brain sciences

MDPI

Opinion

Perceptual Doping: A Hypothesis on How Early Audiovisual Speech Stimulation Enhances Subsequent Auditory Speech Processing

Shahram Moradi [1,*] and Jerker Rönnberg [2]

[1] Department of Health, Social and Welfare Studies, Faculty of Health and Social Sciences, University of South-Eastern Norway, 3918 Porsgrunn, Norway
[2] Department of Behavioral Sciences and Learning, Linnaeus Centre Head, Linköping University, 581 83 Linköping, Sweden
* Correspondence: shahram.moradi@usn.no

Abstract: Face-to-face communication is one of the most common means of communication in daily life. We benefit from both auditory and visual speech signals that lead to better language understanding. People prefer face-to-face communication when access to auditory speech cues is limited because of background noise in the surrounding environment or in the case of hearing impairment. We demonstrated that an early, short period of exposure to audiovisual speech stimuli facilitates subsequent auditory processing of speech stimuli for correct identification, but early auditory exposure does not. We called this effect "perceptual doping" as an early audiovisual speech stimulation dopes or recalibrates auditory phonological and lexical maps in the mental lexicon in a way that results in better processing of auditory speech signals for correct identification. This short opinion paper provides an overview of perceptual doping and how it differs from similar auditory perceptual aftereffects following exposure to audiovisual speech materials, its underlying cognitive mechanism, and its potential usefulness in the aural rehabilitation of people with hearing difficulties.

Keywords: audiovisual speech training; audio speech training; perceptual doping; hearing loss; auditory speech identification; Ease of Language Understanding Model (the ELU model); cognitive function

Citation: Moradi, S.; Rönnberg, J. Perceptual Doping: A Hypothesis on How Early Audiovisual Speech Stimulation Enhances Subsequent Auditory Speech Processing. *Brain Sci.* **2023**, *13*, 601. https://doi.org/10.3390/brainsci13040601

Academic Editor: Kaisa Tiippana

Received: 8 March 2023
Revised: 27 March 2023
Accepted: 30 March 2023
Published: 1 April 2023

1. Introduction

Language understanding is fundamentally a multisensory process, as face-to-face communication in humans aids language understanding by providing not only verbal cues but also non-verbal cues such as emotional expression, body gesture, and facial expression. The importance of face-to-face interactions is evident in adverse listening conditions when access to verbal cues is limited because of external noise or hearing loss.

Sumby and Pollack [1] were the first to show that the addition of visual speech cues to an auditory speech signal enhances speech intelligibility, especially in noisy conditions. Subsequent studies demonstrated the advantage of audiovisual speech perception over audio-only speech perception when perceiving speech signals in degraded listening conditions [2,3]. Furthermore, more recent studies revealed that the addition of visual speech cues to an auditory speech signal reduces the cognitive demands required to process the speech signals in noisy listening conditions, both for normal hearing listeners [4] and for people with hearing loss [5].

In addition, studies have also found audiovisual aftereffects following exposure to audiovisual speech materials [6,7]. Bertelson et al. [6] found that prior stimulation to a video of a face articulating /ada/ or /aba/ accompanied by an auditory ambiguous sound halfway between /d/ and /b/ (A?Vd or A?Vb) caused a phonetic recalibration effect on subsequent auditory ambiguous sound perception. Those who were first exposed to A?Vd subsequently perceived ambiguous auditory sounds halfway between /d/ and

/b/, often as /d/, and those who were first exposed to A?Vb subsequently perceived ambiguous auditory sounds mainly as /b/. Such a phonetic recalibration effect was not observed after exposure to congruent unambiguous AbVb or AdVd conditions. Bertelson et al. [6] reasoned that early exposure to incongruent A?Vd or A?Vb speech stimuli biased or recalibrated subsequent auditory ambiguous speech tokens in favor of visual components of a prior incongruent audiovisual speech signal.

We found a specific audiovisual facilitation effect on subsequent auditory speech-processing tasks. Moradi et al. [4], in a between-subject experimental study, found that participants exposed to gated audiovisual consonants, words, and sentence-final word identification tasks performed better on a subsequent auditory sentence identification in noise task (Hearing in Noise Test [HINT] [8]) than those exposed only to auditory consonants, words, and sentence-final word identification tasks.

In a randomized control study, Lidestam et al. [9] divided normal hearing participants into three groups: (1) a group exposed to gated audiovisual consonants and words, (2) a group exposed to gated auditory consonants and words, and (3) a control group who only watched a video clip. The HINT scores for each group were recorded before and after exposure to the gated audiovisual stimuli, auditory stimuli, or video clip. Only the group exposed to the gated audiovisual consonants and words subsequently performed better on the HINT, not the other two groups.

Using the same gated audiovisual and auditory gated speech stimuli as employed by Lidestam et al. [9], Moradi et al. [5] studied the efficiency and maintenance of gated audiovisual speech training on auditory HINT performance in elderly hearing-aid users. The results showed that gated audiovisual speech stimulation resulted in better performance on the HINT. Importantly, the audiovisual training effect on HINT performance was maintained after one month.

Finally, using data from the n200 study [10], which comprised 200 hearing-impaired hearing-aid users, Moradi et al. [11] found that prior audiovisual speech stimulation generated larger benefits over auditory speech stimulation in terms of the subsequent processing of auditory speech stimuli for the correct identification of consonants and vowels and the correct discrimination of vowel durations.

We have dubbed this rapid type of perceptual learning "perceptual doping", arguing that even a short exposure to audiovisual speech stimuli recalibrates or retunes phonological and semantic processing maps in semantic long-term memory in a way that facilitates subsequent auditory processing of speech stimuli for correct identification (Moradi et al. [5,11]).

This audiovisual facilitation effect on subsequent auditory speech processing cannot be explained by concepts like perceptual learning only or using lexical knowledge to learn how to categorize speech sounds [12]. Perceptual learning reflects enhanced performance in a task achieved via repeated stimulation of that task. In Lidestam et al. [9] and in Moradi et al. [4,5], the enhancement effect on auditory sentence-in-noise identification was only observed after exposure to audiovisual speech materials and not after auditory stimulation alone. In addition, the materials used in prior audiovisual speech exposure were consonants and words, while the outcome auditory task was sentence-in-noise identification. Furthermore, the talkers in the prior audiovisual exposure and subsequent auditory sentence-in-noise identification tasks were different. Norris et al. [12] showed that listeners benefit from their lexical knowledge in the perceptual learning process to interpret ambiguous speech sounds. Assuming a perceptual doping notion, the recalibrated phonological and lexical maps following exposure to a congruent audiovisual speech signal help listeners to better identify subsequent auditory speech signals. So, the benefit from existing lexical knowledge for subsequent auditory speech identification is not the main point in the perceptual doping notion. We did not examine the extent to which lexical knowledge impacts the benefit provided by prior audiovisual speech stimulation on the subsequent processing of auditory speech signal for correct identification. Van Linden and Vroomen [7] showed that both visual speech cues and lexical knowledge play a similar

role in the phonetic recalibration effect. Future studies are needed to evaluate the benefit provided by lexical knowledge and prior audiovisual speech stimulation on the subsequent processing of auditory speech signals. Further, the perceptual doping notion also means that the locus of the effect cannot be ascribed to repetition priming or "pop out" ([13]). We argued that the better task performance in subsequent auditory sentence-in-noise identification was merely the result of prior exposure to audiovisual speech materials that retuned phonological and lexical maps more distinct, with sharp boundaries, and easily accessible, consequently easing the subsequent mapping of auditory speech input onto phonological and lexical items during sentence-in-noise identification.

Here, we reason that the perceptual doping notion differs from the audiovisual phonetic recalibration effect. First, the audiovisual speech materials in our prior research were different from the work by Bertelson et al. [6]. The audiovisual speech stimuli in Moradi et al. [4,5] and Lidestam et al. [9] were congruent but degraded by a speech signal (background noise) that was presented to the participants in a gating format. In Moradi et al. [11], speech items were auditory and audiovisual consonants and vowels that were presented to the participants in silence in a gating format. A facilitation effect (i.e., perceptual doping) was observed only after exposure to audiovisual speech items but not to auditory ones. Second, theoretical assumptions of the perceptual doping notion are different from the audiovisual phonetic recalibration. According to the perceptual doping hypothesis, simple exposure to audiovisual speech stimuli retunes phonological and lexical maps that subsequently facilitate auditory processing of speech signals for correct identification. On the other hand, audiovisual phonetic recalibration assumes a recalibration of an existing phonetic representation by shifting the subsequent ambiguous auditory speech sound toward the visual component of a prior incongruent and ambiguous audiovisual speech signal. In short, we reason that the congruency of audiovisual speech signals differentiates the perceptual doping notion from the phonetic recalibration effect.

2. Cognitive Mechanism behind the Perceptual Doping Phenomenon

The perceptual doping effect can be understood in terms of the Ease of Language Understanding (ELU) model. In particular, within the ELU framework, there is a perceptual-linguistic component that assumes a Rapid, Automatic, Multimodal Binding of PHOnological information (RAMBPHO [14–16]). RAMBPHO serves as an input buffer that binds, integrates, and processes multimodal input in order to map it with corresponding phonological and lexical representations in semantic long-term memory. In fact, RAMBPHO is the default mode for the implicit processing of speech signals that directly and implicitly unlock the multimodal phonological features of speech signals for accurate mapping onto phonological and lexical representations in semantic long-term memory. This default mode of processing incoming speech signals takes place during the first 100–400 ms of the speech signals being presented [15]. We speculate that the initial audiovisual speech stimulation recalibrates the default mode of processing speech signals in the RAMBPHO input buffer such that congruent auditory and visual speech signals reduce the uncertainty of speech input, particularly in degraded listening conditions. In fact, van Wassenhove et al. [17] found that visual speech cues have a predictive role for the auditory component of a congruent audiovisual speech signal for identification. They reported that visual speech cues speed up the neural processing of congruent auditory input during the first 100 ms of the audiovisual speech signal being presented. In addition, Zion-Golumic et al. [18] revealed that audiovisual over auditory speech presentation results in an enhanced capacity of the auditory cortex to process the temporal features of speech signals in degraded listening conditions. Further, Mégevand et al. [19], in a study of the electrical activity of the human brain via implanted electrodes, found that visual speech cues in an audiovisual speech signal improved phase-tracking and reduced the amplitude of evoked responses to congruent auditory speech signals. Frei et al. [20] also reported that visual speech cues increase neural tracking of the speech cues, particularly in the right auditory clusters, which subsequently results in better speech in noise comprehension in older adults with hearing

loss. We speculate that after audiovisual speech stimulation, the improved capacity of the auditory cortex to track down the most critical features of speech, particularly in degraded listening conditions, does not vanish and is expected to persist for a longer period.

3. Perceptual Doping and Aural Rehabilitation of People with Hearing Difficulties

Most recent studies used auditory (and cognitive) training to enhance speech intelligibility in people with hearing loss. Stropahl et al. [21], in their review of auditory training for improving speech processing skills in people with hearing loss, concluded that intense auditory training might enhance non-trained auditory speech tasks in those with hearing loss.

However, the maintenance of auditory (and cognitive) training effects on auditory speech tasks requires further research. To our knowledge, only a few studies have investigated the efficiency of audiovisual speech training on subsequent auditory speech processing tasks. As mentioned above, Moradi et al. [5] showed that only a short-term exposure (around 30–40 min) to gated audiovisual consonants and words for identification resulted in better auditory performance in the HINT, a non-trained task, in elderly hearing-aid users. In addition, the positive effect of audiovisual speech stimulation remained after one month of training. Rao et al. [22] studied the effect of ReadMyQuips™ (RMQ), an audiovisual training program, on HINT performance in elderly people with hearing loss. The results showed that RMQ improved HINT scores in the experimental group that participated in the RMQ training program.

Tye-Murray et al. [23] studied auditory and audiovisual speech training to listening to noise and speech-reading performance in children with hearing loss. The results showed that both auditory and audiovisual speech training improved both listening and speech-reading performance in children with hearing loss. In addition, the effect of auditory training was more evident in the listening performance, while the effect of audiovisual speech training was evident for both listening and speech-reading performance.

Sato et al. [24] studied the feasibility of in-home audiovisual speech training using a tablet computer to improve speech intelligibility in people with hearing loss. The participants used either a hearing aid or a cochlear implant. The participants listened to audiovisual monosyllable words that were spoken by a female talker. The training took 3 months, and speech intelligibility was recorded for untrained words, trained words, and monosyllables. Results showed that audiovisual speech training using a tablet computer improved untrained and trained words after the 3-month training with audiovisual speech stimuli.

We reason that the addition of visual cues to auditory (and cognitive) training speech materials boosts the effectiveness of aural rehabilitation programs for people with hearing loss in terms of both efficiency and the maintenance of training effects on the listening capacities of persons with hearing loss. This speculation requires further investigation; future studies should consider the transfer of the learning effect from trained to non-trained speech materials, subjective hearing satisfaction in daily life, cognitive function in people with hearing loss, and the maintenance of training effects.

3.1. Perceptual Doping and Controlling Experimental Setup for Collecting Auditory and Audiovisual Speech Data

As prior audiovisual speech exposure boosts the subsequent auditory identification of speech stimuli, caution should be taken when collecting auditory and audiovisual speech data in within-subjects studies when the modality of presentation (auditory or audiovisual) is randomized across participants. In addition, caution should be taken if the participants are first tested in an audiovisual and then an auditory modality. In fact, the perceptual doping may cause a type II error (in failing to reject a false null hypothesis) in a within-subjects design, which may result in non-significant differences between auditory and audiovisual speech data or even lead to better performance of auditory speech stimuli than audiovisual stimuli. We suggest that future studies adopting a within-subjects ex-

perimental design consider collecting data using a fixed-order presentation (instead of a randomized-order presentation) by collecting auditory speech data first and then audiovisual speech data. Nevertheless, although a between-subject experimental design can be used to control perceptual doping and priming effects, individual differences across participants (e.g., differences in terms of speech-reading ability, auditory acuity, and audio and video integration ability) are methodological disadvantages that should be taken into account when comparing auditory and audiovisual groups.

3.2. Suggestions for Future Studies

(1) The extent to which a short period of exposure to audiovisual speech stimuli facilitates visual-only and/or audiovisual speech processing for correct identification is an interesting research topic for future studies. Knowledge is scarce concerning how audiovisual speech stimulation can enhance the processing of visual-only speech cues for correct identification, particularly in people with hearing loss. In face-to-face communication, people with hearing difficulties rely more on visual speech cues, as access to the auditory component of audiovisual speech signals is limited by background noise or hearing loss.

(2) The extent to which audiovisual speech training can improve cognitive function is another interesting research topic for future studies. Fergusson and colleagues [25,26] were the first to show that auditory training can improve cognitive function in people with hearing loss. Hence, the question arises: can audiovisual training result in better cognitive functioning in people with hearing loss than with auditory (or auditory-cognitive training)?

Author Contributions: Conceptualization: S.M. and J.R.; writing original draft preparation: S.M. and J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grant no. 2017-06092 (Anders Fridberger) and the Linnaeus Centre HEAD grant no. 349-2007-8654 (Awarded to Jerker Rönnberg), both from the Swedish Research Council.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sumby, W.H.; Pollack, I. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* **1954**, *26*, 212–215. [CrossRef]
2. Erber, N.P. Interaction of audition and vision in the recognition of oral speech stimuli. *J. Speech Hear. Res.* **1969**, *12*, 423–425. [CrossRef] [PubMed]
3. MacLeod, A.; Summerfield, Q. Quantifying the contribution of vision to speech perception in noise. *Br. J. Audiol.* **1987**, *21*, 131–141. [CrossRef] [PubMed]
4. Moradi, S.; Lidestam, B.; Rönnberg, J. Gated audiovisual speech identification in silence vs. noise: Effects on time and accuracy. *Front. Psychol.* **2013**, *4*, 359. [CrossRef] [PubMed]
5. Moradi, S.; Wahlin, A.; Hällgren, M.; Rönnberg, J.; Lidestam, B. The efficacy of short-term gated audiovisual speech training for improving auditory sentence identification in noise in elderly hearing aid users. *Front. Psychol.* **2017**, *8*, 368. [CrossRef]
6. Bertelson, P.; Vroomen, J.; De Gelder, B. Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychol. Sci.* **2003**, *14*, 592–597. [CrossRef]
7. Van Linden, S.; Vroomen, J. Recalibration of phonetic categories by lipread speech versus lexical information. *J. Exp. Psychol. Hum. Percept. Perform.* **2007**, *33*, 1483. [CrossRef]
8. Hällgren, M.; Larsby, B.; Arlinger, S. A Swedish version of the Hearing In Noise Test (HINT) for measurement of speech recognition. *Int. J. Audiol.* **2006**, *45*, 227–237. [CrossRef]
9. Lidestam, B.; Moradi, S.; Pettersson, R.; Ricklefs, T. Audiovisual training is better than auditory-only training for auditory-only speech-in-noise identification. *J. Acoust. Soc. Am.* **2014**, *136*, EL142–EL147. [CrossRef]

10. Rönnberg, J.; Lunner, T.; Ng, E.H.; Lidestam, B.; Zekveld, A.A.; Sörqvist, P.; Lyxell, B.; Träff, U.; Yumba, W.; Classon, E.; et al. Hearing impairment, cognition and speech understanding: Exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. *Int. J. Audiol.* **2016**, *55*, 623–642. [CrossRef]
11. Moradi, S.; Lidestam, B.; Ng, E.H.N.; Danielsson, H.; Rönnberg, J. Perceptual doping: An audiovisual facilitation effect on auditory speech processing, from phonetic feature extraction to sentence identification in noise. *Ear Hear.* **2019**, *40*, 312. [CrossRef]
12. Norris, D.; McQueen, J.M.; Cutler, A. Perceptual learning in speech. *Cogn. Psychol.* **2003**, *47*, 204–238. [CrossRef]
13. Signoret, C.; Johnsrude, I.; Classon, E.; Rudner, M. Combined effects of form- and meaning-based predictability on perceived clarity of speech. *J. Exp. Psychol. Hum. Percept. Perform.* **2018**, *44*, 277–285. [CrossRef]
14. Rönnberg, J.; Holmer, E.; Rudner, M. Cognitive hearing science: Three memory systems, two approaches, and the Ease of Language Understanding model. *J. Speech Lang. Hear. Res.* **2021**, *64*, 359–370. [CrossRef]
15. Rönnberg, J.; Signoret, C.; Andin, J.; Holmer, E. The cognitive hearing science perspective on perceiving, understanding, and remembering language: The ELU model. *Front. Psychol.* **2022**, *13*, 967260. [CrossRef]
16. Rönnberg, J.; Lunner, T.; Zekveld, A.; Sörqvist, P.; Danielsson, H.; Lyxell, B.; Dahlström, O.; Signoret, C.; Stenfelt, S.; Pichora-Fuller, M.K.; et al. The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* **2013**, *7*, 31. [CrossRef]
17. van Wassenhove, V.; Grant, K.W.; Poeppel, D. Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1181–1186. [CrossRef]
18. Golumbic, E.Z.; Cogan, G.B.; Schroeder, C.E.; Poeppel, D. Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". *J. Neurosci.* **2013**, *33*, 1417–1426. [CrossRef]
19. Mégevand, P.; Mercier, M.R.; Groppe, D.M.; Golumbic, E.Z.; Mesgarani, N.; Beauchamp, M.S.; Mehta, A.D. Crossmodal phase reset and evoked responses provide complementary mechanisms for the influence of visual speech in auditory cortex. *J. Neurosci.* **2020**, *40*, 8530–8542. [CrossRef]
20. Frei, V.; Schmitt, R.; Meyer, M.; Giroud, N. Visual speech cues enhance neural speech tracking in right auditory cluster leading to improvement in speech in noise comprehension in older adults with hearing impairment. *Authorea* **2023**. [CrossRef]
21. Stropahl, M.; Besser, J.; Launer, S. Auditory training supports auditory rehabilitation: A state-of-the-art review. *Ear Hear.* **2020**, *41*, 697–704. [CrossRef] [PubMed]
22. Rao, A.; Rishiq, D.; Yu, L.; Zhang, Y.; Abrams, H. Neural correlates of selective attention with hearing aid use followed by ReadMyQuips auditory training program. *Ear Hear.* **2017**, *38*, 28–41. [CrossRef] [PubMed]
23. Tye-Murray, N.; Spehar, B.; Sommers, M.; Mauzé, E.; Barcroft, J.; Grantham, H. Teaching children with hearing loss to recognize speech: Gains made with computer-based auditory and/or speechreading training. *Ear Hear.* **2022**, *43*, 181–191. [CrossRef] [PubMed]
24. Sato, T.; Yabushita, T.; Sakamoto, S.; Katori, Y.; Kawase, T. In-home auditory training using audiovisual stimuli on a tablet computer: Feasibility and preliminary results. *Auris Nasus Larynx* **2020**, *47*, 348–352. [CrossRef]
25. Ferguson, M.A.; Henshaw, H.; Clark, D.P.; Moore, D.R. Benefits of phoneme discrimination training in a randomized controlled trial of 50- to 74-year-olds with mild hearing loss. *Ear Hear.* **2014**, *35*, e110–e121. [CrossRef]
26. Ferguson, M.A.; Henshaw, H. Auditory training can improve working memory, attention, and communication in adverse conditions for adults with hearing loss. *Front. Psychol.* **2015**, *6*, 556. [CrossRef]