

ARE TRANSCRIPTIONS OF SPEECH MATERIAL RECORDED BY MEANS OF BUGS RELIABLE?

Loredana Cerrato, Andrea Paoloni
Fondazione Ugo Bordoni
Via B. Castiglione, 59 - 00142 Rome, Italy
pao@fub.it
<http://www.fub.it>

ABSTRACT

In this paper we will present the results of a perceptual test run to evaluate the effects of some *postperceptual analysis* of speech recorded by means of concealed microphones (i.e. bugs).

Our main aim was to verify the hypothesis that in Italian post-perceptually at least, the grammaticality of a perceived utterance overrides other factors, in particular semantics, when the speech signal is weak or noisy.

The results of our experiment might give some insight to the problems related to the task of transcription of speech recorded by means of bugs for forensic purposes.

1. Introduction

Transcriptions of conversations recorded by means of concealed microphones are being employed as evidences in crime prevention on an increasing basis. The concealed microphones or bugs are very small microphones (few millimeters) able to catch the conversations and to transmit them through the telephone line to a place where the conversation are recorded. Unfortunately the quality of such recordings is very poor and, as a consequence, conversations are not often intelligible [1]. On the contrary considering the application of the transcriptions it is very important that they should be reliable. The poor quality of the signal is due to the several type of noises that affect the recorded signals, which can be summarised as follows:

- background signals, that can be noise-like such as in car, street noise, or speech-like noise such as an environment with competing speakers [2];
- channel distortions due to the transmission channel (either the radio link or the telephone line);
- noises produced by the record devices that often use lower speed to get more storage capacity [3].

But the difficulty of transcriptions depends also on the subjective perception of the transcriber; in fact the human speech comprehension system in some circumstances makes a postperceptual processing of

signals perceived in high conditions of noise, and this leads to the possibility of the so called perceptual illusions, in other words in some cases the desire to “hear” some words or the expectations to hear some words are so strong that even when the signal is weak and of poor quality, the perceptual system processes the items which actually were not uttered.

Another problem in the transcription of speech can be the speaker’s dialect [4] but this difficulty is out of the task of this paper.

Therefore the main aim of our investigation is to verify the hypothesis, already proved on English speech material [5], that in Italian post-perceptually at least, the grammaticality of a perceived utterance overrides other factors, in particular semantics, when the speech signal is weak or noisy.

2. Speech material

To collect the speech material needed for the test we reproduced in our labs the situation of bugging and we recorded some sets of sentences uttered by 4 speakers (2 male and 2 female) in a noisy room. Our stimuli were structured as follow:

A first set of 3 groups “built up” sentences made of:

- 12 normal, grammatical sentences;
- 12 grammatical, but semantically anomalous sentences;
- 12 ungrammatical strings, but semantically plausible.

A second set of 39 sentences made of triplet of ambiguous sentences (similar in their syllabic structure, in the sequence of phonemes or made of similar words) extracted from real conversation (i.e. slip of the hears, slip of the tongues).

3. Experimental set up

The speech material was organised in two series of tests:

- a listening and transcribing and
- a listening and recognition test.

The listening and transcribing test consisted of a within-subject design in which 20 participants listened two groups of 12 sentences (listed in appendix A) produced by 4 speakers (no sentences were repeated) in 3 different S/N conditions:

- -3db S/N;
- -6db S/N;
- -9db S/N.

The speech material was then transformed into telephonic signal through a filter of 300-3400Hz.

The 20 subjects, all reported having normal hearing, (who were recruited among the police forces) were asked to write down, by means of a dedicated computer software, the transcription of each sentence. Thanks to our experimental device we gathered the transcription of each sentence made by each listener, together with the indication of the number of times each sentence was listened.

In the recognition test subjects listened to 13 sentences, one for each of the 13 groups (listed in appendix B), uttered by 4 speakers in a S/N ratio=-6 and had to indicate on an answer sheet the sentence they recognised among the three possibilities given in the choice.

4. Results

The intelligibility of sentence transcriptions for each condition has been measured at sentence and word levels.

With respect to the sentence level, the percentage of correctly transcribed sentences, reported in figure 1, shows that only for the signal with a S/N =-9 there were some mistakes, and these mistakes were mainly for ungrammatical, but semantically plausible strings.

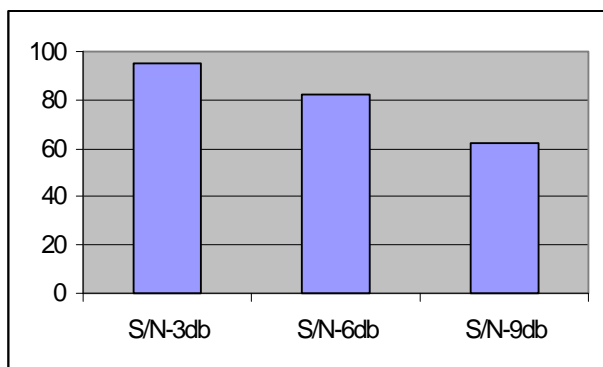


Figure 1. Percentage of correctly transcribed sentences for the three conditions.

With respect to the word level, Word Error Rate (WER) of all sentences for each condition has been

measured by means of an automatic dynamic alignment software. As shown in figure 2, the values of WER are quite low for sentences in conditions 1 and 2, while it rises for the Noise condition 3.

Furthermore, the mean of re-listening times has been calculated in order to have a measure of an overall intelligibility. Table 1 shows that there is a rising of re-listening times only for the condition S/N=-9 db.

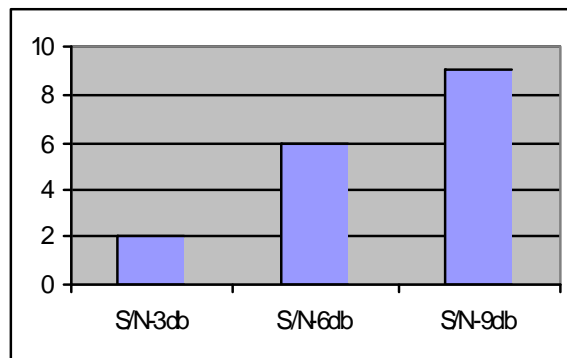


Figure 2. Word error rate for all utterances in each condition

condition	re-listening times
S/N=-3 db	1.8
S/N=-6 db	2.4
S/N=-9 db	3.8

Table 1. Mean of re-listening times for the three conditions.

For the recognition test we just calculated the correct answer. The results, reported in table 2 show that the normal, grammatical sentences were always correctly recognised; the grammatical, but semantically anomalous sentences were recognised 80% of the times, and the ungrammatical sentences were recognised only in 60% of cases.

sentence type	percentage of recognition
grammatical sentences	98%
grammatical, but semantical anomalous	80%
ungrammatical but semantically correct	60%

Table 2. Percentage of correct answer for the recognition test

When the stimulus heard was the ungrammatical but semantically plausible sentence, in 40% of the cases the subjects indicated the grammatical corresponding sentence on the answering sheet.

For instance, if the stimulus was a sentence like:

He look very nice in him red shirt
and the possible choices were:

He looks very nice with his red shirt
He look very nice in him red shirt
He took them ice in a wet dirt

The subjects indicater the grammatically correct answer: He looks very nice with his red shirt as they re-process the sentence in a correct way. When the stimulus was a ungrammatical and semantically non plausible sentence the possible choice was then random.

5. Comments

Our results, show that syntax plays a quite important role in the perception of speech: only the ungrammatical sentences obtained lower intelligibility (40%) compared to the intelligibility of grammatical sentences and 18% compared to the intelligibility of grammatically but semantically implausible sentences.

These results provide a clear demonstration that the human speech comprehension system is better disposed to constructing a postperceptual analysis of a semantically anomalous but grammatical sentence than one that is ungrammatical.

From our results we can also state that on average under a level of S/N equal to -6db it is not possible to understand exactly what it's being said.

This was also evident in a previous test we run on a different speech corpus [6].

Of course the intelligibility depends on the content of the utterances, but under this threshold level, being quite impossible to perceive what is actually said, it is more likely that the effects of "perceptual illusion" takes place. One evident case is the example translated and reported in Fig.4.

One of the stimuli the subjects heard was a grammatical sentence without any semantic sense, while (the normal grammatical and semantically plausible sentence and the ungrammatical, but semantically plausible sentence) were both related to drug pushers.

Being the subjects police officers, 50% of the answers resulted uncorrect as they believed to hear drugs rather than mugs.

Those person pushes drugs Those people push Of course idle crush
--

Fig. 4. Example of a forced-choice answer

This shows that in some circumstances the desire to "hear" these words or the expectation to hear some words are so strong that even when the signal is weak and of poor quality, the peceptual system processes the items which actually were not uttered [7].

So in conclusion it doesn't seem very easy to answer the initial question: are transcriptions of speech material recorded by means of bugs reliable? Sometimes it is possible to report exactly what it is heard, but some other times due to the bad quality of the speech recorded, due to the post-perceptual processing effects and due to the expectations of the listeners, it is possible that what is transcribed is not what was originally produced by the speakers.

APPENDIX A

Sentences used in the transcription tests

TEST1

"Gino ha ingoiato l'avvocato"
"Lisa é un'ammaliatrice"
"Mauro assaggiava la mortadella"
"Ieri li hanno convocati"
"Ti fa andare da sola"
"Tanto non arriva"
"Tutto fumo"
"Anche a passarci accanto"
"Perchè volevano che ritrattasse"
"Come sei stato bravo"
"Hanno condotto male le indagini"
"Ha smontato la bicicletta"

TEST 2

"tutto nudo"
"manco a portarci il cane"
"perchè non era la verità"
"tramite conto slavo"
"hanno riprodotto male le immagini"
"ha scippato la vecchietta"
"Nino ha chiamato l' avvocato"
"Lisa ha una mitragliatrice"
"Mario bastonava la sorella"
"Ieri lo hanno collocato"
"mi fa male la gola"
"io non ho una lira"

APPENDIX B

Sentences used in the recognition tests

Tutto crudo

Lutto fumo
Tutto nudo

La portinaia controllava l'entrata nel palazzo
L'apartheid concomitava l'errata del codazzo
La portinaia controllando l'entrata nel palazzo

Bisogna effettuare i prelievi domani
Agogna espettorare i criteri dei cani
Bisogna effettuare i prelievi domani

La pioggia non danneggia l'agricoltura
La loggia amareggia l'apicoltura
Le piogge non danneggiano l'agricoltura

Il supplemento esce una volta la settimana
Il condimento cresce una donna egiziana
Il supplemento escono una volta a settimana

Traducono abbastanza bene dal cinese
Inducono rimostranza del bel paese
Traducono abbastanza buono dei cinese

Le elezioni sono state indette in maggio
Le colazioni sono state dirette nel faggio
Le elezioni hanno indire in maggio

L'autore di questi racconti è già molto noto
L'amore per i mesti racconti è come una moto
L'autori di questi racconto è già molto noto

Tu devi verniciare il mio cancello
Tu credi ordinare il mio mantello
Tu devo vernicerò il mio cancello

Il caviale lancia le marmotte dei fax
I maiale mangiammo le pannocchie di mais
Il maiale mangia le pannocchie di mais

Il pozzo petrolifero era in fiamme
Il tozzo soporifero frenano in gambe
Il pozzo petrolifera erano in fiamme

Nuotava come un pesce eppure affoga
Mangiava sole e cresce pura la foca
Nuotano come un pesce eppure affogò

Quegli individui trafficano droga
Dell'indistinto raffica poca
Quegli individuo traffica droga

References

- [1] Paoloni A., Cerrato L., "Sulla trascrizione delle intercettazioni ambientali", in *Rivista Italiana di Acustica*, 1996, Vol.20 n.4, p.159-164.
- [2] Parsons T., "Separation of Speech from Interfering Speech by Means of Harmonic Selection", *Journal of Acoust. Soc. Amer.*, 60, 1976, pp. 911-919.
- [3] Hollien H, (1992), "Noisy tape recordings in Forensics", in *ETRW Speech Processing in Adverse Conditions*, 167-170.
- [4] Taylor D. S., "Non-Native Speakers and the Rhythm of English", *Internat. Rev. Applied Ling. In Language Teach.*, 1981,19, pp. 219-226.
- [5] Bird H., Slips of the ear as evidence for the postperceptual priority of grammaticality, *Linguistics*, 1998, 36-3, pp.469-515
- [6] Cerrato L., Delogu C., Paoloni A., "Un test di valutazione dell'intelligibilità delle registrazioni ambientali" in *Atti del XXVI Convegno nazionale AIA*, Torino maggio 1998, pp. 297-301.
- [7] Warren R., Warren P., "Auditory Illusions and Confusions", *Scientific American*, 223, 1970, pp. 30-36.