Speech in the Healthcare Industry

By Dr. Caroline Henton • January/February 2004

Medical care providers, ranging from receptionists to insurance providers to specialist surgeons, currently have a wide array of applications available that may increase their productivity by using speech. Other beneficiaries include the recently established National Patient Safety Network and developers in the growing field of bio-informatics. There are advantages to using speech technology in medical and healthcare applications, such as record keeping and transcription, as well as many linguistic dangers of using similar tools to place pharmaceutical orders, with potentially lethal consequences.

Approximately 15 years ago, researchers in automatic speech recognition (ASR) realized that one of its most effective uses was in routine data entry. To this end, they focused on developing specialized limited-vocabulary tailored applications that stood a better chance of productization and success in niche markets, such as medical (or legal) note-taking and record-keeping. By restricting users' utterances to single words, and avoiding no-match or out-ofgrammar utterances, recognition rates in these scenarios is likely to be more accurate. The software has indeed had greater adoption rates and provides greater help than, say, spoken command-and-control of the PC desktop. Speech-enabled medical documentation systems allow physicians to use ASR to create and dispatch patient notes, medical records, referral letters and, most recently, place prescription orders.

X-RAYS TO X-FILES

Ten years ago ASR was bundled into a handheld device that resembled a personal memo recorder so that radiologists were able to record their analyses of X-ray plates. The data would then be loaded into appropriate fields in medical records on a larger computer. The maturity of this type of application was illustrated recently when *Speech Technology Magazine* recognized Ramapo Radiology Associates with a Most Innovative Solution award in 2003. A combination of Dragon's NaturallySpeaking (from ScanSoft) with VoiceBrook's VoiceOver tools makes it possible for radiologists to deliver prompt diagnoses for better patient care, rather than spend time on repetitive, routine administrative tasks.

Ramapo's description of their product encapsulates this successful deployment. "Speech recognition solutions can effectively replace traditional transcription, reducing cost and speeding response to referring physicians," said Dr. Robert Tash. "Document creation in realtime can be achieved without significantly altering the radiologists daily workflow. In addition, speech recognition software is always available, and the rapid turnaround of reports is a major benefit for us. We are very pleased with our results with speech recognition technology and consider it a vital tool."

Managing healthcare information such as patient names and insurance records is a successful and safe use of speech technology. The challenges of recognizing and verifying personal and other proper names are essentially no greater than in other routine record-keeping applications (Henton, 2003). Well-designed user interfaces combine ASR and graphical user interfaces and custom templates. Macros avoid repetitive tasks to reduce the time taken to create documents by as much as 50 percent. And transcription is real-time. Physicians working in shared practices, hospitals, clinics and other specialty groups benefit from expedited exchange of, and access to, dictated records, notes and prescriptions in a centralized document database. Medical professionals can save time, accelerate reimbursements, cut processing costs and increase revenues.

DO NO HARM

In an emerging and potentially powerful application of speech technology, physicians can now speak prescription orders into a wireless handheld device, like a PocketPC©. Embedded speaker-independent, non-continuous recognition ASR is then used to enter the spoken items in pre-determined fields. After recognition has been performed, text appears on the small screen for confirmation and the prescription is relayed to a central server for rapid filling at the pharmacy of the patient's choice. It is anticipated that physicians and pharmacists should review all prescriptions placed wirelessly at the end of the day, but we are all aware of the public area noise levels, the size restrictions on PDA screens, and the tedium of having to review forms.

Typical orders spoken by harried doctors, walking along the busy hospital corridors take the form: "Ibuprofen. 600 milligrams. Every 4 hours. A.C. For pain." Given the many opportunities for mistakes (in the speech recognition, in mixed-up drug and/or patient identification, in dosage, etc.) this scenario may provoke chills in many of us. How might the linguistic diversity due to physicians who do not speak English as a native language affect the effectiveness of these speech-driven devices?

YOU SAY TRACHEA, I SAY TRACHEA

Medical dictation systems must support far greater than normal vocabularies – more than 250,000 words to include medication names, medical procedures, diagnoses, diseases, etc. Shaw wasn't considering this when he called America and Britain "two countries divided by a common language" (Henton, 2002), but the divisions are as strong here as elsewhere in English. The list below presents a few well-known differences in the terminology used (to designate semantically the same thing) and the varying pronunciations of these scientific/medical terms by American and British speakers. All pronunciations appear according to the International Phonetic Alphabet (IPA) transcription standard; primary stress is indicated by a raised bar before the stressed vowel.

US English	UK English
aluminum	aluminium
anesthesiologist	anaesthetist
pharmacy	chemist's
podiatrist	chiropodist

	US Pronunciation	UK Pronunciation
amine	ə'mi n	'æ m i n
amnesia	æ m 'n i ʒ ə	æ m 'n i z i ə
antibiotic	æ n r aɪ b aɪ 'ɑ r ɪ k	æ n t i b aɪ 'ɒ t ɪ k
asthma	'æ z m ə	'æ s m ə
estrogen	'e s t r ə dʒ ə n	'i s t r ə dʒ ə n
laboratory	l 'æ b ə r ə t ɔ r i	l ə 'b ɔ r ə t r i
migraine	m 'aɪ g r eɪ n	m 'i g r eɪ n
nausea	n 'ɑ ʒ ə	n 'ɔ s i ə
premature	p r i m ə t 'ʊ ə	p r 'æ m ə t ʃ ə
pulmonary	p 'ʊ l m ə n e r i	p 'ʌ l m ə n r i
respiratory	r 'e s p ə r ə t ɔ r i	r ɪ s p 'ɪ r ə t r i
quinine	k w 'aɪ n aɪ n	k w 'ɪ n i n
trachea	t r 'eɪ k i ə	t r ə k 'i ə

The impact of these significant pronunciation divergences – in stress placement, varying numbers of syllables and in vowel length – on speech recognition is perhaps not the most obvious one. ASR providers should know these variants and load appropriately different grammars (with their associated pronunciation models) into the localized software used in the U.S., Canada or the UK. The real problem lies with physicians and medical technologists who have learned English (perhaps as a second or other language) outside North America or the British Isles, but who are resident in the U.S. or the UK. Linguistic speculation accounts for these varying pronunciations by assuming that (native) speakers of English draw different analogies according to their perception of the morphological origins of these neologisms, and by regularizing with the stress patterns preferred in their dialect. Speakers of Indian or Singaporean English will have learned primarily British English but they may practice in Chicago or Vancouver; similarly, Australian English doctors and dentists who studied in Hong Kong may have moved to London. Their accented varieties of English will be one impediment to reliable recognition built for other standard accents, and their learnt/preferred pronunciation of the terminology will add another layer of potential confusion or failure.

UNSPEAKABLE NAMES

For legal purposes names and trademarks need to be spelled correctly. However, it is not possible to legally dictate how they are pronounced. This has important and varied repercussions when names are (re)produced using text-to-speech (TTS). In naming a new company or product, it is now de rigeur to combine upper and lower-case characters in one alphabetic string, with no white space, or to alter the spelling for eye appeal. This typographical rulebreaking also comes from company mergers, giving rise to such unwieldy strings as exemplified in the following list of some pharmaceutical giants and their product brand names. Bold face sequence show non-English spelling names; the hash mark (#) shows a TTS normalized text string that breaks

the normal spelling (phonotactic) rules of English, which may in turn cause the TTS system to produce an unpredictable or weird interpretation.

	What the TTS system sees
AstraZeneca	Astrazeneca #
Aventis	
Cozaar ©	
GlaxoSmithKline	Glaxosmithkline #
GlaxoWellcome	Glaxowellcome #
Kaiser	
Merck	
Pfizer	
re-up-take	reuptake #
Rituxan®	
SmoothCaps™ #	Smoothcaps #
Tarceva™	
TNKase™	Tnkase #
Trastuzumab	
Zilactin	
Zyrtec	

Some drug names are familiar enough to physicians and patients alike that they should not present pronunciation/recognition difficulties for an automated spoken system (e.g. aspirin, codeine, Valium™). For native speakers of English, however, other drug and/or compound names range from fairly unambiguous, to opaque/ambiguous, to those speakers having no idea with regard to either pronunciation or stress placement. The three lists below illustrate these issues, in descending order of difficulty for humans, and by deduction, those which present increasing difficulties for TTS systems:

Transparent drug names: Allegra©, Clarinex, Diovan©, Flonase©, Lunelle©, Minipress©, Nasonex©, Nexium©, Paxil, warfarin, Zoloft©.

It is noteworthy that branded drug names have benefited from some marketing insights, since they are generally more accessible than the largely generic drug names that appear in the lists below.

Opaque/ambiguous drug names

	Pronunciation	Reason
Halcyon©	h æ l s i ə n	obscure word (origin: Gk.)
Levitra©	l ə v i t r ə or	cognate with 'in vitro'
	l e v i t r ə	cognate with 'levitate'
lorezepam	l ɔ r ' ɛ z i p æ m or	no cognate
	l ' ɔ r z ı p æ m	cognate with 'lore' + stress of 'marzipan'?
rifabutin	r aı f ə b ' u r ı n or	no cognate
	r ı f ə b ' ʌ r ı n or	'rifa' + cognate with 'Butlin'?
	r aı f ə b j ' u r ı n	
rifampin	r ' ı f ə m p ı n or	no cognate
	r ı f ' æ m p ı n or	cognate with 'rif'ampicin'
	r aı f ə m p ı n	
Vasotec©	v ' æ z oo t e k or	
	v ' e ı s ə t e k	

Unknowable	Pronunciation problems		
amiodarone	Stress?	æ m i oo d ə r oo n or	
		æ m i ' ə d ə r oo n or	
		ə m i ' ə d ə r oo n	
carbamazepine	Stress?	- p ı n or	– p aı n
ceterizine	Stress?	s ɛ - or	k ɛ -
		- z ı n or	– z aı n
cimetidine	Stress?	s ı - or	s aı -
		- d ı n or	– d aı n
desloratadin	Stress?		
drospirenone	Stress?		
estradiol	Stress?	ɛ s t r ' æ d i oo l or	
		' e s t r ə d i o l	
norgestinate		n ɔ r dʒ - or	n ɔ r g -
thioridazine	Stress?	θ i ɔ r ı - or	θ aı ɔ r ı -
		- z ı n or	– z aı n
vardenafil	Stress?		

In a vain attempt to help speakers with unpredictable stress placement and/or vowel quality in drug names, pharmaceutical companies and health management providers (HMOs) sometimes give pronunciation hints, in a random dictionary-style transcription. For example, the following are taken from product advertisements and prescription leaflets from the HMO:

ceterizine (se-teer-i-zeen)
Levitra (Luh-VEE-Trah)
Naproxen (Na-PROX-en)
SIMVASTATIN (SIM-va-stat-in)
STRATTERA™ (Stra-TAIR-a)
Zyrtec (Zur'-tek)

This information is completely unsystematic: note three different renditions of unstressed syllables, of post-positioned single quote to indicate stress or upper case, and the unjustified or inconsistent use of upper case in general. It is not helpful to native nor non-native speakers of English, or to those confused by quasi-phonetic notation.

Problems with the unknowables (the great majority) remain unalleviated by drug manufacturers providing such pseudo-pronunciations. More often than not, we are left to our own (wobbly) intuitions about stress placement, short vowel /I/, long vowel /i/, or diphthong /aI/; 'hard' or 'soft' letter "c" i.e. /s/ or /k/, etc. Anyone who has listened to a radio doctor's call-in show, where people question a physician about the drugs they have been prescribed, knows that lay people (us) stumble and hesitate with the pronunciation of the drugs they're taking, and ultimately resort to spelling them for the doctor.

Given these many (socio)linguistic variables, is it impossible to attribute a degree of certainty in attempts to recognize many names of drugs. All commercial recognizers rely on certainty/confidence factors to supply a match. Recently Walter Rolandi (2003) supplied a useful, critical analogy for this recognition problem: "Imagine an English-only speaker being asked a question by someone speaking in French ... The English speaker instantly knows that what the other person said was not English, i.e. that the speakers' utterance was not in the listener's grammar ... having a recognizer capable of accurately determining whether ... an utterance is in its grammar would be a significant step toward more intelligent voice user interfaces."

Having medical and healthcare-based systems capable of accurately determining whether diseases, procedures, and the names of drugs have been recognized accurately by speaking them back using TTS (to prompt checking and re-entry by hand if necessary) would not only be an intelligent and significant step. It is a vital, preventative step if these devices are to be used more widely by all medical practitioners. Computerized order entry systems typically offer physicians and medical institutions the ability to "streamline workflow, reduce error, save time, money and lives" (www.validus.com). With the many and varied linguistic and phonetic barriers given, it is not clear how errors can be avoided, let alone reduced, and how lives may be saved.

RX FOR REMEDIES

There are still three hurdles to wider adoption of digital dictation devices to increase efficiency for health-care professionals. First, there are understandable concerns about confidentiality/security. Second, the fragility or fallibility of recognition accuracy. Third is the lack of immediate spoken guidance cum confirmation. What can we suggest to mitigate these factors? The first is the easiest: users need to be sensitized to the need to enter the data in a quiet, semiprivate location. Walking out from a consultation, or from a patient's room, or standing near the nurses station in the center of a bustling ward are not ideal environments in which to speak delicate, private facts about a patient's prognosis or prescriptions. These are also very noisy places, which in turn will affect the accuracy of the recognizer adversely, leading to repeated attempts and giving rise to increased frustration rather than efficiency. The second problem will then be tolerated, if not solved. The last, and most important improvement in these speech scenarios, is for the user to have some guidance and immediate confirmation of what they have spoken.

Many early adopters in U.S. radiology departments have since abandoned spoken record keeping, because the need for repetition and high failure rates were simply too frustrating. According to Philips Speech Recognition Systems, however, their product SpeechMagic™ (available in 22 languages) is now used in some European countries by more than 60 percent of radiologists (STM NewsBlast, December 10, 2003). The product has recently expanded into other specialized areas, such as cardiology, pathology and surgery. Clearly the speech recognition component has improved over the past 15 years. And perhaps the working conditions of these non-U.S. professionals provide better, quieter, privacy.

There remain skeptics in the U.S. medical profession who simply do not trust that doctor-patient confidentiality is not being violated, and who also do not trust the accuracy of the speech recognition. This may be because the ability to talk back is NOT there. None of the current instantiations include TTS, which is capable of talking back. TTS can guide users to speak a personal or product name correctly (i.e. the way the name has been entered phonemically in the recognizer's dictionary), and it can safely confirm entries that have been made using ASR and/or the graphical interface. Every doctor, specialist and pharmacist would welcome such a system if it contained such features and IF their HMO accountants or company paid for the installation, training and setup fees.

References

Henton, C. (2002) You say 'zee', and I say 'zed'. Issues in localizing voice-driven applications. *Speech Technology Magazine*, May/June, 28-31.

Henton, C. (2003) The name game: pronunciation puzzles for TTS. *Speech Technology Magazine*, September/October, 32-35.

Rolandi, W. (2003) When you don't know when you don't know. *Speech Technology Magazine*, July/August, p.28.