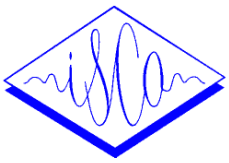

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

INTELLIGIBILITY OF THE FRENCH SPOKEN IN FRANCE COMPARED ACROSS LISTENERS FROM FRANCE AND FROM THE IVORY COAST

Christian BENOIT

**INSTITUT DE LA COMMUNICATION PARLÉE, U.A. CNRS N° 368,
INPG/ENSERG - Université STENDHAL, BP 25X - 38040 GRENOBLE, FRANCE**

ABSTRACT

An identical intelligibility test was run in Grenoble with twenty French listeners and in Abidjan with twenty Ivorian listeners for whom French is a second language. The listeners had similar educational backgrounds, but differed in their linguistic competence with regards to the comprehension of French as spoken in France.

Each subject had to manually transcribe 500 syntactically simple and semantically unpredictable sentences consisting of the most frequent French monosyllables. There were five stimulus series, one of spectrally distorted natural speech, and four of synthetic speech, produced by cross-varying two coding techniques and two prosodic models. The natural speech was obtained from the same French speaker whose voice served as reference in the text-to-speech synthesizers. The two automatic prosodic models were designed on the basis of French as spoken in France.

Results were compared across the two groups of subjects. Overall intelligibility scores, measured as the number of correctly transcribed sentences, fully discriminate the performances of the Ivorian and the French listeners. Nevertheless, the perceived linguistic complexity is independent of the listener's origin. The linguistic complexity is here quantified in terms of the number of independent decision units each subject dealt with when transcribing responses. This number is calculated as the ratio r = log (Ps) / log (Pw) where Ps is the proportion of correct sentences and Pw the proportion of correct words. This index is strongly dependent on the linguistic redundancy contained in well-formed sequences. It is here used in order to estimate the ability with which listeners take advantage of syntactic information when decoding semantically anomalous but syntactically correct sentences.

10.21437/ICSLP.1992-307

I. INTRODUCTION

French is a second language and the official *lingua franca* in Ivory Coast. French is widely and perfectly spoken by Ivorian students who make all their studies in French. However, its Ivorian specificity makes it sound pecular and rather difficult to understand by most of French listeners who are seldom accustomed to it. On the opposite, Ivorians are much more familiar with the French that is spoken in France (specifically in Paris), because of medias in general, and of the French representation at University for students too. This obvious comprehension asymmetry between the two variants is not studied here, as the intelligibility tests only involved the French variant spoken – or modelled – in France.

In this article, I compare intelligibility scores between the two French-speaking communities, obtained from two identical tests run in Grenoble (France) and in Abidjan (Ivory Coast). Overal intelligibility is first compared in terms of correctly transcribed sentences. I also evaluate the ability of the two groups of listeners to take profit from the linguistic redundancy that is introduced in syntactically normal sentences by the observance of lexical, morphological, and syntactic rules.

In a previous experiment [1], results were obtained from an intelligibility test at the sentence level in which natural degraded and synthetic speech was presented to twenty French listeners, all being students without any hearing impairment. A subjective measurement of the sentences linguistic complexity was presented too. Its quantification was reached by means of the index r = Log (Ps) / Log (Pw), i.e., the ratio of the logarithms between the proportions of correct sentences (Ps) and of correct words (Pp). This index was first suggested in order to quantify the effect of context in CVC words and in short sentences, as perceived by young adults [2], and later on by young children and older adults [3]. These studies show that the above mentioned index corresponds to the number of independent units listeners deal with during their decoding process in order to understand speech. These abstract and subjective entities should be related to the so-called *decision units in the perception of speech* [4] , as a result of parallel top-down and bottom-up processes human run when understanding spoken messages.

II. THE STIMULI

On the basis of a simple paradigm [5], five syntactic structures were selected by partners from a collaborative European (ESPRIT SAM) project, due to their simplicity, and to their crosslinguistic suitability – with minor changes – to any Indo-European language [6][7]. Sentences were randomly generated from grammatical lexicons in which the most frequent monosyllables had been previously compiled. Examples of such sentences are given below in Table 1.

20 sentences were generated in each of the five structures. The experiment involved two French diphone-based synthesizers using the same dictionary with different coding techniques. Both of them generated the 100 sentences under two prosodic conditions: a constant-flat prosody (all phoneme duration = 100 ms and Fo = 100 Hz on each voiced frame); and the automatic prosody modelling of each synthesizer, both of them being supposed to model the French spoken in France.

The original speaker of the diphones was also recorded to compare the 100 natural sentences thus obtained (slightly degraded) with the 400 synthetic ones. A latin square order was used to pair the five voices with the the five sets of syntax along five tapes of 100 sentences each. A pseudo-random order of recording avoided successive presentations of the same syntactic structure and of the same voice. A 10 s pause was inserted between two sentences, and a 1 mn pause was inserted every twenty sentences.

Table 1. Examples of semantically unpredictable sentences used in the test. The term-to-term translation in English is given beside between parentheses.

Structure 1 (intransitive)
La robe entre vers la science rouge.
The skirt comes into the red science.

Structure 2 (transitive)
Le verre vrai ouvre le coin.
The true glass opens the corner.

Structure 3 (imperative)
Tourne peu la date ou la main.
Lightly turn the date and the hand.

Structure 4 (interrogative)
Quand le texte pose-t-il la fille crue ?
When does the text puts the crude girl?

Structure 5 (relative)
La chose lance le train qui pense.
The thing throws the train that thinks.

III. THE LISTENERS

In Grenoble as well as in Abidjan, the twenty selected listeners were students at University, with no hearing impairment and with good ability in orthography and handwriting so that their global linguistic competence was homogenous within each group. None of them had special background in Speech Sciences or Linguistics. All were paid for participating in the experiment.

IV. THE TEST

The two tests were run according to a strictly identical protocol. In each country, five sub-groups of four listeners were submitted to five test sessions during each of which they had to handwrite the 100 heard sentences. The same latin square order as above was observed for presentation of tapes to each sub-group during the five sessions. It allowed any average of results across syntactic structures, voices, listeners, or sessions. In each sub-group, the four listeners were sitted apart in a sound-proof chamber. They wore high-quality headphones and were given answer sheets. A short training phase preceded the first test session. It included an acoustic presentation of 20 similar extrasentences in natural speech, followed by an orthographic and acoustic presentation of 25 synthesized sentences so that they were aware of the semantic vacuity and of the acoustic poorness of the messages. They were recommended to write down as much as possible of what they could understand or guess from the messages that were all supposed to have been generated by a computer. They received no other information about the linguistic or acoustic conditions of the test.

V. THE RESULTS

The responses from the 2 (tests) x 20 (listeners) x 5 (sessions, or magnetic tapes) x 100 sentences – i.e., 20,000 sentences, or around 130,000 words – were typed on a computer in order to be analyzed through various factors. The reader is suggested to refer to a previous article [1] where detailed results from the Grenoble test are presented. The data here presented are limited to a comparison of the most relevant observations: The global intelligibility of sentences and of words for each listener, averaged on all acoustic, syntactic, and session conditions on one hand; and the calculated index of linguistic complexity as presented in the Introduction, on the other hand.

5.a Global Intelligibility: A Comparison of Phonetic Decoding Abilities

The average articulation scores were measured from the whole test, for each subject, as the percentage of correctly transcribed units. The first chosen unit was the word, all words being monosyllables too. A word was considered as correct when it was written at the correct location in the correct grammatical category, with the same morphological features, and in an homophonic form (e.g., words *mer* "sea" or *mère* "mother" were considered as equivalent, whereas *mer* and *maire* "major" were not, though all of them are homophonic nouns /mɛʀ/, but the last two are not of the same gender). Words belonging to two different grammatical categories, such as *vieux* ("old": Noun or adjective in French) had previously been stored in the grammatical lexicon corresponding to the most frequent category (adjective, in this case) in order to decrease ambiguities. The second chosen unit was the sentence. A sentence is correct if all the words are correct, according to the preceding criterion.

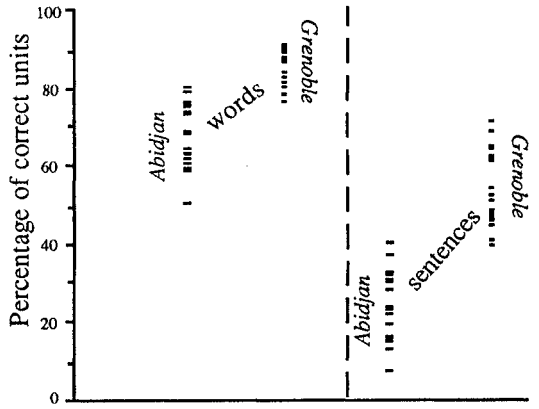

Figure 1. Average intelligibility scores, as percentages of correct words (left pannel), and of correct sentences (right pannel), obtained from 20 Ivorian listeners (left column in each pannel), and from 20 French listeners (right column in each pannel) on 500 sentences, i.e., on 3335 words.

Averaged intelligibility scores are presented in Figure 1, as proportions of correct words, and of correct sentences, from Ivorian and from French listeners. The difference in global performance between the two groups is obvious: Only the "best" Ivorian listener reaches a score in percentage of correct sentences higher than the "worst" French listener.

This global result must however be considered with care: The experimental conditions here described are (fortunately!) far from everyday situations of communication, where most of spoken messages are less unpredictable, i.e., with more semantic and pragmatic bearing, and less acoustically degraded. For instance, in such clear conditions as those provided by television, it can be anticipated that messages spoken in "French French" are as well understood by Ivorian listeners as by French listeners. Nevertheless, this results clearly shows that speech industrials, and, prior to them, speech researchers have still some work to do before their synthesizers be understood as clearly by all possible French-speaking users.

The overall result must also be reconsidered depending on the acoustic degradation, and probably on the prosody modelling. As shown on Figure 2, natural speech allows the two groups to be totally discriminated. Acoustic conditions are not too bad in this case, and prosody is pretty natural. A small overlap is observed in the case of synthesizer n° 1, i.e., in poor acoustic conditions. The use of synthesizer n° 2 leads to high confusions between the two communities, i.e., when acoustic transmission is very poor.

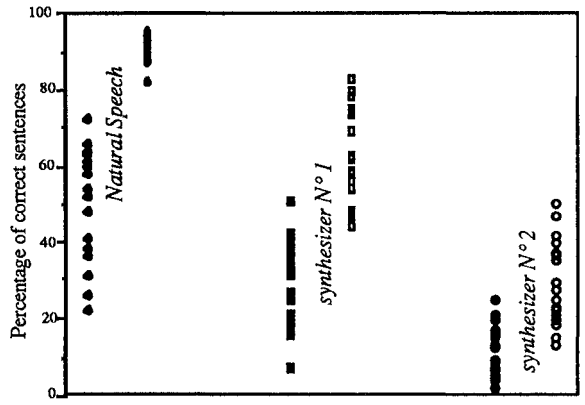

Figure 2. *Comparison of percent correct sentences between the two communities, over the acoustic conditions: Natural (slightly) degraded speech; and two types of synthesizer. In these last two cases, results are averaged on the two prosodic conditions. Ivorian listeners on the left, and French on the right, in each pair of columns.*

In fact, the prosody modelling of synthesizer n° 1 allows French listeners to increase sentence intelligibility (from 58.0%, on average, with a constant flat prosody, to 65.4%, see [1]), whereas very little improvment is observed with Ivorian listeners (from 27.4% to 28.0%, on average, between the same conditions). The prosody modelling of synthesizer n° 2 tends to slightly decrease intelligibility from the constant flat prosody condition, with French listeners (from 28.5% to 26.8%), as well as with Ivorians listeners (from 11.9% to 11.0 %). There is thus a strong relationship between acoustic quality and prosodic quality over the three coding conditions that were used. Therefore, two reasons might *a priori* explain this decrease in differences between communities understanding as intelligibility decreases: *The more similar the overal intelligibility by the two communities either the poorer the acoustics or the poorer the prosodic information.* Furthermore, due to the wide dispersion of results between subjects of both community, the data do not allow me to significantly clarify this point. Furher experiments should be run in order to isolate the effect of prosodic information on speech intelligibility from the effect of acoustic degradation on the improvement of intelligibility due to prosody, so that comparisons could be made between two communities.

5.b Perceived complexity: comparison of linguistic abilities

Till now, differences have been observed between communities in their "low level" performances, in terms of acoustic phonetic decoding, and/or in terms of prosodic decoding, which involve mainly bottom-up strategies. It is thus interesting to compare "high level" performances, in terms of a linguistic competence that involve top-down strategies.

In order to evaluate cognitive differences between the two groups of listeners, I compared the linguistic complexity as they perceived it. To do so, I calculated, for each listener, the ratio of the logarithms of both proportion of correct units, r = Log (Ps) / Log (Pw), where Ps and Pw are resp. the proportions of correct sentences and of correct words. r is a strong index of the discrepancy between i) the theoretical binomial law of word errors distribution, these words being considered as independent from each other, and ii) the observed distribution of word errors in the test sentences, where simple syntactic rules involve relations between some of them in French: concord in gender between the determinant, the adjective, and the noun; morphological agreement between those and the verb; order in grammatical units in the sentence, etc. It has been showed that the index r somehow corresponds to the number of independent units of decision taken by a listener for the comprehension of any string from an homogenous corpus [1][2]. It has also been mathematically showed that the index is sensitive to small variations in the proportions Pw or Ps, especially under highly degraded or under highly clear acoustic conditions [1]. This is the reason why **r** was here applied to the whole set of 500 written sentences by each listener, whatever the voice, or the syntactic structure, or the session, in order to increase its accuracy. Therefore, it does not here reveal the role played by each syntactic structure in their "ease" to comprehension of individual words: It quantifies the global advantage given by the whole syntactic structures to the comprehension of sentences. Such an holistic approach is coherent, as the five structures are representative of French basic syntax, and as sentences are more or less of the same length (6.67 words per sentence on average).

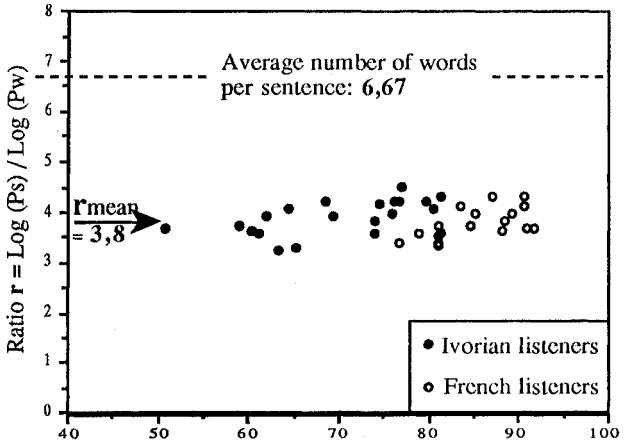

Figure 3. *Variation of the "complexity index" r = Log (Ps) / Log (Pw), as a function of global intelligibility (Percentage of correct words, among 3335 per speaker).*

Figure 3 shows the calculated values of the index **r**, for the forty subjects, as a function of the proportion of correctly transcribed sentences. **r** remains somehow constant over the listeners: 3,8 on average, in comparison to 6.67 words per sentence. All listeners use a similarly large ability to take profit from linguistic redundancy, just as if sentences were made of between 3.3 and 4.5 "independent units". There is thus no noticeable difference between the two communities in their linguistic abilities, whatever the intelligibility level of the sentences.

VIII. CONCLUSION

The acoustic conditions were highly degraded. The linguistic content of the speech material was not overcomplex. However, the results from the above described experiment allow me to state this: *Even if decoding acoustically – or prosodically – degraded messages is more difficult for Ivorian listeners than for French listeners, both groups reveal a similar competence in linguistic decoding, at a symbolic level.*

As far as I am aware of, no study has ever been carried out on the intelligibility of French compared across two French-speaking communities. The analysis and the systhesis of the French spoken in Quebec has widely been undertaken in Northern countries. However, literature is very poor in studies on the French variants spoken in Southern countries. Nevertheless, talking machines will be widely used in Africa as soons as it will be in France or in Canada, since it has already been the case with many technological products.

Acknowledgements

This work has been supported by the ESPRIT "SAM" (n° 2589) European Project and by the *Agence de Coopération Culturelle et Technique* (thanks to the *Observatoire des Industries de la Langue*). This study could not have been carried out without the collaboration of the Applied Linguistics Institute of Abidjan, and most especially Zakari Tchagbalé and Christian Chanard. Special thanks to Véronique Rissoan, too, for her tremendous effort in typing in 20,000 unpredictable sentences that were unpredictably transcribed by forty different hand-writers...

References

[1] Benoît, C. (1990), "An intelligibility test using sematically unpredictable sentences: Towards the quantification of linguistic complexity", *Speech Communication*, 9, 293-304.

[2] Boothroyd, A., & Nittrouer, S. (1988), "Mathematical treatment of context effects in phoneme and word recognition", *Journal of the Acoustical Society of America*, 84, 101-114.

[3] Nittrouer, S., & Boothroyd, A. (1990), "Context effect in phoneme and word recognition by young children and older adults", *Journal of the Acoustical Society of America*, 87, 2705-2715.

[4] Miller, G.A.M. (1962), "Decision units in the perception of speech", *IEEE Trans. Inf. Theory*, 8, 81-83.

[5] Nye, P.W., & Gaitenby, J. (1974), "The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences", Haskins Lab. Stat. Rep. on Speech Research, 37/38, 169-190.

[6] Grice, M. (1989), "Syntactic structures and lexicon requirements for Semantically Unpredictable Sentences in a number of languages", *Proceedings of the ESCA workshop on Speech I/O Assessment and Speech Databases*, Noordwijkerhout, Hollande, 151-154.

[7] Benoît, C., van Erp, A., Grice, M., Hazan, V., & Jekosh, U. (1989), "Multilingual synthesizer assessment using semantically unpredictable sentences", *Proceedings of the Eurospeech '89 Conference*, Paris, 633-636.