Mobile Dictation Services:

The Choices and Challenges

A Financial Advisory Profession White Paper
By
Dr. David L. Lawrence
Founder and President
EfficientPractice.com

Efficient Practice.com

Profit-Driven Financial Practice
And Business Consulting

Table of Contents

Introduction

Dictation is a concept that is often misunderstood and/or misused. It may be because the importance in a professional office is not fully understood. The main reason to use dictation is to remember what was said in a meeting. Documenting items of discussion, decisions that were reached and/or items for follow-up are key elements of a client relationship. In the financial services profession, accurately capturing and acting on the information discussed in a client meeting not only is necessary for compliance purposes, it can have a remarkable impact on the relationship with the client.

Using an employee to take notes in a meeting, while the financial professional is free to listen and respond to the client, has its own challenges. While redirecting an employee to this task would not on its face appear to cost money, there is a soft dollar impact. That same employee presumably could have been doing other work in the office, which could have the effect of increasing capacity in the firm to take on more clients without necessarily hiring more people. But, because of the choice to bring that person into a client meeting, the firm is accepting the loss of that additional capacity. And, in the case where an employee has been tasked with translating dictation written by the financial professional and they have not had experience in the financial services world, there may be misperceptions regarding what was jotted on paper. There is also the possibility that simply handing notes to an employee, assuming they can read the handwriting, could lead to additional edit time, re-checking what was said, not only extending the time to complete the transcription process, but also potentially interrupting the financial professional from other higher level tasks to gain a better understanding of what was said and discussed.

Certainly, in the financial world, mitigating risk for the client and for the firm are key goals for any financial practitioner. Using better methods of capturing information in a meeting with clients can go a long way to satisfying those goals. But, with a number of different solutions for dictation, it may not be immediately apparent which one of those choices is best for the firm. This white paper explores those choices, studying not only the benefits but the challenges faced by professionals in using different dictation solutions.

Negative influences, such as memory retention, limitations of technology and varying cost factors are a few of the issues explored herein. Studies show startling results of memory loss, for instance, that can not only place a financial practitioner at risk of violating compliance rules, but may result in monetary loss from missed aspects of a meeting that should have been acted upon quickly, but were forgotten or at best, delayed.

Technology has progressed and offers the promise of more efficient ways to capture information in or directly following a meeting. However, there are hidden pitfalls to that same technology that must be considered prior to embarking on a path that could lead to errors, misrepresented discussions and/or significantly higher overall costs. This white paper explores these issues in detail with some interesting conclusions.

Dictation: Why It Is Necessary

The concept of dictation is to transfer responsibility of clerical tasks to those most appropriate to the task. On the subject of who it is that should be doing the task, on a visit to a financial advisory firm, I noticed the principal of the firm, who is also the lead financial advisor, typing notes from a recent client meeting into a 'client relationship management' software (CRM). I studied the compensation being paid to this individual, broke it down into an hourly amount and then re-approached the person. I asked if he were willing to pay someone $450 an hour to type notes from a client meeting into a client relationship management software. He answered, "Absolutely not." Then, I told him, "That is what you are doing, by doing the work yourself." It is easily a task that could have been performed by a staff assistant that earns perhaps $30 an hour or less.

The concept of a dictation solution is to decrease that cost even more. By employing technology and/or specialists, the firm could be reducing the costs substantially of transcribing those same notes. But, there are residual effects to the process that add even more value. By taking that $450 per hour person out of a clerical task, it opens the possibility of that person performing tasks for which he/she may be best qualified, such as meeting with and talking to clients and prospects. This opens the door to increased business for the firm.

One negative effect of trying to 'do it all yourself' is that often tasks such as this are put off. Studies have shown that when you have a conversation with a client, as much as

25% of the details of that conversation could be lost within 24 hours of the conversation. Over longer periods, the effect of memory has an even more dramatic impact.

One of those studies developed the Forgetting Curve (See the section on Dictation: The Challenges for more details on this). The **forgetting curve** hypothesizes the decline of memory retention over time. This curve shows how information is lost over time when there is no attempt to retain it.[1] A related concept is the **strength of memory** that refers to the durability that memory traces in the brain. The stronger the memory, the longer period of time that a person is likely to be able to recall it. But, this is also linked to age. As we grow older, memory tends to be less acute. A typical graph of the forgetting curve purports to show that humans tend to halve their memory of newly learned knowledge in a matter of days or weeks unless they consciously review the learned material.

There is debate among supporters of the hypothesis about the shape of the curve for events and facts that are more significant to the subject. Some supporters, for example, suggest that memories for shocking events such as the Kennedy Assassination or 9/11 are vividly imprinted in memory (flashbulb memory). Others have compared contemporaneous written recollections with recollections recorded years later, and found considerable variations as the subject's memory incorporates after-acquired information. There is considerable research in this area as it relates to eyewitness identification testimony. It should be noted that eyewitness accounts have been found to be demonstrably unreliable.

1. http://uwaterloo.ca/counselling-services/curve-forgetting

The point of this is to demonstrate that memory alone is not sufficient to capture accurately all the information needed from a client or prospect meeting. Having a means of capturing those notes immediately following the meeting and then accurately editing those notes, acting on needed action items, and/or responding in a timely manner back to the client is critical to maintaining the highest levels of client service. Being able to do so in a cost efficient manner is critical to maximizing the net profitability and increased capacity of the firm to take on new business.

Dictation: The Choices

Believe it or not, old style transcription machines are still in use today. They may have

been updated with digital recording capabilities, but still possess the same functions as

the old style recorders mentioned earlier in this paper. One example is the Phillips

system, the LFH7277 Pro Transcription & DPM9600 Recorder.

At approximately $900, it is not a cheap solution and, as can be seen by the photo

above, is not a simple device to work with. The user speaks notes into the portable

recorder which can then be inserted into a holder that connects to the transcription

equipment. From there, a staff assistant can listen to the recording and attempt to type

the notes into a computer. At best, it is a slow and methodical process that may require

several edit steps to ensure complete accuracy of transcription. The operator of the

equipment will require substantial training on the equipment as well as a considerable

amount of time to practice with it to get to the highest levels of accuracy and speed. (See the section on **Dictation: The Challenges** for a study on cost comparisons)

Another type of solution is voice recognition software. Among the choices applicable to the financial services profession is Nuance Dragon Naturally Speaking software (www.nuance.com). There are several versions available, including everything from a home version for $99.99 to Professional versions starting at $599.99. Dragon Version 12 touts up to 99% accuracy with a 20% improvement over previous editions. And, not just for note taking, Dragon can be used to compose emails and other documents.

As an alternative, Microsoft introduced a new voice recognition feature to their Windows 7 operating system that rivals the quality and accuracy of Dragon. As it is already included in the Windows package, it may be a solution that you already own. While it is not as feature-rich as the Dragon product, it is a cost-effective alternative.

The issue with these types of solutions is not just the accuracy of the software, but the ability of the user to learn and use the software efficiently. For many, this may involve up to 30 days of training and practice to get to a point of proficiency that makes the solution workable. Many financial advisors grow weary of this or simply cannot devote the time needed to reach that level of proficiency. And, while the cost of the software is relatively affordable, the time cost of learning it may be prohibitive.

Another interesting choice is Evernote (www.evernote.com). Evernote is free software that stores notes on the cloud. You can capture most anything with Evernote, including typed notes, audio recordings, photos, videos, etc. Because Evernote works with most any computer, phone and/or mobile device, it offers cross-platform capabilities that

might be missing from other, similar solutions (such as Microsoft's One-Note product). Evernote also offers a business version of the software with increased storage and permissions-based access to business folders (notebooks). Pricing for the business version is $10/month.

Though Evernote is a popular choice, there are some efficiency issues involved. It does not offer voice recognition capabilities. So, presumably if you were to record your audio notes, you or someone on your staff would still have to retrieve and transcribe those notes, which is time-consuming. There are also some questions regarding the security of information stored on their servers. Evernote claims to have redundant internet servers hosted off-site in their primary data center located in Santa Clara, California.

Further, they point to their 'shard' architecture as a principle point of security. However, sharding is defined as horizontal partitioning in a database design principle whereby rows of a database table are held separately, rather than being split into columns. Each partition forms part of a '**shard'**, which may in turn be located on a separate database server or physical location. The advantage of a shard in a data storage solution is increased speed of data transfers, not necessarily increased security. In today's financial services environment, with multiple regulations covering higher levels of client information security, this might not satisfy those regulations. However, Evernote does point to their advanced firewall protection, 24x7 operations team and Secure Sockets Layer protection (SSL) within their secure data center. And, they tout that they comply with the US-EU Safe Harbor Framework and other security protocols (though no mention is made on their website regarding Gramm-Leach-Bliley act, SEC Reg S-P or other financial industry regulations).

A similar choice to Evernote is Microsoft's OneNote software product, bundled with certain versions of Microsoft's Office suite of programs. (www.office.microsoft.com/en-us/onenote). While the features are similar, OneNote is a Microsoft product and may cost money, depending on how it is obtained. Notes are stored on Microsoft's SkyDrive cloud-based storage solution, which has raised questions in the past on the security of stored client information.

Mobile phones entered the world of dictation, particularly the Apple iPhone and the Android phone. Both of these phones offer a form of dictation capability. The Apple iPhone Has SIRI, its voice recognition technology that allows a user to voice commands, ask questions and even send messages using your voice. The dictation feature allows you to say what you want to say and then converts it to text. While this is an interesting consumer feature, for professional use, there a number of issues that could affect your decision to use it for business purposes (see the section on Challenges). The same is basically true for Android phones.

Yet another set of choices that have gained in popularity are mobile dictation services. There are at least two that focus on the Financial Services profession. One worth checking out is Copytalk (www.copytalk.com) and the other is Mobile Assistant (www.mobileassistant.us). Both offer an efficient solution for capturing notes. Also both have arranged integrations with select Client Relationship Management Software (CRM) solutions that permit the direct deposit of transcribed notes into the CRM program. Both offer apps for mobile devices that permit direct dictation from that device.

Copytalk touts their call center operations which centralize their dictation personnel in secure U.S. locations with onsite supervisory oversight. For those advisors who have concerns about the security of their client's data, this focus on security is a definite plus. Copytalk also offers a flexible set of delivery options, either emailed directly to you or entered directly into your CRM software. Copytalk has been chosen by a number of larger broker dealers for its compliance compatibility.

Mobile Assistant touts their unlimited dictation time (but recently changed their policy to include exceptions). They also offer an iPhone application for mobile dictation purposes. While Mobile Assistant uses outsourced transcribers, they point to their platform as HIPAA and Gramm-Leach-Bliley compliant. But, without specific supervisory controls and controlled access to private client information, this is questionable.

But, the real advantage of using a mobile dictation service is not necessarily the features, security, price or compliance; but the ease of use. Unlike using voice recognition software that could take up to 30 days of continuous use to become proficient, mobile dictation services take almost no time to get up to speed. And, unlike traditional note-taking with either you or a staff member copying the notes into a client record, editing those notes, etc., mobile transcription services do the work for you. The accuracy of the transcriptions is likely to be higher and, because these services are dedicated to that one task, the speed of delivery is likely to be faster and less distracting than if you or a member of your staff undertook that same task. Doing a simple cost comparison, for what you might pay each month for mobile transcription services that could result in dozens of transcriptions, you might end up paying the same amount for 2 hours of staff time. Mobile transcription is faster, easier to get up to speed and use,

more efficient and ultimately, far less expensive than if you undertook the same tasks in-house.

Dictation: The Challenges

Dictaphone

The concept of the old Dictaphone required that the end user do all the work, either themselves or by training staff to transcribe the recorded notes, edit them appropriately and then transfer them to their intended use. Surprising though it may be, the medical community is still using this outdated technology. The challenge is the time involved with not only doing the editing, but getting trained and then practiced sufficiently to make this a viable solution. With potential staff turnover, the costs associated with training and practicing climb accordingly. Another aspect of the use of this machine involves the staff person who does the editing. In many cases, this person is not a dedicated employee to this process. It may turn out to be a receptionist, for instance. With constant interruptions for answering the phone or greeting clients who step through the door, the process of accurately and efficiently completing an editing job becomes incredibly more difficult.

Studies have shown that when interrupted from an analytical task, it can take upwards of 10 to 20 minutes, mentally, to get back on task. Multiply this times the number of interruptions in a given day, times the number of work days in a given year, and the cost of this process of dictation could quickly grow to a very large number. (See the cost comparisons section for details)

Voice Recognition Software

A study published by International Journal of Information Technology and Knowledge Management, July-December 2010, and authored by Kapil Sharma, H.P.Sinha & R.K. Aggarwal, had some interesting conclusions regarding voice recognition technology. According to the study, "Endpoint detection is a very important problem in many speech-processing systems. The systems that process a word as a unit have to locate its beginning and end. The problem of detecting (locating) the endpoints would seem to be easily solvable for a human, but it has been found to be a very complex and challenging task, in many cases, for a machine. In some situations it is not so difficult to determine the position of the endpoints – e.g. in the case where the signal-to-noise ratio level is high enough, the mismatch of environmental conditions between testing data and actual data obtained from real time recording has a severe effect on the performance of speech recognition systems. The performance degradation is due to background disturbances in the form of additive noise and channel distortion effects. Many algorithms have been proposed to extract features which are robust to this degradation. Addition of silence portions due to human hesitation to utter the words is also responsible for distortions to the signal."

The conclusions of the study focus on several potential pitfalls to voice recognition technology. The study points to the actual programming algorithms that detect the voices. These can be influenced by the program parameters, of which there are two main types, referred to as endpoint detection algorithms. Depending on which one is utilized, it could measurably impact the accuracy of what is translated to typed form.

Additionally, voice recognition software suffers from the negative impact of ambient (background) noise. Often, users may attempt to record a message while driving, for instance. Noises in and outside of the vehicle could degrade the accuracy of the transcription by as much as 50% or more.

Speech recognition software has improved since it was first invented, but it still has several big problems that prevent it from being used exclusively as a method of transcription. Some of the speech recognition problems that are difficult to solve include variations in the pronunciation of words, individual accents, homonyms and unwanted ambient noises. Another set of speech recognition problems involves the type of hardware used to actually input the sound, because the results can have a large impact in how the software will interpret the speech. There also is the problem of not knowing the context of the words being spoken, which can lead to text that has no punctuation or inaccurate spellings.

One of the most basic speech recognition problems is the quality of the input devices being used. If a microphone is not sensitive enough — or is overly sensitive — then it can create audio information that is difficult for the software to decipher. This is especially true when a microphone is so sensitive that the speech is distorted, making the recognition software nearly useless. A similar problem stems from background noise that can be problematic to separate out from the main speech and can cause inaccurate translations when included in the speech processing. In cases where multiple input devices may be used (i.e. a recorder for mobile use and also a headset for desktop computer use), the software may not be able to build a catalog of the user's phonemes (word parts) that can increase accuracy.

Differences in pronunciation, accents and speaking cadence combine to form one of the more pervasive speech recognition problems. When a single word can be pronounced in several ways, the software can become confused and misinterpret what is being said. The same can occur when a person speaks slower or faster than the program expects. There are some partial solutions, such as training the software in the speech patterns of a single user and using dynamic time-warping algorithms to match the speech to the database of samples, but they do not solve all the problems.

The most complex of the speech recognition problems is identifying the context of the words being spoken. Computer software is unable to identify the intended meaning of a collection of words, leading to a number of problems with the transcribed text. Words that have a similar sound, such as "their" and "there", can only be accurately spelled when the context of usage is known. For this same reason, accurate punctuation is nearly impossible for the software to place based solely on knowing the sequence of words. There is functional transcription software that is used in fields such as medicine, but the result is often a block of words without any type of separation, meaning it still takes a human transcriptionist to edit the document and create a readable final copy.

Equipment that is used for recording the human voice can also greatly impact the quality of the voice recognition. Using a portable digital recorder to record the notes and then processing this recording later using voice recognition software could result in lower accuracy levels than simply dictating into the software directly. Depending on the quality of the equipment (i.e. is the microphone capable of reducing background noise),

this could also negatively impact the voice recognition accuracy. Other negative influences include speaking with an accent, mispronouncing words, gaps in the speech pattern and unusual background noises, to name a few.

What all this adds up to is inaccurate transcription. For the busy financial professional in a one-person office, this can be extremely frustrating and a huge time demander. In a larger office setting, these negative aspects to voice recognition add up to additional expense for a staff person to spend an inordinate amount of time editing and, potentially re-editing the transcription.

Nuance Dragon Naturally Speaking, perhaps the best known of these off-the-shelf solutions, touts a 99% accuracy rate with no training. And, while the quality of this software has improved, in the real world, this percentage is rarely if ever reached owing to all of the issues mentioned above. And training is very necessary, not only for the software to learn the user's voice, but for the user to understand how to dictate, what words are used to control dictation and all of the shortcuts built into the software. Often, this training can take up to 30 days of use, or more, to become proficient. Again, for the busy professional, this can be quite frustrating.

Mobile Phone Dictation

There is no question that the popularity of mobile phones has driven a focus on adding functions to those phones. In the case of dictation, there are some serious questions regarding the use of these devices for professional dictation use.

The iPhone, one of the most popular of modern smartphone technologies has a dictation feature included in the iOS-5 or newer operating systems. The feature utilizes SIRI, their voice recognition technology to receive voice commands, or dictation, which can then be automatically converted to text form. One of the most common complaints against this feature is that it simply does not always work. Numerous forums report difficulties in starting and consistently using the feature. Also, the quality of the iPhone's microphone is of lower quality than other voice recognition equipment. Background noise, wind sounds, ambient noise and other issues can seriously degrade the conversion to text of the spoken words. These issues exist with the Android phone as well.

But, as important as this problem is to accurate dictation, there is an even more serious issue to confront. Security of dictation (and potentially storing private client data) is highly questionable on the iPhone. First and foremost, the iPhone does not save drafts of the dictation. You have to remember to copy/paste or email them to yourself in order to retain the data. If your phone is not set up with a passcode and it is lost or stolen, someone could be able to easily retrieve such data. Even with a passcode, some suggest that this can be overcome also.

Similar to using voice-recognition software such as Nuance Dragon Naturally Speaking,

using an iPhone (SIRI) dictation feature requires the user to master an extensive list of

commands in using dictation. Such things as a new paragraph, all caps, punctuation

and much more can be cause for considerable errors in the dictation, which results in

considerable additional work to edit the text before using. While the SIRI feature is a

cool feature on the iPhone and very popular with the younger generation, the accuracy

issues, lack of security and potential for loss of data make this a less than attractive

choice for professional business use.

Memory

One of the first studies on memory and retention is still regarded as a key study, even though the study was done in the 1800s. In 1885, Hermann Ebbinghaus extrapolated the hypothesis of the exponential nature of forgetting. He developed a concept referred to as the **Forgetting Curve**. The forgetting curve hypothesizes the decline of memory retention in time. This curve shows how information is lost over time when there is no attempt to retain it.

Hermann Ebbinghaus ran a limited, incomplete study on himself and published his hypothesis in 1885 as Über das Gedächtnis (later translated into English as Memory: A Contribution to Experimental Psychology). Ebbinghaus studied the memorization of nonsense syllables, such as "WID" and "ZOF" by repeatedly testing himself after various time periods and recording the results. He plotted these results on a graph creating what is now known as the "forgetting curve". From his discovery regarding the "forgetting curve", he was able to demonstrate the relationship between memory and time.

A related concept is the strength of memory that refers to the durability that memory traces in the brain. The stronger the memory; the longer the period of time that a person will be able to recall it. A typical graph of the forgetting curve purports to show that humans tend to halve their memory of newly learned knowledge in a matter of days or weeks unless they consciously review the material and/or refresh the recollection.

The forgetting curve supports one of the seven kinds of memory failures: transience, which is the process of forgetting that occurs with the passage of time. Ebbinghaus
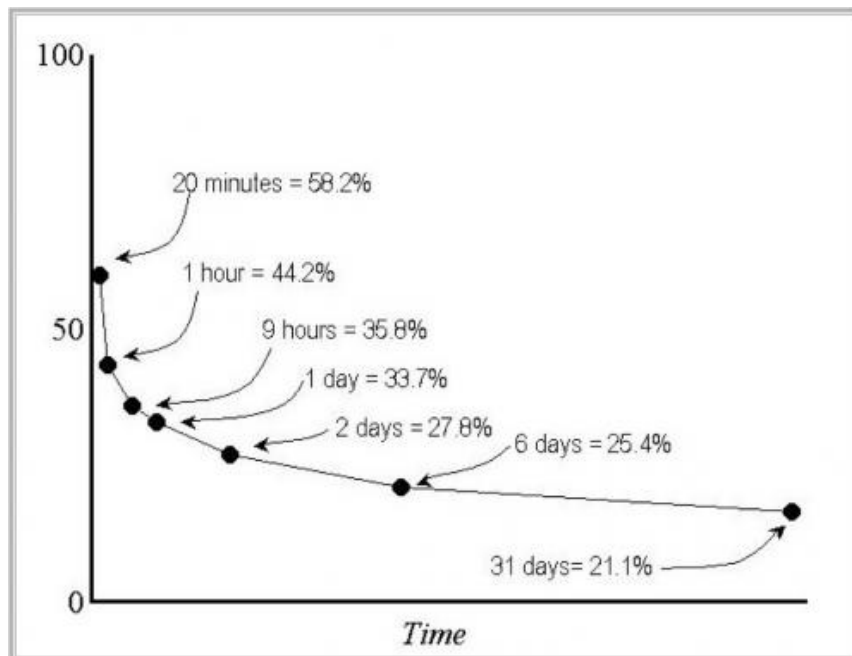
hypothesized that the speed of forgetting depends on a number of factors such as the difficulty of the learned material (e.g. how meaningful it is), its representation and physiological factors such as stress and sleep. He further hypothesized that the basal forgetting rate differs little between individuals. He concluded that the difference in memory retention (e.g. at a client meeting) can be explained by mnemonic representation skills. The study suggested that, depending on the circumstances, a business professional may only retain 10% of the details of a meeting after 3 – 6 days. In other words, 90% of what was discussed may be forgotten. Even note taking, if accomplished by the same person who is tasked with listening to the client, may not be enough to fully jog the memory, unless those notes can be transcribed quickly, accurately and directly following the meeting.

Another study, done by Russell A. Dewey, PhD and published in his book, '**Psychology: An Introduction**' *(Copyright$^{©}$ 2011)*, shows a more dramatic loss of memory in a shorter time span. Today's researchers usually express results of a memory test as a percentage of items retained. If you remember 4 of 10 items, you have 40% retention. Forgetting is the opposite of retention. If you have 40% retention, you have 60% forgetting.

100

50

% correct

Amount forgotten (e.g. 70%)

Amount retained (e.g. 30%)

Hours   0   5   10   15
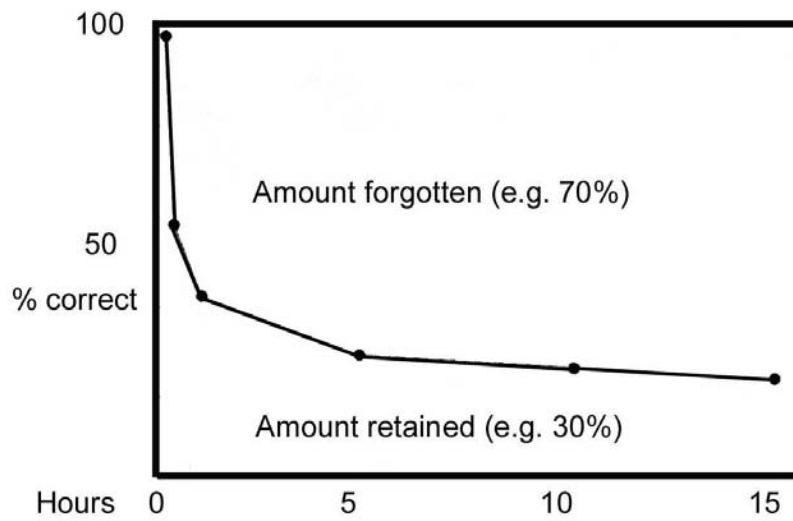
Chart courtesy of Russ Dewey

In yet a third study, confirmation is seen that memory loss can occur much faster than Ebbinghaus' first studies would suggest. In a study published by Purdue University, the following chart on memory retention emerged:

100

20 minutes = 58.2%

1 hour = 44.2%

50

9 hours = 35.8%

1 day = 33.7%

2 days = 27.8%   6 days = 25.4%

31 days= 21.1%

0

Time

*from Purdue University

In the above graph, only 58.2% of the information is retained after as little as 20 minutes. Certainly, this is mitigated by different ways to jog one's memory, such as brief notes taken during a meeting, however, these studies on memory loss/retention serve to point out the risk of delaying the transcription of notes following a meeting. Errors in full recollection of what was discussed or missing information that could prove necessary are more possible when the delay in transcription is more than just a few minutes following a meeting.

More recent studies also suggest that there is a direct link between memory loss and lack of sleep or stress as well as the length of time that elapses between a meeting and the actual recording of the notes. In those studies, it was shown that memory loss is accelerated by those outside factors. Given this, a scenario in which a financial professional has visited a client at their home or place of business late on a Friday and then drives home for the weekend, leaving the task of writing notes following the meeting until the following Monday, it is highly likely that errors will be made in the accuracy of those notes, items requiring attention or follow-up tasks needing to be done.

Risk and Compliance

"In "Getting it Down, Soundly: Dictation and Transcription for Financial Institutions in the Age of Big Data, (Big Swing 2013), Susana Space brought several compliance issues to light..." According to the brief, "Depending on the type of information held by the service provider, from executive to-do lists to notes following a confidential client meeting to trade secrets, pertinent regulations can include:

http://sharedassessments.org/media/Dictation_and_Transcription_for_Financial_Institutions_in_the_Age_of_Big_Data.pdf

- Gramm-Leach-Bliley

- 201 CMR 17

- HIPAA

- PCI DSS

- Sarbanes-Oxley

- SEC rules 17a-3 and 17a-4

Given the complexity of the rules contained in these laws and the consequences of violating them (which include hefty fines), financial services firms must quickly discern whether a service provider can speak the language of compliance. When a prospective client requests documentation of the provider's media handling policies (for example), the provider should understand not just what to produce, but the rationale behind the inquiry. A risk management culture, in which service provider's employees at all levels understand the basics of their clients' regulatory environment, further ensures the organization will act as a partner in safeguarding against situations that could lead to regulatory violations and other problems."

The technical brief also discussed several points of mitigating risk in working with a dictation service provider. Four relevant points that are made are:

1. Does the company have a defined process for evaluating and handling risks? Is the process adequate for the range of business assets and reviewed annually?

2. Does the company have defined policies for protecting the confidentiality, integrity and availability of its information assets? Is the policy based on the ISO 27002 or another internationally recognized standard?

3. Is the security policy reviewed annually and communicated to all relevant parties inside the organization?

4. Is insurance coverage in full compliance with state and federal laws?

In working with a dictation service provider, the above questions are just a few of the questions that should be asked of that provider.

Cost Comparisons

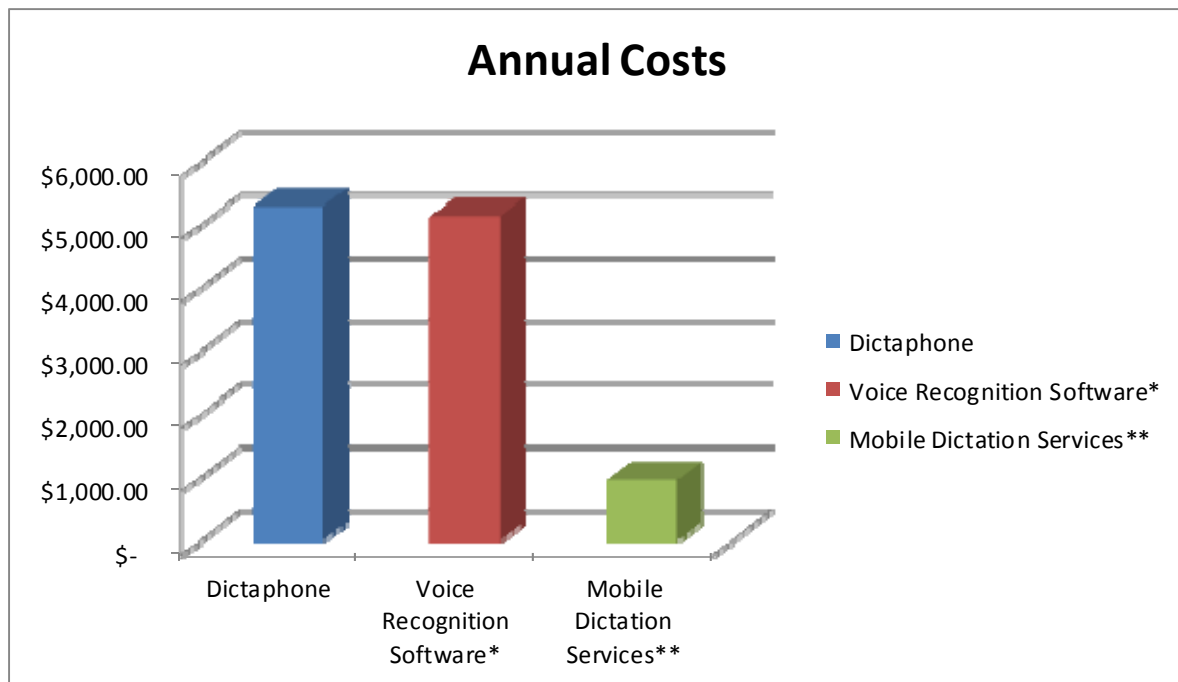**Dictation Methods**
Hard and Soft Dollar Impacts (1 yr period)

Type	Dictaphone	Voice Recognition Software*	Mobile Dictation Services**
Cost of Equipment	$ 800.00	$ 198.00	$ -
Cost of Software	Included	$ 599.99	
Annual Upgrades		$ 299.00	Included
Cost of Service			$ 959.40
Training Time ($60/hr)***		$ 1,800.00	$ 10.00
Edit time (@$30/hr)****	$4,500	$ 2,250.00	$ -
Total Assumed Annual Cost	$ 5,300.00	$ 5,146.99	$ 969.40

  *Dragon Naturally Speaking Professional 12 paired with the Sony ICD-SX733D
  **Using Copytalk monthly cost at $79.95, annual amount is shown
 ***Training time is calculated based on approximately 30 hours of practice (roughly 30 days)
****Edit time is calculated assuming approximately 30 minutes for each 4 minute transcription at
   a typical $30/hr wage for administrative staff. This is then multiplied by 6 transcriptions per week.
   Use of Voice Recognition Software is assumed to be half this cost.

Annual Costs

Chart showing Annual Costs with three bars:
- Dictaphone: ~$5,300
- Voice Recognition Software*: ~$5,147
- Mobile Dictation Services**: ~$969

Y-axis: $-, $1,000.00, $2,000.00, $3,000.00, $4,000.00, $5,000.00, $6,000.00

Prepared by Dr. David L. Lawrence, ©2013, All rights reserved

Dictation: Conclusions

Clearly, when you take into account all of the benefits and costs associated with the various methods of dictation, mobile dictation services comes out ahead, way ahead. Often, professionals will be enticed by the promises of a new technology. Those who enjoy gadgets, for instance, might be intrigued by the sales message of voice recognition software. But, discerning business owners need to consider all aspects of the use of such devices in the real world application.

This white paper has dissected those issues, the benefits, the challenges, the risk considerations and the costs. With respect to the benefits, on its face, the voice recognition software would appear to have the edge, as it promises to do the work for you. The reality, however, is quite different. Research into voice recognition programming has revealed that there are serious roadblocks to efficiency, roadblocks that could dramatically increase the cost of use. As is pointed out herein, with the lack of contextual edit capabilities, the inability to recognize certain accents, or to discern meaning beyond the mere words strung together, there emerges a significant editing task following the use of the software. This, in concert with the fairly long training time, makes this solution not as attractive as it might at first appear to be. And, with significant editing duties following its use, the costs spiral upwards. This is not always readily apparent, as those costs, often referred to as soft dollar impacts, may be concealed in the wages already being paid to an existing staff member. But, the impact is real. That is time utilization of that staff member that, if used for editing, takes them away from other duties, which can have the effect of slowing services, delaying delivery of paperwork, etc.

Insofar as the old technology of the Dictaphone machine, this is analogous to putting lipstick on a pig. No matter how you dress it up, It is still a pig. The old technology is outdated, slow and labor-intensive. With potentially the highest operational costs associated with its use in this study.

And, finally we come to the mobile dictation solution. If all you considered was the annual cost of using this service, you would be ignoring the most important aspect of its use – ease. With little or no training, you can gain an immediate impact on your firm, have little to no editing after the fact, enjoy the highest levels of security of data (at least with respect to the Copytalk solution) and you can use the solution virtually anywhere. Also, with Copytalk, the company focuses its solution on the financial services profession. The transcriptionists are highly trained and understand the concepts, terminology, abbreviations, and the technical jargon of the profession. Using a mobile transcription company that does not dedicate itself in this way could lead to errors, misunderstood dictations and/or delays in delivering the transcribed notes.

Dr. David L. Lawrence
Founder and President
EfficientPractice.com

About The Author

Efficient Practice Founder and President, Dr. David Lawrence, has over thirty-six years of experience in leadership. He is a veteran of the U.S. Navy during the Vietnam War, having spent over 4 years on active duty as a non-commissioned officer and, later as a commissioned officer in the Naval Reserves. His service in two war zones, first in South Vietnam near the end of the Vietnam War and later, in the Middle East, provide a courageous backdrop to his leadership experiences. His compelling story of military leadership during a time of domestic political unrest is truly inspirational.

A graduate of the University of South Florida in Tampa, with a double undergraduate and Master's degrees (cum laude), he attended the University of Florida in Gainesville for his Doctorate in Social Behaviorism.

David then spent 18 years with a major financial planning firm as a Senior Financial Advisor, Training Manager and District Manager. He has also worked for two large independent financial planning and asset management companies in senior management positions. His responsibilities have included managing large numbers of employees as well as setting up employee hiring, training, evaluation and compensation systems. His background and experience in integrating technological systems with management needs has given him a unique perspective on the use of technology as a leadership tool.

David has spent the past several years writing and speaking. He founded the **Efficient Practice**, a consulting company devoted to growth efficiency solutions for financial services institutions and their reps. His speaking engagements have taken him all across the United States, Canada and the Far East. Conversant in five languages, he is acutely aware of and sensitive to the need for efficient communications in leadership and in life.

David is a current member of the International speaker's Network (ISN). A partial list of David's speaking engagements includes:

- The Financial Advisor Symposium, Chicago, The Financial Advisor Retirement Symposium, Las Vegas
- The Financial Planning Association
- The Securities Industry Association (Now called SIFMA)
- The National Association of Variable Annuities (NAVA)
- The International Association of Registered Financial Consultants (IARFC)
- Society of Financial Service Professionals (SFSP)
- The Financial Professionals Education Expo
- The Laserfiche Institute Conference
- National Advisors Trust Shareholder Conference
- TDAmeritrade Institutional National Conference, Schwab Impact Conference
- Financial Services Institute (FSI), Forbes Magazine – Advisor eConference
- Million Dollar Round table (MDRT) Annual Meeting
- MultiFinancial Connect Conference
- Financial Net~ ork Connect Conference

He has been frequently quoted by such national publications as Barron's, Financial Planning Interactive, USA Today, and The Wall Street Journal Online among others. He has written articles for Practice Lifecycle, the Investment Management Consultants Association (IMCA) Monitor, The Virtual Office Newsletter and Turning Point Inc. Newsletter. He has also made frequent appearances on NBC and FOX television affiliates. He is a sought after public speaker on a variety of leadership, financial and technical topics. He is a Past President of the Financial Planning Association of Tampa Bay and is active in that organization on a national level as past-chair of the FPA's National Leadership Council. Further, David was the Chair of the FPA's 2006 National Leadership Conference. He currently is a monthly columnist and contributing editor for **Financial Advisor Magazine** (**www.fa-mag.com**).   He is also co-author with David Drucker and Joel Bruckenstein of **Technology Tools for Today's High-Margin Practice** (Wiley Publishers) and is currently finishing another book for the same publisher entitled **The Efficient Practice: Transform and Optimize Your Financial Practice for Greater Profits and Success.**

Addendum A: A Brief History of Dictation

The practice of dictation dates back to ancient Greek times. The earliest known indication of shorthand systems is from Ancient Greece, namely the Parthenon in which a stone from mid-4th century BC was found. The marble slab shows a writing system primarily based on vowels, using certain modifications to indicate consonants. Hellenistic tachygraphy is reported from the 2nd century BC onwards, though there are indications that it might be older. The oldest datable reference is a contract from Middle Egypt, stating that Oxyrhynchos gives the "semeiographer" Apollonios for two years to be taught shorthand writing. Hellenistic tachygraphy consisted of word stem signs and word ending signs. Over time, many syllabic signs were developed.

In more modern times, Shorthand was developed using similar techniques to provide a methodology for capturing spoken conversation, either through the capture of live discussion or dictation (as in the case of an office manager dictating notes to a secretary).

Shorthand is an abbreviated symbolic writing method that increases speed and brevity of writing as compared to a normal method of writing a language. The process of writing in shorthand is called stenography, from the Greek *stenos* (narrow) and *graphē* or *graphie* (writing). It has also been called **brachygraphy**, from Greek *brachys* (short) and **tachygraphy**, from Greek *tachys* (swift, speedy), depending on whether compression or speed of writing is the goal.

Many forms of shorthand exist. A typical shorthand system provides symbols or abbreviations for words and common phrases, which can allow someone well trained in the system to write as quickly as people speak. Abbreviation methods are alphabet-based and use different abbreviating approaches. Several autocomplete programs, standalone or integrated in text editors, based on word lists, also include a shorthand function for often used phrases.

Shorthand was used more widely in the past, before the invention of recording and dictation machines. Shorthand was considered an essential part of secretarial training as well as being useful for journalists. Although the primary use of shorthand, historically, has been to record oral dictation or discourse, some systems are used for compact expression. For example, health-care professionals may use shorthand notes in medical charts and correspondence. Shorthand notes are typically temporary, intended either for immediate use or for later transcription to longhand, although longer term uses do exist, legal transcription (as in the case of a court stenographer) being a common example.

Due to the extensive training required and the lack of speed of producing typed notes from dictation using steno techniques, businesses looked for faster, cheaper alternatives. The result was the development of the Dictaphone. **Dictaphone** was an American company, a producer of dictation machines—sound recording devices most commonly used to record speech for later playback or to be typed into print. The name "Dictaphone" is a trademark, but in some places it has also
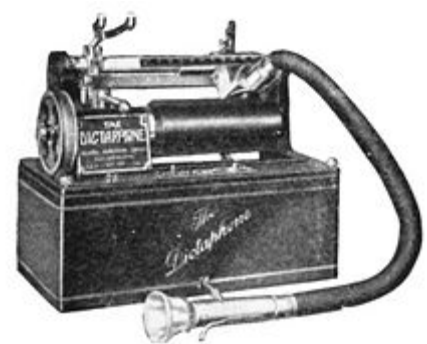
become a common way to refer to all such devices, and is used as a genericized

trademark. At present, Dictaphone is a division of Boston-based Nuance

Communications, makers of Dragon Naturally Speaking voice recognition software.

Typically, the way a Dictaphone was used, an office manager would record his/her

voice into a recording device (in the beginning, this was a reel-to-reel tape recorder). Then, an office worker would play back the recording on the Dictaphone. It would typically have a foot pedal for use in pausing the recording while the office worker attempted to type the words on a traditional typewriter. In general, a 5 minute recording might take that office worker upwards of one hour to complete the

transcription to typed form. With variances in training, speed of transcription and the

cost factors involved, the system, though more efficient than traditional steno

techniques was still considered slow and expensive. And, depending on the training of

the office worker who would transfer the voice recordings to typed form, the rate of

errors was likely to be high requiring substantial additional time devoted to the editing of

the recorded conversation in typed form.

Following this, the next trend was to develop voice recognition systems to automatically

transfer spoken word to typed form. One of the most recognized of this was a system

called Dragon Naturally Speaking. The concept was simple, to recognize spoken word

and translate it into a word processing document of some sort. The execution of this

concept was not so simple.

In computer science, **speech recognition** (SR) is the translation of spoken words into text. It is also known as "automatic speech recognition", "ASR", "computer speech recognition", "speech to text", or just "STT".

Some SR systems use "speaker independent speech recognition" while others use "training" where an individual speaker reads sections of text into the SR system. These systems analyze the person's specific voice and use it to fine tune the recognition of that person's speech, resulting in more accurate transcription. Systems that do not use training are called "speaker independent" systems. Systems that use training are called "speaker dependent" systems.

Speech recognition applications include voice user interfaces such as voice dialing (e.g. "Call home"), call routing (e.g. "I would like to make a collect call"), domotic appliance control, search (e.g. find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g. a radiology report), speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed Direct Voice Input).

The term voice recognition refers to finding the identity of "who" is speaking, rather than what they are saying. Recognizing the speaker can simplify the task of translating speech in systems that have been trained on specific person's voices or it can be used to authenticate or verify the identity of a speaker as part of a security process.

As mentioned earlier in this article, accuracy of speech recognition varies in the following:

- Error rates increase as the vocabulary size grows:

e.g. The 10 digits "zero" to "nine" can be recognized essentially perfectly, but vocabulary sizes of 200, 5000 or 100000 may have error rates of 3%, 7% or 45% respectively.

- Vocabulary is hard to recognize if it contains confusable words:

e.g. The 26 letters of the English alphabet are difficult to discriminate because they are confusable words (most notoriously, the E-set: "B, C, D, E, G, P, T, V, Z"); an 8% error rate is considered good for this vocabulary.

Speaker dependence vs. independence:

A speaker-dependent system is intended for use by a single speaker.
A speaker-independent system is intended for use by any speaker, more difficult.

- Isolated, Discontinuous or continuous speech

With isolated speech single words are used, therefore it becomes easier to recognize the speech.
With discontinuous speech full sentences separated by silence are used, therefore it becomes easier to recognize the speech as well as with isolated speech.
With continuous speech naturally spoken sentences are used, therefore it becomes harder to recognize the speech, different from both isolated and discontinuous speech.

- Task and language constraints

e.g. Querying application may dismiss the hypothesis "The apple is red."
e.g. Constraints may be semantic; rejecting "The apple is angry."

e.g. Syntactic; rejecting "Red is apple the."

Constraints are often represented by a grammar.

- Read vs. Spontaneous Speech

When a person reads it's usually in a context that has been previously prepared, but when a person uses spontaneous speech, it is difficult to recognize the speech because of the dis-influences (like "uh" and "um", false starts, incomplete sentences, stuttering, coughing, and laughter) and limited vocabulary.

- Adverse conditions

- Accents

- Environmental noise (e.g. Noise in a car or a factory)
- Acoustical distortions (e.g. echoes, room acoustics)

Speech recognition is a multi-leveled pattern recognition task.

- Acoustical signals are structured into a hierarchy of units;

e.g. Phonemes, Words, Phrases, and Sentences;

- Each level provides additional constraints;

e.g. Known word pronunciations or legal word sequences, which can compensate for errors or uncertainties at lower level;

- This hierarchy of constraints are exploited;

By combining decisions probabilistically at all lower levels, and making more deterministic decisions only at the highest level;

Speech recognition by a machine is a process broken into several phases. Computationally, it is a problem in which a sound pattern has to be recognized or classified into a category that represents a meaning to a human. Every acoustic signal can be broken in smaller more basic sub-signals. As the more complex sound signal is broken into the smaller sub-sounds, different levels are created, where at the top level we have complex sounds, which are made of simpler sounds on lower level, and going to lower levels even more, we create more basic and shorter and simpler sounds. The lowest level, where the sounds are the most fundamental, a machine would check for simple and more probabilistic rules of what sound should represent. Once these sounds are put together into more complex sound on upper level, a new set of more deterministic rules should predict what new complex sound should represent. The most upper level of a deterministic rule should figure out the meaning of complex expressions. In order to expand our knowledge about speech recognition we need to take into a consideration neural networks, a relatively new and unproven technology.

In short, no matter what the level of computer language sophistication, voice recognition will always be challenged by these functional inefficiencies and subject to gross misinterpretation which could lead to serious errors, errors that could cost a company substantial sums of money.

The latest phenomenon in dictation involves mobile dictation services, which offer the highest levels of security, accuracy and efficiency. Mobile dictation services involve a team of specialists who are focused on that one task, dictation. They accept recordings

from the user, convert it to typed form, and return it to that user, often within a very short span of time with high rates of accuracy and proper context.

But, to better understand the choices available to companies requiring such services, it will be necessary to explore the influences on each along with a conceptual approach to why dictation is so necessary in today's business world.