Voice Recognition Feature of the German Express Courseware: Conceptualization, Specification and Prototyping -- Model Elaboration through Phases

Tatiana Levi, Steve Stokowski, Nikolaus Koster and Andreas Rycshka

Foreign Service Institute, US Department of State, Arlington , Virginia

Abstract

In this paper we discuss how adding speech recognition (SR) affected the development of a computer-based language training program -- how our understanding of the capabilities and limitations of the recognizer evolved and how this understanding affected design and implementation decisions. The program developed is called German Express, it is a CD-ROM based, interactive, language-training course.

1. Introduction

German Express is an introductory course to the language and culture of German-speaking countries. Its target audience is Foreign Service personnel who are novice learners of German. The course contains a set of modules that are divided into lessons. Each lesson has a sequence of screens, each screen is called an activity. In a general way, a lesson sequence progresses from presentation, to practice, to production, to daily-living communicative activities.

The program's content is limited to the essential vocabulary, phrases, sentences, and grammar that are necessary for functional communication. The focus of the content is on word and phrase discrimination and on word order and syntax transformations. The program uses SR in short response and simulated conversation activities to elicit and reinforce listening comprehension and oral production of basic question and answer vocabulary, phrases, and grammar.

Because a learner would be relatively isolated and not receive any immediate teacher feedback, we thought it very important to make the SR component operate as reliably as possible. Making the recognizer reliable meant uncovering and excluding content that affected recognizer performance. Our initial knowledge of SR and its potential was based on seeing and using commercial language-learning software and a French SR module from the US Military Academy.

Course Description

Information from a needs analysis of Foreign Service personnel stationed in German-speaking countries revealed the communicative functions necessary for content development. The analysis information also revealed the relatively minimal amount of detailed grammar knowledge necessary for learners to be functionally successful at the novice level. Thus, as course designers, we decided to have the speech recognizer respond positively to functionally acceptable language rather than just strictly, grammatically correct language. For learners who wanted more grammar information, a third-party reference grammar was provided as a linked resource to the German Express program. [1]

The courseware design applies both the audio-lingual and communicative-proficiency language teaching methods. The audio-lingual component consists of drills and repetitions in listen-and-repeat and listen-record-and-compare activities. These activity types are among the first in a lesson sequence and serve to build the vocabulary base for later communicative activities.

The communicative-proficiency component consists of daily-living activities where the learner interacts with video prompts and feedback to participate in everyday communication scenarios -- greetings, introductions, simple shopping transactions, etc.

A typical communicative activity would run as follows: A character in a video prompts the learner for some expected language. The learner speaks and the recognizer evaluates the utterance. If the utterance is recognized as correct, the interaction moves on. If the utterance is recognized as incorrect, the video character asks for clarification (repetition) and/or the program displays a set of possible choices.

2. Initial Assumptions and Requirements

Our main reason to use a recognizer was to encourage speech production. By repeatedly modelling and asking the user to produce speech we aimed to build a learner's confidence in production and train pronunciation.

Our initial plan was to use the recognizer as a means of prompting and capturing the spontaneous, twist-and-turn flow of "natural" conversation. To do this, we envisioned allowing the learner to start at the beginning of any branch of a dense, multiple-branch conversion tree. When traversing a branch the system would respond to any recognized and appropriate utterance with a randomly drawn response from a set of appropriate responses. Therefore, in this model the learner could navigate a given conversation tree multiple times, each time possibly navigating the tree in a different branch order and receiving different appropriate responses. The hoped for result would be that the learner felt the interaction was a reasonable approximation of a natural flowing conversation.

Our requirements from the recognizer:

- To stimulate learners' speech production.
- To simulate a natural conversation flow.
- To provide an independent assessment of speech .
- To provide immediate evaluative and error-specific feedback based on learner's production.
- To motivate and stimulate learners through feedback.
- To provide flexibility inherent to the natural communicative situations – to accept functionally acceptable utterances and not require full grammatical preciseness.
- To have voice-only, rather than keyboard or mouse, input and thereby more closely simulate real-world communication.

3. Elaboration Process

There were three phases in the development of the design: concept, elaboration and refinement.

Concept Phase

The initial ideas were based on the functional, communicative language identified in the needs analysis -- greetings, introductions, ordering a meal -- as well as a number of language teaching functions. The idea was to construct a continuous listening, mixed-initiative multi-path conversation tree which would simulate a natural exchange between two people.

Elaboration

In the elaboration phase the development team clarified, operationally defined, and expanded the initial ideas. This revealed the work effort required for the mixed-initiative conversation trees. Since these activities were a small part of a much larger course, the effort that could be dedicated to their development needed to be scaled to the overall design workload and timeline. The initial conversation tree was re-worked into a turn-taking, push-to-initiate conversation model.

Refinement

In the refinement phase, discrete elements, such as corrective feedback and selected content phrases, were tested with native speakers and then with a group of beginning German students. These tests identified some limits based on the technical capabilities of the recognizer and some limits based on content selection. We discovered that both the questions posed and the distracters selected were equally important to provide good, helpful, reliable feedback to learner.

To assure better feedback, we needed to develop and include additional recognition grammar that captured the most common inappropriate responses. Also, after a first inappropriate response, we presented the learner with a multiple-choice selection.

4. Recognizer Challenges and Solutions

As a result of field tests we made changes to the recognizer's input parameters, the recognition grammar, the recognition dictionary, and our feedback statements.

Input Parameters

The complexity of the expected input language varied across activities. In one particular activity, the learner was expected to say rather long, polysyllabic German street names. This task was too challenging for most beginning students. The language they produced was full of restarts, articulation artifacts, and pauses or silences. These silences often were long enough to qualify as end-point detection, thereby ending the recognition event.

Through testing, we modified the grammar and extended the end-point parameter to better handle this task. However, because of self-imposed implementation constraints (the way the recognizer was programmed at the course level), we chose not to incorporate these technical changes. Instead, we modified the course content, choosing shorter, less challenging street names.

Recognition Grammar

There were a number of activities that called for the collection of known and unknown input. An example of an unknown input is a piece of personal information, such as a person or street name not known in advance. We were able to modify the recognition grammar with wildcards to act as placeholders for the unknown language and still have the recognizer return reliable confidence information.

Recognition Dictionary

The recognition dictionary did not contain all the items necessary, especially food or specialty items. We added these words to the dictionary by recombining syllables of existing dictionary works to create those needed.

Feedback Statements

We tailored feedback statements to give better conversation context and more direct corrective feedback.

5. Conclusion

Communicative SR activities function reliably and allow for the real opportunity to use the newly acquired language skills in a communicative setting. The final design implementation provided good immediate, evaluated feedback of the learner's language production. Learner feedback on VR activities was very positive.

In using a recognizer for the first time there was a fair amount of work in learning dialog design and the capabilities and limitations of the recognizer. This ongoing learning process required multiple iterations of content development and recognizer implementations. Areas affected include: recognizer end-point length, grammar wild cards, dictionary lexical and phonetic items, and context specific feedback.

Based on our experience, we recommend: First find good speech-recognition dialog design and implementation models. Review and assess the features and limitations of these models. Write a clear set of requirements based on the review. Prototype in stages until the full-set of desired features is implemented.

References

[1] Smith, G. (1997) *The German Electronic Textbook.* Retrieved May 9, 2004 from: http://www.wm.edu/CAS/modlang/grammnu.html

[2] McLaughlin, B. (1987*). Theories of second language learning,* Arnold, London.

[3] Huang, X.., Acero, A., Hon, H. (2001) *Spoken Language Processing: A Guide to Theory, Algorithm and System Development,* Prentice Hall, New Jersey.

[4] Bernsen, N. et al. (1998) *Designing Interactive Speech Systems: From First Ideas to User Testing*, Springer Verlag, New York.