ASSETs+
Alliance for Strategic Skills addressing Emerging Technologies in Defence

Co-funded by the
Erasmus+ Programme
of the European Union

Remote everything?
To what extent can unmanned assets interact with humans in the field for defence operations?

#EuropeanDefenceChallenge

www.assets-plus.eu

CAN I TRUST YOU? AN AI FOR SECURE, AUTHENTICATED COMMUNICATION

Authors:

Daniel Farias,
University of Cádiz

Carlos Thurn
NORDAKADEMIE Hochschule der Wirtschaft

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

CAN I TRUST YOU?
AN AI FOR SECURE, AUTHENTICATED COMMUNICATION

Daniel Farias & Carlos Thurn

Universidad de Cádiz & NORDAKADEMIE
Hochschule der Wirtschaft

Contents

1. Introduction

The European Defense Challenge of 2021 / 2022 focusses on the topic of technology, specifically the idea of ‘remote everything’. The goal is to contribute to the question “To what extend can unmanned assets interact with humans in the filed for defense operations?” This paper is aiming to tackle a key challenge of this topic: *secure* communication.

‘Remote everything’ conveys many advantages like the safety of our soldiers and citizens, however, it also comes with some concerning risks. In a more connected world, the threat of cyberattacks rises with every new disruptive technology. The interaction of humans and unmanned assets requires a method of communication which by principle can be accessed from the outside and is therefore vulnerable. In order to ensure secure communication in the field for defense operations, it is essential to implement proper authentication methods to identify authorized parties of the communication process so that we can entrust the system with critical information.

The following chapter will explain the general approach of the idea and provide an outline of the paper.

2. General Approach

Building on the first proposal for the project, this paper will focus on exploring specific solutions for the outlined problems. How will this be achieved?

The idea is to create a method of secure communication between two or more individual parties in order to allow interaction of human and machine for the application in the defense field. In order to achieve this, a combination of encryption and identity authentication will be implemented.

While the authentication process will ensure that no undesired parties take part in the communication process, encryption will prevent others from reading or seeing the content of the communication. Both measures can be approached in different manors. The goal of the following two chapters is to explore the topics of encryption and authentication with more depth in order to perform an educated decision in what direction to go. Public or private key encryption was identified to be well suited for this application. The authentication will be achieved by an AI, trained to recognize certain patterns of (human) behavior.

This project is expected to deliver a reliable method of communication, especially in the field of defense operations. Thus, unmanned assets can safely be controlled and maintain close contact with qualified and authorized personnel. The last part of this paper aims to deliver a working prototype that demonstrates the sought for technological capabilities.

The following chapters will explore powerful tools like Natural Language Processing, AI and Encryption in order to ensure secure communication. Furthermore, the paper can be used to later create a functional prototype based on the above-mentioned topics.

3. Authentication Process

This section of the paper will solely be focusing on the authentication process. This process as well as the encryption of the actual data will ensure the security of the communication. The idea is to train an AI with the behavior of the authorized user so that the behavior can later be recognized for security purposes. The analysis can entail many different aspects of human behavior which will be explored in the following sub-chapter. Moreover, the following part will serve to educate the reader before the next chapter, where the selected principals will be applied to a prototype for the project. For further understanding, the following graphic will show the planned communication process:

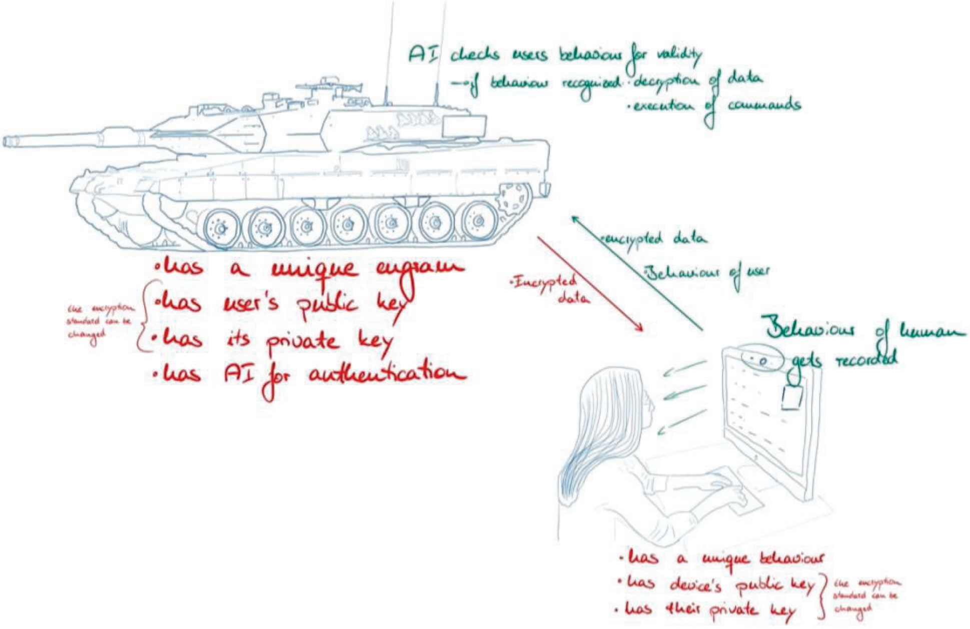
AI checks users behaviour for validity
→ if behaviour recognized · decryption of data
· execution of commands
· has a unique engram
· has user's public key
· has its private key
· has AI for authentication
the encryption standard can be changed
· Encrypted data
· encrypted data
· Behaviour of user
Behaviour of human gets recorded
· has a unique behaviour
· has device's public key
· has their private key
the encryption standard can be changed

Figure 1: General overview of the authentication process

As detailed in the graphic, during the communication process the behavior of the user gets recorded and sent to the asset along with its instructions. The instructions only get decrypted and executed, once the sending entity has been recognized as an authorized user. The next chapter will explore how and what behavior can be recognized. However, this approach does not protect against buffer overflow attacks. Therefore, the sent data should also include an authentication code which can quickly be interpreted to be correct or incorrect. If correct, the data gets accepted under the condition that it does not exceed a predetermined limit so that buffer overflow attack can easily be avoided. The pre-authentication is done so that the actual authentication AI does not need to check data from incorrect sources.

3.1 Behavioral Analysis of a Human

The analysis can cover many different aspects of human behavior. The more behavior gets recorded and analyzed, the more secure the authentication can be, however, there are also downsides to analyzing this many data.

In order to record the behavior, different equipment is needed. To work with voice, one needs a microphone, to work with video, one needs a camera and in order to analyze mimic or different movements, one could use a dot projector. Every single one of the parts costs money and presents a potential point of failure so the risk that the authentication fails due to hardware malfunction rises with every piece, and such failure is unacceptable in heated combat situations. Moreover, the more data gets recorded, the more data needs to be sent to the remote-controlled asset. This means longer processing at both ends and slower response time in general. Therefore, it is recommendable to focus on one or two technologies and perfect the authentication with limited data to be analyzed.

Voice data presents the advantage of being relatively simple, easily recordable and interpretable and has the unique feature of also being able to transport voice commands.[1] Therefore, the asset can use the authentications data for interpretation of commands as well. Video data is similar in terms of being easily recordable, however the size of the data increases by a large factor but modern technology has optimized video analysis and can work with large amounts of data.[2] The behavior of eye movement also fits in the category of video analysis but is separated from the category because it requires a close focus on the eyes, more resolution and a deeper analysis.

When it comes to recording and analyzing mimic, specific movement and interaction with the surroundings the process is more complicated due to several different factors. First, the hardware infrastructure needed for recording these behaviors is uncommon and expensive. Moreover, they require separate pieces of hardware which need to work together and join the data into one package. The required computing power today is unfeasible for quick reaction missions and therefore will be disregarded in this paper.

4. Natural Language Processing[3]

The application of *natural language processing* (NLP) can be seen in many household products like Siri, Google Translate, and Echo (Alexa). As a definition, NLP is nothing but a collection of procedures which involve the application of statistical methods, including or not including insights from linguistics, to be able to understand text with the purpose of solving real-world tasks.

There are various libraries which the project can work with, i.e., spaCy, nltk, Allen NLP, and others. However, this project will be focused on the Spark NLP library, where we can find a lot of powerful tools and one liner codes like *natural language understanding* (NLU), which is a Python wrapper around Spark NLP. It gives you all of Spark NLPs features in 1 line of code and supports all the common Pythonic Data Structures like Pandas and Modin Dataframes.

[1] Compare: (Tizeta Zewide, 2009)

[2] Compare: (Supriya Rao, 2008)

[3] Web: (John Snow LABS, 2022)

The reason for suggesting the library Spark NLP for this paper is because it has many more features than all the other libraries combined, giving no topic or method restrictions and offers all the necessary tools to work on the project and a potential prototype later. The following table can be used as reference for a clear comparison:

NLP Feature	Spark NLP	spaCy	NLTK	CoreNLP	Hugging Face
Tokenization	Yes	Yes	Yes	Yes	Yes
Sentence Segmentation	Yes	Yes	Yes	Yes	No
Steeming	Yes	Yes	Yes	Yes	No
Lemmatization	Yes	Yes	Yes	Yes	No
POS tagging	Yes	Yes	Yes	Yes	No
Entity Recognition	Yes	Yes	Yes	Yes	Yes
Dep parser	Yes	Yes	Yes	Yes	No
Text matcher	Yes	Yes	No	No	No
Date matcher	Yes	No	No	No	No
Sentiment detector	Yes	No	Yes	Yes	Yes
Text classification	Yes	Yes	Yes	No	Yes
Spell checker	Yes	No	No	No	No
Language detector	Yes	No	No	No	No
Keyword extraction	Yes	No	No	No	No
Pretrained models	Yes	Yes	Yes	Yes	Yes
Trainable models	Yes	Yes	Yes	Yes	Yes
Question answering	Yes	Yes	No	No	Yes
Text Style Transfer	Yes	Yes	No	No	Yes
Finance models	Yes	Yes	No	No	Yes
200+ Languages supported	Yes	Yes	No	No	Yes
Summarize Test	Yes	Yes	No	No	Yes
Text Generation (GPT2, T5)	Yes	Yes	No	No	Yes

Some of the many functions that can be used for this project are the following:

4.1 Context Based Spell Checking

The purpose of this function is to check the spelling of the words that form a sentence transmitted in a message. Sometimes, spelling mistakes or typos can be made by the human operator for any reason, which can provoke the other receiving party to misinterpret the intended message. Therefore, it is important to make sure that the words are as clear as possible. Here is an example below:

```
nlu.load(‘spell’).predict(‘The emney has bin spoted neer the sity’)
```

Spell	Token
The	The
enemy	emney
has	has
been	bin
spotted	spoted
near	neer
the	the
city	sity

4.2 Named Entity Recognition

This is a very powerful function that allows us to identify named entities inside a text. For defense operations, this tool is important for both the human operator and the machine to establish words that are key to understanding the environment in which the asset is operating, and to get a better situational awareness for the human operator with keywords like “soldiers”, “enemies”, “allies”, “tanks”, “helicopters”, “refugees”, “anti-air guns”, and so on.

```
nlu.load(‘ner’).predict(‘Text or message goes here’, output_level=’chunk’)
```

For example: “Captain John Doe and his Advanced Mobilization Team squadron have managed to increase the security levels by 60% in 16 hours, during their strategic operation at the center of the besieged X city, from country Y.”

Type	Description
PERSON	Captain John Doe
ORG	Advanced Mobilization Team
OUTPUT	Security
PERCENT	60%
TIME	16 hours
EVENT	Strategic operation
LOC	X city
GPE	Country Y

4.3 Transformer Based Sequence Classification with NLU

Following the same system as the last function, this one classifies a whole sentence or sequence into a topic or type of message. This is particularly useful when one wants the machine to be able to understand the context of the message sent by the user and act according to it. For example, if the user says something like “The enemy must not know about the movement of our troops during the operation”, the machine can interpret that as an “stealth mission”, by which it can take the necessary steps to help both the operatives and the troops carrying out the operation in the ground by communicating whether the enemy is near or if their location has been compromised. Therefore, in the context of the authentication process, the AI can work with these sentence categorizations and analyze them accordingly due to the fact that different situations induce different behaviors. So that it can more accurately ensure the authentication, by using categorized behavior.

The code would look something like this, along with a quick example:

```
seq_pipe = nlu.load(‘en.classify.distilbert_sequence.industry’)
seq_pipe.predict(‘The intel gathered by the team was a success …
```

```
classified_sequence   classified_sequence_confidence   document
  [Intelligence mission]                     [0.9676024]   The intel gathered by the team…
```

4.4 Transformer Based Token Classification with NLU

Working with the same system as the last two functions, this one classifies abbreviations or tokens. Sometimes we use abbreviations to write a message quicker and more effectively. In the military context, there are hundreds if not thousands of terms by which the parties involved can communicate with. Terms like *COP* (combat outpost); *IDF* (indirect fire); *IED* (improvised explosive device); or *AWOL* (absent without official leave) are some of the most used ones during a military operation. So, the machine must be able to understand all these messages around this context.

When dealing with the authentication process, this comes useful for when the user needs to send a quick message to the machine through the AI, in case of a dire situation that needs to analyze the different kinds of behaviors that they convey related to the messages and form statistics with it for future operations and authentications.

Here’s the code and a quick example:

```
tok_pipe = nlu.load(‘en.ner.mil_terms’)
tok_pipe.predict(‘Some military contractors went AWOL’)
```

classified_token	token
O	Some
O	military
O	contractors
O	went
Absent Without Official Leave	AWOL

4.5 Speech Recognition

Speech recognition is widely used in today's technology, which is a subfield of computational linguistics. As the name suggests, it allows devices to recognize human speech. This is achieved by first taking the analog input of a microphone and converting it to a computable, digital signal. This signal is used as a base for interpretation by the program. Then, through natural language processing, using the methods before mentioned, the software is able to send an output to the end party, stating whether or not the input speech matches the behavior of an authorized user.

In other words, this system follows an algorithm which can be interpreted as follows:

1. Speech Recognition
2. Speech to text
3. Text analysis
4. Categorization of content
5. Output scoring (percentage)

The fourth point, categorization of content, means that once the program is done working with the input analysis, it separates the content into blocks of context-specific situations. In this case, the program must ensure that the result of the recognized human speech is adapted to different combat behaviors based on certain keywords. This is related to the last point, which is meant to score the output of the algorithm into a percentage which represents how sure the AI is about the user being an authorized entity. Please refer to Figure 2 for further understanding.

4.6 Voice Recognition

If the user wants the unmanned asset to be able to recognize the voices that he or she, as operatives, send to it (or that the machine itself hears in its surroundings), it is necessary to have a function that allows it to receive and process the signals transmitted by humans with their voice.

One direct and practical function that this paper is going to focus on is the famous *Fourier Transform*[4], which is the foundation for decomposing real world waveforms into sinusoids that allows the user to work with a better way to interpret a signal. In this case, the signal of one's speech.

The paper will not go into details about the *Fourier Transform* itself. However, when it comes to the authentication process, this powerful tool allows the AI in charge of this process to use the recognized voices as input data in order to carry out the encryption/decryption operation in the software. To be more specific, if the user sends a message to the machine as audio, then the AI will take that data and work around with trained algorithms, following the proposed engram system, that will ensure that this message is directly received by the correct entity, in this case the machine.

5. AI for Authentication of human behavior

It is widely known that *Artificial Intelligence* or AI, has become increasingly important in all aspects of modern technology in the past decade. As such, this paper is going to focus on making sure that the core foundation on which this authentication process is based, is up to the task at hand.

But what exactly is *Artificial Intelligence*?[5] Humans are not particularly made for monotonous tasks, but computers can perform simple monotonous tasks efficiently and reliably. The main difference between computers and human beings is that the latter has the ability to reason and use common sense to adapt to new situations, while computers cannot reason and lack common sense to adapt itself. Due to the various senses a human being has, one can hear through the ears a set of voice signals, which then the brain interprets as a meaningful sentence. Therefore, *Artificial Intelligence* aims to create a machine that mimics the behavior of an ordinary human being while handling complex tasks.

There are different types of AI that serve as a categorization separated by two main points, one based on capabilities, and the other one based on functionality.

Based on Capabilities

- Narrow AI: a type of AI that performs dedicated tasks with intelligence
- General AI: performs intellectual tasks with a certain degree of efficiency like a human
- Strong AI: represents a level of intelligence that can surpass that of a human

Based on Functionalities

[4] Web: (Jair, 2018)

[5] (Nagpal, 2018)

- Reactive Machines: the most basic type of AI that doesn't store memories for future actions. They only focus on current scenarios, reacting with the best possible action for it
- Limited Memory: this type of AI can store data for a limited time period
- Theory of Mind: has the ability to understand human emotions and socially interact with them
- Self-Awareness: a type of AI that is self-conscious and has self-awareness with a super intelligence, superior to the smartest human mind. This is only a concept and a theory.

The proposed software library that will be used as foundation for the training operation and algorithms is Tensor Flow, which has an intuitive interface that lets the user dive into using complex machine learning ideas. This machine-learning framework was developed by Google and was used mainly for speech recognition, search, photos, and Gmail, among other applications.

In this particular case, the first prototype would receive data from an audio and video input source in order to generate an engram of behavior for an authorized human. This AI will need to be trained with a substantial amount of data from real world situations over an extended period of time in order to reliably identify authorized users. Therefore, it would be categorized as a narrow AI and with limited memory because should not evolve in the context of actual military missions but be able to spot a copied behavioral data stream. However, it should further be improved in the lab with data taken from all real live mission situations to increase accuracy and security.

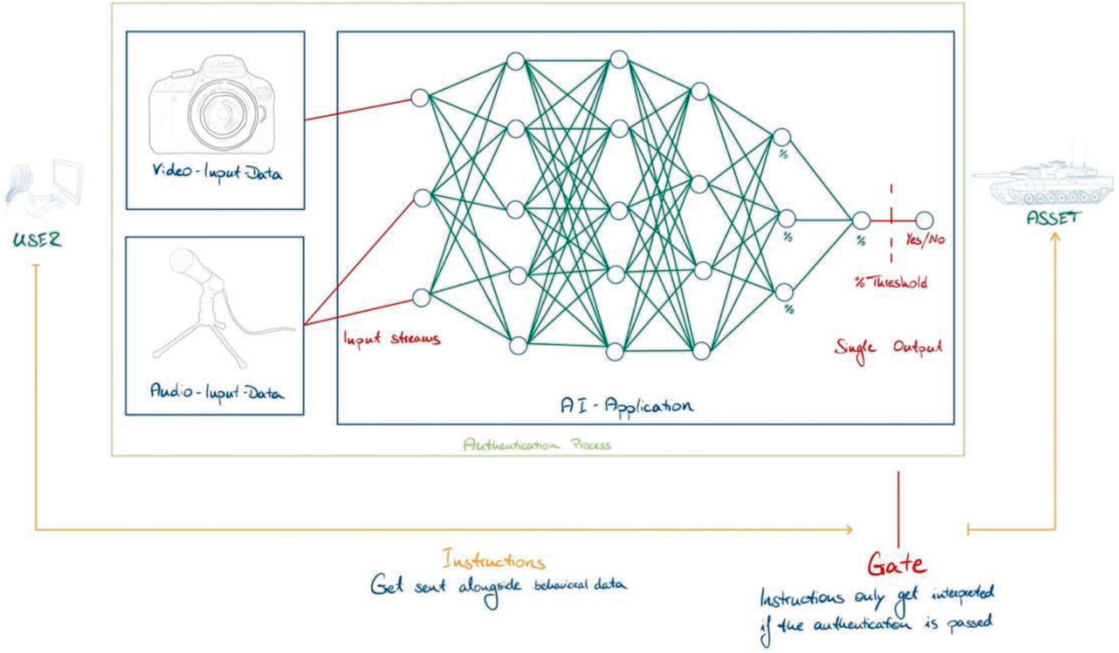

Figure 2: Detailed authentication process

In summary, based on the previous chapters, the AI will have two input sources at first which it should evaluate with different criteria. A behavioral profile or engram of a person will include a recognizable voice, a recognizable grammatical and verbal behavior taken from speech recognition data, as well as the behavior of the human body recorded on video. All three categories should be recognized by the AI within a certain percentile margin which results in a final score for whether or not the presented data corresponds to an authenticated user. If the score moves above a certain threshold, the final output of the AI will be a "Yes". If not, it outputs a "No" and the whole transmission gets disregarded because it was sent by an unauthorized source.

6. Encryption

Beside the AI authentication process, general encryption is also very important so that sent data cannot be intercepted and read by the enemy. This would be a significant security breach and create multiple potential threats. If the sent data were accessible by the enemy, they could easily find out the pre-authentication bit which would then lead to a lot more data to be interpreted by the assets' AI because the enemy would now be able to also attempt to send interpretable instructions by sending false behavioral data along with it. By encrypting the whole data set and changing the authentication bit dynamically, the enemy should not be able to send interpretable data to the asset.

Moreover, the data must be encrypted so that the enemy cannot access the behavioral data nor the given instructions. Both would be devastating security risks in combat situations. The latter can lead to the enemy knowing the asset's next move while the first could enable the enemy to analyze the sent behavioral data every time and try to recreate a pattern which could possibly pass the authentication AI.

For the encryption and decryption process in this context, public or private key encryption can be used depending on the overall planned interoperability.[6] [7] There are also other encryption methods, and it is recommendable to use the one currently in use for military operations due to proven security.[8] This paper will not go into more detail about encryption as it is primarily focused on the authentication process within remote controlled military context.

[6] Compare: (Küchlin, 1987)
[7] Compare: (Don Davis, 1990)
[8] Compare: (Amit Sahai, 2010)

7. Conclusion

The goal of this paper was to contribute to the question “To what extend can unmanned assets interact with humans in the filed for defense operations?” This was achieved by outlining the necessary steps to ensure secure communication in a military context. After introducing the authentication process in chapter three, the paper continues to elaborate Natural Language Processing as the basis for chapter five and the AI for authentication of human behavior. Lastly, chapter six outlines the importance of proper encryption which is already used in the field and can be adapted for the use in this context.

The proposed authentication process will have substantial benefits and allow humans to communicate with unmanned assets more securely. As introduced in chapter one, this becomes more and more important in a more connected and technologically advanced world. However, it is crucial for this authentication process to be tested and validated thoroughly in practice to ensure an error rate of 0 %. Due to the fact that this paper only focused on the theoretical aspect of secure communication, it is necessary to further explore the idea practically. This can be tested in small scale using a Raspberry Pi or better an NVIDIA Jetson Nano which is better suited for AI applications. After validating the prove of concept in this enclosed setup, it is recommendable to further test the idea on a larger scale during an extended period of time of several years to thoroughly train the AI and eliminate any problems. Only after intense testing, this system should be implemented in the actual field for further testing.

As stated in the first proposal, the proposed idea does not have a direct socio-economic, political, ethical, or environmental impact. It is, however, important to quickly state some important indirect impacts. The implementation of proper authentications measures ensures that humans keep communicating with unmanned assets securely. With this basis, it is not necessary to create unmanned assets which operate completely autonomous without human control which would create many ethical problems. Using the proposed secure communication, the stated ethical problem does not need to be created. Moreover, secure remote-control operations save human lives by not having to rely on humans in the cockpit of a plane or tank but rather in a safe base many miles away from the combat situation. Moreover, being able to remove the cockpit from an asset creates more room to be used for crucial technology. Both mentioned points have a large socio-economic impact on military applications. Removing the cockpit from an asset also reduces weight and drag which in turn reduces emissions and has an environmental impact.

We hope our idea can make an impact to save lives and make communication with unmanned assets more secure.

8. Sources

Amit Sahai, H. S. (October 2010). Worry-free encryption: functional encryption with public keys. *CSS '10: Proceedings of the 17th ACM conference on Computer and communications security*, 463-472.

Don Davis, R. S. (1990). Network security via private-key certificates. *ACM SIGOPS Operation Systems Review, Volume 24, Issue 4*, 64-67.

Jair. (2018). *Stack Exchange*. Von Voice Recognition with Fourier Transform: https://datascience.stackexchange.com/questions/38501/voice-recognition-with-fourier-transformation-with-audio-input-in-python abgerufen

John Snow LABS. (2022). *NLP Webinar*. Von https://colab.research.google.com/github/JohnSnowLabs/nlu/blob/master/examples/webinars_conferences_etc/ny_nlp_meetup_2022/NY_NLP_webinar.ipynb#scrollTo=yiU5oCWGz31e abgerufen

Küchlin, W. (August 1987). Public key encryption. *ACM SIGSAM Bulletin, Volume 21, Issue 3*, 69-73.

Nagpal, I. G. (2018). *Artificial Intelligence and Expert Systems.* Dulles, Virginia: Mercury Learning and Information.

Supriya Rao, N. C. (31. October 2008). Peoplee detection in image and video data.

Tizeta Zewide, S. A. (December 2009). Audio data model for mulit-criteria query formulation and retrieval.