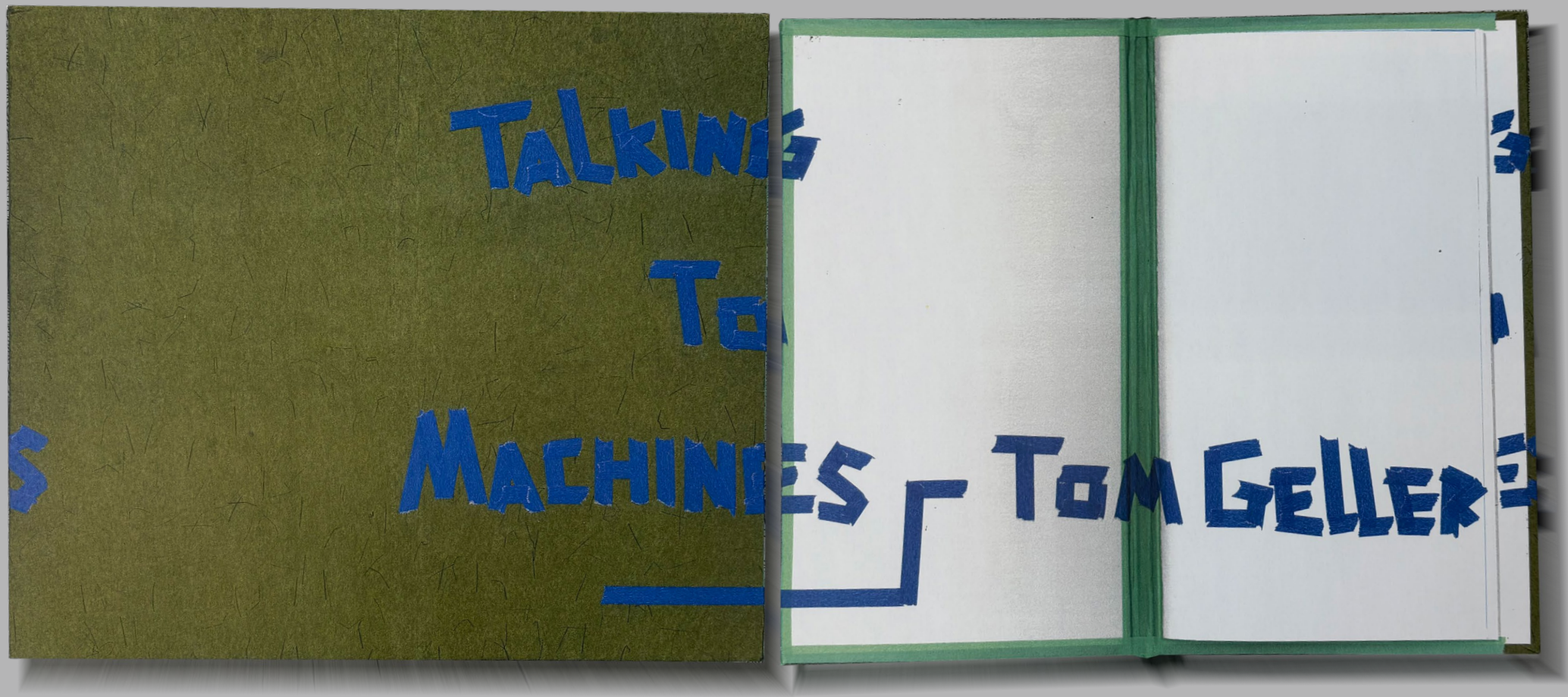
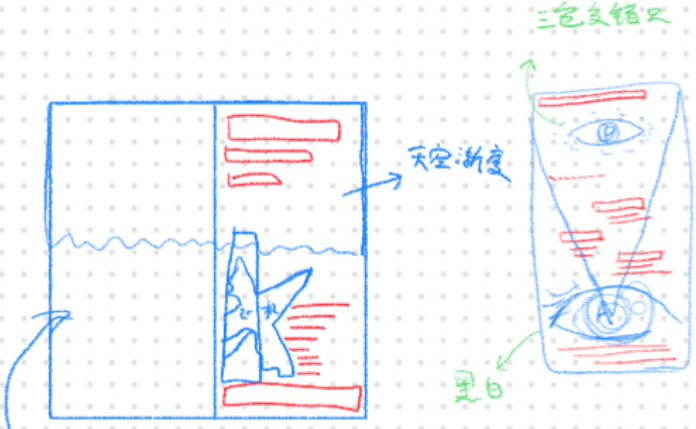
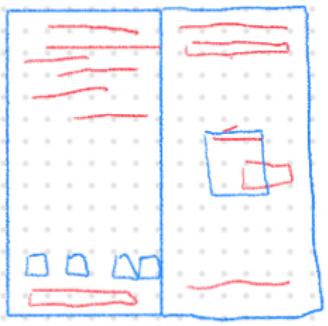
Project 01
Publishing Forms

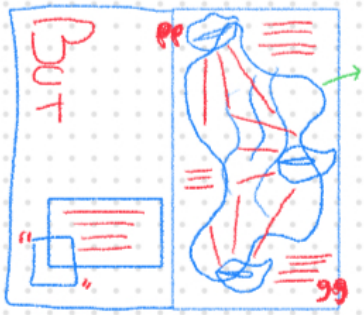
Yuxin Jiang
—35810467



Draft



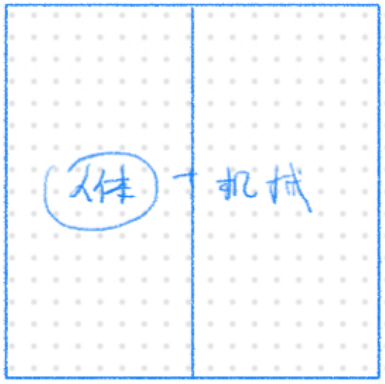




不过，从发出语音命令到执行命令之间还有很长的路要走。这一过程的第一步是将语音音频转换为语音指令。

意义。语音识别的两大主要应用--听写和命令识别--迫使研究人员追求兼顾词汇、口音和语境需求的并行方法。

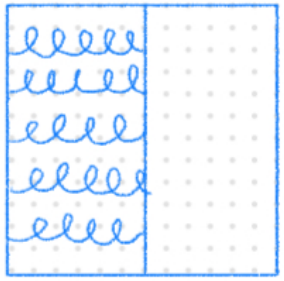
~~基于语法的语音识别是针对程序对说话人将说什么有非常好的理解的情况而优化的~~



Although initial training is not as necessary as before, dictation pro- grams still train themselves as you use them. "The program looks at three things to get to know its user," says Ma- honey. "First is acoustics. How do you actually say words? That's what the ini- tial training used to focus on. Second is the kind of words you use—your spo- ken writing style. Third is a variety of user preferences, such as how you say numbers and names, and how you like them to be formatted. So if you were to say, 'I gave you two dollars and forty- two cents,' the program knows to tran- scribe that as 'I gave you $2.42.'"

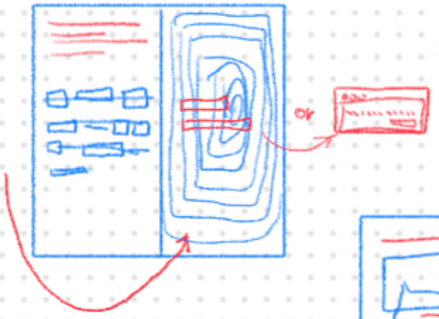
虽然初始培训不像以前那样必要，但听写程序仍然在你使用它们时进行自我训练。 Ma- honey说："该程序研究了三件事来了解其用户。""首先是声学。 你到底是怎么说话的？ 这就是初始培训过去的重点。 其次是你使用的单词类型——你的spo-ken写作风格。 第三是各种用户偏好，例如您如何说数字和名称，以及您希望如何格式化它们。 因此，如果你说，"我给了你两美元和四十二美分"，程序知道将它转录为"我给了你2.42美元。"'

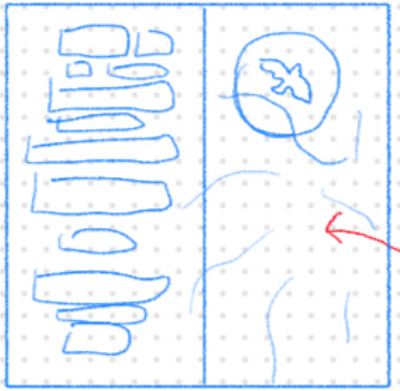
Further Reading



去年 10 月，当苹果将 Siri 添加到 iOS 操作系统中时，iPhone 用户开始通过手机与 Siri 对话。

手机进行对话。该程序可将 "今晚 6 点与丽莎共进晚餐 "等口语命令转换为电话预约、网络搜索等功能，是迄今为止最广泛传播的认知辅助工具。配备 Siri 功能的 iPhone 4S 在首个周末就售出了 400 多万部。虽然我们可能认为它只是简单的语音识别，但它的能力远远超出了简单的转录。



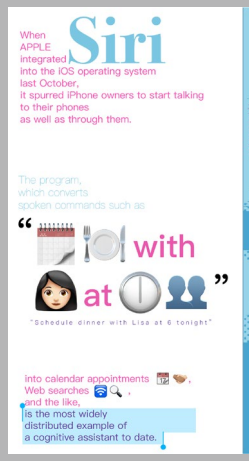


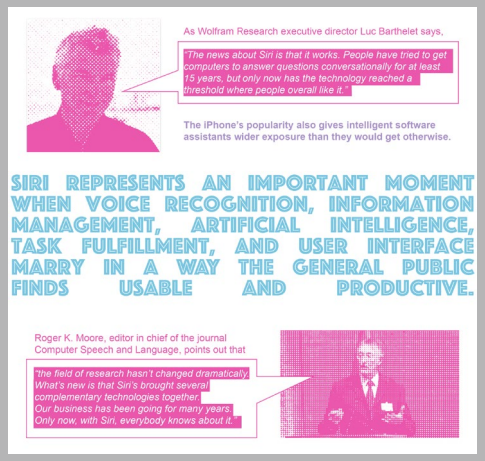
虽然我们可能认为它只是简单的语音识别，但它的能力远远超出了简单的转录。

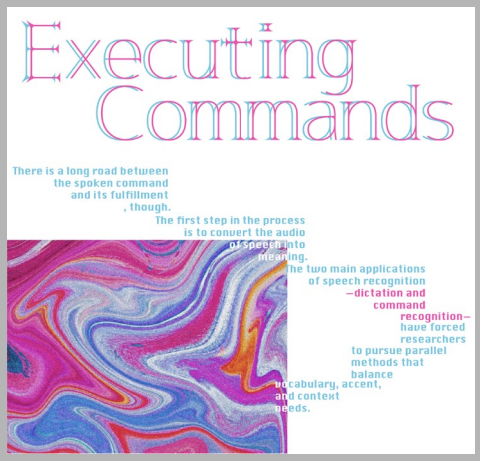
当语音识别、信息管理、人工智能、任务执行和用户界面以大众认为可用和高效的方式结合在一起时，Siri 代表了一个重要的里程碑。正如 Wolfram Research 公司执行董事吕克-巴特莱（Luc Barthelet）所说："关于 Siri 的新闻就是它能工作。人们试图让电脑以对话方式回答问题至少已有 15 年之久，但直到现在，这项技术才达到了让人们总体上喜欢它的门槛。iPhone 的流行也为智能软件助手提供了更广阔的发展空间。计算机语音与语言》杂志主编罗杰-K-摩尔指出："研究领域并没有发生重大变化。新的是 Siri 将几种互补技术结合到了一起。我们的业务已开展多年。只是现在，有了 Siri，大家都知道了。

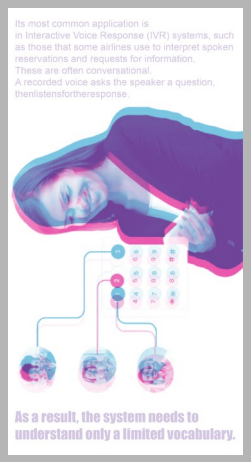
Book Design

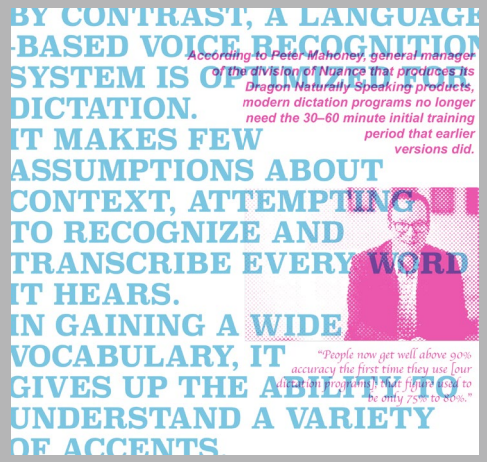
Talking to Machines explores the evolution of speech recognition technology, focusing on Apple's Siri. I wanted to present the content of the article in a visual way that would allow the reader to understand the ideas more quickly and enhance the readability of the article.

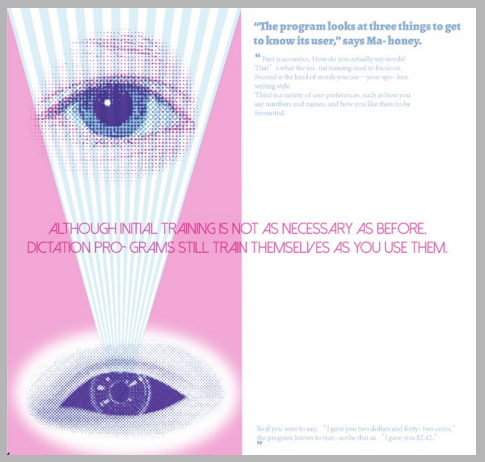


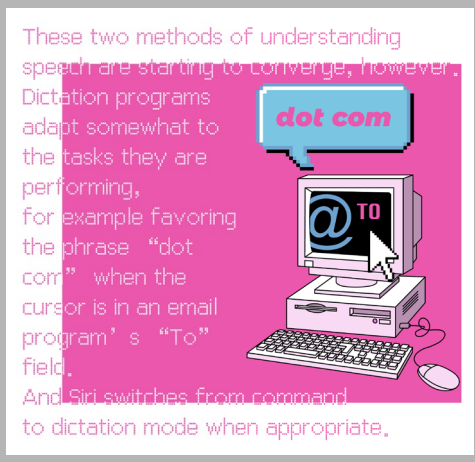












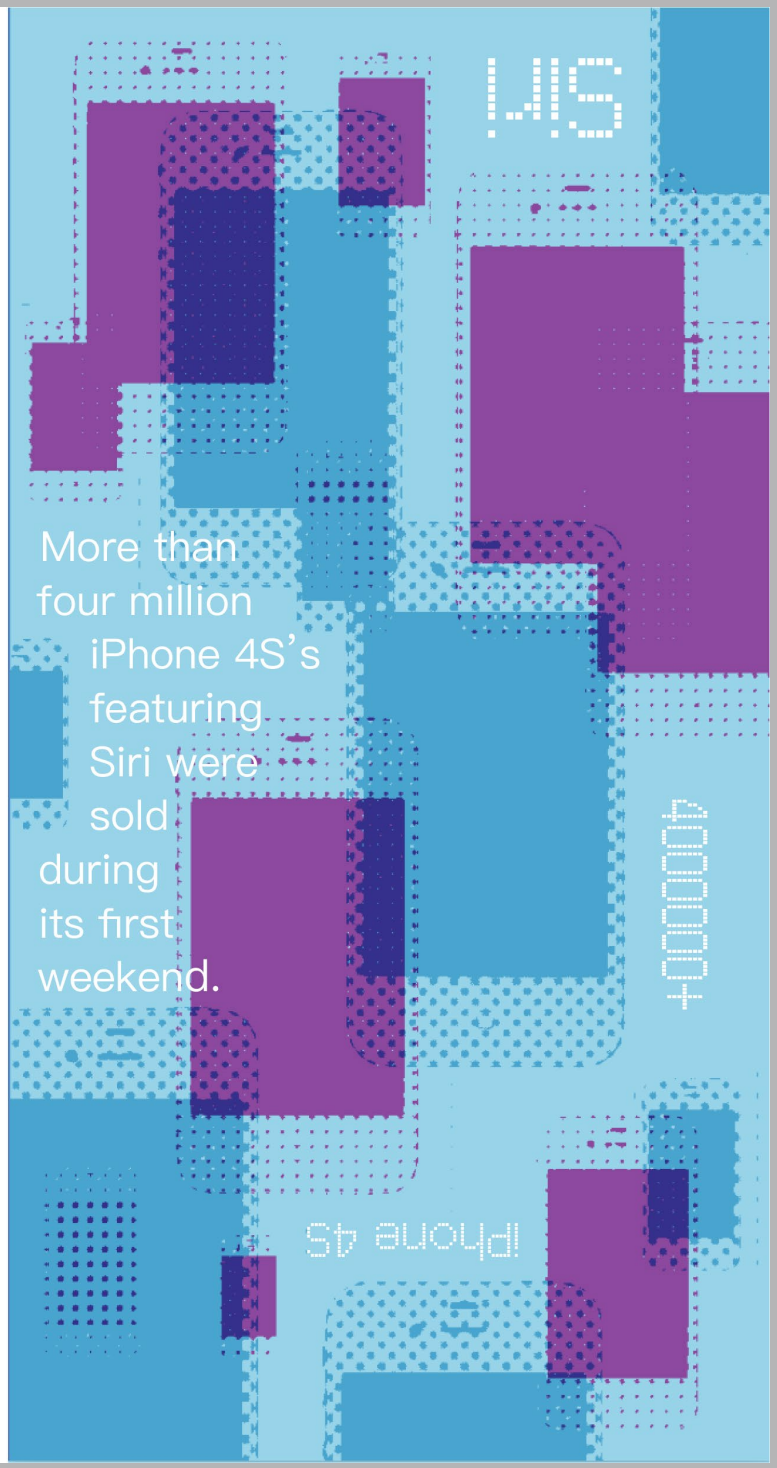
When APPLE integrated **Siri** into the iOS operating system last October, it spurred iPhone owners to start talking to their phones as well as through them.

The program, which converts spoken commands such as

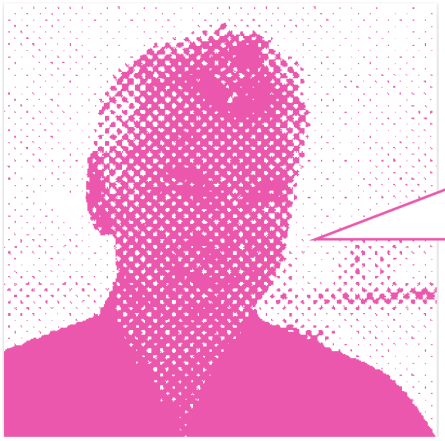
“ with at ”

“Schedule dinner with Lisa at 6 tonight”

into calendar appointments , Web searches , and the like, is the most widely distributed example of a cognitive assistant to date.



ALTHOUGH
USERS
MIGHT
SEE IT AS
SIMPLE SPEECH
RECOGNITION,
ITS ABILITIES
GO FAR
BEYOND
SIMPLE
TRANSCRIPTION.



As Wolfram Research executive director Luc Barthelet says,

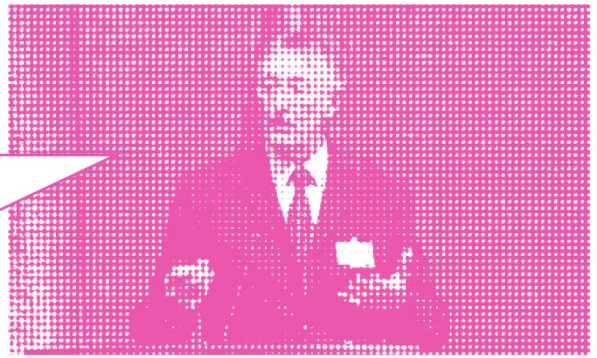
"The news about Siri is that it works. People have tried to get computers to answer questions conversationally for at least 15 years, but only now has the technology reached a threshold where people overall like it."

The iPhone's popularity also gives intelligent software assistants wider exposure than they would get otherwise.

SIRI REPRESENTS AN IMPORTANT MOMENT WHEN VOICE RECOGNITION, INFORMATION MANAGEMENT, ARTIFICIAL INTELLIGENCE, TASK FULFILLMENT, AND USER INTERFACE MARRY IN A WAY THE GENERAL PUBLIC FINDS USABLE AND PRODUCTIVE.

Roger K. Moore, editor in chief of the journal Computer Speech and Language, points out that

"the field of research hasn't changed dramatically. What's new is that Siri's brought several complementary technologies together. Our business has been going for many years. Only now, with Siri, everybody knows about it."



Executing Commands

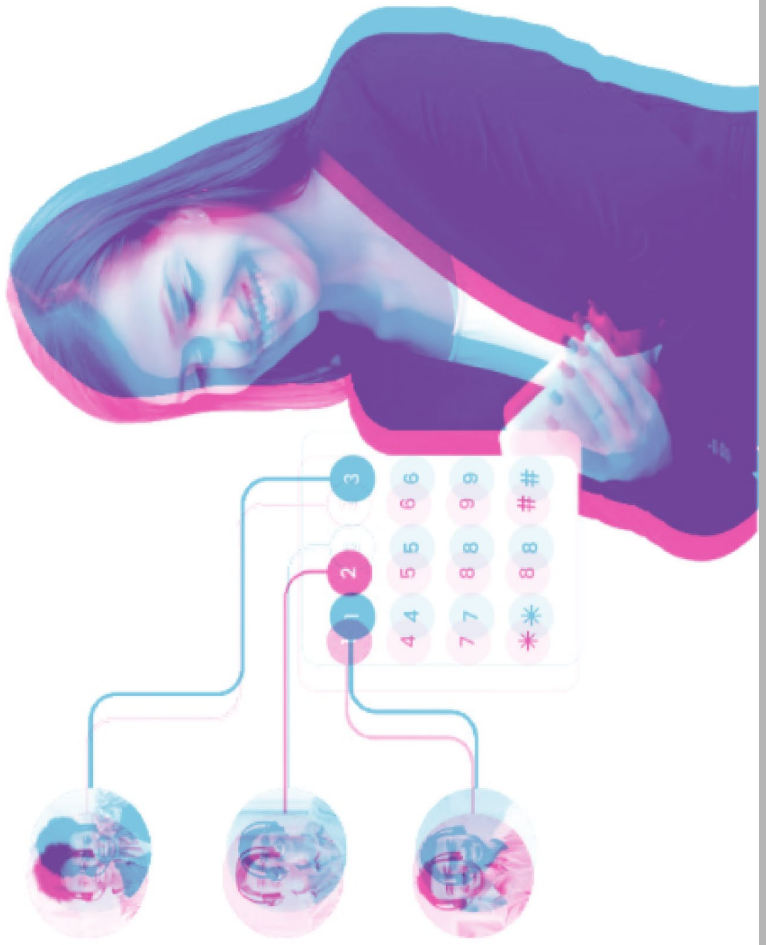
There is a long road between the spoken command and its fulfillment , though.

The first step in the process is to convert the audio of speech into meaning.

The two main applications of speech recognition —dictation and command recognition— have forced researchers to pursue parallel methods that balance vocabulary, accent, and context needs.

Grammarbased voice recognition is optimized for situations where the program has a very good idea of what the speaker will say.

Its most common application is in Interactive Voice Response (IVR) systems, such as those that some airlines use to interpret spoken reservations and requests for information. These are often conversational. A recorded voice asks the speaker a question, thenlistensfortheresponse.



As a result, the system needs to understand only a limited vocabulary.

BUT

According to Dan Faulkner, vice president of product and strategy for the Enterprise Business Unit at Nuance, responses can vary widely no matter how restricted the domain is.

When a phone prompt system tells you to 'please say yes or no,' we might be pretty confident the speaker will say 'yes' or 'no.'

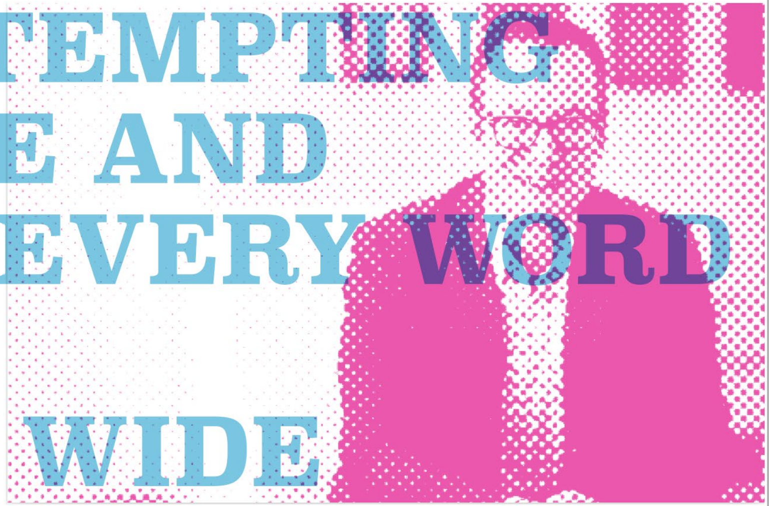
But 'yes' could be 'yeah,' 'that's correct,' or 'yup,'

and in the Southern states some people will say 'yes, ma'am' and 'no, ma' am.'

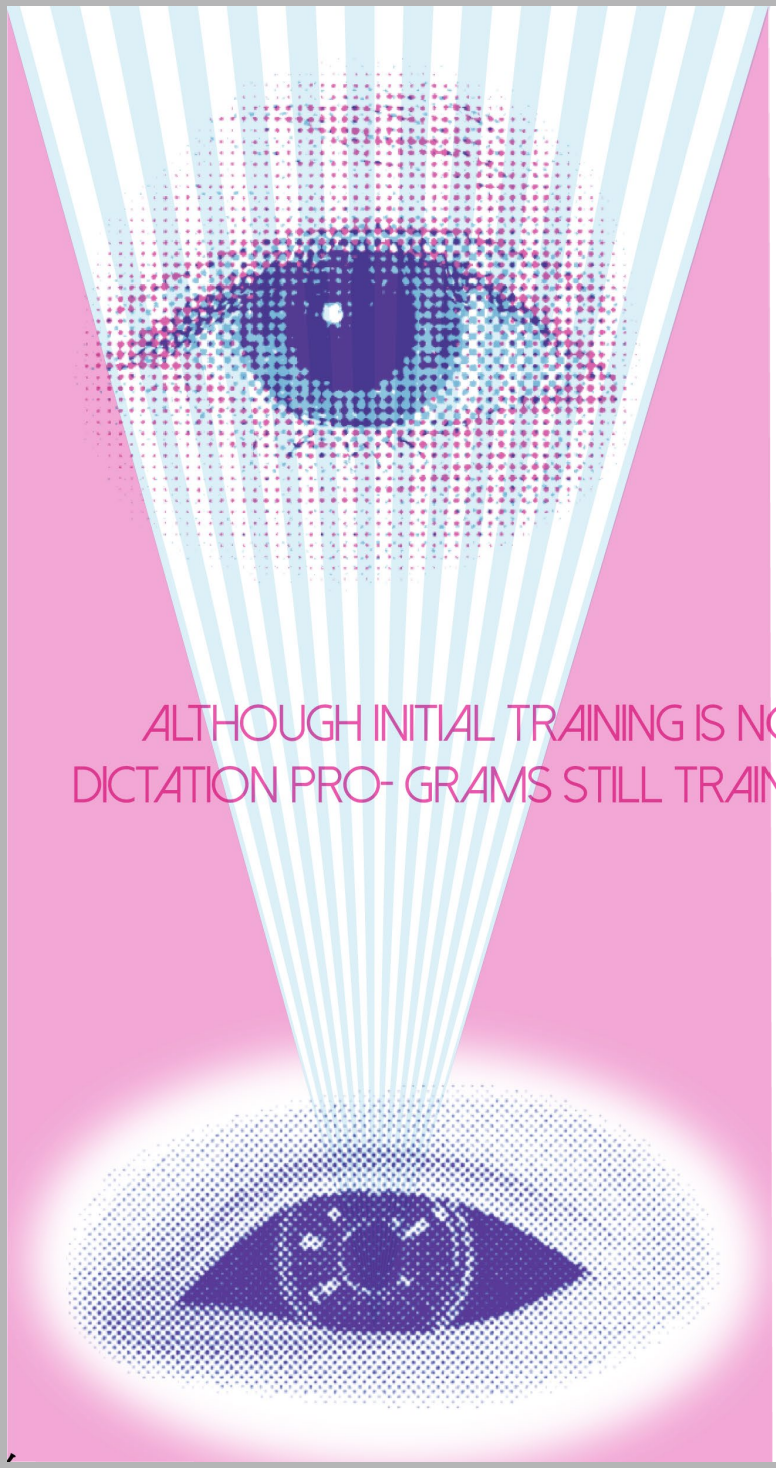
So even for something as simple as 'yes' or 'no,' you need quite an extensive list of phrases.

BY CONTRAST, A LANGUAGE-BASED VOICE RECOGNITION SYSTEM IS OPTIMIZED FOR DICTATION. IT MAKES FEW ASSUMPTIONS ABOUT CONTEXT, ATTEMPTING TO RECOGNIZE AND TRANSCRIBE EVERY WORD IT HEARS. IN GAINING A WIDE VOCABULARY, IT GIVES UP THE ABILITY TO UNDERSTAND A VARIETY OF ACCENTS.

According to Peter Mahoney, general manager of the division of Nuance that produces its Dragon Naturally Speaking products, modern dictation programs no longer need the 30–60 minute initial training period that earlier versions did.



"People now get well above 90% accuracy the first time they use [our dictation programs]; that figure used to be only 75% to 80%."



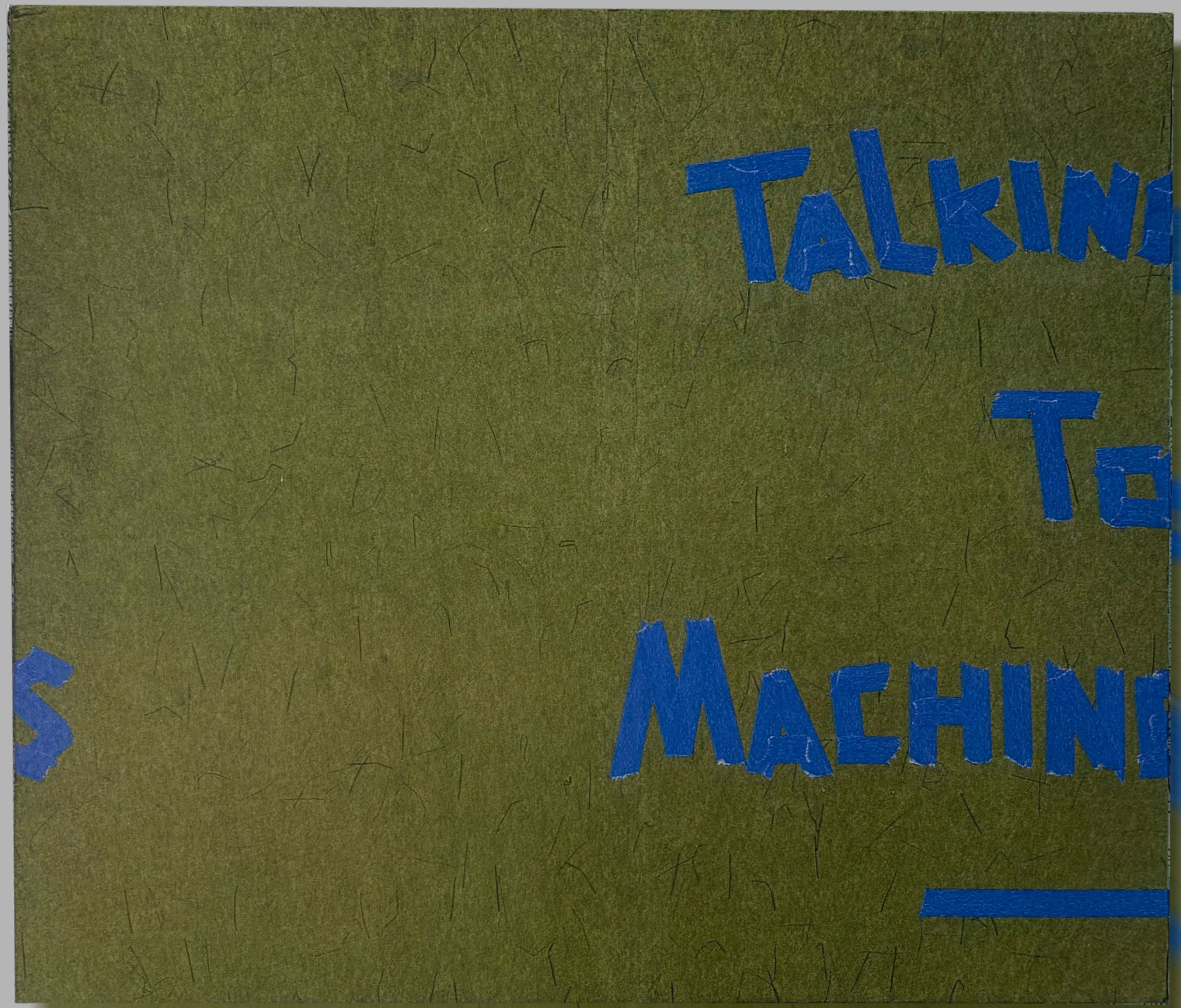
“The program looks at three things to get to know its user,” says Ma- honey.

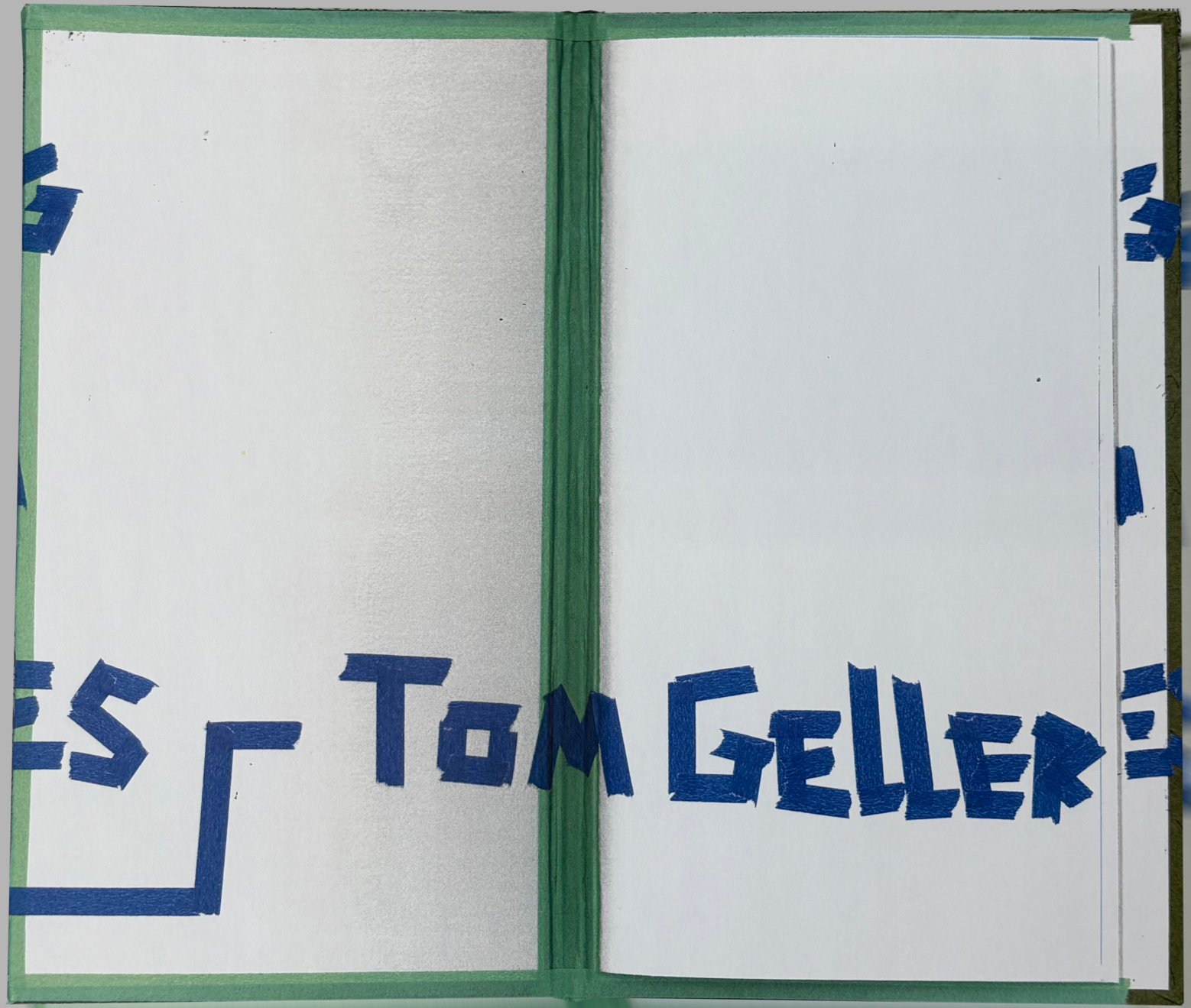
“First is acoustics. How do you actually say words? That’s what the ini- tial training used to focus on.
Second is the kind of words you use—your spo- ken writing style.
Third is a variety of user preferences, such as how you say numbers and names, and how you like them to be formatted.

ALTHOUGH INITIAL TRAINING IS NOT AS NECESSARY AS BEFORE, DICTATION PRO- GRAMS STILL TRAIN THEMSELVES AS YOU USE THEM.

So if you were to say, ‘I gave you two dollars and forty- two cents,’ the program knows to tran- scribe that as ‘I gave you $2.42.’”

These two methods of understanding speech are starting to converge, however. Dictation programs adapt somewhat to the tasks they are performing, for example favoring the phrase "dot com" when the cursor is in an email program's "To" field. And Siri switches from command to dictation mode when appropriate.


TALKIN
TO
MACHIN


TOM GELLER

Siri

When APPLE integrated Siri into the iOS operating system last October, it spurred iPhone owners to start talking to their phones as well as through them.

The program, which converts spoken commands such as

"Schedule dinner with Lisa at 6 tonight"

into calendar appointments, Web searches, and the like, is the most widely distributed example of a cognitive assistant to date.

More than four million iPhone 4S's featuring Siri were sold during its first weekend.

ALTHOUGH USERS MIGHT SEE IT AS SIMPLE SPEECH RECOGNITION, ITS ABILITIES GO FAR BEYOND SIMPLE TRANSCRIPTION.

As Wolfram R... research executive director Luc Barthelet says,

"The news ab... computers to... 15 years, but... threshold whe...

...out Siri is that it works. People have tried to get ...answer questions conversationally for at least ...only now has the technology reached a ...re people overall like it."

The iPhone's... assistants w...

...popularity also gives intelligent software ...der exposure than they would get otherwise.

SIRI REPRESENTS AN ... IMPORTANT MOMENT WHEN VOICE RECOG... NITION, INFORMATION MANAGEMENT, ARTI... FICIAL INTELLIGENCE, TASK FULFILLMENT, ... AND USER INTERFACE MARRY IN A WAY ... THE GENERAL PUBLIC FINDS USABLE ... AND PRODUCTIVE.

Roger K. Moore, editor in chief of the journal Computer Speech and Language, points out tha...

"the field of research hasn't changed dramatically... What's new is that Siri's brought several complementary technologies together. Our business has been going for many years. Only now, with Siri, everybody knows about it."

Executing Commands

There is a long road between the spoken command and its fulfillment , though. The first step in ... is to conv... the process ...rt the audio ...speech into ...ning. The two main applications of speech recognition —dictation and command recognition— have forced researchers to pursue parallel methods that balance ...ccabulary, accent, and context ...eeds.

Grammar-based voice recognition is optimized for situations where the program has a very good idea of what the speaker will say.

As a result, the system needs to ...only a limited vocabulary.

BUT

According to Dan Faulkner, vice president of product and strategy for the Enterprise Business Unit at Nuance, responses can vary widely no matter how restricted the domain is.

When a phone prompt system tells you to 'please say yes or no,' we might be pretty confident the speaker will say 'yes' or 'no.'

But 'yes' could be 'yeah,' 'that's correct,' or 'yup.'

and in the Southern states some people will say 'yes, ma'am' and 'no, ma'am.'

So even for something as simple as 'yes' or 'no,' you need quite an extensive list of phrases.

BY CONTRAST, A LANGUAGE-BASED VOICE RECOGNITION SYSTEM IS OF... DICTATION. IT MAKES FEW ASSUMPTIONS ABOUT CONTEXT, ATTEMPTING TO RECOGNIZE AND TRANSCRIBE EVERY WORD IT HEARS. IN GAINING A WIDE VOCABULARY, IT GIVES UP THE ABILITY TO UNDERSTAND A VARIETY OF ACCENTS.

According to Peter Mahoney, general manager of the division of Nuance that produces its Dragon Naturally Speaking products, modern dictation programs no longer need the 30–60 minute initial training period that earlier versions did.

"People now get well above 90% accuracy the first time they use [our dictation program]. That figure is up to ... by 95% to 98%."

ALTHOUGH INITIAL TRAINING IS NOT AS NECESSARY AS BEFORE, DICTATION PRO- GRAMS STILL TRAIN THEMSELVES AS YOU USE THEM.

"The program looks at three things to get to know its user," says Ma- honey.

"First is position. How do you actually say words? That's what the ini- tial training used to focus on. Second is the kind of words you use — your spe- lcific writing style. Third is a variety of user preferences, such as how you say numbers and names, and how you like them to be formatted."

So if he were to say, "I gave you two dollars and sixty-two cents," the program knows to type out "I gave you $2.62."

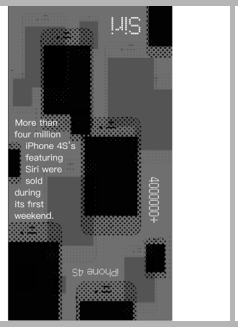
These two methods of understanding speech are starting to converge, however. Dictation programs adapt somewhat to the tasks they are performing, for example favoring the phrase "dot com" when the cursor is in an email program's "To" field. And Siri switches from command to dictation mode when appropriate.

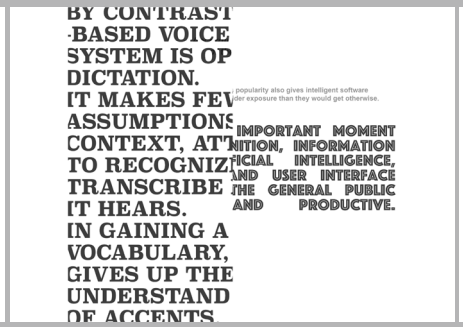
dot com

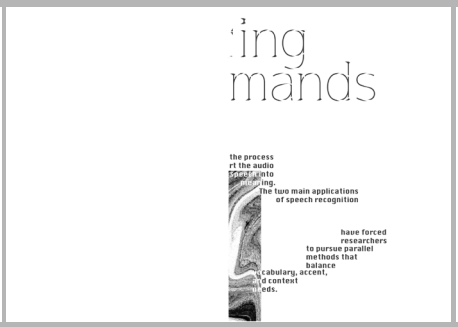
@ TO

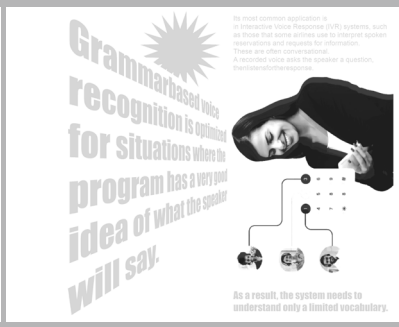
RISO Printing

Blue Color

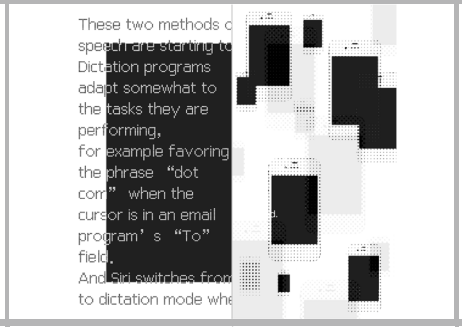


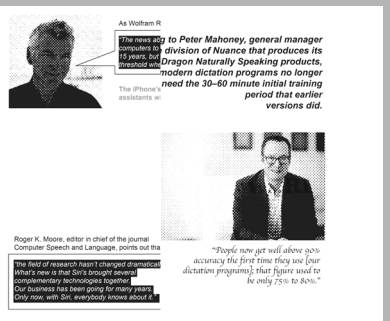


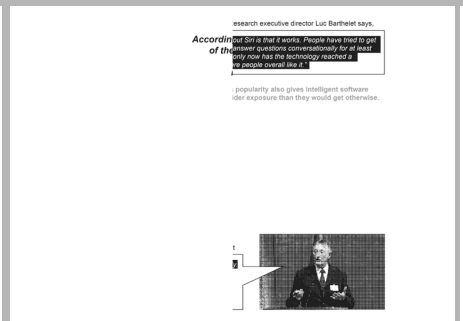


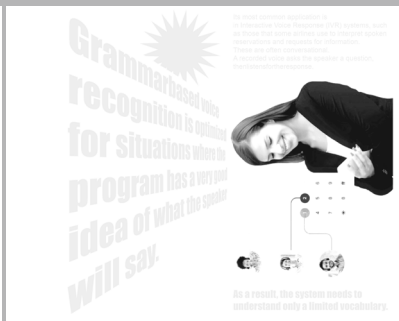


Pink Color



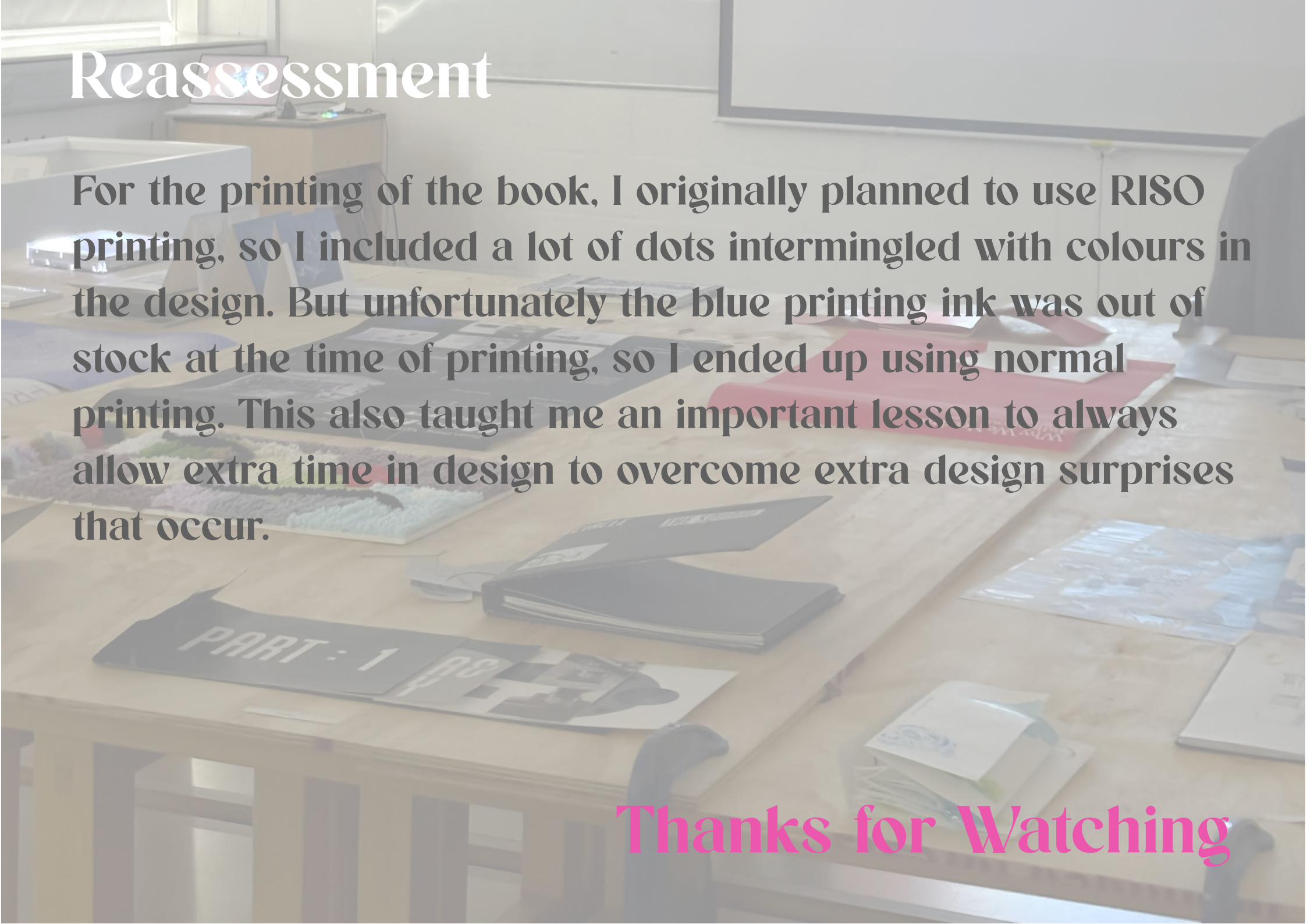






Reassessment

For the printing of the book, I originally planned to use RISO printing, so I included a lot of dots intermingled with colours in the design. But unfortunately the blue printing ink was out of stock at the time of printing, so I ended up using normal printing. This also taught me an important lesson to always allow extra time in design to overcome extra design surprises that occur.



Thanks for Watching