Increasing Air Traffic Control simulations realism through voice transformation

Mathieu Serrurier, Sylvain Neswadba, Jean-Paul Imbert

Increasing Air Traffic Control simulations realism through voice transformation

Mathieu Serrurier
IRIT toulouse France
serrurie@irit.fr

Sylvain Neswadba
DSNA RD Toulouse France
sylvain.neswadba@gmail.com

Jean-Paul Imbert
DSNA RD Toulouse France
imbert@cena.fr

Abstract. Improving realism in simulations is a critical issue. In some air traffic control (ATC) simulations we use a pseudo-pilot which pilots up to fifteen aircraft. Thus, having the same voice for different aircraft in the case of pseudo-pilot decreases the realism of the simulation and may be confusing for the controllers especially in study context. In research context, a virtual aircraft piloted in a flight simulator is sometime needed in addition to the pseudo pilot. For simulation needs, the flight simulator aircraft must be merged with pseudo-pilot's one. This is not possible without voice modification since the controller can distinguish the pilot voice. In this paper we propose a method for transforming the voices of the pilot and the pseudo-pilot in order to have one particular voice and cabin noise for each aircraft. The two experiments that have been conducted show that, through our voice modification algorithm, the realism of the simulation is enhanced and the voice biases disappears.

1 Introduction

Gaver shows in [2] that sounds help the engagement of the users with a system. This is especially true in simulation systems where immersion highly depends on their realism and their capacity to reproduce the target environement. In air traffic control (ATC) the sounds essentially come from the communications with the pilots.

Simulations are a critical part of ATC. They serve at least three purposes. The first one is the training of the student controllers. The second one is the evaluation of new protocols or maneuvers. The last one is the testing of new interfaces or softwares. Three human actors perform the simulation :

- **Controllers :** The controllers are in charge of a virtual traffic. The conditions are as close as possible to the real ones. As in reality the voice is the unique way of communication among controllers and aircraft.
- **Pseudo-pilots :** Since it is too costly to have a pilot in a simulator for each virtual aircraft, pseudo-pilots are used. A pseudo pilot is a human operator that pilots simultaneously up to fifteen aircraft. Instructions for the aircraft are transmitted to the traffic simulator through a dedicated HMI (human machine interface). The pseudo-pilot is in charge of the voice communication of all the aircraft he pilots.
- **Pilot in a flight simulator :** It is usual, in a research context, to have one aircraft piloted through a flight simulator in order to increase the realism of the simulation or test particular scenarios, in this case, the simulation scenario is focused on the aircraft controlled by the pilot. The pilot in the simulator is only in charge of the voice communication of unique the aircraft he pilots.

The architecture of an ATC simulation is presented in Figure 1. Controllers, pilot and pseudo-pilot are interacting through a traffic simulator and a simulated radio based on Voice over IP (VoIP)) .The communications between controllers and pilots or pseudo-pilots are entirely vocal. The signal is encoded in $8Khz$ and $8bits$ PCM using μ-law optimisation.

It is obvious that the voice is a crucial issue in ATC simulations. The architecture described previously introduces biases in the simulation. First of all, having the same voice for all the aircraft piloted by a given pseudo-pilot may be very confusing and damaging for the controller. Therefore, the aircraft that is piloted by a pilot in a simulator is the only one that is associated with a unique voice. This makes it easily identified by the controllers. Controllers will thus focus on this particular aircraft since difficulties in a simulation often come from the aircraft piloted through the flight simulator. Finally, the quality of

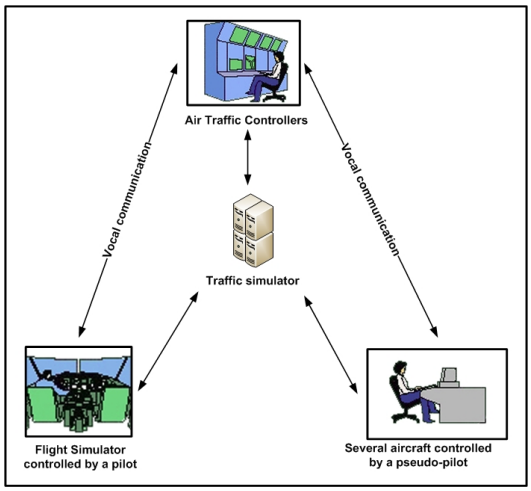

Figure 1: ATC simulation architecture

the communication by VoIP is greater than the quality of the signal in real life where radios are limited to 5khz frequency and are disturbed by VHF (Very High Frequencie) and cabin noises.

In this paper we propose to morph the voice of the pseudo-pilot and the pilot in order to have a distinct voice for each aircraft. Voice morphing has been studied essentially for two purposes. The first one is for vocal recognition [5]. The system of recognition is tuned for a specific reference voice. The voice morphing is then used to transform the voice of the user into the reference one in order to perform the recognition. These methods require much information on user and reference voice. Voice morphing can also be used for masking voice for security purpose [4]. After being morphed the voice has to be no more identifiable as the owner's one. The obtained voice is no more human and therefore this method is not suitable for our case. We propose to use a method named formant shifting which changes the fundamental characteristics of the voice. The method requires few information on the initial and the target voices. The only constraint of the morphing is to obtain a voice that is human. We also modify the signal in order to limit the frequency to $5khz$ and add cabin and VHF noises. Speech synthesis has been considered in order to have a different voice for each aircraft. However, this approach has been given up due to a lack of realism.

The paper is organized as follows. We first describe the algorithm of voice morphing and the other treatment made on the signal. Then two experiments are presented. In the first one, we show that by increasing the realism of the simulation we increase the performance of the controllers by allowing them to quickly identify the calling aircraft. The second experiment shows that the voice morphing algorithm enables to mask the voice of the pilot in the simulator in order to increase the difficulty to identify him.

2 Signal processing

As pointed out in the introduction the voice is transmitted through a VoIP protocol. The signal is processed at two stages. In the frequency space, obtained after fast Fourier transformation, the formant shifting algorithm is applied in order to change the characteristic parameters of the voice. The signal is impaired by restricting the frequency and clipping the magnitude. In the time dimension, cabin noises are added to the signal. Due to the manipulation in the frequency space, the signal has to be resampled. As a tradeoff between performance and quality, the signal is resampled 16 times. Each modification depends on some parameters that constitute a voice modification profile. A different modification profile is assigned for each aircraft, even the one controlled by the pilot on the simulator. The processing of the signal is made just before the VoIP stage. The pseudo-pilot selects the aircraft that needs to answer the controllers, then the voice is automatically transformed according to the parameters associated with the aircraft. The software is implemented in JAVA. It runs without specific hardware, directly on the machine that runs the pseudo pilots, usually a dual core computer. The processing is made in real time.

2.1 Formant shifting

The formant shifting is the most important part of the voice transformation process. Formant shifting is an extension of the pitch shifting algorithm which shifts all the signal in the frequency space. Formant shifting shifts independently each formant in the frequency space. Formant shifting allows us to create more different voices than pitch shifting. The formants are high energy areas in the frequency spectrum. Roughly speaking, they are the peaks in the signal. The first formant is usually bounded between 90 and $450Hz$. This is the formant that changes when people are singing for instance. It is usual to consider that the print of a voice is defined by the first four formant frequencies (f_1,f_2,f_3,and f_4). The formant shifting algorithm consists in shifting the initial formant of the signal into some target formant that corresponds to other realistic voice formants. Let us consider the following initial voice v with $f_1 = 300hz$, $f_2 = 900$, $f_3 = 1800$, $f_4 = 2900$ and the target voice v' with $f_1^{'} = 250$,

$f_2^{'} = 1000$, $f_3^{'} = 1650$ and $f_4^{'} = 3000$. When shifting the formant the signal in the $[0, f_1]$ frequency interval will be linearly compressed in the $[0, f_1^{'}]$ frequency interval, the signal in the $[f1, f2]$ frequency interval will be linearly expanded in the $[f_1^{'}, f_2^{'}]$ frequency interval and so on. The algorithm is based on the Stephane Bernsee pitch shifting algorithm [1]

The only preprocess step needed for formant shifting

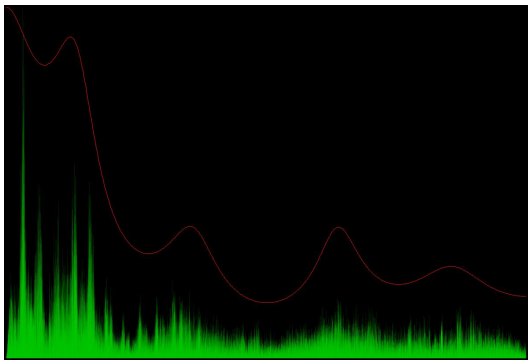

Figure 2: Linear predictive coding for a voice sample

is to determine the formants of the initial voice. In order to determine the position of the formants, we use the linear predictive coding algorithm [3] (see fig. 2). Linear predictive coding is used to represent the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. Having the spectral envelope of the signal, the formants can be easily localized since they correspond to the local maximums of the envelope (see fig. 3). Thus, each aircraft of the simulation will be associated to distinct target formants.

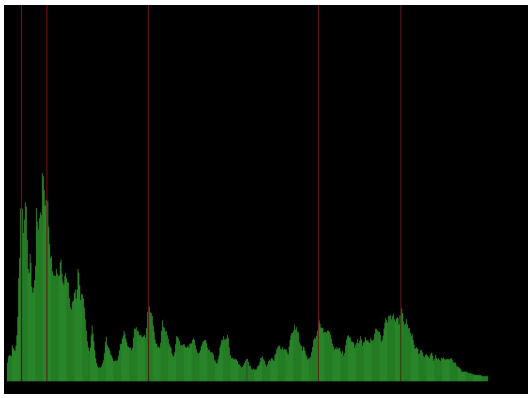

Figure 3: Formant detection for a voice sample

2.2 Signal processing

In order to have a digital signal more realistic, three additional operations are done in the frequency space :

- **Signal equalization**. In order to simulate the environment of the cockpit, the signal is equalized according different possible equalization schemes.
- **Frequency reduction**. The frequencies upper than 5Khz are removed according to the capacity of cockpit radio.
- **Magnitude clipping**. The magnitude of the signal is clipped in order to simulate poor quality radio. The amount of clipping is a parameter that may vary for each aircraft.

2.3 Cabin noises

In real world, different noises disturb the signal of the pilot voice. They can come from background, VHF or the reactors of the plane which can be heard throw the radio. Each aircraft has a particular cabin noise that is easily identifiable by the controller. In order to simulate this, environmental sounds are added to the signal at the end of the process. These sounds have been previously sampled and contain different reactors sounds and cockpit sounds. The sample and the intensity of the cabin noises differ for each aircraft.

3 Experiments

In order to check the effectiveness of our voice morphing algorithm, two experiments have been made. In the first one, we show that, by increasing the realism of the simulation, we increase the controller's efficiency for identifying aircraft. In the second one we measure the ability to hide the pilot's voice during the simulation. The voice are modified by the following ways:

- Determine the profile of the speaker voice.
- Chosing randomly target formant in a restricted range (up tu 10
- Chosing randomly a cabin noise and its level
- Chosing randomly a signal equalization scheme (from a database), a frequency reduction and a magnitude clipping range.

Users use an headse for hearing sounds. No specific hardware and software are used.

3.1 Experiment 1

The hypothesis of the first experiments is that modifying the voice implies that the aircraft are easier identifiable by the controller. In real world, controllers identify an aircraft through the callsign announced by the pilot together with the voice of the pilot and the cabin noises.

3.1.1 Protocol

The experiment has been performed by 20 participants without known hearing problems. They are aged from 20 years to 40 years. 4 are controllers. 16 are male, 4 are female. They are not necessarily controllers or pilots. This is not a biais in our experiments since no competence in ATC are needed. The experiment focus on the capacity of a user to distinguish voices. We consider a simulation with five aircraft with distinct callsigns (AirFrance 2035, Twinjet 07, BritAir 46 EK, Fox Bravo X-ray). These aircraft are piloted by a unique pseudo-pilot. A set of sentences has been recorded for each aircraft. The goal of the subject is to associate as quickly as possible a sentence to the corresponding aircraft. The callsign of the aircraft is pronounced in each sentence. The callsign can be indicated at the beginning, in the middle or at the end of the sentence, as in reality.

The experiment has two phases. In the first one, the voice is not modified, and then the voice is the same for all the sentences regardless to the aircraft. In the second one, the voice is modified according to a set of modification parameters (voice formant, cabin noises, equalizer, ...) that is associated to each aircraft. The starting phase is chosen randomly. These parameters don't change during the experiment. The order of the phrases is chosen randomly. Ten sentences are pronounced for each aircraft. Five labels corresponding to each aircraft are presented to the subject. The labels are in circles, all with the same size. The circles are dispatched regularly around the launch button. When the subject clicks on launch button a sentence is played. The subject must then click as quickly as possible on the label corresponding to the aircraft without making any error. There is no need to listen all the sentence to select a label. The time between the two clicks is measured. The position of the labels is fixed randomly at the beginning and doesn't change during the experiment. If the subject makes an error, the sentence is not replayed. The GUI of the experiment is presented in figure 4

3.1.2 Results

As expected, when the voice is not modified, the subject needs to wait for of the callsign. When the voices are modified, the aircraft can be recognized by the voice or the cabin noises before the callsign is announced. With 10 sentences per aircraft the subjects have time to memorize the signal parameters for each aircraft and then can identify it more quickly. The results are summarized in the following table :

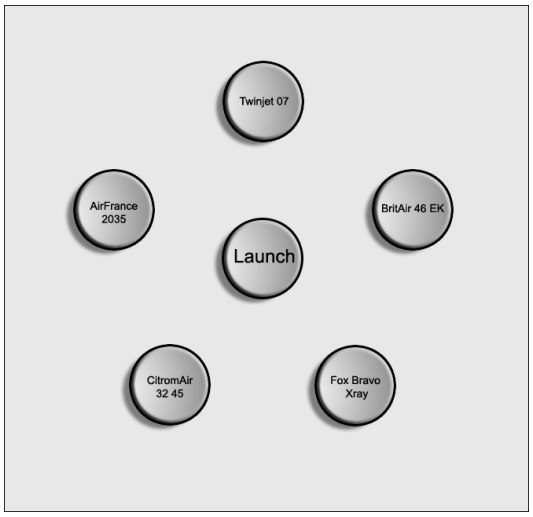

Figure 4: GUI of experiment 1

Results	without modification	with modification
Avg time in s	4.35	3.69
Error (%)	0.5	2.7

The average time for the phrases without modification and with modifications are respectively 4.35s and 3.69s. This corresponds to a performance gain equal to 18%. Student test (threshold=0.00025) shows that the difference between the two average rates is statistically significant. Some subjects refer that, even though they have identified the aircraft thanks to the voice, they wait for confirmation with the callsign in order to avoid error. The error rate increases from 0.5% to 2.7%. This error rate is essentially due to the goal of the experiment and remains acceptable for an ATC simulation since controllers may ask for confirmation of the callsign when they are not sure. Moreover, in real life, the poor quality of the radio signal may lead to some error. The experiment does not allows us to determine to what extend the increase of performance is due to the voice transformation or to the noise added. Even the cabin and radio noises are probably more easily identifiable, the voice transformation is fundamental for the realism of the simulation.

It is interesting to consider the progression of the average time during the experiment. The figure 5 shows the evolution of the average time for identification on a sliding window of size 15 (i.e. the average time for the last 15 extracts for all the subjects). The first observation that can be made is that average recognition

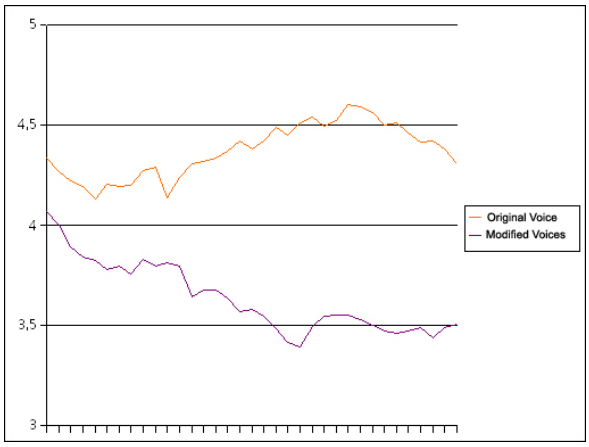

Figure 5: Average identification time (in s) evolution among time for (sliding window of size 15)

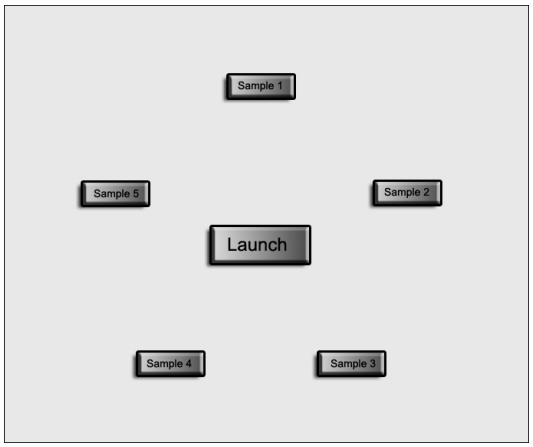

Figure 6: GUI of experiment 2

time remains stationary when the voices are not modified. This illustrates that there is no learning process. On the contrary, when considering the modified voices, the average idetification time decreases along the experiments. It shows that the subject learns to match a callsign with a specific voice and cabin noise.

3.2 Experiment 2

The hypothesis we want to validate is the following : if we change with our algorithm the flight simulator pilot voice and the pseudo-pilot voice in order to have one different voice per aircraft, controllers are no more able to determine which aircraft is piloted through the flight simulator. If the hypothesis is validated, it will avoid the bias of pilot identification.

3.2.1 Protocol

The experiment has been performed by 20 participants without known hearing problems. They are aged from 20 years to 40 years. 4 are controllers. 16 are male, 4 are female.They were not necessarily controllers. Eight people (1 female and 7 male, not necessarily pilot or pseudo-pilot) have recorded each 10 sentences, with the aeronautic phraseology. Some of them have taken part in the experiment. The participants knew the voice of people who had recorded the sentences. This is not a strong bias since it works against our hypothesis and it is similar to the situations encountered in ATC simulations.

The subjects are split into two groups. In the first group, the voices are not modified. In the second group all the voices are modified. The experiment set as follows. A sequence of five sentences is proposed to the subject. Four of these sentences are from a pseudo-pilot (the same for the four sentences), one is from the pilot. For the second group, each sentence pronounced by the pseudo-pilot is modified with different parameters (formant, equalizer, cabin noises, ...) . The voice of the pilot in the simulator is modified too. The five sentences are played randomly. After hearing all the voices, the subject tries to identify the sentence that has been pronounced by the pilot in the flight simulator. This consists of identifying the voice that has pronounced only one sentence. If the subject makes an error, the sentences are not replayed. There are 28 sequences of 5 sentences. The GUI of the experiment is presented in figure 6. The choice of the pair pilot/pseudo-pilot is made randomly with the following constraints : a pair can only be chosen once, A sentence can be played two times, but not in the same sequence and with different modifications.

3.2.2 Results

The subjects in the group 1 and the group 2 have an average error rate equal to respectively 5.5% and 46.5%. Even with modifications, the experiment shows that the female voice is easily identifiable. The results are summarized in the following table.

Error (%)	Group 1	Group 2
with female voice	5.5	46.5
without female voice	7.1	56.2

Student's test (threshold=0.00025) shows that the difference between the two error rates is statistically significant. As expected the error rate for group one is

close to 0. This is not surprising since without modifications, the pilot voice is easily identifiable. The theoretical maximum error rate for group two is 80% (it corresponds to a random choice of the answer). Although this maximum is not reached in the experiment, the algorithm performs very well and tends to validate our hypothesis. The fact that the maximum is not reached can be explained by different reasons. First some parameters such as pronunciation rhythm can not be modified by our algorithm. Female voice is also easily identifiable. However, some biases in our experiment may not appear in real situation. First of all, in our experiment people that are not accustomed with aeronautic phraseology have some hesitations when reading sentences. Moreover, in air traffic simulation, controllers do not focus on recognizing the pilot as in our experiment.

4 Discussion and perspectives

In this paper, we have proposed a method to increase realism of air traffic simulation by modifying the voice of the pseudo-pilots. The algorithm of voice modification is based on the shifting of the characteristic parameters of the considered voice. The voice transformation is made in real time and does not disturb the simulation.

The first experiment is interesting in many ways. It illustrates the fact that we can use cabin noises and voice to easily identify the aircraft. It is worth stressing that this increase of performance has been obtained by decreasing the quality of the signal. Even if the identification of the aircraft is not a critical issue for simulation, by making it easier the workload of controllers is decreased. Its also shows that the performances of controllers can be increased by increasing the realism sound context of the simulation. This increase of realism is very useful in the context of simulation for student controllers.

The second experiment shows that the modification of voice allows us to avoid the bias due to identification of the flight simulator pilot. This also makes the simulation less confusing for controllers since there is one different voice for each aircraft, as in real world.

The software is currently used for real simulations. The controllers judge the modified voice as realistic and report that it highly increases the realism of the simulation. The participants of simulation report unformally the following feedback:

- controllers told us that voice modification induces a quicker and better immersion in the simulation.
- following a simulation with controllers who weren't aware of the voice modification, they couldn't say how many pseudo-pilots were involved (there was only one).

Voice tranformation indudes an extra cost for the pseudo-pilot. Indeed, it is mandatory to select the aircraft before speaking. When pseudo pilot interface allows for aircraft selection, as it is in our case, the pseudo-pilot reported that this cost remains acceptable in comparison with the improvement of the simulation.

Future work will consist in making the female voices less indentifiable. On the other hand, it would be interesting to propose automatic tuning of voice parameters in order to have voices that are as different as possible.

References

[1] S. Bernsee. Pitch shifting using the fourier transform. http://www.dspdimension.com/admin/pitch-shifting-using-the-ft/.

[2] W. Gaver. Sound support for collaboration. In *ECSCW'91: Proceedings of the second conference on European Conference on Computer-Supported Cooperative Work*, pages 293–308, Norwell, MA, USA, 1991. Kluwer Academic Publishers.

[3] J. Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, 1975.

[4] P. Perrot, G. Aversano, and G. Chollet. Voice Disguise and Automatic Detection: Review and Perspectives. In *Progress in Nonlinear Speech Processing*, pages 101–117. Springer Berlin / Heidelberg, 2007.

[5] H. Ye and S. Young. Perceptually Weighted Linear Transformation for Voice Conversion. In *Eurospeech*, pages 2409–2412, 2003.