LANGUAGE MODELING IN AIR TRAFFIC CONTROL

Claudiu-Mihai GEACAR[1], Dan-Cristian ION,[2] Adrian-Mihail STOICA[3]

În mod normal, comunicațiile între piloți şi controlorii de trafic aerian se realizează cu ajutorul sistemelor vocale radio. Pe măsură ce volumul traficului creşte, rutele devin tot mai diverse, iar aeronavele uşoare şi ultra-uşoare devin tot mai accesibile, sistemul clasic începe să-şi arate punctele slabe.

Recunoaşterea vocală poate fi utilizată pentru ca o alternativă la sistemele radio convenționale şi, împreună cu un sistem de comunicații digitale, poate duce la creşterea siguranței şi eficienței transporturilor aeriene.

Modelarea limbajului este un element cheie în construcția unui sistem de recunoaştere vocală. Modelele de limbaj prezentate în această lucrare au arătat faptul că recunoaşterea vocală constituie un mijloc viabil de comunicare, putând ajuta la depăşirea dezavantajelor sistemelor clasice de comunicare prin voce.

Traditionally, pilots and air traffic controllers use radios to communicate with each other. As air traffic increases, routes get more and more diverse and light and ultra-light aircraft are becoming more and more popular, the classic system is beginning to show its weaknesses.

Voice recognition can be used to provide an alternative to the classic system and, together with a data link system, can increase the safety and efficiency of air transport.

Language modeling is a key aspect in building a voice recognition system. The language models presented in this paper have demonstrated that speech recognition is a viable means of communication that can help overcome the drawbacks of classic voice radio communication systems.

Keywords: air traffic control, voice recognition, language model

1. Introduction

Pilot/ATC communication is normally done using voice radios. The capacity and capabilities offered by this system were sufficient at the beginning and up to the end of the 20th century. As air traffic and routes increased and small, private aircraft became more available, this system began to show its weaknesses. The most important ones are channel congestion and language issues.

Channel congestion is caused by the increasing number of aircraft using the same radio frequency while in the same ATC sector. As this number increased, the growing amount of information became more and more difficult to transfer using that single frequency.

[1] PhD Student, Faculty of Aerospace Engineering, University POLITEHNICA Bucharest, Romania, e-mail: claudiugeacar@gmail.com

[2] PhD Student, Faculty of Aerospace Engineering, University POLITEHNICA Bucharest, Romania, e-mail: dancristianion@gmail.com

[3] Prof., Faculty of Aerospace Engineering, University POLITEHNICA Bucharest, Romania, e-mail: amstoica@rdslink.ro

Channel congestion can cause aircraft delays and, most importantly, it can affect the safety of the passengers and flight crew. One of the most dangerous situations is encountered when two people try to talk on the same frequency at the same time, when everybody listening to that frequency will only hear a squeal.

The most tragic accident caused, among other factors, by channel congestion, was the one in Tenerife, on March 27, 1977, when a KLM Boeing 747 crashed while taking off into the top of a Pan Am Boeing 747 that was on the same runway and resulting in the loss of 583 lives (making it the deadliest accident in aviation history).

Language issues are becoming even more common today. As the world becomes more and more "global", language becomes a key factor in the efficiency of pilot/ATC communication. The two main types of language issues are related to similar phonetics of words or numbers and to different levels of proficiency in using the English language. Other types of issues are the use of non-standard phraseology or the use of other languages than English.

The effects of language issues have urged the need for ways to increase the reliability of radio voice communications. Efforts have been made in order to improve crew English knowledge, to ensure proper use of ICAO standard phraseology and to avoid confusing situations.

Rules and recommendations have been issued in order to avoid confusing situations, such as:

- clearly stating if three-digit numbers ending in a "zero" refer to altitude or heading;
- use of ICAO letter pronunciation in order to avoid confusion between B and G or C, D and the number "three";
- the avoidance of confusing statements, such as "made a …" and "mayday" or "hold in position" and "holding position";
- grouping of similar words ("climb to two thousand" can easily be mistaken for "climb two two thousand").

The language issue remains open, with a growing concern caused by the increasing traffic and availability of affordable light aircraft.

The concept of using voice recognition in air traffic control is not new. Voice recognition is currently used for the training of air traffic controllers in ATC simulators, and research has been made [1] for the use of voice recognition in preventing runway incursions.

Other recent research [2], [3], [4] has focused on specific speech recognition tasks such as language recognition and language model robustness. However, these papers are based on the recognition task being performed at the receiver, which in turn poses several speech recognition issues, the most important ones being the high speaker variability (in terms of phonetic differences) and radio interferences, resulting in poor recognition performance. In order to compensate

such differences, the speech recognition system requires large amounts of training data, covering all possible acoustic variants, which make such systems rather difficult to implement. Another key aspect is that such an implementation does not solve one major issue of classic voice radio communication systems, which is channel congestion, because the system still uses voice radio as a means of transmitting voice signals.

This paper presents another approach of the ATC speech recognition problem, for the first time by the authors' knowledge, which is performing the speech recognition at the transmitter, avoiding the use of voice radio and the associated problems (channel congestion and radio interferences). Because the system is pre-tailored to a specific user, the initial training requires a significantly reduced amount of data compared to previously proposed systems, making it easier to implement, faster and more robust. This concept [5] proposes two applications, one being the broadcast of messages using a data link to all recipients on a specific frequency, which are displayed as text on dedicated displays. The other solution is a voice-controlled interface to Eurocontrol's LINK2000+ system. The language models described in this paper follow these solutions.

Voice (or speech) recognition is the process of converting spoken words into machine-usable information (like binary codes of a character string). Voice recognition was initially developed in the United States as a substitute for the filling of medical transcripts. Today, voice recognition is widely used, some of its most important applications being:

- medical: people with physical disabilities can use voice commands to control various equipment;
- military aviation: voice-commanded cockpit (radio tuning, autopilot setting, display control), battle management;
- civil aviation: air traffic controller training, English language proficiency testing [6].

Usually, the difficulty of speech recognition increases as the vocabulary gets larger or the words have similar phonetics. In the case of word sequences, language models or artificial grammar are used in order to limit the number of possible word combinations.

The simplest language model can be defined as a finite network, where the words that can be used after a specific word are explicitly defined. More general models, that can approximate natural speech, use context-dependent grammar.

The leading standard in voice recognition are Hidden Markov Models [7] (HMM). HMM is a stochastic model, in which the generation of phoneme strings is represented probabilistically as Markov processes. HMM is used together with neural networks (used to determine the sound-word correspondence probabilities) in so-called hybrid systems.

The theory behind HMM models is described in the following proposition.

Proposition. *Consider a N-state first order Markov chain. The system can be described as having one of the distinct states 1,…,N at any given discrete time t. The state of the system at time t is noted q_t. Now, the Markov chain can be described using a state transition matrix $A=[a_{ij}]$, where*

$$a_{ij} = \Pr(q_t = j \mid q_{t-1} = i), 1 \le i, j \le N, \tag{1}$$

with the constraints

$$a_{ij} \ge 0 \tag{2}$$

and

$$\sum_{j=1}^{N} a_{ij} = 1, \forall i \tag{3}$$

Assuming that q_0, the system state at t=0 is given by the initial state probability $\pi_i = \Pr(q_0 = i)$, then for any state sequence $\mathbf{q}=(q_0, q_1, \ldots, q_T)$, the probability of this sequence being generated by the Markov chain is

$$\Pr(\mathbf{q} \mid A, \pi) = \pi_{q_0} a_{q_0 q_1} a_{q_1 q_2} \ldots a_{q_{T-1} q_T} \tag{4}$$

We will assume that $\mathbf{q}$ is not observable. Instead, we will assume that each observation $\mathbf{O_t}$ (the cepstrum – the coefficients of the Taylor series of the LPC spectrum of the speech signal) is generated by the system state q_t, $q_t \in \{1, 2, \ldots, N\}$. We also assume that the generation of $\mathbf{O_t}$ in any of the possible states i is stochastic and characterized by a probability set $B = \{b_i(\mathbf{O}_t)\}_{i=1}^{N}$, where

$$b_i(\mathbf{O}_t) = \Pr(\mathbf{O}_t \mid q_t = i). \tag{5}$$

If the state sequence $\mathbf{q}$ that generated the observation sequence $\mathbf{O}=(\mathbf{O}_1, \mathbf{O}_2, \ldots, \mathbf{O}_T)$ is known, the probability of the sequence $\mathbf{O}$ being generated by the system is

$$\Pr(\mathbf{O} \mid \mathbf{q}, B) = b_{q_1}(\mathbf{O}_1) b_{q_2}(\mathbf{O}_2) \ldots b_{q_T}(\mathbf{O}_T) \tag{6}$$

The joint probability that $\mathbf{O}$ and $\mathbf{q}$ are generated by the system can be written as

$$\Pr(\mathbf{O}, \mathbf{q} \mid \pi, A, B) = \pi_{q_0} \prod_{t=1}^{T} a_{q_{t-1} q_t} b_{q_t}(\mathbf{O}_t) \tag{7}$$

The stochastic process, represented by the observation sequence $\mathbf{O}$, is given by

$$\Pr(\mathbf{O} \mid \pi, A, B) = \sum_{q} \pi_{q_0} \prod_{t=1}^{T} a_{q_{t-1} q_t} b_{q_t}(\mathbf{O}_t) \tag{8}$$

which describes the probability of $\mathbf{O}$ being generated by the system without assuming the knowledge of the state sequence for which it was generated.

The hidden Markov model is thus defined by $\lambda = (\pi, A, B)$, also known as model or model parameter set.

For this particular paper, the states consist of individual words (in the abstract sense, disregarding the acoustical information), and the observation

sequence is made up of individual messages (or sentences). The probabilities involved in the model construction are determined using the frequency of individual words across the training data.

The research presented in this paper is focused on the building of generic language models that can be used globally for all the user types and a specific set of models, namely one that can be used for the pilot messages.

The language models were built and tested using the HTK [8] software suite, developed by the Cambridge University Engineering Department.

The models are built for evaluation and research purposes and are not exhaustive in terms of vocabulary and / or phraseology.

2. Language modeling process

Language modeling is a key aspect in building a voice recognition system. Language models represent the core of the system, all the other models being based on or linked to it.

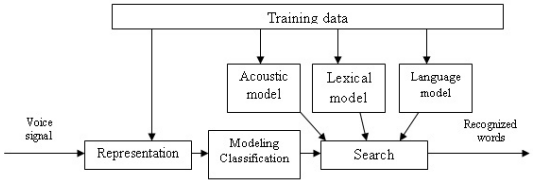

Fig. 1. Speech recognition process

The first step in building a language model is determining the vocabulary. The vocabulary contains all the words that the speech recognition system will be able to recognize.

Message broadcast model

In this case, the vocabulary was determined using the ICAO Radiotelephony Manual (ICAO Doc 9432) and the standard ICAO pronunciation and phraseology rules, in consistency with ICAO Annex X. A brief example of standard ICAO voice radio communication rules are the pronunciation of letters (for example, “A” is pronounced “ALPHA”, “B” – “BRAVO” and so on) and also the pronunciation of numbers (one figure at a time, with one specific case, where “NINE” is pronounced “NINER”, in order to avoid the phonetic similarity with “FIVE”).

As mentioned earlier, the model built is not complete in terms of vocabulary and phraseology. The vocabulary does not contain all aircraft models (for evaluation purposes only Airbus and Boeing are included), nor airline telephony designators (e.g. Tarom, Lufthansa etc.). These can be added later to the model with minor modifications.

LINK2000+ model

In the case of the LINK2000+ system, the base for the vocabulary was Eurocontrol’s „ATC Data Link Operational Guidance for LINK 2000+ Services”,

which describes the functioning of the system and also the message types that the system can handle. Also, the standard ICAO pronunciation rules were used for letters and numbers.

The automated messages, which require no user interaction, were not implemented in the model.

The second step in building a language model is defining the training data. Training data is used in the process of determining the so-called *n-grams* (sequences on *n* symbols – words, syntactic categories etc.), which define the occurrence of specific words in the training data.

The third step is the computation of the *n-gram* probabilities, generating the actual language model.

3. Generic language models

Message broadcast model

The generic language model was built taking into account all the user types as described in ICAO Doc 9432. This includes ground vehicles, aerodrome ATC units (ground, tower, approach, arrival, departure, delivery), area ATC units (radar and control) and flight crews.

The vocabulary was determined by analyzing the phrases in the ICAO Radiotelephony Manual and the standard pronunciation and phraseology rules. As a result, the dictionary contains 650 unique words, covering all situations from regular messages to emergency situations.

The words were determined by isolating all "standard" words like request types, clearances from the phrases (e.g. «callsign» REQUEST CLIMB TO FLIGHT LEVEL «flight level») and then determining all the words that make up the variable parts of the message.

Each word in the dictionary is assigned a unique identification number, which will be used in the later stages of the model building algorithm.

The next step in the language model build is the training data. The training sentences must be chosen so that they cover, in a statistical sense, all the words in the dictionary and all of the sentences in the phraseology rules.

The training sentences were constructed taking again into consideration the ICAO Radiotelephony Manual and choosing the combinations of words in the variable parts of the messages so that they are statistically balanced over the model dictionary.

For the generic broadcast model, the training data consists of 632 sentences, representing complete conversations between flight crews, air traffic controllers, ground maintenance crews and ground equipment both in normal and emergency conditions.

The training data is then used to determine the *n-grams* (for this model, n=3 – *tri-grams*), which constitute the grammar part of the language model.

The *n-grams* (again, having a unique identification number assigned) are used in conjunction with the training data in obtaining the so-called *frequency of frequency table*, which contains the frequency of each word in the training data based on the frequency of each *n-gram*.

This table is then used in the final step of the algorithm, which is the computation of state (word) sequence probabilities.

The analysis of the final language models shows that the *tri-grams* with the highest frequencies are the ones that make up aircraft registrations. This is normal as, according to ICAO regulations, all radio messages must contain the registration number of the aircraft of origin or the destination aircraft.

In order to reflect the magnitude of the occurrence of aircraft registration data in the registration data, the number of occurrences of such word sequences (in the tri-gram sense, that is sequences of three words) is shown below, along with some other higher-occurring word sequences:

- <Sentence start> YANKEE ROMEO occurred 309 times;
- YANKEE ROMEO <Random registration> occurred, on average, 56 times;
- Flight level clearance keywords occurred, on average, 16 times;
- Other general clearances occurred, on average, 6 times.

The testing of the built generic model was performed using a test set of 67 sentences, derived from the standard ICAO phraseology and containing 993 words. The test set contains at least one of each type of messages included in the training set. The test set also contains sentences with a slightly different grammar construction (the aircraft registration also contains numbers, while in the training sentences, aircraft registration contained only letters).

The performance of the language model is evaluated in terms of accuracy (number of words in the test set successfully determined) and perplexity (a measure of on average how many different equally most probable words can follow any given word).

The resulting accuracy was 93%, with a perplexity of 25. The low perplexity is an indicator of good model performance, meaning that, on average, the recognition system has to choose the best word out of 25 possible words. Given that the dictionary contains 650 words, this value indicates that this model could be used in a real speech recognition system with good results.

LINK2000+ model

The generic LINK2000+ language model was built using the same algorithm mentioned above. However, given the different phraseology rules, the actual data used for model generation is entirely different.

The first difference is given by the vocabulary size. The LINK2000+ system uses a strictly defined set of messages, divided into uplink (from ATC to aircraft) and downlink (from aircraft to ATC) messages. The set covers all routine

messages, such as route clearances, altitude and heading requests and information requests.

As a result, the model vocabulary contains 134 distinct words. In order to limit the vocabulary size, the free text function (used by the system in order to allow the sending of messages not covered by the standard set) was not implemented in the language model.

Similar to the broadcast model, each word in the vocabulary is assigned a unique identifier.

The training data was built in order to give an optimum statistic cover of all word combinations in the message set, taking into account the phraseology rules. As a result, the training data consists of 1931 sentences.

For the *n-gram* building step, a *four-gram* model was built, given the fact that the message length was strictly defined.

In this case, given the fact that most messages contain routine route clearances, the training data consists mostly of this particular type of messages. That is why 289 sentences begin with the word ”REQUEST”, and an average of 200 sentences start with other clearances words such as “CLIMB”, “DESCEND”, “CROSS” or “MAINTAIN”. Other high-occurrence word sequences consist of flight levels or flight route designators.

The test data set contains 83 sentences and 634 words. Similar to the broadcast model, the test data contains sentences with a different construction to the ones used for training, for robustness evaluation purposes.

The resulting accuracy was 95%, with a perplexity of 33. The higher accuracy score is a direct result of the reduction in the vocabulary size, which reflects in a similar reduction in the number of words that the system has to choose from during the recognition task.

These values are not a strict indication of how well this model would perform in a real recognition application, but they offer an estimate on how well it models the language used in the application it was designed for.

4. Pilot language models

Because speech recognition is a very demanding process in terms of hardware and software resources and in order to reduce de amount of data required, it is often preferred to design smaller models that will be used in specific applications inside the main recognition task.

For this reason, we have designed two particular models, in this case language models applicable to pilots, based on the global models described above.

The vocabulary of the pilot models is the same as the one of the generic models, containing 650 individual words in the case of the message broadcast model and 134 words for the LINK2000+ model.

The training data for the message broadcast model consists of 298 sentences, representing the sentences of the main data set which are specific to pilot messages.

The training sentences are divided into 2485 *tri-grams*, with aircraft registrations having again the highest frequencies.

The test data contains 35 sentences, composed of 516 words, which also include sentences with a different construction than the training set. The resulting accuracy is 94%, with a perplexity of 24. The slight increase demonstrates that the reduction in the amount of data is reflected in the increase of performance of the model.

The training data for the LINK2000+ pilot model contains 900 sentences, containing only the downlink (aircraft to ATC) messages.

The test set consists of 23 sentences and 152 words. The resulting accuracy was 96%, proving again that the modeling accuracy is directly linked to the size of the modeling data.

6. Conclusion

The results of the research described above indicate that speech recognition represents a viable alternative to the classic voice radio communications that are used today in air traffic control.

Table 1

Modeling precision comparison

Model	Dictionary size	Training data (sentences)	Test data (sentences)	Accuracy
Broadcast-generic	650	632	67	93%
Broadcast-pilot	650	298	35	94%
LINK2000+-generic	134	1931	83	95%
LINK2000+-pilot	134	900	23	96%

Moreover, the accuracy levels obtained suggest that the optimum solution is building of separate, individual speech recognition systems for each user type, in order to reduce both the hardware and software requirements and to increase the system performance. Thus, the solution for the implementation of voice recognition systems in pilot-ATC communication is a voice command interface for the LINK2000+ system, developed specifically for pilots and air traffic controllers.

The main advantages in using speech recognition are the reduction in radio frequency usage, because the data link for this system does not require a high bandwidth. The messages can be broadcasted using the unique identification number of each word in the dictionary, further reducing the amount of data transmitted.

However, in order to use speech recognition in a safety-critical application such as ATC communications, the model has to be tested using acoustic data for training and testing and fine-tuned in order to obtain higher levels of accuracy.

Also, an error-handling system has to be designed in order to manage errors in the speech input or the recognition process itself. The basic error-handling system could be the visual validation of the recognized message by the sender which, in case of any misinterpretation, could repeat the message, correcting the information. Another possible correction mechanism involves a message validation algorithm that would present the speaker with a list of choices in case of ambiguities, enabling the correct transmission of the message.

In any case, the speech recognition system is not designed to completely replace the voice radio communication system, especially in time-critical situations, where stress, fatigue and other factors could affect the vocal characteristics of the speaker, reducing the performance of the system. In case of poor performance of the system or in time-critical situations, the voice radio can be used in order to ensure proper transmission and reception of messages.

Acknowledgment

The work has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/6/1.5/S/19.

The authors would also like to thank the anonymous reviewers for their observations and suggestions.

R E F E R E N C E S

[1] *Alicia Lechner, Kevin Ecker, Patrick Mattson*, „Voice recognition - software solutions in realtime ATC workstations"

[2] *R. Córdoba, J. Ferreiros, R. San-Segundo, J. Macias-Guarasa, J.M. Montero, F. Fernandez, L.F. D'Haro, J.M. Pardo*, "Air Traffic Control Speech Recognition System Cross-Task & Speaker Adaptation" in IEEE A&E Systems Magazine, September 2006, pp. 12-17

[3] *R. Córdoba, G. Prime, J. Macías-Guarasa, J.M. Montero, J. Ferreiros, J.M. Pardo*, "PPRLM Optimization for Language Identification in Air Traffic Control Tasks" in Eurospeech 2003, pp. 2685-2688

[4] *Fernando Fernández, Ricardo de Córdoba, Javier Ferreiros, Valentín Sama, Luis F. D'Haro*, "Language Identification Techniques based on Full Recognition in an Air Traffic Control Task" in ICSLP 2004, pp. II-1565-1568

[5] *Claudiu-Mihai Geacăr*, "Reducing pilot / ATC communication errors using voice recognition", in Proceedings of ICAS 2010, 2010

[6] *Jennifer Balogh, Jared Bernstein, Masa Suzuki, Matthew Lennig*, "Automatically Scored Spoken Language Tests for Air Traffic Controllers and Pilots" in VERSANT White Paper, Pearson Education, 2011

[7] *B. H. Juang; L. R. Rabiner,* "Hidden Markov Models for Speech Recognition" in Technometrics, **vol. 33**, No. 3, Aug., 1991, pp. 251-272.

[8] *** The HTK Book, User Guide, 2009.