

Developing IVVR Applications for Teaching and Customer Self Service

**K. W. (Bill) Scholz
NewSpeech, LLC**

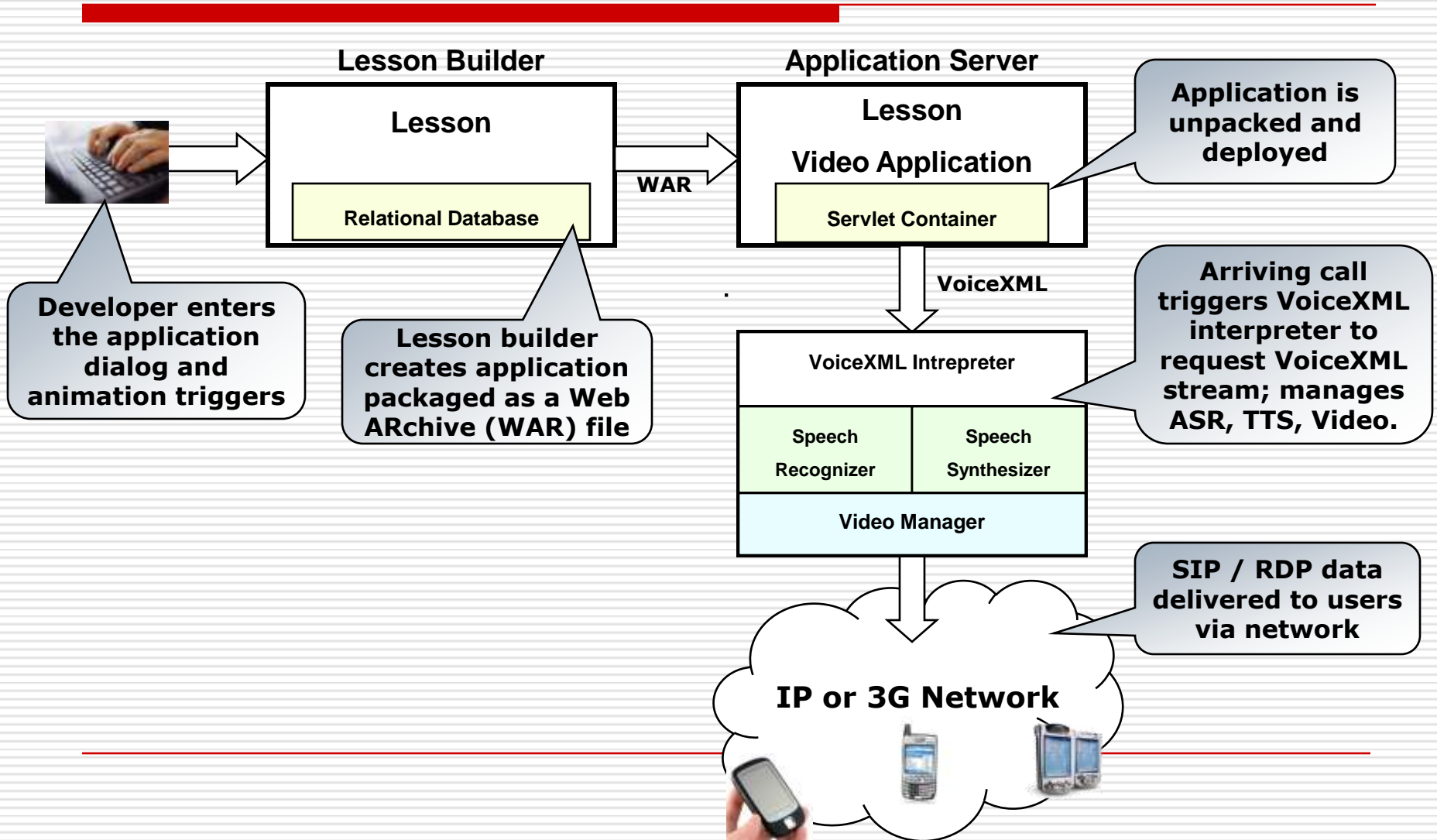
Agenda

- Introduction – the IVVR mobile paradigm
 - Application development and delivery
 - Dialog flow specification
 - Animation specification
 - Application generation
 - Deployment and distribution
-

Introduction

- Overview: interactive voice-video interface to support interactive learning and customer self-service distributed by 3G and IP network
 - Network-delivered IVVR presented on mobile handsets can be effective for:
 - driving individual learning, particularly language learning
 - providing customer self-service
 - providing access to information that can be structured as a simple computer-human dialog
 - Device / appliance trouble shooting and repair
 - Situation management (e.g., managing a fender-bender)
 - Public institution access (e.g., airport check-in)
-

Development & Delivery

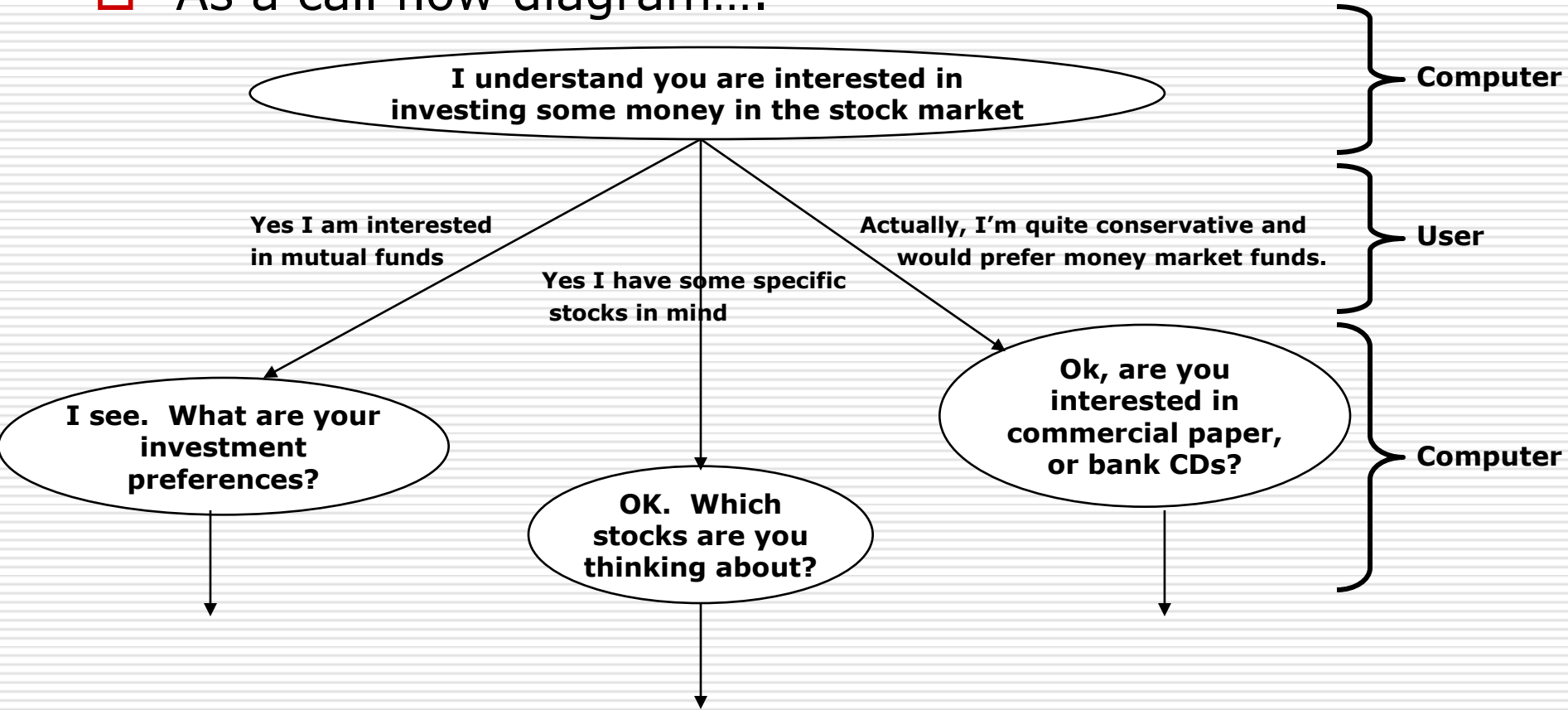


From Dialog flow to application

- ❑ Specify dialog flow: the text for all prompts and responses is entered
 - ❑ Simulate the dialog flow to help find flaws:
 - Off-line simulator on developer's computer
 - On-line execution using TTS, no video or recordings
 - ❑ Generate recording script (text, filenames, characters)
 - ❑ Record audio; optionally preview on line
 - ❑ Generate video by mapping phoneme strings to lip sync and gestures and motions to keywords
 - ❑ Build application with completed VoiceXML, speech grammars, and Audio-Video files
-

Dialog Flow Specification

□ As a call flow diagram....



Dialog Flow Specification

□ As a call flow chart....

Turn	Q2Start	
Prompt	I understand you are interested in investing some money in the stock market.	Goes to
Response 1	Yes, I am interested in mutual funds.	Label_A >
Response 2	Yes I have some specific stocks in mind.	Label_B >
Response 3	Actually, I am quite conservative and would prefer money market funds.	Label_C >

Turn	Label_A	
Prompt	I see. What are your investment preferences?	Goes to
Response 1	I prefer growth stocks. But not too risky.	Label_A4 >

Turn	Label_B	
Prompt	OK. Which stocks are you thinking about?	Goes to
Response 1	Well I am thinking about some of the oil stocks, the majors like shell or ekson mobile.	Label_B4 >

Turn	Label_C	
Prompt	I understand. Are you looking at commercial paper or at bank CDs?	Goes to
Response 1	Commercial paper is ok.	Label_C4 >

Video Generation

- ❑ Recording script generated to manage audio recording
 - ❑ Avatar images and background settings chosen for each character
 - ❑ Keywords in prompts & responses mapped to video snippets of gestures, moods, and eye motions
 - ❑ Recognizer identifies key phonemes requiring mouth movements and maps these to video snippets
 - ❑ Sound dubbed onto audio track
 - ❑ Completed SWF master converted and compressed for shipping via 3G network
-

Recording script

Prompt and response text, filenames, avatar characters

PROMPT TEXT	AVI Filename	Speaker
I understand you are interested in investing some money in the stock market.	MB0002_C102	M10_Professional
I see. What are your investment preferences?	MB0002_C103	M10_Professional
Any interest in biotech? Or how about international plays?	MB0002_C104	M10_Professional
Yes, I am interested in mutual funds.	MB0002_U102	F8_Sporty Casual
Yes I have some specific stocks in mind.	MB0002_U202	F8_Sporty Casual
Actually, I am quite conservative and would prefer money market funds.	MB0002_U302	F8_Sporty Casual
I prefer growth stocks. But not too risky.	MB0002_U103	F8_Sporty Casual
etc.	etc.	etc.

Lip sync – text to mouth movement

Text strings processed by speech recognizer into phoneme strings

....soon....



Phonemes mapped to artist renderings of mouth movements

Mouth movements video snippets merged with avatar



Gestures, moods, eye motions

- Flow chart snippet used to guide avatar construction

a_characters	a_eyemotions	a_holdmood (Preset Mood)	Mood Code	a_mood	Hold Gestures Code	a_holdgestures (Preset Hold Gestures)	Gestures Code
F1_business professional	eyes look up for a moment	normal_mood	mood_normal	normal_mood	H_L_01_hold	H01_LEFT finger upper left_hold	
F2_business professional	eyes look down for a moment	02_pleasant_mood	mood_01	01_very pleased_mood	H_R_01_hold	H01_RIGHT finger upper right_hold	
F3_airline/rentalcar/uniform	eyes look left for a moment	04_bad mood_mood	mood_02	02_pleasant_mood	H_L_02_hold	H02_LEFT finger upper right_hold	H_L_01a
F4_airline/rentalcar/uniform	eyes look right for a moment	05_attentive/interested_mood	mood_03	03_curious_expression	H_R_02_hold	H02_RIGHT finger upper left_hold	H_L_01_hold
F5_business professional	eyes looks left and hold w/ blink state.	08_frustrated_mood	mood_04	04_bad mood_mood	H_03_hold	H03_finger pointing oneself_hold	H_R_01a
F6_business professional	eyes looks right and hold w/blink state.	09_angry_mood	mood_05	05_attentive/interested_mood	H_04_hold	H04_finger up elbow down_hold	H_R_01_hold
F7_senior business professional	eyes looks up and hold w/blink state	13_indifferent/cool_mood	mood_06	06_apologetic_expression	H_05_hold	H05_finger to object in front_hold	H_L_02a
Etc.							

Keyword-triggered animation

Greeting hi mr barrett i'm hank leno from digital international
keyword: 1: hi 2: international 3: 4: 5:

hi hank it's a pleasure to finally meet you
keyword: 1: meet 2: pleasure 3:

Label 1 me too thanks
keyword: 1: thanks 2: 3:

i heard many good things about you
keyword: 1: many 2: 3: 4: 5:

Label 2 oh thank you very much
keyword: 1: 2: 3: 4: 5:

as i mentioned in my email you have a very interesting technology

Animation Features - Windows Internet Explorer

Keyword international

Mood 1 03_curious_expression

Gesture 1 H01_LEFT finger upper left_up

Eye Motion 1

Background 1

Cancel Enter

Lesson Builder Template Processing

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE vxml PUBLIC "-//W3C//DTD VOICEXML 2.1//EN"
"http://www.w3.org/TR/voicexml21/vxml.dtd">
<vxml version="2.0" application="^6_root.vxml"
xmlns="http://www.w3.org/2001/vxml" xml:lang="en-US">
<meta name="Generator" content="LessonBuilder" />
<!-- Lesson name: ^1 -->
<property name="sensitivity" value="^3" />
<property name="bargein" value="^4" />
<form id="Preliminary_Startup">
<block>
<log dest="file:^1.log">,Start: ^1 converse mode.</log>
<^5 />
<goto next="#^2" />
</block>
</form>
```

Template 1: 'Converse' Mode Head

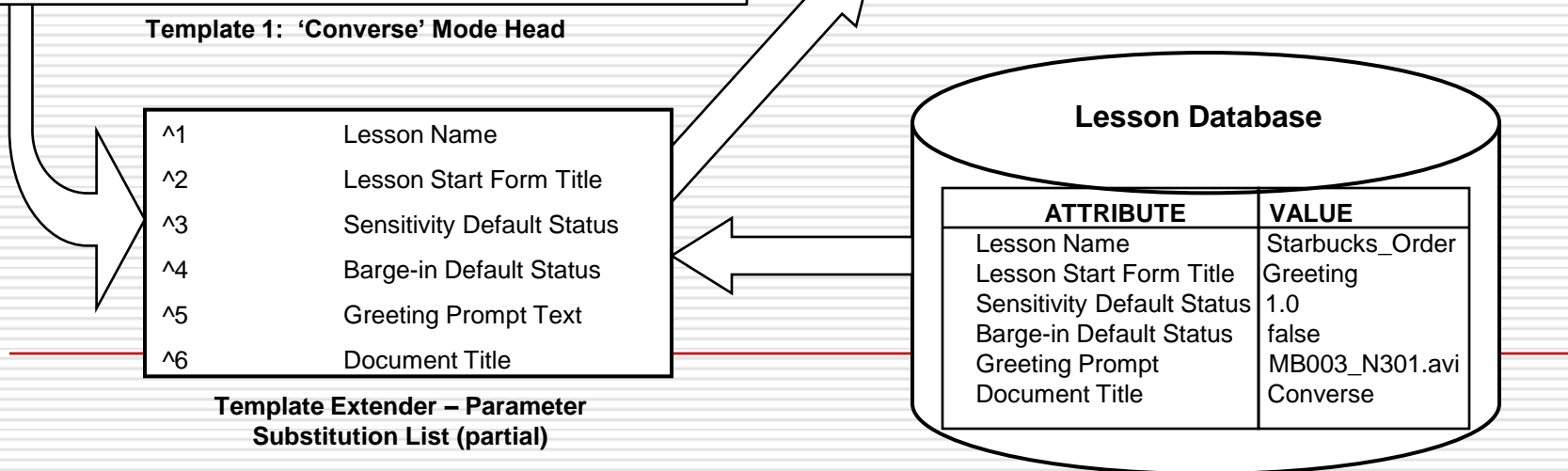
^1	Lesson Name
^2	Lesson Start Form Title
^3	Sensitivity Default Status
^4	Barge-in Default Status
^5	Greeting Prompt Text
^6	Document Title

Template Extender – Parameter Substitution List (partial)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE vxml PUBLIC "-//W3C//DTD VOICEXML 2.1//EN"
"http://www.w3.org/TR/voicexml21/vxml.dtd">
<vxml version="2.0" application="Converse_root.vxml"
xmlns="http://www.w3.org/2001/vxml" xml:lang="en-US">
<meta name="Generator" content="LessonBuilder" />
<!-- Lesson name: Starbucks_Order -->
<property name="sensitivity" value="1.0" />
<property name="bargein" value="false" />
<form id="Preliminary_Startup">
<block>
<log dest="file:Lesson1.log">,Start: Lesson1 converse mode.</log>
<audio expr="videoPromptPath('MB0003_N301.avi')" />
<goto next="#Greeting" />
</block>
</form>
etc. etc.
```

Generated VoiceXML Code

Lesson Database	
ATTRIBUTE	VALUE
Lesson Name	Starbucks_Order
Lesson Start Form Title	Greeting
Sensitivity Default Status	1.0
Barge-in Default Status	false
Greeting Prompt	MB003_N301.avi
Document Title	Converse



Example: teaching ESL


- Video – student practicing English with mobile handset



Summary

- The Interactive Voice-Video Response paradigm
 - Application development
 - Dialog flow specification
 - Prompt recording and lipsync generation
 - Animation specification, keyword-triggers
 - Generate vxml, grammars, and avi
 - Build into industry-standard web archive
 - Delivery: J2EE app server, VXML Media Server, SIP/RTP to client handset
-

Thank You.



NewSpeech, LLC.
Advanced Speech Solutions
Karl W. 'Bill' Scholz, Ph.D.
President

203 Vassar Circle
Strafford, PA
19087

cell: 610-212-8016
fax: 610-989-0989
bill.scholz@comcast.net
<http://www.newspeechsolutions.com/>