XXI IMEKO World Congress "Measurement in Research and Industry"
August 30 – September 4, 2015, Prague, Czech Republic

UNCERTAINTIES CAUSED BY HUMAN FACTORS – CAN AN EMOTION DETECTOR HELP?

Andreas Kamcke [1]

[1] Gigaset Communications GmbH, Bocholt, Germany, andreas.kamcke@gigaset.com

Abstract – Uncertainties in results caused by human factors of test persons are well known in the area of testing the speech quality of telephone connections. The presence of this kind of uncertainty components were already known when this test methodology was introduced. For elimination or reduction a new approach by using emotion detection is presented. First results and limits are given. In order to get a universal method standardization is urgently needed.

Keywords: Emotion Detection, Human Factors, Telephone, Subjective Tests, Uncertainties

1. INTRODUCTION

Some measurement and tests are based on opinion scores of test persons, which can be influenced by sentiments. This kind of measurement is found when quantities containing human factors are used. Here the result is not related to a quantity which is independent from other quantities and is not influenced from physical earth properties. Instead of this wide independent characteristics of human beings have to be taken in to account. When testing these special quantities by using instruments the equipment has to be taken into consideration the human characteristics as well. Usually only an exactly defined part of the human characteristics are wanted to be considered. Others, e.g. sentiments driven parts are seen as disturbing influence leading to uncertainties in the result. The general question comes up on how to exclude the unwanted parts and if a device can detect emotions better than a human being.

As an example testing of end-to-end speech quality of telephone connections is presented in detail below. Here the human characteristics are psycho-acoustical ones. The unwanted ones are sentiments leading the test persons to deviate its opinion score. This principle problem was already known when this kind of test methodology was introduced many years ago. Some approaches were done in order to eliminate or at least to reduce the unwanted effect, hence the uncertainty.

A new method has been introduced. It is based on the use of so called "Emotion Detectors". First results are available. Methods needed in the fields of electro-technique, informatics, signal processing, acoustics and telephonometry are well known, however, the main problem of this approach is found in open points on the psychological side. Furthermore, this new method can be followed by using different approaches in details which may have impact on the results. In order to reach comparable measurement results standardization is needed. That would guarantee the use of uncertainty values without the need of additional computation. It would also allow that results are directly comparable. No care needs to be taken on different restrictions resulting from different methods.

2. SUBJECTIVE TESTING OF SPEECH QUALITY

In the contribution the measurement of speech quality is taken as example for uncertainties in human factors. Because measurements in telecommunications including telephonometry are a very special branch of measurements some more details are given here.

The term "speech quality of telephone connections" is very often found also in daily life when differences of telephone sets or telephone networks are discussed. Usually every contributor follows its very own view on what quality is. Following a scientific definition on and advice for measuring speech quality the following main determinations are found:

- Speech quality is understood as the quality which is perceived by human beings.
- Speech quality is not based on the opinion of a single person but is define as a mean opinion score.
- The test persons are naïve in respect of telecommunications and acoustics.

The term "subjective test" is based on the fact that these tests are made by asking subjects, i.e. human beings, for a score. In contrast to that, tests done by using objects, i.e. measurement instruments, are called "objective tests". A detailed description on the process of subjective testing is found in [1].

It has to be noted that speech quality was the first defined quality in telecommunications. That is because the speech service was the first telecommunications service and the lonely one for many decades.

Because the telephone was invented more than 140 years ago people tend to believe speech quality and its measurements are well defined for many years. In reality speech quality of telephone connections need a permanent observation. Every change in the used technology causes a change in the quantities contributing the main influence in quality. Some examples may illustrate this. In the late 1980ies in Europe the existing analogue long distance lines were replaced by digital lines. Subscribes were surprised that a long distance connection sounds as a local connection.

The typical long distance indications namely loudness loss and additional noise disappeared. Another example: Nowadays existing networks are replaced by IP-based ones. The influence on speech quality is observed every day. Heavy echo effects are often found now. Before this change in telecommunication technology echo effects were rare and sufficiently suppressed by echo cancellers. A last example in this context: The introduction of a new Codec in public mobile networks needs detailed speech quality tests in order to find out which Codec candidate promises the best end-to-end quality. ("Codec" is the technical term in telecommunication for the combination of a coder and a decoder also known as A/D converter and D/A converter). The three examples can be summarized as: Speech quality depends on the technology used in the connection. It has to be reconsidered whenever the technology changes at least in parts of the technical components.

3. LISTENING ONLY TESTS

This clause provides a short description on how listening only tests are carried out. A detailed description is found in [1].

The tests takes place in a well prepared room were acoustical influences are suppressed as ever possible. That is important especially for the noise floor.

Subjects taking part in these listening tests are chosen at random from the normal telephone users. There have to be naïve with respect of telecommunication techniques and they have not participated in any subjective test during the last half year.

In order to avoid confusions by unsufficient understanding of the test sentence the experience show that it is absolutely necessary subjects have to be native speakers of the language used in the test.

Once the text of the test sentences are stable defined a read out version will be recorded. During the test campaign these sentences are played via headphones. The subjects vote on the quality by giving their results as a number in the range between score 1 to score 5. The meaning of these numbers is given in Table 1.

Table 1. Score on the Quality of Speech.

Score	Quality of Speech
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

A single score does not give any useful information. Only the mean is a valuable result. Therefore, the single scores are not reported. The result is always a Mean Opinion Score (*MOS*) on the quality of speech in a telephone connection which represents the quality assessment by human beings.

Since the very first step towards quality assessment of voice transmissions by using human subjects during subjective listening tests, it was clear that used test sentences have to have emotional neutrality in order not to influence the result. Emotions resulting by wording or the content of the test sentences can drive the subjects to pass a different vote. The read out and recording process add other influences like emphasis variation of loudness which are also to be considered as a source of emotions. It will influence the single votes hence the *MOS* will get a specific uncertainty component.

In order to underline the importance of elimination or at least reduction of this uncertainty in the *MOS*-results it may be reminded that objective tests, i.e. measurements by using instruments, are designed by estimating the *MOS*-result of a subjective test. Hence objective tests intend to give the same results as equivalent subjective tests. If a certain objective test is trained on *MOS*-values together with a relevant uncertainty resulting from subjective tests, the result of the objective test got also these specific uncertainties!

In order to get emotion-free test signals a lot of different approaches were followed in the past. Among other approaches test sentences were replaced by artificial voice signals prepared by concatenated syllables. It was found this was more than naïve subjects could handle. A second idea was to use real speech but the sentences contained no meaning. Naïve subjects tend to interpret the meaningless test sentences instead of concentrate on the quality. At the moment the written format of the test sentences were examined by a specialist. He deselects all test sentences having not enough emotional neutrality. Only test sentences which passed this procedure will be taken in a subjective test campaign.

Even if today's method is the best of all what was used in the past, it introduces a break in the methodology. The result is strictly received by computing a mean, however, the voting on the emotional neutrality is done by the opinion of a single person. On the other hand when a human judgement on emotions is good enough for this task it is still an open question how good an instrument can do this work.

In order to improve the current situation in ETSI TC STQ the idea came up to use an instrument for this selection or for elaboration of an uncertainty factor representing the influence in the result coming from emotional effects. It was the assumption that this instrument could be an Emotion Detector based on the observation there are many publications about this new device. The result of a subjective test can then be free from the influence of a single person.

4. EMOTION DETECTORS

On the web one will get easily 1.590.000 references for "Emotion Detector". It is important to know that this term is only a headline for different devices detecting elements which lead to human sentimental reactions while reading text, listening to speech and or music, looking on pictures movies and movements. Furthermore speaker's emotions can be detected as well. The following gives only a very raw impression on what is going on in the field of emotion detection.

The following extremely short description of an Emotion Detector may be sufficient to understand the content of this contribution. An Emotion Detector consists of a computer examining inputs and computing outputs representing results on how emotive the input is, sometimes added with advices

on how to deal with the input. It serves totally different applications:

- Research, namely in psychological research
- Business. E.g. business generation by examining the transmitted content of telecommunication connections.
- Emergency telephone calls. E.g. online classification of emergency calls.
- Military and political applications.
- Others.

These instruments promise to become a solution for the missing automatic emotion check during subjective tests, including emotion checks for future voice services with extended bandwidth.

However, the question comes up if such Emotion Detector can judge emotions better than a human expert.

As shown in clause 3 ready to use test sentences for subjective tests are produced in two steps (elaboration of the text and then recording) with their specific sources of sentimental injections. An Emotion Detector has to check written text and recorded spoken text. It is obvious that an instrument for checking written text differs from an instrument for checking recorded spoken text. The effort for the second instrument is much higher than the effort for the first one. Furthermore totally different technical approaches are needed for the two instruments.

It has also to be noted that an Emotion Detector for telecommunication measurements is not defined so far, neither for written text nor for recorded spoken text.

4.1. The output of an Emotion Detector

Without studying details of the output of an Emotion Detector one get easily more than one solution, even if the task of the Emotion Detector is restricted on the discussed versions for written text and for recorded spoken text.

In any case it is a must that the output can be restricted on a certain language. That is important because subjective tests are done by participation of native speakers. Therefore, only sentiments are relevant when they can be observed during the use of a test sentences in this language. Unimportant are sentiments that might be observed when the same test sentence is used in a different language.

It is also a strong requirement that the detection can be restricted to a certain time period like the years of office of a president. With that, test sentences like “The first name of the president is Balduin” would not be classified as emotional text. This sentence does not result in protests by the listeners during a listening test campaign when the text is used in the right time. Furthermore that requirement offers to reconsider old test campaigns by checking the test sentences on the emotional basis that was found in the time when the test campaign took place.

During discussions on the design of Emotion Detectors it became clear that the expectations on the output of such a Detector are very different. Some expect a list with all detected emotions and a value for each of them. Others expect a single value about the sentimental probability. Also expectations exists that lead to a mix from the two above. The variation in the expectation is based on the fact that in literature no clear view is found about a set of base statements concerning emotions. Some experts understand the emotions are devided in two classes: A small number of basic emotions and other emotions which result from combination of basic emotions. However, it is observed that experts who follow this approach use different sets of basic emotions. Other experts see emotions independent, but categories can be created. Different understandings about the building of groups exist. Furthermore, not all experts follow the approach that emotions can compensate each other. Some understand the mathematical sum of the strength of positive and of negative emotions lead to a result. Others have different understandings like multidimensional approaches. The existence of a zero point for emotions is also unclear. All of that lead to the statement the main problem of the approach by using an Emotion Detector is found in open points on the psychological side.

In this situation the designers of an Emotion Detector have to implement at least one output and an algorithm for the output value(s). That might be a negative detail for the overlaying question if an automat can do the job of detecting emotional components in text better than a human being.

4.2. Emotion Detector for written text

In [2] a trainable Emotion Detector for written text is described. An investigation is included about methods promising usable design of an Emotion Detector for spoken text. Based on the finally selected maximal entropy method a Detector was designed generating a unit free number, as output. This number is called sentiment sore *SSc*. It indicates the probability of sentiments. *SSc* = 1,0 means 100% probability of sentiments, where *SSc* = 0,0 means 0% probability of sentiments. With that, test sentences can be automatically excluded from the usage in a test campaign of subjective tests.

The designed Emotion Detector allows to be trained on different languages. It is also possible to train the machine on a certain time period, e.g. the year 1990. Then the received *SSc* number reflects the probability for this year.

No indication is given in [2] on how precise the resulting *SSc* is.

4.3. Emotion Detector for recorded spoken text

The second step of automatic deselection of test sentences is done by detection of recorded spoken text. In [3] an investigation is given about methods to be seen as useful approaches towards the design of an Emotion Detector for recorded spoken text.

The work is still going on and no result is available at time of preparation this abstract.

4. FIRST RESULTS

In [4] first results are reported. So, the sentence “The grass curled around the fence post.” got a *SSc* value of 0,29 where a *SSc* value of 0,81 was given for the sentence “Help the woman get back to her feet.” These sentences were already used in subjective tests. They were given by a laboratory to the owner of the Emotion Detector for written text in order to evaluate the device.

No more results are available at time of preparation this abstract. With that, no contribution on the answer to the question can be given, if an automat can detect emotion better than a human being.

4. CONCLUSIONS

The problem of emotional uncertainties is shown by the example of subjective testing. The existence of this kind of uncertainty was clear when the test methodology of subjective testing was introduced many years ago. Some approaches were done in order to eliminate or at least to reduce the unwanted effect, hence the uncertainty. The current method is done by using an expert how can deselect sentimental test sentences, hence he eliminates the uncertainty. A device would additionally offer to get a value about the uncertainty when it cannot be eliminated.

A new method has been introduced now. It is based on the use of a device, a so called "Emotion Detector". Very first results are available.

Different approaches in the design of an Emotion Detector are possible. In order to reach comparable measurement results an early involvement of standardization is needed. That would allow to follow a general approach and to use the findings also for handling of uncertainties of human factors in applications outside telecommunications.

However, it is too early or a final answer on if or if not a device can detect emotional components better than a human being.

ACKNOWLEDGMENTS

The author likes to thank Prof. Holub and his students from CTU Prague for all the work they have done in order to develop a prototype of an Emotion Detector for that purpose. Many thanks also to all interest delegates for the valuables discussions in ETSI TC STQ an in ITU-T SG12.

REFERENCES

[1] Recommendation ITU-T P.800 "Methods for subjective determination of transmission quality", 1996.

[2] J. Holcin, "Test sentences Design with Minimum Emotional Uncertainty", *unpublished Master Thesis at CTU FEE Prague*, 2014.

[3] L. Mikulka, J. Prasek, J. Pytela, A. Svoboda, "Emotional detector model for speech", *unpublished paper of CTU FEE Prague*, 2014.

[4] J. Holub, A. Kamcke, "Emotion Detector for written text" *ETSI meeting document STQ(14)46_025*, June 2014.