INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY 8, 247–257, 2005

Bitter Pills to Swallow. ASR and TTS have Drug Problems

CAROLINE HENTON
CTO, Talknowledgy, 330 High Street, Santa Cruz, CA 95060
carolinehenton@hotmail.com

Received April 6, 2004; Revised October 26, 2004; Accepted October 26, 2005

Abstract. In an emerging application of speech technology, physicians speak prescription orders into wireless handheld devices that record the order using automatic speech recognition (ASR). Orders are recognized using ASR, then displayed on the screen for verification and, after checking, are relayed to the pharmacy for filling. Given the frequent opportunities for mistakes, this scenario provokes chills in many of us. Many physicians do not speak English as a native language, or they have a strongly accented variety of English.

Some drugs are familiar enough that they should not present recognition/pronunciation difficulties (e.g. aspirin, codeine). Hundreds of drug names are easily confusable; other names range from opaque to ambiguous to 'no idea' with regard to pronunciation.

The problems with the 'unknowable' drug names (the majority) are sometimes alleviated by drug manufacturers providing a 'pronunciation hint', in a dictionary-style phonetic transcription. This is not necessarily helpful to native or non-native speakers of English, or those unfamiliar with such renditions. More often, we are left to our own (wobbly) intuitions about stress placement, vowel length or 'hard' or 'soft' letter 'c', etc. Speaking the potentially infinite class of drug names correctly is an extra challenge for any TTS system.

This paper presents data for drug names in the confusable, transparent, opaque, and unknowable categories. The clear and present dangers for patients receiving the wrong drugs and for everyone using such speech-automated systems are examined.

Keywords: speech technology in healthcare, linguistic diversity, pronunciation problems

Introduction

Medical care providers, ranging from receptionists, insurance providers, and dental hygienists 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, the Center for the Advancement of Patient Safety at U.S. Pharmacopeia, and developers in the growing field of bioinformatics. There are some advantages to using speech technology in relatively benign medical and health-care applications, such as record-keeping and transcription. There are also many linguistic dangers in using similar tools to place pharmaceutical orders, with potentially lethal consequences.

Prescription drugs with similar-sounding names are used commonly. The risks to patients in hospitals and in the community are growing by the year. Pharmaceutical manufacturers need to ensure clear written and audible differentiation in naming new drugs, so that the burden to distinguish drug names does not fall so heavily on practitioners. And if human beings have increasing difficulty telling names apart, and ordering the correct medication, it is obvious that such mistakes would only be compounded by placing prescriptions using spoken interfaces with automatic speech recognition (ASR) and text-to-speech (TTS).

Approximately fifteen years ago, researchers in ASR realized that one of the most effective uses for ASR 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-of-grammar' utterances, recognition rates in these scenarios is likely to be more accurate, and the software has indeed had greater adoption rates and provides greater help than, say, spoken command-and-control of the PC desktop. This is borne out in a recent examination of the long-term adoption of speech recognition in medical applications, where over a ten-year period, "... use of speech technology for software command and control ... decreased. This is probably the most oversold area of speech technology" (Grasso, 2003). Indeed.

Despite relatively poor recognition accuracy rates (80% to 95%), speech recognition is being used more frequently for dictating medical reports. The low accuracy rate may add 25% or more time needed to dictate reports (Grasso, 2003). Normally this would lessen the adoption of *any* technology; *but* it is still less expensive to take the extra time than to use a transcription service, and it accelerates records' accessibility. So speech-enabled medical documentation systems are being used more widely, because they allow physicians to create and dispatch patient notes, medical records, referral letters and, most recently, place prescription orders. The systems are not necessarily more accurate than those of a decade ago, but they *do* save money.

Examples are given here of the types of linguistic, usability, and confusability problems surrounding speech and speech-driven technology in healthcare provision. Some suggestions are given to improve drug-naming procedures, the basic functionality of digital dictation devices for health-care workers, as well as some ergonomic tips leading to better safety, user satisfaction, and greater efficiency.

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 (with or without the mobility of a handheld device) was illustrated when Ramapo Radiology Associates were given a 'Most Innovative Solution' award by Speech Technology Magazine in 2003. A combination of speech-driven tools makes it possible for radiologists to deliver prompt diagnoses for better patient care, rather than spend time on repetitive, routine administrative tasks. The product encapsulates a successful deployment: traditional transcription can be effectively replaced using speech recognition solutions. In turn this reduces costs and speeds responses to referring physicians. Real-time document creation can be achieved without altering the radiologists' daily workflow significantly. ASR software is never sick, nor does it does need vacation nor personal leave. About the only shared feature of ASR with a transcriptionist is that it's capable of talking back.

Early in 2004, Royal Philips Electronics announced introduction of mobile speech recognition solutions for radiologists. Radiologists are now able to dictate reports independent of their workstation, using mobile input devices. In a wireless LAN set-up, ECR visitors dictate radiology reports with a PDA. The sound file is transmitted automatically to the speech recognition server, corrected by an assistant, and returned to the PDA for final approval by the author. Frequently radiologists dictate on different workstations and change locations between dictations. By using mobile input devices, physicians are no longer tied to a rigid workflow; they may work more freely and efficiently.

Managing healthcare information such as patient names and insurance records is a relatively benign use of speech technology. The challenges of successfully recognizing and verifying personal and other proper names are discussed in Henton (2003). Difficulties with drug names are of a potentially similar order as recognizing proper names. Well-designed interfaces combine ASR and graphical user interfaces. Custom templates and macros avoid repetitive tasks, reducing the time taken to create documents by as much as 50%; and transcription is real-time. Physicians working in shared practices, hospitals, clinics and other specialty groups may benefit from expedited exchange of, and access to, dictated records, notes and prescriptions in a centralized document database. Medical professionals may save time, accelerate reimbursements, cut processing costs, and increase revenues.

Do No Harm

Physicians can now also 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 (where font size is not controllable by the user), and the tedium of having to review forms, regardless of the size of screen on which they are displayed.

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." Such structured speech input avoids the limitations of continuous speaker-independent ASR, to some extent, by recognizing speech as a series of pieces of coded information, isolated for that intended entry field. Frequent opportunities for mistakes (in the drug names, in the speech recognition, in mixed-up drug and/or patient identification, in dosage, etc.) may still provoke chills in many of us. Doctors' bad handwriting may be an occasional impediment to pharmacists filling prescriptions precisely and quickly. Many physicians do not speak English as a native language, or they have a strongly accented variety of English that is hard for other speakers of (different varieties of) English to understand. How might this linguistic diversity affect the effectiveness of these speech-driven devices?

Dangerous Liaisons

Problems that arise in the pharmaceutical industry's under-regulated drug-naming conventions permeate through the healthcare system to cause confusion, and may have serious consequences for practitioners, pharmacists and patients. The leading body in medication error reporting is the U.S Pharmamacopeia (USP), a non-government organization that establishes "standards to ensure the quality of medicines and other health care technologies". USP published its first list of similar drug names in 1995. By 1999, the list had more than 750 unique drug names reported to the USP MER Program. A succinct description of the issues appeared in the introductory paragraph of a report entitled "USP Quality Review: Use Caution–Avoid Confusion," (USP, 1999):

> "Confusion over the similarity of drug names, either written or spoken, accounts for approximately one-quarter of all reports to the USP Medication Errors Reporting (MER) program. This issue involves confusion between similar brand names, between similar generic names, and between brand and generic names. Such confusion is compounded by illegible handwriting, incomplete knowledge of drug names, newly available products, similar packaging or labeling, and incorrect selection of a similar name from a computerized product list."

In a further update, issued in 2001, USP stated:

> "The USP Practitioners' Reporting Network has released an updated list of similar drug names... more than 125 additional name sets have been captured... Although some of the names may not sound alike as read or look alike in print, when written or communicated orally, the names have caused or could cause a medication error."

A *prima facie* confusing name set is Naprosyn and naproxen, where the brand name is merely one (potentially silent) /k/ phone and stress pattern different from the generic name. Some people believe them to be different drugs: they assume that the latter pronunciation is simply a (Southern) accentual variant for the former. There are many other brand and/or generic drug names which differ only by one or two letters, as shown by the examples in the table below. The data come from the web site for 'Voice of the Injured' (2004) and from a USP report, issued in February 2004.

Drugs with similar names may not be perceived as a public health threat, but USP has reported cases of patient harm and fatalities. In March 2004, a professor of clinical pharmacology stated at a seminar on the safety and quality in Australian health care estimated that "80,000 hospital admissions each year were medication related. This included 22% of emergency admissions involving elderly people." (MBF News, 2004). While it is not possible to determine whether these admissions involved bad reactions to drugs, or bad interactions, or prescriptions for the wrong medications, it is nevertheless a disturbingly high number. Under-differentiated drug names, where frequently no second-checks are in place, give overworked pharmacists nightmares. They continue to be a major headache for the U.S. Food and Drug Administration (FDA). And they place patients, particularly older ones, at ever-increasing risk.

Given these many reports of humans making mistakes over the telephone, or simply from reading the

Table 1. Similar-looking or—sounding drug names, analyzed according to the linguistic and medical origins of the similarities.

Drug names	Linguistic (and medical) origins
Ambien ~ Ammens	Differ by only one syllable.
Amiodarone ~ Amrinone	Potentially confusable in poor handwriting. Name changed to 'Inamrinone' in 2000.
Biperiden ~ Risperidone	Differ by only one syllable; initial upper case consonants confusable in handwriting; and /s/ likely to be lost over the telephone or on handheld device.
Cardene SR ~ Cardizem SR	Differ by only one syllable; final nasal consonants indistinguishable over the phone, and in poor handwriting.
Citalopram ~ Escitalopram	Differ by only one syllable; /s/ likely to be lost over the telephone. Both are antidepressants.
Cytosar-U® ~ Cytoxan®	Differ by only two letters. Same therapeutic class: both are used to treat leukemia; both are injectables, stored in the same cabinet.
Dactinomycin ~ Daptomycin	Differ by only one syllable; plosives /t/ and /p/ indistinguishable over the phone.
Endocet® ~ Indocid®	Close vowels /ɛ/ and /ɪ/ sound identical before nasal consonants. /ɛ/ and /ɪ/ sound identical before final stops.
Ephedrine ~ Epinephrine	Names look and sound alike; similar indications; ignorance that one is *not* a brand name for the other.
Feldene ~ Seldane	Differ by only two letters and sounds; fricatives /f/ and /s/ indistinguishable over the phone, and in handwriting.
Flutamide ~ Flumadine	/t/ and /d/ both pronounced in US English as flap [ɾ]. Syllable metathesis in flutamide /f l u ɾ ə m aɪ d /would produce /f l u m ə t aɪ d/, which is very close to the pronunciation for flumadine /f l u m ə d aɪ n/.
Kaletra™ ~ Levitra®	Same number of syllables; same stress pattern
Lodine ~ Codeine	Differ by only one letter; upper case L and C may be confused in handwriting
MS Contin ~ Oxycontin	Differ by only two syllables; fricatives /s/ and /k s/ indistinguishable over the phone
Retrovir ~ Retinovir	Differ by only one syllable; rhoticized[1] syllables are easily metathesized[2] and regularly mispronounced by non-native (Asian) speakers of English

label, confusable drug names guarantee even less accuracy in any devices using speech recognition. If TTS is deployed to confirm the order, the synthetic speech is unlikely to be any more distinct than a human speaking. It is clear that this sort of speech application remains challenging—not just in terms of speakers being able to pronounce the names properly, but also in ensuring that the drugs can be pronounced so that they are not confusable with other drugs that have similar spellings and pronunciations. None of these already alarming indications take into account the effect of different speakers' dialects and accents. These linguistic ramifications are outlined next.

You Say Trachea, I Say Trachea

Medical dictation systems must support far greater than normal vocabularies (>250,000 words) to include medication names, medical procedures, diagnoses, diseases, etc. George Bernard Shaw was not considering this issue in particular when he called America and Britain "two countries divided by a common language" (for further examples of such linguistic separators, see Henton, 2002), but the divisions are as strong here as elsewhere in English. The table below presents a few well-known differences in the terminology used (to designate semantically the same thing) and

Table 2. Differences in US and UK English terminology.

US English	UK English
Aluminum	Aluminium
Anesthesiologist	Anaesthetist
Chiropractor	Osteopath
Pharmacy	Chemist’s
Podiatrist	Chiropodist

Table 3. Differences in pronunciations of medical terms in US and UK English. 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 Pronunciation	UK Pronunciation
Amine	ə m ˈi n	ˈæ m i n
Amnesia	æ m n ˈi ʒ ə	æ m n ˈi z i ə
Antibiotic	æ n ɾ aɪ b aɪ ˈɑ ɾ ɪ k	æ n t i b aɪ ˈɒ t ɪ k
Asthma	ˈæ z m ə	ˈæ s m ə
Estrogen	ˈɛ s t r ə dʒ ə n	ˈi s t r ə dʒ ə n
Laboratory	l ˈæ b ɚ r ə t ɔ r i	l ə b ˈɔ r ə t r i
Migraine	m ˈaɪ g r eɪ n	m ˈi g r eɪ n
Nausea	n ˈɑ ʒ ə	n ˈɔ s i ə
Premature	p r i m ə t ˈʊ ɚ	p r ˈɛ m ə tʃ ə
Pulmonary	p ˈʊ l m ə n ɛ r i	p ˈʌ l m ə n r i
Respiratory	r ˈɛ s p ə r ə t ɔ r i	r ɪ s p ˈɪ r ə t r i
Quinine	k w ˈaɪ n aɪ n	k w ɪ n ˈi n
Trachea	t r ˈeɪ k i ə	t r ə k ˈi ə

the varying pronunciations of these scientific/medical terms by American and British speakers.

The impact of these significant pronunciation divergences—in stress placement, varying numbers of syllables, and in vowel length—on speech recognition is probably fully appreciated. ASR providers should know these variants and load appropriately different ‘grammars’ (with their associated pronunciation models) into the localized software used in the US, Canada or the UK. The real problem lies with physicians and medical technologists who have learnt English (perhaps as a second or other language) outside North America or the British Isles, but who are resident in the US 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 learnt 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 individuated orthographically; it is not however 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 practice of the information age is described as, “. . . the astonishing and quite dangerous drift back to the *scriptio continua* of the ancient world, by which words are just hoiked[3] together as “all one word”. . . .” (Truss, 2003, p. 170). Such typographical rule-breaking 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. Boldface sequences show non-English spelling in 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. Some sophisticated TTS systems may solve such problem sequences as “SmoothCaps” by rule, but most will need a dictionary-based approach.

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

Table 4. Problem spelling sequences (in bold face) that never occur in 'regular' English orthography and which therefore have no cognates in English, and which may need hand-crafted entry into dictionaries for ASR and TTS.

Name	What the TTS system sees
AstraZeneca	Astrazeneca#
Aventis	
Co**zaar**©	
GlaxoSmithKline	Glaxosmithkline#
GlaxoWellcome	Glaxowellcome#
Kaiser	
Merck	
Pfizer	
Re-up-take	reuptake#
Ri**tuxan**®	
Smoo**thC**aps™	Smoothcaps#
Tarceva™	
TNKase™	Tnkase#
Trastuzumab	
Zilactin	
Zo**ō**cor	Zocor
Zyrtec®	

Transparent Drug Names: Advil®, Clarinex®, Diovan®, Flonase®, Lunelle™, Minipress®, Nasonex®, Nexium®, Paxil®, warfarin, Singulair®, Zoloft©.

Transparent names typically contain syllables that are easily identified and pronounceable by native speakers of English (e.g. clari+nex, flo+nase, mini+press, etc.). They may be separated further into those that have positive semantic associations (because they are based in familiar etymology or morphemes), and those which are less successful (meaningless) semantically:

Positive semantics: Allegra®, Alliant, Bellafem, Claritin, Prozac, Wellbutrin
Unclear semantics: Avacor, Imitrex, Naproxen, PremPro, Zōcor, Zyrtec

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

Table 5. Brand and generic drug names which are opaque or ambiguous in their intended pronunciations.

Opaque/ambiguous drug names			
	Pronunciation		Reason
Halcyon©	h ˈæ l s i ə n	*or*	obscure word (origin: Gk.)
	h ˈæ l s i ɑ n		
Levitra©	l ə v ˈi t r ə	*or*	cognate with '*in vitro*'
	l ˈɛ v ɪ t r ə		cognate with 'levitate'
lorezepam	l ə r ˈɛ z i p æ m	*or*	no cognate
	l ˈɔ r z ɪ p æ m		cognate with 'lore' + stress of 'marzipan' ?
rifabutin	r aɪ f ə b ˈu ɾ ɪ n	*or*	no cognate
	r ɪ f ə b ˈʌ ɾ ɪ n	*or*	'rifa' + cognate with 'but-in'?
	r aɪ f ə b j ˈu ɾ ɪ n		
rifampin	r ˈɪ f ə m p ɪ n	*or*	no cognate
	r ɪ f ˈæ m p ɪ n	*or*	cognate with 'rif''ampicin' (another drug)
Vasotec©	v ˈæ z oʊ t ɛ k	*or*	
	v ˈeɪ s ə t ɛ k		

Lives are at Stake

When visiting the dentist, patients are now asked to fill in a form providing a 'Medical History Update'. Questions include "Circle any of the following drugs you are taking: *Ainequan, Asendin, Elavil. . . .Tofranil*." When this author asked the dental hygienist, a native Californian speaker of US English, how to pronounce the first in the list, she did not know. It is hardly encouraging for 'lay people' when those working within the healthcare profession are at a loss for how to say the drug names. Patients should be doubly wary when a polyglot linguist can identify neither English spelling sequences nor common morphemes within drug names.

Recognizing that health caregivers are less than one step removed from lay people in their language intuitions, medical textbook publishers have joined the ranks of 'helpful hint' providers. It appears to be standard to include pronunciation help in current reference guides for practicing nurses and those in training. The following table lists a selection of drug names and the entries they received in four nursing handbooks:

Note that Saunders disambiguates a 'soft ' c' from a 'hard 'c' by giving it an 's' or a 'k' pseudo-pronunciation; however unstressed vowels are rendered with a misleading variety of representations: cf. "eh" vs. "a"; and the post-positioning of the stress mark is confusing. Springhouse chooses to show stress with bold face—comparable in effect and linguistic

Table 6. Brand and generic drug names for which pronunciations are unknowable.

Unknowable drug names		Pronunciation problems
Amiodarone	Stress?	ˈæmioʊdəroʊn *or* æmiˈɑdərən *or* əmˈiədəroʊn
Carbamazepine	Stress?	-pin *or* -paɪn
Ceterizine	Stress?	sɛ- *or* kɛ- -zin *or* -zaɪn
Cimetidine	Stress?	sɪ- *or* saɪ- -din *or* -daɪn
Cypionate	Stress?	sɪ- *or* saɪ-
Desloratadin	Stress?	
Drospirenone	Stress?	
Esomeprazole	Stress?	ɛsəmˈɛprəzoʊl *or* isoʊmˈiprəzoʊl
Estradiol	Stress?	ɛstrˈædioʊl *or* ˈɛstrədiəl *or* ˈɛstrədiɑl *or* ɛstrədˈiɑl
Itraconozole	Stress?	ɪ- *or* aɪ- -kɑn- *or* -koʊn-
Montelukast	Stress?	mɑntˈɛləkæst *or* mɑntilˈukæst
Norgestinate		nɔrdʒ- *or* nɔrg-
Thioridazine	Stress?	θiɔrɪ- *or* θaɪɔri; -zin *or* -zaɪn
Vardenafil	Stress?	

Table 7. Brand and generic drug names for which pronunciations are given in drug reference handbooks and medical textbooks.

Drug name	Pronunciation	Source
Celecoxib	(sel-eh-cox'-ib)	Saunders
Danaparoid	(dan-ah-**pear**-oid)	Springhouse
Dilantin	(di lan' tin)	Clayton & Stock
Flovoxamine maleate	(floo-VOKS-uh-meen MAL-ee-ayt)	Mosby's
Isocarboxazid	no entry	Clayton & Stock
Kanamysin	(kan-a-mye'-sin)	Saunders
Mupirocin	(**mew**-pie-ro-sin)	Springhouse
Phenytoin	(fen' e toe in)	Clayton & Stock

inaccuracy to Mosby's showing stress in all upper case.

Most curiously, the least accessible drug name (isocarboxazid) in the list above gets no help from Clayton and Stock's publisher. The drug is not listed at all in the three other nursing handbook/reference guides consulted, so its pronunciation remains a total mystery. In the 24th edition of the *Nursing 2004 Drug Handbook,* Springhouse publishers have given up entirely on the pronunciation enterprise. So we find the following unwieldy polysyllabic compounds and drugs listed, with no idea how to say them:

basiliximab
cisatracurium besylate
tranylcypromine sulfate
sargramostim

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

The information in Table 8 is unsystematic: note three different renditions of unstressed syllables, of post-positioned single quote or upper case to indicate stress, and the unjustified or inconsistent use of upper case in general. Such *ad hoc* hints are helpful neither to native nor non-native English speakers, nor to those confused by quasi-phonetic notation.

Problems with the 'unknowables' (the great majority) are not alleviated by such pseudo-pronunciations. More often than not, we are left to our own (wobbly) intuitions about stress placement, short vowel /ɪ/, long

Table 8. Brand and generic drug names for which various pronunciation notation is given by drug manufacturers and HMOs.

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

vowel /i/, or diphthong /aɪ/; '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.

Consumers may start down the slippery slope to mystification when they see T.V advertisements for prescription drugs. These ads encourage viewers to "ask your doctor whether Levitra is right for you." and to ask themselves if they "need the purple pill". The advertisements reveal *nothing* about the conditions or disease(*s*) these drugs are supposed to treat. If the curious, worried and under-informed person then wishes to find out more about these drugs, they may find more details published in magazines and newspapers. There the information congestion is only worsened by the core product information typically appearing in a tiny font, frequently printed in pale grey or otherwise transparent ink. The (ill/worried/curious) consumer then slides further into despair while trying to decipher illegible and unintelligible uses and side-effects of drugs they are unsure they ever needed.

Given these many (socio)linguistic and typographic variables, it is 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 towards more intelligent voice user interfaces."

Rolandi's concerns are mirrored in an article plaintively titled "Can you understand me now?" (Andolsen, 2004). The author predicts that "Soon records and information managers will be required to address the output of voice recognition systems as an integral part of their electronic records retention decisions." Unfortunately he relates a far-from-satisfactory conversation with a health provider's 800 number, where the caller had to explain something complex to re-order a prescription. Unusually, this caller was experienced with such systems and did not show the over-weaning anthropomorphism frequent with speech interfaces, where the capabilities are overestimated and the caller thinks they are talking to a human. The caller knew that he was talking to a voice recognition system; he also knew that the system couldn't answer his complicated question. The human-machine dialogue rapidly dissolved into a frustrating sequence of, "I would like to speak to a human being." and "I am sorry. I do not understand what you have asked. Please repeat your request." Only when the caller spoke the 'abracadabra' phrase "customer service representative" did the system respond "I will connect you with a customer service representative", and the door to the conversational cave opened. Speech recognition systems need more expansive and forgiving grammars; it also seems that humans need to know exactly what to say. If a caller is trying to determine the balance in an air-miles account, their not saying the magic words is not a disaster. But when trying to order prescription drugs and medical supplies, the lock-out caused by not knowing the requisite phrasing *is* a potential disaster.

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 here, it is not clear how errors can be avoided, let alone reduced, and how lives may be saved.

Further 'real world' skepticism, is provided in comments from a trainer of medical professionals to use speech recognition to create documents (personal communication). While enjoying a high degree of success with physicians across the US, he nevertheless reports major barriers to faster adoption of speech recognition.

1. Doctors are not good at proof-reading, no matter what size the screen is.
2. Doctors struggle for six months on average to learn how to create a medical record; or they just don't learn and their work is "junk").
3. Some of the better doctors have progressed from speech recognition to forms and macros.
4. Some of the best EMR and document-creation programs fail because of the disparity in the intellect/interest level of the creators.
5. If the computer has been set up correctly with the right specifications, and a personal vocabulary has been prepared for the individual from previous documents, and the pace of their voice understood, and, in some cases, they have been given exercises in speaking clearly, many of the errors disappear. For those errors that do not go away after all this, strong written/spoken forms take up the balance.
6. The biggest issue that speech recognition faces is the differences in the use of the English language (i.e. accents and dialects).

The key question then becomes "Should doctors use speech recognition?" Without a doubt, and yet for some different reasons than those stated above. Doctors should use speech recognition and become medical transcribers in order to become in turn better healthcare providers. His experience with national clients provides a good cross-section of medical professionals and the documents they create.

R_x for Remedies

A first, obvious and critical step needs to be taken by the pharmaceutical industry, and enforced by the FDA. They must jointly ensure that when a new drug is named, and before it is released, the name should not look or sound similar to another drug already on the market. Reports of similar names have triggered changes in the U.S., such as the name change from *Amrinone* to *Inamrinone* (USP Quarterly Review, 2000). In Australia, however, "Experience has shown that it is almost impossible to have the name of a prescription medicine changed once it has been released to the market" (Prof. R. Day, MBF News, 2004). Sometimes, confusion can be avoided if the physician remembers to use the generic name when ordering a drug, rather than its brand-named version. This practice might have prevented confusion for example between Zostrix® and Zovirax®, both of which are topical creams applied to the skin; their generic names are 'capsaicin' and 'acyclovir', respectively. But questions about the pronunciation of the generic names return us to the 'unknowable' problem discussed earlier; and physicians may be over-worked, or too tired to look up the generic names, even if they are aware of potential confusable name pairs.

Pharmacists should always call the physician any time there is uncertainty; they should familiarize themselves with all new drugs (both brand and chemical names) entering the market; and they should ask sales representatives about new drugs being approved. Physicians, nurses and pharmacists should all also try to stay current with the literature, not least so they might know how to pronounce the drug names. Acknowledging the frequency of mistakes and in attempt to divert the risk of more, some hospital facilities provide 'name-alert' warnings in the pharmacy computer system, and place alert stickers on the containers to help technicians. They may reformat the labels for easier legibility and distinctiveness, to include larger or bolder print; they may add 'tall man' lettering on the automatic dispensing machine computer screens and bins for each drug; and they may provide additional information on the labels. They may even label each drug in a confusable set with "This drug is not the same as ... ", and they may put up a warning sign at the pick-up station in the pharmacy.

None of these precautions will of course help patients who choose to order their drugs by mail or through the Internet, or who wish to place a refill order over the telephone or electronically. They place themselves in real danger if they cannot read the details about the medication, directions and warnings from the label. Recent announcements by two companies cast some rays of hope—especially for the illiterate, the elderly, and the sight-impaired, who may have extra difficulty reading the contents and tiny instructions printed on a medicine bottle (HLT News, 1 April 2004). With billions of prescriptions filled each year, pharmacies have been struggling with the issue of label readability and comprehension for many years.

iVoice's patent-pending technology allows a pharmacist to load a wave file wirelessly into "a prescription medicine container to store speech data for prescription instructions to the patient. The patient will press a button on the container to hear the instructions at any time. This system may reduce patient error and in particular make it easier for older patients and patients with reading impairments to comply with their prescriptions." (HLT News, 13 March 2004).

Wizzard's plans to use speech technology include as many as thirty different speech synthesized languages; audio volume control for people with hearing impairments; spoken reminders for 'time to take your medicine'; audio 'exceeding dosage amount' warnings; and audio refill prescription reminders. When such products are mass-marketed, the additional costs in making the bottle (and to the patients) should be minimal. The drawbacks include the need for a pharmacist or health caregiver to read label information into a recording device, when human error and non-native English speech can of course reintroduce problems. Busy pharmacists should also take the time to double-check that each bottle is 'speaking' the correct information; this step only adds to their workload. In addition, patients may need to buy a \$325.00 'reader', or use disposable pill bottle that cost \$10.00—not a 'minimal' sum. Despite three similar systems being available since 2000, adoption of the technology has been slow, and representatives of the pharmaceutical still know little about the products.

There remain several (non-linguistic) impediments to the wider adoption of digital dictation devices and talking bottles. First, healthcare professionals are understandably concerned about confidentiality/security. The potential for being overheard is greater when speaking into a handheld device because people typically raise the amplitude of the voice when speaking into a device in a busy or noisy environment (cf. cell phone users). Second, there is the fragility or fallibility of recognition accuracy, and the 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, semi-private 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 ameliorated, if not solved. The last, and most important improvement, in these speech scenarios is for users to have some usage guidance and immediate confirmation of what they have spoken.

Many early adopters in US radiology departments have since abandoned spoken record-keeping, because the need for repetition and high failure rates were simply too frustrating. A recent survey of 31 authors of papers on medical speech recognition applications underscores this less-than-rapturous reception. The authors viewed speech technology more favorably than when they had first published their papers (ten years ago on average). "However, the adoption of speech applications did not always correspond with their enthusiasm. The survey also suggested that hands-busy, eyes-busy, and mobility requirements are not always enough to offset the current limitations in speech technology. There may need to be other benefits, such as decreased medical costs and increased quality of care, or other factors, such as using a limited vocabulary." (Grasso, 2003).

Nonetheless, Philips Speech Recognition Systems reported that their dictation system is used in some European countries by more than 60% of radiologists (Speech Technology Magazine electronic newsletter NewsBlast, 10 December 2003). The product has expanded into other specialized areas, such as cardiology, pathology, endoscopy, and surgery. Clearly the speech recognition component has improved over the past fifteen years. And perhaps the working conditions of these non-US professionals provide better, quieter, privacy.

The trademark law restricts trademarks that are spelled alike or sound too similar to existing products. But until linguists can define a widely accepted distance metric for both areas (spelling and sound similarity), this area will remain contested. Additional research is required to produce trademark metrics as they apply to the naming of drugs. There remain many skeptics in the US medical profession and in the speech technology community, 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. Speech technology, like any other technology, works best when there is a compelling reason to use it. Some health-care providers may not feel this compulsion without the added reassurance of the application's ability to 'talk back' accurately. None of the current instantiations include TTS, which *does* talk 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). Further, TTS adds a safety check: by speaking the entry back, it can confirm and disambiguate potentially confusable entries that have been made using ASR, a human's spoken recording, and/or the graphical interface. Approximately 3 billion prescriptions are filled annually in the United States; and each year medication errors may be responsible for thousands of deaths. Every doctor, specialist and pharmacist would welcome the technological improvements in the health care system if they contained such counter checks and additional features as described here; and *IFF* their HMO accountants or limited partnership paid for the installation, training and set-up fees. As indicated earlier: if it cuts costs, it will come. Patients may then at least have fewer bitter *bills* to swallow.

Glossary of Acronyms

ASR	Automatic Speech Recognition
EMR	Electronic Medical Reporting
FDA	Food and Drug Administration
HMO	Health Management Organization
IPA	International Phonetic Alphabet
MER	Medication Errors Reporting
PDA	Personal Digital Assistant
TTS	Text-to-Speech
USP	United States Pharmacopoeia

Acknowledgments

I would like to thank Ken Christopher, Sandy Disner, Daryle Gardner-Bonneau, Larry Malakhoff, two anonymous reviewers, and an e-mail correspondent who works with physicians, but who wishes to remain anonymous, for their comments and contributions to this paper. All errors remain my own.

Notes

1. Rhoticized syllables contain a sequence of vowel(s) and an [r], e.g. "heart".
2. Metathesis is a phonological process whereby one sound, syllable or letter is transposed with another; for example "perfect" for "prefect".
3. "hoiked" is a British English term, meaning "to bring together (*out*, etc.) esp. with a jerk; yank." (*Shorter Oxford English Dictionary*, 5th. Edition, 2002).

References

Andolsen, A.A. (2004). Can you understand me now? *The Information Management Journal, 38*(1):52–56.

Associated Press (March 13, 2004) Company invents talking chips for prescription bottles. http://www.macon.com/mld/macon/news/nation/8175551.htm

Clayton, B.D. and Stock, Y.N. (2004). *Basic Pharmacology for Nurses*. 13th Edn. Mosby.

Grasso, M.A. (2003). The long-term adoption of speech recognition in medical applications. In *Proceedings of the 16th IEEE Symposium on Computer-Based Medical Systems (CBMS 2003)*, 257–262.

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

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

HLT News. (April 1, 2004). Speaking bottles, http://www.hltcentral.org/news/index.

Hodgson, B.B. and Kizior, R.J. (2003). *Saunders Nursing Drug Reference 2004*. W.B. Saunders.

MBF News (March 30, 2004) Similar sounding drug names raise medication error risk. http://www.mbf.com.au/about/mediareleases/medicationrisk.html.

Nursing 2004 Drug Handbook. 24th edn., Lippincott, Springhouse Publishing Company.

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

Springhouse Nurse's Drug Guide 2004. Lipincott.

Skidmore-Roth, L. (2003). *Mosby's Nursing Drug Reference*. Mosby

Truss, L. (2003). *Eats Shoots & Leaves. The Zero Tolerance Guide to Punctuation*. London: Profile Books.

Voice of the Injured (2004). Prescription errors due to similar drug names. http://www.voiceoftheinjured.com/a-mp-similar-drug-names.html.

United States Pharmacopeial Convention, Inc.(1999). Use Caution, Avoid Confusion. U.S. Pharmacopeia Quality Review no. 76 (2001). www.usp.org/patientSafety/briefsArticlesReports/practitionerReportingNews/prn1122004-02-25.html.

U.S. Pharmacopeia Quality Review No. 73 (March, 2000).

U.S Pharmacopeia (2001). USP identifies more than 750 unique drug names reported to cause confusion. http://vocuspr.vocus.com/vocuspr30/xsl/uspharm/Profile.asp?Entity=PRAsset&EntityID=3139&XSL=PressRelease.