
SESAR
digital academy
SESAR Digital Academy Webinar:
Automated Speech Recognition for Air Traffic Control
Moderated by Olivia Nunez
ATM expert, SESAR JU
Hosted by
SESAR

Today's speakers

The need for research on ASR in ATM

- Hon. Prof. Dr. Hartmut Helmke, HAWAII Project Lead, DLR

Illuminating flights using speech recognition

- Raquel Garcia Lasheras, R&D Engineer, CRIDA

Application of ASR in the tower environment

- Ramona Santarelli, Engineer, ENAV

The view of an Air Navigation Service Provider: The costs and benefits of ASR for Austro Control

- Christian Kern, Director of Operations, Austro Control
- Christian Windisch, Senior Air Traffic Management Expert, Austro Control

Question	Answer
If you let the AI ignore speech it is not "sure of", can you increase the accuracy to 99.99%?	Yes, we can achieve an error of 0%, but than the recognition rate would be also zero. ASR is only in very seldom cases very sure. So this is a theoretical question. We distinguish therefore between command recognition rate and command recognition error rate and the rest is rejection rate.
Have you also heard that CPDLC doesn't really work reliably. Too many re-transmission attempts and slow?	Voice remains the primary means of communication between controllers and pilots. However, it is foreseen that CPDLC will gradually increase (and latency will go down dramatically (<1 second) when LDACS is introduced). We anticipate that ASR will be equally useful when CPDLC "takes over" (ASR will be used in support of CPDLC, so that both ATCOs and pilots can "dictate" messages rather than type, much as we now use ASR to compose text messages on our phone). Most of the ASR applications currently under development are mainly targeting the current voice environment, but we also have ongoing work on applications in support of CPDLC.
Does it make sense to use Text-To-Speech along with ASR to use a global and normalized accent between ATCos and pilots?	Not clear, how this should work. If you have already the message in text form, i.e. in digital format, there is no need to transform to speech again and then try to recognize it again. You can directly send in digital form. In principle of course, it would help. But I do not see the use case.
The pilot voice may change, from right seat to left. (This question is in reply to a live question on whether voice timbre/pitch could be used to recognise the callsign after the first transmission)	Indeed, while in most cases it will not change, in some cases it does. In any case, use of pitch/timbre for callsign recognition of callsign is not currently under research. In should be used in addition, callsign recognition is still easier than speaker recognition.

Can you share the complete reference/link to a research paper, if this exist? Thanks!	www.haawaii.de and https://www.haawaii.de/wp/dissemination/references/
There has been discussion regarding different accents. However, pitch is also a concerns. Typically female voices have distinctively higher pitch which can have impact on the recognizer. Has this been investigated?	When training the ASR system we use operational recordings and the proportion of female/male in pilots (5/95) and controllers (30/70) has an impact on the database. We know the problem and are trying to improve the proportion of female recordings. AcListant® achieved higher rates on recognition of female speaker, they were closer to phraseology, maybe not statistical significant.
Yes, ASR can be a filter for ATCO messages to harmonize them ...	But then the acceptance goes down. I know this from other projects (no ASR), they had to stop a project many years ago, because no ATCO used it.
Other than content recognition, is it possible to use ASR for identity recognition of the speaker?	Yes, but then it is called speaker identification, ASR is Speech-to-Text.
The focus seems to be on workload reduction, but what is the impact on ATCO Situation Awareness? Because with ASR to create system entries, the ATCO is not composing the entry herself. Research shows that SA is not as good for things that are done for you than those that you do yourself.	The potential impact on SA is a major concern as the levels of automation are increased; all SESAR solutions perform a comprehensive Human performance Assessment that considers these aspects. In this case, the ASR is not really doing something for the ATCO, but merely avoiding the ATCO to have to type what he has already said, and our results so far suggest SA is maintained or improved.

Most discussions today has been centred around speech processing on the ATCO side. Is there any research towards incorporating voice recognition at the aircraft side?	We do not have any research in SESAR on incorporating ASR in the cockpit, but it is definitely something we expect to be able to address in the future. HAAWAII is addressing pilot recognition. ATCO2 project is doing that. Airbus is also addressing recognition in the cockpit.
What is the impact of VHF specific noise on the dataset and the accuracy of the recognizer?	For Vienna and Praha we used data directly from the Voice com system, so when you recognize only ATCOs voice, it is not a problem. When you want to recognize also the pilot's voice, you run in a lot of quality problems.
Hartmut - if you need a special "accent" training, I see an issue on the plurality of nationalities we have in our premises. Furthermore, how does this apply to the pilot side? They are by definition "multi-accented"	You just need a lot of training data, Never "seen" a speaker from India might result in problems, when you have an Indian accent. In principle, however, accent and speaker independence are the aim.
How the ASR decodes the waypoint names, as they don't follow conventional name	When you use the AIP to train the system with all the waypoints available in the TMA, it works.
Christian(s. I consider a simultaneous physical input into a system not only a necessity but as well an additional confirmation for my action and a another way of making it rememberable for me as a controller. Did you get any feedback from controllers during your tests and reviews concerning this?	The feedback of the controllers were; either they want a confirmation click before the command is "send" to the system; others had the feedback they don't want any confirmation. At the end we had the "compromise" you see the command in the label for some seconds, if you don't correct it, it is send automatically into the ATM system.

Has voice data been considered as a data source for training of AI/ML models for other decision support tools?	Yes of course, voice data is the only data source which directly shows the ATCOs work. Radar data are only indirect measures
As a controller support tool will ASR presentations be different to Planner, Executive or MSPs according to their roles?	Most probably, yes. That will be dependent on the ATM system itself. ASR gets the data from COM and ATM System and sends back recognized commands to the ATM system. What is presented to the ATCO depends on the ATM system.
@Hartmut my thinking is that individual ASR instances tuned and matched/referenced to say an individual ATCOs login profile to improve the accuracy.	Yes, that helps, if you know who is speaking you might increase recognition rate by 5% relative, so WER might go down from 10% to 9.5%, but if you have the wrong ATCO, so login is wrong. WER dramatically goes up; for more information please see our paper at SID 2017: https://www.sesarju.eu/sites/default/files/documents/sid/2017/SIDs_2017_paper_27.pdf
Are you currently making use of audio generation methods to create additional training data and/or to augment audio from the simulation environment?	No, we are now mostly using data from the ops room. There we have to support the ATCO and ATCO are different in lab and in ops room at least when you focus on speech recognition. But artificial data can have benefits.
Would be the system able to highlight also flights which got the wrong frequency?	Yes, when the HMI in the ATM system is presenting this then yes. ASR system sending recognized commands to the ATM system and ATM system is presenting it in the aircraft label.
Did you also assess the impacts of delay between the moment in which an ATCO issues the instruction and the time the read back is received using ASR compared with the delay using RTF?	Normally a read back error would be corrected by the controller still. In my expectation ASR would act as an additional safety net and highlight an unheard read back error (hear back error).

So is speech recognition considered being a part of the VCS or the ATM System or both?	Depends on the application and on the interfaces you provide
In case of a fatal accident as a consequence of the automatic voice recognition error, who would be liable? Have all the liability potential issues been considered?	That should not happen. ASR is just a support for the ATCO. The ATCO is always the boss.
Has there been any investigation of "aligned to profile" ASR where a specific instance, "tuned" per individual is called upon?	Sorry, question is not understood or do you mean speaker dependent models? Then see reference 7 at https://www.malorca-project.de/wp/?page_id=424 M. Kleinert, H. Helmke, H. Ehr, Chr. Kern, D. Klakow, P. Motlicek, M. Singh, and G. Siol, "Building Blocks of Assistant Based Speech Recognition for Air Traffic Management Applications". 8th SESAR Innovation Days, Salzburg, 2018.
Can we incentive ATCos (accent / Grammar / phraseology) so that ASR gives 0 errors ?	When ATCOS better adapt to phraseology, the recognition rates will be higher, but not 100%. One ATCO, who had an incident some years before, and since then now speaks according to the book achieved good performance and also 99% of recognition rates.
Were the trials conducted with high traffic density and stressful situations included? How it made an impact on the numbers/capacity benefits?	Only high traffic scenarios. In Today COVID Traffic Situation you will not observe benefits. We had a runway closure and also an emergency with a sick person on board, have a look on the references on www.haawaii.de
Were the capacity increases taken into account in a multi-tasking type scenario, where the ATCO is clicking the entries with the mouse	Yes, the ATCO decides how (s)he want to work

while talking and while the pilot read-back was ongoing, or was it a case of the ATCO talking, then updating the ATM system after the transmission?	
Chrsitian, How do you measure this gains ?	DLR did this during the validations on the SIM. They had many different measuring methods. Some tests during the sim runs to measure the ATCOs stress and workload and some tests after the runs.
In some regions, the environment is not solely English. Is a multilingual environment possible with ASR? How do you train the ASR to recognise additional languages?	In Spain, the model is multilingual, trained with communications in both languages, Spanish and English
What % of speech recognition rate has been defined as the minimum to decrease workload of ATCO and increase safety instead of creating additional disruption?	We showed benefits with 85%. We never had worse rates at DLR, but in SESAR 2020 wave 1, controllers were already happy with 75%. 50% was not enough Maybe 99% is even too much, because then you rely too much, but here we need research, but in the meantime we should already benefit from already available technology.
What type of ASR model did you have the most success with (least WER%) ?	See reference on www.haawaii.de Assistant Based Speech Recognition, however, is a must. We concentrate on Command Recognition Rates, WER is at the end only a hint for good semantic performance. We achieved e.g. better command recognition rates with worse WER
Is the (cyber)security perspective considered in this research, e.g. authentication / verification (voice pattern matching, etc.) to mitigate impersonation of fake ATCO commands	A security assessment is being performed within SESAR projects, nevertheless the voice recognition is performed by the CWP, so the already CWP security will be in place (login, card,...) and initially no further security need is expected.

Was raw voice recording from comms between pilot and ATCO used for training of ASR models? If so how did you overcome the noisy nature of this comms data?	We used recordings from the backup-recording system from our COM System, quality is not great, but it was enough for training the ASR.
Is ASR compatible with the transition towards a full datalink environment, with less and less VHF Speech? Will this technology not be available once no longer needed?	Yes in 2122 ASR is not needed any more, when data link is available, just a joke. In Haawaii project, we are with ISAVIA who want to benefit from ASR even already having 70% of messages sent via CPDLC. ASR and Data Link are complementary.
How did you build your huge datasets: Manual transcription of operational recording? Recording platforms? Other? And how did you create it respecting GPRD (as voice is a biometric data)?	120% of project costs are with respect to GDPR. just a joke, but it was really a challenge to get the OK of the ATCOs etc. Yes, manual transcription is needed, but also see MALORCA (https://cordis.europa.eu/project/id/698824/results) approach of automatic transcription and then check against radar data.
How do you see the potential beyond ASR, in particular feeding recognised ATC commands into DataLink for optimisation of spectrum, reduction of workload, etc.	Speech Recognition coexist with CPDLC. You can much faster speak than clicking a complex command with mouse. You only have to tell the system if that what you will say should go out via frequency or via CPDLC, e.g. you can press a key on keyboard.
How did you managed to train your ASR to recognize the various non-native accents of the pilots?	We need data, data, data For example, chinese pilot not being in training data will not be recognized or will be recognised with less accuracy.

Don´t you have problems in understanding due to ATC radio noise? And do you hit these recognition rates of 95 % with the same ATCOS or can you go across various ATCOS?	Here we had different ATCOs, Austrian, German, Irish, Croatia, Swedish, Danish, Czech, but no Italian, or Spanish or Chinese. Noise is really a problem, but not a showstopper.
Will the presentation be available afterwards?	Yes