

Estimating the Effectiveness of Automatic Vocabulary Expansion

TR 29.3452
October 6, 2001

James R. Lewis

IBM Voice Systems

West Palm Beach, Florida

Abstract

This report documents the motivation, method and results of seven experiments conducted to investigate the properties of automatic document analysis (for the purpose of automatic vocabulary expansion). The results indicated that automatic vocabulary expansion on corrected text improved the accuracy of text dictated in the future, as long as the future text was similar to the analyzed text. None of the manipulations had a measurable effect (either good or bad) when the analyzed text was uncorrected dictation. These results were the same regardless of whether the system interpreting the speakers' audio used a trained or untrained acoustic model.

ITIRC Keywords

Automatic document analysis
Automatic vocabulary expansion
Recognition accuracy
Recognition errors
Error correction
Speech dictation
ViaVoice '98
ViaVoice Millennium

Contents

Introduction	1
General Method	3
Participants	3
Materials and General Procedure.....	3
Experiment 1: Normal Vocabulary Expansion without Updating the User Cache Language Model.....	5
Motivation.....	5
Method.....	5
Results	5
Experiment 2: Normal Vocabulary Expansion with User Cache Language Model Updating	7
Motivation.....	7
Method.....	7
Results	7
Experiment 3: Normal Correction	9
Motivation.....	9
Method.....	9
Results	9
Experiment 4: Automatic Vocabulary Expansion of Uncorrected Dictation for Enrolled Speakers	11
Motivation.....	11
Method.....	11
Results	11
Experiment 5: Automatic Vocabulary Expansion of Uncorrected Dictation for Unenrolled Speakers	13
Motivation.....	13
Method.....	13
Results	13
Experiment 6: Automatic Vocabulary Expansion of Corrected Dictation for Unenrolled Speakers	15
Motivation.....	15
Method.....	15
Results	15
Experiment 7: Automatic Vocabulary Expansion for Dictation Contaminated with Linguistijunk.....	17
Motivation.....	17
Method.....	17
Results	17
Discussion.....	19
References.....	21
Appendix A. Test Scripts	23
Ford	24
Nick.....	25
Laser	26
Tax	27
Bart.....	30
Hoff.....	31
Ad.....	32
Health	33
Appendix B. Linguistijunk Text	35

Introduction

Recognition accuracy is an important attribute of speech dictation systems, but at all except the highest levels of recognition accuracy the speed of correction is as or more important for determining throughput (Lewis, 1999). Careful examination of the correction behavior of people using the major speech dictation products has shown that users typically prefer to make their corrections by typing directly into the text rather than selecting text and using a correction dialog (Karat, Halverson, Horn, & Karat, 1999). In most speech dictation systems, though, the system only 'learns' from corrections that users make with the correction dialog. An alternative way for a system to identify words that are out-of-vocabulary (OOV) and to add document-specific information to the language model (LM) is to run a program designed for the specific purpose of document analysis.

Most modern speech dictation systems (such as IBM ViaVoice¹, starting with ViaVoice '98) provide a document analysis function (called, in ViaVoice, Vocabulary Expansion). One purpose for this function is to let users analyze existing documents during installation for the purpose of improving the product's out-of-box recognition accuracy by identifying OOV words in a user's selected documents and biasing the LM toward the user's writing style. If it were possible to modify this function to allow automatic analysis of documents after a dictation session, then users might be able to make their corrections directly in the text of their dictated documents and still have the system learn about the changes. The system would acquire the same information as that formerly obtained through the correction dialog.

This report describes seven experiments conducted during late 1998 and early 1999 to investigate the potential effects of automatic document analysis. The specific experimental questions were:

- Vocabulary expansion without enrolled user cache (LM) updating: What if the automatic document analysis only added new words, but did not affect the user's personal LM (also known as the 'user cache') for users who have enrolled (provided audio for a customized acoustic model)?
- Vocabulary expansion with enrolled user cache (LM) updating using corrected dictated text: What if the automatic document analysis also updated the user's personal LM?
- Correction by enrolled users: How would automatic document analysis compare with the standard correction procedure for the purpose of improving future dictation accuracy?
- Vocabulary expansion with enrolled user cache (LM) updating using uncorrected dictated text: What if enrolled users didn't correct the text before automatic document analysis?
- Vocabulary expansion with unenrolled user cache (LM) updating using uncorrected dictated text: What if unenrolled users didn't correct the text before automatic document analysis?

¹ IBM and ViaVoice are registered trademarks of International Business Machines Corp.

- Vocabulary expansion with unenrolled user cache (LM) updating using corrected dictated text: What if unenrolled users did correct the text before automatic document analysis?
- Vocabulary expansion with contaminated user cache (LM): What if an external audio source caused the accidental production of random text that the system automatically analyzed?

The outcomes of these experiments affected the decision regarding the deployment of development resources to the coding of an automatic document analysis feature for future versions of ViaVoice.

General Method

Participants

The participants were four males and four females (all adult native speakers of American English), all of whom had provided recordings of test scripts (using SoundForge², 22 kHz, 16-bit mono PCM recording) and enrollments for a previous study of ViaVoice '98 enrollment.

Materials and General Procedure

The computer used to process the recordings was an IBM Pentium Pro 200 (48 MB memory, Windows 95, hundreds of MB of free hard disk space) -- the same system on which participants had enrolled and recorded test scripts. The version of ViaVoice installed on the computer had a customized function designed to let an experimenter define a wave file for speech-to-text transcription.

The eight test scripts (selected from a small library of test scripts) included words, punctuation, and formatting commands. The dependent measure in these experiments was the simple word accuracy for each test script (the number of words correctly recognized divided by the total number of words in the test script). The text of the test scripts appears in Appendix A.

To create a set of texts related to another set, I divided the first four scripts approximately in half. The rationale for doing this was to define the first half as a treatment set and the second half as a related set (making the reasonable assumption that the two halves of a test script would have a stronger relationship with each other than with any other test script). The second four scripts contained material unrelated to the first four, and acted as a control set of texts. Table 1 shows the number of words and OOV measurements for each test script.

Before running the experiments described in this report, the tester ran all recorded files through the customized transcription function to determine the baseline accuracy for each file (the four files that the experimenter would treat in the experiments, the four files of related text, and the four control files).

² SoundForge is a trademark of The Sonic Foundry, Inc.

Table 1. Word Counts and Out-of-Vocabulary Measurements for Test Scripts

Script	Use	Words	# OOV	% OOV	OOV Words
<i>Ford-1</i>	Treated	193	0	0.0	<none>
<i>Nick-1</i>	Treated	134	3	2.2	Nickell, Roswell, lifecast
<i>Laser-1</i>	Treated	211	7	3.3	corneal (2), LVC, PRK, excimer, keratectomy, photorefractive
<i>Tax-1</i>	Treated	423	0	0.0	<none>
<i>Total</i>		961	10	1.0	
<i>Mean</i>		240.3	2.5	1.4	
<i>Ford-2</i>	Related	160	1	0.6	Gilbreth
<i>Nick-2</i>	Related	154	2	1.3	Roswell (2)
<i>Laser-2</i>	Related	190	0	0.0	<none>
<i>Tax-2</i>	Related	448	1	0.2	noncharitable
<i>Total</i>		952	4	0.4	
<i>Mean</i>		238	1	0.5	
<i>Bart</i>	Control	276	3	1.1	Bartholomew, Kurtz, spate
<i>Hoff</i>	Control	303	0	0.0	
<i>Ad</i>	Control	250	0	0.0	
<i>Health</i>	Control	73	0	0.0	
<i>Total</i>		902	3	0.3	
<i>Mean</i>		225.5	0.8	0.3	

Experiment 1: Normal Vocabulary Expansion without Updating the User Cache Language Model

Motivation

The goal of this study was to characterize the effect of normal vocabulary expansion (finding and adding acoustic models for OOV words) but without updating the language model for the speaker (in other words, updating the user cache). The term "normal vocabulary expansion" refers to the analysis of previously typed documents. This had the effect of isolating the influence of simply adding OOV words to the user cache and providing an estimate of the effect of this very conservative approach to vocabulary expansion.

Method

The strategy in this experiment was to run the treatment text (the first half of the first four documents) through the ViaVoice Vocabulary Expander after restoring a participant's enrollment (saved previously with no information in the user cache). To make it easier to run the text through Vocabulary Expander, the tester created a single file from the first half of the scripts in the first set (Ford1, Nick1, Laser1, Tax1), using the correctly typed source text. The tester answered "No" to the prompt that asked whether the analyzed text was representative of the speaker's style. This caused the system to add all OOV words (with pronunciations provided by the tester), but no LM information, to the user cache.

Results

As expected, the effect of vocabulary expansion on the same text run through the Vocabulary Expander was significant ($t(7)=4.64, p=.002$). This effect, although statistically significant, is not very meaningful because users normally won't dictate exactly the same text that they process with the Vocabulary Expander. It is interesting (though not surprising) that the amount of improvement in the treated text (1.5%) was virtually identical to the percentage of OOV words calculated in the first table for the treated texts (1.4%).

The accuracy of related text showed a statistically reliable improvement of 0.7% ($t(7)=2.72, p=.03$ -- a reduction in error of 6.4%). The percentage of OOV calculated for this text in the first table was 0.5%. However, only half of this amount (0.25%) contained words added to the vocabulary when doing vocabulary expansion on the first half, so the remaining portion of the improvement must be due to other factors.

The accuracy of unrelated text showed a statistically nonsignificant decline of 0.2% ($t(7)=1.12, p=.30$).

Even without updating LM data in the user cache, vocabulary expansion improved the dictation accuracy of text related, but not identical, to the text processed with the vocabulary expander. The treatment had no significantly adverse effect on the dictation accuracy of unrelated text.

Experiment 2: Normal Vocabulary Expansion with User Cache Language Model Updating

Motivation

The goal of this experiment was to characterize the effect of letting the vocabulary expansion procedure add OOV words and LM information from the analyzed text to the user cache. This had the effect of combining the influence of adding OOV words and all available LM information to the user cache -- a less conservative approach to vocabulary expansion.

Method

The procedure for this experiment was identical to that of Experiment 1, except the tester indicated "Yes" to the prompt that asked if the analyzed text was representative of the speaker's style. Doing this added all OOV words to the user cache and updated the user cache with the LM information from the analyzed text.

Results

As expected, the effect of vocabulary expansion on the same text run through the Vocabulary Expander was significant ($t(7)=6.50, p=.0003$). This effect, although statistically significant, is not very meaningful because users normally won't dictate exactly the same text they ran through Vocabulary Expander.

The accuracy of related text showed a statistically reliable improvement of 1.0% ($t(7)=2.08, p=.08$ -- a reduction in error of 9.1%).

The accuracy of unrelated text showed a statistically nonsignificant decline of 0.4% ($t(7)=1.11, p=.30$).

Vocabulary expansion with user cache updating improved the dictation accuracy of text related, but not identical, to the text run through the vocabulary expander. The treatment had no significantly adverse effect on the dictation accuracy of unrelated text. Adding the LM information seemed to have a slight effect of increasing the variability of improvement (indicated by the difference in the observed significance level of the t -tests conducted on the accuracy changes in the related texts -- .03 in Experiment 1 and .08 in Experiment 2).

Experiment 3: Normal Correction

Motivation

The goal of this experiment was to characterize the effect of normal correction on the future dictation accuracy of related and unrelated texts. This provided an estimate of the benefit gained by the practice of correcting dictated text using the ViaVoice correction dialog.

Method

The procedure to assess the effect of normal correction was a little different. The tester ran one of the recordings for the first set of scripts through the customized transcription function. After transcribing the file, he corrected all misrecognitions in the first half only, and did this for all four treated files. This resulted in the types of changes to the user cache that happen as a consequence of normal correction -- adding OOV words and providing LM updating for the text in the immediate vicinity of the correction (generally plus and minus two words from the target word). The tester then ran all eight test scripts for that participant through the customized transcription function.

Results

As expected, the effect of correction on redictating text that a user had already dictated and corrected was significant ($t(7)=6.29$, $p=.0004$). This effect, although statistically significant, is not very meaningful because users normally won't dictate exactly the same text that they have already dictated and corrected.

The accuracy of related text showed a statistically reliable improvement of 0.8% ($t(7)=2.50$, $p=.04$ -- a reduction in error of 7.3%).

The accuracy of unrelated text showed a statistically nonsignificant decline of 0.1% ($t(7)=0.26$, $p=.80$).

Thus, normal correction improved the dictation accuracy of text related, but not identical, to text previously dictated and corrected, and with roughly the same magnitude of effect as vocabulary expansion without LM updating of the user cache (the situation studied in Experiment 1). The treatment had no significantly adverse effect on the dictation accuracy of unrelated text.

Experiment 4: Automatic Vocabulary Expansion of Uncorrected Dictation for Enrolled Speakers

Motivation

Experiments 1 and 2 investigated the effect of performing vocabulary expansion on existing documents, finding and adding all OOV words and either adding (Experiment 2) or failing to add (Experiment 1) the LM data from the analyzed documents to the user cache. Experiment 3 investigated the effect of normal correction on dictated text. None of these experiments addressed the possibility that a user might save a file of uncorrected dictation with the intention of correcting misrecognized text at a later time. Because the plan of record was to perform automatic document analysis when the user saved a file, there was concern that leaving the uncorrected text in the document might contaminate the LM data in the user cache, leading to reduced accuracy in future dictation. The purpose of Experiment 4 was to characterize the effect of this very likely situation on future dictation.

Method

The procedure for this experiment was a blend of the procedures used in Experiments 2 and 3. The tester ran one of the recorded files for the first set of scripts through the customized transcription function. After transcribing the file, he copied the uncorrected dictated text from the first half only into a separate file. He did this for all four files, adding the uncorrected first half for each document into the same file. Then, after restoring a participant's enrollment (with no data in the user cache), he ran the file with the uncorrected dictation through Vocabulary Expander. This had the effect of contaminating the user cache with the uncorrected dictation, simulating what would happen if automatic vocabulary expansion was performed when a user saved a dictation file without having corrected the misrecognitions in the dictated text.

Results

Surprisingly, the effect of this manipulation on the accuracy of the text analyzed with the Vocabulary Expander was positive and marginally statistically significant, with an average improvement in accuracy of 0.5% ($t(7)=1.91, p=.10$). Even so, this effect is of little practical significance because users would rarely dictate exactly the same text twice.

The accuracy of related text showed a statistically nonsignificant improvement of 0.2% ($t(7)=0.66, p=.53$).

The accuracy of unrelated text showed a statistically nonsignificant change of 0.0% ($t(7)=0.09, p=.93$).

Even though contaminated with uncorrected misrecognitions, the simulated automatic vocabulary expansion did not cause any apparent degradation in recognition accuracy for the related or unrelated text.

Experiment 5: Automatic Vocabulary Expansion of Uncorrected Dictation for Unenrolled Speakers

Motivation

Despite the success of Experiment 4, there was a fear that if recognition accuracy was lower (for example, if users failed to enroll), automatic vocabulary expansion of uncorrected text might cause a serious degradation in accuracy. The purpose of this experiment was to provide an estimate of the effectiveness of automatic vocabulary expansion for the situation in which speakers have not enrolled and do not correct their errors immediately, thus putting uncorrected dictation into the user cache at a rate higher than that of Experiment 4.

Method

Other than using a speaker-independent (unenrolled) acoustic model, the method of this experiment was the same as that of Experiment 4.

Results

The accuracy of related text showed a statistically and practically nonsignificant improvement of 0.1% ($t(7)=0.45$, $p=.67$). The accuracy of unrelated text showed a statistically and practically nonsignificant improvement of 0.2% ($t(7)=0.89$, $p=.40$). Even though contaminated with an average of about 15% word errors, the simulated automatic vocabulary expansion did not cause any apparent degradation in recognition accuracy for either related or unrelated text.

The worst case performance was from Speaker 7, who had a tendency to speak very quickly and not very clearly. Her overall accuracy was about 70% (30% word error rate). For her, the accuracy of related text after running her uncorrected text through vocabulary expander was 0.2% higher than her baseline unenrolled accuracy. For the control text, her accuracy was 1.5% higher than her baseline unenrolled accuracy. This suggested that even if the word error rate were around 30% and automatic vocabulary expansion occurred on uncorrected text, there would be no negative effect on future accuracy when dictating either related or unrelated (control) text.

Experiment 6: Automatic Vocabulary Expansion of Corrected Dictation for Unenrolled Speakers

Motivation

The purpose of this experiment was to investigate the possibility that if the accuracy was lower (for example, if users failed to enroll), automatic vocabulary expansion of corrected text might improve accuracy to an even greater extent than occurred in Experiment 2.

Method

Other than using a speaker-independent (unenrolled) acoustic model, the method of this experiment was the same as that of Experiment 2.

Results

The accuracy of related text showed a statistically significant improvement of 1.0% ($t(7)=5.45, p=.0009$). The accuracy of unrelated text showed a statistically nonsignificant change of 0.5% ($t(7)=1.14, p=.29$). The amount of improvement in accuracy for the related text was exactly the same as that observed in Experiment 2 (which used full enrollments and had significantly lower base word error rates). The treatment had no significantly adverse effect on the dictation accuracy of unrelated text.

The data did not support the hypothesis that the amount of improvement would be greater as a function of lower baseline accuracy. Given correct text (Experiments 2, 3 and 6), automatic vocabulary expansion improved the accuracy of subsequently dictated text related to the treated text, whether or not users had enrolled. Given uncorrected text (Experiments 4 and 5), automatic vocabulary expansion did no apparent harm, whether or not users had enrolled.

Experiment 7: Automatic Vocabulary Expansion for Dictation Contaminated with Linguistijunk

Motivation

The purpose of this experiment was to address a concern about a scenario that, while unlikely, could happen with automatic vocabulary expansion. The proposed design for ViaVoice's automatic vocabulary expansion was to analyze documents only when users saved dictation files. In most cases, if a file contained a lot of linguistijunk (quasi-random words produced by the recognizer when it interprets background noise as dictated text), then a user wouldn't save it. If a user did save such a file, though, then the linguistijunk would contaminate the user cache. The prevailing belief was that, because linguistijunk would randomly follow patterns established in the language model, contaminating the user cache with linguistijunk would not affect recognition accuracy. Despite this belief, it seemed prudent to investigate what would happen.

Method

To investigate the amount of damage such contamination could do to recognition accuracy, I started a dictation session and put the microphone next to a speaker playing music. Over the course of about 30 minutes, this produced a file with 1582 words of linguistijunk (see Appendix B). To get linguistijunk treatment scores, we did the following for each speaker:

- Restored the speaker's enrollment
- Ran the linguistijunk file through vocabulary expander
- Ran the test scripts selected for that speaker through the transcription system
- Scored the results

Note that the concept of related text doesn't apply to linguistijunk contamination because the linguistijunk has no relationship to any of the test texts. For this evaluation, the comparison was that between the baseline and contaminated word accuracy scores for fifteen test scripts (selected using the criterion of sampling from a wide range of baseline accuracies).

Results

The correlation between the baseline and contaminated accuracy scores was very high ($r=.99$, $p=.0000001$). The mean difference between the baseline and contaminated accuracy percentages was a nonsignificant 0.22 ($t(14)=1.0$, $p=.34$).

This experiment addressed the concern about a potentially harmful consequence of automatic vocabulary expansion. Because the language model certainly plays a significant role in what the recognizer produces as linguistijunk, the introduction of this type of text into the user cache had no harmful (or beneficial) effect.

Discussion

An interesting pattern emerged from the results of these seven experiments. Automatic vocabulary expansion on corrected text improved the accuracy of text dictated in the future, as long as the future text was similar to the analyzed text. None of the manipulations had a measurable effect (either good or bad) when the analyzed text was uncorrected dictation (including linguistijunk). These results were the same regardless of whether the system interpreting the speakers' audio used a trained or untrained acoustic model.

Based on these results, we decided to include automatic vocabulary expansion in the ViaVoice product line, starting with ViaVoice Millennium. The major consequence of this to a user of this speech dictation product is that it is no longer necessary to use the correction dialog to receive the long-term benefits of correcting dictation errors. Users can, if they choose, make their corrections directly in the text of the document, using either the keyboard or dictation. This design addresses one of the major usability problems reported by Karat et al. (1999). Furthermore, because there are fewer dialogs to manipulate when editing text directly in a document, it is likely that the average speed of correction would also improve, leading to a substantial improvement in text throughput when using speech dictation to produce a document (Lewis, 1999).

References

Karat, C. M., Halverson, C., Horn, D., and Karat, J. (1999). Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *CHI '99 Conference Proceedings* (pp. 568-575). Pittsburgh, PA: Association for Computing Machinery.

Lewis, J. R. (1999). Effect of error correction strategy on speech dictation throughput. In *Proceedings of the Human Factors and Ergonomics Society* (pp. 457-461). Santa Monica, CA: Human Factors and Ergonomics Society.

Appendix A. Test Scripts

This appendix contains the test scripts used in the experiments. Words appear in normal typeface. Punctuation and formatting commands are in bold face. The phrase <**BREAK
HERE**> indicates the halfway points for the treated scripts (Ford, Nick, Laser and Tax).

Ford

June twenty first nineteen ninety seven **NEW-LINE**
Four o'clock p.m. **NEW-PARAGRAPH**

Dear Mrs. **CAPITAL-LETTER Ford COLON : NEW-PARAGRAPH**

I recently acquired some copies of an old book entitled **CAPITAL-LETTER** Typing **CAPITAL-LETTER** Behavior **COMMA** , by August Dvorak **COMMA** , published in nineteen thirty six **PERIOD** . I have in my possession over five hundred copies of this book **COMMA** , which I would be glad to sell to you for five thousand dollars **PERIOD** . If you want to pursue this matter **COMMA** , you **UPPERCASE-ON MUST CONTACT ME UPPERCASE-OFF** no later than the end of business next week **EXCLAMATION-POINT !** If I don't hear from you by June twenty eighth I will look for another buyer **PERIOD** . By the way **COMMA** , I thought you might find the following observations interesting **PERIOD** . This book's a classic **PERIOD** .
NEW-PARAGRAPH

OPEN-QUOTE "What is the oldest **COMMA** , greatest invention from the past **QUESTION-MARK ?** In considering **COMMA** , you may suddenly exclaim **COMMA** , I have it **DASH -** it really is language **EXCLAMATION-POINT !** During primitive centuries **COMMA** , while men were inventing certain sounds to help control one another's actions **COMMA** , they stumbled upon devices which would carry such human speech further than the unaided human voice **PERIOD** . When people were still savages **COMMA** , a few crude pictures served that purpose **PERIOD** . Some ridiculous drawing **DASH -** perhaps a crude map that designated a meeting place **DASH -** was the only love letter a primitive girl might expect from her boyfriend **PERIOD**. **CLOSE-QUOTE** "
NEW-PARAGRAPH

<**BREAK HERE**>

OPEN-QUOTE "Outside the schools **COMMA** , typewriters have been widespread for over a generation **PERIOD** . Yet as a student typist you now make your first formal contact with a writing machine **EXCLAMATION-POINT !** In all probability as a youngster you **COMMA** , too **COMMA** , once took a good position and practiced easy **COMMA** , sideways writing movements across a sheet of paper **PERIOD** . Rhythm **COMMA** , as always **COMMA** , helps your speed **PERIOD** . However **COMMA** , if you desire finer and faster writing **COMMA** , you can use a typewriter **PERIOD** . Business men who require high quality or quantity writing buy machines to produce it **PERIOD** . Why else did men invent **DASH -** amid other possibilities **DASH -** the typewriter **QUESTION-MARK ?** The words that you have so painfully organized are swiftly and precisely organized for you by this machine **PERIOD** . **CLOSE-QUOTE** "
NEW-PARAGRAPH

Don't you agree that this is interesting material **QUESTION-MARK ?** Although almost no one uses the Dvorak keyboard **COMMA** , the book might be of historical interest to your customers **PERIOD** . If you need to contact me about this **COMMA** , please call me at **PHONE-NUMBER** four nine five nine zero three three during business hours **PERIOD** . **NEW-PARAGRAPH**

Respectively yours **COMMA** , **NEW-PARAGRAPH**

Frank Gilbreth

Nick

September thirteenth nineteen ninety six **NEW-LINE**
Ten thirty a.m. **NEW-PARAGRAPH**

Dear Mr. Nickell **COLON : NEW-PARAGRAPH**

I read your latest article **OPEN-QUOTE “ CAPITALIZE-ON Alien Autopsy Hoax CAPITALIZE-OFF CLOSE-QUOTE ”** in the **CAPITAL-LETTER Skeptical CAPITAL-LETTER Inquirer** and found it to be **UPPERCASE MOST UPPERCASE INFORMATIVE EXCLAMATION-POINT !** Do you have a set fee for lecture engagements **QUESTION-MARK ?** I am in the process of hiring speakers for a series of seminars and hope three thousand dollars is a great enough incentive to have you speak for an hour to a group of **QUOTE “reality buffs QUOTE ”** from Roswell **COMMA , CAPITAL-LETTER New Mexico PERIOD .** The following paragraphs are what piqued my interest **PERIOD . NEW-PARAGRAPH**

OPEN-QUOTE “None of us were of the opinion that we were watching a real alien autopsy COMMA , or an autopsy on a mutated human which has also been suggested PERIOD . We all agreed that what we were seeing was a very good fake body **COMMA ,** a large proportion of which had been based on a lifecast **PERIOD .**

<BREAK HERE>

Although the nature of the film obscured many of the things we had hoped to see **COMMA ,** we felt that the general posture and weighting of the corpse was incorrect for a body in a prone position and had more in common with a cast that had been taken in an upright position **PERIOD . CLOSE-QUOTE ” NEW-PARAGRAPH**

OPEN-QUOTE “The Roswell myth should be permitted to die a deserved death PERIOD . Whether or not we are alone in the universe will have to be decided on the basis of better evidence than that provided by the latest bit of Roswell fakery **PERIOD .** Television executives have a responsibility not to confuse programs designed for entertainment with news documentaries **PERIOD . CLOSE-QUOTE” NEW-PARAGRAPH**

If speaking at our seminar is a possibility please call me at **PHONE-NUMBER** five nine six eight seven six zero during business hours **PERIOD .** I can assure you an audience of about two hundred people **COMMA ,** and you will find this group’s enthusiasm will make this engagement worthwhile **PERIOD . NEW-PARAGRAPH**

Looking forward to your reply **COMMA , NEW-PARAGRAPH**

Eugene Emory

Laser

Photorefractive keratectomy **OPEN-PAREN (PRK CLOSE-PAREN) COMMA** , commonly referred to as Laser Vision Correction **OPEN-PAREN (LVC CLOSE-PAREN)** is a remarkable new medical development which can eliminate or significantly reduce the need for glasses or contact lenses in patients with nearsightedness **PERIOD** . Laser vision correction utilizes the micro **DASH** – precision of a computer controlled excimer laser to reshape the outer surface of the cornea and thus correct nearsightedness in a matter of seconds **PERIOD** . **NEW-PARAGRAPH**

After many years of research studies **COMMA** , laser vision correction has been approved by the **CAPITALIZE-THIS** Food and **CAPITALIZE-ON** Drug Administration **OPEN-PAREN (FDA CLOSE-PAREN) PERIOD** . **CAPITALIZE-OFF** The laser is already in use in forty-five countries with several hundred thousand people already having had their nearsightedness treated **PERIOD** . **NEW-PARAGRAPH**

The laser produces an invisible beam of ultraviolet light which can remove microscopic amounts of corneal tissue without causing damage to surrounding cells **PERIOD** . This elegant surgical tool is so precise it can etch microscopic layers off a human hair **PERIOD** . To achieve most corrections **COMMA** , a microscopic amount of corneal tissue **DASH** – approximately a third the thickness of a human hair **DASH** – is sculpted from the surface of the cornea **PERIOD** . The beam is precisely directed onto the surface of the eye in a series of ten pulses per second **PERIOD** . Each pulse removes a quarter of a micron of tissue **OPEN-PAREN (one hundred thousandths of an inch CLOSE-PAREN) PERIOD** . The laser correction is thus extremely precise **PERIOD** . **NEW-PARAGRAPH**

<BREAK HERE>

The laser is specifically programmed for your prescription **PERIOD** . The treatment which lasts less than thirty seconds is painless **PERIOD** . The discomfort after correction **COMMA** , which is usually mild **COMMA** , lasts approximately thirty-six **DASH** – forty-eight hours and is generally relieved with pain medications **PERIOD** . **NEW-PARAGRAPH**

Laser vision correction has achieved excellent results with the majority of patients no longer dependent on corrective lenses after the treatment **PERIOD** . After one year **COMMA** , over ninety-eight percent of people with mild to moderate nearsightedness who had undergone laser vision correction in the **CAPITALIZE-ON** FDA Clinical Trials **CAPITALIZE-OFF** had vision good enough to pass a drivers' license vision test without corrective lenses **PERIOD** . Since the final result depends on your own healing process **COMMA** , the end result cannot be guaranteed **PERIOD** . However **COMMA** , the lower the degree of correction required **COMMA** , the better the probability of achieving the desired results **PERIOD** . Some people will not be able to totally eliminate their need for corrective lenses but should be able to use much lower strength prescription and use for only certain tasks **DASH** – thus significantly improving ones lifestyle **PERIOD** . **NEW-PARAGRAPH**

As with any surgical procedure there are potential risks and side effects **PERIOD** . They are rare but should be discussed by your surgeon or eye care specialist **PERIOD** . See Q&A **PERIOD** .

Tax

BOLD-ON UPPERCASE-ON

NEWS ABOUT THE NEW TAX LAW OF 1997

UPPERCASE-OFF BOLD-OFF NEW-PARAGRAPH

Several provisions in the new tax agreement recently signed by President Clinton will have an impact on personal financial planning and charitable giving **PERIOD** . Here are a few notes on the changes **COMMA** , with the caveat **BOLD-ON** that everyone should seek professional tax advice before taking any action **BOLD-OFF PERIOD** . Congress did not simplify things **COMMA** , and the dates of your 1997 transactions will subject you to different rules before and after the passage of the bill **PERIOD** . **NEW-PARAGRAPH**

BOLD-THIS UPPERCASE-THIS NOTE COLON : In most of the following items **COMMA** , the new law includes qualifications and conditions too detailed to be discussed here **PERIOD** . These are general remarks **PERIOD** . **BOLD-ON** You should always consult with your professional advisors on these matters. **BOLD-OFF NEW-PARAGRAPH**

BOLD-ON UPPERCASE-ON

CAPITAL GAINS TAX CUT

UPPERCASE-OFF BOLD-OFF NEW-PARAGRAPH

OPEN-QUOTE “Capital **CAPITALIZE-THIS** Gain **CLOSE-QUOTE** ” is the term used to describe an increase in the value of an asset **OPEN-PAREN** (an **QUOTE** “appreciated asset **QUOTE** ” **CLOSE-PAREN**) **PERIOD** . Generally speaking **COMMA** , when you sell something that has increased in value **OPEN-PAREN** (say **COMMA** , stocks or art or land **CLOSE-PAREN**) you must pay a tax on the difference between what you got for it and what you paid for it **PERIOD** . If you owned the asset for less than one year **COMMA** , that **QUOTE** “short **DASH** – term **QUOTE** ” gain is taxed at the same rate as your regular income **PERIOD** . If you owned the asset for one year or more **COMMA** , your **QUOTE** “mid **DASH** – term or long **DASH** – term **QUOTE** ” gain is taxed at rates that have now changed **PERIOD** . **NEW-PARAGRAPH**

An appreciated asset is a good thing to give because your gift will be valued at its new higher price when you take your charitable deduction **PERIOD** . Suppose you have enjoyed some of the growth in the stock markets in recent years **COMMA** , and a stock you bought three years ago for one thousand dollars is now worth three thousand dollars **PERIOD** . Instead of paying capital gains tax on the two thousand dollar gain from your sale of the appreciated asset **COMMA** , you can deduct the entire three thousand dollars from your taxable income **PERIOD** . The amount you save **OPEN-PAREN** (by not paying taxes **CLOSE-PAREN**) is determined by the tax rates that have just changed **PERIOD** . See the chart below **PERIOD** . **NEW-PARAGRAPH**

The top rate for long **DASH** – term capital gains is reduced from twenty-eight percent to twenty percent and the long **DASH** – term holding period has been increased to eighteen months **PERIOD** . **UPPERCASE-THIS BUT**: the twenty-eight percent rate still applies to gain on the sale of assets held more than one year but less than eighteen months, and it still applies to gain on the sale of collectibles such as art and antiques held for more than one year **PERIOD** . The gain on the sale of assets held less than twelve months is still taxed at regular income tax rates **PERIOD** . Real estate is subject to different rules not discussed here **PERIOD** . **NEW-PARAGRAPH**

<BREAK HERE>

It is important to note that only the **BOLD-THIS** rate of tax on capital gains has changed **COMMA** , so the amount of capital gains tax avoided by donors may be reduced **PERIOD** . You can still deduct the full fair market value of any assets you contribute **COMMA** , if you have held them at least a year **COMMA** , so all the appreciation you've accumulated will help reduce your income taxes when you use it to make a gift **PERIOD** . **NEW-PARAGRAPH**

For some donors who seek UNDERLINE-ON lifelong income from a trust or annuity UNDERLINE-OFF **COMMA** , the new law will make it possible to receive parts of that income at a capital gains rate that is significantly lower than their ordinary income tax rate **PERIOD** . **NEW-PARAGRAPH**

UPPERCASE-ON BOLD-ON

NEW TAXES ON THE APPRECIATION OF YOUR HOME
BOLD-OFF UPPERCASE-OFF NEW-PARAGRAPH

The new law repeals two longstanding tax benefits for homeowners **PERIOD** . On or before May sixth 1997 **COMMA** , anyone selling their principal residence could postpone the recognition **OPEN-PAREN** (taxability **CLOSE-PAREN**) of any gain in its value by purchasing another home of equal or greater value **PERIOD** . Moreover **COMMA** , homeowners age fifty-five or older were entitled to a one **DASH** – time exclusion of up to one-hundred-twenty-five thousand dollars in gain on the sale of their home **PERIOD** . **NEW-PARAGRAPH**

Instead **COMMA** , the new law permits the exclusion of up to two-hundred-fifty thousand dollars **OPEN-PAREN** (five-hundred thousand dollars if married and filing a joint return **CLOSE-PAREN**) of gain realized on the sale of a principal residence after May sixth 1997 **PERIOD** .

While this will benefit many people **COMMA** , it will be a significant new tax burden on those who have held their residence for a number of years in a rising real estate market **PERIOD** . Any gain in excess of the limits will be taxed as a capital gain **PERIOD** . If that is your case **COMMA** , you may benefit significantly from making a deferred gift of a portion of the value of your residence to reduce your gain to an amount below the taxable threshold **PERIOD** . In this way you can turn a difficult tax burden into both a charitable gift and increased retirement income **PERIOD** . **NEW-PARAGRAPH**

BOLD-ON UPPERCASE-ON

ESTATE AND GIFT TAX PLANNING

UPPERCASE-OFF BOLD-OFF NEW-PARAGRAPH

The **QUOTE** “unified credit **COMMA** , **QUOTE** ” **OPEN-PAREN** (the first six-hundred thousand dollars of your estate that is not subject to estate taxes **CLOSE-PAREN**) **COMMA** , will increase to six-hundred-twenty-five thousand dollars in 1998 and will rise each year until it reaches one million dollars in 2006 **PERIOD** . The tax exemption on noncharitable gifts to individuals **OPEN-PAREN** (that is **COMMA** , you and your spouse can each give ten-thousand dollars **SLASH** / year to anyone **COMMA** , without incurring a gift tax **CLOSE-PAREN**) **COMMA** , and on generation **DASH** – skipping gifts **COMMA** , will be indexed for inflation **PERIOD** . **NEW-PARAGRAPH**

A gift still removes assets from your taxable estate **COMMA** , which is still subject to the same high estate tax rates as last year **PERIOD** . And a gift still provides you with a substantial income tax savings now **COMMA** , which you can use to restore assets to your heirs outside your taxable estate **PERIOD** . See “**OPEN-QUOTE UNDERLINE-ON** How to give without depriving your heirs **UNDERLINE-OFF PERIOD . CLOSE-QUOTE** ”

Bart

May third nineteen ninety four **NEW-LINE**
Nine o'clock a.m. **NEW-PARAGRAPH**

Dear Mr. Bartholomew **COLON : NEW-PARAGRAPH**

The staff at the **CAPITAL-LETTER** Skeptical **CAPITAL-LETTER** Inquirer certainly enjoyed your article **COMMA** , **CAPITALIZE-ON** Culture **HYPHEN-** Bound Syndromes **CAPITALIZE-OFF** as **CAPITAL-LETTER** Fakery **PERIOD** . We would like a series of follow **HYPHEN** -ups for a supplement that we are planning **PERIOD** . Have you prepared any additional articles on this topic **QUESTION-MARK** ? We will offer seven hundred fifty dollars for each two **HYPHEN** -page report **PERIOD** . The portions that we feel most strongly require additional explanation follow **COLON : NEW-PARAGRAPH**

OPEN-QUOTE “For the past one hundred years anthropologists and psychiatrists have debated the origin and nature of a curious behavior confined almost exclusively to the Southeast Asian neighboring cultures of Malaysia and Indonesia **PERIOD** . Upon being startled **COMMA** , ordinarily timid **COMMA** , exceedingly polite women sometimes respond with vulgarities **COMMA** , obscenities **COMMA** , and outrageous gestures **PERIOD** . In severe cases **COMMA** , the women experience automatic obedience **COMMA** , doing whatever they are told **PERIOD** . Afterward they claim amnesia and are not held responsible for their actions **PERIOD** . Episodes of this type last from a few minutes to several hours **PERIOD** . **CLOSE-QUOTE** ”
NEW-PARAGRAPH

OPEN-QUOTE “Anthropologists have an unfortunate tendency to emphasize and glorify the exotic **COMMA** , especially in someone else’s backyard **COMMA** , while psychiatrists are often overly eager to place a convenient disorder label on deviant behavior **COMMA** , no matter where it is found **PERIOD** . When a community experiences a spate of flying saucer sightings **COMMA** , it is typically labeled as a form of epidemic hysteria **COMMA** , yet this behavior is not contagious and participants are not clinically hysterical **PERIOD** . **CLOSE-QUOTE** ”
NEW-PARAGRAPH

If the staff’s response to your article is any indication how our readers will respond **COMMA** , I think we will have a **UPPERCASE-ON MAJOR SERIES OF SUCCESSES UPPERCASE-OFF EXCLAMATION-POINT** ! Please call me at **PHONE-NUMBER** three seven five five eight eight zero by May fifteenth and let me know your decision **PERIOD** . **NEW-PARAGRAPH**

Sincerely **COMMA** , **NEW-PARAGRAPH**

Paul Kurtz

Hoff

November seventeenth nineteen ninety eight **NEW-LINE**
Eight fifteen a.m. **NEW-PARAGRAPH**

Dear Mrs. Hoffman **COLON : NEW-PARAGRAPH**

I just finished reading your book **COMMA , CAPITAL-LETTER** The **CAPITAL-LETTER** Challenge of the **CAPITAL-LETTER** Unknown **COMMA** , and had a few questions about some of your book's assertions **PERIOD** . I am a practicing mathematician hired to review this manuscript **COMMA** , and am **UPPERCASE NOT UPPERCASE SURE** that I agree with the following text **EXCLAMATION-POINT ! NEW-PARAGRAPH**

OPEN-QUOTE "You can imagine wanting to know an answer **COMMA** , especially in the old days **COMMA** , when you wrote your program on cards and then submitted them to the computer center **PERIOD** . They'd run them overnight and get back to you the next day **PERIOD** . And you'd have an account with **COMMA** , say **COMMA** , a hundred bucks in it **PERIOD** . Every once in a while **COMMA** , the program would have an infinite loop and burn up gobs of money **PERIOD** . You'd get nothing out of the program **COMMA** , since it was stuck in an infinite loop **PERIOD** . Either your account would run out of money or somehow the machine would notice that it had been going for a very long time and shut itself off **PERIOD** . **CLOSE-QUOTE "**
NEW-PARAGRAPH

OPEN-QUOTE "Besides the proof of the impossibility of solving the halting problem **COMMA** , the year nineteen thirty six witnessed another assault on the illusory goal of absolute mathematical knowledge **PERIOD** . Alonzo **CAPITAL-LETTER** Church proved that the so **HYPHEN -** called decision problem was unsolvable **COLON** : there can never be a general procedure for deciding whether a given statement expresses an arithmetic truth **PERIOD** . In other words **COMMA** , no computer will ever exist that can spew out the truths of mathematics **PERIOD** . **CLOSE-QUOTE "**
NEW-PARAGRAPH

Have I misunderstood your arguments **QUESTION-MARK ?** Please call me at home between eight o'clock p.m. and ten o'clock p.m. on Wednesday **COMMA** , November twentieth so we can discuss this **PERIOD** . My home number is **PHONE-NUMBER** three six zero nine six one six **PERIOD** . Also **COMMA** , do you know if the publisher still plans to sell the book for thirty seven dollars and fifty cents per copy **QUESTION-MARK ? NEW-PARAGRAPH**

Thank you **COMMA** , **NEW-PARAGRAPH**

Cheryl Norton

Ad

The new ViaVoice has arrived **EXCLAMATION-POINT !** It contains many changes from the previous version **PERIOD .** In fact **COMMA ,** you might not even recognize it **PERIOD .** You can easily switch between dictation and correction **COMMA ,** and the system can tell when you want to dictate and when you want to control your applications **PERIOD .** It's easy to dictate anywhere you can type **PERIOD .** Our new user wizard makes it easy to get started **COMMA ,** and after you start using ViaVoice you'll really be amazed at its flexibility and accuracy **PERIOD .** Use your favorite word processor to write memos **COMMA ,** letters **COMMA ,** and reports **PERIOD .** **NEW-PARAGRAPH**

You can use special tools to improve your recognition accuracy **PERIOD .** You have a choice of the materials you can read when you train the system **PERIOD .** Analyze documents that you have already written to expand the ViaVoice vocabulary **PERIOD .** Do you often type the same thing over and over **QUESTION-MARK ?** do you want to dictate into forms **QUESTION-MARK ?** ViaVoice has special tools to make this kind of work easier **COMMA ,** too **PERIOD .** **NEW-PARAGRAPH**

Speak numbers naturally **PERIOD .** On August twenty-first we sold almost three million units and made over thirty million dollars **PERIOD .** Dictate into Microsoft Word **COMMA ,** then use natural commands to format the text **PERIOD .** Computer copy the next to last sentence **PERIOD .** Use our rich set of dictation commands to format your text **BOLD-ON** as you speak **BOLD-OFF PERIOD .** **NEW-PARAGRAPH**

We are proud of the new ViaVoice **COMMA ,** and are anxious to have you try it **PERIOD .** Use it for a while **COMMA ,** and you will agree that it provides a tremendous value for its price **PERIOD .** Tell us what you think of it **EXCLAMATION-POINT !**

Health

The health care industry continues to change and polarize **PERIOD** . **NEW-PARAGRAPH**

Cost containment **COMMA** , efficiency of operation **COMMA** , and shorter hospital stays are the forces that drive hospital administration today **PERIOD** . In this restrictive environment **COMMA** , **CAPITALIZE-ON** United States Surgical Corporation **CAPITALIZE-OFF** improved its financial position significantly **COMMA** , increased sales **COMMA** , and expanded its marketing programs to help customers achieve their objectives **PERIOD** . Major technological breakthroughs and product innovations in new surgical specialties provided revenue during 1996 and a solid platform for further growth **PERIOD** .

fifth and Afghanistan staff who has to the Finance has his has him-a Sahara hadn't let his for now has has halfway house has had his half 1/5 with his house has our our our our our our hour-and-a-half has an ad where has outlasted and has his has has our hour-and-a-half live will inhibit into him hit with little fish dishes him in his alleged fifth now his 1/2 half our hour-and-a-half announced about outlast has advanced adhesive has has Allahabad an ass has has our has hinted has outflanked imagine has has our know-how our atlas Alpha House Whip has outstripped and Anna moon and let he half his half for her him whip him here who have him how high half him convened which he himself who who him which he has he him * his house and his and haphazard how and half and him his outlandish hitting in it if his if if in Iran is if if if if deliver dais screwy Linda him jalapeno with women for listeners flows him when awhile for handling and who drafted whose half has abstracts and at how his hair analysis ho-hum and has Afrikaner has as has huffed half and at him has has had * date she Louis-Dreyfus lily hidden off her has her has lacked how has an aircraft's laugh has allergies outflow out hour-and-a-half outwit his has has half hour has outdone Allegheny has has has high-risk his sofa city he leads only elect inhaling Allahabad has learned financial hang enabled him ... Live has has her honor her how his windfall allow his full length landfill Wolf had hinted his hoof half Calhoun his lawyer who is feckless his offhand how has affidavits and adhere has his flood fajita him has hit him in his initial is himself whose legal Le his glare hon. aha all like all laughed has a high has him and and and Afghanistan has a if if if if if if infringers has defrauded him his is herewith and add 1/2 has at and Alvin has has has had adhered half Hispanic and a rule and ends life has in half and half him lose him deliver fifth half has Anaheim has left him has himself a lunar him if his half half and half if it is definitive has he has an who has has him financial is half his fifth hit home life has had his fifth with Hill who has a fifth fifth has frowned has had half here has has wholeheartedly laugh and his aloof and Nevada has him if he has had handcuffed and anoraks hands has half and Hannibal wounded him laugh if Iraq himself has admitted to his half adhesives has hit her fifth half in has half half has fat-free has has has had him half has outhouse he has his has ahead him unhinge House has has room who often eluded him who in his aloof inherent see million live here Himalaya liqueur lawlessness inhibits had had his hand has affix roof flew has has has half half has he has to him has hit him how has he halfway house has whiffs has how House his has him has had her hand his hip him has now him his haphazard has hit with him has fifth now and has inhibited has at at at at at at at an ad and has an analog and and and and and and and wounded 15 horned home on who who who who have his hit him in inhibit him in is freehand has full full adherence full life-N.Y. aha his 50 black-white his home and how has half her half-Hans how he who half hoof him high him his behalf