International Academic Conference on Teaching, Learning and Education
T L E C O N F

3rd International Academic Conference on Teaching, Learning and Education

MUNICH, GERMANY
29-31 July, 2021

Spoken Language Fluency Via Live Situational Language Practice On an AI-Driven App

Brendan McMahon - PGCE, Dip TEFL., MBA
Smoothweb Technologies Limited, Hong Kong

Abstract

A Foreign Language Learning Platform is provided which uses NLP in situational dialogs and simulates active conversations to facilitate voice-based language acquisition by a user.

The platform presents multiple themed situational dialogs that simulates and determines stress and interaction of live dialogs by the user with the platform. For example, the situational dialogs may include a Japanese native language speaker learning to book a taxi in English language or ordering food in English language or asking for directions in English language.

At the core of the platform is a proprietary approach that bypasses a requirement for an exact response from the user allowing multiple responses to the dialog which intensifies the live interaction in the dialogs. To simulate a live language exchange as closely as possible, the NLP model references multiple alternatives so there is more than one response to a situation. Multiple language form alternatives include but are not limited to, nouns, verbs, synonyms, vocabulary, grammar, and other parts of speech.

The platform can also offer hints to speakers during the dialogs via assisted hints as well as multiple testing options pre- and post-dialog.

Keywords: emerging technologies, AI-driven app, real-time interaction, AI teacher and coach, Situational language learning through dialog

Introduction

Smoothweb Technologies (Smoothweb), a commercial organisation, with R&D capabilities and expertise in software development and AI voice research, is at the forefront of AI voice research and development to commercialise a proprietary language learning platform supported by several patents held by the company, filed in 2020 (pending), which support management of STT (speech to text) and TTS (text to speech) models, enhanced error reduction capabilities, the ability to listen in to a dialog and to provide feedback on content and context, with optional navigation by voice commands.

The language learning model – called ARGOT - simulates real-time spoken interaction on a mobile app with a wide choice of languages to choose from. The aim is to provide live, simulated performance scenarios for a student with an AI buddy/tutor. The user studies theme-based situational dialogs via pre-testing and post-testing and goes "live" where he or she can take part in a dialog in real-time with an 'AI Buddy'. The app evaluates and encourages while reducing the anxiety of live speech situations. It gives the student more confidence to speak up. A dashboard also provides feedback on spoken performance and engagement in dialogs.

ARGOT's voice first approach differs from other language learning mobile apps on the market such as Duolingo, which are based on response by click or swipe or one-word answers in multiple choice 'gamified' applications.

ARGOT will provide learning opportunities for multiple users: Individuals who set their own daily targets for learning from a large database of graded situational dialogs, and organisations, training departments, schools, and colleges where bespoke and off the shelf scenarios can be provided for specific speaking scenarios – retail, customer service, immigration, spoken fluency needed for college entrance or career promotion in organisations.

1. A review of recent approaches to language learning and understanding in text and speech innovations

As part of a review into current language simulation models, we undertook desk research which has uncovered a myriad of online language learning solutions and language apps in the market. We also benchmark the market leading app – Duolingo – to identify best practice and opportunities for improving on these apps with respect to active speaking skill development. The research provides insights to the various approaches to language learning and understanding. Based on these findings, our aim with ARGOT is to offer a differentiated and sustainable language learning app.

Language 'learning' computers
Several computer hardware and software language translation solutions take questions or utterances from a user and output the speech after analysing dialog context using statistics and various proprietary methodologies. The algorithms can gauge the level of the learner by matching input to language context (from the database). This system can be used in context for shopping, information. Commercially, this approach can be seen in multilingual portable translators where a passive database is manually searched in the native language for the output to be in the target language.

Digital Assistant Apps
Other products offering digital assistants have storage in memory on a device. Nodes and actions can combine groups of words for specific purposes (ordering in a restaurant, bar, station). Actions are inferred and the AI performs a task (e.g., program invocation). The aim is to reduce error by associating a word with a specific pronunciation.

Conversational understanding systems based on toolsets
A number of speech solutions are based on processor and memory system instructions. For instance, the technologies respond to speech input and display a result on a screen for users, specifically as output from search by text or voice (i.e., movie or music domains).

Foreign language acquisition and learning based on context awareness using a smart device
Other products are based on language learning and service from a real-world data of the user and his/her interests and using a scheduler, information for setting a time to study, to send language learning content to a device. Content is pushed such as an exemplary dialog with content related to the learner and real-life situations with additional video, multimedia content to increase learner interest.

Intelligent human machine conversation framework with STT and TTS
This area of innovation relates to a multilingual speech processor with sentiment for use in live situations where a response to a situation is needed by a control room (e.g., emergency, disaster, trespassing). It uses STT and TTS to supply a response from a live enquiry.

System and method for language instruction using visual and/or audio prompts
An instructional approach offers language instructions - speech segments- phonemes, words - with specific applicability to a vehicle with direct instructions, more commonly known as car navigation systems. It does not aim to reproduce situations or dialogs and is only instructional in nature.

Two language free dialog system and method for language learning
STT intent recognition is done to convert STT to the target language and analyse intent. The output is then used to perform a task such as moving a map or invoking an agent to retrieve a piece of information - e.g., advertising content, a music application.

System and method for providing a natural language voice user interface in an integrated voice navigation services environment
This system uses speech-based models and dialog-context models at different levels and prompts when the user cannot find an answer as a chatbot-driven voice search and navigation application. Response is from a stored database.

Findings with implications for the ARGOT language learning model
The multiple approaches and contexts for interactive voice solutions have been designed to solve a specific language exchange whether via a local set of terms and references in a locale – e.g. shopping centre, security context, digital signage, or for a specific user – music subscriptions service, car navigation.

Approaches are confined to recommended expressions and canned output. Also, the systems do not generally cope with situational dialog or language systems graded by difficulty. Errors are answered by outputting a best guess reply unlike ARGOT which is a language learning system designed to analyse and reply in a live dialog situation or give hints to correct the speaker. ARGOT also provides a summary of feedback report on accuracy. ARGOT provides testing, rewards, and recycles language based on performance to improve speaking and language ability. Error reduction is built into the model to give hints.

While these are more innovative than earlier citations, in this context the real-life situations are driven by the real-life context of the learner. There is no context of an integrated set of situational dialogs or ties to an overall curriculum as provided by ARGOT for instance. There is no reinforcement through the practice-speak-test method ARGOT uses. The systems also do not mention multilingual capability.

The language learning Approach using NLP technologies in tandem with language learning methodologies

Following on from the analysis of technology applications and toolsets, we looked at current language learning apps in the market which have been designed to simulate language learning in a more natural approach incorporating elements of AI and technology tools.

Case study: Duolingo - the market leader in language learning apps
Duolingo has made notable progress in providing a proprietary approach via a learning algorithm and spaced learning based on a standard curriculum (CEFR) with gamification to motivate, reward, and encourage repeat use. Duolingo has succeeded in building a brand around a behavioural approach rather than manipulation of language use and intent via tools and technology alone.

For some time now, language apps such as Duolingo (and Memrise, Busuu among others) have captured the attention of a younger and more independent learner in the context of a world driven by social media and instant news. These apps have a certain appeal in the soundbite learning approach. While the apps are hugely successful - Duolingo, for instance, claims over 30 million subscribers -, the jury is out on the effectiveness of language acquisition.

Observers of Duolingo's commercial success approve of the advertising revenue approach and Duolingo spends a lot of time and effort supporting its market position as an educational provider. It talks about its pedagogical foundations - CEFR-based language acquisition - and testing strength (for which it holds a patent), although the latter has not yet achieved the uptake expected.

Duolingo is arguably more of an entertainer - not a bad thing - offering a gamified approach to learning with 5-15 minute daily lessons as its mantra. Typically, a lesson offers 8 sections with drills to acquire new vocabulary. The repetition and reinforcement with spaced learning allows students to build up a vocabulary base and improve language capability in context of the app. However, there is little or no active production of new language in dialog / speaking, for instance. Supported by messaging and reminders to practice 5 minutes a day, the business model is more about avoiding churn of the customer base. The 'free forever' promise is, however, attractive to many who are not deterred by the constant adverts that appear between each lesson.

The gamified approach plays out a consistent pattern of tests. This predictability and incremental goal learning appeals to users who are keen to learn.

2. ARGOT - Blueprint for an AI approach to language learning though active speaking

Learnings from the research carried out underpin the development of the ARGOT model, which combines multiple elements to produce a robust technology solution with an AI backbone. This has the potential to be a competitive solution to existing models, adding value to learners though its simulation of situational dialogs in the context of now familiar gamified learning on mobile apps.

ARGOT specifically provides the following: A Foreign Language Learning Platform is provided which uses NLP in situational dialogs and uses **simulated active conversations** to facilitate **voice-based language acquisition by a user.**

The platform presents multiple simulated series of **situational dialogs** that simulate and determine stress and interaction of live dialog by the user with the platform. For example, the situational dialogs may be such as a user having Japanese as a native language learning to book a taxi in English language or ordering food in English language or asking for directions in English language.

The platform uses a **search mechanism** in the situational dialogs that bypasses a requirement for an exact response from the user. The platform controls a degree of error in the situational dialogs to offer the user assisted/curated dialog practice as well as multiple testing options.

ARGOT's **error reduction** examines multiple language forms including but not limited to, nouns, verbs, synonyms, vocabulary, grammar, and other parts of speech to allow multiple responses to a question or dialog element.

The platform uses Speech-To-Text (STT) and a Text-To-Speech (TSS) in situational conversations, dialogues to provide several benefits:

- **Conversational engine** for ability, proficiency; context via a way to determine through TTS alternatives (e.g., 'Where is the pool?' example), thereby allowing us to understand and provide testing scenarios, materials for learners' individual needs.
- A **context model** (anaphoric reference) to cope with usage of language in situations where meaning is inferred by a listener e.g., through illocutionary force (wanting the listener to believe you have done something and the listener does).
- Ability in a dialogue model to **request extra information** to complete a request - i.e. force a dialogue to complete.
- **Cope with error** in speech, multi-turn intents, to prompt a speaker for more information or clarification.
- Provide a system for imparting/teaching foreign languages augmenting and replacing the need for live tutors, with the benefit of **higher productivity, motivation**, and **reduced costs** of teachers / service provision.

- The platform uses **real-time help and support** to the user involved in the situational dialogs for learning a foreign language. For example, when a user proficient in Japanese language attempts the situational dialogs in the English language, he or she may be presented with hints in written (e.g., both in Japanese and in English language) and in verbal form (e.g., English Language) when the user could not provide the correct answer.
- The platform provides **voice navigation** and helps the user to search to find resources on the learning platform/application for further practice and expansion of skills and content.
- Machine language capabilities using a proprietary approach in English language to develop models for training and elaboration in multiple languages leading to reduced costs and time in replicating situational dialogs in multiple languages.
- **Gamification** of drills using speech/voice (not just click and drag) or passive language skills such as listening or reading.

3. Building the Content model for ARGOT

ARGOT can gauge the level of the learner by matching intent to language context. ARGOT can be used in context for retail, information search, social situations and business and educational situations. The components and resulting benefits that differentiate the ARGOT language acquisition model are:

- **Low code/no code model:** This allows non-engineers to create content on the Argot model with simple language input for question and answer, dialog creation and contextual situations with multiple pathways.
- The ARGOT system allows for multiple synonyms, differences in local cultural expressions and mixed language identification to be mapped, and post-launch of new content with minimal cost and time to scale and extend.
- Proprietary language model: This saves time and allows deployment of multiple languages via reference to open-source multilingual data sets – e.g., Google and Bing - obviating the need, time, and costs of training a new language on separate data sets / training data.
- **Data tagging:** This allows efficient relational database creation for tagging language by parts of speech, topics, language levels. Learners benefit through voice search to find similar situational dialogs and multiple levels to navigate and learn if desired.
- **Efficiencies:** Given the benefits of time and cost savings that accuse from the faster time to deploy, Lower cost of training and scale, and hardware variable costs.

4. ARGOT status and roadmap

ARGOT currently supports three languages in its pilot phase - English, Spanish and Japanese. ARGOT offers any native speaker the option to study using the direct or indirect method. This allows beginners to login to the app with the support of their native language for instructions if desired. More advanced learners may prefer instructions and navigation to be in the target language for study.

The pedagogy on direct versus indirect methods of teaching leans more towards the former as a more effective approach, especially in structured courses of classroom language teaching over a period. We took the decision to allow flexibility at this stage to gauge the behaviour and engagement through providing options for students. Additionally, the 'App approach' to studying languages - Duolingo, Memrise, Busuu, and so on, offer navigation and direction in the native not the target language of the learner.

Once the learner logs in to ARGOT, he or she can select from multiple situations. The student can choose to practice and study the vocabulary and change roles, hide or unhides the dialog. Hints and advice are given by the AI buddy. When ready to go live, the student takes one of the roles in the dialog and the AI buddy is the counterpart. Roles can be swapped for further practice.

The student is required to complete the dialog with options for hints or language to be revealed if necessary. The dialog may be repeated multiple times. The AI buddy gives feedback at the end and a recording can be kept. Future development will make use of sentiment analysis through tone of voice and hesitation, accuracy of expression, and so on.

A key goal is for the learner to have a non-threatening spoken language course and understanding of the personal power and command of the spoken language in multiple situations with the help of AI.

For individuals, a dashboard provides performance ranking data related to friends or other learners of a similar level. For businesses, reports can be provided to a line manager as part of an intensive or ongoing training program.

Finally, the content is mapped against accepted curriculum standards and levels and against a matrix of popular language tests.

Conclusion

The case for building an AI language app - ARGOT - has merit given the combination of AI proprietary technologies, the demonstrable benefits of time, cost savings and scalability and extensibility. Additionally, the demand in the market for an active skill tool or solution has never been stronger. Proven successes such as Duolingo have pointed the way to an ongoing demand for language learning, especial accessible via a mobile app. With cultural and social changes flipping classroom learning and with multi-modal approaches to language acquisition, the use of AI voice as a tool for improving spoken fluency and confidence is of ongoing interest.

The downside to the hype is that currently, AI has its limitations as it is not NLU - natural language understanding - where sophisticated illocutionary force or disguised intent is practised by native speakers. Learning or language learning systems are currently limited to multi-turn dialogs where outcomes are more expected and conversations do not take on the characterised by lively debate with arguing and rejection, dismissal of multiple parties. It is still a very useful tool for language acquisition where recommended expressions are given in the context of learning via graded exercises.

Acknowledgment

This paper is written based on research and development by Smoothweb Technologies Limited, an AI company based in Hong Kong developing multiple voice applications for language learning as well as other domains where real time AI voice dialogs can facilitate more productive and efficient exchange of information, i.e., in commerce, health, transport, retail.

We are grateful to Professor Obari, Ph.D in Computer Science, M.A. TESOL of Aoyama Gakuin University, Tokyo and Mr. Andrew Hemingway, M.A., Director of Language Programs at Hertford College, Oxford University for support and insights related to the development of ARGOT and for ongoing collaboration.