World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:9, No:8, 2015

Possibilities, Challenges and the State of the Art of Automatic Speech Recognition in Air Traffic Control

Van Nhan Nguyen, Harald Holone

Abstract—Over the past few years, a lot of research has been conducted to bring Automatic Speech Recognition (ASR) into various areas of Air Traffic Control (ATC), such as air traffic control simulation and training, monitoring live operators for with the aim of safety improvements, air traffic controller workload measurement and conducting analysis on large quantities controller-pilot speech. Due to the high accuracy requirements of the ATC context and its unique challenges, automatic speech recognition has not been widely adopted in this field. With the aim of providing a good starting point for researchers who are interested bringing automatic speech recognition into ATC, this paper gives an overview of possibilities and challenges of applying automatic speech recognition in air traffic control. To provide this overview, we present an updated literature review of speech recognition technologies in general, as well as specific approaches relevant to the ATC context. Based on this literature review, criteria for selecting speech recognition approaches for the ATC domain are presented, and remaining challenges and possible solutions are discussed.

Keywords—Automatic Speech Recognition, ASR, Air Traffic Control, ATC.

I. Introduction

STEADILY increasing levels of air traffic world wide poses corresponding capacity challenges for air traffic control services. According to the "Outlook for Air Transport to the Year 2025" report of International Civil Aviation Organization (ICAO) [55], passenger traffic on the major international routes is expected to grow about 3 to 6 percent each year through to the year 2025. Thus, ATC operations has to investigate, review and improve in order to be able to meet with the increasing demands [9]. In ATC operations, communication between controllers and pilots is one of the key components. The quality of this communication significantly affects the performance as well as the safety of ATC operations.

Integration of automatic speech recognition (ASR) technologies in the ATC domain has been investigated in order to improve the performance of controller-pilot communications and to increase the automation of ATC systems. The introduction of automatic speech recognition to ATC and the steadily improvement in accuracy and performance of ASR technologies have opened many potential opportunities to investigate, review and improve ATC operations. For example, facilitating applications such as simulating the work environment of controllers for testing and training, controller workload measurement and balancing, assistant systems that support controllers in operational environment by catching potential dangerous situations that might be missed by the controllers, and providing suggestions as well as safety information to the operators.

Automatic speech recognition (ASR) technology, which is capable of translating human speech into sequences of words, has advanced significantly over the past decades. By 2015, ASR technologies has been successfully used in many applications like dictation, command and control, voice user interfaces such as voice dialing or call routing, medical applications, personal assistants on mobile phones, home automation, and automatic voice translation into foreign languages [52].

However, integrating ASR technologies into the ATC domain comes with many challenges such as call sign detection, poor input signal quality, the problem of ambiguity and the use of non-standard phraseology which dramatically reduce the recognition rate and the performance of speech recognition systems. Although the integration of ASR technologies into the ATC domain was introduced in the early 90s (or earlier) [30], it still has not been able to provide acceptable results in terms of recognition rate and overall performance.

With the aim of providing a comprehensive overview of current state-of-the-art speech recognition technologies, challenges as well as possibilities for applying ASR in the ATC domain, we have conducted a thorough literature review.

Based on the literature we identify five major existing challenges which make the integration of ASR technologies to the ATC domain difficult, and suggest possible approaches to address the challenges and improve the recognition rate of ASR systems in the ATC domain. Criteria for selecting ASR systems which well suited for use in ATC domain were also identified. The main contribution of this paper is to provide a fundamental starting point for researchers who are interested in integrating ASR systems in the ATC domain for both operational and simulation environments.

The remainder of the paper is structured as follows: Section II describes the methodology for conducting the literature review, before we present general introduction to automatic speech recognition, classification of ASR approaches as well as history of the field in section III. In Section IV we presents a brief introduction to air traffic control, possible applications of ASR in the ATC domain, criteria for selecting ASR approaches for the ATC domain, and an extended literature review of ASR research relevant to ATC. Finally, in Section V and Section

Authors are with Faculty of Computer Science, Østfold University College, PO Box 700, 1757 Halden, Norway, emails: nhan.v.nguyen@hiof.no, h@hiof.no.

VI we identify remaining challenges for ASR in ATC, discuss possible solutions to these challenges, and conclude the paper with a summary and outlook for this field.

II. Methodology

The literature review was conducted using the following keyword phrases: "Speech Recognition in Air Traffic Control OR Voice Recognition in Air Traffic Control", "Speech Command Recognition OR Voice Command Recognition", and "Medium Vocabulary AND Continuous Speech Recognition AND Speaker Independent". Searches were performed in ACM Digital Library, IEEEXplore Digital Library, Google Scholar and Google Search. From the search results we identified and reviewed 60 papers that focus on speech command recognition systems, the use of medium sized vocabularies, continuous speech, and speaker independent recognition, as well as speech recognition specifically in the context of air traffic control.

The purpose of including the last keyword phrase "Medium Vocabulary AND Continuous Speech Recognition AND Speaker Independent" is to capture articles about speech recognition techniques well suited for use in air traffic control (See Section IV for more details).

TABLE I
Search results summary. Keyword Phrase 1: "Speech Recognition in Air Traffic Control OR Voice Recognition in Air Traffic Control", Keyword Phrase 2: "Speech Command Recognition OR Voice Command Recognition", Keyword Phrase 3: "Medium Vocabulary AND Continuous Speech Recognition AND Speaker Independent"

Keyword Phrase \ Search Engine	ACM	IEEE	Google Scholar	Google
Keyword Phrase 1	2	4	12	9
Keyword Phrase 2	1	7	10	9
Keyword Phrase 3	4	6	13	10

The literature review provides background for identification of suitable speech recognition systems for air traffic control, as well as a discussion of remaining challenges and possible solutions for these types of applications.

III. Automatic Speech Recognition (ASR)

Speech recognition is the process of converting a speech signal into a sequence of words. It also called Automatic Speech Recognition (ASR) or Speech-to-Text (STT). In recent years, the technology and performance of speech recognition systems have been improving steadily. This has resulted in their successful use in many application areas such as in-car systems or environment in which users are busy with their hands (e.g., "voice user interfaces") [34], hospital-based healthcare applications (e.g., systems for dictation into patient records, speech-based interactive voice response systems, systems to control medical equipment and language interpretation systems) [15], home automation (e.g., voice command recognition systems) [1], speech-to-text processing (e.g., word processors or emails), and personal assistants on mobile phones (e.g., Apple's Siri on iOS, Microsoft's Cortana on Window Phone, Google Now on Android). Speech recognition has also been widely used in air traffic control for many applications such as air traffic controllers' work load measurement [10], speech interface for air traffic control terminals [20], automated analysis and transcription of ATC voice communications [9], replacing the "pseudo-pilot" in air traffic control simulation and training by "automated pilot" which can recognize and understand the controller's speech using speech recognition modules [45].

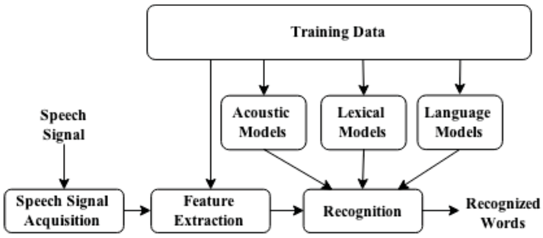

Fig. 1. General structure of speech recognition system

A. Modules of Speech Recognition Systems

The general speech recognition approach can be described in two steps. 1) Given an acoustic observation, identify a feature vector sequence $X = X_1, X_2, ..., X_n$ using a feature extraction module. 2) Given this vector, find the corresponding word sequence $W = W_1, W_2, ..., W_n$ that has the maximum posterior probability $P(W \mid X)$ [35], expressed using Bayes theorem in (1).

$$W = \underset{w}{\operatorname{argmax}} P(W \mid X) = \underset{w}{\operatorname{argmax}} \frac{P(W)P(X \mid W)}{P(X)} \quad (1)$$

Fig. 1 shows the general structure of a speech recognition system. The system consist of six main modules: Speech Signal Acquisition, Feature Extraction, Acoustic Modeling, Language Modeling, Lexical Modeling, and Recognition.

1. Signal Acquisition: The signal acquisition module is responsible for obtaining the speech signal to be analyzed, for example by using microphones.

2. Feature Extraction: The feature extraction module is responsible for converting the speech signal into a feature vector. The performance of the ASR system depends heavily on this process. There are many feature extraction techniques such as Principal Component Analysis (PCA), Mel Frequency Cepstral Coefficients (MFCC), Independent Component Analysis (ICA), Linear Predictive Coding (LPC), Autocorrelation Mel Frequency Cepstral Coefficients (AMFCCs), Relative Autocorrelation Sequence (RAS), and Perceptual Linear Predictive Analysis (PLP). [23], [28], [57]. Studies have shown that Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC) are techniques extensively used in speech recognition [52].

3. Acoustic Models: The acoustic model plays a critical role in improving accuracy of the ASR system by linking the input features with the expected phonetics of the hypothesis sentence [28] [35]. In (1), P($X \mid W$) represents the acoustic model, which is is the probability of acoustic observation of X when the word W is uttered.

4. Language Models: The main task of a language model is detecting connections between the words in a sentences with the help of lexical models. ASR systems usually use an n−gram language model to provide context for distinguishing words and phrases that sound similar. The use of a language model not only makes speech recognition more accurate but also helps to reduce the search space for recognition [35]. In (1), $P(W)$ represents the language model, which is the probability of word W uttered.

5. Lexical Models: A lexical model is also known as a pronunciation dictionary. It is developed to provide pronunciations of words in a given language. The lexical model links the acoustic-level representation with the word sequence which is output by the speech recognizer [7].

6. Recognition: The recognition module takes input from the feature extraction module and then uses acoustic models, language models and lexical models to recognize which words were spoken.

B. *Classification of Speech Recognition Systems*

Speech recognition systems can be classified by type of speech utterance, type of speaker model and type of vocabulary that the systems can recognize [52].

1. Types of Speech Utterance: In ASR, an utterance is the smallest unit of speech and it is the sound of a word or set of words. Types of utterance can be classified into four classes as follows:

- Isolated Words - according to Radha et al., “isolated word recognizers usually require each utterance to have quiet on both sides of the sample window. It doesn’t mean that it accepts single words, but does require a single utterance at a time” [52]. It is also known as “Isolated Utterance”. This type of speech recognizer is comparatively simple and easy to develop because word boundaries are obvious.
- Connected Words - connected word recognizers are quite similar to isolated word recognizers, but require smaller pauses between utterances. It also known as “connected utterances”.
- Continuous Speech - continuous speech recognizers require special techniques for determining utterance boundaries, and allow speakers to speak almost naturally [52]. Although this kind of system is very difficult to develop, it has been widely used in many applications because of its flexibility.
- Spontaneous Speech - spontaneous speech recognizers are capable of recognizing unrehearsed speech, words being run together, “ums” and “ahs”, and even slight stutters [52]. Because of the large linguistic variation of spontaneous speech, recognition is extremely difficult. However, it has been shown that acoustic and language models with very large training data sets are able to overcome the problem of variation to some degree. This has resulted in increased recognition rates in spontaneous speech recognition systems [22].

Speech recognition systems for isolated words and connected words are considered relatively easy to develop because word boundaries are easy to find and the pronunciation of a word tends not to affect others. In contrast, continuous speech and spontaneous speech is more difficult to handle for a number of reasons. Challenging aspects of this type of ASR includes word boundary detection, the problem of coarticulation, and varying speech rates.

2. Types of Speaker Models: Because of the uniqueness physical bodies and personalities among people, speakers usually have distinct voice characteristics. ASR speaker models can be divided into two classes depending on how they handle these differences; speaker dependent and speaker independent models. [52].

- Speaker Dependent Models - speaker dependent systems depends on knowledge of a specific speaker’s voice characteristics. This kind of system must usually be trained for a specific user before it can recognize the speech of the user. Although these systems are easy to develop and achieve high accuracy, they are not used widely because they are usually not as flexible as speaker adaptive or speaker independent systems.
- Speaker Independent Models - speaker independent systems does not require knowledge of specific speakers, and can recognize speech from practically any people speaking a given language. Apple’s Siri assistant is an example of a system using a speaker independent model. Compared with speaker dependent systems, these systems are more flexible, however they offer less accuracy and are more difficult to develop.

Speaker dependent systems are commonly used for speech-to-text software (e.g., word processors, emails and dictation applications), while speaker independent systems are more commonly found in telephone applications (e.g., call centers). There is a third type of speaker model called a speaker adaptive model. These systems are developed to adapt its operation to the characteristics of new speakers. Implementing speaker adaptive systems is more complex than speaker dependent systems, but easier than the use of speaker independent models.

3. Types of Vocabulary: Another distinguishing factor of ASR systems is the size of the vocabulary they are able to recognize. The size of vocabulary affects the complexity, performance and the accuracy of the system [52]. In the literature, these vocabularies are usually classified into five classes as follows:

- Small vocabulary - tens of words
- Medium vocabulary - hundreds of words
- Large vocabulary - thousands of words
- Very-large vocabulary - tens of thousands of words
- Unlimited vocabulary - the system is able to suggest recognized words based on the phonems even when the word is not found in the (very large) vocabulary.

Generally, the smaller the vocabulary the easier it is to implement the ASR system.

C. *Performance of Speech Recognition Systems*

Accuracy and speed are the two most common metrics for measuring speech recognition system performance. Word Error Rate (WER) is usually used for measuring accuracy, whereas speed is usually rated with Real Time Factor (RTF) [52]. WER can be computed by using (2):

$$WER = \frac{S + D + I}{N} \quad (2)$$

Where S is the number of substitutions, D is the number of deletions, I is the number of insertions and N is the number of words in the reference.

If the input of duration I requires time P to process, RTF can be computed by using (3):

$$RTF = \frac{P}{I} \quad (3)$$

Other measures of performance include Concept Error Rate (CER), Single Word Error Rate (SWER) and Command Success Rate (CSR).

D. History of Automatic Speech Recognition

The history of ASR started in 1952 with an isolated digit recognition system for a single speaker. It was built by Davis, Biddulph, and Balashek of Bell Laboratories [11]. Over the last 60 years, technology development has led to a dramatic improvement of speech recognition systems. Juang and Rabiner [39] describes the development during the first four decades:

- 1960's - speech recognition systems were able to recognize small vocabularies (10 - 100 words) of isolated words with the help of filter-bank analyses and simple time normalization methods.
- 1970's - by using simple template-based, pattern recognition methods, researchers were able to build connected words, speaker independent speech recognition systems which can recognize medium vocabularies (100 - 1000 words).
- 1980's - large vocabulary (1000 - unlimited number of words) further advances in speech recognition problem was addressed using Hidden Markow Models (HMM) and stochastic language models.
- 1990's - with the helps of stochastic language understanding, statistical learning of acoustic and language models, and finite state transducer framework (and the FSM Library), researchers were able to build large vocabulary systems for continuous speech recognition and understanding.

In beginning of the new millennium, speech recognition systems were expanded to recognize very large vocabularies [52] [51]. Spontaneous speech recognition has started to receive attention from many researchers. In addition, researchers have started to use multimodal speech recognition, in which visual face information, particularly lip information is utilized. Results from multimodal speech recognition research show that performance can be improved compared with using audio only [21].

Currently (2015), we are able to build unlimited vocabulary speech recognition systems which can solve a large number of tasks, including the multiple languages problem [36], [51], [52]. Although artificial neural networks has been explored since the 1980's, they have so far not been able to compete with the Gaussian Mixture Model/Hidden Markov Model (GMM-HMM) approaches, which continues to be the dominating approach [13]. Nowadays, the introduction of deep learning [14], [32] and hybrid approaches [29], [67], [68] has overcome most of these difficulties and significantly increased the recognition rate of ASR systems.

IV. Air Traffic Control (ATC)

A. Introduction to Air Traffic Control

According to the Oxford English Dictionary, Air Traffic Control (ATC) is "the ground-based personnel and equipment concerned with controlling and monitoring air traffic within a particular area" [60]. The main purpose of ATC systems is to prevent collisions, provide safety, organize aircraft operating in the system and expedite air traffic [18]. With the steady increase in air traffic, ATC has become more and more important. This increase has also resulted in more complex procedures, regulations and technical systems [54].Thus, air traffic control systems have to be continuously improved to meet the evolving demands in air traffic.

In ATC, air traffic controller (ATCO) have an incredibly large responsibility for maintaining the safe, orderly and expeditious conduct of air traffic. Given the important roles of air traffic control and air traffic controllers, there is an ongoing need to strengthen training and testing of the operators. Further, being able to simulate the working environment of controllers enables increased safety through the use of support systems that can assist controllers and improve procedures, and by analyzing controller-pilot communications. In the past few years, the advances in technology and performance of ASR systems has offered many promising ways to deal with these needs.

B. Applications of ASR in ATC

Because voice communication plays a critical role in ATC, many researchers have been interested in using automatic speech recognition technology for various applications in ATC operations as well as for simulation environments [41].

1. Air Traffic Control Simulation and Training: Air traffic control simulation provides facilities for testing and evaluation of new systems and concepts, and training of traffic controller students to handle realistic scenarios. Current air traffic control simulation typically requires "pseudo-pilots" who will act as real pilots in the simulation of controller-pilot communications with air traffic controller students. The use of "pseudo-pilots" make air traffic control simulators less flexible and comes at a relatively high cost.

By introducing speech technologies in ATC simulation and training the "pseudo pilots" can be replaced with so-called "automated pilots". The "automated pilot" will understand and process air traffic controllers' speech using a speech recognition module and generate responses that is sent back to the controllers using a speech synthesis module. The use of "automated pilot" instead of "pseudo-pilot" can dramatically reduce the cost of ATC systems and make the systems more flexible [61].

2. Air Traffic Controllers Workload Measurement and Balancing: In ATC systems, air traffic controller workload is the key factor that limit the capacity of the whole system. With the increase in air traffic, measuring and balancing air traffic controller workload becomes important.

However, measuring controller workload is currently not an easy task because workload is difficult to measure directly. It is a costly process that requires manual observation and analysis of spoken communication. With the help of ASR systems, detecting spoken control events that the controller has to perform becomes easier, thus facilitating more direct measurements of controller workload. The detected events can be used for automated controller workload balancing [9], [10].

3. Controller-pilot Speech Analysis and Transcription: With the help of ASR systems in transcribing controller-pilot communications, it is possible to analyze large quantities of voice data for ATC research and analysis [41]. This analysis can be used for investigating and improving procedures and regulations, detecting air traffic controllers' events for workload measurement and balancing of controller workloads.

4. Backup Controller: An ASR system combined with other information sources in the ATC context (e.g., radar information, minimum safe altitudes, restricted zones, and weather information) could be used as input for a system called a "backup controller" to catch potentially dangerous situations that might be missed by the controller. It can also provide suggestions and safety information to the controllers in real time [41], [65].

C. Criteria for Selecting ASR Systems for ATC

Applying automatic speech recognition in the ATC domain comes with many challenges and opportunities because of the unique characteristics of communication between controllers and pilots, such as small vocabulary sizes, high accuracy requirements, close to real time demands, and standardized formats for communication [41]. Based on these characteristics, studies has suggested that an ASR system that is suitable for ATC should be a speaker independent system which can recognize medium sized vocabularies and continuous speech [54], [65].

1. Speaker Dependence: Although Air Traffic Control Command Recognition (ATCCR) applications require only one controller at the same time, there are situations where multiple controllers are required in the operational environment.

Additionally, in the context of simulation and training, the system has to be able to recognize many air traffic controller students without the requirement to retrain or reconfigure the system. Thus, speaker independent systems are best suited for these applications, despite the reduced recognition accuracy of such systems [37], [65].

2. Continuous Speech Recognition: Although isolated words and connected words recognition systems usually have higher accuracy than continuous speech recognition systems, they are not well suited in the context of ATC. This is because they require the controllers to pause between each word when giving commands. Isolated words and connected words recognition systems will therefore cause delay in pilot-controller communication. A continuous speech recognition system, which permits the controller to speak in a natural way without pauses [54], is the system of choice when applying ASR in ATC [37].

3. Vocabulary Size: In the ATC domain, vocabularies used in communication between controllers and pilots follows International Civil Aviation Organization (ICAO) Standard Phraseology. The entire vocabulary of words (excluding names of specific places and call signs) is only about a few hundred words [65] [37] [17]. Thus, a medium sized vocabulary speech recognition system is adequate in the context of air traffic control.

4. Performance: In ATC, it is not important that ASR systems can recognize every single word, but it is important that the conveyed concepts are correctly detected. For example, the ASR system is not required to recognize all of the words in the following sentence: "Good morning Lufthansa one zero one descend level one two three", however it has to be able to extract the concept "DLH101 DESCEND FL 123".

The Concept Error Rate (CER) metric is used to measure the systems ability to extract the concepts from speech [30]. The CER of an ASR system which can be applied in ATC should not exceed those of pilots or pseudo pilots, which is 0.73% [54]. In addition, the system should be able to recognize and understand the concepts in real time without causing delays in communication between controllers and pilots or pseudo pilots.

D. State-of-the-art of ASR suitable for use in ATC

Based on the previously mentioned criteria for selecting ASR system for the ATC domain, the number of suitable systems is limited. In this section, we highlight progress made so far for ASR systems that match these criteria. Although some of the systems were not developed for ATC or the English language, the approaches and technologies of the systems are still applicable to the ATC domain. The research presented in this section are grouped into three: The Hidden Markov Model approach, hybrid approaches and other approaches.

1. The Hidden Markov Model (HMM): In ASR, HMM has been the dominant approach over the last two decades.

Although the method has it's own weaknesses, it is still popular because it can be trained automatically, it is simple and computationally feasible.

In 1994, Daniel Jurafsky et al. used HMM combined with a Viterbi decoder, a bigram language model and a phonetic likelihood estimator to develop the Berkeley Restaurant Project (BeRP), which is a medium-vocabulary, speaker-independent, spontaneous continuous speech recognition system which functions as a knowledge consultant [40]. The recognition error rate and understanding error rate were quite high at 32.1% and 34% respectively.

Three years later, Jones et al. developed a continuous speech recognition system using syllable-based HMMs [38]. The authors concluded that the introduction of syllable-level bigram probabilities, word- and syllable-level insertion penalties, and the investigation of different model topologies can improve the recognizer performance. Compared with 35% of the baseline accuracy for monophone recognition, the proposed system achieved over 60% recognition accuracy.

Recognition of non-English languages have also been investigated by many ASR researchers, including Arabic, Tamil, Estonian, Amharic and Malayalam.

An acoustic training system for building acoustic models for a medium vocabulary speaker independent continuous speech recognition system for the Arabic language was developed Nofal et al. [47]. Cross-word triphones HMMs were used for acoustic modeling, and the models were trained using maximum likelihood estimation. The best word error rate was 0.19%.

A continuous speech recognition system for the Tamil language using a monophone-based HMM was developed by Radha et al. in 2012 [53]. The system used Mel Frequency Cepstral Coefficients (MFCC) for feature extraction. The results were relatively good, with the system yielding 92% word recognition accuracy and 81% sentence accuracy.

Thangarajan et al. built a small vocabulary word based and a medium vocabulary triphone based continuous speech recognizers for the Tamil language using HMM based word and triphone acoustic models [62]. 92.06% and 70.08% accuracy were achieved with new speakers on test sentences for the word-model and triphone-model respectively.

Thangarajan et al. used syllable modeling for developing a continuous speech recognition system for the Tamil language [63]. A small vocabulary context independent word model and medium vocabulary context dependent phone model were developed. The models were trained using SphinxTrain, a HMM-based acoustic model trainer from Carnegie Mellon University (CMU) [58]. The Word Error Rate of the proposed system was 10.63%.

A limited-vocabulary Estonian continuous speech recognition system using HMM was proposed by Alumäe et al [2]. Clustered triphones with multiple Gaussian mixture components were used to model words. The recognizer yielded 82.9% accuracy with a medium-sized vocabulary. If the real-time requirement was discarded, the correctness increased to 90.6%.

Although HMM has been the dominant technique for acoustic modeling in speech recognition for over two decades, it has two main weaknesses: it discards information about time dependencies, which creates problems for recognizing speech with varying speeds, and is prone to overgeneralization. De Wachter et. al (2007) [12] attempted to overcome these problems by relying on straightforward template matching. The authors extended the Dynamic Time Warping (DTW) framework with a flexible subword unit mechanism and a class sensitive distance measure. This resulted in an error rate reduction of 17% compared to the HMM results.

Gebremedhin et al (2013) built a syllable based, medium vocabulary size, continuous Amharic speech recognition for weather forecast and business report applications based on HMM [27]. To do this, they introduced a new approach for reducing the number of acoustic models that are required to build a syllable based Amharic ASR by combining similarly pronounced syllables. Finite state transducers were also explored to specify the grammar rules. The recognition accuracy of 93.6% was achieved on a 4000 words test set.

Kurian and Balakriahnan developed a continuous speech recognition system for the Malayalam language using PLP (Perceptual Linear Predictive) Cepstral Coefficient [42]. The developed system was evaluated with different number of states of HMM, Gaussian mixtures, and tied states. The word recognition accuracy and sentence recognition accuracy were 89% and 83% respectively.

Edward C. Lin implemented a 1000-word vocabulary, speaker independent, continuous live-mode speech recognizer in a single FPGA (A field-programmable gate array) [44]. A 4-state HMM is used to represent triphones in the implemented system. Although the implementation is extraordinarily small, it can still achieve almost the same accuracy as the state-of-the-art software recognizer at 10.9% Word Error Rate.

In order to address the problem of automatic speech recognition in the presence of interfering noise, Gales et al. developed a robust continuous speech recognition system using parallel model combination [24]. The model used in the system is a standard HMM with Gaussian output probability distributions.

Novotnŷ et al. developed a speech command recognition system using hidden Markov models of context dependent phones (triphones) and mel-frequency cepstral coefficients analysis of speech (MFCC) [48].

Although HMM-based ASR systems have not achieved the required accuracy in the ATC domain (0.73% CER), the steady improvement in term of accuracy and performance makes HMM-based ASR systems potential candidates for use in ATC. Approaches facilitated by the characteristics of the ATC domain can be applied to improve the accuracy of the systems in order to achieve the required results.

2. Hybrid Approaches: Although HMM is the dominant method for speech recognition over the last two decades, it still has it's weaknesses. Many research initiatives have been conducted to overcome those weaknesses, for instance by proposing hybrid approaches. Combining HMM and Artificial Neural Networks (ANN) is a new research area that has received focus from many researchers. A survey of hybrid ANN/HMM models for automatic speech recognition was conducted by Edmondo Trentin et al. [64].

Hussien Seid et al. developed an Amharic speaker independent continuous speech recognizer based on an HMM/ANN hybrid approach [56]. With the help of the CSLU Toolkit [33], the model was constructed at a sub-word level using context dependent phonemes. This resulted in the achievement of 74.28% word and 39.70% sentence recognition.

Shantanu Chakrabartty et al (2000) proposed a hybrid Support Vector Machine (SVM), Hidden Markov Model approach for continuous speech recognition [6]. The architecture of the proposed system is based on the MAP (maximum a posteriori) framework [25].

Wroniszewska et al developed a voice command recognition system based on the combination of genetic algorithms (GAs) and K-nearest neighbor classifier (KNN). 94.2% recognition rate was achieved [67].

The ability to overcome the existing weaknesses of HMM and the improvement in terms of accuracy and performance with hybrid speech recognition systems makes this a good candidate for applications in environments like ATC.

3. Other Approaches: Although it has been proven that Support Vector Machines (SVM) have problems which make them difficult to apply to speech recognition, Padrell-Sendra et al. proposed a pure SVM-based continuous speech recognizer, using the SVM to make decisions at frame level, and a Token Passing algorithm to obtain the chain of recognized words [49]. The proposed system achieved a better recognition rate than traditional HMM-based systems (96.96% vs 96.47%).

Pellom et al. proposed fast likelihood computation techniques in nearest-neighbor based search for continuous speech recognition systems [50]. The authors concluded that the combination of the two techniques with partial distance elimination (PDE) reduced the computational complexity for likelihood computation by 29.8% over straightforward likelihood computation.

Leung et al. proposed a neural fuzzy network and genetic algorithm approach for Cantonese speech command recognition [43].

Beritelli et al. (2006) proposed a noise robust, low-complexity algorithm for voice command recognition using Vector Quantization-Weighted Hit Rate (VQWHR) and Dynamic Time Warping (DTW). The authors concluded that the proposed algorithm was robust to various types of background noise [4].

Although HMM and hybrid approaches have been used very widely for speech recognition, they are still facing challenges like computational complexity and background noise. Approaches such as Support Vector Machines or combinations of Vector Quantization-Weighted Hit Rate (VQWHR) and Dynamic Time Warping can deal with those challenges to some degree, and is still being explored by many researchers.

The following discussion is based on the state-of-the-art of ASR presented in section III and the specifics of the ATC context presented in section IV.

V. Discussion

The discussion is divided in two. First, we identify challenges of applying automatic speech recognition (ASR) in the ATC domain, and second we suggests possible approaches which can be used to address the challenges and improve the recognition rate of ASR systems in general.

A. Challenges of ASR in ATC

There are five major challenges to overcome in order to successfully apply ASR in ATC. While some challenges are unique to the ATC domain, such as call sign detection and the use of non-standard phraseology, others are general challenges of ASR systems such as poor input signal quality, the problem of ambiguity, and the use of dialects, accents and multiple languages. The latter challenges becomes even more pronounced when ASR is introduced to a high-risk domain such as ATC.

1. Call Sign Detection: Because of the variety of ways to refer to the same flight call sign and the use of airline aliases (e.g., “Speedbird” for British Airways Plc (United Kingdom), “Norstar” for Norwegian Long Haul (Norway), “Pacific” for Jetstar Pacific Airlines (Vietnam)), call sign detection is an extremely challenging task of ASR in ATC domain. It especially increases the CER (Concept Error Rate), but also affects WER (Word Error Rate) because of the requirement to identify all airline aliases, and then train the system for these alternative names.

2. Poor Input Signal Quality: The input signal quality can be affected by both technological and human factors. While technological problems such as background noise in cockpits and communication via radio links physically reduce the quality of input signal, human related problems such as spontaneous speed, high speed and/or slurred speech increase ambiguity in the ASR process. Both factors lead to increased misrecognition rates. The above-mentioned technological problems can to a certain degree be resolved by using noise canceling microphones and high quality radio links. However, solving human related problems would be very challenging because it is not likely that we can force controllers and pilots to significantly change the way they speak in order to adapt to ASR systems.

3. The Problem of Ambiguity: In the ATC domain, the problem of ambiguity (e.g., the number two-four-five can refer to a speed, heading or flight level) and references to confusable entities such as call signs or flight levels is one of the main factors which contribute to the reduction in the speech recognition rate of ASR systems [41], especially with regards to CER.

4. The Use of Non-Standard Phraseology: The use of non-standard phraseology leads to errors in controller-pilot radio messages. Studies have shown that about 80% of all pilot radio messages contain at least one error [26]. In addition, only a small number (less than 30%) of the examined utterances fully conform to the ICAO recommended phraseology [31]. This starting point adds to the difficulty of introduction of ASR in ATC.

5. Dialects, Accents and Multiple Languages: Because Air Traffic Control services are global services, ASR systems must be able to recognize foreign accents, different dialects and commands with a combination of multiple languages. For example, German controllers may say “Guten morgen Lufthansa one two three descend level one two zero”, where “Guten morgen” is good morning in German.

B. Approaches that can be used to improve the accuracy of ASR in ATC

Although the ATC context poses many challenges to ASR systems, it also offers many distinct opportunities such as the use of context knowledge, the structured format of controller-pilot communications, and small vocabulary sizes. The use of post-processing approaches to reduce the uncertainties and ambiguities which resulted from the speech recognition process in order to improve recognition accuracy is a very well-known approach in ASR. There are three main post-processing approaches that are well suited for the ATC domain; syntactic analysis, semantic analysis and pragmatic analysis.

1. Syntactic Analysis: Syntactic analysis is the process of representing the language domain of the speech recognition system by a grammar, and then parsing inputs to eliminate invalid words or sentences [46]. Finite State Networks, Augmented Transition Networks (ATNs) and heuristics are the three methods that can be used to implement syntax in ASR [37].

In the ATC domain, syntactic analysis can be performed with the help of grammar files, which is made easier because of the structured format of controller-pilot communications and the predefined vocabularies. These grammar files define structure of sentences used in the operation. By using these grammar files, improved recognition can be achieved by focusing on the words likely to be spoken next in a sentence.

The ASR system use the grammar files to compile lexical trees which will be used recognize a statement by parsing the tree. For example, the simplest form of an ATC command consists of a call sign (e.g., SpeedBird, Norstar) followed by a goal action (e.g., descent, heading, fly direct) and a goal value (e.g., FL 90, 260 (degrees)) [17]. After a call sign is detected, the speech recognition system should expect to find a goal action. Thus, words which are not goal actions (e.g.,“Ahs”, ‘Ums”) can be eliminated through the syntactic analysis.

The ability to eliminate invalid words and sentences of syntax analysis offers great potential to address the poor quality input signal challenge. In addition, syntax analysis can be used to deal with the problem of ambiguity.

With the help of the list of known ATC vocabularies, syntax analysis is able to correct misrecognized words, for example due to the problem of ambiguity, by replacing them with valid words with similar pronunciation.

2. Semantic Analysis: Semantic analysis is the process of testing the meaningfulness of sentences recognized by a speech recognition system. The method has been used to improve speech recognition performance by many researchers [8], [16], [69]. In the ATC domain, semantic analysis can be performed with the help of grammar files. Semantic knowledge is static, so it can be obtained and implemented into the syntax.

One possible method of using context knowledge is N-best list. The speech recognizer first analyzes the input signal and transforms it into a N-best list, and the list is then reduced by eliminating word sequences that parse syntactically, but are not actually meaningful [30] [37].

Because semantic analysis have the ability to eliminate words and sentences which are not meaningful even when they are parsed syntactically, it can be used to assist syntax analysis in dealing with the problem of ambiguity, poor input signal quality, and even the use of non-standard phraseology.

3. Pragmatic Analysis: Pragmatic analysis is the process of predicting likely future words based on the previously recognized words and the state of the system [37] [59].

A few methods exist that can be used to perform pragmatic analysis in the ATC domain with the help of context knowledge. One example is work by Schaefer, who developed a context-sensitive speech recognition system for air traffic control simulation using a cognitive model of the ATC controller. The model can continuously observe the present situation and generate a prediction of sentences the controller is most likely to say next [54].

Further, by using a so-called “Dialog Model” combined with context knowledge, the ASR system is able to predict the form and content of the next utterance from the previously recognized utterances [66]. The dialog models allow the system to consider only a subset of the application’s full grammar and vocabulary, so both performance and accuracy of the ASR system can be improved.

In addition, radar information and flight plans could be used to reduce the list of likely aircraft call signs that a controller may refer to in a sector to only those in the sector or about to enter the sector [41]. With the ability to reduce the list of likely aircraft call signs, pragmatic analysis can be used to mitigate the challenge in call sign detection.

Finally, knowledge based rules, Finite State Networks, and knowledge state databases can also be used to implement pragmatic analysis in the ATC domain.

4. Other Approaches: Although, the three suggested post-processing approaches cannot address all the ATC challenges completely, they offer great potential to improve the recognition rate of ASR systems in the ATC domain.

Issues related to the use of dialects, accents, and multiple languages remain difficult to address. One possible way forward is to use detector modules for identifying which dialects, accents and languages which are spoken. This approach has been demonstrated by Fernandez et al. [19], who devised an ATC speech understanding system which can understand both English and Spanish. They achieved this by using a language detection module, which is capable of detecting the languages spoken by air traffic controllers. Detecting dialects and accents for tuning of a speech recognition system has been investigated by other researchers (see for example [3] and [5]).

VI. Conclusion

In this paper, we have presented a thorough review of the Automatic Speech Recognition literature, including a look at the research history, and a presentation of the state-of-the-art of ASR approaches.

Further, we have presented possible applications of ASR in air traffic control, and identified central criteria for ASR approaches applicable to the ATC domain.

Following a detailed review of current ASR research approaches, we identified existing challenges applications of ASR in ATC, and discussed possible solutions to these challenges.

Because of the operation critical nature of systems in the ATC domain, there are still challenges that remain before ASR systems can be applied fully both in training, testing and ATC operations. However, as we have pointed out in this paper, research is steadily providing better results, both in terms of accuracy and speed.

Combining state-of-the art ASR approaches with contextual information to include syntactic, semantic and pragmatic analysis in the recognition process, and the identification of dialects, accents and languages holds great promise for the application of automatic speech recognition in the air traffic control domain.

Acknowledgment

The authors would like to thank The Institute for Energy Technology (John E. Simensen and Christian Raspotnig), Edda Systems AS, and WP3 of "User driven innovation in smart homes for health and energy" at Østfold University College for support in the work with this paper.

References

[1] H. AlShu'eili, G. Sen Gupta, and S. Mukhopadhyay. Voice recognition based wireless home automation system. In *Mechatronics (ICOM), 2011 4th International Conference On*, pages 1–6, May 2011.

[2] Tanel Alumäe and Leo Võhandu. Limited-vocabulary estonian continuous speech recognition system using hidden markov models. *Informatica*, 15(3):303–314, 2004.

[3] Hamid Behravan. *Dialect and accent recognition*. PhD thesis, 2012.

[4] Francesco Beritelli and Salvatore Serrano. A robust low-complexity algorithm for voice command recognition in adverse acoustic environments. In *2006 8th International Conference on Signal Processing*, volume 3. IEEE, 2006.

[5] Fadi Biadsy. *Automatic dialect and accent recognition and its application to speech recognition*. PhD thesis, Columbia University, 2011.

[6] Shantanu Chakrabartty, Guneet Singh, and Gert Cauwenberghs. Hybrid support vector machine/hidden markov model approach for continuous speech recognition. In *Circuits and Systems, 2000. Proceedings of the 43rd IEEE Midwest Symposium on*, volume 2, pages 828–831. IEEE, 2000.

[7] Rahul Chitturi, Venkatesh Keri, Gopalakrishna Anumanchipalli, and Sachin Joshi. Lexical modeling for non native speech recognition using neural networks. In *Proceedings of the International Conference on Natural Language Processing (ICON–2005)*, page 79. Allied Publishers, 2005.

[8] Noah B. Coccaro. *Latent Semantic Analysis As a Tool to Improve Automatic Speech Recognition Performance*. PhD thesis, Boulder, CO, USA, 2005. AAI3190360.

[9] José Manuel Cordero, Manuel Dorado, and José Miguel de Pablo. Automated speech recognition in atc environment. In *Proceedings of the 2nd International Conference on Application and Theory of Automation in Command and Control Systems*, pages 46–53. IRIT Press, 2012.

[10] José Manuel Cordero, Natalia Rodríguez, José Miguel, and Manuel Dorado. Automated speech recognition in controller communications applied to workload measurement. *Third SESAR Innovation Days*, 2013.

[11] KH Davis, R Biddulph, and Stephen Balashek. Automatic recognition of spoken digits. *The Journal of the Acoustical Society of America*, 24(6):637–642, 1952.

[12] M. De Wachter, M. Matton, K. Demuynck, P. Wambacq, R. Cools, and D. Van Compernolle. Template-based continuous speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(4):1377–1390, May 2007.

[13] Li Deng, Khaled Hassanein, and M Elmasry. Analysis of the correlation structure for a neural predictive model with application to speech recognition. *Neural Networks*, 7(2):331–339, 1994.

[14] Li Deng, Geoffrey Hinton, and Brian Kingsbury. New types of deep neural network learning for speech recognition and related applications: An overview. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 8599–8603. IEEE, 2013.

[15] Scott Durling and Jo Lumsden. Speech recognition use in healthcare applications. In *Proceedings of the 6th international conference on advances in mobile computing and multimedia*, pages 473–478. ACM, 2008.

[16] Hakan Erdogan, Ruhi Sarikaya, Stanley F Chen, Yuqing Gao, and Michael Picheny. Using semantic analysis to improve speech recognition performance. *Computer Speech & Language*, 19(3):321–343, 2005.

[17] Eurocontrol. All clear? the path to clear communication. icao standard phraseology a quick reference guide for commercial air transport pilots. http://www.skybrary.aero/bookshelf/books/115.pdf, 2011.

[18] AJV-0 VP Mission Support Federal Aviation Administration. Air traffic control - chapter 2. general control, faa 7110.65 2-1-1. Technical report, February 19, 2014.

[19] F Fernández, J Ferreiros, JM Pardo, V Sama, R de Córdoba, J Marias-Guarasa, JM Montero, R San Segundo, LF d'Haro, M Santamaría, et al. Automatic understanding of atc speech. *Aerospace and Electronic Systems Magazine, IEEE*, 21(10):12–17, 2006.

[20] J. Ferreiros, J.M. Pardo, R. de Crdoba, J. Macias-Guarasa, J.M. Montero, F. Fernndez, V. Sama, L.F. d'Haro, and G. Gonzlez. A speech interface for air traffic control terminals. *Aerospace Science and Technology*, 21(1):7 – 15, 2012.

[21] Sadaoki Furui. 50 years of progress in speech and speaker recognition. *SPECOM 2005, Patras*, pages 1–9, 2005.

[22] Sadaoki Furui, Masanobu Nakamura, Tomohisa Ichiba, and Koji Iwano. Why is the recognition of spontaneous speech so hard? In *Text, Speech and Dialogue*, pages 9–22. Springer, 2005.

[23] Santosh K Gaikwad, Bharti W Gawali, and Pravin Yannawar. A review on speech recognition technique. *International Journal of Computer Applications*, 10(3):16–24, 2010.

[24] M.J.F. Gales and S.J. Young. Robust continuous speech recognition using parallel model combination. *Speech and Audio Processing, IEEE Transactions on*, 4(5):352–359, Sep 1996.

[25] J. Gauvain and Chin-Hui Lee. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *Speech and Audio Processing, IEEE Transactions on*, 2(2):291–298, Apr 1994.

[26] Claudiu-Mihai Geacăr. Reducing pilot/atc communication errors using voice recognition. In *Proceedings of ICAS*, volume 2010, 2010.

[27] Yitagessu Birhanu Gebremedhin, Frank Duckhorn, Rüdiger Hoffmann, and Ivan Kraljevski. A new approach to develop a syllable based, continuous amharic speech recognizer. In *EUROCON, 2013 IEEE*, pages 1684–1689. IEEE, 2013.

[28] Wiqas Ghai and Navdeep Singh. Literature review on automatic speech recognition. *International Journal of Computer Applications*, 41(8):42–50, 2012.

[29] A. Graves, N. Jaitly, and A.-R. Mohamed. Hybrid speech recognition with deep bidirectional lstm. In *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*, pages 273–278, Dec 2013.

[30] Hartmut Helmke, Heiko Ehr, and Matthias Kleinert. Increased acceptance of controller assistance by automatic speech recognition. *Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013)*, 2013.

[31] Horst Hering. Technical analysis of atc controller to pilot voice communication with regard to automatic speech recognition systems. *EEC note*, 1, 2001.

[32] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE*, 29(6):82–97, 2012.

[33] John-Paul Hosom. The cslu toolkit: A platform for research and development of spoken-language systems. *Center for Spoken Language Understanding (CSLU), OGI Campus, Oregon Health & Science University (OGI/OHSU), visitado em Janeiro de*, 2002.

[34] Zhang Hua and Wei Lieh Ng. Speech recognition interface design for in-vehicle system. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 29–33. ACM, 2010.

[35] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, and Raj Foreword By-Reddy. *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice Hall PTR, 2001.

[36] Xuedong Huang, James Baker, and Raj Reddy. A historical perspective of speech recognition. *Commun. ACM*, 57(1):94–103, January 2014.

[37] Karlsson Joakim. The integration of automatic speech recognition into the air traffic control system. Technical report, Cambridge, Mass.: Flight Transportation Laboratory, Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology,[1990], 1990.

[38] Rhys James Jones, Simon Downey, and John S. Mason. Continuous speech recognition using syllables. In *In Proc. Eurospeech '97*, pages 1171–1174, 1997.

[39] Biing-Hwang Juang and Lawrence R Rabiner. Automatic speech recognition–a brief history of the technology development. *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara*, 1, 2005.

[40] Daniel Jurafsky, Chuck Wooters, Gary Tajchman, Jonathan Segal, Andreas Stolcke, Eric Foster, and Nelson Morgan. The berkeley restaurant project. In *ICSLP*, volume 94, pages 2139–2142, 1994.

[41] H.D. Kopald, A. Chanen, Shuo Chen, E.C. Smith, and R.M. Tarakan. Applying automatic speech recognition technology to air traffic management. In *Digital Avionics Systems Conference (DASC), 2013 IEEE/AIAA 32nd*, pages 6C3–1–6C3–15, Oct 2013.
[42] Cini Kurian and Kannan Balakriahnan. Continuous speech recognition system for malayalam language using plp cepstral coefficient. *Journal of Computing and Business Research*, 3(1), 2012.
[43] KF Leung, FH Frank Leung, HK Lam, and Peter Kwong-Shun Tam. Neural fuzzy network and genetic algorithm approach for cantonese speech command recognition. In *2003. FUZZ'03. The 12th IEEE International Conference on Fuzzy Systems*, volume 1, pages 208–213. IEEE, 2003.
[44] Edward C Lin, Kai Yu, Rob A Rutenbar, and Tsuhan Chen. A 1000-word vocabulary, speaker-independent, continuous live-mode speech recognizer implemented in a single fpga. In *Proceedings of the 2007 ACM/SIGDA 15th international symposium on Field programmable gate arrays*, pages 60–68. ACM, 2007.
[45] F Marque, SK Bennacef, F Neel, and S Trinh. Parole: a vocal dialogue system for air traffic control training. In *Applications of Speech Technology*, 1993.
[46] LG Miller and S Levinson. Syntactic analysis for large vocabulary speech recognition using a context-free covering grammar. In *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, pages 271–274. IEEE, 1988.
[47] M. Nofal, E. Abdel-Raheem, H. El Henawy, and N.A. Kader. Acoustic training system for speaker independent continuous arabic speech recognition system. In *Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology, 2004.*, pages 200–203, Dec 2004.
[48] Jan Novotnỳ, Pavel Sovka, and Jan Uhlíř. Analysis and optimization of telephone speech command recognition system performance in noisy environment. *Radioengineering*, 13(1):1, 2004.
[49] JM Pardo, J Ferreiros, F Fernandez, Valentin Sama, R De Cordoba, Javier Macias-Guarasa, JM Montero, R San-Segundo, LF D'Haro, and Germán González. Automatic understanding of atc speech: Study of prospectives and field experiments for several controller positions. *IEEE Transactions on Aerospace and Electronic Systems*, 47(4):2709–2730, 2011.
[50] B.L. Pellom, R. Sarikaya, and J.H.L. Hansen. Fast likelihood computation techniques in nearest-neighbor based search for continuous speech recognition. *Signal Processing Letters, IEEE*, 8(8):221–224, Aug 2001.
[51] Omprakash Prabhakar and Navneet Kumar Sahu. A survey on: Voice command recognition technique. *International Journal of Advanced Research in Computer Science And Software Engineering*, 3(5), 2013.
[52] V Radha and C Vimala. A review on speech recognition challenges and approaches. *doaj. org*, 2(1):1–7, 2012.
[53] V. Radha, C. Vimala, and M. Krishnaveni. Continuous speech recognition system for tamil language using monophone-based hidden markov model. In *Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology*, CCSEIT '12, pages 227–231, New York, NY, USA, 2012. ACM.
[54] D. Schaefer. Context-sensitive speech recognition in the air traffic control simulation. *EEC Technical/Scientific Report No. 2001-004*, 2001.
[55] ICAO Secretariat. Outlook for air transport to the year 2025. *Report No. Cir*, 313, 2007.
[56] Hussien Seid and Björn Gambäck. A speaker independent continuous speech recognizer for amharic. *INTERSPEECH 2005*, 2005.
[57] Benjamin J Shannon and Kuldip K Paliwal. Feature extraction from higher-lag autocorrelation coefficients for robust speech recognition. *Speech Communication*, 48(11):1458–1485, 2006.
[58] CMU Sphinx. Cmu sphinx: Open source toolkit for speech recognition. *Retrieved*, 8(13):2010, 2010.
[59] Georg Stemmer, Elmar Nöth, and Heinrich Niemann. The utility of semantic-pragmatic information and dialogue-state for speech recognition in spoken dialogue systems. In *Text, Speech and Dialogue*, pages 439–444. Springer, 2000.
[60] Stevenson. Oxford dictionary of english.
[61] Glenn Taylor, J Miller, and Jeff Maddox. Automating simulation-based air traffic control. In *Interservice/Industry Training, Simulation, and Education Conference*, volume 2193, 2005.
[62] R. Thangarajan, A. M. Natarajan, and M. Selvam. Word and triphone based approaches in continuous speech recognition for tamil language. *WSEAS Trans. Sig. Proc.*, 4(3):76–85, March 2008.
[63] R Thangarajan, AM Natarajan, and M Selvam. Syllable modeling in continuous speech recognition for tamil language. *International Journal of Speech Technology*, 12(1):47–57, 2009.
[64] Edmondo Trentin and Marco Gori. A survey of hybrid ann/hmm models for automatic speech recognition. *Neurocomputing*, 37(1):91–126, 2001.
[65] Thanassis Trikas. Automated speech recognition in air traffic control. Technical report, Cambridge, Mass.: Massachusetts Institute of Technology, Dept. of Aeronautics and Astronautics, Flight Transportation Laboratory, 1987, 1987.
[66] Karen Ward. A speech act model of air traffic control dialogue. 1992.
[67] MARTA WRONISZEWSKA and JACEK DZIEDZIC. Voice command recognition using hybrid genetic algorithm. *TASK QUARTERLY*, 14(4):377–396, 2010.
[68] Dong Yu and Li Deng. Deep neural network-hidden markov model hybrid systems. In *Automatic Speech Recognition*, pages 99–116. Springer, 2015.
[69] Bartosz Ziółko, Suresh Manandhar, Richard C Wilson, and Mariusz Ziółko. Semantic modelling for speech recognition. *Proceedings of Speech Analysis, Synthesis and Recognition. Applications in Systems for Homeland Security, Piechowice, Poland*, 2008.