COINCIDENCES AND DIVERGENCES BETWEEN AUDIO TRANSCRIPTION AND TEXTUALIZATION

Coincidências e divergências entre transcrição e textualização de áudios

Maria Inês Beltrati Cornacchioni Rehder [1], Lucilene Aparecida Forcin Cazumbá [2], Rosângela Mitsue Kato Assis [3], Patrícia Jorge Soalheiro de Souza [4], Maria da Conceição Farias Freitas Tandel [5]

ABSTRACT

Purpose: investigate coincidences and divergences between audio transcription and textualization in order to check for statistical evidence which may be a justification as to the best procedure to be applied. **Methods:** retrospective study. 30 audios were selected randomly among the 239 available audios, proceeding from telephone intercepts of the same lawsuit. We considered: the number of words and time in minutes for each audio, and the comparative analysis for maintaining the main content highlights. Three Speech Pathologists transcribed and textualized different parts of the file, ensuring independence. A Speech Pathologists, who did not attend the previous step, conducted content analysis. For statistical analysis we used Wilcoxon-Mann Whitney test in R environment, with Tinn R interface. Significance level 5% (0.05). CEP: 274-742. **Results:** the mean number of words used in of audio file transcription was 27% greater than the number of words used in textualization, p=0.52. The mean time in minutes required to perform the transcription was significantly higher p=0.013. In the comparative analysis as for the maintenance of the main content highlights, we found that on average 93% of highlights were kept, p=0.61%. **Conclusion:** the similarities among the of transcription and textualization processes were compared with the median number of words and the maintenance as for the median number of content highlights. There was divergence as for the implementation time, significantly lower in textualization. Considering the data obtained in this study, textualization process proved to be the most suitable in audio de-recording.

KEYWORDS: Voice; Speech, Language and Hearing Sciences; Language; Forensic Sciences

(1) Parole Fonoaudiologia; CEFAC – Saúde e Educação; Núcleo de Perícias em Fonoaudiologia Forense; Universidade Federal de São Paulo, São Paulo, Brasil.

(2) Ministério Público do Estado de São Paulo; Núcleo de Perícias em Fonoaudiologia Forense; Universidade Católica de Brasília, Brasília, Brasil.

(3) Mogi das Cruzes Fonoaudiologia; Núcleo de Perícias em Fonoaudiologia Forense, Rio Claro, São Paulo, Brasil.

(4) Universidade de São Paulo - USP, São Paulo, SP, Brasil; Núcleo de Perícias em Fonoaudiologia Forense, Rio Claro, São Paulo, Brasil.

(5) Instituto de Geociências e Ciências Exatas, UNESP - Universidade Estadual Paulista, Rio Claro,SP,Brasil.

Conflict of interest: non-existent

INTRODUCTION

Forensic Science has been developed, studied and practiced in many countries for decades, contributing to justice in the most different knowledge areas. From the procedures involving Forensic Science with highlight in Human Communication, we highlight video and audio content analysis, transcripts, textualizations, communicative profile analysis, graphological-technical examinations, facial identification, and establishment of a causal link among hearing and/or vocals and occupational therapy. Among these, we highlight the identification

of speakers, used in civil and criminal proceedings as judicial evidence[1-3].

In Brazil, although recent, Speech Pathologists are now being inserted in specialized sectors of institutions working directly with the law, and as the Department of Justice and the Institutes of Forensic Expertise e and more States. Forensic Speech Pathology is, then, described as the interface between law and science, applying technical and scientific knowledge of human communication in judicial issues, and aiming to clarify the facts under legal interest, by using the grounds of Speech Pathology and its specialties, which include the areas related to hearing, voice, speech, orofacial motricity, oral written language [4].

Transcription and textualization, or de-recording, as it is known in the legal environment, are commonly performed procedures in the analysis of audio arising from wiretappings [5]. Authors experienced in the subject [6,7], as well as the commands of the Civil and Criminal Process Code, stressed out the need for the procedure to be performed by a person who holds *"expertise"* on a given area of knowledge, duly registered on the board of class 8 and/or, in cases involving human communication, demonstrating knowledge in the areas related to syntax, semantics, morphology, lexicology, dialectology, sociolinguistics, psycholinguistics, in addition to articulatory phonetics and acoustic phonetics [9-11]. The linguistic experience of the person performing it is also considered, with the proposal to build up the most relevant pieces of content, through the reproduction of discourses, intentions, situations, relationships and correlates chronologically chained [1,12].

The de-recording of audio material can be made using the transcription, which consists in transforming in writing exactly what is being heard, keeping the phonemic content and traces of prosody; or textualization which is based on a written narrative on the speaker's communicative intents. Studies on the contribution of textualization and/or transcription in a given audio material can guide the application of justice in either procedure, depending on the nature of the process at issue [13-16]. A timely request can accelerate the processes, benefiting the judicial power [8,17].

The need for the partial transcription or in its entirety for the audios and, if the same ones are considered expertise or documental proof [18-21] are ongoing discussions in the high level court in the country, considering the large volume of material to be analyzed, due to technological advancement, and cases of impact on the national political and economical scenario. Therefore, for this study, we will be considering the differences and similarities in carrying out transcription and textualization procedures as a way to contribute for the choice between one of the subjects in the process or law enforcement officers.

This study aims to investigate coincidences and divergences between audio material transcription and textualization in order to check the best applicability.

METHODS

This research was duly registered with the Brazil Platform having the approval of CEP (Committee of Ethic in Research) under number 274-742.

This is a retrospective study. The audio material is used as a sample comes from wiretapping records that were used to identify speakers in the same lawsuit. As this is not a comparative study, the samples retain the code of secrecy, since it does not identify any given process as well as the speakers. The researcher in charge for this study is committed to maintaining the confidentiality signing a Term of Usage Commitment and Data Disclosure.

For the making up the sample, the audios coming from the same process were submitted to random sampling statistical treatment using the R and Tinn-R[22] software. 239 audios that comprised the database were first registered by time, where the lowest audio had 0.13 min and the larger one 10.12 min. 30 audios were selected randomly among the 239 available. The boxplot is a chart representing the distribution of a data set based on the median and other quartiles, was used to describe the boxplot representation of the sample (Figure 1).

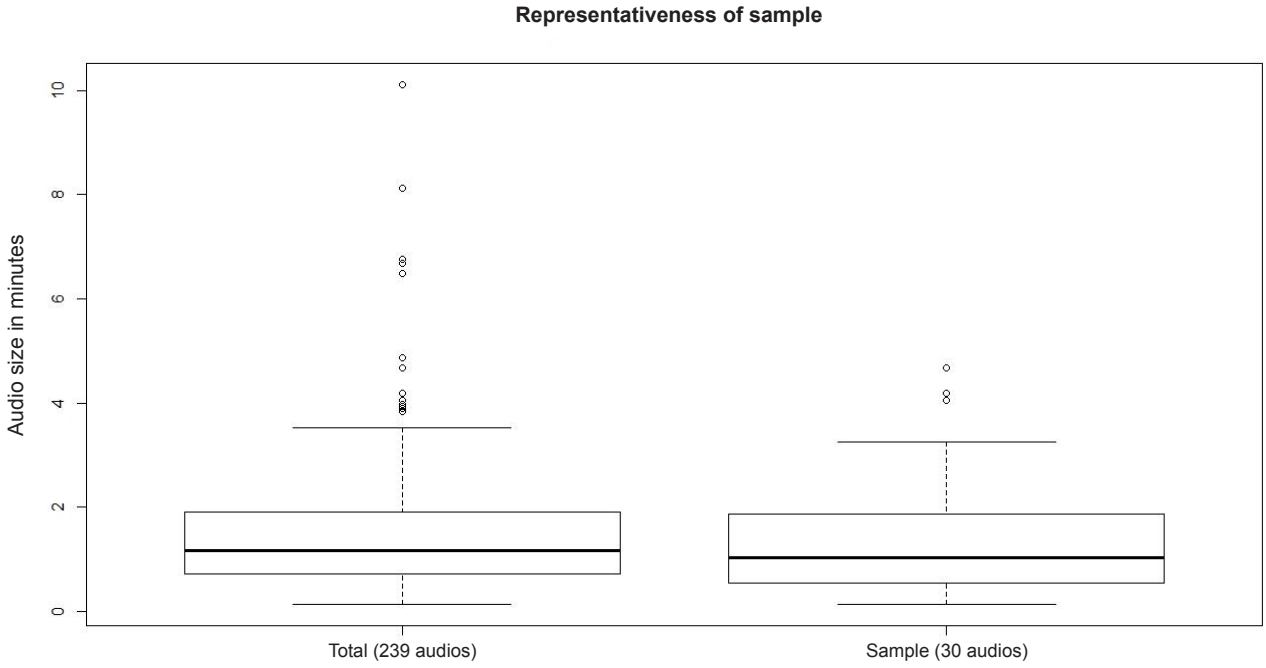

Figure 1 - Representativeness of sample by audio time distribution in minutes.

To check the coincidences and divergences among transcription and textualization procedures we considered:

1. Number of words and time in minutes for each audio.
2. Comparative analysis on the maintenance of the main audio content highlights.

Transcription and textualization were performed independently by three Speech Pathologists with training in Forensic Speech Pathology, from the same educational institution. It was up to each one to perform textualization for 10 audio files and transcription for other 10 files, different from the previous ones. The Speech Pathologists have had contact just with only the files intended for them by a raffle, avoiding the methodological induction bias that could occur if they had access to other de-recordings. Furthermore, they should mark the time taken to perform each task and the number of words used (Figures 2 and 3). The choice of the parts was done by draw. Standards of Conversation Analysis proposed by Marcuschi [18] were used to carry out transcription and textualization.

Comparative analysis regarding the maintenance of the main audio content highlights was performed by a Speech Pathologist, PhD in Human Communication Disorders and with Training in Forensic Speech Pathology. The Speech Pathologist listed the main content highlights contained in the transcription and verified whether they were kept in the corresponding textualization (Figure 4).

The results were compiled in tables and subsequently analyzed statistically. We performed the statistical analysis under R environment, with Tinn R interface. We used the Wilcoxon-Mann-Whitney[22] test for comparative analyzes, with 5% (0.05) significance level.

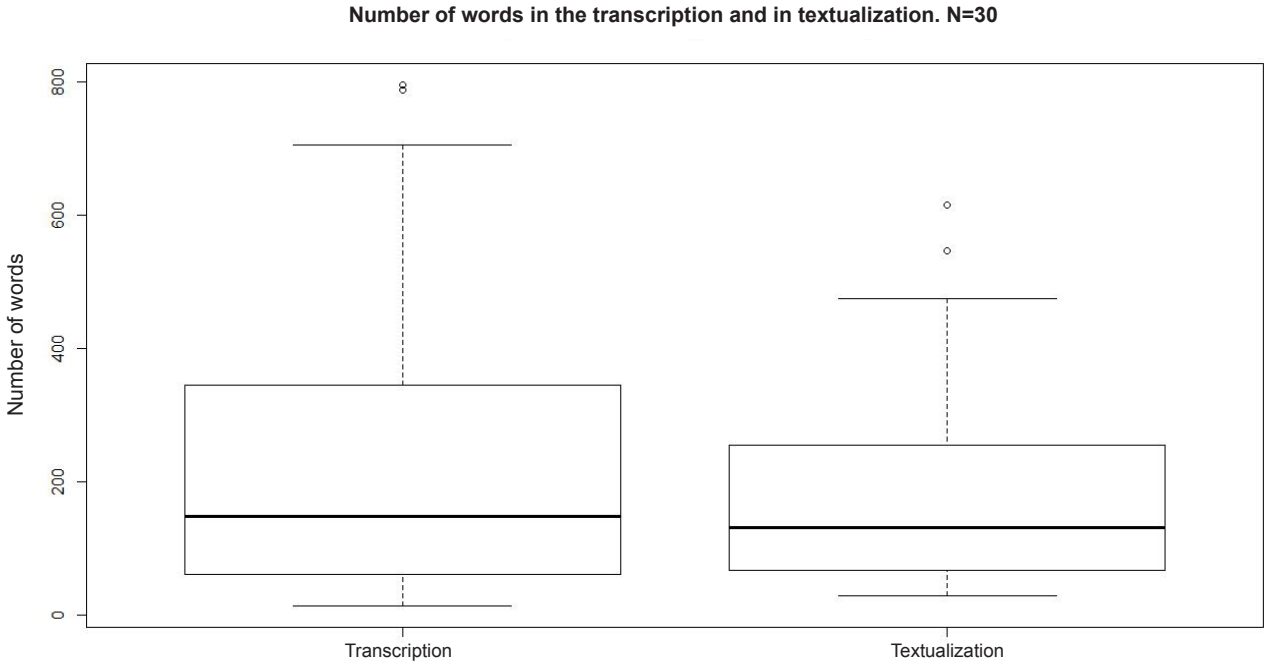

Figure 2 - Illustrative image in boxplot format for the distribution of the data set on number of words in the transcription and number of words in textualization

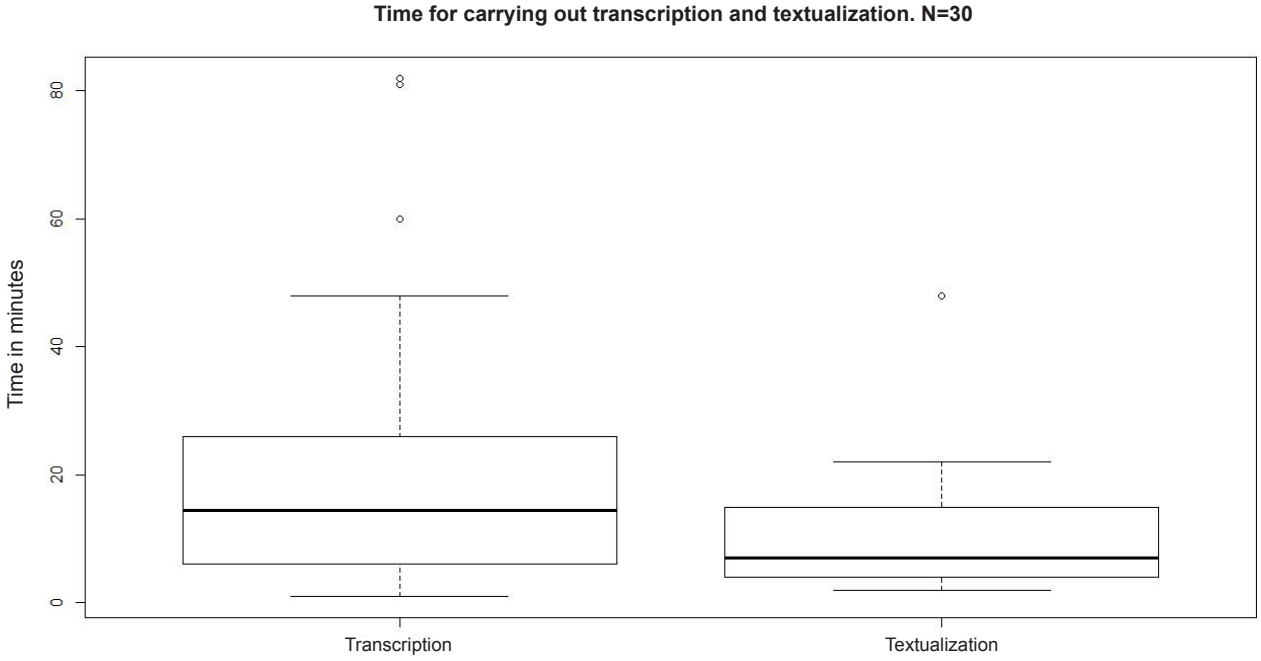

Figure 3 – Illustrative image in boxplot format for the distribution of the data set on the time in minutes, spent for carrying out transcription and textualization

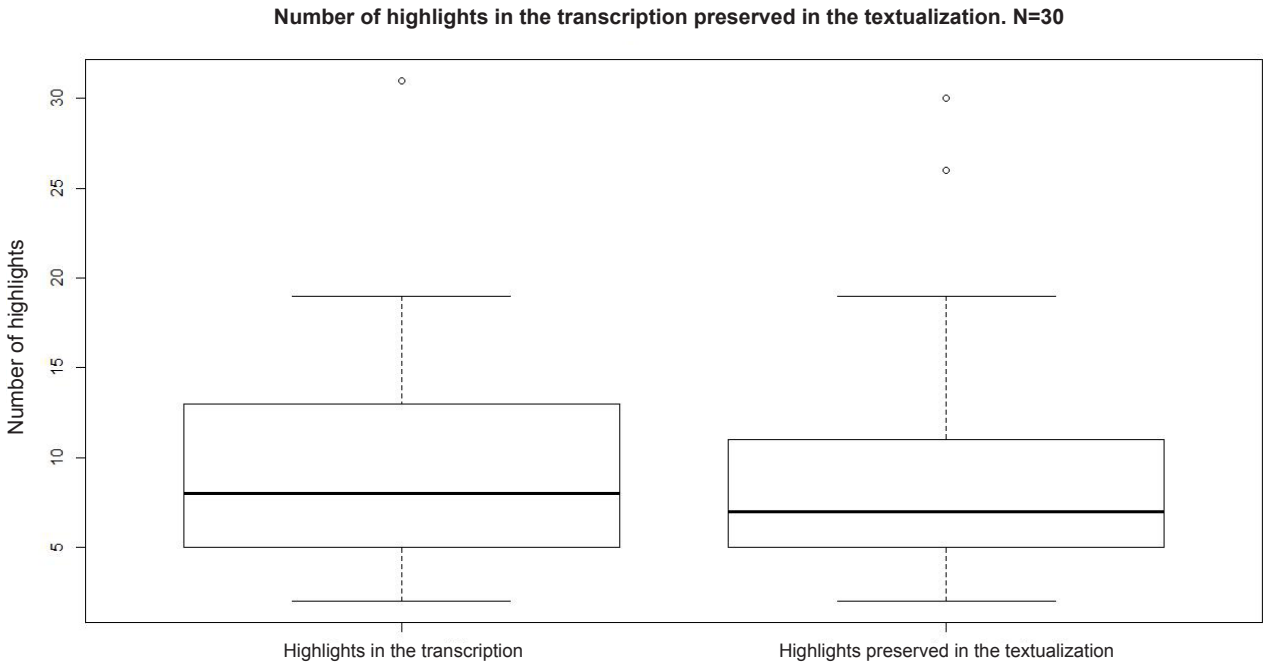

Figure 4 - Illustrative chart in boxplot format on the distribution of the data set on the number of highlights in the transcription and the number of highlights preserved in the textualization

RESULTS

The mean number of words used in of audio file transcription was 27% greater than the number of words used in textualization, p=0.52. The mean time in minutes required to perform the transcription was double the time required to perform the textualization, p=0.013. The evaluators transcribed on average 12.44 words per minute and textualized 18.79 words per minute. The time for completing textualization was on average half the time needed to perform the transcription. Such data are shown in Table 1, illustrative charts in boxplot format 2 and 3.

In the comparative analysis as for the maintenance of the main content highlights contained in the transcriptions found in the textualization, it was possible to verify that 983% of the highlights were kept, p=0.61%. These data are shown in table 2, illustrative chart boxplot 4 format.

Table 1 - Comparison between transcription and textualization considering the number of words and the time in minutes used for carrying out each sample. N=30

Variables	Size of audio (min)	No. of words Transcription	No. of words Textualization	Time for Transcription (min)	Time for Textualization
Minimum	0,13	14	29	1	2
Mean	1,44	244	184	21,5	10,3
Standard Deviation	1,24	229	158	21,7	9,4
Maximum	4,68	796	616	82	48
p		0,52		0,013 *	

Table 2 - Comparison between the number of highlights in transcription and preserving highlights in textualization. N=30

Variables	No. of highlights Transcription	No. of highlights kept in Textualization
Minimum	2	2
Mean	10,3	9,2
Standard Deviation	7,5	6,6
Maximum	31	30
P	0,61	

No. audio	Time spent in minutes for execution	No. of words
interlocutors	V1 x V2	
V1		
V2		
V1		
V2		

Figure 5 - Model for audio transcription

No. audio	Time spent in minutes for execution	No. of words
interlocutors	V1 x V2	
Text:		

Figure 6 - Model for audio textualization

No. Sample	Content highlights - transcription	Maintenance of highlights in textualization Yes/No

Figure 7 - Model for comparative analysis between transcription and textualization

DISCUSSION

The search for experts in the human communication field has been increasingly common in Brazil, a possible reflection on the commitment of the Brazilian judiciary in order to make justice more precise and transparent [23,24].

In this context, there are several requests for audio transcription and textualization arising mainly from wiretappings [23,24]. It falls to the expert not only to translate the audio material for writing, but also, in some cases, to identify the speakers' voices [25,26].

This study sought to elucidate coincidences and divergences among audio transcription and textualization in order contribute with the judiciary, law enforcement officers and/or subjects of the process, in the option of requesting either procedure.

According to the data found during the analysis of the number of words used during transcription and textualization, the following data were obtained: the total number of words used in the transcription was, on average, 27% higher than in textualization. However the coefficient of variation was very high, 35%, which leads to no statistical significance of this difference. These data are shown in table 1 and illustrative chart 2. The insignificant difference indicates that the text size and thus the reading time would be about the same in both procedures, and therefore, not constituting an important factor in choosing one of the two processes [27,28].

It is noteworthy that the act of transcribing is a procedure that is directly related to the linguistic baggage of the subject and the ability to interpret what is being heard, factors directly related to their academic grade and social-cultural level[27]. This

research sought to minimize the intra-subject differences suggesting that the transcription and textualization tasks were to be performed by professionals with academic training and converging socio-cultural level.

Analyzing the time spent for each task, it was possible to verify that the transcriptions were performed, on average, twice as long as the textualizations, being it a statistically significant difference (P=0.01). Such data are important enough to argue that when the time factor is involved, from the time when material quality denoted not being impaired, this procedure can be used without compromising the result. These data are shown in table 1, illustrative chart 3.

The evaluators transcribed on average 12.44 words per minute and textualized 18.79 words per minute. The time for completing textualization was on average half the time needed to perform the transcription. Time is an important factor considering the need to expedite legal proceedings and minimize costs. The need for a faster procedure is real respecting the limitation of human and technical resources made available to the authorities [20]. This study shows that the choice for textualization significantly save time for audio treatments [29,30], i.e., with the time factor being significant, the judiciary power should opt for textualization.

The slower speed in transcription is probably due to the very process that by itself requires all words to be accurately reproduced [31]. Although textualization depends on the skill and knowledge of the textualizer and Portuguese idiom domain, more swiftness was attributed to the fact of this power interpreting the contents by means of context, while keeping the highlights, without the need for literal understanding of all words. Such data corroborate other studies that classify the transcription as a complex process that involves numerous aspects such as conversation, performance time, nonverbal actions, speaker/listener relationship and physical orientation [32,33].

Regarding the maintenance of content highlights between transcription and textualization, this study indicates that, on average, 93% of the highlights were held between the two methods. The difference among the medians as for the number of existing highlights in both methods, was not statistically significant (P=0.61), which confirms the similarity of content between the two processes. These data are shown in table 2 and illustrative Figure 4. Different highlights may cause serious problems of understanding, preventing, sometimes, the establishment of coherence [34]. This study showed no significant difference, demonstrating that there is no damage to the preservation of contents, opting for either procedure.

Whereas transcription and textualization must retain the content of the links, to avoid changing the original message and the findings showing that the central highlight of the messages is maintained, it is possible to emphasize that textualization, due to its execution speed, becomes more feasible in content analysis of intercepted calls, providing the speed of the procedure and performing a greater number of analyses [10,28].

Thus, although many studies seek to understand the automatic speech recognition and transcription programs, the results found in this study show that speech can not be easily analyses because it involves relevant factors, such as linguistic knowledge, practice of their transcriber, time to perform the work and perception of non linguistic signs [35,36].

CONCLUSION

According to this study's results, which aimed to verify coincidences and divergences between audio transcription and textualization, it is concluded that the similarities among the processes of transcription and textualization are related with the median number of words and the maintenance of the median number of content highlights. There was divergence as for the implementation time, significantly lower in textualization. Considering the data obtained in this study, textualization process proved to be the most suitable in audio de-recording.

RESUMO

Objetivo: investigar coincidências e divergências entre transcrição e textualização de áudios, a fim de verificar se há evidências estatísticas que possam servir de subsídio quanto ao melhor procedimento a ser aplicado. **Métodos:** estudo retrospectivo. Foram selecionados aleatoriamente 30 áudios, entre os 239 áudios disponíveis, provenientes de interceptações telefônicas do mesmo processo judicial. Foram considerados: o número de palavras e tempo em minutos para a realização de cada áudio, e a análise comparativa da manutenção dos focos principais de conteúdo. Três Fonoaudiólogos transcreveram e textualizaram terços diferentes do arquivo, garantindo independência. Um Fonoaudiólogo, que não participou da etapa anterior, realizou a análise de conteúdo. Para a análise estatística foi utilizado o teste de Wilcoxon-Mann Whitney no ambiente R, com interface Tinn R. Nível de significância de 5% (0,05). CEP: 274-742. **Resultados:** o número médio de palavras utilizadas na transcrição dos arquivos de áudio foi 27% maior que o número de palavras utilizadas na textualização, p=0,52. A média do tempo em minutos necessários para realizar a transcrição foi significantemente maior p=0,013. Na análise comparativa da manutenção dos focos principais de conteúdo, foi possível verificar que em média 93% dos focos foram mantidos, p=0,61%. **Conclusão:** as semelhanças entre os processos de transcrição e textualização foram com relação ao número mediano de palavras e a manutenção do número mediano de focos de conteúdo. Houve divergência quanto ao tempo para a realização, significantemente menor na textualização. Considerando os dados obtidos neste estudo, o processo de textualização mostrou ser o mais indicado na degravação de áudios.

DESCRITORES: Voz; Fonoaudiologia; Linguagem; Ciências Forenses

REFERENCES

1. Amino K, Sugawara T, Arai T. Effects of the syllable structure on perceptual speaker identification. IEICE Tech. Rep. 2006;105:109-14.

2. De Sordi NAD, Axt G, Fonseca PRP. Manual de procedimentos do programa de história oral da justiça federal. Brasília: Conselho da Justiça Federal. 2007. [acesso em 17 Jul 2013]. Disponível em: http://www2.cjf.jus.br/jspui/bitstream/handle/1234/19/manual%20historia%20oral.pdf?sequence=1

3. Rose P. Forensic Speaker Identification .1st ed. London. Taylor & Francis; 2002.

4. Porto AC, Gonçalves CS. Proposta de análise perceptivo-auditiva de voz e fala para uso em fonética forense. Rev. do IGP. 2007;3(3):23-5.

5. Cabette ELS. Interceptações Telefônicas. 2ª ed. São Paulo: Saraiva; 2011.

6. Bonfim EM. Curso de Processo Penal. 8ª ed. São Paulo: Saraiva; 2013.

7. Tourinho Filho FC. Manual de Processo Penal. 16ª ed. São Paulo: Saraiva; 2013.

8. Oliveira EP. Curso de Processo Penal. 17ª ed. Belo Horizonte: Del Rey; 2013.

9. Flôres O, Silva MR. Da oralidade à escrita: uma busca da mediação multicultural e plurilinguística. Canoas, Rio Grande do Sul: Ed. Ulbra; 2005.

10. Gold E, French P. An international investigation of forensic speaker comparison practices. 17th International Congress of Phonetic Sciences, 2011 Aug 17-21; Hong Kong.

11. Klus K, Trawiñska A. Forensic speaker identification by the linguistic-acustic method in KEÚ and IES. Problems of forensic sciences. 2009;78:160-74.

12. Silva W, Melo MSDS. As estratégias argumentativas em crimes de extorsão: uma análise de discursos patêmicos em golpes de falso sequestro. Rev Diálogo Letras. 2013;2(1):74-96.

13. Barletta JB, Fonseca ALB, Oliveira MIS. Transcrição e observação como estratégias para aprimoramento da competência clínica. Rev Bras Terapia Cognitiva. [periódico na internet]. 2011; [acesso em 04 nov 2013]: 7(2):17-24. Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1808-56872011000200004&lng=pt&nrm=iso>. ISSN 1808-5687.

14. Gago PC. Questões de transcrição em análise da conversa. Veredas (UFJF). 2002;6(2):89-113.

15. Marega LMP, Jung NM. A sobreposição de falas na conversa cotidiana: disputa pela palavra? Veredas on line [periódico na internet]. 2011 [acesso em 17 jul 2013]; 1:321-37. Disponível em: http://www.ufjf.br/revistaveredas/files/2011/05/ARTIGO-231.pdf

16. Oliveira RS, Surreaux LM. Análise da fala sintomática: diferenças entre transcrição fonética e transcrição de base enunciativa. Salão de Iniciação

Científica:Análises discursivas e textuais. 2010 out. 18-22 : UFRGS, Porto Alegre, RS. Porto Alegre: UFRGS. 2010.
17. Santos Junior AG. Adeus à transcrição fonográfica: um estudo de caso. Rev Perícia Federal. 2003;4(16):25-8.
18. Marcuschi LA. Análise da conversação. 5ª ed. São Paulo: Ática; 2003.vv
19. Flores V. Entre o dizer e o mostrar a transcrição como modalidade de enunciação. Rev Organon. 2006;20(40-41):61-75.
20. Prado G. Limite às interceptações telefônicas e a jurisprudência do superior tribunal de justiça. 2ª ed. Rio de Janeiro: Lumen Juris; 2006.
21. Avólio LFT. Provas ilícitas. Interceptações telefônicas, ambientais e gravações clandestinas. 4ª ed. São Paulo: Revista dos Tribunais, 2010.
22. Faria JC, Grosjean P, Jelihovschi E. Tinn-R: GUI/ Editor for R language and environment statistical computing. [software]. 2008. [cited 2013 jul 10]. Available from: http://sourceforge.net/projects/tinn-r
23. Kistenmacher D, Vandresen T. A interceptação telefônica e a garantia constitucional da inadmissibilidade das provas ilícitas. Rev. da Unifebe [periódio na internet]. Dez-Jan 2009 [acesso em 17 Jul 2013];(7):[11p.]. Disponível em: www.unifebe.edu.br/revistaeletronica/2009/artigo029.pdf
24. Almeida AV. Interceptação de comunicações telefônicas no direito militar. TJMPS. 13 Jul 2009. [acesso em 21 mai 2013]. Disponível em: www.tjmps.jus.br/exposições/art012.pdf
25. Romito L, Galatà V. Towards a protocol in speaker recognition analysis. Forensic Sci Int [serial on the Internet]. 2004 Dec 2. [cited 2013 Jul 17]. 146: [about 3 p.]. Available from: www.ncbi.nlm.nih.gov/pubmed/15639553
26. Gomes MLC. Richert LC, Malakoski, J. Identificação de locutor na área forense: a importância da pesquisa interdisciplinar. Encontro do CELSU; 2012 out 24-6. Cascavel: UNIOESTE. 2012
27. Pollack I, Pickett JM, Sumby WH. On the identification of speakers by voice. J Acoust Soc Am. 1954;26(3):403-6.
28. Mondada L. Commentary: transcript variations and the indexicality of transcribing practices. Discourse studies. 2007;9(6):809-21.
29. Tiwari M, Tiwari M. Voice: how humans communicate? J Nat Sci Biol Med. 2012;3(1):3-11.
30. Tosi O. Methods of voice identification for law enforcement agencies. Identification News. 1981;2:235-46.dd
31. Davidson C. Transcription: Imperatives for Qualitative Research. International Journal of Qualitative Methods.[periódio na internet] 2009 [acesso em 4 nov 2013];8(2). Disponível em: http://ejournals.library.ualberta.ca/index.php/IJQM/article/view/4205
32. Bucholtz M. The politics of transcription. Journal of Pragmatics. 1999;32(2000):1439-65.
33. Bucholtz, M. Variation in transcription. Discourse Studies. 2007;9(6):784-808.
34. Koch IV, Travaglia LC. A coerência textual. 14ª ed. São Paulo. Contexto; 2002.
35. King S, Frankel J, Livescu K, McDermott E, Richmond K, Wester M. Speech production knowledge in automatic speech recognition. J Acoust Soc Am. 2007;121(2):723-42.
36. Harnsberger JD, Hollien H, Martin C, Hollien KA. Stress and Deception in Speech: Evaluating Layered Voice Analysis. J Forensic Sci. 2009;54(3):642-50.

Received on: July 29, 2013
Accepted on: April 01, 2014

Mailing address:
Maria Inês Beltrati Cornacchioni Rehder
Avenida 15, 531. Rio Claro/SP
CEP: 13500-330
E-mail: mariainesrehder@uol.com.br