Knowledge You Need Before Purchasing Voice-Directed Applications for Your Warehouse

Introduction

Today's business applications are more complex and time-critical than ever before. Adaptability and multi-tasking have become industry buzzwords in recent years, and nowhere is this more necessary than in warehouse operations. The ability to efficiently identify, receive, store, retrieve and ship goods and materials is critical to the overall health of a company. Complicating this task is a shrinking labor pool of skilled workers and an influx of workers for whom English is not their first language.

Increasingly, companies are recognizing that voice systems offer significant benefits for a number of warehouse applications.

Voice systems recognize spoken words as data entry and provide audio (verbal) directions and confirmations to workers. For the non-technical operator, voice is the easiest and most transparent of all data entry and order picking technologies available. In fact, voice has been called the most natural of all data entry methods since it's one of the first skills we learn.

Voice System Overview

Voice systems leave a worker's hands free and eyes free to focus on the task at hand. Operators don't have to hold a bar code scanner or other type of reader and don't have to take their eyes off their work to read prompts on a screen. This single fact is an indication of the immediate time and productivity gains that can be realized with a well-designed voice system.

Voice System Applications

In the warehouse, voice systems are most appropriate in operations such as:

Picking
Put-Away
Shorts Processing
Replenishment
Returns Processing

For put-away, replenishment and picking operations, a shorthand code can easily be used on the product as well as the storage location. The voice-enabled terminal would direct the operator to a specific location -- for example, Row L5, Bin 15. An operator involved in picking would verbally confirm the location, possibly by reading a second code printed on the location label, and then be informed of the quantity and product code to pick. The operator would then speak the quantity and product code to confirm the pick. The system would verify the entry then move to the next task. For put away, the same basic scenario would play out with the voice prompt directing the put away and the worker verifying code and quantity.

For shorts and returns processing, again, voice is a natural data entry method. Particularly useful in returns processing where a variety of actions might be taken -- such as un-kitting, returning to vendor, repair, or restocking -- phrases rather than codes are the natural entry method. Finding a bar code symbol on a menu or locating the screen entry that indicates "damaged during shipping" takes far more time than simply saying it does – therefore they can focus on handling the return rather than on data entry.

In applications where there is a context to the data, such as address information or condition, voice excels. Voice systems are frequently used in manufacturing quality programs to record defects but, for the same reasons of

allowing eyes- and hands-free operation as well as natural input (a word instead of a code). Additionally, voice systems benefit any application where the item must be handled and evaluated to record its condition, disposition or destination.

Voice systems typically employ a "shorthand" code for a location or product code -- or even the actual product description -- when doing voice entry. Voice is not well suited for reading long strings of part numbers. It's as easy to make a "speako" as a "typo." Voice systems do offer verbal confirmation to help reduce errors, but it's more advantageous to use shorthand codes.

For applications where this kind of "shorthand" is not feasible, for example, where data entry requires a long purchase order number, shipping container code (SSCC-18), EPC code, or other long string of numbers or characters, the use of a mobile computer with a bar code reader or RFID reader may make more sense. Speaking long strings of letters or numbers may take more time than scanning them. The right set of tools, in this case voice and AIDC readers, leverages the appropriate technology to increase productivity and accuracy in different applications.

Operator Interaction

Voice systems require only a short learning curve and training regimen. Although voice systems are designed to recognize only specific words or phrases in addition to alphanumeric characters -- which might seem to be a chore to memorize -- voice prompts simplify a worker's routine by telling him or her exactly what to do and what input is required. The time required for training is mostly in training the voice system, not the worker; days of trainings now become hours.

Industrial voice systems are speaker-dependent. That is, they may recognize some words spoken by everyone but will only recognize every word spoken by one person at a time. Systems are therefore "trained" to recognize a specific operator's pronunciation of each word in the accepted vocabulary. Today's voice systems start with a basic recognition vocabulary and modify it for each individual. Training consists simply of repeating a list of letters, numbers and words several times.

Because voice is speaker dependent, the language spoken by the worker is irrelevant. Voice solution companies offer core vocabularies in a variety of common languages including English, French, Spanish, German and others. In one warehouse application, Vietnamese, Somolian, Bosnian, Spanish and English are all used by the voice system.

Although voice terminals are speaker dependent -- that is, they are designed to recognize a specific worker's voice input -- they do not have to be "reserved" for that individual. Personalized voice files are downloaded to each terminal as it's checked out at the beginning of each shift which means that any worker can use any terminal.

Voice Recognition Operation

Voice systems have been around for more than 20 years and have evolved from limited desktop devices to small engines that can be housed in a handheld or wearable computer connected via WiFi to the host. Voice recognition has achieved a sophistication that allows systems to accurately operate in noisy industrial environments and to ignore non-data entry speech. They are also able to distinguish problematic words like "five" and "nine" as well as distinguish all the sound-alike letters of the alphabet such as "b," "d" and "e."

Today's voice systems recognize continuous -- or natural -- speech. Whereas many "discrete speech" systems from only ten+ years ago would require an operator to pause briefly between words, for example, "product…damaged…in…shipping," that's no longer necessary. Continuous speech voice systems recognize the idiosyncrasies of natural speech patterns where two or more words may be pronounced almost as one. For example,

few people would say "*returned product*" distinctly. More likely it would sound like "*returnproduct*". Continuous speech voice systems would know what that meant.

Voice systems are not typically limited by their capacity to receive continuous speech data without a pause. In fact, most people would run out of breath before exceeding the ability of today's voice systems to recognize a continuous speech entry.

Voice systems can also offer distinct "grammars" for different applications. That is, the same word can mean different things depending on the job. "Left" might mean a direction or that something remains. The vocabulary of acceptable words can be designed to eliminate some of the confusion but the use of "grammars" to assign different significance to the same word -- or even homonyms -- based on the specific application helps ensures accuracy and simplifies creation of vocabularies.

Architectures

There is a trend toward open systems these days and voice systems have kept pace. Many systems use Voice XML (VXML) to create application packages and host filtering software.

The open system approach to voice reduces the total cost of ownership. This can be seen on many fronts. Open systems offer companies greater flexibility when it is time to add features or change the behavior of the system. Modifications can usually be done at a lower cost since VoiceXML is open to everyone, and there is a growing pool of programmers with VXML experience. Open systems also provide a level of assurance that support will not suddenly disappear which may happen with proprietary systems.

The history behind the 802.11 architecture gives a powerful message with open architectures: before 802.11, wireless networks were expensive and offered no real upgrade path. Today, an access point costs 80% less than it did 5 years ago and additional features are available that were not possible in the past. Furthermore, the new systems are still backward compatible with the first 802.11 devices. This is the power of an open architecture; VoiceXML architecture looks just as promising.

Implementing Voice Technology

Voice recognition systems, unlike bar code and RFID readers, do not take digital data as input. Input must be translated into a digital code recognized by the processing system, and then associated with a validation or action code in the host. Similarly, direction from the host has to be translated from digital to verbal form.

There are two aspects to voice system software: applications and host communication. In both cases, there are typically proprietary and open system solutions.

Voice Tools

Voice system vendors may offer a "toolkit" with pre-written applications for a variety of common tasks as well as the necessary tools for users to write their own applications. Pre-written applications are typically used as the basis for customization by the user and are a significant head start to completing the final product. In some cases, these "toolkits" are based on proprietary architectures.

There is a significant trend towards open systems which, in addition to lowering the customization and maintenance costs, can offer a more consistent look and feel in mobile and stationary terminals. Applications can also be written using open system tools for both socket-based and web-based connectivity.

The range of features available from different vendors may vary and, insofar as developing the application software will be the most time-consuming aspect of the implementation, care must be taken to evaluate the various offerings.

Connecting to a WMS

There are two methods of connecting voice terminals to host WMS programs: direct and via a gateway.

In the direct method, a company writes its own XML logic for voice terminals using various XML pages, each of which would be linked to a specific WMS screen or operation.

Using the gateway option, a company would use a vendor-supplied software package to establish host to terminal communication.

While both approaches achieve the same results, the difference is in up front programming time versus gateway package cost, operational maintenance and modification, and IT personnel learning curve.

As with voice tools, it is important to fully evaluate available options to make an informed cost/benefit decision.

Choosing a Voice-Enabled Device

Choosing the right voice-enabled mobile computing device requires an understanding of your 24 hour operations.

In some companies, picking and other operations that could benefit from voice are performed 24 hours a day (or for each full schedule of shifts, whatever that happens to be). For these applications, dedicated voice-enabled client devices are usually the most cost-effective solution. Dedicated voice client devices do not have capabilities or features that allow them to be used for any other type of data entry.

In most companies, however, many activities occur in blocks of time to coincide with inbound and outbound logistics schedules, company policies, employee availability, or other operational requirements. Having dedicated voice-enabled client devices that cannot be used in every shift is not a cost-effective approach. In these situations, it is often more economical to consider voice-enabled mobile computers that are also equipped for bar code scanning, key entry and offer graphical user interfaces and touch screens that can be used in both voice and non-voice directed applications. This would be a typical case where voice is used for picking during the first shift and standard terminal emulation is used for replenishment on the second or 3rd shift.

Even in companies that could employ dedicated voice client devices, predicting changes in volume and staffing may leave some devices unused during slack times or create a shortage during high volume periods. Neither scenario is an optimal one. It may be advantageous for these companies to consider a minimum number of dedicated voice client devices and complement them with voice-enabled multi-purpose mobile computers. This arrangement can help smooth out the demand for one particular type of data entry device or another.

The suitability of a particular device for your application depends not only on your current but your future operations. Dual use and multimodal mobile computers offer you the choice of operations and the flexibility to change with your operations.