



Project: Daddy Cool

DOCUMENT TYPE: Design
DATE: 15 October 2001
REFERENCE: P-117

FILE NAME:
PAGES INCLUSIVE OF HEADER: 47

Title: User Interface Design Specification

Version: 1.1

Status: Draft

Author(s): Adam Jordan
Scott Maddux
Tim Plowman
Victoria Stanbach
Jody Williams

Confidential See section 1.1 for details

History List

Version	Date	Author	Remarks
V0.1	23 July, 01	Adam Jordan, Jody Williams	Initial draft
V0.2	24 July, 01	All Authors	Revised initial draft as a group
V0.3	18 Aug, 01	Adam, Victoria, Jody	Filled in the flowcharts and much of the text.
V1.0	2 Sept, 01	Jody, et al	First complete, still rough, draft
V1.1	10 Oct, 01	Jody, et al	Incorporated comments on first draft, updated flowcharts, added command sequence diagrams

Trademark Information

AgileTV, Agile.TV, PLEX, AgilePLEX, AgileEngine, and AgileVoice are registered trademarks of AgileTV Corporation.

CONTENTS

1	INTRODUCTION.....	5
1.1	AUDIENCE	5
1.2	CONFIDENTIALITY	5
1.3	DOCUMENTATION CONVENTIONS	6
1.4	REFERENCES	6
1.5	GLOSSARY	6
1.5.1	Acronyms	6
1.5.2	Terms	6
2	SYSTEM OVERVIEW	8
3	INTERFACE DESIGN GOALS, PRINCIPLES, AND CONSTRAINTS.....	9
3.1	CONSIDERATIONS FOR TV INTERFACE GRAPHIC DESIGN	9
3.2	CONSIDERATIONS FOR TV INTERFACE BEHAVIOR	9
3.3	CONSIDERATIONS FOR SPEECH INTERFACE	9
4	MARKETING REQUIREMENTS	10
4.1	PRODUCT FEATURES	10
4.2	BRANDING AND POSITIONING	10
5	INTERFACE OVERVIEW	11
5.1	TELEVISION SCREEN	11
5.1.1	AgileTV Global User Interface.....	11
5.1.2	Navigation	13
5.1.3	Guidelines for Developing Speech-Enabled Applications.....	13
5.1.4	Agile User Center	14
5.2	SPEAKER ID AND PERSONALIZATION	15
5.3	SPEECH CONTROL	15
5.4	REMOTE CONTROL	15
5.5	VOICE LINK RECEIVER DISPLAY	15
6	TELEVISION SCREEN INTERFACE – FUNCTIONALITY AND FLOW.....	16
6.1	GLOBAL INTERFACE ELEMENTS	16
6.1.1	Immediate Speech Feedback	16
6.1.2	Help Overlay.....	28
6.1.3	Feedback Overlays	29
6.1.4	Main Menu Overlay	30
6.1.5	Speakable Text Icon	31
6.2	NAVIGATION	31
6.3	TELEVISION SCREEN INTERFACE TOOLKIT	31
6.3.1	Dialog boxes.....	31
6.3.2	Menus	31
7	TELEVISION SCREEN INTERFACE – GRAPHIC USER INTERFACE (GUI)	32
7.1	GRAPHIC USER INTERFACE TOOLKIT	32
7.1.1	Layout.....	32
7.1.2	Brand Indicator	32
7.1.3	Feedback Tab.....	33
7.1.4	Dialog Box.....	35
7.1.5	Text Box	36
7.1.6	Typeface	36
7.1.7	Background Imagery	37
7.1.8	Selection Highlight	37

7.1.9	Speakable Text Indicator	37
7.2	BUILDING SPECIFIC GSUI ELEMENTS	38
7.2.1	Immediate Speech Feedback	38
7.2.2	Help Overlay.....	38
7.2.3	Feedback Overlays	38
7.2.4	Main Menu Overlay.....	38
7.2.5	Speakable Text Indicator	39
8	SPEECH CONTROL – COMMANDS AND GUIDELINES.....	40
8.1	COMMAND ENVIRONMENT.....	40
8.2	GLOBAL COMMANDS	40
8.3	AGILETV GLOBAL SPEECH USER INTERFACE COMMANDS	40
8.4	NAVIGATION COMMANDS	40
8.5	COMMANDS FOR REMOTE CONTROL BUTTONS	41
8.6	APPLICATION-SPECIFIC COMMAND DEVELOPMENT GUIDELINES.....	41
8.7	AMBIGUITY RESOLUTION GUIDELINES.....	41
9	USAGE SCENARIOS	42
10	USER RESEARCH.....	43
10.1	MAIN MENU.....	43
10.2	GLOBAL INTERFACE ELEMENTS	43
10.2.1	Immediate Speech Feedback.....	43
10.2.2	Context Sensitive Help	43
10.2.3	General Navigation Issues	43
10.2.4	Context-sensitive Graphic Overlays	43
10.3	LATENCY	43
10.4	AGILE USER CENTER	44
10.5	VOICELINK REMOTE COLORS	44
10.6	VOICELINK REMOTE FORM.....	44
11	TO DO.....	45
12	ENGINEERING ISSUES AND DESIGN QUESTIONS	47

1 INTRODUCTION

This document is a design specification. It describes the global user interface to speech-enabled applications on the AgileEngine, and a set of guidelines for developing speech-enabled applications such as the Interactive Program Guide (IPG), Video on Demand (VOD), and Walled Garden.

The global speech user interface consists of an input system to receive speech, a feedback system to give cable subscribers information on the progress of their spoken requests, a set of visual cues on the television screen to help cable subscribers understand what they can say, a help system, and a model for navigation among applications. The interface is extensible to make it easy to add new applications.

This document describes the design with respect to cable subscribers. The implementation of this design is described in separate software and hardware design documents. See the references section for the complete document list.

The guidelines for developing speech-enabled applications are a means of providing a consistent user experience across applications. They are based on user research and our experience with adding speech to button-based applications.

The rest of this document describes the motivation and marketing requirements behind the design, the functionality and flow of the generic user interface, the look and feel of the components of the Agile system, and the guidelines for developing speech-enabled applications.

Note: User research will be done on every element of the design. This document will change based on the results of this research. In addition, the technical limitations imposed by the DCT2000 set-top box are not yet completely understood, and changes may be required to make the design work in this environment.

The major components that are still under construction are:

- The Help System – we may need more than one overlay
- Medium confidence recognition feedback
- Guidelines for developing speech-enabled applications
- The Agile User Center
- Speaker ID and Personalization
- The role of the agent (will we have an Agile AI, and how will he be used?)

Section 5 also gives a status for each of the television screen interface components.

1.1 Audience

The primary audience for this document is Engineering. It should be used as input to the technical specifications listed in the references section below.

1.2 Confidentiality

This document contains details of data received and discussions held under Non Disclosure Agreement (NDA).

This document, its contents, and findings must only be distributed to members of the project team and others covered by the NDA.

1.3 Documentation Conventions

Dates in this document are formatted as DD MMM YY, e.g. 12 Nov 00, or in full, e.g. 12th November 2000, to prevent confusion with day and month ordering in digit only dates. This same format should be used in all project documentation and communications.

1.4 References

1. Voice Link Receiver & Voice Link Remote Requirements Document
2. AgileTV Phase I Interactive Program Guide Speech Control Requirements Specification
3. AgileTV Phase I Interactive Program Guide TLD
4. AgileTV User Center Software Design Specification (This is not yet written)
5. AgileTV DIVA Speech Control Requirements
6. AgileTV Channel Surfer Application Specification
7. AgileTV User Center Design Specification (This is not yet written).

1.5 Glossary

This section describes the acronyms, abbreviations, and terms used in this document.

1.5.1 Acronyms

GSUI	Global Speech User Interface
GUI	Graphical User Interface
IPG	Interactive Program Guide
MSO	Multiple System Operator
PTT	Push To Talk
STB	Set Top Box
UI	User Interface
VOD	Video On Demand

1.5.2 Terms

<i>VoiceLink Receiver</i>	Compresses speech input obtained from the VoiceLink Remote and sends it to the Agile Engine via the Set Top Box.
<i>VoiceLink Remote</i>	The remote control used to obtain speech input from cable subscribers.
<i>Walled Garden</i>	The interactive applications provided by the MSO. Many of these will be Internet-based.
<i>Headend</i>	The control center for a cable TV system.
<i>Grammar</i>	A structured collection of words and phrases bound together by rules that define the set of all utterances that can be recognized by the speech engine at a given point in time.

Application Wrapper

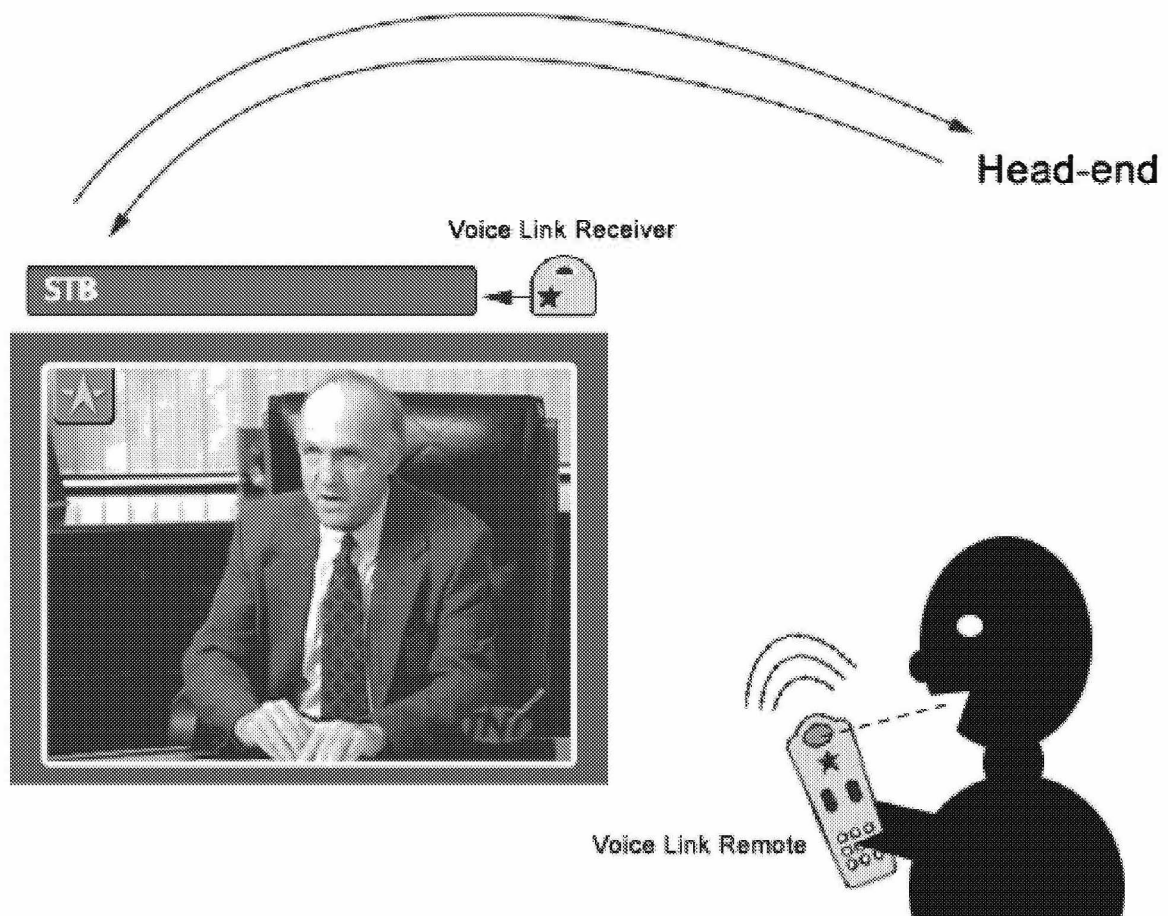
The component of an application that resides on the Agile Engine and determines whether an utterance has been recognized or not.

Application Server

The component of an application that carries out cable subscriber's requests. This component may or may not reside on the Agile Engine.

2 SYSTEM OVERVIEW

Speech control augments a subscriber's digital cable service. With digital cable service, the subscriber has a television connected to a digital set top box (STB) supplied by a local cable company. The augmentation for speech control requires additional hardware, a VoiceLink Remote control with a built-in microphone and a VoiceLink Receiver that connects to the set top box to send compressed voice back to the Headend via the STB.



3 INTERFACE DESIGN GOALS, PRINCIPLES, AND CONSTRAINTS

The design is based on a set of assumptions about the people who will use the system, how they want to interact with their television sets, and on the limitations imposed on the design by current technology. This section describes these assumptions and the design goals based on them. We will add to this list as we learn more about how cable subscribers interact with speech-enabled applications.

3.1 Considerations for TV Interface Graphic Design

- Large fonts & rich graphics (icons) are easier to read from 14 feet away, the average TV viewing distance.
- Brief amounts of, and polite, friendly text appeal to TV viewers.
- Dark background colors with light colored fonts have more visual impact on TV.
- Large, intuitive buttons are more noticeable on TV as well as remote & speech friendly.
- Scrolling is bad on TV, better to use paging mechanisms.

3.2 Considerations for TV Interface Behavior

- Should be simple & easy to use.
- Limit forced choices.
- Make forced choices obvious, in the forefront.
- Limit the covering of screen content.
- If covering the screen, limit the length of time and provide the cable subscriber with a way to remove.
- If providing “help” information, provide in a manner that is unobtrusive to screen viewing.
- Provide consistent visual reference to speakable text.

3.3 Considerations for Speech Interface

- Do not expect cable subscriber to remember speech commands.
- If possible, show speech commands on screen; otherwise, allow simple mechanism for cable subscriber to see what commands are available.
- Speech commands should be as short and direct as possible.
- Speech commands should be consistent across the entire Agile user interface.

4 MARKETING REQUIREMENTS

4.1 Product Features

This section is under construction.

- Initial versions of the product will be English-only.
- Future versions will support multiple languages.
- Speech is the ultimate navigation solution, and any products we do should reflect that.
- Speech-enabled applications should be easier to use than button-enabled applications.
- It should be easy for cable subscribers to discover what they can say.
- The system should be easy for novice users but not burdensome for advanced users.

4.2 Branding and Positioning

The branding objective for the Global Speech User Interface is to be flexible. The brand is simply applied to the surface of the core architecture and logic. Our ultimate goal is to brand the AgileTV voice control product with our own unique name to hold value with the cable subscribers. The Global Speech UI is flexible enough to reflect any brand. This enables us to use an MSO brand or our own brand.

There are three main uses of branding in the AgileTV product:

1. The logo on the Instant Speech Feedback graphic tab.
2. The logo on the dialog boxes and user center.
3. The logo on the remote, whether on the PPT button or on the body of the remote.

Consistent use of colors and typefaces on screen and on the remote and receiver also reinforce the branding of the system. A secondary branding element, the speakable text indicator, appears next to words that may be spoken. This indicator could change to reflect a logo or a color to reinforce branding. The same colored dot or logo on the remote could also be tied in to the PTT button.

Version 3.5 of our prototype software uses AgileTV branding elements: the AgileTV star, the AgileTV range of blue colors, and the AgileTV background abstract imagery. We do this solely because these branding elements are available, tie into our demo version IPG & VOD and reflect our corporate brand and name. Our goals are to develop a new name and brand that communicate that our product is speech control of the TV, that are easy to remember, and easy to pronounce.

5 INTERFACE OVERVIEW

This section contains high-level descriptions of the global speech UI and the related hardware and software components. The overview covers the graphical user interface on the cable subscriber's television screen, the Voice Link Remote control, the Voice Link Receiver, the AgileTV User Center, speaker ID and personalization, and the approach we're taking to speech-enable applications.

5.1 Television Screen

This section gives an overview of the Global Speech User Interface and the voice-enabled applications that will be the first to use it. The interface is implemented as overlays that appear on the television screen. All of the interface elements are mandatory for our beta release. We are still doing user tests on the GSUI to be sure it works well for cable subscribers. Some of the elements are working well and some need more work. A status is included where appropriate to indicate the completeness of the design.

5.1.1 AgileTV Global User Interface

The AgileTV Global User Interface is the onscreen interface for the AgileTV system. The purpose of the Global UI is to support the use of speech as a means of controlling digital TV content. The interface provides a feedback system to give cable subscribers information on the progress of their spoken requests, a set of visual cues on the television screen to help cable subscribers understand what they can say, a help system, and a model for navigation among applications.

The functionality and visual design of the Agile Global UI is consistent across all speech-enabled applications and services. The visual design may include the use of an agent as an assistant to introduce concepts, to guide the cable subscriber through the functionality of the system, and to provide a style for the Agile UI.

Specific content in the Global UI may be context-sensitive and customized to the particular application or service.

The elements of the AgileTV Global UI are:

- Immediate speech feedback
- Help Overlay
- Feedback Overlays
- Main Menu Overlay

This section describes the role of the agent and then gives an overview of each of the elements of the AgileTV Global UI.

5.1.1.1 The Role of the Agent

Status: under construction

The role of the agent is still under discussion, but there are some general rules about how the agent might be used.

1. The agent only appears associated with the Agile UI.
2. The agent does not appear integrated with 3rd party applications.
3. The agent is associated with positive feedback, not negative feedback.

4. The technical constraints of the DCT2000 restrict both the functionality and the visual appearance of the agent. There will be limited or no transparency, limited or no translucency, limited or no animation, and limited detail, limited resolution, and limited color depth. These technical constraints imply that the agent will have a visual impact on the design rather than a functional one.

The tasks we might use an agent for are:

1. Instructional
 - a. Out Of Box (OOB) equip. set-up
 - b. Video (on-line, VHS, barker channel etc.)
 - c. On-screen Help
 - d. Agile User Center
2. To enhance fun, e.g. the cable subscriber says “whassup” and the agent responds with one of a dozen replies.

5.1.1.2 Immediate Speech Feedback

Status: close to final

Immediate Speech Feedback provides simple, graphic, and quickly understood feedback to the cable subscriber about the basic operation of the speech interface. It tells the cable subscriber when the Agile system is listening or processing an utterance, and whether or not the utterance has been understood.

5.1.1.3 Help Overlay

Status: under construction

The Help Overlay provides a list of context-sensitive speech commands for the current speech-enabled application and is accessible at all times. The Help Overlay also provides brief instructions about what onscreen text is speakable and links to More Help and the Main Menu Overlay.

5.1.1.4 Feedback Overlays

Status: misrecognition overlays near final, other recognition feedback overlays under construction

There are 2 types of Feedback Overlays – Recognition Feedback Overlays and Application Feedback Overlays. Recognition Feedback Overlays inform the cable subscriber that there has been a problem with recognition. The type of feedback that is given to the cable subscriber includes generic “I don’t understand” messages, lists of possible recognition matches, and more detailed help for improving recognition. Application Feedback Overlays inform the cable subscriber about errors or problems with the application that are not related to unsuccessful recognition.

5.1.1.5 Main Menu Overlay

Status: close to final

The Main Menu Overlay provides the list of speech-enabled digital cable services that are available to the cable subscriber. The Main Menu Overlay is meant to be faster and less intrusive than switching to the MSO's full-screen list of services.

5.1.1.6 Speakable text Icon

Status: design is close to final, implementation is difficult (see Issue 4 in section 13)

The speakable text icon shows which on-screen text is "speakable". The icon is placed next to all on-screen text that the cable subscriber may speak to issue a particular command.

5.1.2 Navigation

Status: some preliminary work done

The navigation framework defines the grammar, functionality, and flow for switching from one application to another and for revisiting previous screens. Cable subscribers do not want a rigid hierarchy. They want to be able to get to any of their favorite screens in a single step. This is known as spherical navigation. Spherical navigation is impractical because there are limits to the number of items a speech recognition system can handle at once. We are working on ways to give cable subscribers more navigation options without degrading recognition accuracy. We are also trying to understand how application boundaries affect the way people think about the previous screens they've seen.

5.1.3 Guidelines for Developing Speech-Enabled Applications

Status: under construction

This section currently contains a brief description of the first applications we are adding a speech interface to and an outline of the speech functionality to be added. It will be expanded soon to include a list of guidelines that apply to all applications.

5.1.3.1 Approaches to Typical Interface Elements

5.1.3.1.1 Lists

5.1.3.1.2 Buttons

5.1.3.1.3 Scroll Bars

5.1.3.1.4 Dialog Boxes

5.1.3.1.5 Radio Buttons

5.1.3.1.6 Pull-Down Menus

5.1.3.2 Approaches to Specific Applications

5.1.3.2.1 Interactive Program Guide

The Interactive Program Guide (IPG) is the application cable subscribers use to find out what's on television. One of the problems with button-enabled program guides is that navigation is awkward. Cable subscribers frequently need to press the page up/down buttons repeatedly until

they find the program they want. Speech provides a streamlined interface where many common functions can be done with a single voice command.

The speech interface for the IPG will allow the use of speech commands to control all IPG functionality. This includes:

1. Selecting on-screen "buttons".
2. Directly accessing any program or channel in the current time slot.
3. Performing every function that can be executed with remote control key presses.

5.1.3.2.2 VOD

Video on Demand is the electronic version of the video store. Cable subscribers pay a fee for each video they want to watch, and they have access to the video for several days. While they have access, they can start the video any time, watch it as many times as they'd like, and use VCR-like controls to fast forward and rewind. One of the problems with button-enabled video on demand systems is that navigation is awkward. Cable subscribers frequently need to press the page up/down buttons repeatedly until they find the movie they want. Speech provides a streamlined interface where many common functions can be done with a single voice command.

The speech interface for the VOD will allow the use of speech commands to control all VOD functionality. This includes:

1. Selecting on-screen "buttons".
2. Directly accessing any movie title in a particular list.
3. Performing every function that can be executed with remote control key presses.

5.1.3.2.3 Walled Garden / Internet

This section is a placeholder for future work on web browser-based applications. Version 3.5 of the prototype software demonstrates one way of organizing a set of these applications.

5.1.3.2.4 Channel Surfer Application

One of the problems with current digital cable television systems is that cable subscribers need to switch to the program guide to find out what's on and then switch back to watch the program. There are some shortcuts, but finding programs and then switching to them still requires many button presses. AgileTV is developing an application that allows cable subscribers to get one-step access to programs they want to watch without ever switching away from the TV screen. For more information, see the *AgileTV Channel Surfer Application Specification*.

5.1.4 Agile User Center

Status: some preliminary work done

The AgileTV User Center is an application that provides:

1. Training and tutorials on how to use the AgileTV system.
2. More help with specific speech-enabled applications.
3. User account management.
4. Settings and preferences for the AgileTV system.

See the *AgileTV User Center Design Specification* for more information.

5.2 Speaker ID and Personalization

Speaker ID is a technology that allows the speech recognition system to identify a particular user from his or her spoken utterances. In order for the system to identify the user, the user must train the system with approximately 45 seconds of speech. When the system is fully trained, it can identify that particular speaker out of approximately 10,000 other speakers. In the near term, Speaker ID will be used to improve recognition accuracy. In the long term, Speaker ID can allow the AgileTV service to show a custom interface and personalized ITV content for a particular trained speaker. Speaker ID can also allow simple and immediate parental control - an utterance itself, rather than a PIN, can be used to verify access to blocked content.

5.3 Speech Control

Status: under construction

Speech control is accomplished through a set of commands that define what a cable subscriber can say. The commands for the global speech user interface are described in detail in section 8.

5.4 Remote control

Status: close to final

The AgileTV remote control combines universal remote control functionality with a microphone and a Talk button. The AgileTV remote is fully compatible with the DCT-2000 (all of the standard DCT-2000 remote buttons are present) allowing the AgileTV remote to replace the cable subscriber's existing STB remote. Additionally, speech commands are transmitted from the remote to the AgileTV system when the cable subscriber presses the Talk button and speaks into the microphone. The requirements for the remote control are documented in the Voice Link Receiver & Voice Link Remote Requirements Document

5.5 Voice Link Receiver Display

Status: close to final

The VoiceLink receiver connects to the STB, receives speech commands, via infrared (IR), from the AgileTV remote control, and sends those commands to the cable headend. The receiver has an LED display on its front panel consisting of 7 LED's: one to indicate power, and six to indicate the signal strength of the sound being received by the remote. Additionally, the first segment of the signal strength display lights when the Talk button on the remote is pressed. The requirements for Voice Link receiver are documented in the Voice Link Receiver & Voice Link Remote Requirements Document

6 TELEVISION SCREEN INTERFACE – FUNCTIONALITY AND FLOW

This section gives a detailed description of the functionality of the television screen interface components developed by AgileTV. This includes the Global Speech User Interface, the navigation model for speech-enabled applications, and a set of television screen interface elements that can be used as a toolkit to build application-specific menus and dialog boxes.

6.1 Global Interface Elements

This section gives a detailed description of the functionality and flow of the Global Speech User Interface.

Version 3.5 of the prototype software implements the functionality described below, and should be considered the definitive source of answers to any questions not covered in the text and flowcharts in this section.

6.1.1 Immediate Speech Feedback

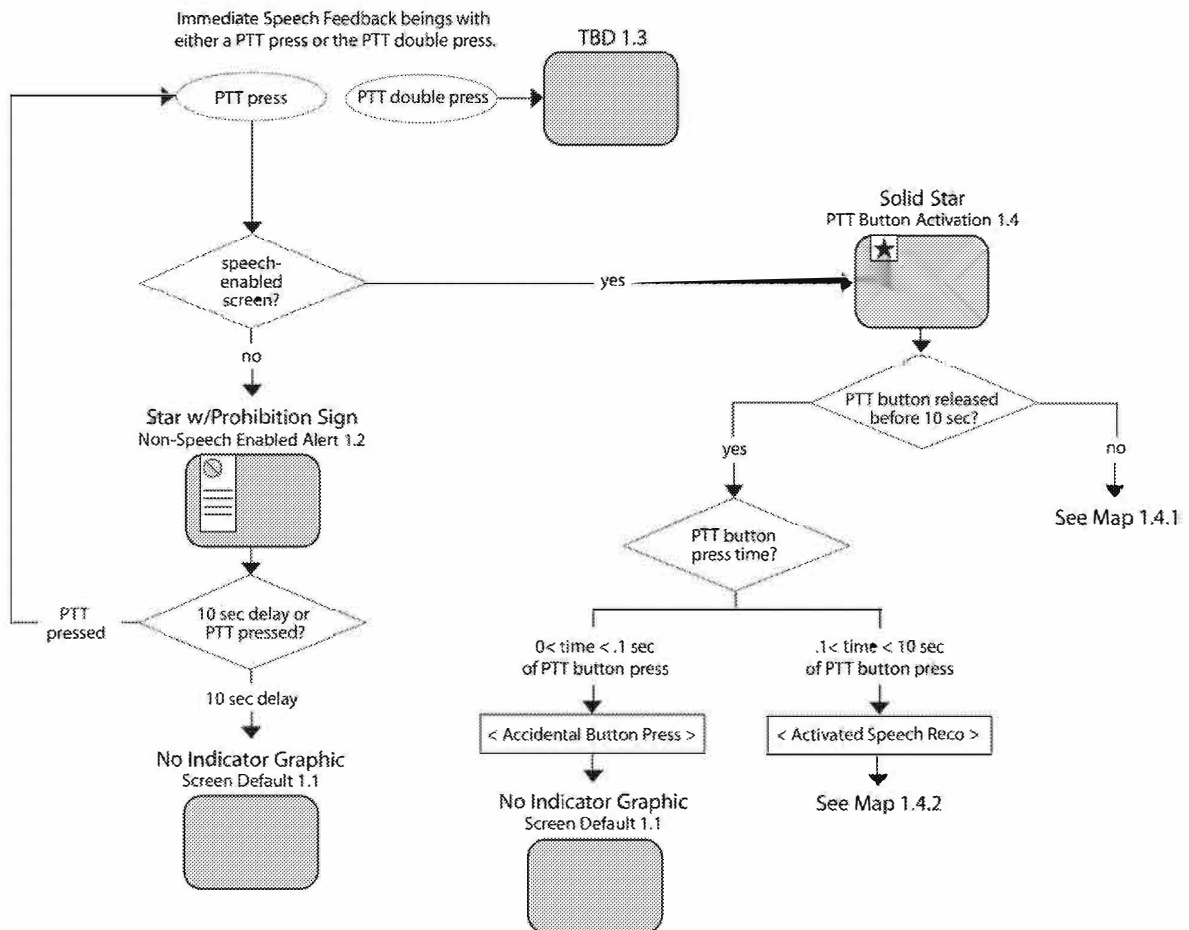
Immediate Speech Feedback provides simple, graphic, quickly understood feedback to the cable subscriber about the basic operation of the speech interface. This subtle, non-textual feedback gives necessary information without being distracting.

Immediate Speech feedback displays the following states:

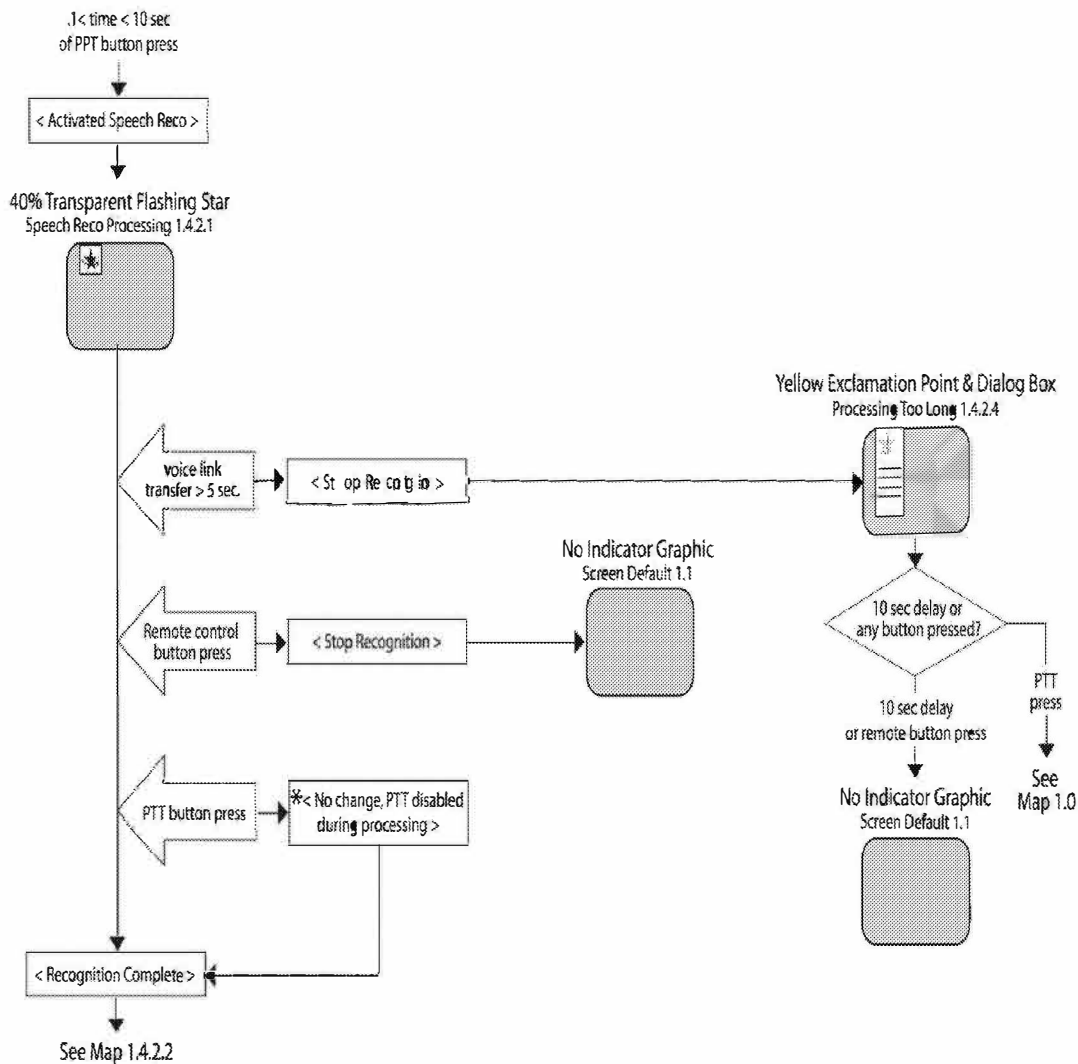
- Talk button pressed down – the system has detected that the talk button on the remote has been pressed and is “listening” to the cable subscriber
- The system is processing an utterance (covering the duration between the release of the Talk button and the resulting action of the Agile system)
- Successful recognition has occurred and the system is executing an action
- Unsuccessful recognition has occurred (after the 1st try, the Recognition Feedback Overlay is also displayed)
- Recognition was stopped. A Feedback Overlay is also displayed to show why this occurred.

These states are shown in the following set of four flowcharts.

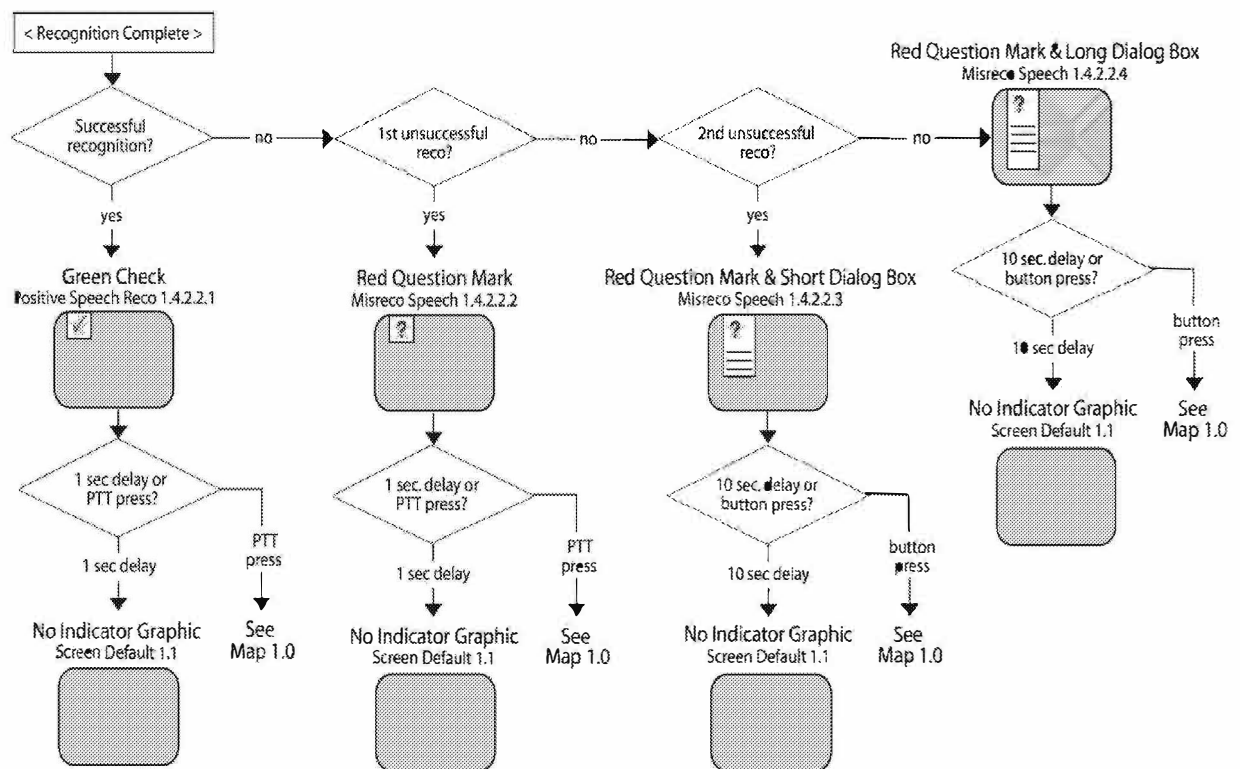
Note that remote control buttons are disabled while the PTT button is pressed, and that once the system has started processing a voice command, the PTT button is disabled until the cable subscriber receives notification that the recognition was successful, unsuccessful, or stopped.



Note that the preceding flowchart includes a double press of the talk button. The action to be taken when this is detected is still TBD. A double press has occurred when there is 400 ms or less between the “key-up” of a primary press and the “key down” of a secondary press.

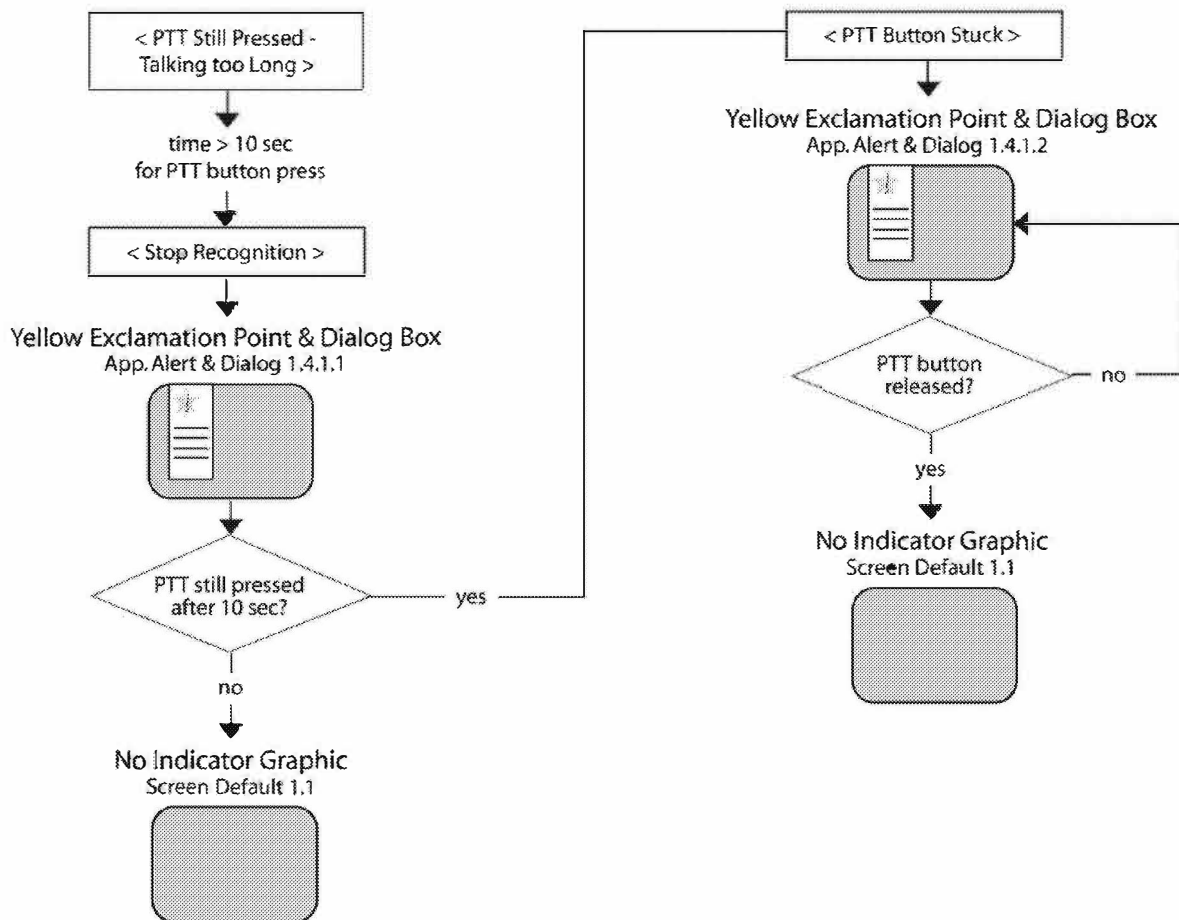


When there is no congestion, there should rarely be a need for the cable subscriber to press a remote control button while a speech command is being processed. When there is congestion, however, the cable subscriber should be able to use the remote control buttons to improve response time. An extensive discussion of when cable subscribers can issue a second command while the first is still in progress and what happens when they do so is given in section 6.1.1.1, Command Sequencing.



The system keeps track of the number of unsuccessful recognitions in a row. This number is reset to zero after a successful recognition and when the cable subscriber presses any remote control button. If this number is not reset, the cable subscriber will continue to see the long recognition feedback message any time there is an unsuccessful recognition. If cable subscribers are having difficulty with the system, the long message is good, even when several hours have elapsed between unsuccessful recognitions.

The recognition feedback only stays on screen for one second, so it isn't necessary to remove it when a remote control button is pressed. When the PTT button is pressed the recognition feedback should be replaced by the speech activation icon (the solid star).



There are two different messages when the talk button is held down for a long interval. The first message covers the relatively normal case where the cable subscriber takes more than ten seconds to speak the command. The second covers the abnormal case where the PTT button is stuck. There is no transition between the two messages. The second message stays on screen until the button is released.

6.1.1.1 Command Sequencing

There is ordinarily a small, expected, latency between a command and its execution, so a cable subscriber issues a command and waits for the results. Speech commands introduce longer latencies because speech requires more bandwidth to the headend, and it can be affected by network congestion. In addition, some applications are implemented outside of the Agile Engine. In these cases, recognition is done on the Agile Engine and the command is then sent on to the application server.

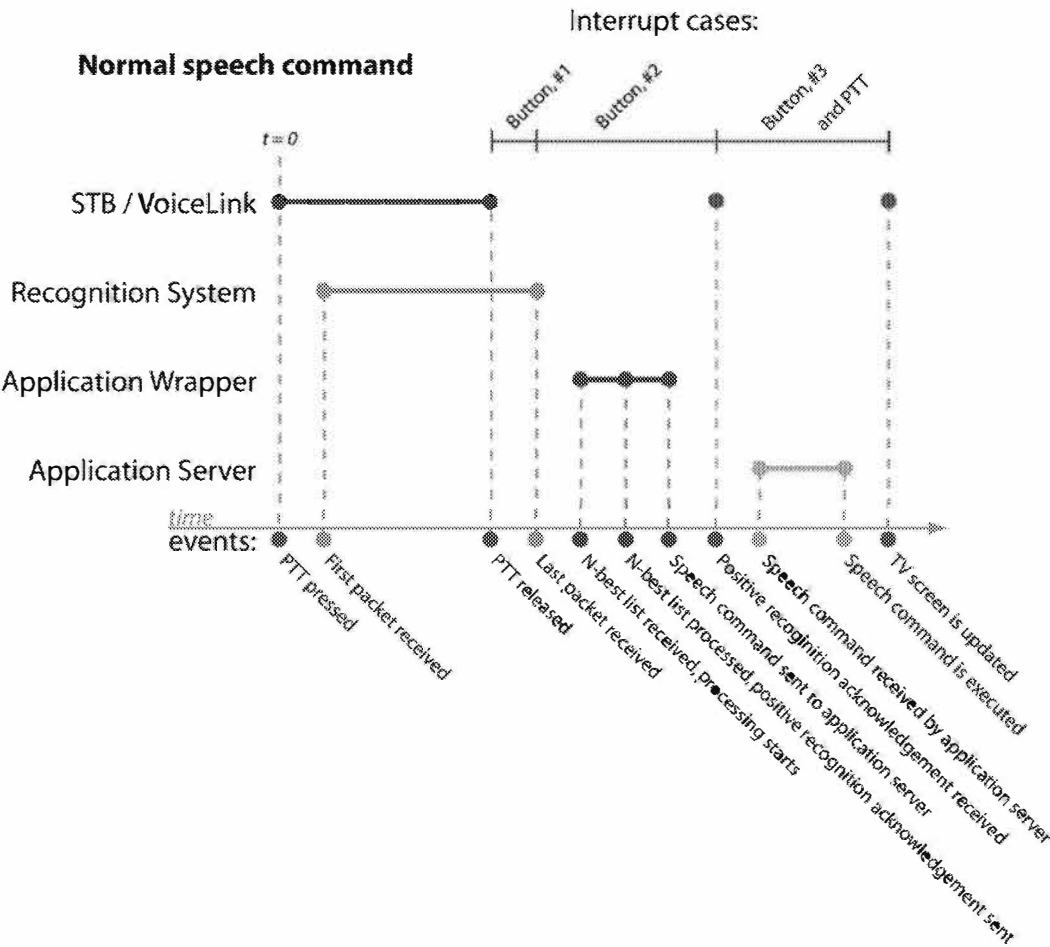
Applications on the Agile Engine and those outside of the Agile Engine should look the same to cable subscribers. In particular, it is highly desirable for the recognition feedback for a speech command and the results of the execution to appear on the television screen at the same time. However, if there is likely to be latency in communicating with an off-Engine application server

or in the execution of the command, the recognition feedback should appear as soon as it's available.

When there is congestion and speech commands are taking a long time to process, the cable subscriber will try to use the buttons on the remote control or to issue another speech command. The sequence diagrams below describe what happens when the cable subscriber attempts to issue another command. There are race conditions in the underlying system, so we can't specify exactly what happens, we can only give guidelines. The guidelines support two general goals:

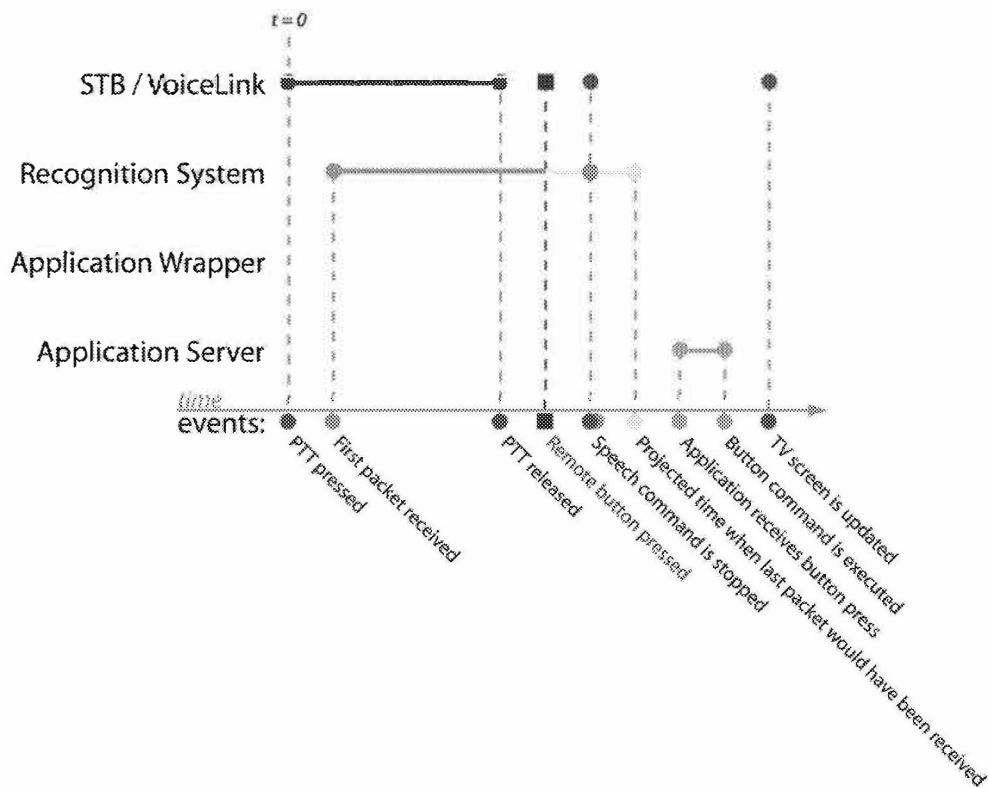
1. The cable subscriber should be in control. If a command is taking too long, the cable subscriber should be able to issue another command. In the sequence diagrams, when a cable subscriber presses a remote control button while a speech command is being processed, the speech command is preempted, where possible, to give control back to the cable subscriber. A detailed description of where preemption is possible and which part of the system is responsible for it accompanies the sequence diagrams.
2. The system should be as consistent as we can make it. To accomplish this, we need to minimize the race conditions in the underlying system. We can do this in two ways:
 - a. We can prevent the cable subscriber from issuing a second voice command until the STB receives an indication of whether the recognition for the first command was successful or not. This makes it highly probable that the application has received the command and is executing it by the time the subscriber sees the recognition feedback. If the command still takes a long time to execute, there are two explanations, either there is a network problem between the Agile Engine and the application server executing the command, or the latency is in the application, not the speech recognition system. Network problems can be handled via the command sequencing described in b). Applications where there can be long latencies should already have built-in mechanisms to deal with multiple requests at a time. For example, it can take a long time to retrieve a web page, and the web browser is prepared to discard the first request when a second request arrives.
 - b. We can require applications to sequence the execution of commands as follows. If the cable subscriber issues commands in the order speech command (A), followed by button command (B), and the application receives them in the order A, B, both commands are executed. If the application receives them in the order B, A, command B is executed, and when command A arrives, it is discarded because it is obsolete.

The sequence diagrams show the points in time where a second command may be issued and describe what should happen when the second command is issued.



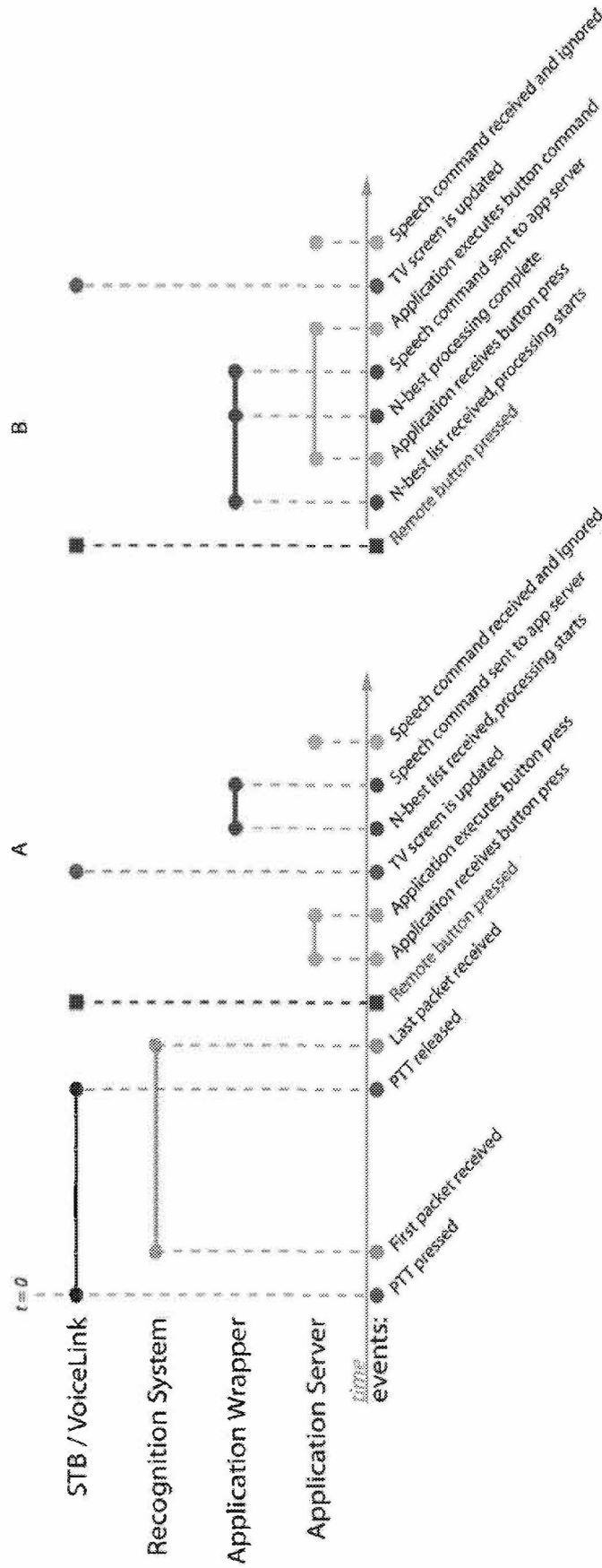
This diagram shows the timeline of a normal speech command. The round dots are events. A bar that spans events indicates activity. For example, the bar between PTT pressed and PTT released indicates that the PTT button is depressed and speech packets are being generated. The labels on the left side of the diagram indicate the components in the system. The application wrapper and the application server are listed as separate components. When the entire application resides on the Agile Engine, the wrapper and the server will be the same component, and command sequencing will be easier. A dot on the same horizontal line as the name of the component means that the event occurred in this component. The labels on the bottom of the diagram describe the events that have occurred. The events are ordered by the time they occurred. There are four cases where a button or speech command can be issued while another command is already in progress. These are shown under the label “Interrupt cases:” at the top right of the diagram. The rest of the diagrams will describe what happens in each of these cases.

Speech command interrupted by button input, case #1



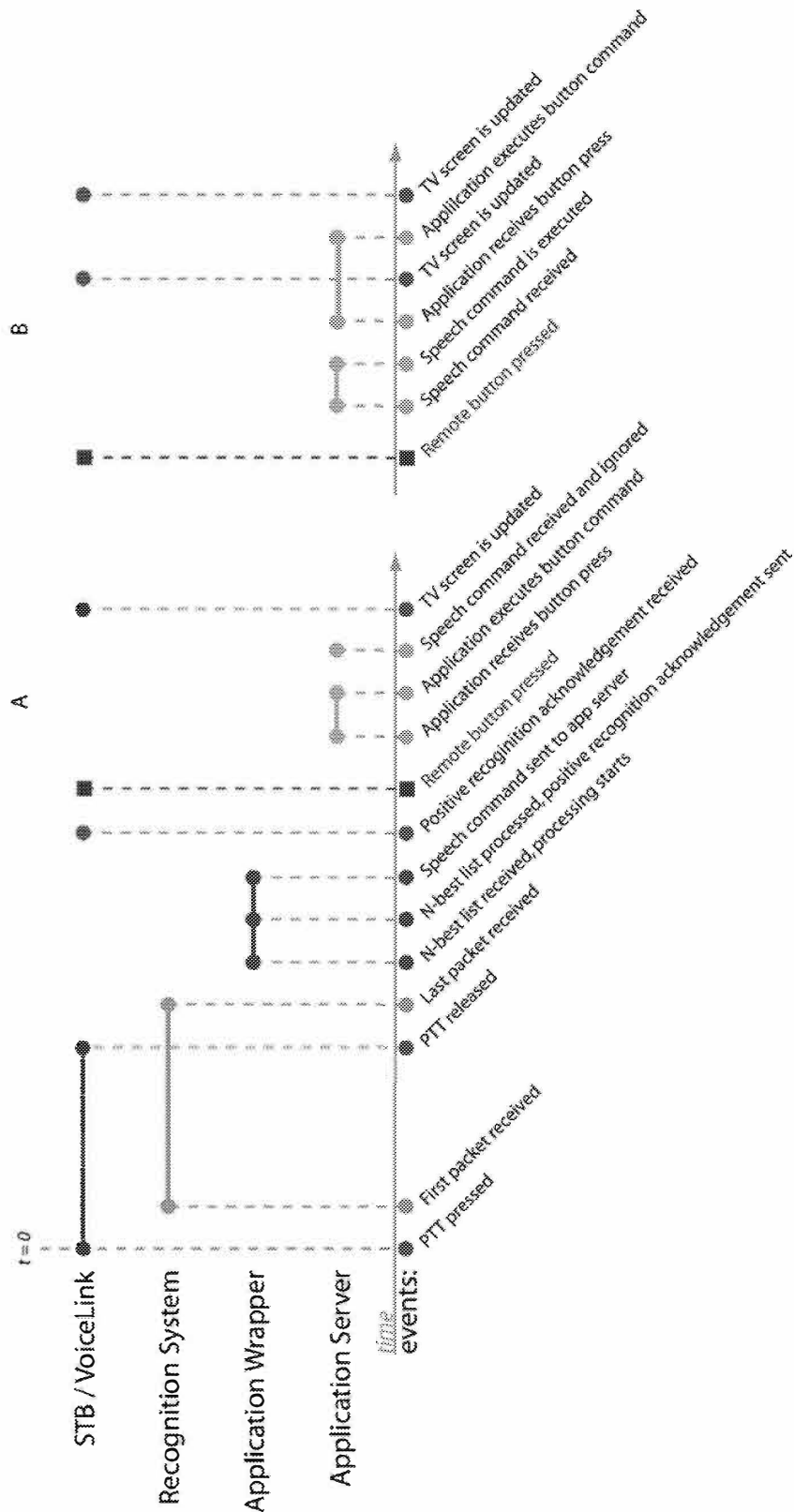
In case 1, the cable subscriber pushed a remote control button before the STB / Voice Link sent all of the packets for the speech command to the Recognition System. The diagram shows that the speech command is cancelled and the remote control button command is executed. The STB / Voice Link and the Recognition System should cooperate to cancel the speech command.

Speech command interrupted by button input, case #2



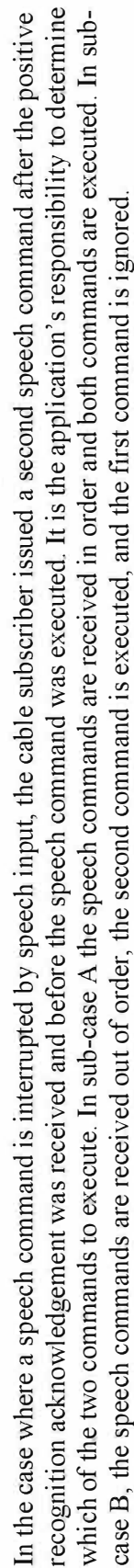
In case 2 the cable subscriber presses a remote control button after the last packet is received by the recognition system and before the n-best list is processed by the application wrapper. In both cases, the speech command is discarded and the button command is executed. This diagram shows that the STB / Voice Link and the Recognition system could have cooperated to cancel the speech command in sub-case A, and the application would not have had to be involved. In sub-case B, the application cancels the speech command because it arrived out of sequence.

Speech command interrupted by button input, case #3



In case 3 the cable subscriber pressed a remote control button after the positive recognition acknowledgement was received and before the speech command was executed. It is the application's responsibility to determine which of the two commands to execute. In sub-case A the speech command is received out of sequence, and it is ignored. In sub-case B, the speech command is received in order, and both the speech command and the remote control button command are executed.

Page 27 of 47



6.1.2 Help Overlay

The Help Overlay displays a short, context-sensitive list of frequently used speech commands for each unique screen of every speech-enabled application. The Help Overlay is meant to accomplish two goals: 1) provide hints to new users to allow them to control basic functionality of a particular speech-enabled application; and 2) provide a reminder of basic commands to experienced users in case they forget those commands.

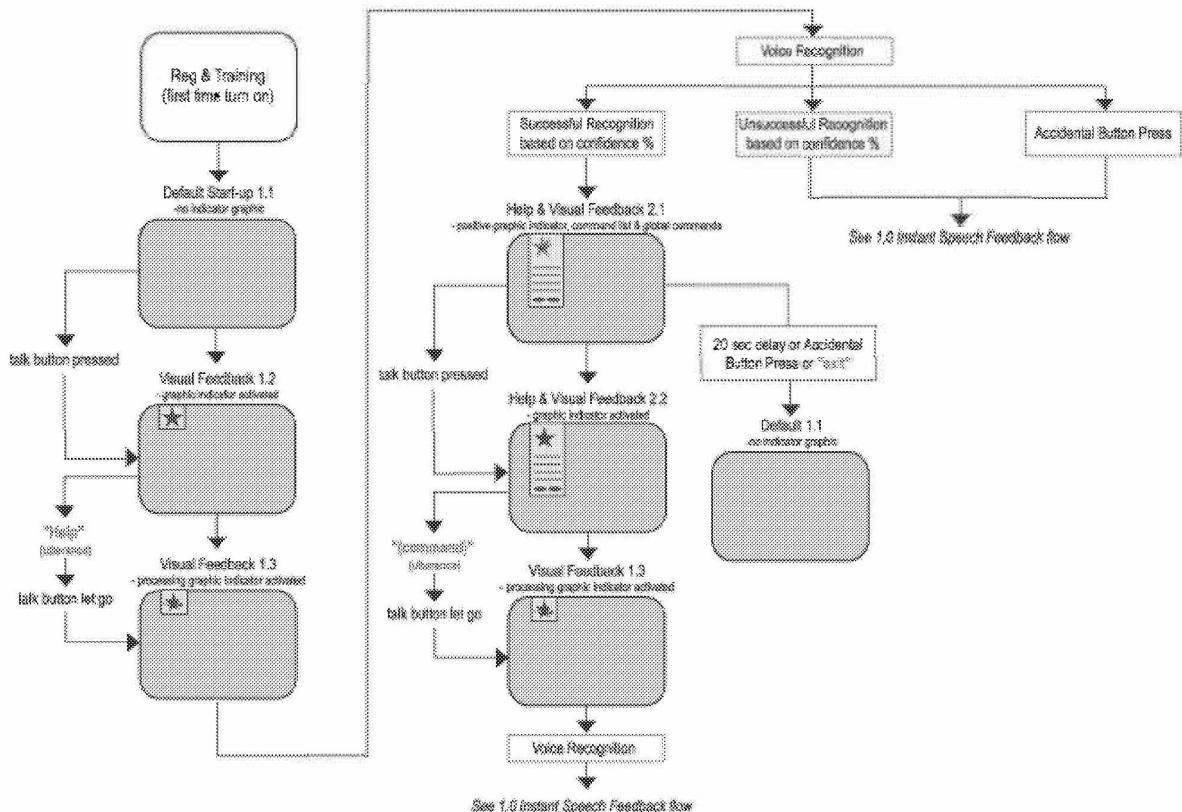
In addition to displaying application-specific commands, the Help Overlay always shows the commands for accessing the Main Menu Overlay and “more help” in the AgileTV User Center. Also, the Help Overlay will explain the Speakable text indicator, if it is activated.

Note that the Help Overlay only shows help for the Agile speech interface, not for application functionality.

The Help Overlay is organized as follows:

- Application-specific commands (approx. 5 basic commands)
- “More help” command (link to AgileTV User Center)
- “Main Menu” command to display Main Menu Overlay
- “Exit” to make overlay disappear

The logic for when the help overlay appears and disappears is shown in the following flowchart:



6.1.3 Feedback Overlays

There are 2 types of Feedback Overlays: Recognition Feedback Overlays and Application Feedback Overlays. Recognition Feedback Overlays inform the cable subscriber that there has been a problem with recognition. Application Feedback Overlays inform the cable subscriber about errors or problems with the application that are not related to unsuccessful recognition.

Recognition Feedback Overlays exist in three states and respond to four different conditions.

The three different Recognition Feedback states correspond to unsuccessful recognitions which occur “in a row”; that is, unsuccessful recognitions which immediately follow one another. This behavior occurs when the cable subscriber tries multiple times to issue a command which is not recognized by the system; the three states offer progressively more feedback to the cable subscriber and links to help.

Recognition Feedback states:

- 1st unsuccessful recognition. The immediate speech feedback icon changes to a question mark. Provides minimal, quickly understand feedback to the cable subscriber.
- 2nd unsuccessful recognition. The Feedback Overlay is displayed with message and link to Help Overlay.
- 3rd (and nth) unsuccessful recognition. The Feedback Overlay is displayed with message and links to Help Overlay and more help in AgileTV User Center.

The different Recognition Feedback conditions correspond to the amount of information the recognizer has about the cable subscriber’s utterance and to the latency in the underlying system:

- Low confidence score. A generic “I don’t understand” message is displayed.
- Medium confidence score. TBD.
- Sound level of utterance too low. The “Speak more loudly or hold the remote closer” message is displayed.
- Sound level of utterance too high. The “Speak more softly or hold the remote farther away” message is displayed.
- Talking too long. There is a ten second limit to the amount of time the PTT button may be depressed. If the time limit is exceeded, the utterance is discarded and the “Talking too long” message is displayed.
- PTT button stuck. If the PTT button has been depressed for 20 seconds, the “PTT button stuck” message is displayed.
- Processing too long. If the VoiceLink and Set Top Box are unable to transfer an utterance to the head end within 5 seconds after the PTT button is released, the “Processing too long” message is displayed.

Application Feedback Overlays are displayed when application-specific information needs to be communicated to the cable subscriber. A different icon at the top of the overlay differentiates Application Feedback from Recognition Feedback.

A description of when a feedback overlay appears and the text it should contain is shown in the following spreadsheet:

Feedback spreadsheet goes here.

6.1.4 Main Menu Overlay

The Main Menu Overlay provides the list of speech-enabled digital cable services that are available to the cable subscriber. The Main Menu Overlay is meant to be faster and less intrusive than switching to the MSO's full-screen list of services.

Services displayed:

- Watch TV – full screen TV viewing
- Program Guide
- Video on Demand
- Walled Garden / Internet
- AgileTV User Center

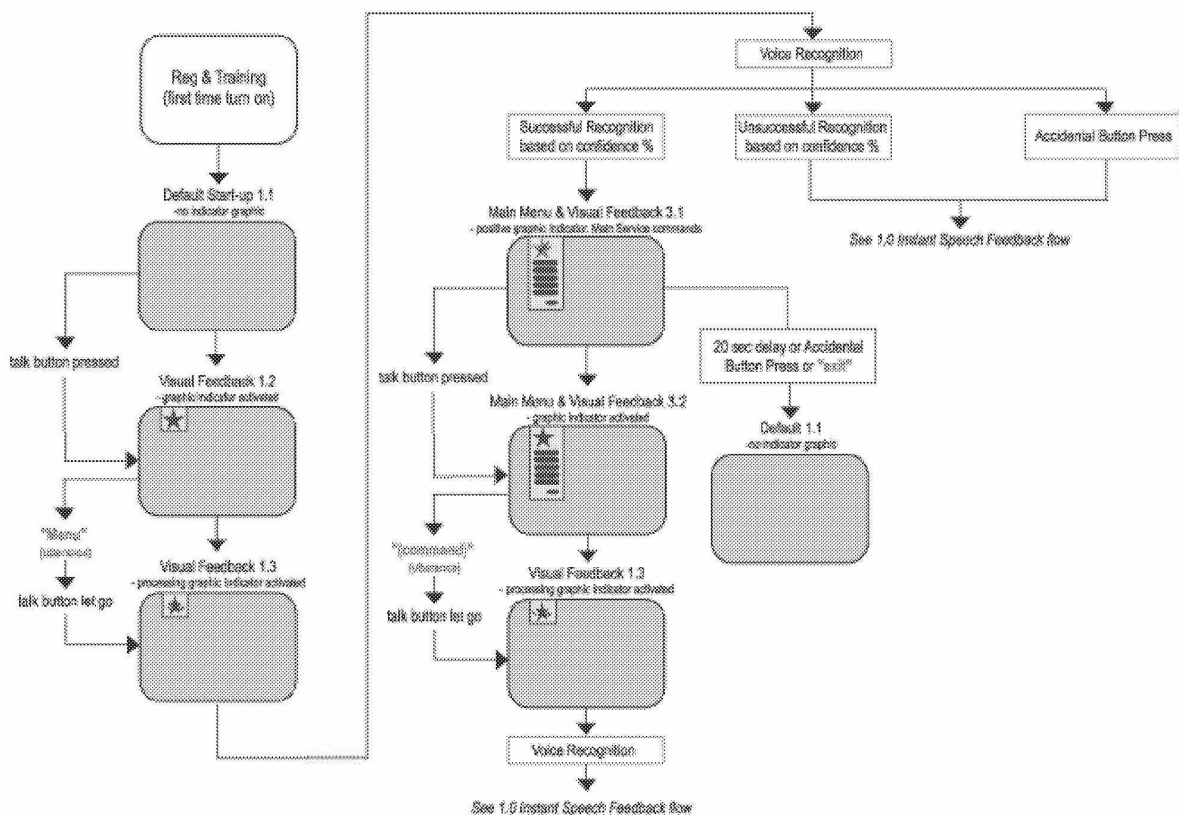
Additional commands displayed:

- “Exit” to make overlay disappear

Functionality:

- Current service is highlighted

The logic for when the menu appears and disappears is shown in the following flowchart:



6.1.5 Speakable Text Icon

The Speakable Text Icon appears to be layered above speech-enabled applications as a part of the Global Speech UI. It is placed next to all on-screen text that the cable subscriber may speak to issue a particular command. This text may be static or dynamic. Static text is used in labels for on-screen graphics or “buttons” that may be selected by moving a highlight with the directional keys on the remote control. As such, most screens will usually have several text-labeled buttons and therefore will require a corresponding number of speakable text icons. Dynamic text is used in content such as the list of movies for the Video on Demand (VOD) application. Each line of dynamic text may include speakable text icons to indicate which words are speakable.

The Speakable Text Icon is currently a green dot, and may be changed to a different icon. It is important that the icon be visible but not distracting.

Additionally, the cable subscriber should have the ability to turn the Speakable Text Icons on and off.

6.2 Navigation

Details still TBD

6.3 Television Screen Interface Toolkit

The Global Speech User Interface describes functionality that is required in every speech-enabled application. It does not include application-specific functionality. The television screen interface toolkit describes the way application-specific functionality fits into the global framework.

6.3.1 Dialog boxes

This section gives speech UI guidelines for getting application-specific information, such as setup options, from cable subscribers.

Details are still TBD.

6.3.2 Menus

This section gives guidelines for application-specific speech UI menus and describes how these menus fit into the global framework.

Details are still TBD.

7 TELEVISION SCREEN INTERFACE – GRAPHIC USER INTERFACE (GUI)

This section describes the graphics for version 3.5 of the prototype. The graphics for the final version of the interface are still TBD. The current graphics are intended to give Engineering an idea of the intent of each of the visual indicators and to show how branding will tie the interface together. The Graphic User Interface is also extremely flexible. Consider the concepts of the toolkit design elements to be the real building blocks for the final design not the art itself.

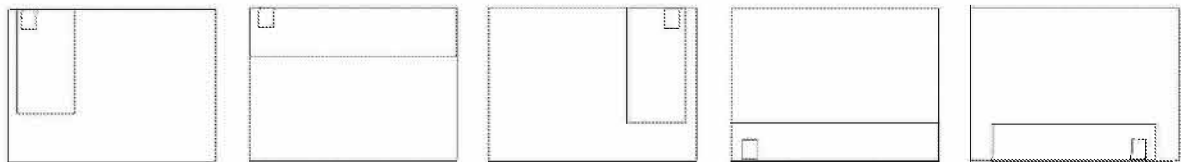
Once we decide on the final graphics, complete style guides will be added to this document.

7.1 Graphic User Interface Toolkit

The GSUI overlays described in section 6 are created from a set of toolkit elements. This section describes the toolkit elements. The way the GSUI overlays are created from the toolkit is described in the section 7.2.

7.1.1 Layout

The MSO has some flexibility to specify where the GSUI should appear. The GSUI is anchored by the immediate speech feedback tab, which should appear along one of the edges of the screen. The anchor point and the size and shape of the dialog boxes may be different for each MSO. Here are some possible layouts:

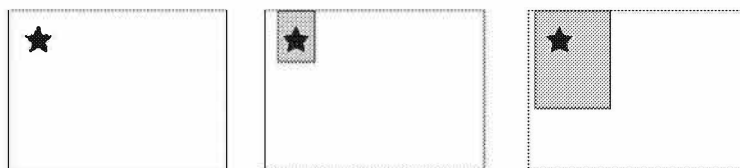


7.1.2 Brand Indicator

Brand Indicator Concept:

The brand identity of the AgileTV system may appear alone or in conjunction with the MSO brand identity. Whenever the brand identity appears, it should be consistent in location, size and color treatment.

The static placement of the brand indicator is key in reinforcing that the GSUI feedback is coming from the AgileTV product. Various states of color and animation on the brand indicator are used to indicate system functionality.



On Application

Feedback Tab

Dialog Box

As exemplified in Proto 3.5:

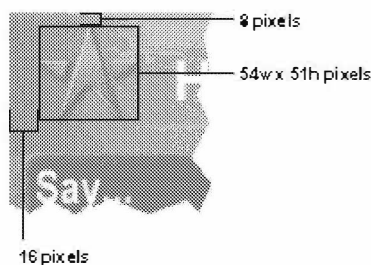
The AgileTV star is used consistently throughout the GSUI as a placeholder for the branding element we will use in the product. Screens containing the brand indicator contain information relative to speech recognition. The rest of this section describes how the AgileTV star should be used as a brand indicator. Similar rules will apply for the final brand indicator.

The AgileTV star has various states of transparency and color to provide visual clues to the state or outcome of a speech request.

There are three main AgileTV star states:

- *40% Transparency Star*
Used as a brand indication. Appears on all aspects of the Agile GSUI: Tab, Dialog Boxes and Full Screens.
- *Solid Star*
This is used to indicate that the remote's Push To Talk button is currently being pressed.
- *40% Transparency Flashing Star*
Used to indicate that the system heard what you said and is processing the information.

The AgileTV star should be positioned in the upper left corner of the screen and remain the same size throughout the GSUI. The official AgileTV CMYK (*Cyan-Magenta-Yellow-Black*) gradation star logo should be used.



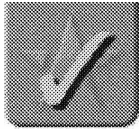
7.1.3 Feedback Tab

Feedback Tab Concept:

The Feedback Tab is the on-screen graphical element used to implement Immediate Speech Feedback as described in sections 5.1.1.2 and 6.1.1. The feedback tab uses a variety of graphics to indicate the status and outcome of a speech request.

As exemplified in Proto 3.5:

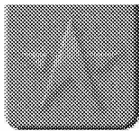
- *Positive Speech Recognition Feedback – green check*



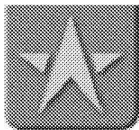
- *Misrecognition Speech Feedback – red question mark*



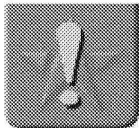
- *Speech Recognition Processing – 40% Transparency Flashing Star*



- *Push to Talk Button Activation – Solid Star*



- *Application Alert – Yellow Exclamation Point*



- *Non-Speech Enabled Alert – Star with Prohibition Sign*



Some design rules about the Tab are:

- The color should always be R: 54, G: 152, B: 217.
- It should always have a transparent background.
- It should always be aligned to the top of the TV screen.

- The size should always be 72w x 67h pixels.
- The brand indicator, in this case the AgileTV star, should always be present.
- The bottom corners should be rounded.
- The star and graphic indicators should be centered in the tab.

7.1.4 Dialog Box

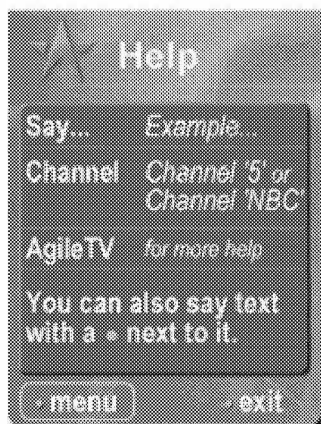
Dialog Box Concept:

The dialog box is used to implement the Feedback Overlay, Help Overlay, Main Menu Overlay, and Command List Overlay described in sections 5 and 6. The dialog box is a bounded simple shape. It may contain a text box to convey information associated with the AgileTV product. It may also contain virtual buttons that can be selected either by voice or by the buttons on the remote control. Different dialog boxes may use different sets of virtual buttons. When two different dialog boxes use a virtual button, it should appear in the same order relative to the rest of the buttons and have the same label in each dialog box.

As exemplified in Proto 3.5:

The dialog box is the transparent background box used to display textual information. The textual information is displayed in a text box, which is overlaid on the dialog box. Some design rules of this box are as follows:

- It should always flush align to the top of the TV screen.
- The bottom corners should be rounded.
- The AgileTV Background Imagery should always be present.
- The box height can fluctuate, but width should stay consistent.
- The box should always appear on the left side of the TV screen.



7.1.5 Text Box

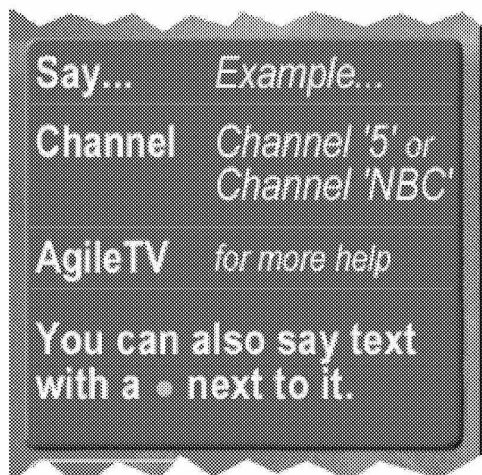
Text Box Concept:

A text box conveys information associated with the AgileTV product. This information should stand out from the background imagery. To accomplish this, the text box is a bounded shape placed within the bounded shape of a dialog box.

As exemplified in Proto 3.5:

Text in a text box is always presented on a solid colored blue box, which is then overlaid on the dialog box. There can be more than one text box per dialog box. For example, the Main Menu Overlay contains one text box for each item in the main menu. Secondary navigation, such as the “menu” and “exit” buttons, can be displayed outside the text box on the dialog box background imagery. Some design rules for the blue box are:

- The color should always be R: 42, G: 95, B: 170.
- The text box should always sit 8 pixels in from each side of the Dialog box.
- All corners should be rounded.
- All text within a text box should be flush left.



7.1.6 Typeface

Typeface Concept:

Use of a single font family with a combination of typefaces helps reinforce the brand identity of the AgileTV product. When different typefaces are used, each should be used for a specific purpose. This helps the cable subscriber gain familiarity with the user interface. Any typeface used should be legible on the TV screen.

As exemplified in Proto 3.5:

Arial Narrow Book Italic, Arial Narrow Bold and Arial Bold are used throughout the GSUI. All text is white except for titles on the dialog boxes, which are R: 204, G: 255, B: 255. The font size

can vary but should never be less than 20pt., kerning should be normal and leading can vary for readability.

Each typeface is used for a specific purpose:

- *Commands* – **Arial Narrow Bold**
- *Command examples and descriptions* – *Arial Narrow Book Italic*
- *Dialog Box Title* – **Arial Bold**

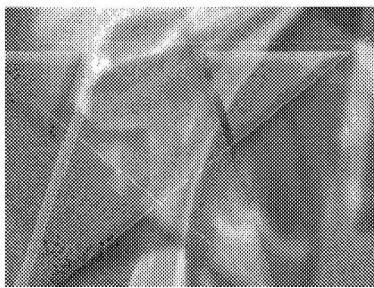
7.1.7 Background Imagery

Background Imagery Concept:

The background imagery should reinforce the branding of the AgileTV product.

As exemplified in Proto 3.5:

The consistent use of the AgileTV abstracted star background imagery helps brand and visually indicate that the information being displayed is part of the AgileTV speech recognition product.



7.1.8 Selection Highlight

Selection Highlight Concept:

The selection highlight is a standard graphical element used to highlight a selected item on-screen.

As exemplified in Proto 3.5:

The selection highlight is a 2 pixel, yellow rule used to outline text or a text box indicating that it is the currently selected item.



7.1.9 Speakable Text Indicator

Speakable Text Indicator Concept:

The speakable text indicator is described in detail in sections 5.1.1.6 and 6.1.5. It should always be the same color and shape. It should be placed next to any speakable text that appears on-screen.

As exemplified in Proto 3.5:

The placeholder for the speakable text indicator is a green dot. The green dot should be consistent in size and color throughout the GSUI and in speech-enabled applications. The only exception to this rule is that the green dot is larger in the help text about the green dot itself.

Some design rules about the green dot:

- The size should always be 6 x 6 pixels.
- The color should always be R: 0, G: 255, B: 0.



7.2 Building Specific GSUI Elements

7.2.1 Immediate Speech Feedback

The Feedback Tab is the graphic element used for immediate speech feedback. This element appears on top of any other GSUI overlay on screen. For example, if the Help Overlay is on screen, and the cable subscriber presses the push-to-talk button, the Push To Talk Button Activation Tab appears on top of the Help Overlay.

7.2.2 Help Overlay

The Help Overlay contains helpful information about the speech user interface and menu and exit buttons. The visual design of the Help Overlay is a dialog box that uses these graphical elements: Brand Indicator, Text Box, Background Imagery, Typeface and Menu Highlight, as well as a dialog box title indicating which service the Help is for. The content in the text box changes relative to the digital cable service being used. The Help Overlay should never change design layout but can increase or decrease in length according to text box needs.

7.2.3 Feedback Overlays

The Feedback Overlay is displayed upon misrecognition of voice commands. The visual design of the Feedback Overlay is a dialog box that uses the following graphical elements: Brand Indicator, Text Box, Background Imagery, Typeface and Menu Highlight, as well as a dialog box title indicating which service the feedback is for. The Feedback Overlay should never change design layout but can increase or decrease in length according to text box needs.

7.2.4 Main Menu Overlay

The Main Menu Overlay is a dialog box that contains a dialog box title, buttons with links to various digital cable services and an exit button. The Main Menu uses the following graphical elements: Dialog Box, Background Imagery, Typeface, Menu Highlight, and Text Box. Each selection on the Main Menu is a Text Box.

7.2.5 Speakable Text Indicator

The Speakable Text Indicator is a single graphical element. It is described in section 7.1.8.

8 SPEECH CONTROL – COMMANDS AND GUIDELINES

This section describes the environment in which a command will be processed, and then describes commands for the global speech user interface, guidelines for application-specific command development, and guidelines for how to resolve ambiguities that arise when commands appear in more than one context.

8.1 Command Environment

Each speech command is processed in a context that includes commands to access to any content named on the screen the cable subscriber is viewing, commands to access application features, commands to access the Global Speech User Interface (GSUI), commands to simulate remote control button presses, and commands to navigate to other applications.

Many of the guidelines in this section were developed to try to minimize the potential for words or phrases from one source to duplicate those from another. For example, the content in the Interactive Program Guide (IPG) application contains the names of television shows. There could easily be a television show named “Exit” which would conflict with using “exit” as the speech equivalent of pressing the exit button on the remote control.

The specification for a command describes the way it fits into the environment. The specification includes the command’s:

1. Scope. This characterizes when the command is available.
2. Language. This defines the words cable subscribers use to invoke the command.
3. Behavior. This specifies what happens when the command is invoked.

Pictures to show how these three characteristics are related are under construction.

8.2 Global Commands

Global commands are always available. Applications may only disable them to force the user to make one of a set of application-specific choices. However, this should be a rare occurrence. Speech interfaces should be designed to make the cable subscriber feel like he or she is in control.

Since AgileTV technology will be incorporated by MSOs over time, there will be installations with both speech-enabled and non-speech enabled applications. In these cases, it is highly desirable for the navigation commands to be speech-enabled and available globally. This allows cable subscribers to move from one application to another via voice. When all of the applications supported by an MSO are speech-enabled, both the navigation commands and the AgileTV GSUI commands should be global.

8.3 AgileTV Global Speech User Interface Commands

The AgileTV GSUI commands should always be available for speech-enabled applications. The language and behavior of these commands will be developed by AgileTV and specified in this document.

The table of AgileTV GSUI commands goes here.

8.4 Navigation Commands

The navigation commands should always be available. The navigation commands include specific commands to allow cable subscribers to go to each application supported by the MSO and general

commands that support the navigation model. For example, “Video On Demand” is a specific command that would take the cable subscriber to the VOD application, and “last” is a general command that would take the cable subscriber to the appropriate screen as defined by the navigation model.

The language for the navigation commands may be different for each MSO because each MSO supports a different set of applications. For early adopters, AgileTV will specify the language for the navigation commands by working closely with the MSO.

The navigation model determines the behavior of the navigation commands. There may be an opportunity to introduce a new navigation model, but AgileTV technology will need to support navigation models that already exist at the MSO. There may be an overall navigation model, and different navigation models for different applications. Where navigation models already exist, navigation is done via remote control buttons. The speech commands for navigation should be the same as pressing the corresponding remote control buttons. When a screen contains virtual buttons for navigation and the cable subscriber invokes the speech command corresponding to the virtual button, the virtual button should be highlighted and the command invoked.

Table to show first draft of navigation commands goes here.

8.5 Commands for Remote Control Buttons

The scope for remote control buttons varies widely. Some remote control buttons are rarely used in any application, for example, the “a”, “b”, and “c” buttons. Some are used in most applications, for example, the arrow keys. Since recognition can be improved by limiting choices, it would be ideal if each context only included speech commands for applicable remote control buttons.

The language for remote control buttons will be defined by AgileTV and specified in this document.

The behavior of the speech commands for remote control buttons should be the same as pressing the remote control buttons. However, when a screen contains virtual buttons that represent buttons on the remote control and the cable subscriber invokes the speech command corresponding to a virtual button, the virtual button should be highlighted and the command invoked.

The table defining the language for remote control buttons goes here.

8.6 Application-Specific Command Development Guidelines

This section is still under construction. It will contain a set of guidelines derived from user tests and from our experience in prototyping applications. The guidelines here now are placeholders.

Cable subscribers should rarely be forced to say one of the choices in a dialog box.

The global commands should always be available unless the cable subscriber is forced to say one of the choices in a dialog box. This should be a rare event.

People commonly say phrases such as *Show me* or *Go to* before they issue a command. Application-specific commands should include these phrases to make applications more comfortable to use.

8.7 Ambiguity Resolution Guidelines

Guidelines TBD

9 USAGE SCENARIOS

This section describes usage scenarios that are driving the design.

Our focus has been on the visual cues cable subscribers need to understand how the AgileTV system is capturing and processing voice commands. Usability studies have shown where people sit, how they sit, how they hold the remote control, what angle we expect them to use with the VoiceLink remote, how they use the buttons on the remote control, and where they keep it when it isn't in use. The studies have also tested our initial designs for the global speech user interface. See §12 for more information on these studies.

10 USER RESEARCH

This section contains the most definitive results from the user testing we've done through August 30, 2001. Many of the design choices we made were based on the results of these tests. More information is available on the www/ConApps website.

10.1 Main Menu

- “Menu” is expected to call up the main menu consistently
- Should have an exit/cancel button or stay until next command

10.2 Global Interface Elements

10.2.1 Immediate Speech Feedback

- Pulsing star progress is useful
- The question mark symbol (?) is clear and provides useful feedback
- Error messages are seen as useful and should probably give all options
- Users are confused as to whether the system is responsive while the error message is up

10.2.2 Context Sensitive Help

- IPG and Watch TV help gives unclear info, very few users understood the instructions
- IPG Help should color code examples differently from the instructional text

10.2.3 General Navigation Issues

- Spherical navigation is highly desired
- Users have a strong desire to scroll in many environments
- Names and categories should be as short, descriptive, and literal as possible
- Something like a site map might be useful
- Users want to say the fewest words possible (“Agile TV User Center” is long)
- Most users catch on to the green dots quickly and consistently
- Users expect to be able to say “forward” as well as “back”
- Saving games and returning to the program in between commercial breaks is desired

10.2.4 Context-sensitive Graphic Overlays

- Green dots are preferred to no green dots

10.3 Latency

- 4 to 5 seconds is pushing it but as long as these delays are interspersed with shorter delays, users don't seem to mind them. TV Guide has set the bar
- Progress indicator is beneficial and better than TV Guide pixelation

- Users are somewhat inured to latency from computer use
- Latency can to some extent be offset by rich content (i.e., it is worth the wait)

10.4 Agile User Center

- Names and descriptions are good if a bit lengthy
- White lettering on blue back is okay in this context

10.5 VoiceLink Remote Colors

- Users are conservative in their tastes – no one likes the colored remotes with the exception of surf. They prefer silver ice, blue ice, and gunmetal gray. These colors convey “entertainment tech” and also do not stand out too much
- The remotes selected tend to remind people of Audi, Passats, and BMWs
- Most remotes will live on the coffee table or the couch
- User preference for buttons Light lettering on a dark background
- User preference for housing dark lettering on a light background
- Colored remotes are identified as the ones most likely to go out of style

10.6 VoiceLink Remote Form

- Most users prefer discrete buttons for mode selection
- Most users will not miss the AUX button
- Button layout and spacing overall is good but users feel like there are a lot of buttons on the remote – shading may help reduce this undifferentiated mass
- Users recognize that they will have to work a little to learn the layout of new remote and are willing to do so.
- Navigation buttons must have directional indicators.
- The IR lens should go up as high as possible.
- PTT button should be a light color to facilitate finding in the dark

11 TO DO

Priority	Completion Date	Description
High		The help system still isn't working for users. We are working on our third attempt at this.
High		We need to determine what to do for medium confidence recognition. Should the resolution be application-specific, or should there be GSUI elements to handle this? Should we develop guidelines for application-specific handling?
High		When a cable subscriber has trained the system to get speaker ID, the GSUI should show that speaker ID is in effect.
High		We have decided that there should be a command list in addition to the Agile help. This needs to be formalized and added to this document.
High		We need a spreadsheet to describe the messages that appear in dialog boxes and the circumstances under which they appear. The text of the messages is lower priority.
High		We have a lot of the guidelines for speech-enabled applications in our heads from user tests and our experience with the prototype. We need to formalize them.
High		The pictures describing scope, language, and behavior of speech commands need to be done, and the tables with the initial grammar definitions need to be added to the doc.
High		The navigation section (5.1.2) is a placeholder and needs to be filled out.
High		We need to work on the actual words in the grammars. Note: this work will be based on future work with the SpeechWorks prototype and a better understanding of the technical limits of speech recognition.
High		We need to develop a phased implementation plan for the Agile User Center.
Medium		The mechanism for speaker ID training needs to be worked out.
Medium		Each MSO probably already has a main menu, so we need to determine what the menu invoked by a speech command should be. Should it be based on the MSO main menu?
Medium		The television screen interface toolkit section of the document needs a lot of work. So far we don't use dialog boxes or application-specific menus. We're leaving these to the applications themselves. Should we remove this section, or do more to fill it out?

Priority	Completion Date	Description
Medium		The Marketing Requirements section needs to be filled out.
Medium		The role of the agent. We need to answer three primary questions and to determine where the agent fits on several interaction scales. The primary questions are: 1. Do we need an agent/host? Victoria to develop pros/cons list and gather examples. 2. What is the user's relationship to agent? 3. What is agent's relationship to other UI components? The interaction scales are: Personal Relationship (Close/Distant), Visual Development (Real/Abstract), Character Development (Simple/Deep), Engagement with user (Low/High), Intelligence (Low/High), Agile Brand Identity (Same/Different), Brand Expression (Low/High).
Low		The approaches to speech-enabling typical interface elements need to be filled out.
Low		We need to determine the action to take on a double press of PTT.

12 ENGINEERING ISSUES AND DESIGN QUESTIONS

1. Design. Do users want spherical navigation? If so, how can we give users the feeling that they can do spherical navigation without vastly increasing the global navigation grammar?
2. Design. Is a favorites list a good feature to add to the navigation model? If we use a favorites list, how do users set favorites?
3. Design. Do we want to implement a history list for navigation? If so, how should last and forward work? How does this fit in with the MSO's navigation model? Should last and forward behave differently once the user has "exited" an application?
4. Engineering. The way the speakable text icon is implemented is still under discussion, and may vary from application to application. There are two basic implementation paths for adding the icon to the background screens of an application: including the icon in the application graphics and using an overlay. To include the icon in the application graphics, all of the graphics for the application would need to be changed and the application would need two sets of graphics, one with the icon and one without. The icon could only be implemented as an overlay if each of the application screens is static, i.e. each graphic stays in the same place every time the cable subscriber looks at the screen. When the icon can be implemented as an overlay, there would be an overlay for each screen. The speakable text icon should also be added to the dynamic text for an application. For example, the icon should appear in front of the names of the movies in the Video on Demand (VOD) application. This dynamic text is usually implemented as an overlay. There seems to be a single implementation path for doing this: using a special character in the dynamic text to represent the icon. We are still experimenting with this.
5. Engineering. There may be some latency in displaying the non-speech-enabled icon. When an application is not speech-enabled, the STB may be able to display the icon immediately. However, when a screen in an application is not speech-enabled, a round trip to the head end will be necessary to determine that voice cannot be used on the screen.
6. Engineering. Will congestion ever be a large enough problem that we need to advise cable subscribers to stop using voice commands for an extended period, e.g. 5 minutes?
7. Engineering/Marketing. What information are we recording about cable subscriber's use of the AgileTV system? Do we need to develop a privacy policy around this?
8. Engineering. We currently don't think we need a timer to reset the unsuccessful recognition count because when people are having trouble with the system, we want them to see the long message. If user tests show that people don't like the long message, we may have to revisit this.