

ERGONOMICAL ANALYSIS OF PARTIALLY DIGITIZED SOUND-RECORDING STUDIO

Feliksas Laugalys

Gamtos mokslų daktaras
Vilniaus universiteto
Bendrosios psichologijos katedros docentas
Didlaukio g. 47, LT-2057 Vilnius
El. paštas: Feliksas.Laugalys@fsf.vu.lt

Albinas Bagdonas

Profesorius, gamtos mokslų daktaras
Vilniaus universiteto Socialinio darbo katedros ir
Specialiosios psichologijos laboratorijos vedėjas
Universiteto g. 3, 903, LT-2734 Vilnius
El. paštas: Albinas.Bagdonas@fsf.vu.lt

This paper provides the comparative ergonomical analysis of traditional and digital sound record studio. There are selected four transitional steps in moving from traditional magnetic tape recoding to digital book : classical (traditional) talking book at the audio studio; installation of digital code in the technology of talking books; audio studio without announcer and studio of full speech synthesis for talking book in this article. A short description of performance of sound studio operator is presented there also. On the basis of Moscow town N enterprises ergonomic analysis of partially digitized sound – recording studio are formulated some ergonomic – psychological recommendations for future improvement.

1. Talking books production and installation of digital technology component

One of most serious problems in development of information processing technologies is information storage as well as efficient and precise reproduction of it. There are many various technical solutions in this field. Most of them are using analogic signal, for example magnetic tape recording, based production of talking books. Implementation of digital components in such production caused technology to move from the “classical” talking book at the audio studio to the studio with full speech synthesis. There are some ergonomic advantages in this new technology (Laugalys F, 1995)¹:

- a) a high-quality acoustic signal is produced for listening
- b) qualitatively new changes of acoustic signal manipulation in the reproduction stage have appeared (Schmidt – Nelsen, 1995)
- c) in the recording stage, sound director change the acoustic attention to the signal output into the acoustic verbal attention to the signal output (facilitating and simplifying in this way signal quality control operations performed by a sound director)
- d) sound director is released from permanent checking of sound quality to switch over to a relatively less complicated control of verbal characteristic of the speaker’s speech
- e) sound record in a final stage of digital studio evolution, becomes available for reproduction by means of conventional personal computer of standard architecture (Tucer and Jones, 1991)

¹ This investigation was granted by DG of the Commission of the European, project Copernicus, Digibook – 806 (Digitized Speech Processing for Efficient Distribution of Texts).

The main purpose of this paper is to determine the main activities in moving from recording of a classical talking book at the sound studio to the new digital studio with full speech synthesis at the Russian blind association N enterprise (Moscow).

2. Necessary steps in the change of the production chain

The shift from the “classical” production of a talking books at the audio studio to a studio with full speech synthesis can be described as a gradual, evolutionary (not revolutionary) movement. This movement consist of four steps (Laugalys F, Bagdonas A., 1996):

1. “Classical birth” of a talking book at the audio studio. In this case a professional announcer is reading the book at the studio and a sound director is recording his speech to master tape for copying (Fig. 1). The copy operator is the

main publisher, who reproduces the magnetic cassettes for magnetic tape players. During this step all central problems in respect to information are not so urgent because this stage is temporary. More important are psychological problems, that can arise the changes towards the digital studio.

2. Implementation of digital code in the technology of talking books. This step is current state of art. In the “classical” technology chain the digital coding of book text is involved (Fig. 2).

Such innovation partially changes the function of sound engineer and brings a full change in function of the copy operator, who transforms into a technical operator, as he is now manipulating computer and copy equipment.

The final product of this technology remains without changes: it is a book on magnetic tape cassettes.

During this step we have to face two psychological/ ergonomic problems (chain “d, e, f”):

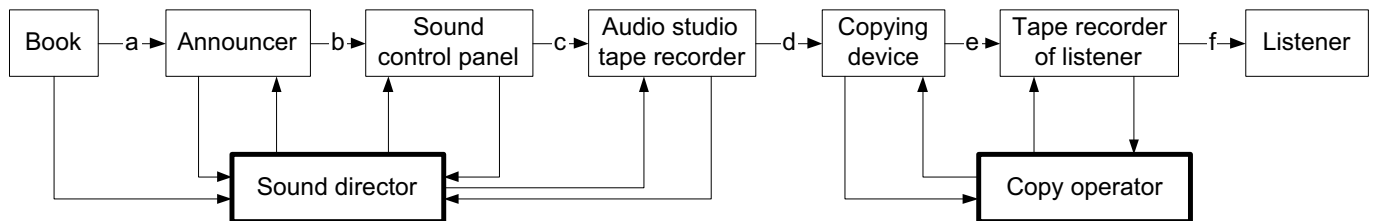


Fig. 1. “Classical birth” of talking book at the audio studio

a – text; b – signal from microphone; c – phonogramm; d – master tape; e – tape for listener; f – acoustic equivalent of text

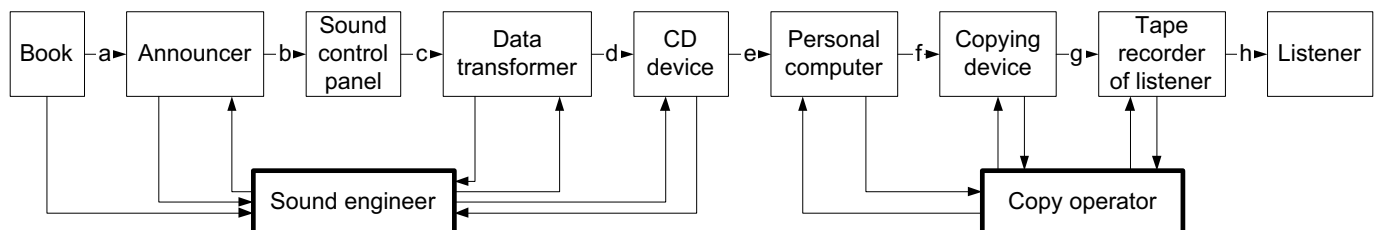


Fig. 2. Implementation of digital code in technology of talking books production

a – text; b – signal from microphone; c – phonogramm; d – digital equivalent of phonogramm; e – CD for memory and copying; f – analogous equivalent of digital code; g – tape for listeners of talking books; h – acoustic equivalent of text

The first: to design the workplace for both operators, to separate and to define their functions.

The second one is in psychophysical: it is necessary to determine the signal/noise relation and the optimal sampling frequency in digital technology. Both these parameters are influencing to subjective comprehension of the text by listener (consumer). Operators must be trained to judge the equality requirements on behalf of the end users.

3. Audio studio without announcer. The job contents of both the copy engineer and the technical operator are changing in this production chain. The main target of changes in this step are for the sound engineer.

The main function of the sound engineer is a mixture of the technique and the art. But at this step he is transformed into a technical operator too – an operator for text input. The text is then transformed into digital form by the equipment.

4. Studio of full speech synthesis for talking books. In this case we have the full use of speech synthesis in the technology of talking books (Fig. 4).

We have further changes in functions of the operator at the text input phase. In the present situation he becomes perform formation operations and must be able to handle various sources of information (databases, books, newspapers, magazines, compact disks, optical disks etc.)

He must also be familiar with current communication equipment: e-mail, fax, post, telephone, internet, mass media (TV, radio). There will also be an interaction between operators on the same level of the technology chain.

During this final step a serious changes in knowledge and attitude of the consumer (listener) can be predicted. Professional teaching of listener about sound handling, is in principle

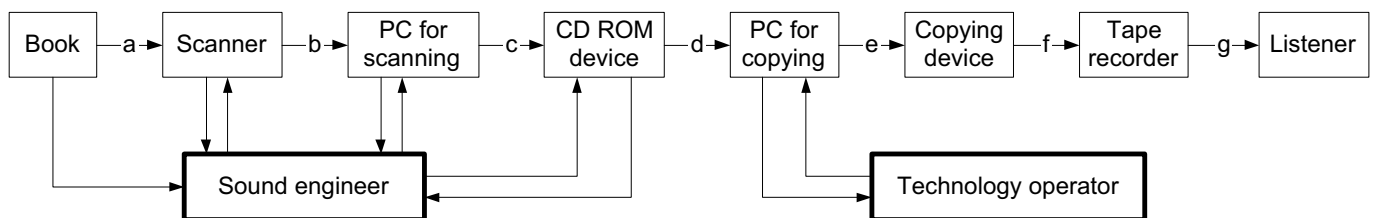


Fig. 3. Audio studio without announcer in technology of talking books

a – text; b – equivalent of text; c – formatting for digital recording; d – CD ROM; e – analogous equivalent of digital code for copying; f – tapes or CD; g – acoustic equivalent for copying

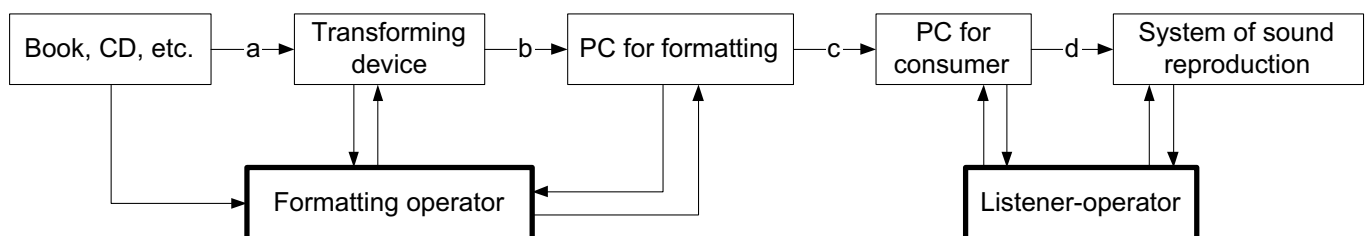


Fig. 4. Studio of full speech synthesis for talking book

a – input information; b – digital equivalent of input information; c – “talking book” at network or CD; d – analogous equivalent of synthetic speech

the same as teaching of the formatting operator at the studio side.

At this stage of technological development we can predict a full informational integration of blind in our society. It is this possible that such an information provision also brings a serious change in information processing and information provision for people with normal vision too. So, speech synthesis has a large potential to be used in our future life (Lau-galys F., Bagdonas A., 1995).

3. Professional activities and implementation of digital technology component

According to diagram (Fig. 4), at the present time Moscow N enterprise is doing a second step toward full speech synthesis studio (Bagdonas A. et al., 1996). The function of sound engineer partially changes. It can be concluded, that psychological parameters of professional activities performed by a sound director operating a digital studio in the recording stage are as follows:

a) introduction of the digital component into the talking book technology modifies considerably the creative work performed by a sound director turning it into the technical work performed by a sound operator;

b) currently, most sound directors have no special training in computer operation, therefore their professional qualifications are to be improved;

c) part of sound directors are ill disposed towards a computer (“computerophobia”);

d) real physical space within which the operator acts when operating a conventional sound-recording studio is changed into unreal virtual space of the new digital studio, not quite understandable to him;

e) sound director’s professional manual manipulative movements with tactual and proprioceptive feedback are turned, in the new digital studio, into the movements with visual and proprioceptive feedback (acoustic feedback is preserved in both cases, only its characteristics are changed).

4. Current ergonomic – psychological state of art of new digital studio

Current ergonomic-psychological imperfections of the new digital studio at the Moscow N enterprise are as follows:

a) the currently used sampling rate is technically proved to be sufficient for the maintenance of communication channel, however psychological proof may be insufficient, as far as the Russian language is concerned;

b) the BEAG RKC-112 sound-recording control panel is used exclusively for checking of the quality of speaker’s speech (of a number of its potentialities only the record quality indicator is used);

c) the “Wave SE-II” program currently used in the digital studio to record the speaker’s speech has been originally designed for recording music, therefore it has a number of shortcomings:

- as its direct purpose is to control the parameters of music recording, not speech, its record quality indicator has the integral sensitivity that is not sufficient to check human voice recording precisely (its sensitivity must be equal to that in the BEAG RKC-112 control panel, an opportunity to control sensitivity would be desirable);
- windows of the time indicator installed in virtual program panel (hrs min sec) are small-sized, consequently, when reading the indications, it is necessary to strain one’s eyes which is both inconvenient and tiring;

- information in the extreme right window of time indicator is used when providing videorecordings with sound. It is not necessary in our digital studio, as it only produces undesirable masking effects;
- the program does not provide an alternative key-board control opportunity from virtual magnetic recorder control panel, therefore it is not available to a blind person;
- a file name in this program cannot exceed the standard room allotted for 8-symbols which is obviously insufficient for making a full bibliographical description of a book: shortenings and codifications are unavoidable, therefore the title of bibliographical source is clear in fact only to the operator who is making the recording (individualised perception);
- in the course of recording of the speaker's speech, the mouse cursor must be placed strictly on the "stop" square, if it is moved a little by chance, the time of operator's reaction to the speaker's error is extended significantly;
- in the course of speech recording, the program's working screen is masked by the "Window 3.1" window located in its background, which impairs the reading characteristics of various indicators (increases the probability of error and extends the time of reading);
- virtual "hook" imitation is not provided in the program. The "hook" is a handy tool for correction of recording errors, it is located on the BEAG RKC-112 control panel (its presence would enable a positive skill transfer in the transition from an analog studio to a digital one);
- in the existing digital studio, sound director cannot as yet work independently without permanent assistance of a programmer.

5. Ergonomic-psychological recommendations

For Moscow N enterprise were proposed such ergonomic recommendations :

- a) in "Wave SE-II" program, an alternative opportunity to effect the virtual control of magnetic recorder panel and of the mouse movements by means of a keyboard must be provided. This will shorten the operator's reaction time and eliminate the necessity of keeping the mouse cursor permanently placed against the "stop" button;
- b) to add a file for making keyboard keys produce sound by means of speech synthesis. This will make sound operator's profession available to blind people;
- c) to make a virtual imitation of the "hook" of "Mechlabor STM 610" studio magnetic recorder;
- d) "Windows 3.1" graphics should be removed from the monitor's working screen in recording mode;
- e) record quality indicator's design and sensitivity should be adequate (this will enable to replace the expensive, intricate and only partially employed analog control panel with a cheap consumer standard amplifier and a loudspeaker);
- f) to increase the area of indication windows in the "Wave SE-II" program;
- g) only three indicators (virtual magnitude recorder control panel, time indicator and record quality indicator) must be left on "Wave SE-II" program working screen in recording mode, all the rest indicators must be removed;
- h) bibliographical source indication by means of 8 symbols must be altered: either the number of symbols must be greater, or, instead of them, a more spacious text file must be provided;
- i) "Wave SE-II" program must be made available to Russian language speakers. This will

increase the number of potential users, inasmuch as the original version of “Windows 3.1” has been already russified;

k) the speaker’s workplace in the digital studio may be the same as in a conventional analog studio;

l) sound operator’s working room may be the room which is currently occupied by a sound director, however it must be modified accordingly in order to meet the requirements for standard computer equipment;

m) to introduce an automatic acoustic record memorising variant and to secure sound record retention in case of an accidental electric power cut off;

n) conventional studio’s sound directors must be retrained, they must be able to work as sound operators;

o) to start arrangements for further changes in the functions performed by a person who operates the recording in the course of digital studio evolution: from sound operator’s profession to input operator’s activities;

p) to start potential user training: with the further evolution of digital technology they’ll have to pass from magnetic recorder to computer, otherwise the talking book will simply become non-available to them.

6. Sampling rate

As we mentioned above, it is necessary to determine the signal/noise relation (Gorenflo et al., 1994) and to measure the sampling rate influence. Both those parameters are influencing the subjective comprehension of speech by listener, particularly sampling rate.

A special psychophysical experiment was created to evaluate the influence of sampling rate on natural speech acceptability and 6 listeners were used. There are 2 periods in this testing: during first period listeners were trained and

informed how to detect changes in sound quality.

The second period was devoted to sampling rate effect measurement. Five sampling rate frequencies were used: about 11, 22, 32, 44 and 48 kHz. The poll of 75 sentences (the third subtest from speech intelligibility test) was recorded on magnetic tape with different sampling rate – each frequency had 15 sentences. Differently sampled sentences were randomised and then presented to listeners to detect changes in sound quality, depending on digitising frequency.

Statistical data analysis show no reliable influences of any sampling rate frequency to be used on natural sound speech quality. So, it can be stated that 11 kHz frequency is sufficiently for normal sounding of natural speech in “classical” technology of talking books with digital components.

In addition it can be said that subjects sometimes high sampling rate frequency speech evaluated as more “painful sounding”, particularly in sounds “s” and “tc”. Low sampling rate frequency speech some times tend to be named as having additional low frequency overtones.

7. Conclusions

1. Moving from classical tape recorder audio studio to a digital record studio with full speech synthesis can be described as gradual evolutionary changing including four stages.

2. Installation of digital component in technology of talking books forced to change functions performed by sound director and by copy operator, evoked industrial equipment renovation and required serious corrections of new production consumers knowledge and attitude.

3. Introduction of digital elements considerably modifies the creative art work per-

formed by a sound director turning it into the technical work of computer manipulations performed by sound operator.

4. The computer software currently used in the partially digitize studio to record the announcer speech has been originally designed for recording music and therefore it has a number shortcoming at new application and must be improved.

5. The currently used sampling rate is technically proved to be sufficient for the maintenance of communication channel and psychological experiment revealed that 11 kHz frequency sampling rate is sufficient for normal sounding of natural speech in classical recording of talking books.

6. The ergonomic – psychological recommendations for improving of the technology current state of art were presented.

REFERENCES

Gorenflo C. W., Gorenflo D. W., Santer S. A. Effects of synthetic voice output on attitudes toward the augmented communicator. *J. of Speech and Hear. Res.* 1994, vol. 37, no 1, p. 64–68.

Bagdonas A., Kasperaitis P., Laugalys F. Studies of the evaluation of the quality of the Russian speech synthesiser and creation of Lithuanian speech synthesiser. Final report, Copernicus 806, Vilnius, December 12, 1996.

Laugalys F. Digital book: Outlook of New Technology // Education and Rehabilitation of Children and Adults with Visual Impairments. Coll. of Theses of Conf. Dedicated to 70-anniversary of A. Baltramejūnas, Vilnius, April 20–21, 1995. Vilnius: LABLS, 1995. P. 43. (In Lith.).

Laugalys F., Bagdonas A. “Talking book”: from tape recording to computerised speech synthesis // Learning Difficulties and Disabilities. Coll. of Theses of Conf., Vilnius, September 12–14, 1995. Vilnius: VU Press, 1995. P. 47–51.

Laugalys F., Bagdonas A. Ergonomic aspects of the digital studio concept // Identity, Freedom, Values and Memory. Proc. of 2nd Internat. Baltic Psychol. Conf., Tallinn, August 23–27, 1996. Tallinn: Union of Est. Psychol., 1996. P. 109.

Schmidt-Nielsen A. Intelligibility and acceptability testing for speech technology // Applied speech technology. Boca et al: CRS Press, 1995. P. 195–231.

Tucker P., Jones D. Voice as interface: An overview. *Int. J. of Human-Computer Interact.* 1991, vol. 3, no 2, p. 145–170.

SKAITMENINĖS GARSO ĮRAŠŲ STUDIJOS ERGONOMINĖ ANALIZĖ

Feliksas Laugalys, Albinas Bagdonas

Santrauka

Straipsnyje pateikiama tradicinės ir skaitmeninės garso įrašų studijos lyginamoji ergonominė analizė. Išskirti keturi pereinamojo proceso etapai: klasikinės (tradicinės) “kalbančios” knygos, teksto skaitmeninio kodavimo, nediktorinės garso įrašų studijos ir garso įrašų studijos, kuriose garso įrašai visiškai grįsti kompiuterine kalbos

sinteze. Glaustai aprašoma kalbos sinteze grįstos garso įrašų studijos operatoriaus veikla. Remiantis realios Maskvos N gamyklos garso įrašų studijos įrengimo ir veiklos ergonomine analize pasiūlytos konkrečios rekomendacijos tolesniam tobulinimui.

Įteikta 2003 02 15