**Original Article**
*Artigo Original*

Marina Englert[1]
Glaucya Madazio[2]
Ingrid Gielow[2]
Jorge Lucero[3]
Mara Behlau[1,2]

# Learning factor influence on the perceptual-auditory analysis

## *Influência do fator de aprendizagem na análise perceptivo-auditiva*

## ABSTRACT

**Purpose:** To investigate the learning factor during a perceptual-auditory analysis of an unusual task in three different groups. **Methods:** 269 listeners, divided into three groups: 73 voice specialists Speech Language Pathologists (EG), 84 voice specialists Speech Language Pathologists (NEG); and 112 non-speech pathologists in the Naive Group (NG). They all completed a listening session that included 18 synthesized and 18 human voices with different types and degrees of deviation (50% of repetition for intra-rater consistency analysis). The task was to classify the voices as human or synthesized. We analyzed the learning factor by comparing the initial error percentage, first 18 voices, with the final, last 18 voices. **Results:** EG presented less error towards the end of the task (25.5%) than at the beginning (28.6%) with statistical difference (p = 0.024). The error percentage of the beginning and the end of the task did not differ for the NEG and the NG (NEG beginning = 36.5%, end = 35.3%; NG beginning = 38.3%, end = 37.7%). **Conclusion:** The EG was the only group to present evidence of learning factor. Therefore, it seems that professional experience positively influences the perceptual-auditory analysis, which reinforces the impact of its training to become a voice specialist. Moreover, the voice specialists seem to be more prepared and more susceptible to use learning strategies to improve their performance during a perceptual-auditory analysis task, even if unusual.

## RESUMO

**Objetivo:** Investigar o fator de aprendizagem durante uma tarefa perceptivo-auditiva para três grupos diferentes em uma tarefa não usual. **Método:** 269 ouvintes, divididos em três grupos: 73 no grupo dos fonoaudiólogos especialistas em voz (GE), 84 no grupo dos fonoaudiólogos não especialistas em voz (GNE) e 112 no grupo leigo (GL), dos não fonoaudiólogos. Todos foram submetidos a uma sessão de escuta que incluiu 18 vozes humanas e 18 vozes sintetizadas com diferentes tipos e graus de desvio, mais 50% de repetição para avaliar a consistência intraindivíduo. A tarefa era classificar as vozes como humana ou sintetizada. Analisou-se o fator de aprendizagem pela comparação da porcentagem de erros do começo, primeiras 18 vozes, e do final, últimas 18 vozes, da sessão de escuta. **Resultados:** O GE foi submetido ao fator de aprendizagem, apresentando menos erros no final da tarefa (25,5%), do que no começo (28,6%), com diferença estatística (p = 0,024). O GNE e o GL não apresentaram diferença da porcentagem de erros no começo e no final da tarefa (GNE começo = 36,5%; GNE final = 35,3%; GL começo = 38,3%; GL final = 37,7%). **Conclusão:** O GE foi o único grupo que apresentou indícios evidentes do fator de aprendizagem. Parece que a experiência profissional influencia de modo positivo a análise perceptivo-auditiva, reforçando o impacto de um treinamento para se tornar um especialista em voz. Ainda, o especialista em voz parece estar mais preparado e mais suscetível a utilizar estratégias de aprendizagem para melhorar sua *performance* durante uma tarefa perceptivo-auditiva mesmo que pouco usual.

**Correspondence address:**
Marina Englert
Universidade Federal de São Paulo – Unifesp
Rua Machado Bittencourt, 361/1001, Vila Mariana, São Paulo (SP), Brasil, CEP: 04044-001.
E-mail: marinaenglert@gmail.com

## INTRODUCTION

The perceptual-auditory analysis is considered the gold standard for the vocal evaluations[1], and it is used for diagnosis, measurement of treatment results and to evaluate immediate effects of vocal exercise[2-4]. However, it is considered to be a subjective assessment, which results in poor inter-rater reliability and inconsistency in the vocal evaluations[1,5].

Listeners with similar auditory experiences, that is, with similar professional activities and backgrounds, present a better inter-rater reliability[6]. Thus, it is necessary to establish basic specifications and/or provide training before any vocal evaluation in order to reduce this common inter-rater variability. Furthermore, evaluators with auditory training and relevant clinical experience have better intra-rater reliability. Therefore, previous training is essential to guarantee better outcomes and more consistent answers on the perceptual-auditory analysis[1,7,8].

It is known that learning occurs when a new information is associated with previously learned and relevant concepts. In other words, to facilitate the learning process, the information must have already been presented to the individual and he must have considered it relevant and useful.

Thus, it can be stated that, in order to guarantee the learning factor, it is necessary to first present the new concept and/or information. Subsequently, other related concepts and information must be presented and added to what was posteriorly showed; hence, the individuals will in fact retain and learn that new information[9].

With this in mind, to guarantee a better perceptual-auditory analysis, regardless of the type of vocal stimulus, the more training, the better[9]. Thus, the more the individual is exposed to this activity, with different types and degrees of vocal deviations, the more he will be able to learn about it and from it. Therefore, he will be more prepared to perform a perceptual-auditory analysis, regardless of the target stimulus.

Even without a full understanding of the learning factor concept related to training, commonly, the perceptual-auditory analysis uses training as a strategy to overcome its lack of reliability[10]. Also, training is considered extremely important for professionals who will work with the human voice[11,12]. Thus, experienced professionals in this work field, that is, voice specialists with a specific training for this task and relevant clinical experience possibly will have a better reliability and greater skills in the analysis of different stimuli. They will be able to add to their experience, information and concepts to which they were presented previously, possibly during their study to become a voice specialist.

Therefore, the aim of the present study was to investigate the learning factor during a perceptual-auditory analysis for three different groups on an unusual task.

## METHODS

This retrospective study performed a new analysis on the data of a previous research that was approved by the Committee for Ethics in Research under the protocol number 1.281.837[13]; all participants signed the informed consent form.

This new analysis focused on other issues not addressed in the previous study that aimed to evaluate the quality of synthesized voices[13]. The present study aimed to analyze the learning factor in the perceptual-auditory analysis of listeners with different levels of perceptive-auditory training and backgrounds.

The stimuli used for this research analysis were 18 human and 18 synthesized voices of the Brazilian vowel "ae", sustained for 1 second. The listening session counted also with 50% of random repetition to verify the intra-rater reliability.

The human voices were selected from a voice bank of a vocal clinic that has over 1000 controlled stimuli of patients. The voices' selection was performed by convenience of three voice specialist speech-language pathologists in order to represent different types (roughness, breathiness and strain) and degrees (mild, moderate and severe) of vocal deviation. Although very common at the vocal clinic, voices with combined characteristics, such as roughness and breathiness, or roughness and strain, were avoided due to the complexity of this type of stimuli. Finally, the human set of voices was composed by six voices for each vocal type, three for each gender, with three degrees of vocal deviation, totalizing 18 stimuli. All voices were of Brazilian adult patients.

The set of synthesized voices was selected from a voice bank with over 200 stimuli, previously produced by a physics-based synthesizer, the VoiceSim. The synthesizer contains a representation of the vocal tract in the form of concatenated tubes through which an acoustic wave propagates; it includes model of trachea, nasal cavity and paranasal sinuses. The vocal tract that was used represented the Brazilian vowel "ae"; the vocal tract excitation was generated by a vocal fold model that considers its movement as a superficial wave that propagates through the mucosa, induced by the air flow.

In order to produce voices with different vocal deviations, different acoustic parameters were manipulated, such as: jitter, signal-to-noise energy ratio, glottal area at the prephonatory position, stiffness coefficient and fundamental frequency[13].

The same three voice specialists speech language pathologists who selected the human voices, also selected the synthesized voices. These produced voices had to be in accordance and paired with the vocal type and degree of deviation of the previously selected human voices. Therefore, a total of 18 synthesized stimuli, nine male and nine female, were selected.

A total of 269 listeners participated in this study; they were all adults more than 18 years old who were invited to participate. The listeners were divided into three groups according to their auditory experience based on their professional background and voice experience.

The Experienced Group (EG) was formed by 73 voice specialists speech language pathologists (5 men and 68 women), with an average of 11.5 years in the profession and 35.3 years old. The Non-Experienced Group (NEG) was composed by 84 non-voice specialists' speech language pathologists (3 men and 81 women), with an average of 6.2 years in the profession and 29.5 years old. The third group was the Naive Group (NG), with 112 listeners that were not speech-language pathologists and with an average age of 32.4 years old.

The perceptual-auditory task for all the 269 listeners was to classify 54 voices as being human or synthesized. This same

voice bank and perceptual-auditory task was used in another study that focused on different analysis[13]. To better address the aim of the present study - analyze the learning factor during a perceptual-auditory analysis for three different groups on an unusual task - the groups were renamed.

In the interest of verifying the consistency of the listeners, a screening was performed. The listeners were considered consistent if at least 13 of the 18 repeated voices had the same answer, that is, if the voice 2 was the repetition of the voice 37, both should have the same response, despite of being right or wrong. Therefore, the listeners should have a percentage of equal answers, of consistency, of at least 72.2%.

The responses of 99 listeners were inconsistent and therefore, excluded from the final analysis. The EG, the group of voice specialists, thus, the group with greater perceptual-auditory experience, had less excluded listeners due to the lack of consistency (20.5%; p <0.05 – Student's t-Test for paired samples). The NEG, the group of non-voice specialists had 39.2% of excluded listeners and the NG, non-speech language pathologists, had 45.5% of excluded listeners (EG vs NEG p = 0.011; EG vs NG p <0.001; NEG vs NG p = 0.382). Thus, the non-voice specialists speech language pathologists and the naive listeners were equally inconsistent and thus excluded from the final analysis.

The final analysis counted with 170 individuals: 58 from the EG (2 men and 56 women, with an average of 11.7 years in the profession and 34.9 years old); 51 from the NEG (3 men and 28 women, with an average of 5.4 years in the profession and 30.3 years old) and 61 from the NG (31 men and 30 women, with an average of 30.3 years old). The Cohen Kappa Coefficient analysis of these 170 listeners presented regular values for all groups (EG = 0.523; NEG = 0.595 and NG = 0.592)

The analysis of the learning factor considered the responses of each group. For this analysis, the occurrence of intra-rater errors at the beginning of the task (the first 18 voices at the Beginning) was compared with the occurrence of errors at the end of the task (the last 18 voices at the end). An error was considered whenever a listener classified a human voice as being synthesized, and vice versa.

It had already been observed[13] that the individuals from the so-called non-experienced and naive groups presented more errors in classifying human and synthesized voices. However, no analysis was performed regarding how these individuals improve or not their evaluation, neither what elements and strategies the so-called experienced group might have used in order to present a better outcome in the identification of human and synthesized voices. Thus, this study analyzed the occurrence of error at the beginning and at the end of the perceptual-auditory analysis, in order to understand if there was a learning factor for each one of the groups, especially for the group of voice-specialists.

## RESULTS

The learning factor was observed only for the EG; the voice specialists had less errors at the end of the perceptual-auditory analysis than in the beginning (Table 1).

**Table 1.** Learning factor for each group

|  | Error % | p-value |
| --- | --- | --- |
| EG |  |  |
| Beginning | 28.6 | 0.024* |
| End | 25.5 |  |
|  |  |  |
| NEG |  |  |
| Beginning | 36.5 | 0.363 |
| End | 35.3 |  |
|  |  |  |
| NG |  |  |
| Beginning | 38.3 | 0.649 |
| End | 37.7 |  |
|  |  |  |
| Total |  |  |
| Beginning | 34.5 | 0.036* |
| End | 32.8 |  |

*Significance level (p≤0.050) - Student's t-Test for paired samples; EG = Experienced Group; NEG = Non-Experienced-Group; NG = Naive Group; Beginning = first 18 voices; End = last 18 voices

## DISCUSSION

The vocal assessment is a perceptual phenomenon[14] influenced by both the previous training and the listeners experience[1,7,8]. Strategies such as training and anchor stimuli are used in order to guarantee a higher reliability of vocal evaluations[10,15].

Listeners with previous training showed advantages on the screening performed in this research, once they had less excluded listeners due to inconsistent answers. Therefore, it seems that the EG are able to maintain consistency probably due to internal standards[15,16] formed during their study to become a voice specialist. The NEG and the NG made more random decisions.

The learning factor analysis considered the error percentage of the first 18 stimuli presented on the task and compared them with the error percentage of the last 18 stimuli. The EG presented less error at the end of the task (Table 1); hence, listeners with some type of experience probably presented the learning factor. The learning factor was also observed for the Total group; perhaps, the positive outcome of the EG affected the total analysis, once the NEG and NG did not present this result (Table 1).

The identification of human and synthesized voices is quite unusual and it is not part of the speech language pathologist clinic routine; despite of this, the voice specialists were able to complete the task with a better performance. They presented fewer mistakes in the identification of human and synthesized voices when compared to the other listeners[13]; plus, they seem to have presented the learning factor, because they had fewer errors at the end of the listening session. In order to complete the unusual task that this research proposed, it is possible that the voice specialist used learning strategies related to the perceptual-auditory analysis – that were certainly learned during their study to become a voice specialist. With this in mind, it may be suggested that the voice specialists were self-regulated learners.

The self-regulated learning depends of motivation, cognitive and metacognitive factors[17-19]. Moreover, the use of learning strategies demands effort; thus, they are only used when the information is considered valid by the individual[19]. Thus, it may be suggested that only the EG used learning strategies once they are the group that understands the importance and relevance of a perceptual-auditory analysis; therefore, they put more effort to perform the task.

In addition, it is known that learning depends on repetition, motivation and emotion[20-22]. One of the motivation strategies is the intention to learn[17]. The voice specialists had motivation; they wanted to study to become a voice specialist. Also, they had experience, which brings repetition that increases the intrinsic redundancy to build auditory standards; therefore, they have a more robust internal standard system to identify voices[15,16].

The intrinsic redundancy involves the acoustic signal that reaches the brain including structures related to its neural pathways and conversion to neural code[23]. The neural connections will become stronger with more stimulation of the auditory pathway, and therefore, more redundancy. Such redundancy guarantees to the voice specialists more strategies to learn during the task and to search for auditory memories that will improve their performance, as observed in this research.

There are several studies about learning styles theories; they affirm that each person learns in a different way and that to guarantee a better performance each personal style must be considered. However, the studies that seek to prove such theory have failed methodologies or do not actually prove its existence[24].

On the other hand, regardless of the learning style, it is known that in order to learn tasks related to, for example, geometry, the best training is with visuospatial activities[24]. Thus, in order to learn how to perform a perceptual-auditory analysis, the best training is with auditory activities, that is, listening to voices.

No type of training or task-specific instruction was given for this study task. The listeners should simply classify the voices as being produced by a human or a synthesizer, without further information and without any previous specific training.

It is known that students with high abilities have better results with less structured instructions when compared to student with low abilities; also, students with low abilities present better performance with more structured instructions[25]. It is worth mentioning that self-regulated learners are more independent to perform a task[17]. With this in mind, it seems that the voice specialists are more prepared and once again it may be mentioned that they are self-regulated learners.

As previously mentioned, this research did not offer any training or instruction regarding how to perform the task. It is possible that if it had done so, the groups of non-voice specialists and the naive listeners could also have learned during the task and also presented the learning factor. This hypothesis should be tested, once it may also provide data regarding the minimum amount of training time to consider a listener as able to perform a valid perceptual-auditory analysis; this training may even be included into the voice specialization courses. Anyhow, it is clear that the auditory training is essential, since it provides learning strategies to ensure a better perceptual-auditory analysis. Further,

when performed by someone without previous training, such analysis is less reliable.

These research data reinforce the importance of performing a perceptive-auditory training for all voice specialists or professionals aiming to work with vocal analysis; previous training gives the evaluator the possibility to use learning strategies and greater cognitive flexibility to identify auditory challenging tasks. Such strategies make it easier for the evaluator to learn and to improve his performance during the task, even though if it is unusual, new and considered hard by most of the listeners.

## CONCLUSION

The voice specialists' speech language pathologists, the experienced group, were the only group that presented the learning factor. Therefore, it seems that the professional experience positively influences the perceptual-auditory analysis, which reinforces the impact of its training to become a voice specialist. In addition, the voice specialists seem to be more prepared for the task and to benefit from the task itself; thus, they use learning strategies and have a better performance, even for an unusual task.

## REFERENCES

1. Oates J. Auditory-perceptual evaluation of disordered vocal quality: pros, cons and future directions. Folia Phoniatr Logop. 2009;61(1):49-56. http://dx.doi.org/10.1159/000200768. PMid:19204393.

2. Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS. Comparing internal and external standards in voice quality judgments. J Speech Hear Res. 1993;36(1):14-20. http://dx.doi.org/10.1044/jshr.3601.14. PMid:8450655.

3. Yiu EM, Murdoch B, Hird K, Lau P, Ho EM. Cultural and language diferences in voice quality perception: a preliminar investigation using synthesized signals. Folia Phoniatr Logop. 2008;60(3):107-19. http://dx.doi.org/10.1159/000119746. PMid:18332632.

4. Roy N, Barkmeier-Kraemer J, Eadie T, Sivasankar MP, Mehta D, Paul D, et al. Evidence-based clinical voice assessment: a systematic review. Am J Speech Lang Pathol. 2013;22(2):212-26. http://dx.doi.org/10.1044/1058-0360(2012/12-0014). PMid:23184134.

5. Sellars C, Stanton AE, McConnachie A, Dunnet CP, Chapman LM, Bucknall CE, et al. Reliability of perceptions of voice quality: evidence from a problem asthma clinic population. J Laryngol Otol. 2009;123(7):755-63. http://dx.doi.org/10.1017/S0022215109004605. PMid:19250586.

6. Bele I. Reliability in perceptual analysis of voice quality. J Voice. 2005;19(4):555-73. http://dx.doi.org/10.1016/j.jvoice.2004.08.008. PMid:16301102.

7. Awan SN, Roy N. Acoustic prediction of voice type in women with functional dysphonia. J Voice. 2005;19(2):268-82. http://dx.doi.org/10.1016/j.jvoice.2004.03.005. PMid:15907441.

8. Kreiman J, Gerratt BR, Ito M. When and why listeners disagree in voice quality assessment tasks. J Acoust Soc Am. 2007;122(4):2354-64. http://dx.doi.org/10.1121/1.2770547. PMid:17902870.

9. Ausubel DP, Novak JD, Hanesian H. Psicologia educacional. Rio de Janeiro: Interamericana; 1980.

10. Chan KM, Yiu EM. A comparison of two perceptual voice evaluation training programs for naive listeners. J Voice. 2006;20(2):229-41. http://dx.doi.org/10.1016/j.jvoice.2005.03.007. PMid:16139475.

11. Awan SN, Lawson LL. The effect of anchor modality on the reliability of vocal severity ratings. J Voice. 2009;23(3):341-52. http://dx.doi.org/10.1016/j.jvoice.2007.10.006. PMid:18346869.

12. Solomon NP, Helou LB, Stojadinovic A. Clinical versus laboratory ratings of voice using the CAPE-V. J Voice. 2011;25(1):e7-14. http://dx.doi.org/10.1016/j.jvoice.2009.10.007. PMid:20430573.

13. Englert M, Madazio G, Gielow I, Lucero J, Behlau M. Perceptual error analysis of human and synthesized voices. J Voice. 2017;31(4):516.e5-18. http://dx.doi.org/10.1016/j.jvoice.2016.12.015. PMid:28089485.

14. Patel S, Shrivastav R. Perception of dysphonic vocal quality: some thoughts and research update. Perspect Voice Voice Disord. 2007;17(2):3-6. http://dx.doi.org/10.1044/vvd17.2.3.

15. Chan KM, Yiu EM. The effects of anchors and training on the reliability of perceptual voice evaluation. J Speech Lang Hear Res. 2002;45(1):111-26. http://dx.doi.org/10.1044/1092-4388(2002/009). PMid:14748643.

16. Selleck MA, Sataloff RT. The impact of auditory system on phonation: a review. J Voice. 2014;28(6):688-93. http://dx.doi.org/10.1016/j.jvoice.2014.03.018. PMid:24962229.

17. Boekaerts M. Self-regulated learning at the junction of cognition and motivation. Eur Psychol. 1996;1(2):100-12. http://dx.doi.org/10.1027/1016-9040.1.2.100.

18. McWhaw K, Abrami PC. Student goal orientation and interest: effects on students' use of self-regulated learning strategies. Contemp Educ Psychol. 2001;26(3):311-29. http://dx.doi.org/10.1006/ceps.2000.1054. PMid:11414723.

19. Souza LFNI. Estratégias de aprendizagem e fatores motivacionais relacionados. Educar Ed UFPR. 2010;36:95-107.

20. Eccles J, Wigfield A. Motivational beliefs, values and goals. Annu Rev Psychol. 2002;53(1):109-32. http://dx.doi.org/10.1146/annurev.psych.53.100901.135153. PMid:11752481.

21. Cardoso L, Bzuneck JA. Motivação no ensino superior: metas de realização e estratégias de aprendizagem. Psicol Esc Educ. 2004;8(2):145-55. http://dx.doi.org/10.1590/S1413-85572004000200003.

22. Pally R. The predicting brain: unconscious repetition, conscious reflection and therapeutic change. Int J Psychoanal. 2007;88(4):861-81. http://dx.doi.org/10.1516/B328-8P54-2870-P703. PMid:17681897.

23. Ferre JM. Processing Power: a guide to CAPD assessment and management. San Antonio: Communication Skill Builder; 1997. 186 p.

24. Pashler H, McDaniel M, Rohrer D, Bjork R. Learning styles: concepts and evidence. Psychol Sci Public Interest. 2008;9(3):105-19. http://dx.doi.org/10.1111/j.1539-6053.2009.01038.x. PMid:26162104.

25. Freebody P, Tirre WC. Achievement outcomes of two Reading programmes: an instance of aptitude-treatment interaction. Br J Educ Psychol. 1985;55(1):53-60. http://dx.doi.org/10.1111/j.2044-8279.1985.tb02606.x.

## Author contributions

*ME was responsible for the study design, data collection and analysis and writing of the manuscript; GM was responsible for the study design, data analysis and the manuscript revision and correction; IG was responsible for the study design, data analysis and the manuscript revision and correction; JL was responsible for the study design, data analysis, manuscript revision and production of the synthesizes voices; MB was responsible for the study proposal and design, data analyses e final revision of the manuscript.*