpected "rules", such as situations where the speech recognition system recognized an acronym correctly, but a trained Transcriptionist would always turn that acronym into a full phrase.

4. Speech Recognition in action

The intelligence that turns talk into text is the heart of any speech recognition system. By recognizing what is said, and interpreting it reasonably, Intelligent Speech Interpretation technology makes sure the transcription process requires the minimum amount of intervention to produce accurate reports.

But as important as the ability to emulate the human ability to interpret speech, workflow automation and the way a speech recognition system is integrated within a facility's existing IT infrastructure are key to generating tangible results.

4.1. The digital dictation is the glue that holds everything else

A stand-alone speech recognition solution on an individual PC does not bring the expected gains in productivity and efficiency. It often causes frustration, as physicians quickly come to realize that they need to spend more time on report creation. Speech recognition needs to be approached as part of a whole document creation platform. Real benefits only come by implementing a digital dictation workflow solution with integrated speech recognition, which takes into account the entire document creation process and not simply the transcription of a dictation. The digital dictation workflow system is the central framework that supports everything else, from voice control to workflow management and it is what the physician will be interacting with on a day-to-day basis. The difference resides in the system's new ability to produce a "recognized text" together with the voice file. This draft report simply needs to be corrected as opposed to being fully transcribed.

4.2. Towards an integrated IT environment

The way this digital dictation system integrates with other medical systems is absolutely key. An integrated IT environment is synonymous for optimal accuracy and reliability of medical data.

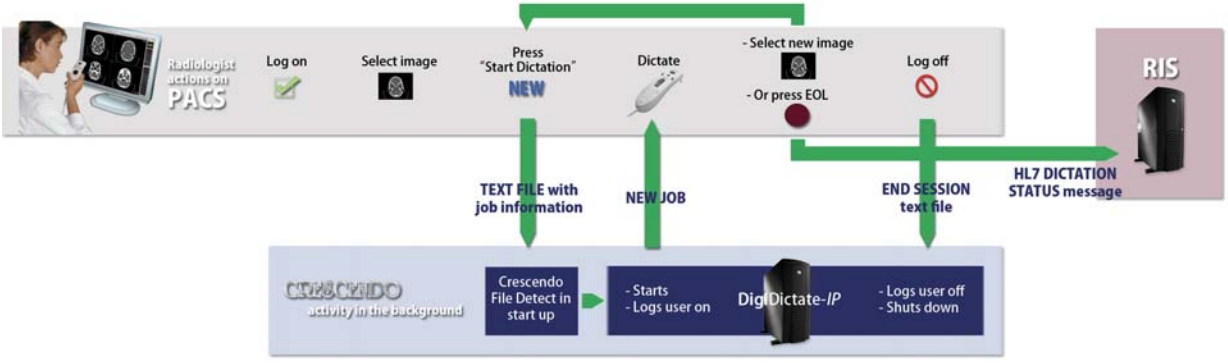
4.2.1. Example of an integrated dictation workflow in Radiology

A dictation system that is integrated with a Picture Archiving and Communications System (PACs) system allows the radiologist to dictate directly into the patient's image file, automatically importing relevant demographics into the dictation and eliminating manually entries.

- The integration between the two environments eliminates the need for the user to log on twice.
- The dictation is started and ended automatically by logging on and off the PACs system.

- Patient information is automatically passed on each time a new image is brought to the screen by the radiologist.
- To start a dictation, they press Record and simply close the image to end the dictation.
- To edit a dictation they select the appropriate image and dictate the required changes.

All dictation controls can be accessed using a PC microphone with an integrated mouse - and optional bar code reader. Any of the user actions, such as the creation of a dictation, bar code reading, recording, locating and editing existing dictations and completing dictation, can be performed according to the physician's preference, either within the PACS application using a PC microphone or the dictation interface running in the background.



Such systems integration can save radiologists significant time while greatly improving reading quality by concentrating all critical data and actions into one single user application. Integrated with a RIS or EMR, the dictation workflow system can then store the document in the corresponding patient file.

4.3. Multi-site support and workflow automation

Professional systems are designed to support thousands of users and dictation throughput daily. Whether dictating physicians are split over several facilities today, or might be a few years from now, the digital dictation system had better be based on multi-facility architecture. At the end of the day, healthcare administrators are looking for a total enterprise solution, fully scaled to serve multiple sites, Departments, and geographical locations without impacting on workflow efficiency. Rather than having to handle one server per site, a centralized speech solution greatly eases system administration while routing, privileges and security can still be customized on a per-site, per-Department and per-user basis.

4.4. Confidentiality

Remote dictation should be as secure as dictating on-site. With or without speech recognition, a digital dictation workflow system is not viable if it is not secure. It needs to provide a level of security that

strictly complies with HIPAA/PIPEDA standards allowing sensitive data to be shared among individuals, Departments and cross-regional locations. All voice files and reports need to be stored and managed centrally in a way that prevents patient-sensitive data from residing on the local PC. Integrated authentication mechanisms, where each user needs to be validated to gain access, and the support of other security technologies (i.e. Virtual Private Networks (VPN), Remote Access Service (RAS) and Windows® NT authentication) are also a critical factor to consider.

4.5. The front-end approach

Front-end speech recognition is a particularly attractive feature for physicians who prefer to look after the full report generation process. Text is generated on-screen from their dictations in real-time, allowing physicians to edit and finalize documents themselves without the involvement of a Transcriptionist - fantastic on weekends or for highly confidential documents. However, while physicians are in control of the entire dictation-to-distribution process, it may not be the most cost-efficient solution for all of them. Some might see their main activities affected by certain tasks or features such as the initial training of the engine or the time-consuming editing process. This is why a system should ideally support the back-end workflow as well, leaving all options open.

4.6. The back-end approach

4.6.1. Two-tiered Workflow

Because most physicians prefer not to be involved in the document creation process, back-end speech recognition is an attractive option. Dictations are automatically processed by the speech recognition server in the background and the Transcriptionist is presented with a transcribed text and the original audio file. Their new role consists of checking the recognition accuracy rather than having to transcribe the entire report. In addition, the location of the Transcriptionist is no longer an issue since documents are automatically routed to the relevant Transcriptionist, wherever they might be working.

4.6.2. Self-correction alternative

Should some physicians decide to do the correction themselves while taking advantage of all the back-end functionalities, the option should be available. This is particularly important for environments such as Radiology, Pathology and Emergency Medicine, where reports are relatively short and usually "normal" (negative findings) and experienced medical Transcriptionists are more and more difficult to find. Physicians can then dictate, review and correct recognized text that has been returned from the speech recognition server, perform synchronized playback and sign off dictations and/or reports using one single application.

4.6.3. Back-end functionalities

Acoustic adaptation

Involvement in the report creation process is reduced to the press of a couple of buttons and the dictation itself so that the process is completely transparent to physicians. Their voice characteristics, pronunciation and dialect are analyzed and a specific profile created accordingly. The acoustic adaptation permanently adjusts to each individual physician and their working environment (e.g. background noise) in a fully automated way for each dictation, minimizing the likelihood of misunderstandings resulting from physician specific pronunciation as well as local dialects or accents of non-native speakers. This continuous adaptation concept minimizes manual interaction and results in continuous improvement in the recognition rate. To further improve the initial performance, unknown words can be added from existing documents using the ConText Tuner. It is possible for the user to add up to 128,000 words. As for input devices, the system supports multiple input devices per user (telephony, mobile, PC microphone).

The voice model training (initial training) typically takes two minutes and is often not necessary for native speakers. For non-native speakers or speakers with a strong accent up to ten minutes of initial training are recommended. Typically, voice model training is carried out using a wizard requiring the physician to read out a given text according to which the voice model is adjusted.

Shared vocabulary and corrections

Due to the network-based and shared lexicon approach, every new word is immediately picked up by the system and made available to all physicians for ever increasing recognition accuracy. Physicians can also access the system from any connected workstation and not only from their respective PCs.

Template automation

The system supports templates, which can be made available to an individual Author, a whole Department, a hospital or a whole region. The same holds true for frequently used paragraphs.

Scalability

A network-based approach also means a system that is highly scalable for the largest facilities and hours of dictation throughput per day. In a multi-facility configuration, ConTexts can be made available in several languages, allowing multilingual facilities to have an identical English user interface, for instance, in all of their offices, while the speech recognition engine, SpeechMagic, supports dictation in the local language.

APPENDIX

Author

Bob Yacovitch, Ph.D. and MIT Graduate
Senior Design Engineer, Crescendo Systems Corp.
1-800-724-2930

About the Author

After graduating with honors in Physics at McGill University in Montreal, Canada, Dr. Yacovitch went on to earn his Doctorate at the prestigious MIT in Cambridge, Massachusetts. In the more-than-twenty years since, he has dedicated himself to the field of high technology software development. After holding management positions at organizations such as XIOS Systems, ACOM Computer Systems and Marconi Communications, Dr. Yacovitch now supervises a team of developers accountable for the most progressive systems to date from Crescendo.

Bibliography

"Recognizing the Meaning", a Philips Speech Recognition Systems White Paper by
David Hegarty, Marcel Wassink and Armin Scheuer

"The State of Hospitals' Financial Health" - White Paper
Advancing Health in America (AHA)

HIMSS 2003 Survey

"Healthcare documentation: A report on information capture and report generation"
The Consensus Workgroup on Health Information Capture and Report Generation

"Survey of hospitals and health systems"
PricewaterhouseCoopers, 2001

"Does digital dictation and speech recognition help ease the shortage of doctors in the UK?"
Armin Scheuer, so2say communications, Berlin, Germany, 2005

"Healthcare Technology Star Trek Style: Will The Force Be With You Or Against You?", a presentation by Nick van Terheyden, MD

Crescendo Systems Corporation

Head Office
1600 Montgolfier
Laval, Quebec
Canada H7T 0A2
Toll free: 1-800-724-2930
Fax: (450) 973 4449

Toronto Office
1 Yonge Street, Suite 1801
Toronto, Ontario
Canada M5E 1W7
Tel: (416) 214-4272
Fax: (416) 369-0515

UK Office
Crescendo Systems Limited
Unit 5 Sunbury International Business Centre, Brooklands Close
Sunbury-On-Thames
United Kingdom TW16 7DX
Tel: +1 44 (0)870 770 1717
Fax: +1 44 (0)870 770 1716

www.crescendo.com

info@crescendo.com