EPiC Series in Computing

Volume 81, 2022, Pages 303–312

Proceedings of 11th International Congress on Advanced Applied Informatics

On the interplay between vocal production effect and learning content types in e-learning settings

Kazuma Ohta[1], Zeynep Yücel[1], Parisa Supitayakul[1], Akito Monden[1], and Pattara Leelaprute[2]

[1] Okayama University, Department of Information Science, Faculty of Engineering, Okayama, Japan
pha33dk8@s.okayama-u.ac.jp, zeynep@okayama-u.ac.jp, pgw45ydd@s.okayama-u.ac.jp, monden@okayama-u.ac.jp

[2] Department of Computer Engineering, Faculty of Engineering, Kasetsart University, Bangkok, Thailand
pattara.l@ku.ac.th

Abstract

The concept of "production effect" from experimental psychology suggests that producing a word aloud during study improves explicit memory as compared to reading the word silently. In this study, we investigate the effect of the different vocal production behaviors on recollection rates concerning varying content types delivered through an e-learning platform and inquire whether there is any possibility of improving the e-learning system by integrating vocal production instructions. As for different sorts of vocal production behaviors, we considered as the usual depiction (*uttering*) as well as lack (*free view*) and suppression (*mouthing*) of vocal production. As for content types, one numerical content and two verbal contents with varying levels of pronunciation difficulty are considered. Our results indicate that there is no statistically significant difference on recollection rates between various vocal production behaviors. However, it is observed that by *uttering*, the content which is relatively harder to pronounce, can be recalled better than the others in a statistically significant way. This unexpected result indicates that there is a potential to increase the performance of learners, who study unfamiliar verbal content (e.g. foreign vocabulary) by integrating vocal production into e-learning systems.

1 Introduction and motivation

Learning or practicing a subject using smart devices such as phone or tablet PC is becoming increasingly popular among students. These multimedia learning systems (henceforth referred to as *e-learning systems*) come in a large variety (e.g. online or offline, individual access or collaborative activity etc.) and have vast advantages (e.g. economical, flexible in time, diverse in content) [1, 2].

The platforms hosting such e-learning systems are equipped with various features, which can potentially help the learner and improve his memory registration. However, most systems rely on only visual stimuli and do not receive feedback from the users, apart from subjective ratings. In that respect, in interaction with the system the role of the learner is rather *passive*.

In other words, he/she is exposed to certain changes in the interface but he/she often does not need to take any action or react in a certain specified way.

In that respect, this study focuses on *active* involvement of the user in the e-lerning process. Specifically, we consider interacting with the e-learning system through uttering (reading aloud) of the content. The reason for considering such an interaction is based on the implications of the concept of the so-called *production effect* from experimental psychology [3]. The production effect refers to the fact that producing a word aloud during study, relative to simply reading that word silently, improves explicit memory.

In that respect, the e-learning system can (i) detect the likeliness of learner to forget a certain piece of the content or disengage [4, 5], and then (ii) it may require the learner to read it aloud before proceeding to future pieces. For the former, it is possible to use the feedback from the learner or to carry out an estimation based on behavioral indicators (e.g. reaction time, eye gaze, facial landmarks etc.) [5]. In this study, we focus on the latter part, i.e. interaction between the learner and the e-learning platform through vocal production. In doing that, we consider different sorts of learning content as explained in Section 3.3 and vocal production behaviors described in Section 3.2. We investigate whether (or how much) recollection rate relating to various learning improves under different vocal production behavior in a statistical way.

2 Background and related work

One of the first studies on the relation between recollection of material that is simply visually presented as compared to material *produced* by the subjects, was carried out by Slamecka and Graf [6]. They observed that the material produced is better recalled than the material merely viewed. They called this phenomenon the "generation effect", which led to a vast amount of studies on priming strategies or manipulations of subjects' behavior or stimuli in relation to memory retention.

In this article, we focus on a certain kind of manipulation, namely subjects' *vocal production* of the material. As briefly mentioned in Section 1, the *production effect*, refers to the fact that producing a word aloud during study, relative to simply reading the word silently, improves explicit memory[1][2]. In most cases, the assessment of memory retention is based on free recall, explicit recognition test, source identification or speeded reading test[3].

The mechanisms, that are hypothesized to be underlying the production effect are termed in the literature as *accounts*. Several popular accounts include decision-based account, memory-based account [7], strength account [8], distinctiveness account [9] and attributional account [10]. In order to understand the validity of these hypotheses or the extent of their effects, it is common to deliberately induce *negative* production effect or to expose the subjects to several disruptions, which are anticipated to interfere with or eliminate the accounts.

In addition, common vocal production (read-out) has been contrasted to different means of information registration such as reading silently, mouthing, whispering, spelling, hearing, writing, typing, and even singing [11, 12, 13]. It is shown that vocal production is superior against all of these different ways, although also some alternatives such as mouthing and whispering are

[1] In this study, we prefer using the term *vocal production effect*, since we consider it to be less ambiguous for our audience.

[2] In that respect, production effect can be considered as a specific case of generation effect.

[3] Free recall refers to subject's listing of the items in the presented task. Explicit recognition refers to recognition of the items among a set involving also distracters. Source identification refers to attributing the items to one of the several (usually two) sets of items (tasks). Speeded reading is the subjects' reading into a microphone a mixed list of items, which is then analyzed to detect the changes in his/her reading pattern.

also seen to have a positive effect on memory to a certain extent [14]. This is considered to be due to the presence of both articulation and audition components in speech, whereas audition is absent in mouthing and extremely limited in whispering. Moreover, writing and typing are also beneficial, whereas spelling probably suffers from the reuse of letters across words.

In addition to its performance as compared to the above-mentioned alternatives, several other specifics of vocal production in relation to the nature of the presented items are explored such as being a meaningful word or not [15], its relation to additional visual stimuli [16], involvement of multiple speakers [17] as well as participant profile (e.g. age or medical disorders) [18, 19, 20] and additional background sounds(e.g. steady-state energetic or fluctuating-informational noise etc.) [21] .

In addition, the benefits of production effect are explored in relation to education and learning [22] and it is shown to be a viable encoding strategy for educational material, due to a lasting effect and extension beyond isolated words (i.e. it applies also to word pairs and sentences). Nevertheless, apart from the delayed testing and diversification of content, most studies addressing educational use of production effect basically address typical cross-subject cross-content laboratory experiments for measuring memory retention. In that respect, very few studies actually correspond to realistic classroom or multimedia-learning settings. But there is still promising evidence that production effect can be used to enhance memory in real-world educational scenarios [15].

3 Experiments

In this section, we will elaborate on participant profiles, task content, and different kinds of vocal production behavior, as well as data recording and memory tests.

3.1 Participant profile

We performed a set experiments for investigating the relation between participants' recollection rate and different sorts of vocal production behavior and learning content. To that end, we recruited 6 participants (1 females and 5 males). The participants are fourth year undergraduate students in various departments of our university and are all mother tongue Japanese speakers. They were informed in a clear manner about the nature and method of the research, volunteered to participate in the experiments, gave their written permission for participation and data recording.

3.2 Vocal production behaviors

The stimulus conveying a piece of information (i.e. visual or audio-visual) is known to make a distinguishing effect on its cognitive registration and recollection rate [19]. Similarly, vocal production effect is proven to make a positive effect on memory and recollection as mentioned in [3]. However, one may sometimes feel uncomfortable to read aloud and prefer either to whisper or to pretend to read-out (i.e. move the lips without uttering). In that respect, we consider the following three sorts of behaviors:

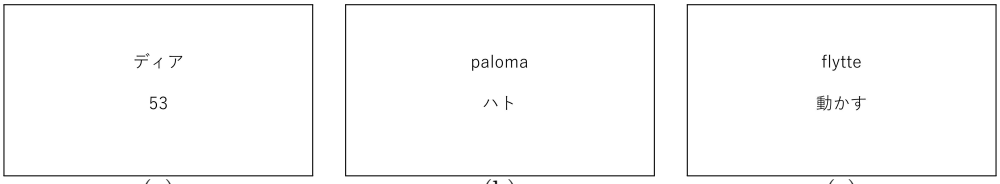
(i) No vocal production (henceforth, referred to as *free view*)

(ii) Moving the lips without pronouncing the text (i.e. suppressed vocal production, henceforth referred to as *mouthing*)

(iii) Reading the text aloud (henceforth, referred to as *uttering*)

3.3 Learning content

As for the learning content, we consider three cases as (i) word-number associations, (ii) word-word associations with a high level of anticipated pronunciation difficulty and (iii) word-word associations with a low level of anticipated pronunciation difficulty.



(a) (b) (c)

Figure 1: Sample stimuli regarding tasks for (a) memorization of word-number associations, and (b), (c) word-word associations from Spanish and Danish to Japanese.

(i) The word-number associations contain official names of moons of Jupiter and their labels[4]. Henceforth, we refer to this learning content as *Num*.

(ii) The word-word associations with a high level of anticipated pronunciation difficulty contain a word (noun or verb) in Japanese (kana or kanji), which is the mother tongue of all participants, and a word in Danish. Henceforth, we refer to this learning content as *Dk*.

(iii) The word-word associations with a low level of anticipated pronunciation difficulty contain a word (noun or verb) in Japanese (kana or kanji) and a word in Spanish. Henceforth, we refer to this learning content as *Es*.

Here, Spanish is anticipated to be easier to pronounce and Danish is anticipated to be harder to pronounce for our subjects, which is later confirmed by subjective evaluations of pronunciation difficulty collected from the participants following the experiment (see Section 4.1).

3.4 Experiment tasks and agenda

A task is composed of watching a slide-show illustrating a *set of images*. Specifically, we prepared one slide-show for each each combination of vocal production behavior and learning content[5]. The slide-shows with the learning content of *Num* (see Figure 1-(a)) contain 6 images each, whereas those with *Dk* or *Es* contain 8 images each. The reason for having different number of images in these sets is due to the scarce number of choices regarding the content type of *Num*[6].

[4]Here, *label* refers to the Roman numeral attributed to each moon in order of their naming. Nevertheless, we replaced the Roman numerals with Arabic numerals, since the participants are more familiar with the latter.

[5]The participants were delivered 3 tasks with the learning content of *Num*. for each of them, they depicted one of the three vocal production behaviors in the order given in Section 3.2. Subsequently, this procedure is repeated first for *Es* and then for *Dk*.

[6]Namely, there are 79 moons of Jupiter but some of them do not have official names (only provisional designations, which are code names involving letters and numbers, e.g. S/2003 J 16). Since we consider only "words" (i.e. official names) in this study, the set of choices is rather limited as compared to Spanish or Danish corpora, which provide an abundant amount of choices.

The participants were asked to view each image for 5 sec. Between each pair of images, they viewed a cross hair fixation target for 1 sec and a blank screen for 0.5 sec for resetting[7]. The tools for delivering the visual stimuli are developed in-house using Python 3.5.2 without any specific dependencies and are considered to be precise enough to provide millisecond resolution.

3.5 Data recording and memory tests

The amount of recalled information was determined through a conventional pen-and-paper memory test. After viewing each set of images, the participants were given a test of paired associated recall. Namely, they were given the list of the words on the top lines of the images (see Figure 1) and asked to fill in the bottom lines (i.e. the associated number or word) as much as they can recall. The test did not have any time limit but the participants were often finished in a few minutes. If the participant recalled the information successfully, we registered the score of that association with a 1, and otherwise with a 0.

In addition, the computer registers the course of the experiment and saves it as a log file, which is composed of image file name and Unix time stamps (in milliseconds) of the instants at which each image is displayed and removed. In addition, we record the upper torso (face, head and shoulders) of the participants using the integrated webcam of the notebook PC, which also displays the slide-shows[8].

4 Methodology and analysis

In this section, we firstly present a statistical analysis on the participants' subjective evaluation of pronunciation difficulty and confirm that it meets the anticipations reported in Section 3.3. Then, we examine the memory test results and investigate whether there is any relation between recollection rate and different kinds of vocal production behavior or learning material.

4.1 Subjective evaluation of difficulty

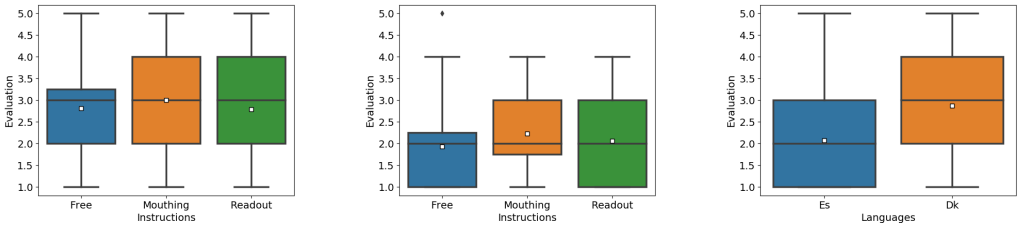
The participants evaluated the difficulty of each foreign language word in the data set on a 5-level Likert scale. Specifically, a label of 5 denotes *difficult to pronounce* and a label of 1 denotes *easy to pronounce*[9]. In order to confirm that there is no significant variation within the sets relating to the same language, that there is a significant difference between the sets relating to different languages and that the latter does not have any dependency on the sort of vocal production behavior, we carried out a statistical analysis on these subjective evaluations. The relating box-plots are presented in Figure 2 and the ANOVA results can be seen in Table 1.

One may see in Figures 2-(a) and (b) that *Dk* content is evaluated on the average with a label around 3, whereas *Es* content receives labels around 2. In these figures, it can be seen that the evaluations do not depend much on the type of vocal production behavior. In addition, these qualitative observations are confirmed through the ANOVA presented in Table 1. Namely, the p-values relating to the *Dk* and *Es* contents with different vocal production behaviors are found as 0.37 and 0.66, respectively, which indicate insignificance. Moreover, it can be seen

[7] In that respect, one set of images is viewed in shorter than 1 min.

[8] Such video data can be used to confirm learners' adherence to the instructions. Moreover, we register the eye gaze of the participants using an eye tracker (the infra-red sensor Tobii 4C operating at 90 Hz). Specifically, it registers the pixel coordinates of the estimated gaze location together with the Unix time stamp in millisecond resolution. Such data can be used as an additional clue for assessing learners' mental/cogntive state (e.g. engaged, fatigued) during a production task [4].

[9] We use the terms "label" and "subjective evaluation" interchangeably.



(a) (b) (c)

Figure 2: (a) Box-plot for subjective evaluation of pronunciation difficulty of *Dk* content with different vocal production behaviors. (b) Similar plot for *Es* content. (c) Box-plot contrasting *Dk* and *Es* contents irrespective of the vocal production behaviors.

in Figure 2-(c), that subjective evaluations vary with respect to the language. This is proven through the ANOVA given in Table 1. Namely, *Dk* and *Es* contents have levels of perceived pronunciation difficulty, which are different[10] in a statistically significant way.

Table 1: ANOVA relating subjective evaluation of pronunciation difficulty.

Fixed	Varying		
Content	Vocal production	F	p
Dk		0.41	0.66
Es		0.99	0.37
Fixed	Varying		
Vocal production	Content	F	p
Free		15.40	$\mathbf{1.65 \cdot 10^{-4}}$
Mouthing		12.13	$\mathbf{7.53 \cdot 10^{-4}}$
Uttering		8.64	$\mathbf{4.12 \cdot 10^{-3}}$
Fixed	Varying	F	p
None	Content	35.39	$\mathbf{7.85 \cdot 10^{-9}}$

After confirming insignificance within *Dk* and *Es* contents, we built two sets of vocabulary pooling the words of the same language presented with different vocal production behaviors into the same set. The relating box-plot is shown in Figure 2-(c) and the p-value associated with it is given at the last row of Table 2 ($p = 7.85 \cdot 10^{-9}$), which confirms that there is a significant difference in the perceived difficulty of pronunciation of the two languages. In order to have a better insight, we also contrast the vocabulary from different languages studied with same sort vocal production behavior and confirm significance (see second block of Table 1)[11].

[10] Here, *difference* refers to *Dk* content being *harder* than *Es* content.
[11] For the sake of brevity, the relating box-plots are skipped.

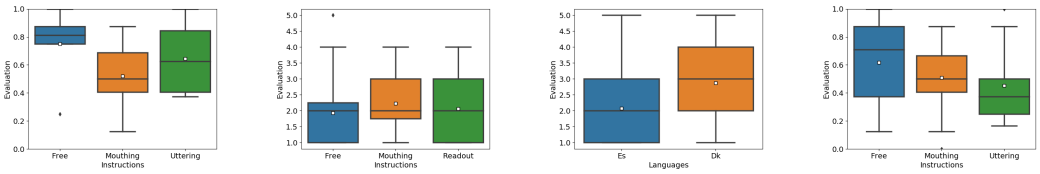
4.2 Memory test results and recollection rate

For identifying any effects on recollection rate induced by vocal production behavior and content type, memory test scores are analyzed with ANOVA. Firstly, we consider the effect vocal production behaviors by focusing on each content type separately. For example, Figure 3 illustrates the box-plots for studying *Dk* of content with varying vocal production behaviors, whereas Figures 4-(b) and (c) display similar results for contents relating to *Es* and *Num*. Moreover, in Figure 4, we care only about the type of vocal production behavior and pool all items from the same content type into the same set.

Here, it can be seen that *free view* has often an advantage over the other two kinds of behaviors. In addition, regarding *Es* content, vocal production behavior of *uttering* introduces an unexpected drawback, although not significant. The F and p-values concerning these figures are given in the first block and the upper row of last block of Table 2. It is understood from these values that different vocal production behaviors do not induce an effect in any statistically significant way, but having more data may help to establish the results more firmly.

Next, the effect of content types is examined in Figure 4 in a similar to way to Figure 3, i.e. at first considering the contents corresponding to each vocal production behavior separately and then pooling the ones carried out with varying sorts of vocal production behaviors into same set.

Here, it is interesting to see that even though *Dk* content is confirmed to be hard to pronounce, it does not suffer from that difficulty in terms of recollection. It is surprising that it surpasses *Num* content, which contains common and well-structured information (at most 2 digit) without any pronunciation issue. Coupled with vocal production behaviors of *free view* and *uttering* (see Figure 4-(a) and (c)), *Dk* content turns out to be recalled more than the other two content types, which is reflected also on the aggregate plot (see Figure 4-(d)). The relating ANOVA results are presented in the second block and last row of Table 2. As expected, the p-values relating the above-mentioned three cases are lower and *free view* and the collective case are regarded to be different in a statistically significant way. Here, it is also worth noting that the collective one attains a lower p-value most probably due to the increase in the number of data points, which indicates that a larger data set is necessary to increase the reliability of these inferences.

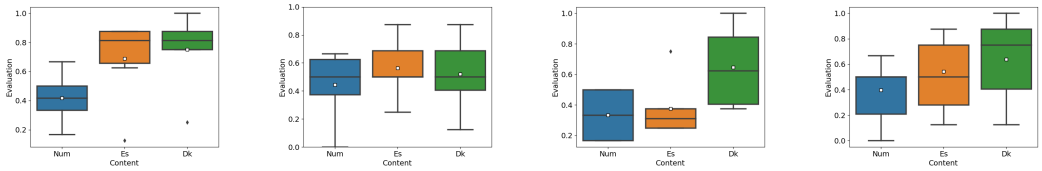


(a) (b) (c) (d)

Figure 3: Box-plots for recollection rates (in %) for varying vocal production behaviors. Content types are (a) Dk, (b) Es, (c) Num. and (d) for any kind of content.

5 Conclusion and future works

In this study, we investigated the effect of the various vocal production behaviors on recollection rates of varying content types. As for different sorts of vocal production behaviors, we considered



(a) (b) (c) (d)

Figure 4: Box-plots for recollection rates (in %) for varying content types with vocal production behaviors of (a) free view, (b) mouthing, (c) uttering and (d) for any kind of vocal production behavior.

Table 2: ANOVA relating recollection rate.

Fixed	Varying		
Content	Vocal production	F	p
Dk		1.12	0.35
Es		2.59	0.10
Num		0.47	0.63
Fixed	Varying		
Vocal production	Content	F	p
Free		3.05	0.07
Mouthing		0.35	0.70
Uttering		3.64	**0.05**
Fixed	Varying	F	p
None	Vocal production	1.98	0.14
None	Content	4.42	**0.01**

the usual depiction (*uttering*) as well as lack (*free view*) and suppression (*mouthing*) of vocal production. As for content types, one numerical content and two verbal contents with varying levels of pronunciation difficulty are considered.

Our results indicate that there is no statistically significant difference on recollection rates between various vocal production behaviors. However, it is observed that with *uttering*, the words which are hard to pronounce (*Dk*) are recalled better than the others in a statistically significant way. This unexpected result indicates that there is a potential to increase the performance of learners, who study unfamiliar verbal content (e.g. foreign vocabulary) by integrating vocal production into e-learning systems. However, if the learners do not follow the instructions properly (e.g. do mouthing rather than uttering) than there will not be any improvement on recollection as shown in Section 4.2. This implies that an efficient integration of these results into the e-learning system requires also confirming the observance of instructions through other means (e.g. speech recognition/processing or facial image processing).

6 Acknowledgement

We would like to thank our anonymous subjects for participating in our experiments.

References

[1] L. Juhaňák, J. Zounek, and L. Rohlíková, "Using process mining to analyze students' quiz-taking behavior patterns in a learning management system," *Computers in Human Behavior*, vol. 92, pp. 496–506, 2019.

[2] A. Hershkovitz and R. Nachmias, "Learning about online learning processes and students' motivation through web usage mining," *Interdisciplinary J. E-Learning and Learning Objects*, vol. 5, no. 1, pp. 197–214, 2009.

[3] C. M. MacLeod, N. Gopie, K. L. Hourihan, K. R. Neary, and J. D. Ozubko, "The production effect: Delineation of a phenomenon," *J. Experimental Psychology: Learning, Memory, and Cognition*, vol. 36, no. 3, p. 671, 2010.

[4] H. V. Willoughby, *The pupillometric production effect: Measuring attentional engagement during a production task*. PhD thesis, Memorial University of Newfoundland, 2020.

[5] Z. Yücel, S. Koyama, A. Monden, and M. Sasakura, "Estimating level of engagement from ocular landmarks," *Int. J. Human–Computer Interaction*, vol. 36, no. 16, pp. 1527–1539, 2020.

[6] N. J. Slamecka and P. Graf, "The generation effect: Delineation of a phenomenon," *J. Experimental Psychology: Human Learning and Memory*, vol. 4, no. 6, p. 592, 1978.

[7] D. P. Mccabe, A. G. Presmanes, C. L. Robertson, and A. D. Smith, "Item-specific processing reduces false memories," *Psychonomic Bulletin & Review*, vol. 11, no. 6, pp. 1074–1079, 2004.

[8] J. D. Ozubko, J. Major, and C. M. MacLeod, "Remembered study mode: Support for the distinctiveness account of the production effect," *Memory*, vol. 22, no. 5, pp. 509–524, 2014.

[9] J. D. Ozubko and C. M. MacLeod, "The production effect in memory: Evidence that distinctiveness underlies the benefit," *J. Experimental Psychology: Learning, Memory, and Cognition*, vol. 36, no. 6, p. 1543, 2010.

[10] G. E. Bodner and A. Taikh, "Reassessing the basis of the production effect in memory," *J. Experimental Psychology: Learning, Memory, and Cognition*, vol. 38, no. 6, p. 1711, 2012.

[11] M. A. Conway and S. E. Gathercole, "Modality and long-term memory," *J. Memory and Language*, vol. 26, no. 3, pp. 341–361, 1987.

[12] S. E. Gathercole and M. A. Conway, "Exploring long-term modality effects: Vocalization leads to best retention," *Memory & Cognition*, vol. 16, no. 2, pp. 110–119, 1988.

[13] C. K. Quinlan and T. L. Taylor, "Mechanisms underlying the production effect for singing," *Canadian J. Experimental Psychology/Revue canadienne de psychologie expérimentale*, vol. 73, no. 4, p. 254, 2019.

[14] N. D. Forrin, C. M. MacLeod, and J. D. Ozubko, "Widening the boundaries of the production effect," *Memory & Cognition*, vol. 40, no. 7, pp. 1046–1055, 2012.

[15] M. Icht and Y. Mama, "The effect of vocal production on vocabulary learning in a second language," *Language Teaching Research*, pp. 1–20, 2019.

[16] M. Icht and Y. Mama, "The production effect in memory: A prominent mnemonic in children," *J. Child Language*, vol. 42, no. 5, pp. 1102–1124, 2015.

[17] D. Knutsen and L. Le Bigot, "Capturing egocentric biases in reference reuse during collaborative dialogue," *Psychonomic Bulletin & Review*, vol. 21, no. 6, pp. 1590–1599, 2014.

[18] O. Y. Lin and C. M. MacLeod, "Aging and the production effect: A test of the distinctiveness account," *Canadian J. Experimental Psychology/Revue canadienne de psychologie expérimentale*, vol. 66, no. 3, p. 212, 2012.

[19] R. Pillai and A. Yathiraj, "Auditory, visual and auditory-visual memory and sequencing performance in typically developing children," *Int. J. Pediatric Otorhinolaryngology*, vol. 100, pp. 23–34, 2017.

[20] M. Icht, O. Bergerzon-Biton, and Y. Mama, "The production effect in adults with dysarthria: Improving long-term verbal memory by vocal production," *Neuropsychological Rehabilitation*, vol. 29, no. 1, pp. 131–143, 2019.

[21] Y. Mama, L. Fostick, and M. Icht, "The impact of different background noises on the production effect," *Acta Psychologica*, vol. 185, pp. 235–242, 2018.

[22] J. D. Ozubko, K. L. Hourihan, and C. M. MacLeod, "Production benefits learning: The production effect endures and improves memory for text," *Memory*, vol. 20, no. 7, pp. 717–727, 2012.