

Voice Technology in the Distribution Center

May 2002

KOM International White Paper Series

SupplyChain



Voice Technology in the Distribution Center

Contributors: Marc Wulfraat, Partner

© 2002 by KOM International, Inc.
All rights reserved. Published 2002.

Restricted Rights

Printed in Canada.



The information contained within this document is proprietary and confidential to KOM International, Inc.

No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, for any purpose without the express written permission of KOM International, Inc.

This document is subject to change without notice, and KOM International does not warrant that the material contained in this document is error-free. If you find any problems with this document, please report them to KOM International.

KOM International and the KOM International logo are registered trademarks of KOM International, Inc. Copyright © 2002 KOM International, Inc. All rights reserved. All other company and product names may be trademarks of their respective owners.

This document contains or may contain statements concerning possible functionality for supplier software and hardware products and technology. The research contained within this document has been collected from end-users of the technology and from the suppliers of these solutions. Kom International, Inc. has taken much care and attention to reflect the greatest accuracy of comments regarding solutions and suppliers within the body of this report and at no time is there intent to indicate an implied preference for any specific solution provider or commercially available technology solution. KOM International disclaims any express or implied intent to provide information that is in any way inaccurate or misleading.

Voice Technology in the Distribution Center

Introduction

The use of speech recognition technology in the distribution center is a relatively new phenomenon. Since the late 1990's, an increasing number of high volume distribution companies have invested in speech technology solutions to increase order accuracy with the goal to improve customer or store service levels. There is skepticism among industry experts as to whether or not voice technology solutions will be widely adapted in the distribution industry. The main concerns are the relatively high solution cost and the difficulty companies have in cost justifying the investment.

This document provides an overview of the use of speech technologies in the distribution center. Topics reviewed include the:

- Evolution of the industry
- Types of speech technology solutions
- How speech technology works in the distribution center
- Types of operations and functions most suited for speech technology
- Benefits of speech technology
- Warehouse Management System Integration
- Costs of a turnkey speech technology solution
- Profiles of the speech technology solution providers for the distribution industry

Speech Technology Industry Evolution

Scientists at Bell Laboratories began working on speech technology prior to World War II. During the cold war years, the U.S. federal government funded research in this area and as computer manufacturers introduced hardware with faster processing speeds and enhanced memory capacities, a number of companies began to make advances in developing software packages that could reproduce and recognize the human voice.

Initially, the majority of investment into voice technology was for applications that manage telephony automation, provide automated directory assistance, and automated call direction or unified messages. Over the past 15-20 years, speech synthesis and speech recognition technologies have also been introduced for use in industrial applications including manufacturing and distribution. This past year, the investment firm Morgan Keegan & Co. Inc. estimated that the market for the three main types of speech technology – speech recognition software, speaker verification software and text-to-speech software – was worth \$153 million in sales in 2000. It expects the market to exceed \$1.5 billion by 2005, a ten-fold increase in just five years.

Today, companies in the banking, insurance and financial service markets are investing significant capital to deploy speech technology to automate customer service call centers with the goal to improve customer service. For example, customers can now conduct financial transactions such as stock trades through simple voice commands spoken to computers. Automobile manufacturers have even started to deploy this technology into car navigation systems to guide motorists.

What is relatively new is the incorporation of speech recognition technology into portable or "wearable" computers. This application is proving to be valuable to the industrial workplace, saving money in labor-intensive applications such as distribution center order picking, maintenance, repair, package sortation and handling, returns processing and inspection. In distribution-centric operations, the technology is primarily being deployed to increase shipment accuracy by improving the quality of order picking operations that can benefit from a paperless hands-free environment.

When the first voice technology solutions began to appear on the market in the 1980's, the price tag was cost prohibitive for most companies. Over time, early adopters began to take notice and invest in these solutions and a competitive market began to take shape. By the late 1990's, the cost of a turnkey speech technology solution dropped to the price point where more companies could afford to justify the investment. Over the past several years, the demand for speech technology in the distribution center has grown by more than 50% annually.

Today, over 150 companies have invested in speech technology for distribution operations including many retailers and wholesalers with labor-intensive high volume distribution operations. Firms that have implemented speech recognition technology in their distribution

centers include: Wal-Mart, Fleming, Kroger, Roundy's, P & C Grocers, Price Chopper, Safeway, Somerfield PLC, SuperValu, K-V-A-T Food Stores, Giant Eagle, Mid-Mountain Food, Perishables Distributors of Iowa (a subsidiary of Hy-Vee, Inc.), Reed Boardall Cold Storage, Nabisco Biscuit Company, Nautica Enterprises, P&O Cold Logistics, Atlas Cold Storage, U.S. Foodservice, PYA Monarch, 7-Eleven U.S.A., Haggard Clothing, United Parcel Service, Henry's Foods, Freedman Group LLP, Corporate Express, Campbell's Soup, Ford Motor Company, Mitsubishi, Saturn, Timken Company, and Robbins Lumber Company, amongst others.

Types of Speech Recognition Technology Solutions

For industrial applications, there are basically two types of speech recognition technology: speaker-independent and speaker-dependent systems.

Speaker-Independent Technology

Speaker-independent technology solutions recognize the spoken words of many people speaking a specific language. A good example of speaker-independent technology is an automated directory service application whereby the computer software has no way of being able to predict the voice pattern of the person calling and millions of different voices need to be synthesized and recognized every day. These solutions require little training but tend to be less effective where there is a diverse speaker population with regional accents and/or multilingual environments and in high-noise environments.

In the speaker-independent environment, a standard pre-defined lexicon of words is established and operators are simply trained on which words to say for specific task confirmations and situations. For example, after picking a product, an operator will always say "got it" to confirm that the task is completed. The computer does not recognize the operator's specific voice, rather it recognizes the standard spoken words "got it". Since there is no notion of a user-specific voice template, operators do not need to establish or download a voice template. Thus all computer processing power is performed by a centralized computer server rather than by mobile computer terminals worn by operators. In technology terminology, this is referred to as thin client / fat server because the central "fat" server does all of the processing.

Speaker-independent technology requires that each and every operator dialogue must be communicated from the operator's headset to the operator's mobile terminal through the RF network to be processed by a central server such that a directive can be returned to the operator in the reverse fashion. This can result in critical failures if a dead zone is encountered (i.e. an area of the distribution center where RF communications may not be available due to the density of product storage in that area on that day). The intensity of the network traffic required through the "fat server" architecture is also inherently a limiting constraint in terms of the number of concurrent users at any time.

Speaker-Dependent Technology

Speaker-dependent technology solutions require that each operator go through an initial enrollment process of training the computer to recognize their specific voice profile for all words that they enunciate during their daily routine. These solutions require a voice template to be recorded for each operator so that the computer understands the exact nature of the person's dialect, tone and voice. It is generally accepted that speaker-dependent technology is more accurate than speaker-independent technology and is now at the point that a person's voice is as secure as his or her fingerprint. Speaker-dependent technology is generally suitable for mission-critical applications such as in high productivity industrial applications or in high security applications such as financial institution transactions.

With speaker-dependent speech technology, each operator's voice template is recorded to a software file that is stored on a computer workstation / Windows NT server connected to a wireless radio frequency network in the distribution center. At the start of each workday, each operator invokes a wireless download of his or her voice template. In turn, the operator's personal voice template file is transferred through the wireless radio frequency network to the operator's wireless computer in a matter of seconds. The mobile computer does the majority of the processing thereby this technology is referred to as a "fat" client / "thin" server technology. The "fat client" approach potentially reduces a significant volume of network traffic which is a key concern for high scalability requirements.

For distribution centers operating multiple shifts, mobile wireless computers can be shared between multiple operators on different shifts. Each time a new shift begins, operators will re-initialize the wireless computer terminal by requesting their personal voice template. It is important to note that no additional training is required at each login. The only equipment that is not shared is the headset for sanitary health reasons.

In industrial environments such as in manufacturing plants, distribution centers, warehouses and sortation facilities; the use of machinery and equipment (e.g. conveyor belts) contribute to significant background noise levels often in excess of 90 decibels. This is an important consideration when evaluating speech recognition technology solutions. In austere operating environments with high noise conditions, it is recommended to use an accurate, high performance, noise-robust interactive speech recognition technology and these are the characteristics of a speaker-dependent technology solution.

Text-to-Speech Versus Digitized Technology

To instruct operators on where to go and the next task to be performed, text commands need to be converted to speech commands that are communicated through the operator's headset. There is essentially two ways to do this: Digitized voice or text-to-speech.

Text-to-speech requires that software be deployed to convert text commands into computer-generated synthetic voice commands. Synthetic voice commands sound like a computer talking such as the voice directives one would hear when calling up operator assistance, airline reservations information, etc.

Digitized speech is a human voice that is digitized into a sound file. To accomplish this, a human being must first record the basic vocabulary or lexicon to be issued by the computer. In general, a digitized human voice is more pleasant to listen to than the synthetically generated computer voice which sounds more robotic.

Digitized speech may or may not be practical or feasible depending on the number of words that need to be digitized. For example, in distribution operations where the computer instructs the operator to pick an item at a location and the description field needs to be articulated as part of the process, it would be difficult to maintain a voice template of all item descriptions. Having said this, most users of speech technology prefer to minimize the level of computer chatter therefore this issue is not applicable to all operations.

How Speech Technology Works in the Distribution Center

Speech technology solutions require that operators in plants and distribution centers wear wireless voice powered computers with headsets. The operators respond to computer-generated speech commands by speaking into the headset to confirm that the command has been executed. The wireless computers are light-weight (i.e. approximately one pound) and are typically worn on waist-belts. These portable computers communicate back to a host system through a wireless radio frequency network based on the current 2.4 GHz communications standards. The host computer can be an enterprise resource planning system (ERP) or more typically, a dedicated warehouse management system (WMS) that creates text-based work instructions issued via traditional paper-based pick lists or radio frequency handheld terminals.

The host computer creates work files throughout the day that are essentially text-based instructions. The data in these work files must be presented to the voice application software where it can then be converted into speech commands through text-to-speech conversion software or through a pre-recorded digitized text file. To accomplish this, an interface must be developed between the host computer and the software that controls the speech technology solution. These tasks are placed into a prioritized task queue for assignment to operators based on the preferred operating methodology(s) deployed.

A simple example might be an outbound sales order that would normally be printed on a paper pick list. Typically, the entire pick list is batch downloaded to an operator's wireless computer to initiate the process. The pick tasks are converted by the application software into computer-generated speech commands such that order pickers are verbally instructed where to work without the need for paper documents. Depending on the capabilities of the host system, the communication of pick results can be communicated back to the host in real-time or through a batch upload at the completion of the order or at the time of an exception occurrence (Ed. Note - Clearly, the former approach of real-time task confirmation is preferred as it enables superior operations and improved replenishment synchronization with picking). Either way, operators confirm transactions by verbally communicating through the headset. The operator's speech is synthesized to text for confirmation by the computer and the process repeats itself until the work is completed.

At a recent site visit to a grocery distribution center utilizing voice technology, I followed an order selector who was being guided by speech instructions via a headset and mobile computing terminal. In this case, the technology being used was of the speaker-dependent type. The process is as follows:

- The operator walks through the pick line with a battery-powered double pallet jack and concurrently selects two pallets of full case grocery cases to be shipped to one of the firm's retail supermarkets.
- This operator happened to be a music aficionado who worked with a large portable "boom-box" pounding out rap music at full volume (Ed. Note – OSHA was not a frequent visitor at this facility).
- As the operator walked down the aisle, the headset's speech instructions navigated him to the next pick location in the travel path.
- The operator confirmed that he was at the right location by verbally communicating random check-digits (e.g. 12 spoken as one-two or check one-two) appearing on a label placed on the rack beam above the product's pick location. Note that check digits are usually recommended to be 2 digit numeric identifiers for ease of use.
- If the check digits match the expected response then the system instructs the operator on the quantity to pick, otherwise a mismatch requires that the computer redirects the operator to the correct location. In some distribution operations, the system may also articulate the item number, description, unit of measure, UPC code, etc. (Ed. Note – in general, solution providers have designed software that allows all dialogues to be configured by the end user). In general, it is best to keep dialogues to an absolute minimum so as not to slow down the operators.
- Different approaches are available as regards how to handle the issue whereby the operator may or may not pick the correct quantity and experimentation may be required to identify the best approach for a given distribution operation.
 - To keep computer chatter to an absolute minimum, some companies have chosen to never have the quantity spoken by the computer unless it is greater than one. This approach is fine if the vast majority of picks are for a quantity of one and the unit of measure being picked is consistent (e.g. cases). As well, this approach requires that there is no mixing of products in the location being picked from (pick or reserve). If either of these conditions is false then more information needs to be conveyed to the operator such as item number, unit of measure, etc.
 - If a quantity > 1 needs to be picked then the computer can direct the operator by indicating the quantity (e.g. Pick 5). The operator can confirm that the quantity has been picked through a confirmation (e.g. "five" or "Grab five").

- Alternatively, a countdown approach can be used whereby the operator is required to count down each unit picked with the intent to increase the probability of picking the right quantity (e.g. five-four-three-two-one). Operators may tend to circumnavigate this by just quickly reading out the countdown but date-time stamps in the task history allow a post-audit of this behavior to take corrective action.
 - One other approach is that quantities > 1 be communicated to operators in standard multiples of X (e.g. X=5). An example might be whereby a quantity of 20 cases to be picked would be communicated to the operator as four commands of 5 cases. This approach enables increased control (versus a single command to pick 20) but this approach may not be popular with order pickers who would like to know in advance if 20 cases are to be selected to enable improved pallet building.
 - The reason that this issue draws attention is that if the goal of speech technology is to eliminate pick errors then it is critical to eliminate errors that are quantity related. In this regard, there are multiple techniques that can be deployed to increase quantity accuracy and each distribution operation needs to be configured according to its business needs.
- In the situation where a pick location is empty or the pick location is short of the required quantity, the operator instructs the computer of the quantity that can be picked so that he can proceed to the next pick task or wait until the next replenishment task has been executed. In the ideal scenario, such an occurrence would immediately trigger a prioritized “hot” letdown to replenish the pick location.
 - At any time, the operator can request an information repeat in case he forgot or did not hear the instructions. For example, the location name or quantity to be picked can be repeated at any time.
 - In distribution operations where additional item information needs to be captured as part of the pick process, the dialogues can be configured to require item-specific data capture (e.g. catch weights) and as well, portable scanners can be connected to the mobile computer to enable scanning of information with longer data streams that are too time-consuming to speak (e.g. lot number, serial number, UPC, etc...).

The most illuminating aspect of this particular site visit was that the system had no difficulty recognizing and interpreting the operator's speech despite the fact that less than 3 feet away, the background rap music was extremely loud. The operator picked his order and built his pallets at a rate of roughly 200 cases per hour, or one case every 18 seconds, with no paper lists, labels or other handheld devices. It is interesting to note that the operator consistently communicated the location's check digit well in advance of actually arriving at the location, thereby enabling faster work, and in this case more incentive pay. This particular company reported a significant reduction in warehouse employee turnover since they implemented the technology.

In this distribution center operation, management decided to implement a random audit procedure to ensure shipment accuracy. Operators are randomly audited by a supervisor at any time during the picking or pre-shipping process, mainly to catch quantity mis-picks. As well, the random audit keeps people honest by monitoring the situation where operators don't want to pick heavy cases even though they have confirmed to the system that they did the work.

Types of Operations and Functions Most Suited for Speech Technology

To date, firms that have invested in speech technology include, but are not limited to the following industries:

- Retail and Wholesale Grocery
- Foodservice
- Food Manufacturing
- General Merchandise / Health and Beauty Care
- Convenience Stores
- Apparel / Garment
- Automobile Manufacturing
- Package Sortation

Speech technology solutions have been deployed in functional operations that include, but are not limited to:

- Full case order picking
- Split case order picking
- Flat garment order picking
- Manufacturing quality assurance processes
- Pallet receiving and breakdown operations
- Putaway and replenishment forklift operations
- Cycle counting
- Package sortation operations that require human intervention for non-labeled packages

As mentioned earlier, in distribution operations where outbound lot number or serial number tracking or other extraneous information is required to be captured, it is too time consuming to expect operators to verbally read and speak extensive character strings. To accommodate this business requirement, a bar code scanner can be plugged into the operator's wireless computer to enable the scanning of bar coded lot or serial number information. The integration of speech and scanning technology in these distribution centers is an example of how the best of both worlds is achieved.

Currently, the vast majority of speech technology implementations are for high volume full case picking operations where there are ergonomic advantages to having hands-free paperless picking operations. The Grocery and Foodservice industries have been early adopters of this technology for several reasons:

- These industries have very low net margins and are characterized by labor-intensive high volume distribution operations, especially in the order picking function. Speech technology provides improvements in shipment accuracy and labor productivity that are of critical concern to these companies.
- Food distributors often have up to 50% of their operators working in cold storage environments such as freezers where gloves are required to be worn. The handling of paper lists or labels or the use of radio frequency equipment is cumbersome and difficult to handle with gloves. A key benefit of voice technology in this context is paperless and hands-free ergonomics.
- The capture of catch weight information (i.e. variable weights associated with meats, cheeses, etc.) can be done verbally during the pick process thereby increasing accuracy and saving time and labor associated with keypunching this hand-written data after orders are picked.
- The labor involved with order checking to validate shipment accuracy can be extremely costly and in many operations the savings in reduced checking labor requirements is enough to cost justify the technology investment with less than a 2 year payback.

There are other distribution industries that are excellent candidates to derive benefits speech technology, but have been slow to take advantage of these solutions. For example, the drug wholesale industry is characterized by razor thin margins, high volume shipments of high value products where order accuracy is critical. Other industries that come to mind are retail mass merchants, automobile aftermarket parts distributors, high volume office supplies distributors, garment distributors, aerospace and defense, liquor and spirits, computer electronics, home furnishings, candy and tobacco, health and beauty care, and so on. In short, companies that stand to gain the most benefit from speech technology in the distribution center are those firms that distribute high volume or high value merchandise where there is low tolerance for errors that incur lost sales or costly returns processing. In short, if shipment accuracy is a mission critical requirement for your distribution operation and operators benefit from working in a hands-free environment, then speech technology represents an excellent opportunity to concurrently improve efficiency and service levels.

In terms of competing technology solutions to improve order accuracy in the distribution center, the two most common alternatives to speech technology are radio frequency / bar code scanning technologies and pick-to-light systems. Both of these alternatives are mature solutions that enable high productivity and high order accuracy in a paperless environment. The pros and cons of each alternative is beyond the scope of this document, however the following points summarize the tradeoffs of each solution:

- Radio frequency / bar coding handheld devices are about half the cost of a speech recognition terminal and therefore represent a less expensive solution which is an important consideration for small and medium sized distribution operations. RF devices make sense when detailed bar code information needs to be captured with each transaction (e.g. UPC, lot/serial number, etc.). When companies mix products and/or units of measure in pick and reserve locations, it becomes crucial to scan the product being handled hence RF/bar coding makes a logical choice for these environments.
- Pick-to-light systems are typically deployed in broken case picking environments where products are picked from and stored in case flow racks. Operators are guided to pick their orders by lights that indicate the pick location and the quantity to pick. Operators push a button to signal the completion of the pick and the next location illuminates. Pick-to-light solutions are commonly used in distribution centers where pick rate productivity ranges between 300 –500 order lines per hour or more such as in high volume health and beauty care distribution centers. In these types of operating environments, pick-to-light technology represents a more suitable paperless solution because the use of speech technology would simply slow operators down to a much slower pick rate. Though it is always dangerous to generalize, pick-to-light is a less flexible solution because it is installed in conjunction with fixed equipment in the distribution center. Therefore, it is difficult to adapt to sudden changes in order profiles, new product lines or significant changes to the distribution operation. On other consideration is that pick-to-light solutions are almost deployed in environments where operators do not need to pass one another during the pick process.

Benefits of Speech Technology in the Distribution Center

A summary of the main benefits of speech technology discussed in this section follows:

- Order accuracy
- Increased productivity
- Faster, hands-free, easier picking
- Real-time inventory information
- Elimination of pick lists and labels and clerical work associated with paper-based documents
- Faster, easier training
- Support for multi-lingual operators in the distribution center
- Improved ergonomics and safety

- Reduced employee turnover

Order Accuracy

The majority of companies investing in speech technology cite order accuracy as the primary benefit to cost justify their capital investment. There is a quantifiable price tag to pay for order picking errors and these companies have determined that a significant reduction of pick errors and therefore reduced checking labor and reduced returns is enough to justify the expenditure.

For self-distributing retailers, the dollar value of a pick error is somewhat difficult to identify because it may or may not result in a lost sale at the retail store. The self-distributing retailers interviewed for this white paper used a 15-20% factor to estimate lost sales at retail resulting from a mis-pick. For example, if an operator mis-picks a product with a retail value of \$30 then lost sales is valued at $\$30 \times 15\% = \4.50 . Self-distributing retailers will debate this factor because it depends on how well stocked the retail stores are and the frequency of store deliveries during the week and so on. Some will argue that the factor is closer to 30% so the cost justification exercise needs to be based on a factor that is agreed to by store operations and the CFO.

For wholesale distributors and manufacturers, the cost of an order error is much easier to identify as it typically involves direct costs to the organization for returns processing including many cost factors such as returned freight expense, accounts receivable labor, additional handling, etc. In these companies, an order error can easily result in \$20 - 50 of lost income or more.

The majority of case studies to date indicate that the deployment of speech technology in the distribution center reduces pick errors between 70 – 90% and short-shipments by about the same amount. One company using voice technology indicated that pick errors (measured as a combination of incorrect item and incorrect quantity mistakes) were running at 3 per 5,000 order line items and that 100% of the mistakes were related to quantity errors. Others have quoted numbers closer to 1 error per 10,000 lines. Much of this depends on exactly how errors are captured and recorded and if indeed customers always report them.

The table below provides an overview as to the source of order selection errors based on a recent survey of 96 grocery wholesalers, food service distributors and food manufacturing companies conducted for WERC by Thomas W. Speh, Ph.D.

Types of Order Selection Errors	
Type of Error	% of Total Errors
Wrong product picked	41%
Wrong quantity – short picked	25%
Customer order error	25%
Damaged product	13%
Wrong quantity – over picked	12%
Wrong pack size	5%
Other	20%

Speech technology solutions improve order accuracy by ensuring pickers select the right product and therefore reduce the number of mis-picks released by the distribution center. According to the above data, mis-picks of the wrong product account for 41% of all order accuracy errors that are incurred by the end customer.

In general, the distributors interviewed for this report indicated that order accuracy increased from 99.3 to 99.8% or higher. Though a 0.5% order accuracy increase may seem insignificant to some readers, consider the following example that illustrates the impacts of this improvement for a self-distributing retailer:

- A company ships 500,000 cases per week with an average value at cost of \$15 per case or \$7,500,000/week.
- At 99.3% order accuracy, the company mis-picks 3,500 cases per week worth \$52,500. Assuming a 20% lost sales factor and a 25% retail markup, this equates to \$683,000 lost sales per annum at retail.
- At 99.8% order accuracy, the company miss-picks 1,000 cases per week worth \$15,000. Based on a 20% lost sales factor and a 25% retail markup, this equates to \$195,000 lost sales per annum at retail.
- In this example, the gain in order accuracy has reduced lost sales at retail by \$488,000 per annum.
- If this company invested \$450,000 into a speech technology solution, they would experience a conservative 1-year payback based on reduced lost retail sales alone.
- This simple exercise has not quantified the many other savings derived from improved accuracy and productivity (e.g. reduced checking labor, reduced training, etc.) which will further improve the payback timeline.

The same exercise for the wholesaler / manufacturer that incurs an average cost of \$20 per return is as follows:

- Reduced mis-picks of 2,500 cases per week x \$20 per miss-pick equates to reduced returns costs of \$2,600,000 per annum.
- If the \$20 cost per return seems excessive (or too low) then substitute the dollar amount that makes most sense for your business. (Ed. Note: there are many industries where \$20 is a significant understatement for the cost of a return).

Increased Productivity

Most companies investing in voice technology do not include labor productivity gains as part of their cost justification exercise. That is not to say that these companies don't wish to improve productivity, rather they are reluctant to make efficiency improvement claims due to the uncertainty of the results. It is fairly intuitive to understand how speech technology yields improved productivity as checking labor can be significantly reduced and in many cases order picking labor productivity improves as well.

Many companies that have implemented voice technology for order picking operations have reported definite order picking productivity gains, typically varying between 5 to 15% and in some cases even higher. Productivity gains tend to be higher in cold temperature environments where operators wear gloves and therefore reading and handling paper documents slows them down.

Whether or not increases in picking productivity can be attributed to the use of speech technology is a question of how productive the operation is to begin with. Three examples below demonstrate this point:

- A health and beauty care distributor runs a pick-to-tote-to-belt operation for all split case merchandise. The average pick rate is 1,000 pieces per hour or the equivalent of 350 order lines per hour. Operators work alongside a conveyor and are dedicated to a zone or aisle throughout the day. In this environment, speech technology is likely to be unsuitable as the operator may actually be slowed down by having to verbally communicate confirmations with every pick. Keep in mind that 350 order lines per hour is one item picked roughly every 10 seconds.

- A grocery distributor is averaging 200 cases picked per hour in its dry grocery distribution center. The operators use adhesive pick labels that identify the location and product to be picked. Labels are placed on cases and the cases are picked to pallets in a conventional operation. There are no obvious constraints such as overcapacity conditions, crowded aisles or congested docks that prevent this number from improving. The implementation of voice technology could quite conceivably improve this pick rate from 200 to 210 if pick labels are eliminated. (Ed. Note – more than one grocery company has cost justified the investment into voice technology despite the fact that the pick labels could not be eliminated for customer service reasons).
- A wholesale distributor specialized in fresh meat must record the case weight for each individual case being shipped to the customer since invoices require that product is billed out on a per pound basis. The most common approach is to keypunch in case weights at the end of the pick process prior to shipping. The value of the merchandise is relatively high and the clerical keypunch errors result in invoicing inaccuracies, a source of customer dissatisfaction if they are over-charged (and joy when they are undercharged). The implementation of speech technology eliminates the need for operators to write down catchweights on pick lists and also eliminates the clerical function of keying in this data.

These examples serve to point out that there is no such thing as a boilerplate plug-and-play spreadsheet to cost justify the investment into speech technology on the basis of productivity improvements. It takes an educated understanding of distribution operations to identify realistic and achievable productivity improvement figures.

The companies interviewed for this report all reported the majority of productivity gains in the form of reduced auditing and checking labor. For some firms, the checking function labor reduction exceeded \$500,000 - \$1,000,000 per annum. The bottom line is that with order accuracy exceeding 99.8%, the time delays and labor spent validating shipment accuracy can be reduced substantially.

Faster, Hands-Free, Easier Picking

Using voice technology, order pickers will communicate the confirmation check digits to the computer terminal well in advance of reaching the pick location. The computer is already instructing the operator on the next pick task while he or she is building the pallet or depositing the merchandise to a shipping container. Thus the operator has the opportunity to “look ahead” at the next pick without having to revert back to the small print on paper documents where digits can easily be inverted during the rushed reading process. The elimination of the paper documents and the need to read and interpret information from those documents is the principle source of providing easier picking operations.

Most grocery retailers / wholesalers and foodservice companies use adhesive pick labels to guide order selectors through the pick path. Operators use the labels to identify the location, item and quantity to be picked. The labels are then placed on selected cases as a means of identifying the merchandise for the customer. With voice technology, pick labels can *potentially* be eliminated thereby enabling operators to work in a paperless hands-free environment. In distribution centers where operators pick full cases at a rate of 200 cases per hour x 8 hours per day x 30 pounds per case = 48,000 pounds per day, the advantages of ergonomic hands-free operations are easy to appreciate. In cold-temperature environments where meat, produce, dairy / deli and frozen food products are stored and operators wear thick gloves to keep warm, the elimination of labels and/or paperwork is even more beneficial.

Real-Time Inventory Information

The implementation of voice technology in the distribution center may *or* may not enable real-time inventory information depending on the capabilities of the underlying enterprise resource planning (ERP) or warehouse management system (WMS) application in place. In general, speech technology solutions are designed to support real-time inventory transactions whereby each inventory “movement” within the distribution center is recorded in the warehouse management system as it happens. Unfortunately, not all enterprise applications and warehouse management systems were designed to support real-time transaction processing and therein lies the limiting constraint.

In high volume distribution center operations, the concept of real-time order picking is of great value. First, the timing of letdowns (i.e. replenishments) can be closely synchronized with picking so that pick locations are replenished as needed rather than having product clutter the aisles during peak picking operations. Products that are letdown into aisles awaiting replenishment risk being damaged by operators driving equipment and also results in lower pick productivity rates.

In the real-time operating environment, the recording of inventory discrepancies (i.e. inventory shortages) at pick locations is immediate thereby enabling a replenishment task to be initiated in response to the discrepancy. This translates into a rapid reduction in the time

required to resolve inventory discrepancies thereby eliminating lost productivity associated with waiting for and chasing inventory after the fact.

Real-time transaction processing enables higher quality inventory information as inventory transaction updates are posted as they happen rather than a delta of time later. In fact, inventory becomes so accurate that discrepancies become difficult to quantify because they become relatively insignificant. This translates directly into less inventory adjustments, less searching for lost inventory, improved order fill rates, and increased sales and reduced inventory levels. The bottom line is that there is a significant efficiency and accuracy value to real-time inventory information in any distribution center.

Enterprise Resource Planning (ERP) applications and warehouse management systems that do not support real-time order picking transactions are usually constrained for one of two reasons.

- The software was never designed to support real-time order picking in the first place as batch order processing was the order of the day when the software was first developed
- The extensive input/output dialogues generated through real-time order picking confirmations in high volume transaction processing requirements would result in a serious degradation of system performance and therefore reduced labor productivity.

For companies facing this situation and seeking to achieve the benefits of real-time inventory, there is not much choice but to either work with your existing software partners to modify your system or seek new software systems. Neither proposition is particularly attractive and will involve significant investments in capital and time. As such, one must determine what the quantifiable benefits of real-time inventory information are and if there is a cost justification to modify or upgrade software systems to get there.

Elimination of Pick Labels/Lists

Pick labels are commonly used in distribution operations that ship merchandise through company-owned private fleets whereby the label provides the driver and the customer with information required to deliver and check in the goods. When merchandise is shipped through third party carriers such as LTL or via UPS, carrier-compliant shipping labels are required on every shipping carton/pallet. As well, when merchandise is shipped to specific retail mass merchants, retailer-specific compliance labels are required on all cartons. Thus, the elimination of pick labels/lists and shipping labels may *or* may not be feasible.

When pick labels and lists can be eliminated and a true paperless pick environment is possible, then the savings in paper and clerical labor can be calculated quite easily. The fact remains that this source of savings represents a small dollar amount relative to the overall benefits of deploying speech technology.

Faster, Easier Training

Most companies that have implemented speech technology discover unexpected benefits in productivity gains relating to the training of new order pickers. A common industry practice is to have an experienced senior order selector walk alongside a new operator for a designated training period of time. Training time for new associates is significantly reduced because the system-directed speech commands instruct the operator where to go in a paperless environment which simplifies training greatly.

The companies interviewed for this report indicated that new operators require an average of 7 weeks to learn order picking to achieve the company-standard productivity rate. Following the implementation of speech technology, this training time requirements dropped to approximately 3 weeks.

Most WMS software applications have the ability to record real-time date / time stamps on each transaction conducted with a radio frequency data terminal or a voice-directed system. Companies with WMS software can easily review a new associate's progress by reviewing summary reports generated with report writers.

Reduced Employee Turnover

Companies that have deployed speech technology have witnessed an unexpected reduction in employee turnover as a side-benefit. Apparently, the younger generation of associates is more interested in their jobs because of the “cool” factor associated with the use of high technology equipment. The use of wireless computers enables associates to view themselves as being part of a more sophisticated or prestigious job function. Apparently, there seems to be a desire to remain in a high technology environment rather than crossing the street to another job opportunity paying slightly more in a less sophisticated environment.

Support for Multiple Languages in the Distribution Center

In many North American distribution centers, it is not uncommon to have 30 – 50%, or higher, direct labor force with a mother tongue other than English. In many of these distribution centers, executives find themselves in the position where they have difficulty communicating to their immediate associates. This can lead to difficulties training new associates who may be forced to deal with English language instructions or paperwork. Speech technology offers some benefit in this area by enabling operators to interface with the computer in their own language.

On this note, it is important to probe the capabilities of the supplier’s speech technology solution to determine if multiple languages can concurrently be supported in the same distribution center. In general, speaker-dependent solutions are strong at supporting multilingual operations because each operator has the ability to record their own voice template in the language of their choice.

Improved Ergonomics and Safety

Not to be underestimated, a key benefit of voice technology is the potential to enable paperless picking which contributes to improved ergonomics and safety. This is especially important in any environment where order pickers are working quickly with heavy merchandise and mobile equipment. The ability for the operator to perform their job without having to look at paper documents and handle them while picking is clearly going to improve ergonomics and safety. As well, operators have the ability to be more alert to their surroundings when traveling with mobile equipment rather than having their eyes focused on paperwork.

Warehouse Management Integration

Over the past several years, a number of strategic alliance partnerships have sprung up between the suppliers of warehouse management software solutions and the suppliers of speech recognition technology. Many of these partnerships are what the software industry would classify as “Barney” relationships (I love you, you love me, we’re a happy family...). Press releases are issued with the suppliers patting each other on the back on how great things will pass because of the new relationship, but behind the scenes there is little or no research and development investment to truly develop an integrated solution that eliminates the need for software customization. The thinking is that they’ll worry about that when the first customer is ready to bankroll the whole effort. As a customer, there is no problem buying into this type of situation, as long as you have deep pockets and are willing to bankroll the whole effort.

Having said this, there are a couple of software firms that have made a serious and impressive effort to integrate voice directed solutions into their warehouse management system functionality. To determine if your warehouse management software vendor has developed any level of integration, a little homework is highly recommended. A proven installation may or may not be easy to find because the software providers have only recently begun to focus on this integrating speech technology into their solutions. In the author’s experience, here are a couple of points to remember:

- In most situations, speech-directed technology is relatively quick and easy to integrate into any WMS solution that already supports radio frequency (RF) functional tasks in the distribution center.
- The real issue is can the WMS software solution take full advantage of the fact that speech technology is being utilized? For example, does the introduction of speech-directed technology truly enable real-time information uploads with each pick transaction? If so, can the timing of replenishments to the pick slots be synchronized in a timelier manner now that real-time picking is enabled?

With voice-directed picking, a number of important side-benefits can be realized given that the WMS solution is in fact capable of real-time picking input/output dialogues to the server (as opposed to batch uploads).

- Situations where inventory discrepancies occur such that a pick short is invoked can immediately be dealt with. The picker will indicate the shortage situation through the headset and the WMS can invoke (1) a prioritized replenishment to refill the pick location (2) a new pick task if a secondary pick location exists or to the original pick location once the replenishment has occurred and (3) a system-directed cycle count to the location where the discrepancy occurred to enable inventory control personnel to research the issue.
- If the WMS is equipped with true industrial engineered labor standards with XYZ coordinates, the WMS can predict the timing of when an operator should arrive at a pick location once a new order is initiated. As such, user-defined settings can be set to enable pick location replenishments to be invoked X minutes in advance of when the replenishment will be needed. In high volume distribution centers, this can truly help to alleviate the situation where replenishments occur well in advance of when they are needed thus resulting in product being staged in operating aisles and the associated aisle congestion, product damage, etc.

As a final note on this topic, one easy way to identify if the WMS software supplier has gone beyond the “Barney” stages with speech-directed technology is to study the software’s maintenance screens. For example, inquire where the switch settings are to indicate that a function is voice-directed rather than RF or paper-directed. Inquire whether other customers have implemented voice technology with the WMS software already and ensure that you speak to any installed accounts to determine if this was a one-off custom installment. Be sure to test-drive the software and remember not to bank on functionality that will appear in a future release. The future is a very big place so supplier promises for new features and functionality may or may not ever materialize as priorities tend to shift the moment a new customer account is signed.

The Cost of Speech Technology Solutions

Today, the cost of speech technology remains the main hurdle as the barrier to entry for most companies. The capital investment required to deploy speech technology is slowly dropping but remains at the price point where many firms have difficulty accepting the turnkey solution cost.

For accurate budget numbers, it is important to work with your suppliers of choice to obtain requests for proposals. The following represents a very high-level estimate at the budgetary requirements for a speech technology solution (all prices quoted in \$US):

- Cost for a wireless computer with headset, 2 batteries, charger and client software license fee is about \$5,000 - \$6,000 per unit. The price per unit typically drops as a function of the quantity of terminals being purchased. The quantity of terminals to be purchased is based on the maximum number of concurrent users on a shift as their mobile terminals can be shared across multiple shifts. For budgeting purposes, spare units should be purchased as lead times on new equipment can be up to six weeks.
- Middleware and software interface to the ERP or WMS – Assume a conservative investment of \$30,000.
- 2.4 GHz RF Network – this depends on the size of the building. Assume a fixed cost for site survey, installation, project management, etc. at \$20,000 plus approximately \$3,000 per every 40,000 square feet for access points, wiring, etc. To be safe, obtain an estimate from a supplier of RF equipment.
- Assume \$30,000 – 100,000 for training, implementation and software modifications as required.

For example, a company operates a 600,000 square feet distribution center with a maximum of 75 concurrent order pickers per shift. A high-level budget figure would be:

- Wireless terminals: $75 \times \$6,000 = \$450,000$.
- Software / Interface: \$30,000.
- RF Network: $\$20,000 + 600,000/40,000 * \$3,000 = \$65,000$.
- Training / Implementation / custom enhancements: \$30,000 - \$100,000.
- Subtotal: \$575,000 - \$645,000.
- Add 10% contingency: \$58,000 - \$65,000
- Total Budget: **\$633,000 - \$710,000.**

Clearly, it is dangerous to generalize and the suppliers of this equipment will object to this type of exercise, but the intent is to demonstrate to the newcomer how to approximate the ballpark budget requirement for this type of solution. Honing in on the exact figures can be done if there is subsequent interest.

Profile of the Suppliers of Speech technology Solutions for Distribution Operations

There are three suppliers for the North American speech technology market for industrial applications: Syvox, Voxware and Vocollect. A brief supplier overview follows.

SyVox

SyVox Corporation
Worldwide Headquarters
2545 Central Avenue
Boulder, Colorado, 80301, USA
Telephone: (303) 938-1110
www.syvox.com

Additional Offices: Brugge, Belgium, The Netherlands and Les Mureaux, France.

Marketing pitch: SyVox Corporation is a global leader in developing and marketing complete, speech-based software solutions that increase the productivity of enterprises that employ mobile workers.

SyVox solutions for warehousing and logistics are branded as SpeechNet™, which consist of the portable mobile speech data terminal and the SpeechNet™ Engine. SyVox solutions are based on the speaker independent technology approach.

With SpeechNet™, associates are trained to speak a predefined lexicon of words such as “Got It” to confirm transactions. Thus training time required to use the system is reduced because the system does not need to understand the speaker’s specific voice template. SyVox claims that their solution is superior as the system allows for non-obtrusive feedback such that changes in a person’s voice (e.g. due to a common cold) will not impact the system’s ability to understand the speech.

The SpeechNet™ solution is also packaged with software logistics modules that manage functional warehouse activities including picking, putaway, cycle counting and cross docking. For companies that have weak functional support for warehouse management, this software will enable voice technology to be used with additional flexibility.

SyVox customer accounts include Reed Boardall, Nautica, P & O Cold Logistics, Atlas Cold Storage, and Nabisco. Founded in 1981 as Speech Systems, Inc., the company changed its name to SyVox in 1998 at the same time that it merged/acquired two small European businesses. In April, 2000, Psion Teklogix, a leading radio frequency / mobile computing hardware provider, invested \$10 million to acquire 20% of SyVox. In October 2000, Psion Teklogix expanded their shareholding to 19% through an additional investment of \$4.7 million. The investment strategy was to develop and market RF-based speech-

powered solutions for organizations deploying the SAP® Logistics Execution System (SAP LES).

SyVox has successfully established numerous partnerships and reseller agreements to increase its market awareness and to increase its mind share. Through a license agreement with Philips Speech Processing (PSP), SyVox is dedicating resources to supporting the major European languages. By adding the PSP European language technology, SyVox is anticipating making further inroads into Europe and capturing global market share.

SyVox was the only supplier of speech technology solutions that did not participate in the research for this report.

Vocollect

Vocollect
Worldwide Headquarters
701 Rodi Road
Pittsburgh, Pennsylvania, 15235, USA
Telephone: (412) 829-8145
www.vocollect.com

Vocollect is a privately held company that has been developing and installing voice-directed systems since the company was founded in 1987. Vocollect claim to be the world leader in providing voice-directed distribution systems with nearly 10,000 Talkman® units deployed in North America, Europe and Pacific Asia. In 2001, Vocollect's revenue grew by 30% as it shipped a record 3,000 units for the year.

Vocollect manufactures all of its Talkman® units and software technology at its global headquarters in Pittsburgh where over 100 people are employed. Talkman® units weigh less than one pound and are sealed units which is important in any environment that has cold temperatures or moisture. According to Vocollect, Talkman® units are also designed with no external antennas for improved industrial durability.

Vocollect solutions are speaker-dependent and the company offers users the choice to implement speech-to-text or digitized voice technology. According to Vocollect, the majority of customers implement the speech-to-text technology option because it is flexible and significantly faster as regards the time required for the computer to articulate commands to operators. The synthetic computer voice can be adjusted for speed and for pitch such that it can emulate a male or female voice.

The company also markets a software product called QuickPick™ which is a system designed to voice-enable any warehouse picking operation regardless of its existing warehouse computer support capabilities. QuickPick™ creates a real-time picking solution for companies that may or may not run with a warehouse management system (i.e. the software essentially manages the picking process).

Vocollect customers include: Wal-Mart, Safeway, Kwik Trip, Maines Paper & Food, Giant Eagle, K-VA-T Food Stores, Wegmans, Kroger, Perishable Distributors of Iowa (PDI; a subsidiary of Hy-Vee), Fleming, SuperValu, Roundy's, Price Chopper, Ralph's, Market Day, Mid Mountain Food, Campbell Soup Company, Dreyer's Ice Cream, Timken Company, Robbins Lumber Company, Super Store Industries (SSI), Mitsubishi, Ford Motor Company, and Saturn, amongst others. Vocollect has also established a number of business partner relationships with several leading WMS providers.

In February 2001, Vocollect obtained a \$12.5 million round of financing from three venture capital firms. The company remains privately held and continues to demonstrate strong top-line revenue growth and increasing market share.

Voxware

Voxware, Inc.
Corporate Headquarters
PO Box 5363
Princeton, New Jersey, 08543-5363, USA
Telephone: (609) 514-4100
www.voxware.com

Additional Offices: Voxware's engineering and research and development are headquartered in Cambridge, MA.

Voxware has concentrated their efforts on providing voice-based solutions to the warehousing and distribution center since their acquisition of Verbex Voice Systems in 1999. Verbex was founded in 1984 and had a history of providing voice systems that were specifically engineered to operate in high-noise environments with a diverse speaker population performing mission critical operations. Voxware originally started as a company that developed speech compression software for, among other applications, the Internet telephony market (i.e. technology to enable Web sites to support speech). In 1999, the company sold its software to Lucent Technologies, acquired the assets of Verbex and turned its focus towards the warehousing and logistics market. Voxware, formerly Verbex, scored initial success in the package handling and postal industries to manage identification, sortation and labeling of parcels and mail.

Of the three solution providers reviewed in this report, Voxware is the only speech technology solution provider to IPO on the NASDAQ in 1996. Today, the company's shares are traded on the over-the-counter market. The most recent financial results for the company were published for the quarter ended March 31, 2002. Voxware reported revenues of \$1.5 million for the quarter ended March 31, 2002, a 408% growth over the same quarter in 2001. The firm is anticipating significant top-line revenue growth and a return to profitability in fiscal 2002. Today, the company employs about 40 people, the majority of which are based in Cambridge, MA.

Voxware's VoiceLogistics™ solution was introduced in April 2000 and includes hardware, application software, speech recognition technology as well as design, development and on-going support. According to the company, there is a wide range of VoiceLogistics™ applications available including replenishment, cycle counting, loading and what the company calls its VoxView™ Management Console.

Voxware solutions utilize speaker-dependent, client-based recognition and the company offers users the choice to implement text-to-speech or digitized voice technology. According to Voxware, one of their competitive advantages is that their technology is developed using the World Wide Web Consortium standard voice XML or VXML protocols. Their applications run on a Web server which serves the pages, containing the application, dialogues and data, to the individual users on the floor of the warehouse. The Management Console, which provides tools for configuration, management and reporting, is browser-based and is deployable on a corporate intranet thereby enabling visibility and enterprise-wide reporting of operational data to any user with a Web browser and secure access to the management console application. All of the dialogues and event-specific logic are expressed in the software logic which enables flexibility and simplicity as regards customizing the behavior of the software.

According to Voxware, the fact that the company is a single source provider for all hardware and software technology solutions enables clients to get problem resolution through a single supplier. Voxware technologies are the only devices approved by the U.S. military to function according to specification at 120 dB noise levels. In fact, Voxware technologies are deployed for use in several important military aerial units including command and control for U.S. military helicopters and airplanes.

Voxware has a diverse customer base that includes grocery, foodservice, apparel, office supplies, consumer package goods, convenience store, and postal sorting operations. According to Voxware, this is testimony to the fact all dialogues, applications and process flows are easily tailored to the needs of the business.

Voxware customer accounts include U.S. Foodservice, PYA Monarch, Henry's Foods, Freedman Group LLP, 7-Eleven U.S.A., Somerfield PLC, Bofrost, the United States Postal Service, FedEx Ground, Canada Post, UPS, Hagggar Clothing, Siegwark Druckfarber GmbH, Migros, and Corporate Express. There are over 5,000 Voxware and Verbex units in active use today. Voxware has also established a number of business partner relationships with several leading WMS providers.

Sources:

The author wishes to extend gratitude to several people whose contribution has enabled this report to be researched and published including Ken Finkel, Dave Vetter, Jon Van Cleve, Mickey Johnson and Winston Chai.

An Overview of Kom International

Kom International provides innovative, unbiased and proven supply chain consulting services that enable companies around the world to manage logistics activities more effectively. Over the past 40 years, Kom International has worked with many of today's Fortune 500 companies across a wide spectrum of industry sectors to deliver world-class customer service levels at the least possible cost.

Kom International's profile:

- 40 years of experience consulting in warehouse, transportation, distribution and supply chain technology solutions
- Over 1,000 clients served
- 3,600 projects successfully completed
- One of the few consulting firms that provide 100% unbiased advice
- Markets served: North America, South America, Europe, and Asia
- Clients range from small family businesses to Fortune 500 firms

Kom International's core supply chain consulting competencies are in the following areas:

- Logistics strategy – optimization of distribution networks
- How supply chain policies affect the bottom line
- Evaluation of distribution operations to improve efficiency / service
- Distribution center design and layout
- Reset and optimization of existing distribution facilities
- Implementation services
- Supply chain technology solution selection and implementation
- Industry productivity benchmarking
- Inventory management to reduce cost of goods
- Transportation and fleet management

Kom International has a reputation for delivering logistics solutions that are practical, flexible and value-oriented. To better understand how Kom International can help your firm achieve world-class logistics, contact us at our worldwide headquarters and request a business assessment at no obligation or visit our Web site at www.komintl.com.

Supply Chain



**Corporate Headquarters
KOM International Inc.**

Place Du Parc,
Box 1113
300 Leo Pariseau, Suite 2300
Montreal, Quebec
Canada H2W 2P4
Tel.: (514) 849-4000
Fax: (514) 849-8888
www.komintl.com

Kom International's offices are located in several countries. Contact us or visit our Web site for a complete list of office locations.

All names, trademarks, products and services mentioned are registered or unregistered trademarks of their respective owners.
© 2002 KOM International, Inc.

043002
Printed in Canada.

Supply Chain