



Auditory-visual speech perception in bipolar disorder: behavioural data and physiological predictions

Doğu Erdener¹, Şefik Evren Erdener², Arzu Yordamlı³

¹Psychology Program, Middle East Technical University, Northern Cyprus Campus, N. Cyprus

²Institute of Neurological Sciences and Psychiatry, Hacettepe University, Ankara, Turkey

³Clinical psychologist in private practice, Bodrum, Muğla, Turkey

vdogu@metu.edu.tr, evrenerdener@hacettepe.edu.tr

Abstract

In this study, we tested two bipolar groups, manic-episode and depressive episode patients and healthy controls over a series of McGurk effect stimuli as well as auditory and visual-only (lip-read) speech stimuli. We hypothesized that the bipolar group's auditory-visual speech integration should be weaker than the control group. We also predicted the manic-period bipolar individuals should integrate visual speech information more robustly than their depressive-episode counterparts. These hypotheses were not supported, and all groups were found to integrate auditory and visual speech information comparably. However, paradoxically, the depressive-episode individuals were unable to lip-read despite that they were able to integrate the two sources of speech information. Here, we discuss and try to make sense of this behavioural data with solid predictions on how corresponding physiological data can decipher our findings. On top of this discussion, physiological predictions and possibilities are presented.

Index Terms: auditory-visual speech perception, bipolar disorder, McGurk effect, clinical psycholinguistics, physiological predictions

1. Introduction

It is now the consensus that speech perception is an auditory-visual process whereby we integrate orofacial movements into what we hear. A now classical example of this is a visual speech illusion called the McGurk effect in which the auditory syllable /ba/ dubbed onto the lip movements for /ga/ is often perceived as /da/ or /tha/ by listeners/users of most languages such as English [1] and Turkish [2] and it works on word and sentence levels as well [3]. Beyond this, the McGurk effect has come to be used as a widespread measure of the degree to which visual speech information influences the resultant percept – but also see [4].

While there are inter-language and developmental differences with respect to the weight of visual speech information [e.g., 5], the visual dimension of speech perception and processing has been put into application in an array of domains. In one area of application is foreign language instruction. This domain is based on the general finding that the visual speech influence is greater when we attend to speech in a foreign language (L2) than in our native language (L1) [5, 6]. This difference between L1 and L2 can be interpreted in a way that the perceptual system makes use of any source of speech information to decipher the non-native speech signal, particularly visual. The visual speech information has a potential to be exploited to be integrated into L2 instructional

materials [7] with an awareness of inter-language differences [8]. While the L2 instruction field appears to be a good example of an ever-growing applied area of auditory-visual speech research, there are other areas of potential applications that have just been emerging, such as the psychopathological domain – the focus of this paper. The next section presents the very limited literature on the issue and paves the way for the current investigation.

2. Auditory-visual speech perception and psychopathology

Despite the steady growth of research in auditory-visual speech perception, the clinical population has largely been neglected. Given that, as in the case of L2, auditory-visual speech research has a lot of potential for a plethora of domains including the clinical one. In the context of speech and hearing pathology, we know that adults and children with hearing problems tend to make use of visual speech more than their hearing counterparts [9], and those with phonological speech disorder have difficulty merging visual speech information with the auditory input [10] – also see [11]. These and similar results demonstrate that the role of visual speech information can profitably be applied to these domains as well. In that respect, one wonders what happens to auditory-visual speech perception in those with mental disorders.

Currently, there is hardly any literature on the issue save for a small number of scattered studies. A couple of studies with schizophrenic patients showed difficulty of auditory-visual integration and that the amount of visual speech influence was inversely related to age [12, 13]. These differences between healthy and schizophrenic patients were also demonstrated via cortical data as well. Surguladze and his colleagues showed that a lip-reading condition activated the superior and inferior posterior temporal areas of the brain in healthy controls, the activation in these areas in their schizophrenic counterparts was significantly less [14]. Normally, these areas are associated with the conversion of visual speech input into a form to be processed by the auditory cortex [15]. The impediment of this conversion is apparently claimed to be associated with schizophrenia [16]. Similar auditory-visual speech discrepancies were observed in people with other mental disorders such as Alzheimer's disease [17]. A similar auditory-visual speech discrepancy was observed in people with Asperger's syndrome [18], this discrepancy relatively dissipates and becomes negligible with age [19].

2.1. The current study: auditory-visual speech perception and bipolar disorder

Here, we investigated the status of auditory-visual speech perception in individuals with bipolar disorder - a disorder marked by alternating – and contrastive – periods of mania and depression. While we wanted to see the difference between healthy and bipolar individuals, we also wanted to see the difference between manic and depressive episodes as well. Due to the paucity of research, if any, that would help us formulate a hypothesis we capitalised on anecdotal observations and a limited number of studies that point to attentional differences between bipolar and healthy groups. These sources show that bipolar individuals seem to have problems with sustained, goal-directed attention compared to non-disordered individuals [20] and that the manic-period bipolar individuals focus on tasks more excessively and in a goal-directed fashion than the depressive-episode bipolar individuals [21]. Based on previous works with other clinical populations, e.g., schizophrenics, we hypothesized that the healthy controls should give more visually-based/integrated responses to the McGurk-type auditory-visual stimuli than their bipolar-disordered counterparts. We further predicted that if the auditory and visual speech information are integrated at behavioural level as a function of attentional focus and excessive goal-directed behaviour [21], then manic-period bipolar participants should give more integrated responses than their depressive-episode counterparts.

2.2. Method

2.2.1. Participants

A total of 44 native Turkish-speaking participants (14 females, 30 males, $M_{age} = 28.8$ years, $SD = 10.2$) were recruited. The bipolar disorder sample consisted of 22 in-patients from a state mental hospital in western Turkey. Twelve were in manic ($M_{age} = 30.9$ years, $SD = 6.78$) and 10 were in depressive ($M_{age} = 41.5$ years, $SD = 11.3$) episodes. All bipolar subjects were on valproate and/or lithium-based medications. As controls, 22 participants ($M_{age} = 21.8$ years, $SD = 1.26$) were recruited from volunteers at Middle East Technical University's (METU) Cyprus Campus. All participants were native speakers of Turkish with normal hearing and normal/corrected-to-normal vision. A written informed consent was obtained from each participant.

2.2.2. Materials and Procedure

The stimuli were produced using recordings of words and non-words in Turkish by a male and a female talker. From these auditory-visual, auditory and visual stimuli were edited. The auditory-visual stimuli were produced via a dubbing procedure, e.g., auditory /soba/ + visual /soga/, percept \rightarrow soda). In the auditory-visual stimuli, the fused responses always produced a percept of a real word in Turkish allowing us to see whether a judgement was visually-based. The auditory and visual-only stimuli were created by stripping them off from visual and auditory portions, respectively. All AO and VO words were real words. There were 24 auditory-visual, 12 auditory and 12 visual-only stimuli which were presented in a quiet testing room at METU and at the state hospital. Each stimulus was presented once thus a total of 48 trials. The ambient sound level was set at 65 dB. The participants were asked to “watch and listen” the stimuli and state what they perceived. To avoid task-related burdens in favour of disordered participants, responses were

manually recorder by an experimenter. The test phase was preceded by a familiarisation phase which simply consisted of two AV trials that were not included in the actual stimulus set.

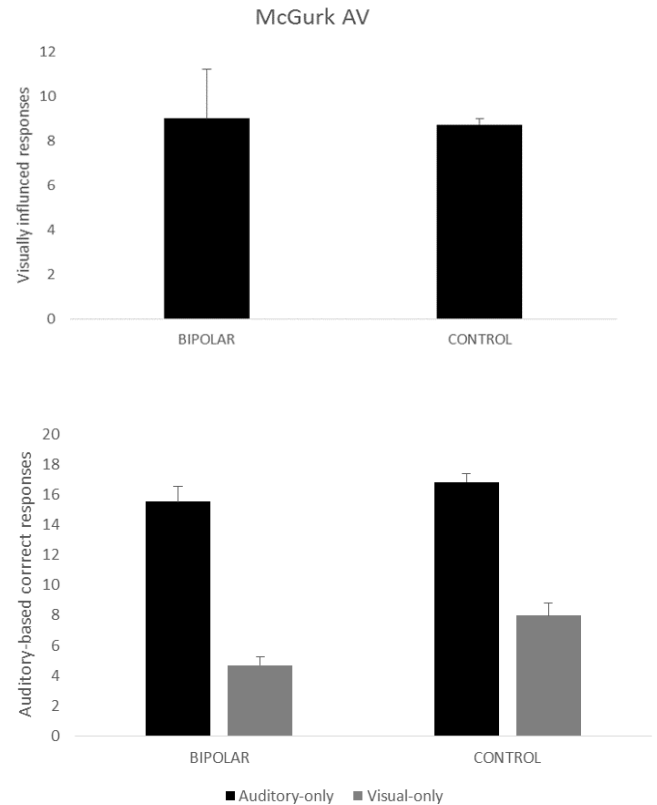


Figure 1: The mean scores for auditory-visual, auditory-only and visual-only conditions with error bars representing the standard error of the mean.

2.3. Results

Two sets of analyses, a t-test analysis to compare healthy controls and the overall bipolar groups and the non-parametric Kruskal-Wallis to compare the two bipolar and one control groups.

2.3.1. Disordered versus Control groups

Independent samples *t*-tests were conducted on auditory-visual, auditory and visual scores comparing the overall bipolar-disordered and the control groups. The Levene's tests showed the homogeneity of variance assumptions were met ($p > .05$) except for the auditory scores ($p = 0.029$).

There was no significant difference between the overall bipolar and control groups, on the auditory-visual, $t(42) = .227$, $p = .82$, and auditory scores, $t(35.602) = .593$, $p = .56$. However, the control group performed significantly better than the disordered group on the VO scores, $t(42) = 2.882$, $p < .005$ (see Figure 1).

2.3.2. Comparing manic, depressive and control groups

Due to small group sizes, in lieu of a one-way analysis of variance, we decided to use its non-parametric version,

Kruskal-Wallis test to compare the two bipolar groups – manic and depressive – and the control group.

The analysis of both auditory-visual scores ($\chi^2 (2, N= 44) = .804, p= .669$) and auditory scores, ($\chi^2 (2, N= 44) = .584, p= .747$) revealed no significant difference amongst the three groups; however, the visual-only scores did, $\chi^2 (2, N= 44) = 7.665, p= .022$. Further to this, we ran a Mann-Whitney U test on the visual-only scores. The three-way comparisons revealed no significant difference between manic bipolar and the control groups, $z (N = 34) = -1.773, r^2 = .080$, as well as between manic and depressive bipolar groups, $z (N = 22) = -.863, r^2 = .418$. On the other hand, the results showed that control group was more prone to visual speech information than their depressive-episode counterparts, $z (N = 32) = -2.569, r^2 = .009$ (See Figure 2 for group means).

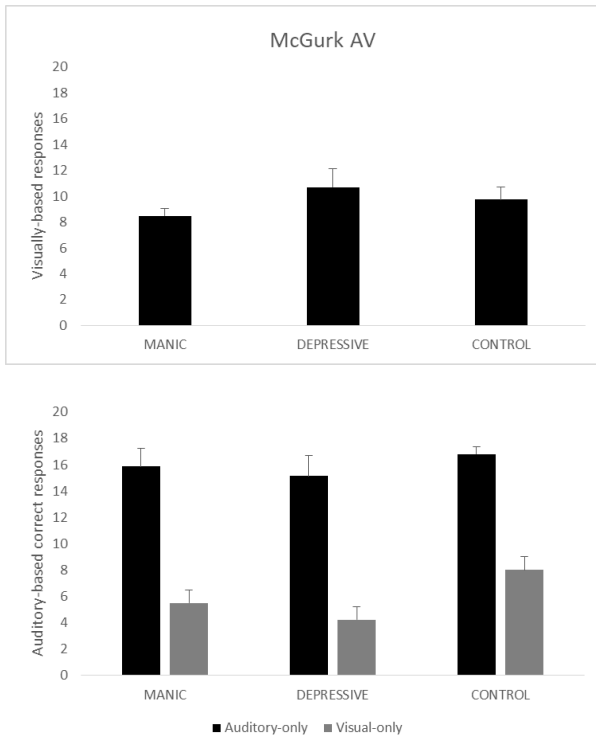


Figure 2: The mean scores for auditory-visual, auditory and visual-only conditions from the two bipolar groups and the control group with error bars representing the standard error of the mean.

2.4. Discussion: behavioural data

The hypothesis that the bipolar group would overall yield less auditory-visual speech integration has not been supported. Similarly, no differences were observed for auditory stimuli, either. Likewise, there were also no differences amongst the two bipolar groups and the control group over both auditory-visual and auditory stimuli. However, the comparison of visual-only (lipreading, henceforth) scores revealed that the healthy group significantly performed better than the overall bipolar group and both control and manic-period groups performed significantly better than the depressive period participants.

While this preliminary data on the auditory-visual speech perception by the bipolar group/s shows that the integration pattern works akin to that of healthy individuals, then, how, paradoxically, is it possible that the depressive-episode bipolar

individuals show poor performance on a lip-reading task? In the context of our data, one can think of two possibilities for explanation: (a) given that speech perception is predominantly auditory, the visual speech information is processed with no integration as would be suggested by auditory-dominant models [e.g., 22]. In this case, what is the level of auditory-visual integration in the bipolar individuals: phonetic [23], phonological [24] as claimed for the healthy population or some other?; (b) given the no difference among the groups on auditory-visual speech integration, this calls for a scrutiny on whether different cortical pathways are activated to achieve integration – phonetic or phonological. Unfortunately, our data does not answer these questions readily as the answer require multi-level data. Thus we believe that, while we need to replicate this study with typical McGurk stimuli as well as using new methods of auditory-visual speech perception research [4], we also should scrutinize this paradoxical picture using physiological methods thus check against physiological / cortical data. Then we will have a better and fuller understanding of the whole picture. The next section proposes what can/should be done to make a better sense of the behavioural data on this issue.

3. Conclusion

3.1. Where does our data lead us to?

The interpretation of the paradoxical finding that the depressive-episode bipolar participants can integrate auditory and visual speech information like their healthy and manic-period counterparts, but their lip-reading is impoverished unlike their manic-episode and healthy control counterparts warrants further research. Our data does not provide sufficient explanation and remains only descriptive at this stage. This problem with lipreading may be a significant secondary symptom of entry into the depressive period of this debilitating disorder thus we believe deserves scrutiny.

As best as we can understand from our data, the poor processing of the lip-read (visual-only condition) information does not affect the eventual auditory-visual speech integration process in the depressive-episode patients. However, the visual speech information is, in fact, significant in that it activates several different cortical areas that are normally associated with auditory information processing. For instance, it activates the superior temporal gyrus which is not activated in response to non-speech movements of the face [15]. When the face performs speech-related movements, the lip-read information is converted to an auditory code in healthy individuals as well as manic-episode patients. However, our data suggests that this does not occur in depressive-episode bipolar individuals. Our behavioural data coupled with that of Calvert et al. [15], may suggest that the way visual speech information is integrated in bipolar individuals is most likely different to the way it occurs in healthy individuals due to different *and* alternative cortical areas being activated. To make sense of the visual speech information on behavioural level, we need both behavioural and cortical data simultaneously. In the next and final sub-section, we are presenting some possibilities thus predictions on how the auditory-visual speech integration may be happening in the bipolar individuals.

3.2. Physiological possibilities and predictions

As stated above, our data provides a behavioural explanation and cannot explain the finding that the two bipolar groups and

the controls do not significantly differ on auditory-visual speech integration, but the depressive-episode group's lipreading performance is much poorer than the other two groups. With this behavioural data, the challenge that we need to make sense of how visual-only input is integrated in those in the depressive-episode remains. In this final section, we describe some possible cortical activations that may account for the auditory-visual speech integration in the bipolar disorder, hopefully paving the way for some preliminary hypotheses. Because we believe that this is the terminal stop of the behavioural data and that next step in this investigation entails physiological data alongside behavioural.

In normal lipreading, an occipital lobe activity is initiated bilaterally, followed by a further cortical activity superior temporal gyrus, adjacent to the auditory cortex and another activity in inferior frontal gyrus (IFG) and/or arcuate fasciculus. One can predict that the auditory-visual speech perception process follows a similar pattern beyond the activation of temporal gyrus. In the case of incongruent auditory-visual speech perception, as in the case of the McGurk effect, for instance, the occipital-lobe activation is followed by processing at inferior frontal gyrus, simultaneously in the cingulate gyrus. The co-processing of the conflicting auditory and visual speech information most likely rapidly activates the IFG and the cingulate gyrus. Once this conflict is neurologically resolved, the speech centre in superior temporal gyrus becomes active [25, 26]. This basically may be a reverse circuitry process tackling and resolving the two conflicting information. Another possibility could be that the activation of the IFG and speech-associated motor areas suppress the perception of McGurk effect thus leading to the perception of auditory input – in perceptual terms, allocating a heavier weighting to the auditory source in probabilistic terms [24]. In simpler terms, the over-activity of the frontal neural circuits may lead to an impossible incoherence thus no integration. All these points remain to be discovered in the next phase of research.

During the depressive episode of the bipolar disorder, the occipital-temporal-frontal connections may be dysfunctional. In other words, the signal from the occipital lobe may not be relayed efficiently (or not at all), leading to a significantly impoverished inability to lip-read. If, on top of this inefficient connection between the occipital lobe and the frontal areas, the frontotemporal areas remain intact, and if there is another path somewhere between occipital-frontal and occipital cingulate gyrus, then the depressive-episode individuals may still be able to integrate auditory and visual speech information. This possibility needs to be explored via imaging data.

Another possibility is that calls for a combination of behavioural and physiological data is as follows. Assuming that the frontotemporal connections are weakened in the dominant hemisphere, the occipital-frontal connection in the non-dominant hemisphere may become active in a compensatory way paving the way for IFG activation thus lipreading. Then the signal may subsequently be relayed to dominant temporal cortical areas. Another possibility – out of many – is that given the strong occipito-occipital connections, the non-dominant occipital cortex links up to the non-dominant frontal cortex ipsilaterally, stimulating a response to the conflicting McGurk stimulus [27]. Then it relays contra-laterally to the frontal lobe arriving in the left temporal lobe – the speech areas, thus the illusory percept.

Methodologically, there are possible ways of testing the above possibilities. For instance, using transcranial magnetic stimulation (TMS), the right frontotemporal areas of the depressive bipolar individuals can be stimulated. If this stimulation interrupts the McGurk effect, then we may be able to discover the pathways involved in processing auditory-visual speech perception in the bipolar individuals. As a cross-check of this, the stimulation of left frontotemporal areas disturbs the McGurk effect in healthy controls, then the above prediction may be verified. Other imaging devices such as fMRI or fNIRS can actually shed light in varying details on the pathways involved in both healthy and disordered processing of both auditory-visual and lip-read speech information.

4. Acknowledgements

We would like to thank the anonymous participants. We also extend our gratitude to our talkers for the stimuli Mehtap Işkın and Tolga Hasan Kaçan. We are also indebted to Prof. Julian Saurin, formerly of METU Cyprus and currently of the Isle of Harris, who provided insights from a philosopher's and a political scientist's perspective on both this paper and an earlier version of the data published previously in the journal *Languages*.

5. References

- [1] H. McGurk and J. MacDonald, "Hearing lips and seeing voices", *Nature*, vol.264, pp.746-748, 1976.
- [2] D. Erdener, "Türkçede McGurk Illüzyonu / The McGurk Illusion in Turkish". *Turkish Journal of Psychology*, vol.30, pp.19-27, 2015.
- [3] M. Sams, P. Manninen, V. Surakka, P. Helin, and R. Kättö, "McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context", *Speech Communication*, vol. 26, no.1-2, pp. 75-87, 1998.
- [4] S. Jerger, M.F. Damien, N. Tye-Murray and H. Abdi, "Children use visual speech to compensate for non-intact auditory speech", *Journal of Experimental Child Psychology*, vol. 126, pp. 295-312, 2014.
- [5] K. Sekiyama and D. Burnham, "Impact of language on development of auditory-visual speech perception", *Developmental Science*, vol. 11, 306-320, 2008.
- [6] K. Sekiyama and Y. Tohkura, "Cross-language differences in the influence of visual cues in speech perception", *Journal of Phonetics*, vol. 21, pp. 427-444, 1993.
- [7] D. Erdener, "Basic to Applied Research: Benefits of Audio-Visual Speech Perception in Teaching Foreign Languages", *The Language Learning Journal*, vol. 44, no. 1, pp. 124-132, 2016.
- [8] D. Erdener, "Teaching Turkish as a foreign language: extrapolating from experimental psychology", *Journal of Language and Linguistic Studies*, vol. 13, no. 1, 156-165, 2017.
- [9] P. Arnold and A. Köpsel, "Lip-reading, reading and memory of hearing and hearing-impaired children", *Scandinavian Audiology*, vol. 25, pp. 13-20, 1996.
- [10] B. Dodd, B. McIntosh, D. Erdener and D. Burnham, "Perception of the auditory-visual illusion in speech perception by children with phonological disorders", *Clinical Linguistics & Phonetics*, vol. 22, pp. 69-82.
- [11] A. Meronen, K. Tiippana, J. Westerholm, and T. Ahonen, "Audiovisual speech perception in children with developmental language disorder in degraded listening conditions." *Journal of Speech, Language, and Hearing Research*, vol. 56, pp. 211-221.
- [12] D. Pearl, D. Yodanis-Porat, N. Katz, A. Valevski, D. Aizenberg, M. Sigler, A. Weizman, and L. Kikinzon, Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups, *Comprehensive Psychiatry*, vol. 50, pp. 186-192, 2009.

- [13] T. White, R.L. Wigton, D.W. Joyce, T. Bobin, C. Ferragamo, N. Wasim, S. Lisk, and S. Shergill. "Eluding the illusion? Schizophrenia, dopamine and the McGurk effect". *Frontiers in Human Neuroscience*, vol. 8, p. 565, 2014.
- [14] S.A. Surguladze, G.A. Calvert, M. J. Brammer, R.Campbell, E.T. Bullmore, V. Giampietro, and A. S. David, "Audio-visual speech perception in schizophrenia: an fMRI study", *Psychiatry Research: Neuroimaging*, vol. 106, pp. 1–14, 2001.
- [15] G. A. Calvert., E.T. Bullmore, M. J. Brammer, R. Campbell, S. C. R. Williams, P.K. McGuire, P.W. R. Woodruff, S.D. Iversen, and A.S. David. "Activation of auditory cortex during silent lipreading". *Science*, vol. 276, pp. 593–596, 1997.
- [16] G.R. Szycik., T.F. Münte, W. Dillo, B. Mohammadi, A. Samii, and D. E. Dietrich. "Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study". *Schizophrenia Research*, vol. 110, pp. 111–18, 2009.
- [17] X. Delbeuck, F. Collette, and M. van der Linden, "Is Alzheimer's disease a disconnection syndrome? Evidence from a cross-modal audio-visual illusory experiment." *Neuropsychologia*, vol. 45, pp. 3315–3323, 2007.
- [18] S. Schelinski, P. Riedel, and K. von Kriegstein. "Visual abilities are important for auditory-only speech recognition: Evidence from autism spectrum disorder." *Neuropsychologia*, vol. 65 pp. 1–11, 2014.
- [19] R. A. Stevenson., J. Siemann, T.G. Woynaroski, B.C. Schneider, H.E. Eberly, S. Camarata, and M.T. Wallace, "Brief report: Arrested Development of Audiovisual Speech Perception in Autism Spectrum Disorders.", *Journal of Autism and Developmental Disorders*, vol. 44, pp. 1470–1477.
- [20] L. Clark, S. D. Iversen and G. M. Goodwin, "Sustained attention deficit in bipolar disorder", *British Journal of Psychiatry*, vol. 180, pp. 313-319, 2002.
- [21] G. Goodwin, and Gary Sachs, "Fast facts: Bipolar Disorder" 2nd ed. Abingdon: HEALTH Press.
- [22] N. Altieri and Cheng-Ta Yang, "Parallel linear dynamic models can mimic the McGurk effect in clinical populations", *Journal of Computational Neuroscience*, vol. 41, pp. 143–55, 2016.
- [23] D. Burnham and B. Dodd, "Auditory–visual speech integration by prelinguistic infants: Perception of an Emergent Consonant in the McGurk Effect", *Developmental Psychobiology*, vol. 45, no. 4, pp. 204-220.
- [24] D. W. Massaro, "Speech perception by ear and eye: A paradigm for psychological inquiry", Psychology Press: New York, NY, 1987.
- [25] R. Campbell, "The processing of audio-visual speech: empirical and neural bases". *Philosophical Transactions*, vol. B 363, pp. 1001–1100, 2007.
- [26] T. Murakami, M. Abe, W. Wiratman., J. Fujiwara, M. Okamoto, T. Mizuochi-Endo, T. Iwabuchi, M. Makuuchi, A. Yamashita, A. Tiksnadi, F.-Y. Chang, H. Kubo, N. Matsuda, S. Kobayashi, S. Eifuku, and Y. Ugawa1, "The motor network reduces multisensory illusory perception", *The Journal of Neuroscience*, vol. 38, no. 45, 9688-9679, 2018.
- [27] L.M. Fernandez , E. Macaluso and S. Soto-Faraco, "Audiovisual integration as conflict resolution: the conflict of the McGurk illusion", *Human Brain Mapping*, vol. 38, 5691–5705, 2017.