

Revised Syllabus

Voice User Interfaces (CS4390/5390)

Spring 2015

Tuesday & Thursday 3:00 – 4:20 , CCS Room 1.0204

Instructor: Nigel Ward
Office: CCS 3.0408
Phone: 747-6827
E-mail: nigel@cs.utep.edu
Office Hours: Wednesdays 2:30-3:30, and by appointment

Motivation: In the past decade there has been an explosion in the creation and commercial deployment of spoken dialog systems user interfaces, recently Siri, Google Now, and Cortana. These systems provide callers with voice access to information and enable them to issue commands and perform transactions. Although standards, design tools, and design methods now exist, building a successful application is still a challenge. The course will cover not only the underlying technologies --- digital signal processing, speech recognition, dialog management, and so on --- but how to use these to produce usable systems.

Course Objectives:

Students will acquire the knowledge and skills needed to create highly usable voice user interfaces (VUI): including:

- key properties of human language and communication
- capabilities and constraints of the underlying technology
- the development process
- implementation, in VoiceXML
- integration with web technologies (PHP, XML, XSLT)
- VUI and rival technologies in the marketplace
- VUI usability fundamentals and heuristics

Students will also learn about emerging technologies and research issues, such as dialog strategy learning, adaptation to the user, and efficient turn-taking.

Students' knowledge and skills in the following areas will be reinforced:

- software engineering
- human-computer interaction
- formal languages
- usability testing
- experiment design

Format: Primarily lectures, augmented with in-class design exercises, student project presentations, and student-led discussions.

Textbook: *Voice User Interface Design*, Michael H. Cohen, James P. Giangola and Jennifer Balogh, Addison-Wesley, 2004 (henceforth VUID).

This will be supplemented by readings handed out in class.

Two other good developer-oriented books to own are *VoiceXML: Introduction to Developing Speech Applications*, James A. Larson, Prentice Hall, 2003, and *VoiceXML: 10 Projects to Voice-Enable Your Web Site*, Mark Miller, Wiley, 2002. The best technology book is *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, Second Edition*. Daniel Jurafsky and James H. Martin, Prentice Hall, 2009.

Prerequisites: This course is designed for graduate and senior-level students in Computer Science, Information Technology, Linguistics, and Psychology. No specific prerequisite knowledge is required, but experience in programming, systems designs, project management, systems integration, and formal notations will be helpful. Students outside Computer Science will need permission from the instructor to register.

Assignments: There will be a number of structured assignments, designed to give experience with various usability engineering activities. Most assignments will be done in teams.

As a final project, individuals or teams will build a substantial dialog system. Past projects included an Auto Body Parts Finder, the SEAL Agenda and Events VUI, the Radio Paradise Song Request and Information Line, a Movie Recommendations and Rental service, a “UTEP Phonemine” Course Information Portal, an Advising Scheduling System, and a Remote Home Control system.

Grading: The weighting will be approximately 45% assignments, 35% examinations, and 20% final project. Graduate students will have one or two more assignments than undergraduates.

Assignments and tests will be challenging. Grading will be on a points-earned basis (points above zero), rather than a points-off basis (points below expectation). Letter grades will be assigned accordingly.

Conduct, etc. Students are expected to be punctual. Assignments due at the start of class will be collected after a one minute grace period; late assignments will receive at most two-thirds credit. Assignments must be submitted in hardcopy.

Tests will be closed-book, except that one single-sided page of hand-written notes may be brought in for the first test, two for the second test, and three for the final. If you leave the classroom for any reason, your test will be graded on only what you did up until that time. No make-up exams or assignments will be given except under the conditions set forth in the Catalog. Students are free to attend class or not, bearing in mind that absence may annoy other students, interfere with learning, and result in a lower grade.

Cooperation among students and among teams is encouraged, but not to the extent that it interferes with each individual’s understanding or with learning-by-doing. Help given and received from other students and sources should be noted in the assignment write-up. More generally, students will follow the UTEP Standards of Conduct, available at <http://studentaffairs.utep.edu/Default.aspx?tabid=4386>.

If you have or suspect a disability and need accommodation you should contact the Center for Accommodations and Support Services at 747-5148 or at cass@utep.edu or visit Room 106 Union East Building.

Important Dates:

Test 1	February 19
Spring Break	March 9-13
Test 2	April 7
Final Exam	May 14, 4-6:45pm

Course Web Site: <http://www.cs.utep.edu/nigel/dialog/>

Approximate Schedule

Part I Introduction

- 1. Course Overview** (1 day)
Course Objectives, Requirements
Historical Perspective
Possible Futures
Assignment A: Voice User Interface Dissection (1hr, 10 points)
- 2. Overview of VUI Design Issues, Technology and Process** (VUID Chapters 1,2, 3) (3 days)
VUIs vs GUIs
VUI components
Development-Process Overview
Assignment B: Project Proposals (2 hours; 10pts)
- 3. Familiarization with the Medium** (VUID Ch 11, 16) (2 days)
The output side: introduction to prompt design
The input side: introduction to recognition grammars

Part II: Tools and Technologies

- 4. Basic VoiceXML** (Larson Chapter 5) (3 days)
Requirements for a VUI language
VoiceXML Basics
BeVocal Café and Voxeo Evolution
Forms
Modularization into multiple files
Assignment C: VUI drop-add form code (1 hour, 10pts)
Assignment D: "Hello world" twice (3 hours, 20pts) **Test 1**
- 5. Interaction Styles** (VUID Chapter 5) (1 day)
Mixed Initiative, Multi-slot forms and Universals
Multimodal Interaction
Assignment E: Restaurant bill advisor, etc. (2 hours, 20pts)
- 6. Dynamic Content and Back-End Integration** (2 days)
Embedded Javascript
Parsing XML, using <data>
Generating VoiceXML from XML using XSLT
Generating VoiceXML server-side using PHP
Assignment F: XSLT-generated Vxml (4 hours, 40 points)
Assignment G: PHP or RSS integration (3 hours, 30pts)

Part III: Design and Development

- 7. Requirements** (VUID Chapters 4, 5, 7) (1 day)
Requirements Gathering

Focus Groups; in-class exercise
Project Triage; in-class exercise
Assignment I: Project Charter (2 hr, 20 pts)

8. Detailed Design (VUID Ch. 8 and 14) (.5 day)
Assignment J: High-Level Design or Quick Prototype (5 hours, 20 pts)

9. Design Principles, Design Elements, Usability (VUID Chapters 9, 12, 13) (.5 day)
Cognitive load, Efficiency, Clarity, Naturalness, Accuracy
Usability Evaluation and Metrics
Assignment K: Quick Prototype or High-Level Design (4 hrs, 20pts) **Test 2**

10. Customers, Markets, Companies, and Opportunities (1 day)
Assignment O: Examine a Speech Company or Product (2 hrs, 10 pts)

11. Persona Design and Implementation (VUID Chapter 6, readings) (1 day)
Persona design (in-class exercise)
Social Psychology (Nass)
Assignment L Persona Tuning (3 hours, 15 pts)

12. Prompt Design and Production (VUID Chapter 17, readings) (2 days)
Audio formats
More on recording (KTEP studio tour)
Prosody; in-class exercise on splicing and prosody
Voice Coaching

13. Development (VUID Chapters 10, 11, 15, 16, 18, 19) (1 day)
Grammar Development and Testing
Application Development and Testing
Data-driven Tuning

Part IV: Technical Underpinnings

14. Dialog Models and Dialog Phenomena (4 days)
Applications that use dialog knowledge
Empirical interlude: beyond state-based models of dialog
The psychology of dialog
Dialog management beyond VoiceXML 2.1
Empirical Interlude: dialog patterns and phenomena
Beyond information: meta-communicative and interpersonal dimensions

15. Sound, Speech, and Digital Signal Processing (1.5 days)
Articulation
Time- and Frequency-Domain Representations
Phonemes and Features
Prosody and Synthesis

16. Speech Recognition, Emotion Recognition (Jurafsky&Martin 2^c, Chapter 9) (1 day)
Spectral and Prosodic Features
Regressions and Classifiers
Probabilistic Acoustic Models and Language Models
Argmax and Search

Part V: Other

- 17. Research Topics** (2 days)
Various topics of interest
Assignment H: present a research paper (4 hrs, 30 pts; grads only)
Assignment P: ask a question about a research paper (1 hr, 10 pts)
Assignment Z: research planning (4 hours, 20 pts grads only)
- 18. Project and Experiment Presentations** (1 day)
Assignment M: Final Reports and Presentations (20 hours, 60pts)
- 19. Review** (1 day)
Assignment N: A Question for the Exam (1 hour, 5pts)

The time estimates for the assignments are for an efficient person working with a well organized team.

Graduate students will be held to a higher standard for Assignment M, and will have two additional Assignments: H and Z.