



HAL
open science

Modulating fusion in the McGurk effect by binding processes and contextual noise

Olha Nahorna, Ganesh Attigodu, Frédéric Berthommier, Jean-Luc Schwartz

► **To cite this version:**

Olha Nahorna, Ganesh Attigodu, Frédéric Berthommier, Jean-Luc Schwartz. Modulating fusion in the McGurk effect by binding processes and contextual noise. AVSP 2013 - 12th International Conference on Auditory-Visual Speech Processing, Aug 2013, Annecy, France. pp.181-186. hal-00941306

HAL Id: hal-00941306

<https://hal.science/hal-00941306>

Submitted on 3 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modulating fusion in the McGurk effect by binding processes and contextual noise

Olha Nahorna, Ganesh Attigodu Chandrashekhara, Frédéric Berthommier, Jean Luc Schwartz,

GIPSA-Lab – DPC, UMR 5216 –CNRS University of Grenoble, France

olha.nahorna, ganesh.attigodu, frederic.berthommier, jean-luc.Schwartz, @gipsa-lab.grenoble-inp.fr
<http://www.gipsa-lab.grenoble-inp.fr>

Abstract

In a series of experiments we showed that the McGurk effect may be modulated by context: applying incoherent auditory and visual material before an audiovisual target made of an audio “ba” and a video “ga” significantly decreases the McGurk effect. We interpreted this as showing the existence of an audiovisual “binding” stage controlling the fusion process. Incoherence would produce “unbinding” and result in decreasing the weight of the visual input in the fusion process. In this study, we further explore this binding stage around two experiments. Firstly we test the “rebinding” process, by presenting a short period of either coherent material or silence after the incoherent “unbinding” context. We show that coherence provides “rebinding”, resulting in a recovery of the McGurk effect. In contrary, silence provides no rebinding and hence “freezes” the unbinding process, resulting in no recovery of the McGurk effect. Capitalizing on this result, in a second experiment including an incoherent unbinding context followed by a coherent rebinding context before the target, we add noise all over the contextual period, though not in the McGurk target. It appears that noise uniformly increases the rate of McGurk responses compared to the silent condition. This suggests that contextual noise increases the weight of the visual input in fusion, even if there is no noise within the target stimulus where fusion is applied. We conclude on the role of audiovisual coherence and noise in the binding process, in the framework of audiovisual speech scene analysis and the cocktail party effect.

Index Terms: audiovisual speech perception, McGurk effect, unbinding, rebinding, perception in noise

1. Introduction

It is known since long that the human brain combines visual and auditory information to better understand spoken language, particularly in the case of perception in noise [1-4]. A classical paradigm to demonstrate audiovisual fusion is provided by the “McGurk effect” in which a conflicting visual input modifies the perception of an auditory input, e.g. visual /ga/ added on auditory /ba/ leading to the percept of /da/ [5].

Audiovisual fusion in speech perception has long been considered as automatic [6, 7]. However a number of recent experiments have provided evidence that it is in fact under the control of attention in a broad sense, considering that various cognitive variables can modulate audiovisual integration [8-13].

1.1 Binding and unbinding in audiovisual fusion

While evidence for the non-automaticity of the fusion mechanism stays compatible with one-stage architecture, some data suggest that audiovisual interactions could intervene at various stages in the speech decoding process [14-16]. Actually, audiovisual fusion could be conceived as a two-stage process, beginning by *binding* together the appropriate pieces of audio and video information, followed by integration per se [17]. The binding stage would occur early in the audiovisual speech processing chain enabling the listener to extract and group together the adequate cues in the auditory and visual streams, exploiting coherence in the dynamics of the sound and sight of the speech input.

To demonstrate the existence of this “binding” process we defined an experimental paradigm possibly leading to “unbinding”. In this paradigm (Figure 1) incongruent “McGurk” (A/ba/ + V/ga/) or congruent “ba” (A/ba/ + V/ba/) targets were preceded by coherent or incoherent audiovisual contexts [18]. The experimental results showed that the McGurk effect (displaying the role of the visual input on phonetic decision) depends on the previous audiovisual context. Indeed, various kinds of incoherent contexts, such as acoustic syllables dubbed on video sentences, or phonetic or temporal modifications of the acoustic content of a regular sequence of audiovisual syllables, can significantly reduce the McGurk effect. Short incoherent context durations (even 1-syllable long) were sufficient to produce a significant amount of unbinding [19]. On the contrary, coherent contexts let the McGurk effect stable, which suggests that there is possibly a “default mode” in which binding occurs (and hence produces the McGurk effect in isolation).

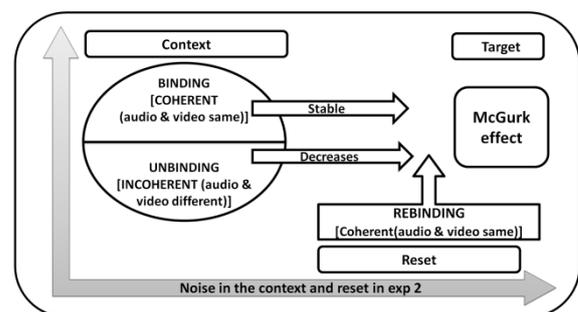


Figure 1: Experimental paradigm for displaying unbinding or rebinding mechanisms modulating the McGurk effect

1.2 Experiment 1- From unbinding to rebinding

Our previous studies clearly show that an incoherent context results in a decrease of the McGurk effect, which is due in our interpretation to an “unbinding” mechanism. An unanswered

question is to know what kind of information is able to reset the system and put it back in its supposedly bound default state [18]. The objective of the first experiment in the present paper is to attempt to answer this question. For this aim, we tested whether applying a period of either coherent material or silence after the incoherent “unbinding” context would enable to recover the McGurk effect (Figure 1). The driving hypothesis of the first experiment is the following: (1) the incoherent context alone should decrease the McGurk effect; (2) the additional reset context, if it is efficient for rebinding, should result in recovering the McGurk effect, therefore the amount of McGurk responses should increase for increasing durations of the reset stimulus.

1.3 Experiment 2- The role of noise in audiovisual fusion

The role of visual speech is particularly important in noise [1-4]. Noise also seems to modulate decision in the case of incongruent stimuli. Indeed, if one applies noise during a McGurk stimulus, the McGurk effect decreases when the extraneous noise is visual, whereas it increases when the noise is auditory [20-24]. In the well-known “Fuzzy-Logical Model of Perception” (FLMP [6]) this is interpreted as due to the increasing ambiguity of the noisy component, which would automatically decrease its role in the fusion process. However, it could also be envisioned that there is a specific weighting process in which a given modality would be positively or negatively modulated in the fusion process depending on the noise in this modality [25, 26]. In the first case fusion would only depend on stimuli while in the second case there would be in addition an evaluation of the perception conditions resulting in a modification of the fusion process per se. Our reasoning here is that if noise is applied in the (context + reset) part of the stimulus in Figure 1 but not on the target itself, if fusion only depends on stimuli, then the McGurk effect should not change since the McGurk target stays clear. If however fusion depends on a weighting process driven by the environment, then application of acoustic noise in the context part should result in increasing the role of vision in fusion, hence increasing the McGurk effect. The second experiment aims at testing the role of noise on context, and its interaction with the binding/unbinding/rebinding processes.

2. Method

Globally, the two experiments consisted in testing the McGurk effect in various kinds of contexts including: (i) a coherent vs. incoherent component to replicate unbinding – with decrease of the McGurk effect – in case of incoherent contexts; (ii) for incoherent contexts, a coherent reset component to test the possibility of rebinding – with recovery of the McGurk effect (Experiment 1); (iii) addition of acoustic noise in one set of conditions, to test if noise added to the (context+reset) part could globally increase the McGurk effect (Experiment 2).

2.1. Stimuli

The stimuli are described in Figure 2. They were typically made (Fig. 2, top) of:

- an incoherent context (2 or 4 acoustic syllables superimposed on excerpts of video sentences matched for equal duration) for both experiments (“*incoherent context*”);

- followed by a reset stimulus consisting in 0, 1, 2 or 3 coherent audiovisual syllables (“*coherent reset*”) or audio silence with fixed image of duration 0, 480, 1000, 1480 ms (“*fixed reset*”) in Experiment 1, while in Experiment 2 we used only the coherent reset;
- finishing by a target which could be either a congruent audiovisual “ba” or a McGurk stimulus consisting in an audio “ba” dubbed on a video “ga”.

A control stimulus, aimed at providing a reference for the McGurk effect, was provided by (Fig. 2, bottom):

- a coherent context (2 or 4 coherent audiovisual syllables) (“*coherent context*”);
- followed by a target which could be either a congruent audiovisual “ba” or a McGurk stimulus.

A series of audiovisual films were presented to participants in two blocks in both experiments. In Experiment 1, there was one block with coherent reset and the other one where the reset consisted in the silence with fixed image. In Experiment 2 there was one block without acoustic noise (“*silent*”) and the other one with acoustic noise superimposed on all context and reset parts of the stimuli (“*noise*”). Noise consisted in speech-shaped noise at 0 dB SNR. The target parts always remained without noise.

Coherent context and coherent reset material was constructed by pairing audiovisual syllables randomly selected within the following syllables (“pa”, “ta”, “va”, “fa”, “za”, “sa”, “ka”, “ra”, “la”, “ja”, “cha”, “ma”, “na”). In Experiment 1, the “fixed reset” was obtained by dubbing auditory silence on fixed image with durations 0, 480, 1000 or 1480 ms. In the incoherent context material, the auditory content was same, but the visual content was replaced by excerpts of video sentences matched in duration.

The congruent “ba” target was used to ensure that participants were performing the speech task correctly and to serve as a baseline to contrast with the McGurk effect. The incongruent McGurk target was produced by carefully synchronizing an auditory /ba/ with a video /ga/, precise temporal localization of the acoustic bursts of the original “ba” and “ga” stimuli providing the cue for synchronization. McGurk targets were presented three times more than congruent “ba” targets, which served as controls.

For each (context+reset) condition (2 context durations; 4 reset durations for incoherent context; 2 reset types in Experiment 1, and 2 noise conditions in Experiment 2; hence altogether 20 context conditions) there were 4 occurrences of a “ba” target and 12 occurrences of a McGurk target. Hence there were 320 sequences in total spread over 2 blocks of 10 min each.

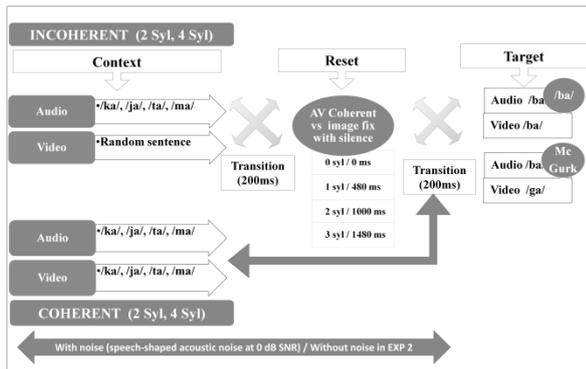


Figure 2. Description of the audiovisual material

2.2. Procedure

All experiments were carried out in a soundproof booth. Stimulus presentation and recording of responses were controlled by the Presentation software. The experiment consisted of two possible responses “ba” or “da” (with one button for “ba” and one for “da,”) and the participants were instructed to constantly look at the screen and, each time a “ba” or a “da” was perceived, to immediately press the corresponding button. The films were presented on a computer monitor with high-fidelity headphones set at a comfortable fixed level. The video stream was displayed at a rate of 25 images per second, the subject being positioned at about 50 cm from the screen. There were 5 different orders of the stimuli in the films, and the order of the two blocks “fixed reset” and “coherent reset” in the case of Experiment 1 and “silent” and “noise” in the case of Experiment 2 was counterbalanced between subjects. The response button was also interchanged between subjects.

2.3 Participants

Twenty subjects participated in Experiment 1 (9 women and 11 men; mean 25.7 years) and twenty in Experiment 2 (13 women and 7 men; mean 34 years). All of them were French native speakers, without any reported history of hearing disorders and with normal or corrected-to-normal vision. Written consent was obtained from each participant and all procedures were approved by the Grenoble Ethics Board (CERNI).

2.4 Assumptions and analyses

The experiment was focused on the role of context, reset and noise on the McGurk effect. For each (context, reset and noise) condition, each target and each subject, the amount of “ba” responses against “ba+da” responses was computed and used as an index of the subject’s perception (between 0 and 1). An $\text{arc}(\sin(\sqrt{x}))$ transformation was applied on these relative “ba” scores to ensure Gaussianity of the dependent variable in the analyses of variance that will be presented in Section 3. Though response times were systematically recorded and processed, they will not be presented here.

We had three main assumptions, all involving McGurk stimuli (let us recall that “ba” targets are just there as controls).

- Firstly, incoherent context should produce unbinding and decrease the McGurk effect (hence increase the amount of “ba” responses) in respect to coherent

context, whatever the context duration (2 or 4 syllables).

- Secondly, for incoherent context, reset should produce rebinding and increase the McGurk effect (hence decrease the amount of “ba” responses), from 0 to 3 syllables of duration of the reset stimulus. However, we had no expectation at the beginning of Experiment 1 whether “coherent” or “fixed” reset would both provide efficient reset.
- Thirdly, noise in Experiment 2 should enhance the role of vision and hence globally increase the McGurk effect (decrease the score of “ba” responses) whatever the context and reset.

3. Results

3.1 Preliminary remarks

As expected, the “ba” target leads to 100% “ba” responses in both experiments and in all conditions. Therefore, for now on, we shall concentrate on McGurk targets.

Preliminary analyses of the role of context duration in the incoherent context conditions in both experiments showed that the incoherent context duration (2 vs. 4 syllables) has only little effect on the McGurk effect, hence we shall average data for the two context durations in the next analyses.

3.2 Assessing the efficiency of coherent vs. fixed resets in Experiment 1

On Figure 3 we display relative “ba” scores for McGurk targets in all conditions for Experiment 1 (averaging over the two context durations, 2 and 4 syllables). Three major facts emerge from this figure.

- *Unbinding with incoherent context.* Let us first look at what happens for the incoherent context without reset, corresponding to the 0-syl condition (left bars, for both types of resets). The score of “ba” responses is around 75-80%, much larger than the score for the “coherent context” condition (rightmost bars), which is less than 50%. This replicates the decrease of McGurk effect from coherent (more than 50% McGurk effect) to incoherent context (less than 25% McGurk effect). This decrease is due in our interpretation to “unbinding”, resulting in a decrease of the visual weight in fusion for the target perception.
- *Poor rebinding with fixed reset.* Looking at the bars in light grey on Figure 3, corresponding to the “fixed reset” condition, it appears that this reset (made of acoustic silence + fixed image) provides almost no rebinding, since the “ba” score only slightly decreases from 0 to 1-syl (that is 480ms duration), then remains stable and stays much larger than the score for coherent context even for the longest reset duration (3-syl corresponding to 1480 ms).
- *Good rebinding with coherent reset.* On the contrary, looking at the bars in dark grey corresponding to the “coherent reset” condition, we observe that the “ba” score regularly decreases with reset duration and reaches the same value as for coherent context, coming back to its “default” state for the largest coherence period of 3 syllables.

To assess the significance of the rebinding effects, we performed an analysis of variance with the factors “subject” (random-effect), “reset” (coherent vs. fixed) and “reset duration” (0, 1, 2 & 3 syllables / 0, 480, 1000, 1480 ms). The three factors were statistically significant (“subject”: $F(19,18)=6.88$, $P<0.001$; “reset”: $F(1,19)=5.45$, $P<0.05$; “reset duration” [$F(3,57)=14.9$, $P<0.001$].

The interaction between “reset” and “reset duration” was also significant [$F(3,57)=7.65$, $P<0.001$], which is in agreement with the difference between variations of scores with reset duration for fixed vs. coherent reset. Post-hoc analyses with Bonferroni corrections show that in the fixed reset condition, there is no difference between scores for the four reset durations. In the case of coherent reset syllables, the score at 0 was significantly higher than with 2 or 3 syllables, and the score at 1 or 2 syllables was significantly higher than with 3 syllables ($P < 0.05$).

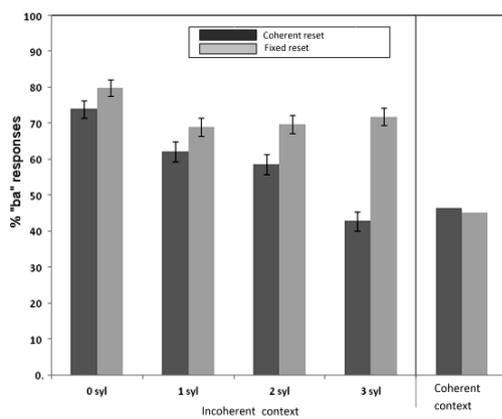


Figure 3. Results for Experiment 1. Percentage of “ba” responses for “McGurk” targets, in the “coherent reset” vs. “fixed reset” conditions for incoherent context with the four reset durations, compared with coherent context. The ANOVA was performed only for the four reset durations in the incoherent context.

3.3 Assessing the effect of noise in Experiment 2

On Figure 4 we display “ba” scores for McGurk targets in all conditions for Experiment 2 (averaging over the two context durations 2 and 4 syllables). Two major facts emerge from this figure.

- *Unbinding/rebinding*: Focusing on the “without noise” condition (black bars), we replicate the results of Experiment 1 in the “coherent reset” condition (remember that only coherent reset is used in Experiment 2). Indeed, the “ba” score is higher (less McGurk effect) for incoherent context without reset (most left) than for coherent context (most right) (unbinding). But it decreases when reset duration increases from 0 to 3 syllables (rebinding). The effects are quantitatively different from Experiment 1 to Experiment 2, which is not unexpected considering the large inter-individual differences in the McGurk effect. But the portrait is qualitatively similar.
- *Modulation by noise*: Comparing black bars (with noise) and grey bars (without noise), it appears that noise decreases “ba” scores (and hence increases the

McGurk effect) for all conditions, and by a large amount (12 to 20%).

To assess the effects of rebinding and noise, we performed an analysis of variance with the factors subject (random-effect), noise (silent vs. noise) and reset duration (0, 1, 2 & 3 syllables). The three factors are statistically significant (“subject”: $F(19, 18) = 13.46$, $P < 0.001$; “noise”: $F(1, 19) = 6.12$, $P < 0.05$; “reset duration”: $F(3, 57) = 14.82$, $P < 0.001$). There was no significant interaction between any pair of factors.

The effect of reset duration confirms the result of Experiment 1 for coherent reset. The effect of noise confirms that applying noise in the context+reset part modulates the target perception even though there is no noise during the target. The lack of interaction between noise and reset duration shows that the role of noise seems more or less stable whatever the reset duration. Altogether, it appears that noise applied in the context part modifies the results of audiovisual fusion, with a global and more or less stable effect leading to an increase of about 15% in the McGurk effect whatever the context.

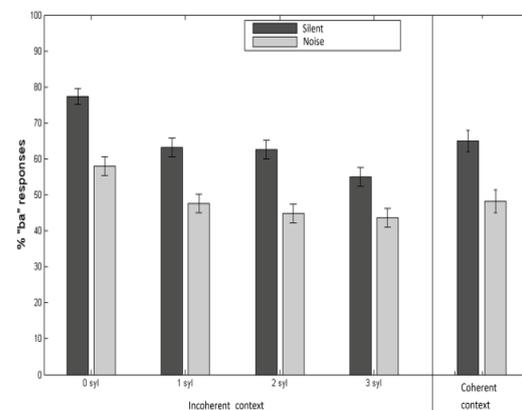


Figure 4. Results for Experiment 2. Percentage of “ba” responses for “McGurk” targets, in the “silent” vs. “noise” conditions for incoherent context with the four reset durations, compared with coherent context. The ANOVA was performed only for the four reset durations and the two noise levels in the incoherent context.

4. Discussion

4.1 Unbinding, rebinding and noise in the audiovisual fusion process

This set of experiments confirms that context may modify the McGurk effect, through a series of mechanisms, which combine unbinding (through incoherent context decreasing the role of the visual input), rebinding (through coherent reset setting back the weight of the visual input) and noise (increasing the role of the visual input).

In Experiment 1, it appeared that a fixed reset has almost no rebinding effect, with the consequence that even for the longer duration (around 1.5s) the subjects stay “frozen” in an unbound state where the McGurk effect is largely decreased. On the contrary, a coherent reset of 3 syllables is enough to completely recover from unbinding and restore the default binding stage.

In Experiment 2, it appeared that noise applied on the contextual part – but NOT on the target – systematically increases the McGurk effect, whatever the content of context and reset. To our knowledge it is the first time that such a result is obtained. This strongly suggests that noise in the McGurk effect, already displayed with noise applied on the target itself [20-24], intervenes not only at the level of the stimuli, but also at the level of the fusion process itself.

From there on, it is possible to come back to the models of audiovisual fusion available in the literature. Classical models consider that phonetic decision operates at a given representational stage and produces an integrated percept combining auditory and visual cues in a given way, possibly mediated by general attentional mechanisms. Our data on the binding process led us suggest that an additional computational stage should be incorporated before decision operates, involving online computation of some assessment of the coherence/ incoherence of the auditory and visual inputs, resulting in a “two-stage model” of audiovisual speech perception [17] (see Fig. 5).

The present results first add some information about the way coherence could be computed, involving a dynamics made of unbinding and rebinding stages with short constant times: indeed, less than one second of incoherence (2 syllables or less) suffices to produce unbinding, and less than one second of coherence (2 syllables or less) suffices to produce complete rebinding.

Furthermore, the results about noise suggest that noise, and probably more generally knowledge about the conditions of communication, also participate to the decision process by providing an enhancement of “efficient” modalities, not contaminated by noise, versus modalities where noise could contaminate the decision process (Fig. 5).

The present data suggest that the role of unbinding/rebinding on one hand, and noise-based selective weighting of each modality on the other hand, could play additional independent roles, according to the lack of interaction between noise and reset in Section 3.3. This will have to be confirmed in future experiments specifically dealing with this question.

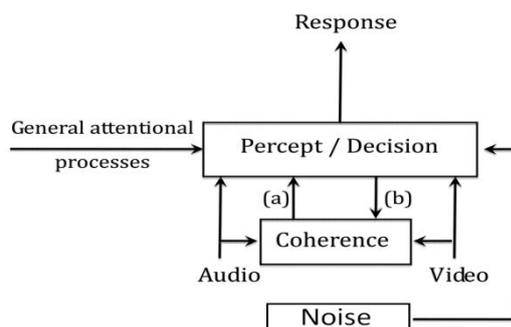


Figure 5: A two-stage model of audiovisual speech perception

4.2 Future experiments

A number of further experiments will have to extend the present data in various directions, involving e.g. more about the dynamics of unbinding and rebinding. Various proposals could also deal with reset mechanisms (such as changing speaker or the global communication setting), or specificity of the binding mechanism (could non-speech incoherent

audiovisual material also produce unbinding?). The role of noise could also be further assessed by using visual noise. Indeed, some studies [10, 24] have manipulated the size or position of the face and found influence on the McGurk effect, showing in both that visual noise may decrease the McGurk effect just as auditory noise increases it. If our conjecture about the role of noise in Figure 5 is correct, this effect should also occur for visual noise added on the contextual part of the stimuli in the present paradigm.

Another important extension concerns intelligibility in noise. The present paradigm was an aim to progress towards the next important question that is to know if unbinding mechanisms would also decrease the beneficial effect of lipreading in noise. Future experiments will deal with targets consisting in ambiguous though coherent stimuli and test if an incoherent audiovisual context is able to remove the visual benefit. This will enable us incorporate the two-stage model inside a general question concerning the cocktail-party effect and what we propose to call “audiovisual speech scene analysis” [18].

5. Acknowledgements

This project has been supported by ANR Multistap (MULTISTability and binding in Audition and sPeech: ANR-08-BLAN-0167 MULTISTAP) which provided a doctoral funding for Olha Nahorna, and by Academic Research Community “Quality of life and ageing” (ARC 2) of the Rhône-Alpes Region, which provided a doctoral funding for Ganesh Attigodu Chandrashekara.

6. References

- [1] Binnie, C. A., Montgomery, A. A., and Jackson, P. L., "Auditory and visual contributions to the perception of consonants", *J. Speech Hear. Res.*, 17:619–630, 1974.
- [2] Erber, N. P., "Auditory and audiovisual reception of words in low-frequency noise by children with normal hearing and by children with impaired hearing", *J. Speech Hear. Res.*, 14:496–512, 1971.
- [3] Grant, K. W., and Seitz, P., "The use of visible speech cues for improving auditory detection of spoken sentences", *J. Acoust. Soc. Am.*, 108:1197–1208, 2000.
- [4] Sumbly, W. H., and Pollack, I., "Visual contribution to speech intelligibility in noise", *J. Acoust. Soc. Am.*, 26:212–215, 1954.
- [5] McGurk, H., and MacDonald, J., "Hearing lips and seeing voices", *Nature*, 265:746–748, 1976.
- [6] Massaro, D., "Speech perception by ear and eye", Hillsdale: LEA, 1987
- [7] Soto-Faraco, S., Navarra, J., and Alsius, A., "Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task", *Cognition*, 92; B13–B23, 2004.
- [8] Alsius, A et al., "Audiovisual integration of speech falters under high attention demands", *Curr. Biol.*, 15:839–843, 2005.
- [9] Buchan, J.N., and Munhall, K.G., "The Effect of a Concurrent Cognitive Load Task and Temporal Offsets on the Integration of Auditory and Visual Speech Information", *Seeing and Perceiving*, 25:87-106, 2012.
- [10] Colin, C., Radeau-Loicq, M. and Deltenre, P., "Top-down and bottom-up modulation of audiovisual integration in speech", *Eur. J. Cogn. Psychol.*, 17:541-560, 2005.
- [11] Mozolic, J.L et al., "Modality-specific selective attention attenuates multisensory integration", *Exp. Brain. Res.*, 184(1):39-52, 2008.
- [12] Navarra, J et al., "Assessing the role of attention in the audiovisual integration of speech", *Informational Fusion.*, 11(1):4-11, 2009
- [13] Tiippana, K., and Andersen, T. S. (2004). Visual attention modulates audiovisual speech perception, *Eur. J. Cogn. Psychol.*, 16:457-472, 2004.
- [14] Bernstein, L., Auer, E., and Moore, J., "Audiovisual speech binding: convergence or association?" In G.A. Calvert et al. (Eds) *The handbook of multisensory processes*, Cambridge: The MIT Press, pp. 203–224, 2004.
- [15] Grant, K. W., and Seitz, P., "The use of visible speech cues for improving auditory detection of spoken sentences", *J. Acoust. Soc. Am.*, 108:1197–1208, 2000.
- [16] Van Wassenhove, V., Grant, K., and Poeppel, D., "Visual speech speeds up the neural processing of auditory speech", *PNAS*, 102:1181–1186.
- [17] Berthommier, F. "A phonetically neutral model of the low-level audiovisual interaction", *Speech. Comm.*, 44:31–41, 2004.
- [18] Nahorna, O., Berthommier, F., and Schwartz, J.L., "Binding and unbinding the auditory and visual streams in the McGurk effect", *J. Acoust. Soc. Am.*, 132:1061-1077, 2012.
- [19] Nahorna, O. "Audiovisual speech binding", PhD Thesis, Grenoble University (to appear), 2013.
- [20] Sekiyama, K., and Tohkura, Y., "McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility", *J. Acoust. Soc. Am.*, 90:1797-1805, 1991.
- [21] Sekiyama, K. and Tohkura, Y., "Inter-language differences in the influence of visual cues in speech perception", *J. Phonetics*, 21:427-444, 1993.
- [22] Sekiyama, K., "Difference in auditory-visual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility", *J. Acoust. Soc. Jap.*, 15:143-158, 1994.
- [23] Fixmer, E., and Hawkins, S., "The influence of quality of information on the McGurk effect", in *Proceedings AVSP'1998*, Terrigal, Australia, pp. 27-32.
- [24] Kim, J., and Davis, C., "Audiovisual speech processing in visual speech noise", in *Proceedings AVSP'2011*, Volterra, Italy, pp. 73-76.
- [25] Heckmann, M., Berthommier, F., and Kroschel, K., "Noise Adaptive Stream Weighting in Audio-Visual Speech Recognition", *EURASIP J. Appl. Signal Processing*, 11:1260–1273, 2002
- [26] Huysse, A, Berthommier, F, and Leybaert, J., "Degradation of labial information modifies audiovisual speech perception in cochlear-implanted children", *Ear. Hear.*, 34:110-21, 2012.