SOUND NAVIGATION: SONIFIED HYPERLINKS

Patrick Susini, Sandrine Vieillard, Emmanuel Deruty, Bennett Smith, Cecile Marin

Institut de Recherche et Coordination Acousique/Musique (IRCAM-CNRS)
1 place Igor Stravinsky, F-75004 Paris, France
susini@ircal.fr

ABSTRACT

This article deals with the idea of hyperlinks in the auditory realm, "sonified hyperlink", which is analog to the visual hyperlinks of the HTML language. The aim of the present article is to propose acoustical recommendations for sounds used to underline a word in spoken text.

1. INTRODUCTION

This study, part of a project to develop techniques for the creation of on-line radio programs based on an individualõs thematic choice, focuses on the ergonomics of sonification in the context of web radio. The sonifications, superimposed on the audio stream, reflect the functions of this medium: hyperlinks, alerts, "assistants", etc. É on the one hand, additional information about the program being listened to or about a similar program proposed in a real time channel, and on the other, unrelated events, such as the arrival of mail, a message from a background process, or a reminder of an appointment. This information should be unambiguous, in spite of the fact that the sonifications occur simultaneously with the radio program. Program information might be visual, such as a web page, or audible, such as related program, or downloadable. If the program is musical, for instance, the information could include a biography, a bibliography, an auditory sample or a portrait of a composer. The goal is to inform the listener simultaneously about supplementary information related to some specific content of the program (figure 1). The choice has been made to "underline", that is, to audibly emphasize, a spoken word or a few words of a radio program by superposing a non-speech signal (the "underlining sound"). This technique has the advantage of simplifying post-production of the recorded text. In an analogy to the hyperlinks of the HTML language, we speak of a "sonified hyperlink". The aim of this study was to quantify how well a given underlining sound fulfills its function of hyperlink without interfering with the program material. In other words, what are the caracteristics of sounds which best achieve the underlining function. The final purpose of the present article is to propose acoustical recommendations for sounds used to underline a word in spoken text. This raises the question of the experimental protocol to use for testing different acoustical configurations of the sounds. Thus, an experimental protocol (material and procedure) has been designed (2) and different acoustical parameters have been tested using synthesized sounds (3). The analysis and the results are presented in the last part of the article (4).

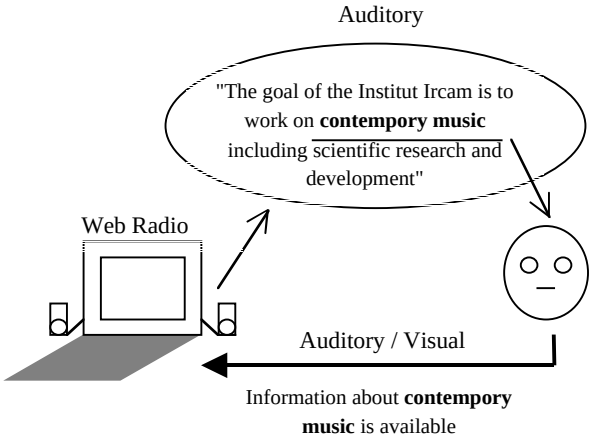

Figure 1˚: *"Sonified hyperlink"*

2. EXPERIMENTAL PROTOCOL

The experimental goal is to determine which sounds best fulfill the function of hyperlink. The two questions addressed are: 1. Is the "target" (the underlined word) correctly identified and understood by the listener? 2. Has the comprehension of the spoken material been affected by the underlining sound? In order to respond to these questions, particular care was taken to create the textual material presented to listeners (2.1) and the experimental procedure was divided into three tests (2.2).

2.1. Material

A number of constraints must be respected in obtaining a reliable group of text excerpts which permit comparison of the different underlining sounds independently of the underlined target and its context in the excerpt.
The criteria controled to be nearly the same among the texts were: - the speaker who read the text (male) - the grammatical structure (adverbial phrase, subject, verb, direct object) - the position of the target (nearly in the middle) - the number of syllables in the text (~ 57) - the target's frequency in the French language (~ 44 millions) - the number of syllables in the target (3). Based on these constraints, ten test texts were developed in French with nearly the same caracteristics except the content. The ten targets (the underlined word) are presented in table 1, in french and in english. Each target was integrated into a text which met the above constraints. The "underlining sounds" are applied to the words and are tested in the

following experimental procedure. One french text with the word "producteur" ("producer") is presented below˚:
" Nous avons le plaisir de recevoir l'auteur du film, Les chevaux de Sophie. Gr ce ce film, le **producteur** a r alis des b n fices importants. C'est donc une r ussite totale la fois artistique et commerciale."

mots cibles	**target words**
capitaine	captain
employe	employee
producteur	producer
paysage	landscape
musicienne	musician
ecrivain	writer
edition	release
document	book
emission	program
sensation	impression

Tableau 1: *Target words in french and in english.*

2.2. Procedure

Three tests were realized successively for each subject. The first two explored the effectiveness of the underlining, and the third measured the "nuisance value" of the underlining sound.

- Test 1: Identification. In the first test, the subjects heard each text only once, and were asked to identify the underligned word, by typing the word on the computer keyboard. For each text, the percentage of correct answers for all subjects was calculated, giving an objective measurement of the effectiveness of the underlining.

- Test 2: Effectiveness. The subject heard the same stimuli as for test 1, and was asked to rate subjectively the effectiveness of the underlining, on a scale *Very poor, Poor, Average, Good* et *Very good*.

- Test 3: "Nuisance value". The subject heard the same stimuli as for test 1, and was asked to rate how distracting/annoying the presence of the underlining sound was, on a scale *Very distracting* to *Not distracting* (the French word used was *g nant*).

3. UNDERLINING SOUNDS

The underlining sounds were synthesized from a *reference* sound (in this study, pink noise), to which were applied different transformations. The goal of the study was to determine which transformations produced sounds which were effective underliners of the target text.
Three groups of acoustical transformations were defined and tested using the above procedure. Of interest was the interaction between independent parameters, to evaluate their interdependence and relative effectiveness. The three groups are presented below and the different values of the parameters are indicated in table 2.
- Group 1˚: Spectral signature. A passband filter was applied to the reference sound. Two parameters were combined : the central frequency (low, medium, high), and the bandwidth (narrow, wide, very wide) of the underlining sound, giving 9 stimuli.
- Group 2˚: Energy and position. Two parameters are combined˚: the energy ratio between the underlining sound and the target (weak, medium, or strong), and the time difference (figure 2) between the underlining sound and the target (anterior, simultaneous, or posterior), giving 9 stimuli.

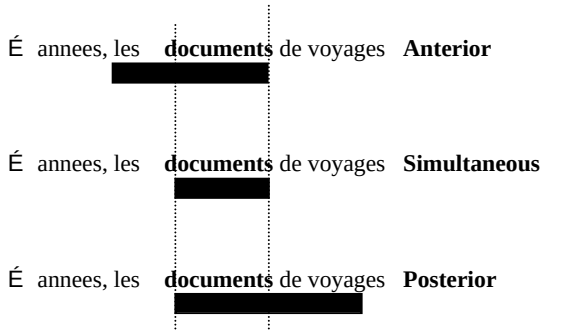

Figure 2˚: *Position˚:three position (A: anterior, S: simultaneous, or P: posterior)*

- Group 3˚: Attack and modulation. Two parameters are combined˚: attack time (gradual, rapid, abrupt) and the post-attack amplitude envelope over time of the underlining sound (constant, or following the amplitude envelope of the target, see figure 3), giving 6 stimuli.

For each of the three underlining groups, the three tests were carried out, using different subjects (about 30) for each group (See table 2). There was one text (the one with "capitaine" as the target word) serving as a control for all tests, where the underlining sound was not modified, that is, the original reference sound (pink noise) was used. This stimulus was therefore the same for all listeners. A summary of the experimental parameter configurations, the parameter values and the analyses is given in table 2.

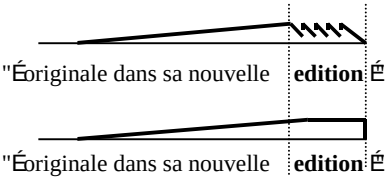

Figure 3: *Attack and modulation: two amplitude envelope with a gradual attack: following the amplitude envelope of the target and constant.*

4. RESULTS

The results of the the objective test (Test 1) are presented for the three underlining groups, and the the more significant results from the subjective tests (Test 2 and Test 3) are presented.

4.1. Test 1: Identification

The results for the three groups show that:
- All the spectral configurations permit a good identification of the target. The success rates approach in general 100% (χ^2=3.56, NS, df=9).
- Increasing the intensity of the underlining sound does not affect the identification rate.

- On the other hand, if the underlining sound preceeds the target (with or without ramp) there is a significant reduction of the identification rate (90% to 30%, figure 4)

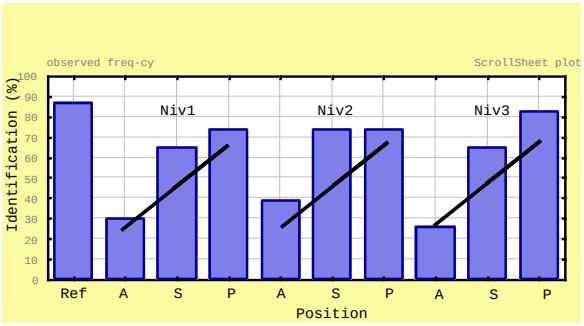

Figure 4: *Effect of the temporal position of the underlining sound on the target's identification (identification task).*

4.2. Test 2: Effectiveness

The results for the three groups show that:
- The underlining sound is judged more effective when for the Narrow Band, NB ($F(2, 58)=10.07$, $p<0.001$). For one group of subjects, this result is even more pronounced if the central frequency (fc) is 1kHz (figure 5).
- The underlining sound is judged less effective when the RMS difference between the underlining sound and the target is greater than around 4 dB ($F(2, 44)=5.0$, $p < 0.05$, figure 6) and as well when it preceeds the target by a ramp, irrespective of whether it is short or long ($F(2, 46)=6.5$, $p < 0.05$, figure 7). This is coherent with the results of Test 1.

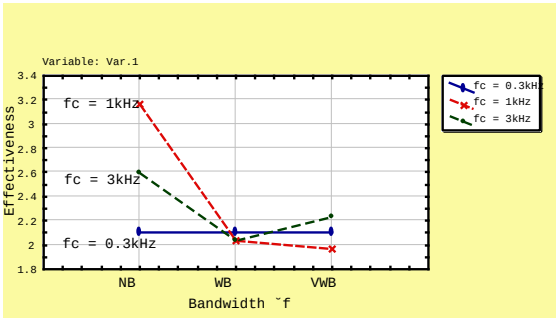

Figure 5: *Effect on the effectiveness of the spectral signature.*

4.3. Test 3: "Nuisance value"

The results for the three groups show that:
- The nuisance value increases when the frequency range of the signal includes those to which the ear is most sensitive: 1 et 3 kHz ($F(2, 58)=6.6$, $p < 0.005$).
- The nuisance value increases when the level of the underlining signal is increased ($F(2, 44)=5.7$, $p < 0.01$).
- The nuisance value increases when the the ramped underlining sound preceeds the word, the more so the longer the ramp time ($F(2, 46)=7.04$, $p < 0.01$, figure 8).

On the other hand, the modulation of the underlining sound appears to have no effect for any test (1, 2, ou 3).

which is more pronounced the longer the duration of the ramp (90% to 5%)

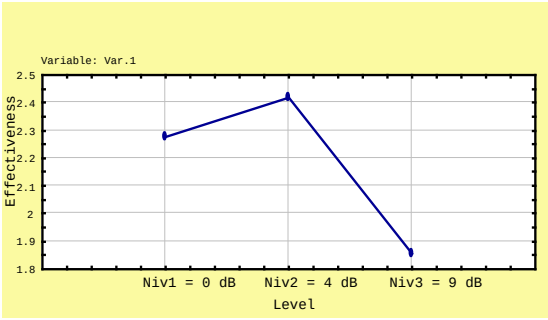

Figure 6: *Effect on the effectiveness of the Level ratio between the underlining sound and the target.*

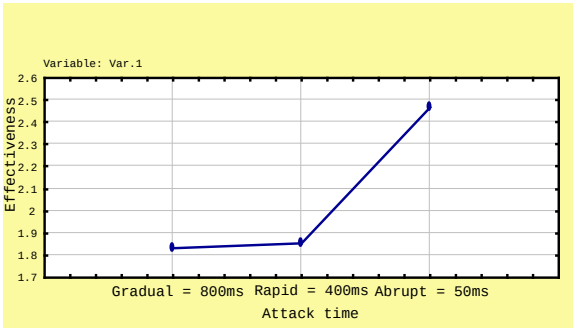

Figure 7: *Effect on the effectiveness of the attack time.*

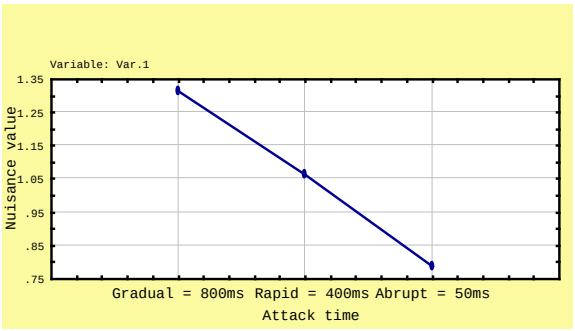

Figure 8: *Effect on the Nuisance Value of the attack time.*

5. SUMMARY

- All the underlining sounds, no matter what their spectral configurations, succeeded in helping the subjects identify the targets in each of the stimulus texts. The effectiveness judgements revealed that a narrow band sound (NB) centered on 1kHz was judged significantly more effective than wider band sounds. As for Nuisance Value, the sounds with spectral content in the 1-3kHz range were judged more distracting, the more so the wider their bandwidth.
- The sound level of the underlining sound relative to the the level of the target should not be too strong, if we wish it to be both effective and not a distracting nuisance.

On the other hand, contrary to expectations:
- There is a clear drop in identification performance when the underlining sound preceeds the target.
- It appears that the modulation of the underlining sound with the temporal envelope has no significant effect.

Experiment	Group 1	Group 2	Group 3
- Parameters	Central frequency (fc) Bandwidth (Δf)	Level (Lvl) Position (Pos)	Attack time (Att) Modulation (Mod)
- Values	fc = [0.3, 1, 3] kHz Δf = [NB, WB, VWB][1]	Lvl = [Niv1, Niv2, Niv3] Pos = [A, S, P][2]	Att = [G, R, A][3] Mod = [NM, M][4]
- Parameter combinations	3 (fc) * 3 (Δf)= 9	3 (Lvl) * 3 (Pos) = 9	3 (Att) * 2 (Mod) = 6
- Subjects	30 subjects	28 subjects	29 subjects
- Objective tests (Recall) - Subjective tests (Effectiveness, difficulty)	Chi-2 (χ^2) analysis Principal component analysis (PCA) Analysis by repeated measure of variance (ANOVA)		

Table 2: Experimental stimuli used and data analyses employed.

[1] NB: narrow band. WB: wide band. VWB: very wide band.
[2] A: anterior. S: simultaneous. P: posterior.
[3] G: gradual. R: rapid. A: abrupt.
[4] NM˚: non modulated. M˚: modulated.