



SPEECH SENTIMENT ANALYSIS FOR HEALTHCARE

A master thesis by

JULIA GARCÍA LORENTE

advised by

Dr. Peter Seifter

and submitted to the

Institute for eHealth

of the

Graz University of Applied Sciences

in partial fulfillment of the
requirements for the degree of a
Master of Science (MSc)

July 2018

Acknowledgements

To my brother, Jaime, who has dedicated so much time and effort dealing with me and helping with this work.

Abstract

Recently, Sentiment Analysis has been one of the topics with a biggest relevance in the field of artificial intelligence. With the help of this tools, computers are able to understand people, this means, to understand the intention on their messages. In particular, Sentiment Analysis technologies allow to know the feeling transmitted in the message, either by facial, speech or voice analysis.

In the following study the benefits of Speech Sentiment Analysis orientated to health care are going to be analyzed and, most notably, orientated to telenursing and telephone triage. Some of the current developed technologies will be tested facing a dataset with 22 samples of opinions transmitting different sentimental orientations. Hereunder, the study will carry on with a proposal of development of an Speech Sentiment Analysis tool specifically orientated to health care. Finally, the support of Voice Sentiment Analysis in combination with Speech Sentiment Analysis will be suggested , with the aim of improving the accuracy of these software for telephone triage assistance.

In conclusion, this studium is hoped to be useful as reference in future development of Sentiment Analysis based softwares orientated to telephone triage and telenursing assistance.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit Stimmungsanalyse als eines der wichtigsten Themen im Bereich der künstlichen Intelligenz mit großer Relevanz für das Gesundheitswesen. Mit Hilfe von Software-Lösungen sind Computer in der Lage, Menschen zu verstehen, das heißt auch die Intention ihrer Botschaften zu verstehen. Neuartige Sentiment-Analyse-Technologien ermöglichen es, das Gefühl zu kennen, das in der Botschaft vermittelt wird, entweder durch Gesichts-, Sprach- oder Stimmanalyse. In dieser Arbeit werden die Vorteile der Sprachstimmungsanalyse, die sich an der Gesundheitsversorgung und vor allem an Telenursing und Telefon-Triage orientieren, analysiert. Ausgewählte, aktuell entwickelte Technologien werden anhand eines Datensatzes mit zwanzig beispielhaften Meinungen, die unterschiedliche sentimentale Orientierungen vermitteln, getestet. Darauf aufbauend wird ein Prototyp einer speziell auf das Gesundheitswesen ausgerichteten, mobilen Anwendung zur Sprachstimmungsanalyse entwickelt. Abschließend wird die Unterstützung von Voice Sentiment Analysis in Kombination mit Speech Sentiment Analysis untersucht. Das Ziel ist die Genauigkeit für die telefonische Triage-Unterstützung durch diese Software zu verbessern. Zusammenfassend dient diese Arbeit als Referenz für die zukünftige Entwicklung von Sentiment Analysis basierter Software, die sich an der telefonischen Triage und Telenursing-Hilfe orientiert.

Contents

Acknowledgements	ii
Abstract	iii
Zusammenfassung	iv
List of figures	vii
List of tables	viii
1 Introduction	1
1.1 What is Speech Sentiment Analysis and what is it used for?	1
1.2 How could Speech Sentiment Analysis be useful in healthcare sector? . . .	2
1.2.1 Speech Sentiment Analysis for consumers and providers in Healthcare	3
1.2.2 Speech Sentiment Analysis for individuals in Healthcare	3
2 How does Sentiment Analysis works?	5
2.1 Objectives	5
2.2 Collecting the data	5
2.3 Building the model	7
2.4 What can be classified with Speech Sentiment Analysis?	9
3 Testing current technologies	11
3.1 Materials and methods	11
3.1.1 Tools	12
3.1.2 Data	13
3.2 Results	20
3.2.1 Microsoft Azure	21
3.2.2 Google Cloud	25
3.3 Conclusions	30

4	Abstract development of a prototype	32
4.1	Objetives	32
4.2	Working of the software	32
4.3	Design of the interface of the App	33
5	Further steps	37
5.1	Speech Sentiment Analysis using voice patterns	37
5.2	Voice properties	38
5.3	Working of voice analysis algorithms	39
5.4	Conclusion	41
	Bibliography	42
	Internet sources	43
A	Complementary data	44
A.1	Raw output of Text Analytics API	44
A.2	Raw output of Google Cloud Natural Language API	47
B	Codes	48
B.1	Cleaning text	48
B.2	Internet behavior	50
B.2.1	MainActivity	50
B.2.2	InstructionClass	57
	Obligatory Signed Declaration	58

List of Figures

1.1	Use of Sentiment Analysis in commercial management	2
1.2	Telenursing and Telephone Triage	4
2.1	Process for collecting structured data.	6
2.2	Example of working of the k-nearest neighbors.	8
2.3	Sentimental Dictionary groups	10
3.1	Comparison the differences between the obtained and the expected result in texts in english and in spanish with Text Analytics API	24
3.2	Comparison the differences between the obtained and the expected result in written and spoken texts with Text Analytics API	25
3.3	Comparison the differences between the obtained and the expected result in texts in Spanish and in English with Google Cloud Natural Language .	29
3.4	Comparison the differences between the obtained and the expected result in spoken and written texts with Google Cloud Natural Language	29
4.1	Instructions for the use of the buttons in the app SpeakToMe	34
4.2	Initial main screen of the app SpeakToMe.	35
4.3	Main screen of the app SpeakToMe recording.	35
4.4	Instructions for the use of the progressbar in the app SpeakToMe	36
4.5	Ascending scale of damage severity in app SpeakToMe	36
5.1	Sound wave's envelope marked in red [Dagupta, 2017]	39
A.1	Screenshot of output from commentary 1	47

List of Tables

3.1	Selected commentaries for the testing of Text Analytics and Natural Language API.	20
3.2	Results of transcription of commentaries with Bing Speech.	22
3.3	Results of Sentiment Analysis with Text Analytics API.	23
3.4	Results of transcription of commentaries with Google Cloud Speech-to-Text API.	27
3.5	Results of Sentiment Analysis with Google Cloud Natural Language API. .	28

Chapter 1

Introduction

1.1 What is Speech Sentiment Analysis and what is it used for?

Speech Sentiment Analysis is called emotion recognition too. In this term is included every computer tool which analyzes an speech transcription (a text) and obtain the general feeling and the intention out of it.

This could sound far, but actually, people are using it everyday in their smartphone, tablet or computer. How many times have we ask...:

"Siri, how will be the weather today?"

"Cortana, remind me to buy fresh fruit"

How could they be able to understand us and even to answer back?. The answer is: with the help of Speech Sentiment Analysis and Natural Language Processing. The fact that a person can speak to the computer and the phone, makes them friendlier and give an idea of intuitiveness. Since they actually fulfill the orders or provide instruction for achieving the objectives, they are more intuitive.

Sentiment analysis has been a hot research topic and has outgrown from business intelligence for commercial products to other disciplines like social sciences, politics, geography, management sciences, health care, stock market etc. [K.Taimoor, 2016]. This appliance allows to know, in a fast way, the opinion of an individual over a concrete product or

service. This, in combination with Big Data Technologies, gives a wide comprehension of what is liked or unliked about this proposal.



Figure 1.1: Use of Sentiment Analysis in commercial management

Regarding to this, in recent years, an idea of insecurity and lack of privacy has been growing among the population, due to the great amount of personal information this tools manage. Actually, the information managed by sentiment analysis tools are minimal: the series and movies you've liked, your favourite music and videos, etc.

1.2 How could Speech Sentiment Analysis be useful in healthcare sector?

This type of analysis reveals a trend of public opinion over a period of time. It can be used to track the performance of a patient, instrument or health center, where the ones with dropping performance can be pointed out. Normally people are reluctant to new procedures of treatment, so this could mean a change in the perception of people. The purpose of health care sentiment analysis is to empower the patients to make safe and preferred health related decisions. [J.C.Gonzalez and E.Valencia, 2017] [K.Taimoor, 2016]

Continuing with written previously, Sentiment Analysis could be used in the field of health in two ways: oriented to groups or oriented to individuals. [T.Russell-Rose, 2014]

1.2.1 Speech Sentiment Analysis for consumers and providers in Healthcare

As in every other field, Speech Sentiment Analysis could be used to evaluate the general opinion about a health service or product. For example, the dynamics of an emergency department, the services of an hospital, or even the effectiveness of a medication.

A great amount of social websites have been created with the aim of expressing the opinions of the consumers about almost every healthcare service provided. Even many hospitals, like AKH (Allgemeines Krankenhaus der Stadt Wien) in Vienna or Hospital Universitario de la Princesa in Madrid, have specific sections in their websites which allows patients to write about their experience in the hospital and provide it with feedback.

All this information can be used later by the very providers to analyze the verdict about the product and the reason for this verdict. [J.C.Gonzalez and E.Valencia, 2017]

1.2.2 Speech Sentiment Analysis for individuals in Healthcare

These tools could be individual oriented too. In the last years, much has been written and spoken about personalized medicine. For that ones of us that don't know exactly what the term *personalized medicine* means, here we have a definition:

"A type of medical care in which treatment is customized for an individual patient."

Regarding to this mentality, we can relate the concept of personalized medicine with the *Telemedicine*. Telemedicine is defined as:

”The remote diagnosis and treatment of patients by means of telecommunications technology.”

In different parts of the world, a new branch of Telemedicine has been developed: Telenursing. This service allows nurses and doctors to communicate with patients out of the clinic. It is really useful with treatment of chronic patients, patients with reduced mobility, and an specially relevant use for this work: telephone triage. Telephone triage refers to symptom or clinical calls. Clinicians perform assessment by asking detailed questions about the problem. The clinician’s task is to estimate the severity of the symptoms. The symptoms are basically classified as life-threatening, emergency, urgent, acute or non-acute. In this way, clinicians are able to manage incoming patients and dispatch them to the needed service: emergencies, general doctors or specialist. For that aim, nurses have a support software where they introduce the answers of the patient generating a recommendation of service as a result, depending on the reason and severity of the problem. [H.Asmat et al., 2017] [J.Li et al., 2017]



Figure 1.2: Telenursing and Telephone Triage

Chapter 2

How does Sentiment Analysis works?

2.1 Objectives

This powerful tools have the ability of working with data of different areas as well as a concrete one. Nonetheless, the behavior of the tool is strongly affected by the chose of the focused data. This means, if the tool is oriented to deal with varied data, the accuracy in front of data of an specific field will not be so precised as a tool oriented to that specific area.

Another decisive aspect of this technologies is the way they treat the received data: group the data, score it, etc.

This two points are the more remarkable for the correct behavior of this tools, but, of course, much more should be taken into account: the collecting method of the data, the format of the results and so on. [J.C.Gonzalez and E.Valencia, 2017]

In the following sections we are going to see a more exhaustive explanation about this two main aspects.

2.2 Collecting the data

For this part there are two important things to take into account: what kind of data we need, and where to take it from.

As we have learned before, sentiment analysis has been used in a wide range of fields: politics, commercial analysis, management sciences, and for develop digital assistants. It is obvious that for purposes like politics or commercial analysis, our tool must work in a concrete field. Nevertheless, tools like digital assistants must feed from a great variety of data so they are prepared for any consult.

Based on this, an initial dataset is build with the following process:

First, we have to select an field, or group of fields, of topics our software will have to face. This kind of data can be found on social networks, forums, etc. This is the raw data that conform the initial dataset. This initial dataset is unstructured due to the information into it has not been processed jet. For that aim, words must be sintactically classified as verbs, nouns, prepositions and so on. With this clasification, words are compared with an external sentiment dictionary and the components that do not add sentimental content are discarded from the set. The same external dictionary allows us to label every word with the proper sentiment and intensity. So finally, we have this structured dataset with words classified by syntax, sentiment and strength. [J.C.Gonzalez and E.Valencia, 2017] [T.Mullen, 2011]

This external sentiment dictionary has been prebuild.

The following image has been obtained from [T.Mullen, 2011].

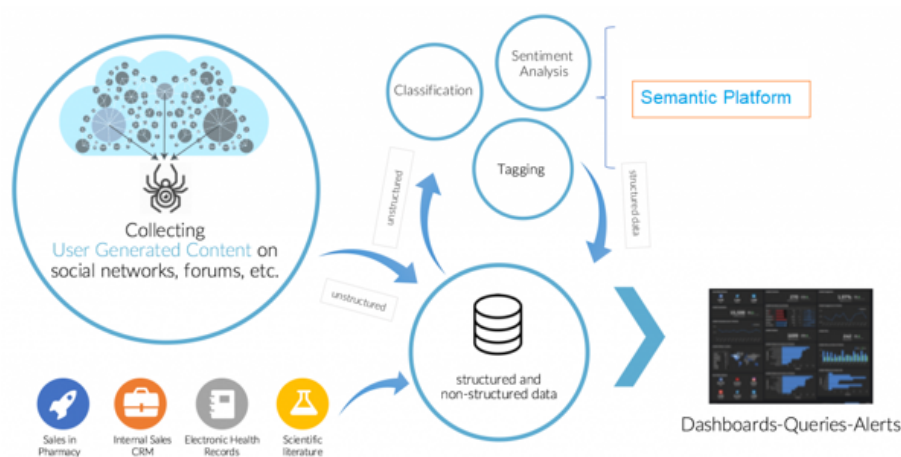


Figure 2.1: Process for collecting structured data.

2.3 Building the model

The process for building a model is simple and it can be divided in to phases. The dataset will be divided in two sections: one section for training the model and another one for testing it, so we have a phase of training and a phase of test. This two divisions in the dataset do not need to be equally big. In fact, it is usual that the training part is much bigger than the test part. In the training phase groups are made basing on the similarities between the data. This means that if we consider similarity between each piece of data as a distance, data that is closer to each other will conform a group. For that, we need to know how many groups we want to conform.

The most commonly used classifiers for Sentiment Analysis are Naive Bayes kNN(k-Nearest neighbour), Centroid Based and SVM(Support vector machine). This methods of calculating groups are iterative and there are different ways of conforming them. If we need i groups, the software takes i individual components as initial centroids, and calculates their distances with the other pieces of data. The centroid changes always looking to minimize the distance with the closest pieces and maximize it with the furthest. The final groups are form with the closest components to the centroids. Another method is to add to the set i auxiliary centroids, so they move following the same rule. The results can be presented with the data as defined part of a group, or with a percentage. This percentage depends on the distance of the data in respect to every centroid, so each piece will have a percentage for each group. This means that, if one piece is really close to the first centroid than to a second one, will have a higher percentage of membership of the first group.

Lets see an example: we have a group of people: women and men, who are blond and dark headed. If we ask the software to build two groups, it surely separate them in women and men, or in blond and dark headed. In the other hand, if we ask the software to build four groups, it will divide them in blond women, dark headed women, blond men and dark headed men.

This training process can be repeated so many times as wished. It is usual too to conform n training groups and the final model will be an average of the results.

The test part of the dataset is used to test the accuracy of the building model.

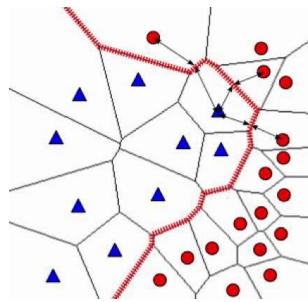


Figure 2.2: Example of working of the k-nearest neighbors.

With this cleaned dataset we can start to develop our model. In this part there are two options: the model can be build with supervised or unsupervised techniques:

- Supervised techniques. This kind of techniques build the model from an initial dataset. This dataset needs to be prelabelled. The set must be properly large, so it will not lead to overtraining errors. This errors appear when the model do not have enough information to learn from, so when it faces new information with which it has never work with, it classifies this new data in the wrong way.
- Unsupervised techniques. This kind of techniques are constantly developing the model due to their feeding with a continuous flux of information. This information does not need to be prelabelled. Obviously, this methods need from a bigger power, but it is really useful facing problems of really diverse and changing areas. This is the method that use translators or digital assistants.

Summarizing, models built with unsupervised techniques are more accurate facing wide different problems, but the ones built with supervised techniques do not need such an amount of power and are able to achieve an excellent accuracy in a defined area. Nevertheless there is another option: semi-supervised techniques. This techniques are also proposed where the probabilistic model require some domain specific user intervention. The hybrid require a small training set to identify initial values for parameters that are later used for building an unsupervised model. [K.Taimoor, 2016]

2.4 What can be classified with Speech Sentiment Analysis?

In Speech Sentiment Analysis is interesting to classify words, sentences and texts. [T.Mullen, 2011]

Classifying words allows not only to understand sentences, but to find *keywords*. This keywords should be the major responsible of the general feeling in the sentence or text, or a synonym. This keywords can be identify with Data-driven methods, unigram methods or topic-based binary text classification. This methods are able to generate a keyword list with a certain accuracy.

In the other hand, sentences and texts are classified in order to know the general feeling about a topic. There are two ways to score a sentence or a text:

- Heuristic methods. This methods are based in databases of words manually created with scores of positivity and negativity. The database can also include relationships between synonyms and antonyms. Some examples of this methods are SentiWord-Net, Whissell's Dictionary of Affective Language or General Inquirer. This are the external dictionaries mentioned in the "*Collecting Data*" phase of the process of Sentiment Analysis. This *dictionaries* allow us to sum the scores of the different words in a sentence or text so the final result is considered the general sentiment.
- Sentiment-oriented data sets. This sets are reviews of complete texts. They divide the units in positive or negative too. Some examples of this sets are the Blitzer et al Multi-domain sentiment dataset and the Pang & Lee datasets. The datasets are formed by self-annotated or hand-annotated data. An example of self-annotated data are commentaries on amazon, where the own author can score the positivity or negativity of the commentary. Moreover, hand-annotated data is not scored by the author himself, but by "*annotators*". Annotators are people specifically dedicated to produce sentiment analysis of text. Hand-annotated datasets are obviously not productive for general sentiment analysis, but lets remember prelabelled data is needed in supervised techniques. [T.Mullen, 2011]

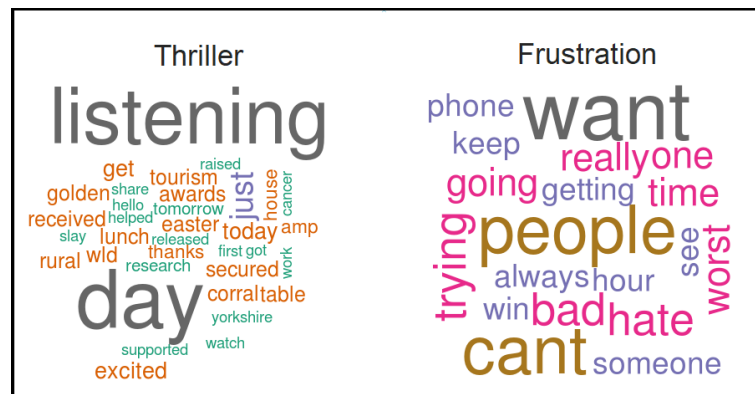


Figure 2.3: Sentimental Dictionary groups

Chapter 3

Testing current technologies

3.1 Materials and methods

In this section current tools for Speech Sentiment Analysis are tested, but before that, lets see a wide framework of these technologies.

Nowadays, it exist a great variety of tools for Sentiment Analysis developed by different companies and for different uses. Here we have some examples: [T.T.Pham, 2017]

- Lingmotif. It is the result of the project FFI2011-25893 begun in 2012 by the tean Tecnolengua of the University of Málaga. This tool is able to determine the semantic orientation of a text: if it is positive or negative and the intensity. Furthermore, lingmotif does not just classify the text, but selects the fragments responsible from the diagnosis. The results are shown in HTML format, and this tool is able either in Spanish and English. It is available for academic use in Mac OS, Windows and Linux. [A.Moreno-Ortiz, 2016]
- Sentiment Analysis API. Developed by Meaning Cloud, it can extract the positive, negative and neutral polarity in the text. The peculiarity in this tool is that it can detect irony and polarity disagreement. Sentiment Analysis API is able to work with English, Spanish, French, Italian, Catalan and Portuguese texts. The console is available online. [MeaningCloud, 2018]

- IBM Watson Alchemy Language. Developed by the International Business Machines Corporation (IBM) in 2017, it analyzes words and shows the results of the text analysis based in these keywords and their score. It works observing relations of words in the sentence and the text, and giving more weight to that keywords that contribute more in a sentimental aspect of the text. Watson is able to relate information of the text not only inside it, but from others online sources. For example, results can show the author of the text or publication date. For developers, it works in Node, Java, Python and iOS SDK. [Cloud, 2017]

Additionally, we find that the two most popular tools in Sentiment Analysis are Text Analytics API of Microsoft Azure and Google Cloud Natural Language from Google Cloud.

None of this tools is orientated specifically to healthcare. However, the aim of this part of the work is to check if this non-specialized technologies are acceptable for use them as the support software in Telephone Triage. Since they are the most popular and therefore they have more possibilities of being the most accurate and intuitive tools, we are testing in this section Text Analytics API and Google Cloud Natural Language. Furthermore, this two are supposed to show the most intuitive results of the analyzed text, so it should be perfect for an appropriate work in a service that needs speed and clarity.

3.1.1 Tools

Then, we are going to explain what are Text Analytics API and Google Cloud Natural Language, and the complementary tools we are using for supporting them in our purpose. Lets remember that nowadays sentiment analysis technologies work only with texts, not with voice recordings, so first, we need a method for recording and transcribing the speech, so the sentiment analysis tool can work with that transcription.

A list of technologies and their details used for this testing part can be found below:

- Microsoft Azure
 - Text Analytics API. It allows us to analyze sentiment, extract key phrases and words and detect the language of the text.

- Bing Speech. This tool allows us to transcribe speech into text. Bing Speech can work with speech in English, Chinese, French, German, Italian and Spanish.
- Google Cloud
 - Google Cloud Natural Language API. This API includes sentiment analysis, entity analysis, entity sentiment analysis, content classification and syntax analysis. It can work with text in Chinese, English, French, German, Italian, Japanese, Korean, Portuguese and Spanish.
 - Cloud Speech-to-Text API. This is the tool of Google Cloud to transcribe speech into text. It can work with a great variety of languages, for example: Afrikaans, Catalan, Danish, German and even Zulu.

3.1.2 Data

Since we do not have access to the kind of files that can result from telephone triage, some data have been collected for this test. The 22 selected files contain opinions and commentaries of products. The files have been recorded from videos in Youtube and commentaries in Amazon. They can be divided depending on the communication medium or language. In one hand we have 11 text from spoken speech and 11 written texts. Since the chosen vocabulary in a written text uses to be more elaborated than in spoken speech, in this way, we will evaluate the accuracy of the tool facing formal and informal language. In the other hand, 11 of these files are in Spanish and 11 in English. In this way, the accuracy either facing different mediums or languages will be tested.

Furthermore, sentiments in the texts from Amazon are already scored by their own authors, but some of the files from Youtube do not have a define score of positivity or negativity of the commentary. To solve that, a inquiry shall be done. In this inquiry we have ask 10 people to score this commentaries from 0 to 10 depending on the positivity or negativity that they perceive in the text. The scale works as follows: a score of 0 means an absolute displeasure by the tester, besides 10 expresses an absolute satisfaction with the product. Due to each tool uses different scales, this is the easiest way to compare the obtained results.

Selected commentaries with their each score and specifications can be seen below.

Number	Commentary	Obtained score	medium and language
1	<p>“Oigan, perdón, pero este es el mejor y más sencillo slime. Sigo sin poder creerlo. Internet win.”</p> <p>“Listen, excuse me, but this is the best and simple slime I’ve ever done. I still can’t believe it. Internet win.”</p>	9,0	Spoken, Spanish
2	<p>“Sabe bien, osea, no sabe a queso, sabe como a papa. Y dicen que gratina, así que vamos a probar eso. Pues miren, no gratina así que digas:” ah, no ma que buen gratín”, no, pero sí medio gratina. Osea, no mucho, pero es lo que hay, amigos. Lo voy a probar. Está superbién, no sabe a queso, pero sabe a win.”</p> <p>“It tastes good, I mean, it doesn’t taste like cheese but like potatoes. It’s said that you can gratin it, so let’s try that. Look, it doesn’t gratin so much you’ll say: “ah, so good gratinated”, no, but it almost does. I mean, it’s not so much, but it’s what we have, friends. I’m going to taste it. It tastes great, it doesn’t taste like cheese, but like win”</p>	5,9	Spoken, Spanish
3	<p>“Bajen Clue, de verdad, no saben la ayuda que les va a hacer. Es gratis, está para iOS y para Android.”</p> <p>“Install Clue, really, you don’t know of how much help it will be. It’s free, you can find it for iOS and Android.”</p>	8,1	Spoken, Spanish

Number	Commentary	Obtained score	medium and language
4	<p>“Apenas llevo 4 meses usándola, creo que voy a perfeccionar estos problemas con el tiempo. En general, creo que sí fue un win, porque sí la puedes tener más tiempo. Bueno, los días que te baja muy cabrón si la tengo que estar cambiando cada 5 o 6 horas, pero ya en los días leves la puedes dejar y eres feliz, y sí, es un superwin.”</p> <p>“I’ve barely been using it for 4 months, so I think I will solve this problems with time. In general, I think it was a win, because you can use it for longer. Well, my hardest days I had to change it every 5 or 6 hours, but the lighter days you can leave it there and be happy, and yes, is a great win.”</p>	6,5	Spoken, Spanish
5	<p>“La verdad es que no es tan difícil como me pensaba. Es fácil hacerlo, es cuestión de practica y lo que mas me gusta es que se queda fijo, lo tocas y no se mancha. Me parece innovador, y solo por eso, solo por que prueben cosas nuevas, me parece una buena iniciativa, y por eso le voy a dar un OK.”</p> <p>“The truth is that it’s not so difficult as I thought. It’s easy to use, it’s a matter of time and what I like more is that it’s set, you can touch it and it doesn’t stain. I find it innovative, and just for that, just because they try new things, I find it a good initiative, so I will give it an OK.”</p>	9,0	Spoken, Spanish

Number	Commentary	Obtained score	medium and language
6	<p>“Es una solución más que podéis utilizar para exfoliar y activar la circulación sanguínea, pero no esperéis nada milagroso o fuera de serie. Al menos, es lo que me ha parecido a mí. Eso sí, no os voy a negar que no exfolia la piel, no es nada agresivo, ni que al final es una solución más que podéis utilizar para el cuidado. Eso sí, a mí me huele un poco raro, es un olor que me molesta. Así que, como yo no lo utilizaría, le voy a dar un intermedio.”</p> <p>“It’s another solution you could use for exfoliate and activate your blood flow, but you musn’t expect a miracle. At least, this is my impression. I won’t deny that it doesn’t exfoliate your skin in an agresive way. But, in truth, it smells a little weird, and it bothers me. So, since I wouldn’t use it, I’m going to give it a maybe.”</p>	2,0	Spoken, Spanish
7	<p>“Por lo que me ha costado, yo lo recomiendo, así que le voy a dar un OK.”</p> <p>“Just for the price, I recommend it, so I’ll give it an OK.”</p>	8,0	Spoken, Spanish

Number	Commentary	Obtained score	medium and language
8	<p>“Para hacer la gracia, para regalárselo a alguien, para pasar un buen rato y eso, sí. Como tratamiento, a menos que os apliquéis un poco el producto sobrante, lo veo un poco... distinto, así que yo le daría un intermedio o un no me gusta.”</p> <p>“Just for the laughs, for giving it to someone, for having fun and that, yes. As a real treatment, except you will use the excess of product, I find it a little bit... different, so I would give it a maybe or a dislike.”</p>	3,0	Spoken, Spanish
9	<p>“So, that was actually sort of scary. I feel like it’s sort of dangerous messed with your nails like that, but it worked really well. I’m super happy with the way they look now and I think that was a pretty good buy for like 15 bucks. Totally awesome.”</p>	8,1	Spoken, English
10	<p>“I swear other people have done this and it’s gone so well while mine is not working out. Guys, this is not working. Kind of smells like bubblegum. I guess I would have to sit and wait a minute and see but it’s already stick. Maybe this brand is not the one, because I swear for all the people it works.”</p>	2,5	Spoken, English

Number	Commentary	Obtained score	medium and language
11	<p>“The texture is actually pretty sticky, but I think it dries up so it’s not sticky anymore. I honestly don’t know how I feel about this product. I guess that’s a good way of repairing wrinkles, but, do I think it’s necessary? Not, not really.”</p>	2,3	Spoken, English
12	<p>“Después de casi 5 años probando productos, ninguno me lo llevo a quitar excepto este, muy contento con el resultado y solo llevo un par de semanas de uso. El precio está muy bien. Volveré a comprarlo.”</p> <p>“After almost 5 years testing products, none of them achieve it but this one. Really happy with the results even when I’ve been using it for just a paar of weeks. The price is good. I’ll buy it again.”</p>	10	Written, Spanish
13	<p>“No es milagroso contra las picaduras de mosquitos. Después de un mes en Colombia tengo tantas picaduras como cualquier otro repelente del mercado, por lo que con los gastos de envío la relación calidad-precio no es muy ventajosa.”</p> <p>“It’s not a miracle against mosquito bites. After a month in Colombia I have as much stings as I would have with any other repellent in the market. So, taking into account the shipping costs, the rate quality-price is not so good.”</p>	6,0	Written, Spanish

Number	Commentary	Obtained score	medium and language
14	“I should’ve known when I saw the ingredients. The active ingredient, allantoin is used to protect chafed, chapped, cracked skin. But it doesn’t reduce the appearance of old/healed scars.”	4,0	Written, English
15	“Llego correcto y rápido. La verdad no viene siendo más que una plantilla de mejor calidad, yo no he notado ninguna diferencia positiva reseñable, precio caro para lo que es.” “It arrived in perfect shape and fast. The truth is that it’s just a better quality slipsole, I have notice no remarkable difference. The price is high for the product. ”	4,0	Written, Spanish
16	“These work, it’s not as accurate as blood test, but as a guide, tis works well. Happy to have bought, would recommend.”	8,0	Written, English
17	“Great oil, has so many uses. I get boils on my face now and then. Normally they will last a month but I use this as I can feel it coming and it’s gone within 2 days. I’m so pleased with it :)”	10	Written, English
18	“I’ve used a lot of earplugs due to being a really light sleeper. I’d have to say these are the best by far. They’ve always stayed in my ears and cut enough for me to stay sleeping. I’ve also found that as my ears are so small I’ve been able to get 6 pairs instead of 3. Bonus!”	10	Written, English

Number	Commentary	Obtained score	medium and language
19	“very good substitute for cigarettes, but appealing enough to get you hooked to it.”	6,0	Written, English
20	“ok it’s the real thing. It smells horrible, like turpentine but you buy it for it’s benefits not for smell. Now about benefits, I cannot say it helped for my skin acne. Perhaps if you use it every day for weeks but I could not do that. There are better products who work faster.”	6,0	Written, English
21	“I like this scrub. However be light handed when using it as the granules are quite large and abrasive. It dows the job but I wouldn’t buy it again.”	8,0	Written, English
22	“this stuff works. One sachet contains enough to las a few uses, but I personally apply this to my t-zone and it works. The very first time I used it on my face all sorts of dirt came out which is both amazing and a great relief.”	10	Written, English

Table 3.1: Selected commentaries for the testing of Text Analytics and Natural Language API.

3.2 Results

In the following section we will evaluate the tools of Microsoft Azure and Google Cloud and analyze the results.

3.2.1 Microsoft Azure

Bing Speech

Let's remember Bing Speech is the tool that allows us to transcribe speech into text. There are two options for using this API: the option 1 make transcriptions of less than 15 seconds of speech and the option 2 is supposed to remain transcribing till it is told to stop. Due to the length of our testing data, the second option is chosen. It is important for the aim of this studium, since voicecalls use to take more than 15 seconds, even for an emergency.

Since this is the testing of the voice-to-text part it is not necessary to test it with all our commentaries, so in the following table it is shown the ones that could be the most problematic.

Number	Commentary	Obtained text
7	Por lo que me ha costado, yo lo recomiendo, así que le voy a dar un OK.	¿Por lo que me ha costado yo lo recomiendo así que le voy a dar un okey?
8	Para hacer la gracia, para regalárselo a alguien, para pasar un buen rato y eso, sí. Como tratamiento, a menos que os apliqueis un poco el producto sobrante, lo veo un poco... distinto, así que yo le daría un intermedio o un no me gusta.	¿Para hacer la gracia para regalárselo a alguien para pasar un buen rato y eso sí cómo tratamiento a menos que sea apliquéis un poco de productos obrante lo veo un poco de distinto así que yo le daría un intermedio aún no me gusta?
9	So, that was actually sort of scary. I feel like it's sort of dangerous messed with your nails like that, but it worked really well. I'm super happy with the way they look now and I think that was a pretty good buy for like 15 bucks. Totally awesome.	So that was actually short of scary. I feel like it's sort of dangerous messed with your nails like that, but it works really well. I'm super happy with the way they look now and I think that was a pretty good, bye for like 15 bucks. Don't tell you some.

Number	Commentary	Obtained text
10	I swear other people have done this and it's gone so well while mine is not working out. Guys, this is not working. Kind of smells like bubblegum. I guess I would have to sit and wait a minute and see, but it's already stick. Maybe this brand is not the one, because I swear for all the people it works	I swear other people have done these on it's going so well. Why mine is not working out guys. This is not working kind of smells like bubble gum. I guess I will have to sit and wait a minute and see but it's already stick. Maybe this run is not the one because I swear for all the people it works.
11	The texture is actually pretty sticky, but I think it dries up so it's not sticky anymore. I honestly don't know how I feel about this product, I guess that's a good way of repairing wrinkles, but, do I think it's necessary? Not, not really.	Texture is actually pretty sticky, but I think it dries up so it's not sticking anymore. I honestly don't know how I feel about this product. I guess that's a good way of repairing windows, but the thing is necessary not not really.

Table 3.2: Results of transcription of commentaries with Bing Speech.

After seeing these results, we can observe some characteristics:

- The tool transcribes in real time, showing some of the intermediate results while transcribing. Between all these intermediate results, it chooses the one that have the best level of "confidence".
- The tool needs from the user to choose the predetermined language. That means that if the API is used with a language different from the default one, it will not make a proper transcription.
- The maximum working time is 15 minutes.
- It does not work well while recording word contractions and some voice intonations.

Text Analytics API

Text Analytics is the Microsoft API for sentimental analysis. Appart of showing the resulted sentimental score, Text Analytics is able to detect the language of the text and provides a list with the key words for the result. The sentimental score is given in an scale of percentaje: 0% is the most negative, 50% is neutral feeling, and 100% is the most positive.

The following table shows the obtained scores in comparison with the original score of each commentary.

Number of the commentary	Obtained score (%)	Original score(%)
1	47	90
2	57	59
3	81	81
4	59	65
5	57	90
6	48	20
7	78	80
8	70	30
9	5	81
10	5	25
11	9	23
12	78	100
13	25	60
14	11	40
15	75	40
16	77	80
17	99	100
18	12	100
19	91	60
20	94	60
21	95	80
22	83	100

Table 3.3: Results of Sentiment Analysis with Text Analytics API.

With the obtained results it is possible to do two comparisons: the accuracy of the tool facing different languages and the accuracy facing spoken and written text. For this comparison, the difference between expected and obtained scores has been calculated in absolute value. This way, if this difference is big, it means the accuracy is not so good. To make this comparison more visual, two graphics have been drawn:

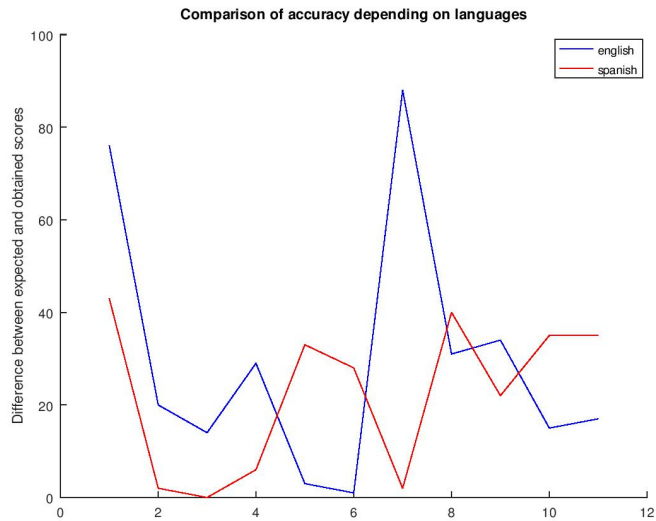


Figure 3.1: Comparison the differences between the obtained and the expected result in texts in english and in spanish with Text Analytics API

In this graphic we can observe an slightly better accuracy with texts in Spanish than with texts in English. The differences between the sentimental score obtained with the API and the original score are less pronounced in Spanish texts, but, since this is an small sample, we can just say that a better tendency could be observed facing texts in Spanish than in English. The two highest differences between scores can be found in English texts. Besides, just 2 of the 11 texts in English have a difference smaller than 10 points, while in the 11 texts in Spanish we can find 4 texts in this range.

In this graphic it is observed that the accuracy does not have a great dependency in the formality of the text. We can not observe a particular tendency in favour of any of them (spoken or informal or written or formal language).

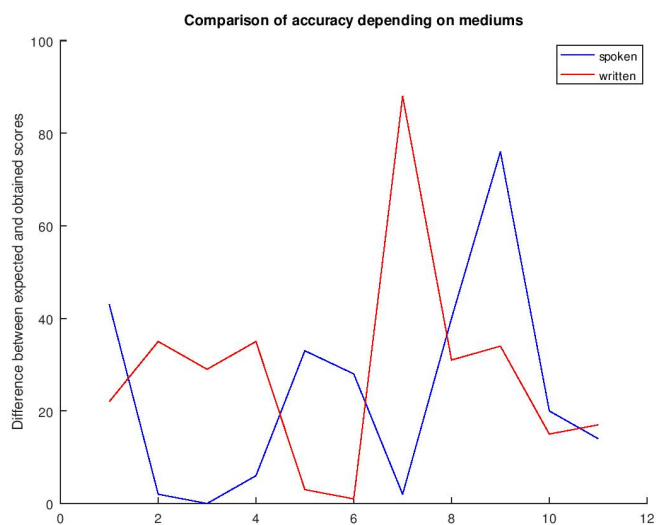


Figure 3.2: Comparison the differences between the obtained and the expected result in written and spoken texts with Text Analytics API

In general, we have only 2 cases were the difference between the expected and the obtained score have been bigger than 50 points. Furthermore, in 11 cases this difference is less than 30 points.

3.2.2 Google Cloud

Cloud Speech-to-Text API

Cloud Speech-to-Text API is the tool of Google for transcribing speech into text. In the following table it is shown the results of transcriptions of the commentaries 7, 8, 9, 10 y 11.

Number	Commentary	Obtained text
7	Por lo que me ha costado, yo lo recomiendo, así que le voy a dar un OK.	por lo que me ha costado yo lo recomiendo así que le voy a dar un OK

Number	Commentary	Obtained text
8	Para hacer la gracia, para regalárselo a alguien, para pasar un buen rato y eso, sí. Como tratamiento, a menos que os apliqueis un poco el producto sobrante, lo veo un poco... distinto, así que yo le daría un intermedio o un no me gusta.	para hacer la gracia para regalárselo a alguien para pasar un buen rato y eso sí como tratamiento a menos que os apliquéis un poco el producto sobrante lo veo un poco distinto así que yo le daría un intermedio o un no me gusta
9	So, that was actually sort of scary. I feel like it's sort of dangerous messed with your nails like that, but it worked really well. I'm super happy with the way they look now and I think that was a pretty good buy for like 15 bucks. Totally awesome.	So that was actually short of scary. I feel like it's sort of dangerous messed with my nails like that, but it works really well. I'm super happy with the way they look now. I think that was a pretty good buy for like 16 bucks. Totally awesome.
10	I swear other people have done this and it's gone so well while mine is not working out. Guys, this is not working. Kind of smells like bubblegum. I guess I would have to sit and wait a minute and see, but it's already stick. Maybe this brand is not the one, because I swear for all the people it works	I swear other people have done this and it's gone. So well while mine is not working out like this is not working. I know if it smells like bubblegum, I guess I would have to see them wait a minute and see but it's a red stick. Maybe this one is not the one because I swear for all the people it works.

Number	Commentary	Obtained text
11	The texture is actually pretty sticky, but I think it dries up so it's not sticky anymore. I honestly don't know how I feel about this product, I guess that's a good way of repairing wrinkles, but, do I think it's necessary? Not, not really.	The Dexter is actually pretty sticky but I think it's rice up. So it's not sticky anymore. I honestly don't know how I feel about this brother. I guess that's a good way of repairing wrinkles. But do I feel it's necessary. Not not really.

Table 3.4: Results of transcription of commentaries with Google Cloud Speech-to-Text API.

Taking into account these results we can make a pair of observations:

- Cloud Speech-to-Text does not transcribe in real time, but it waits till a long silence of a pair of seconds for showing the results of that block. After recording a paragraph it chooses the most probable transcription like Bing Speech. The user can keep speaking immediatly after a transcription without stopping the record.
- This tool need too from the user to choose the language of the transcription.
- The maximum working time is 3 minutes.
- Cloud Speech-to-Text works well facing word contractions and has a good accuracy in open spaces with noise.

Google Cloud Natural Language

Google Cloud Natural Language includes in his analysis of the text syntax analysis and language recognition. The sentimental analysis is either of each sentence in the text and of the whole text. The sentiment analysis give two results: Score takes a range between -1 and 1 and expresses the positivity or negativity of the feeling, besides, Magnitude takes a range from 0 to infinite that express the strength of this feeling. Google Cloud Natural Language can recognize relations between words. This means for example, that

it recognize the words "J.K.Rowling" and "novelist" as the same term in a concrete text.

For evaluating this API we will proceed in the same manner as with the Language Analytics API. Since we do not have a punctuation for the strength of the feeling as Magnitude, just the value Score will be used for this comparison. In the following table it is shown the results of the sentiment analysis with Google Cloud Natural Language.

Number of the commentary	Obtained score (%)	Original score(%)
1	0.1	0.9
2	0.1	0.19
3	0.2	0.62
4	0.4	0.3
5	0.3	0.8
6	-0.4	-0.6
7	0.6	0.6
8	-0.1	-0.4
9	0.6	0.62
10	-0.5	-0.5
11	0	-0.58
12	0.5	1
13	-0.6	0.2
14	0.3	-0.2
15	0	-0.2
16	0.6	0.6
17	0.4	1
18	0.2	1
19	0.9	0.2
20	0	0.2
21	0	0.6
22	0.4	1

Table 3.5: Results of Sentiment Analysis with Google Cloud Natural Language API.

Let's see a visual representation of comparison with these results:

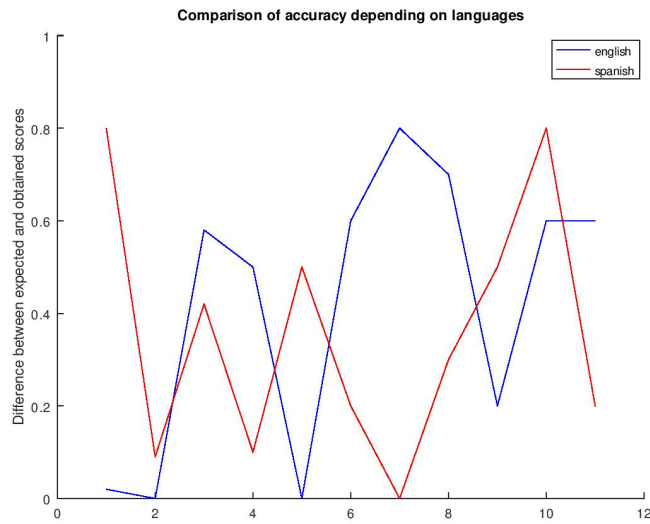


Figure 3.3: Comparison the differences between the obtained and the expected result in texts in Spanish and in English with Google Cloud Natural Language

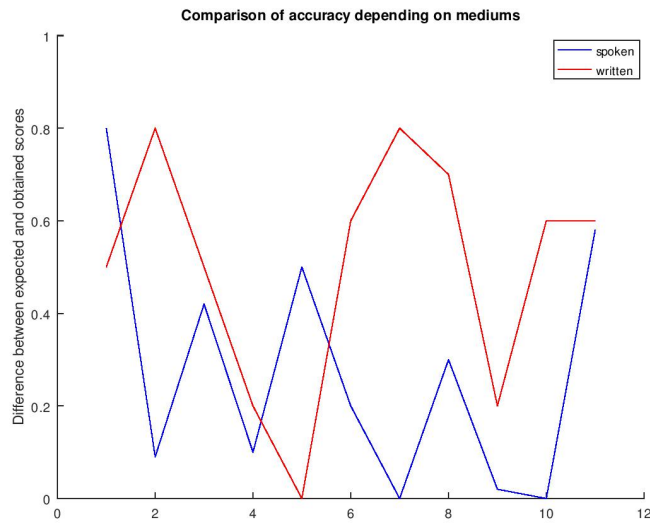


Figure 3.4: Comparison the differences between the obtained and the expected result in spoken and written texts with Google Cloud Natural Language

It is not possible to see a pronounce tendency in any of the comparisons. This means that the API can work with the same accuracy with independence of languages or formality of the speech. Just 4 of the results have a difference smaller than 0.1 points, and in 10 of them, this difference bigger than 0.5 points. On the other hand, knowing that the range

of values of the scores is -1, 1, this means that the maximal difference that could exist between the expected and the obtained score is 2 points, and the maximum that have been found in this test has been 0.8 points, so on the whole, the obtained results are good.

3.3 Conclusions

Summarizing, we have found that Language Analytics API is a very complete tool, but it have a slightly dependency on languages, that Google Cloud Natural Language does not have. Neither of the APIs present a dependency on the formality of the text, what is a good point in favor of both of them.

The accuracy is stronger in Google Cloud Natural Language, but in the other hand, it has been observed that the console of Google Cloud Platform needs more resources and internet power than the console of Microsoft Azure, what makes its use more restrictive. Besides, the criteria of sentimental punctuation in Google Cloud Natural is peculiar with some words, such as: "nails" that it considers as a negative word, or "awesome" that is considered neutral. Furthermore, this API analyzes just the sentimental content in names, as it would be more logical to analyze the content on adjectives, which usually provide the most part of the feeling to a description.

Finally, the last against point could be the bad integration between the transcription tools and the sentimental analysis APIs, since the ideal case would be them to be integrated in just one tool. Apart of that, we find that language must be specified in both transcription APIs, since sentiment analysis APIs use to be able to recognize languages, what seems to be a waste of functionalities. For solving this, sentimental analysis APIs should be integrated with transcription tools that are able to automatically recognize languages.

In conclusion, Language Analytics does not have the adequate accuracy for the use in telephone triage, but it could be good for management analysis, analysis of satisfaction with treatment implementation or medicines, etc. In the other hand, Google Cloud Natural Language seems to have the characteristics for being used as support in telephone triage, with a good internet connection and the adequate knowledge for interpreting the results.

Chapter 4

Abstract development of a prototype

4.1 Objectives

In this section we are going to propose a technical solution for a software for support in telephone triage for telenursing. With the intention of having an intuitive friendly environment for this technology, it has been decided to develop an app. Nowadays, people are used to deal with mobile and computer applications in their daily life. Due to this fact, we have considered that it will take less effort to learn how to use it and to work with an app than with an online tool or a console.

4.2 Working of the software

The application must have access to the phone calls or/and to the microphone for recording the voice of the patient. While obtaining the record, the transcription must appear in real time. Every transcription will be storage as a new element in an array, and when the call finishes, the array must be cleaned. This steps can be done developing an app with Android Studio. Android Studio give us the possibility of develop a transcription tool that automatically recognize the language. A proposal code can be seen in the Appendix.

Since Google Cloud Natural Language has demonstrated to have the proper accuracy for this sentiment analysis, we can combine it with the transcription part. For that aim, we

must remove the separations between elements and conform a unique array with just one element. This will be the text used for the sentiment analysis.

Another option is to develop an specialized model and use it for the sentiment analysis. Every element of the array should be separated in words, and a new array will be conformed with all the words of the transcriptions. Now, it is time to clean the unstructured array from unnecessary words like prepositions and punctuation. This cleaning can be done with Python with the package nltk. A possible option of cleaning code in Python can be seen in Appendix 2.

Now that we have an structured array of words we can use it with the developed model. The outgoing results will be given in percentage with the 0% being the most negative and the 100% the most positive result.

4.3 Design of the interface of the App

For a clear and friendly view of the results of the analysis a simple interface that allows an easy interpretation of the output data must developed. For the better interpretation of this part of the studium, the author has build a really simple app called SpeakToMe.

Basically, we are going to need a pair of things: a text box, two buttons and a scale. The text box will allow to read the transcription of the call in real time. Two buttons are necessary to control the start and the end of the recording. Finally, the scale will show the results in an intuitive way.

The buttons must behave in the following way. The button that marks the beginning of the recording is called from now on Control, and the button that clean the data of the ended call is Delete. Control will be responsible of beginning the record of voice for the transcription. For more comfort, this button must be clicked just one time for start the recording, so the clinician would have the hands free for taking notes or any other activity. Clicking this button a second time will just stop the microphone of recording, so if the clinician click it again the previous transcription will not be substituted, and the microphone will start recording more data and transcribing it. This new data will be added to the one recorded before.

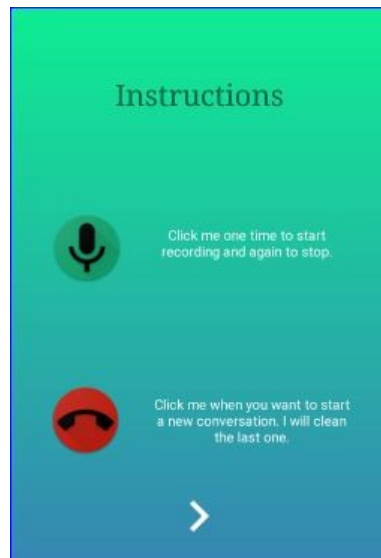


Figure 4.1: Instructions for the use of the buttons in the app SpeakToMe

In the text box is going to be shown the transcription of the current or immediate previous record. Every time the Control button is clicked for another recording, the text in the text box is cleaned, although every transcription is storage in the array of each call.

- Threaten level. It defines a situation where the life of the patient is in danger. This patient in a crisis, that needs from immediate assistance and a delay would have fatal consequences.
- Emergency level. This level refers to a situation where the state of the patient is really serious but his life is not in a risk. The patient is not critical, but they need of complementary urgent explorations for their diagnosis and proper immediate treatment.
- Moderate level. This are patients with stable vital signals and that do not have acute pain.
- Slight level. This are the banal patients, that do not present urgent injuries and do not require of complementary explorations.[hvn, 2011]

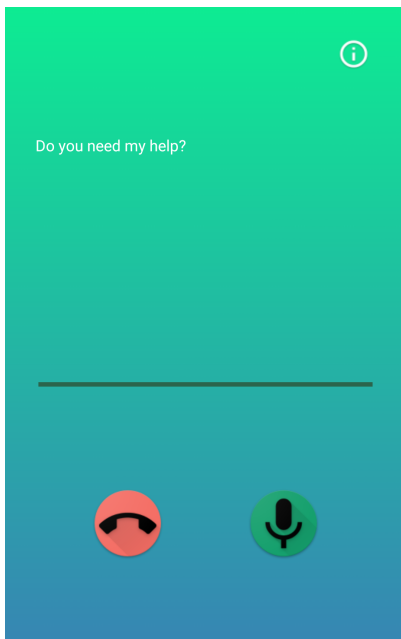


Figure 4.2: Initial main screen of the app SpeakToMe.

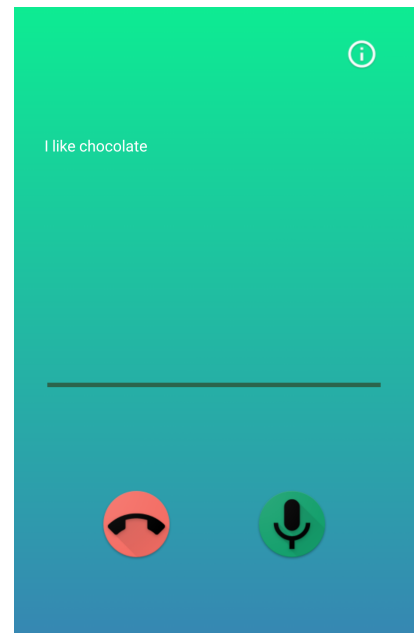


Figure 4.3: Main screen of the app SpeakToMe recording.

Depending on the method used to do the sentiment analysis, the results can be given in a range of -1 to 1 or in percentage. Since the progress bar works with percentages, we must adapt the results to the scale. The scale of the interface will be divided in 4 regions. If the result 0% is the most negative result possible and 100% the most positive in our scale, these will be the ranges for divisions:

- 0%-25% : Threaten level.
- 25%-50% : Emergency level.
- 50%-75% : Moderate level.
- 75%-100% : Slight level.

Working with a specific built model, the results do not need to be adapted, because the output of the model will be a percentage. In the other hand, if we connect the app to Google Cloud Natural Language, the output will be in the range of -1 to 1, so the -1 would correspond with the 0% of the scale, the 0 would be 50%, 1 would be 100% and so on.

For beginning the sentiment analysis, we must click in the progress bar. This will activate the order for the analysis, and the results will be shown by the progress of the bar.

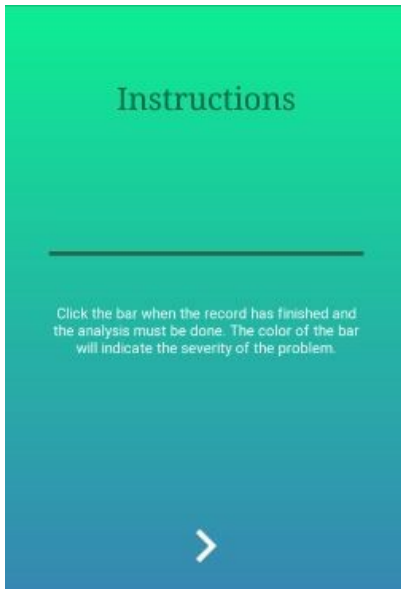


Figure 4.4: Instructions for the use of the progressbar in the app SpeakToMe



Figure 4.5: Ascending scale of damage severity in app SpeakToMe

Chapter 5

Further steps

5.1 Speech Sentiment Analysis using voice patterns

An ideal implementation of Speech Sentiment Analysis is a software that takes into account either the content of the speech and the intonation of the voice. Not taking advantage of the voice of the individuals is a waste of information, of sentimental meaning in current telephone triage software. Obviously, the expressions people use for transmitting opinions are important to translate our feelings, but sometimes "Actions speak louder than words", as we usually say. It would be useful too for distinguishing between patient and clinician in the conversation, so, in this way the clinician should not have to take care of clicking the button just when the patient speaks, but the software would do it automatically.

The voice intonation usually transmits better how a person feels. It is usual to say "be careful with what you say", but the modulation of the voice is more natural and difficult to control, and uses to be the reason why people unconsciously know about the mood of the speaker. When someone is tired or sad, he uses a low monotonous intonation. When someone is nervous or scared, they use a high-pitched voice, and try speak faster. When the person is nervous or excited, voice starts to tremble, and in the other hand, when we are pensive we speak slower, with long pauses between each word. [Dagupta, 2017] So, as we can see, voice is a great indicator of our mood, and that is why we should even give more weight to voice analysis in Speech Sentiment Analysis.[S.Maghilnan and Kumar, 2017][A.Davletcharova et al., 2015]

The use of voice patterns in Speech Sentiment Analysis in combination with the content of the speech could result in a great powerful tool in healthcare. It would make telephone triage more accurate, besides it could be used in the daily life of patients in treatment for adapting their routines to their mood and make it more bearable.

5.2 Voice properties

The human voice is a wave, and as every wave, can be characterized by 4 principal attributes, that have an special name for the wave "voice": pitch or frequency, timbre, loudness and duration.

The pitch is the audible range of frequencies of the human voice. It includes frequencies between 100 Hz and 200 Hz, where highest frequencies correspond to high-pitched voices and fast oscillations, and low frequencies to low-pitched voices and slower oscillations. Low frequencies are usual in men, besides high frequencies are common in women and children. The voice range is divided in 7 subranges, from the highest-pitched to the lower, we have: soprano, mezzo-soprano, contralto, countertenor, tenor, baritone and bass.

Timbre allows to distinguish different voices, even with the same pitch and loudness. Timbre makes people voices and instruments sound different, what characterize a sound.

Loudness or energy, is the property that makes a sound loud or quiet. In the graphical representation of a sound with the same pitch and timbre, loudness can be observed in the amplitude of the wave. This means that if the area inside an imaginary enveloping line of the wave is calculated, the bigger areas will correspond to the loudest sounds and the smaller to the quietest.

Finally, the duration of the sound is the time that a particular vibration (a wave with the same pitch, timbre and energy) last. This allows us to distinguish changes in syllables and silences. [Dagupta, 2017]

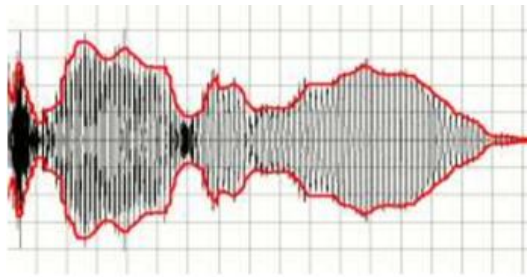


Figure 5.1: Sound wave's envelope marked in red [Dagupta, 2017]

5.3 Working of voice analysis algorithms

Algorithms of voice analysis, usually take the pitch, the timber, and the energy of the voice. Pitch allows us to divide the record on units depending on the frequency: if the frequency is low, the units of division will be larger, and on the contrary, when frequency is high, this units must be smaller for being able to proper analyze the information. The timber will be the key for distinguishing the voice of a person from someone's else, and to determine who is the patient and the clinician. At last, the energy of the voice will be the property that allows to build the voice patterns.

So, the process for creating a pattern out of a record would be as follows:

- First, the speakers in the conversation must be discriminated by the timber of the voice. It would be inefficient analyze both the voice of the patient and the clinician, so we are just using the voice of the patient for the analysis. For identifying who is the clinician and who is the patient, the clinician should make a voice sample every day before start using the software, so it exist a reference record.
- Then, we must analyze the frequency of the recorded waves. The range of frequencies of pitch in from 100 Hz to 200 Hz. While analyzing the record, lower and higher frequencies out of this range must be cleaned. This is not possible to do with the raw signal, but with its histogram of frequencies, that allows to know how much of the signal belong to each frequency. Once we have a clean signal, we can divide the new raw data with the rule of sampling frequency or with an adaptive threshold. If we remember the rule of the sampling frequency:

$$B \geq 2 \times \Lambda$$

- B: minimal sampling frequency.
- Lambda: frequency of the original wave.

This formula is the key for digitally recording the whole wave without losing information, and taking into account, the highest frequency of human voice is 200 Hz, the sampling frequency must be 400 Hz. On the other hand, adaptive threshold allows to make divisions depending on the frequency of the speech in that piece of signal.

- Finally, when the signal is divided in pieces, the energy of the speech must be analyzed. This energy in the temporal domain, can be calculated as the the squared area above the curve that envelope the signal. The formula for this calculation is:

$$Energy = \sum Signal^2 \times \Delta t$$

- Signal: raw signal in temporal line.
- t: time.

With these values for energy, one or more thresholds can be set to distinguish low intonation from high intonation and the pattern is build.

This models would remark the strongest of acutest part of the speech record, and make patterns out of this information. For making it more visual, this patterns would be similar to Morse code: the ”_” would be, for example, every fragment of the speech in an acute intonation, and ”.” would be the lowest tones. Every expressed emotion would be related with a pattern or group of patterns, for example, anger begins with ”.” and finishes with ”_”, happiness would have a different pattern and so on. The advantage of building this models is that specialized data is not needed in this case, due to we are working just with the voice, and we do not need from the model to learn specialized words. This means that any dataset with voice records could be used, as long as it contains speech of people expressing opinions.

5.4 Conclusion

Summarizing, working more deep in Speech Sentiment Analysis with intonation patterns worth the effort and could lead to more powerful and thorough tools in healthcare and specially in telephone triage. We have to take into account that the intonation differs into different countries and accents, so the same pattern is not useful between counties and sometimes even between regions.

Bibliography

- A.Davletcharova, S.Sugathan, B.Abraham, and A.Pappachen (2015). Detection and analysis of emotion from speech signals. *Procedia Computer Science*.
- Dagupta, P. (2017). Detection and analysis of human emotions through voice and speech pattern processing. *International Journal of Computer Trends and Technology (IJCTT)*, pages 1–2.
- H.Asmat, S.Shinwari, T.Cooksley, R.Duckitt, Jeune, I., and C.Subbe (2017). Measuring impact of telephone triage in acute medicine. page 1.
- J.Li, L.Li, R.Lake, and A.Georgiou (2017). The quality, safety and governance of telephone triage and advice services- an overview of evidence from systematic reviews. *BMC health serv res*.
- K.Taimoor (2016). Sentiment analysis for health care. *ResearchGate*.
- S.Maghilnan and Kumar, M. (2017). Sentiment analysis on speaker specific speech data. In *International Conference on Intelligent Computing and Control*.

Internet sources

A.Moreno-Ortiz (2016). Lingmotif. URL: http://tecnolengua.uma.es/?page_id=8.

Cloud, I. (2017). *IBM Cloud Docs. Alchemy Language*.

hvn (2011). Anexo. Technical report, hvn.

J.C.Gonzalez and E.Valencia (2017). Voice of the patient analysis. MeaningCloud, URL: <https://www.meaningcloud.com/blog/voice-of-the-patient-analysis>.

MeaningCloud (2018). *Sentiment Analysis API documentation*.

T.Mullen (2011). Introduction to sentiment analysis. Ict-master, URL: <https://lct-master.org/files/MullenSentimentCourseSlides.pdf>.

T.Russell-Rose (2014). Sentiment analysis in healthcare. Data and Analytics, URL: <https://www.slideshare.net/tgr2uk/sentiment-analysis-in-healthcare-37463007>.

T.T.Pham (2017). 50+ useful machine learning and prediction apis. KDnuggets, URL: <https://www.kdnuggets.com/2017/02/machine-learning-data-science-apis-updated.html>.

Appendix A

Complementary data

A.1 Raw output of Text Analytics API

In the following section an example of the output code of the results from Text Analytics API is shown, so we can observe the format of the output of this API and the response to our sample data. The API generates a JSON file whose lines are shown here.

Results for commentary 1:

```
{  
  
  "languageDetection": {  
  
    "documents": [  
  
      {  
  
        "id": "00242ff7-3161-4f92-bcef-fc5ea5a5bf53",  
  
        "detectedLanguages": [  
  
          {
```

```
        "name": "Spanish",
        "iso6391Name": "es",
        "score": 1.0
    }
]
}
],
"errors": [],
},
"keyPhrases": {
    "documents": [
        {
            "id": "00242ff7-3161-4f92-bcef-fc5ea5a5bf53",
            "keyPhrases": [
                "sencillo slime",
                "Internet win",
                "perdon",
                "poder"
            ]
        }
    ]
}
```

```
],  
  
  "errors": []  
  
},  
  
"sentiment": {  
  
  "documents": [  
  
    {  
  
      "id": "00242ff7-3161-4f92-bcef-fc5ea5a5bf53",  
  
      "score": 0.46752721071243286  
  
    }  
  
  ],  
  
  "errors": []  
  
}  
  
}
```

As observed above, the results can be divided in three regions: the first lines is the language detection, the second region is for key words detection, and the third region is the one for sentiment analysis. In the region of language detection and of sentiment analysis, apart of the raw results, the algorithms gives an estimation of the accuracy of the obtained results in a scale from 0 to 1.

A.2 Raw output of Google Cloud Natural Language API

In the following section screenshot example of the output of the API is shown so we can observe the format of the output of Google Cloud Natural Language API and the response to our sample data. It was not possible to catch directly the output code, so only images will be shown.

```
{
  "documentSentiment": {
    "magnitude": 1.5,
    "score": 0.1
  },
  "language": "es",
  "sentences": [
    {
      "text": {
        "content": "Oigan, perdón, pero este es el mejor y más sencillo slime.",
        "beginOffset": 0
      },
      "sentiment": {
        "magnitude": 0.8,
        "score": 0.8
      }
    },
    {
      "text": {
        "content": "Sigo sin poder creerlo.",
        "beginOffset": 61
      },
      "sentiment": {
        "magnitude": 0.5,
        "score": -0.5
      }
    }
  ]
}
```

Figure A.1: Screenshot of output from commentary 1

In this case, the result is divided in section and subsections: the first section is for language detected, and the second section is divided into subsection. Each subsection is dedicated to each sentence on the text, and includes the positivity of the feeling and its intensity in a scale of -1 to 1. As seen before, the language detection have a value for the accuracy of the analysis.

Appendix B

Codes

B.1 Cleaning text

The following code is a proposal for conforming a structured dataset with Python and its nltk.

```
#lets try the code with the comment number 10.
#before start working in the console of python, we have installed
#the next packages from
#the computer terminal:
#pip install nltk

#nltk: natural language toolkit

import nltk
nltk.download()
#install all the packages just in case

#we have an string object called text with the comment 10 in it:
#text="So, that was actually sort of scary. I feel like it's sort
#of dangerous messed with your nails like that, but it worked
#really well. I'm super happy with the way they look now and I
#think that was a pretty good buy for like 15 bucks. Totally
```

```
#awesome."

##Tokenize the text

from nltk.tokenize import word_tokenize
tokens=word_tokenize(text)
#if we want to see the tokens: print(tokens)
#lets delete punctuation signs and monosyllables
clean_tokens=[i for i in tokens if len(i)>1]

##Removing stop words
from nltk.corpus import stopwords
sr=stopwords.words('english')
#this last line depends on the language of the text
#we need to have everyword in lower-case letters because
#stopwords just works with lower-case letters
for i in range(len(clean_tokens)):
    word=clean_tokens[i]
    clean_tokens[i]=word.lower()

#and now lets clean clean_tokens
for i in range(len(sr)):
    if sr[i] in clean_tokens:
        while sr[i] in clean_tokens:
            clean_tokens.remove(sr[i])

#calculating the frequency of each word in the text(how many
#times each one appears)
freq= nltk.FreqDist(clean_tokens)
#if we want to see the frequencies of each word:
#for key, val in freq.items():
#    print(str(key)+ ':' + str(val))

##Lemmatizing
#lets conform a list with the lemmas of the words in clean_tokens
```



```
#to normalize them.
from nltk.stem import WordNetLemmatizer
lemmas=list()
lemmatizer= WordNetLemmatizer()
for i in range(len(clean_tokens)):
    lemmas.append(lemmatizer.lemmatize(clean_tokens[i]))
```

B.2 Internet behavior

Interface developed with Android Studio. In this section it is included just the codes of the logical part(.java), without the visual part.

B.2.1 MainActivity

```
package com.example.user.speaktome;

//sorry for the amount of commentaries, but I need them

import android.Manifest;
import android.content.Intent;
import android.content.pm.PackageManager;
import android.media.AudioFormat;
import android.media.AudioRecord;
import android.media.MediaRecorder;
import android.net.Uri;
import android.os.Build;
import android.provider.Settings;
import android.speech.RecognitionListener;
import android.speech.RecognizerIntent;
import android.speech.SpeechRecognizer;
import android.support.v4.content.ContextCompat;
import android.support.v7.app.AppCompatActivity;
```

```
import android.os.Bundle;
import android.view.MotionEvent;
import android.view.View;
import android.widget.Button;
import android.widget.ProgressBar;
import android.widget.TextView;

import java.util.ArrayList;
import java.util.Locale;

import static android.os.Build.*;

public class MainActivity extends AppCompatActivity {

    //checking permission on runtime. This code ask you to
    //habilitate the access to audio since the first time you are
    //using the app
    private void checkPermission() {
        if (VERSION.SDK_INT >= VERSION_CODES.M) {
            if (!(ContextCompat.checkSelfPermission(this,
Manifest.permission.RECORD_AUDIO) ==
PackageManager.PERMISSION_GRANTED)) {
                Intent intent = new Intent(Settings.ACTION_
APPLICATION_DETAILS_SETTINGS,
                    Uri.parse("package:" + getPackageName()));
                startActivity(intent);
                finish();
            }
        }
    }

    @Override
    protected void onCreate(Bundle savedInstanceState) {
        super.onCreate(savedInstanceState);
        setContentView(R.layout.activity_main);
    }
}
```

```
//this line just allows us to see the screen application
//if we have given access to the microphone
//we need this permission to use SpeechRecognizer class
checkPermission();

//lets create a microphone for checking if it's being
//used or not
final AudioRecord micro = new AudioRecord(
MediaRecorder.AudioSource.MIC, 44100,
AudioFormat.CHANNEL_IN_MONO,
AudioFormat.ENCODING_DEFAULT, 44100);
//defining a TextView object that will be our text in
screen
final TextView textView = findViewById(R.id.textView);

//Creating an Speech Recognizer object
final SpeechRecognizer mSpeechRecognizer =
SpeechRecognizer.createSpeechRecognizer(this);
//and an Intent for listening to the speech
final Intent mSpeechRecognizerIntent =
new Intent(RecognizerIntent.ACTION_RECOGNIZE_SPEECH);
mSpeechRecognizerIntent.putExtra(
RecognizerIntent.EXTRA_LANGUAGE_MODEL,
RecognizerIntent.LANGUAGE_MODEL_FREE_FORM);
//EXTRA_LANGUAGE_MODEL informs the recognizer which
//speech model is preferred, and LANGUAGE_MODEL_FORM
//uses a language model based on free-form speech
//recognition
mSpeechRecognizerIntent.putExtra(
RecognizerIntent.EXTRA_LANGUAGE,
Locale.getDefault());
//EXTRA_LANGUAGE takes an optional IETF language tag.
//Since we have asked for the default one,
//it will be the one of the phone. For example, for me it
//will be Spanish.
```

```
//defining an Array for collecting all the transcriptions
final ArrayList<String> transcriptions =
//new ArrayList<String>();

//Defining an object ProgressBar
final ProgressBar results = findViewById(R.id.progressBar);

//setRecognitionListener should be called before any
//command to the created SpeechRecognizer
mSpeechRecognizer.setRecognitionListener(new
RecognitionListener() {
    @Override
    public void onReadyForSpeech(Bundle bundle) {
    }

    @Override
    public void onBeginningOfSpeech() {
    }

    @Override
    public void onRmsChanged(float v) {
    }

    @Override
    public void onBufferReceived(byte[] bytes) {
    }

    @Override
    public void onEndOfSpeech() {
    }

    @Override
    public void onError(int i) {
    }
}
```

```
@Override
public void onResults(Bundle bundle) {
    //getting all the matches
    //the result is an array of Strings with possible
//recognition results, where the first element is
//the most likely candidate
    //so matches is the array which will content the
//text(since 0 element is the one supposed to be
//correct,
    // we will use that one)
    ArrayList<String> matches = bundle
        .getStringArrayList(SpeechRecognizer.
RESULTS_RECOGNITION);

    //displaying the first match
    if (matches != null) {
        //matches [0] is the correct speech recognition
        textView.setText(matches.get(0));
        //adding each correct transcription to the
//transcriptions array
        transcriptions.add(matches.get(0));
        //showing them on run screen just to make
//sure it's saving properly
        int x;
        for (x = 0; x < transcriptions.size(); x++) {
            System.out.println(transcriptions.get(x));
        }
    }
}

@Override
public void onPartialResults(Bundle bundle) {

}

@Override
```

```
        public void onEvent(int i, Bundle bundle) {

            }

    });

    //after defining TextView, lets set Listening... if the
    //button is pressed or You will see input here if the
    //button is not pressed
    findViewById(R.id.button).setOnTouchListener(new
    View.OnTouchListener() {
        @Override
        public boolean onTouch(View view, MotionEvent
    motionEvent) {

            //we will start or stop listening when we touch
            //the button, and write the message when remove
            //the finger

            switch (motionEvent.getAction()) {
                case MotionEvent.ACTION_UP:
                    if (micro.getRecordingState() ==
    AudioRecord.RECORDSTATE_RECORDING) {
                        textView.setHint("Listening...");
                        textView.setText("");
                    }
                    //when the user removed the finger
                    //mSpeechRecognizer.stopListening();
    textView.setHint("Do you need my help?");
                    break;

                case MotionEvent.ACTION_DOWN:
                    if (micro.getRecordingState() ==
    AudioRecord.RECORDSTATE_STOPPED) {
                        mSpeechRecognizer.startListening(
    mSpeechRecognizerIntent);
                    } else {
                        mSpeechRecognizer.stopListening();
                    }
                }
            }
        }
    });
```

```
        }
        //finger is on the button
        //mSpeechRecognizer.startListening(
//mSpeechRecognizerIntent);
textView.setText(""); textView.setHint("Listening...");
break;
    }
    return false;
}
});

findViewById(R.id.delete).setOnTouchListener(new
View.OnTouchListener() {
    @Override
    public boolean onTouch(View v, MotionEvent event) {
        if (MotionEvent.ACTION_DOWN == 0) {
            transcriptions.clear();
            textView.setText("");
            textView.setHint("Do you need my help?");
        }
        return false;
    }
});

//with the results of the analysis lets move the
//progressBar
results.setOnTouchListener(new View.OnTouchListener() {
public boolean onTouch(View u, MotionEvent uevent) {
if (MotionEvent.ACTION_DOWN == 1) {
results.setProgress(50);

        }

        return false;
    }
}
});
}
```

```
    //Finally we are going to adjust the behavior of the info
//button. It should show the first layout of instructions
    public void info(View view){
        Intent i =new Intent(this, InstructionClass.class);
        startActivity(i);
    }
}
```

B.2.2 InstructionClass

```
package com.example.user.speaktome;

import android.app.Activity;
import android.content.Intent;
import android.os.Bundle;
import android.view.View;

public class InstructionClass extends Activity {
    protected void onCreate(Bundle savedInstanceState) {
        super.onCreate(savedInstanceState);
        setContentView(R.layout.instructions);
    }
    public void info1(View view){
        Intent a =new Intent(this, InstructionClass2.class);
        startActivity(a);
    }
}
```

The code in InstructionClass.java is the same for InstructionClass2.java and InstructionClass3.java just changing the layout.instructions for layout.instructions2 and layout.instructions3, and Intent a for e and o respectively.

Obligatory Signed Declaration

concerning this thesis titled

SPEECH SENTIMENT ANALYSIS FOR HEALTHCARE

“I hereby declare that the present thesis was composed by myself and that the work contained herein is my own. I also confirm that I have only used the specified resources. All formulations and concepts taken verbatim or in substance from printed or unprinted material or from the Internet have been cited according to the rules of good scientific practice and indicated by footnotes or other exact references to the original source.

The present thesis has not been submitted to another university for the award of an academic degree in this form. This thesis has been submitted in printed and electronic form. I hereby confirm that the content of the digital version is the same as in the printed version.

I understand that the provision of incorrect information may have legal consequences.”

Graz, July 4, 2018
