

## Auditory-visual speech perception in bipolar disorder: a preliminary study

Arzu Yordamlı<sup>1</sup>, Doğu Erdener<sup>2, §</sup>

<sup>1</sup> Clinical Psychology Graduate Program, Department of Psychology, Near East University

Near East Boulevard, Nicosia 99138 KKTC via Mersin 10, Turkey

<sup>2</sup> Psychology Program, Middle East Technical University, Northern Cyprus Campus,

Kalkanlı, Güzelyurt (Morphou) 99738 KKTC via Mersin 10, Turkey

§Corresponding author, e-mail: [vdogu@metu.edu.tr](mailto:vdogu@metu.edu.tr)

## Author Note

## Abstract

The focus of this study was to investigate how individuals with bipolar disorder integrate auditory and visual speech information compared to non-disordered individuals and whether there were any differences in auditory and visual speech integration in the manic and depressive episodes in bipolar disorder patients. It was hypothesized that bipolar groups' auditory-visual speech integration would be less robust than the control group. Further, it was predicted that those in the manic phase of bipolar disorder would integrate visual speech information more than their depressive phase counterparts. To examine these, the McGurk effect paradigm was used with typical auditory-visual speech (AV) as well as auditory-only (AO) speech perception on visual-only (VO) stimuli. Results. Results showed that the disordered and non-disordered groups did not differ on auditory-visual speech (AV) integration and auditory-only (AO) speech perception but on visual-only (VO) stimuli. The results are interpreted to pave the way for further research whereby both behavioural and physiological data are collected simultaneously. This will allow us understand the full dynamics of how, actually, the auditory and visual (relatively impoverished in bipolar disorder) speech information are integrated in people with bipolar disorder

*Keywords:* auditory-visual speech perception, bipolar disorder, speech perception

## Auditory-visual speech perception in bipolar disorder

Speech perception is not solely an auditory phenomenon but an auditory-visual one as initially evidenced in noisy listening conditions by Sumbly and Pollack (1954, whose results were recently verified via a comparison between screened and live talking face stimuli) and later, in clear listening conditions by the McGurk Effect (McGurk & MacDonald, 1976). In a typical demonstration of the McGurk effect, an auditory syllable /ba/ dubbed onto the lip movements for /ga/ is often perceived as /da/ or /tha/. Thus, this illusory effect unequivocally shows that speech perception involves visual speech information in the form of lip and mouth movements. McGurk effect is described in the literature as a demonstration of how humans integrate auditory and visual speech information to yield a single percept. As such, the study of speech perception process in terms of the relationship between auditory and visual information gives us a thorough understanding of the phenomenon. Not only did the McGurk effect show the role of visual speech information in clear listening conditions, but more importantly, has come to be used as widespread research tool that measures the degree to which visual speech information influences the resultant percept and the degree of auditory-visual speech integration. The effect is very robust and most participants perceive the effect and it has come to be used as –almost – a generic metric to measure the influence of visual speech information in auditory-visual speech perception research. McGurk effect is an auditory-visual speech illusion the responses to which vary greatly across languages. The extent to which visual speech information affects the resultant percept varies such that in some languages the effect is observed robustly such as English, Italian (Bovo, Ciorba, Prosser & Martini, 2009) and Turkish [REDACTED] but not readily so in some others such as Japanese and Mandarin (Sekiyama & Tohkura, 1993; Sekiyama, 1997; but also see Magnotti et al, 2015). Further, there are cross-

language differences coupled with developmental factors such as language-specific speech perception (see Burnham, 2003) in perceiving McGurk effect thus the degree of auditory and visual speech integration (e.g., Sekiyama & Burnham, 2008; ██████████). The degree to which visual speech information is integrated into the auditory information appears also to be a function of age. While the (in)coherence between the auditory and visual speech components are detectable in infancy (Kuhl & Meltzoff, 1982), the McGurk effect is also evident in infants (Burnham & Dodd, 2004; Desjardins & Werker, 2004), visual speech influence increases with age (McGurk & MacDonald, 1976; Desjardins, Rogers, & Werker, 1997; Sekiyama & Burnham, 2008) as a result of a number of factors such as language-specific speech perception – relative influence of native over non-native speech perception (██████████). Unfortunately, there is a shortage of research in auditory-visual speech perception in the context psychopathology. In the context of speech pathology and hearing, we know that children and adults with hearing problems tend to utilize visual speech information more than their hearing counterparts (Arnold & Köpsel, 1996). Using McGurk stimuli, ██████████ (2008) tested three groups of children: those with delayed phonological acquisition, those with phonological disorder and those with normal speech development. The results showed that children with phonological disorder had greater difficulty in integrating auditory and visual speech information. This shows that the extent to which visual information is used can have the potential to be used as an additional diagnostic and prognostic metric in both recognition and treatment of speech disorders. Auditory-visual speech perception in mental disorders is almost a completely uncharted area. Few, and rather scattered studies with no clear common focus, have emerged recently. These studies in the context of different mental disorders or developmental disabilities demonstrate a paucity of auditory-visual speech

integration (see below). This particular area of research is of specific importance from both pure and applied science sense. In an applied sense, an understanding of why people with these disorders mentioned here have difficulty in auditory-visual speech integration has the potential to render us with additional behavioural criteria and diagnostic and prognostic tools. Thus this preliminary research has both basic and applied motives and justifications. In one of the very few auditory-visual speech perception studies with clinical cases, schizophrenic patients showed difficulty integrating visual and auditory speech information and the amount of illusory experience was inversely related with age thus with chronicity of the illness (Pearl et al., 2009; White et al., 2014). These auditory-visual speech perception differences between control and schizophrenic perceivers were shown to be salient at cortical level as well. It was, for instance, demonstrated that while silent speech (i.e., a visual-only speech or 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 (Surguladze et al., 2001; also see Calvert et al, 1997). Evidence also suggests that the problem in auditory-visual speech integration was due a dysfunction in the motor areas (Szycik et al., 2009). Such auditory-visual speech perception discrepancies were found in other mental disorders, too. For instance, Delbeuck, Collette and Linden (2007) reported deficits in auditory-visual speech integration in Alzheimer's disease patients and with a sample of Asperger's Syndrome individuals, Schelinski, Riedel and von Kriegstein (2014) found a similar result. In addition, Stevenson et al (2014) found that the magnitude of deficiency in auditory-visual speech integration was relatively negligible at earlier ages, with the difference becoming much greater with increasing age. A comparable developmental pattern was also the case with a group of children with developmental language disorder (Meronen, Tiippana, Westerholm, & Ahonen,

2013). In this investigation, we attempted to study the status of auditory-visual speech perception in the context bipolar disorder – a disorder characterized by alternating and contrastive episodes of mania and depression. The paucity of data from clinical populations disallows us to advance literature-based, clear-cut hypotheses. So we adopted the following aims: (a) to preliminarily investigate the status of auditory-visual speech perception in bipolar disorder; (b) to determine whether, if any, differences exist between bipolar-disordered individuals in both manic and depressive episodes. In par with these aims, we predicted that (1) based on previous research with other clinical groups, the control group here should give more visually-based/integrated responses to the auditory-visual (AV) McGurk stimuli than their bipolar-disordered counterparts; (2) if the auditory and visual speech information are fused at behavioural level as a function of attentional focus, and excessive goal-directed behavior (Goodwin & Sachs, 2010), then bipolar participants in the manic episode should give more integrated responses than the depressive subgroup. We based this latter prediction on the anecdotal evidence (in the absence of empirical observations) as reported by several participants here that the patients are almost always focus on tasks of interest when they go through a manic episode whereas they report relatively impoverished attention to tasks during the depressive phase of the disorder.

## Method

### Participants<sup>1</sup>

A total of 44 participants (14 females, 30 males,  $M_{age}=28.8$  years,  $SD=10.2$ ) were recruited after removing missing data and also data from outliers. The bipolar disorder sample consisted of 22 in-patients at Manisa Mental Health Hospital in Turkey. Of these, 12 were in the

manic ( $M_{age}=30.9.9$  years,  $SD=6.78$ ) and 10 were in the depressive ( $M_{age}=41.5$  years,  $SD=11.3$ ) episodes at the time of testing. As in-patients all bipolar disordered subjects were on valproate and/or lithium-based medications. A further 22 healthy participants ( $M_{age}=21.8$  years,  $SD=1.26$ ) were also recruited from amongst volunteers at Middle East Technical University, Northern Cyprus Campus. All participants were native speakers of Turkish with normal hearing and normal/corrected-to-normal vision. Furthermore, no psychiatric or previous mental health disorders were reported by the members of the control group. A written informed consent was obtained from each participant.

## Materials and Procedure

The McGurk stimuli were created by using words and non-words spoken by two native speakers of Turkish – a male and female talker. These material were created for an earlier study [REDACTED]. The raw stimuli were then edited to auditory- visual (AV), auditory-only (AO), and visual-only (VO) stimuli. The AV stimuli were created by dubbing incongruent auditory components onto video components; (e.g. Auditory /soba/ + Visual /soga/ “soda”) and fusion of the auditory and visual components were planned to yield a real word (see the preceding example). This allowed us recognise whether a perceiver’s judgement was visually based or not.

The AO and VO stimuli were created by deleting visual or auditory portions. All AO stimuli were real words. There were a total of 24 AV, 12 AO and 12 VO stimuli. Participants were instructed to “watch and listen” to the video file presented in each trial. The sound level was around 65 dB and testing sessions were completed in a quiet, comfortable room provided by the hospital administration. Responses from participants were manually recorded by the experimenter. A computer-based response collection method did not appear feasible given the



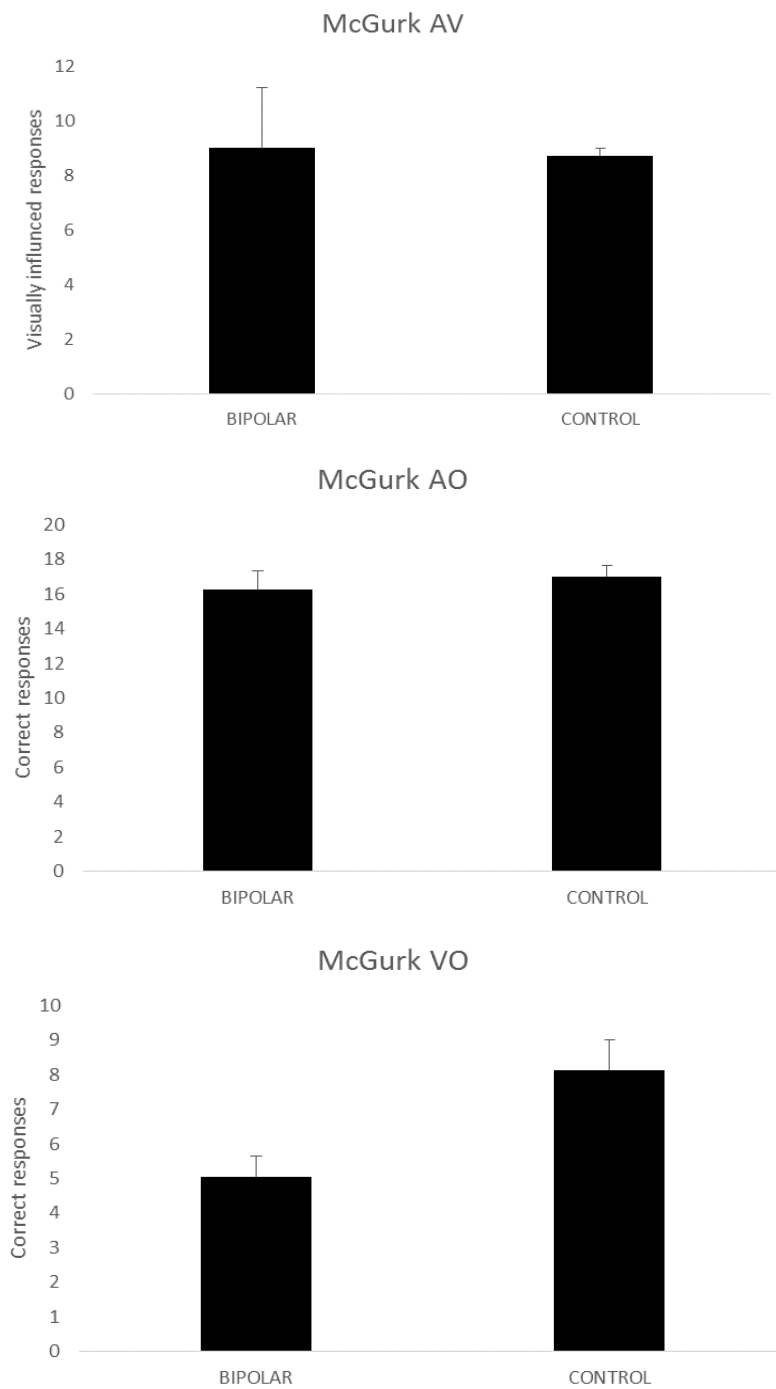
delicate nature of the experimental group thus to avoid any potential task-related burdens. The test phase was preceded by a familiarization trial in each experimental condition. None of the participants had any difficulty in comprehending or completing the task requirements.

## Results

Two sets of statistical analyses were conducted in this study: a comparison of disordered and non-disordered groups by means of a t test analysis, and a comparison of the disordered subgroups, namely, those with bipolar disorder going through manic versus depressive stages versus the non-disordered control group using the non-parametric Kruskal-Wallis test due to small sample size.

### The t-test analyses

A series of independent t tests were conducted on AV, AO, and VO scores comparing the bipolar-disordered and the control groups. The homogeneity of variance assumptions as per Levene's test for equality of variances were met for all except McGurk AO scores ( $p = .029$ ). Thus for the AO variable the values for when equal variances were not met are reported. The results revealed no significant differences between the disordered and non-disordered groups over the AV,  $t(42) = .227, p = .82$ , and AO scores,  $t(35.602) = -.593, p = .56$ . However, in the VO condition, the non-disordered group performed better than their disordered counterparts,  $t(42) = -2.882, p < .005$ . The mean scores for these measures are presented in Figure 1.



*Figure 1.* The mean scores of AV (upper panel), AO (central panel) and VO (lower panel) scores obtained from the bipolar groups combined and the control group. Error bars represent the standard error of the mean.

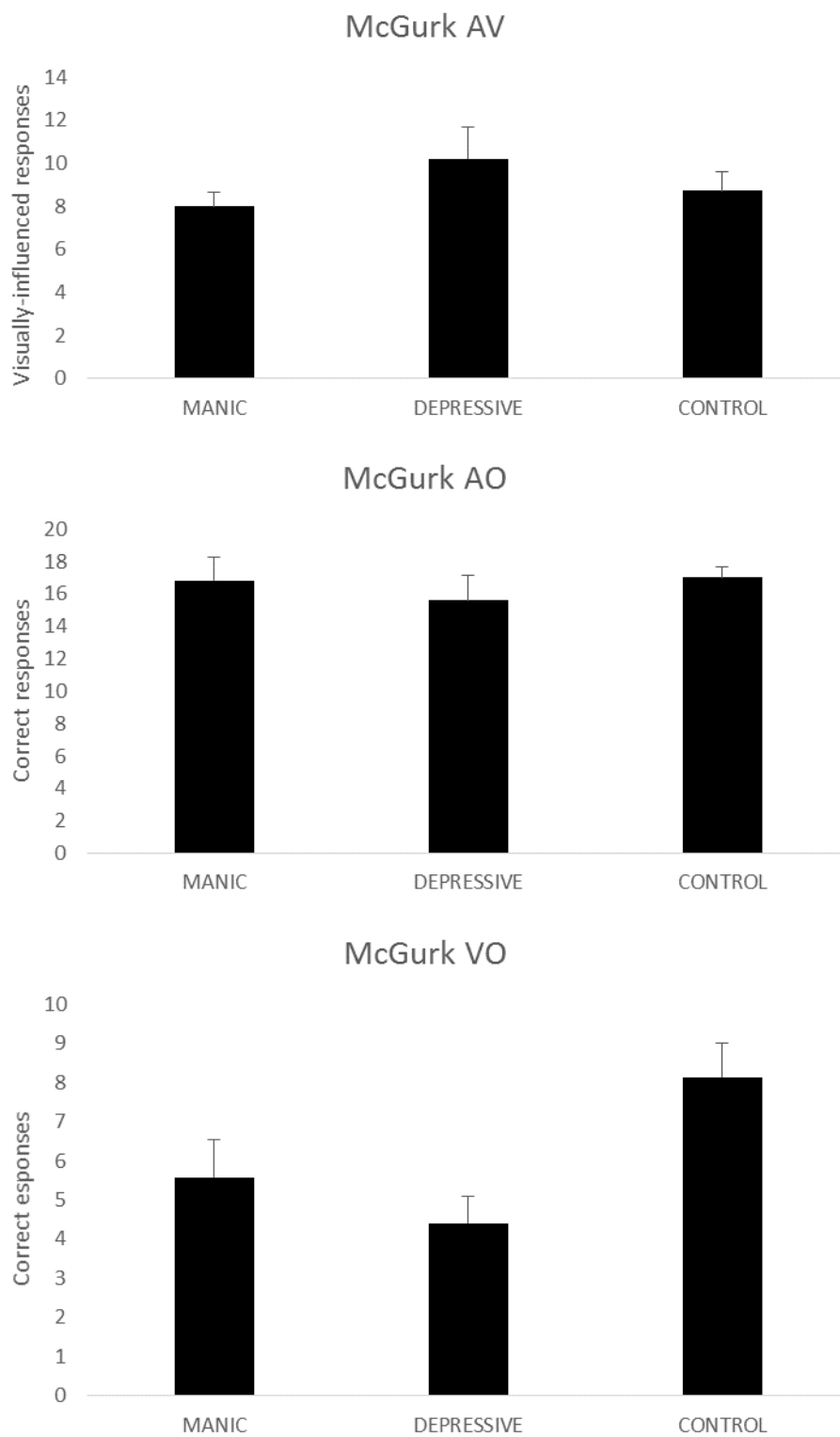
## The Kruskal-Wallis Test

Given the small sample size of groups ( $n < 30$ ), we resorted to three Kruskal-Wallis nonparametric tests to analyse the AV, AO, and VO scores. The analyses of AV,  $\chi^2(2, N = 44) = .804, p = .669$ , and AO scores,  $\chi^2(2, N = 44) = .584, p = .747$ , revealed no significant results. The Kruskal-Wallis analysis of the VO scored produced significant group differences,  $\chi^2(2, N = 44) = 7.665, p = .022$  (see Table 1). As there are no any post-hoc alternatives for Kruskal-Wallis test, we ran two Mann-Whitney U tests with VO scores. The comparison of VO scores between manic and depressive bipolar groups,  $z(N = 22) = -.863, r^2 = .418$  and manic bipolar and the control groups failed to reach significance,  $z(N = 34) = -1.773, r^2 = .080$ . The third Mann-Whitney U comparison between the depressive bipolar and the control groups, on the other hand, showed a significant difference,  $z(N = 32) = -2.569, r^2 = .009$ . Figure 2 presents the mean data for each group on all measures.

Table 1.

*The results of Kruskal-Wallis tests with McGurk auditory-visual (AV), auditory-only (AO), and visual-only (VO) data.*

	McGurk AV	McGurk AO	McGurk VO Median	8.000
18.000	6.500			
<b>Chi-square</b>	.804	.584	7.665	
<b>df</b>	2	2	2	
<b>Sig.</b>	.669	.747	<b>.022*</b>	



*Figure 2* The mean scores of AV (upper panel), AO (central panel) and VO (lower panel) scores obtained from the manic and depressive episode bipolar subgroups and the control group. Error bars represent the standard error of the mean.

## Discussion

It was hypothesized that the bipolar disorder group would be less susceptible to McGurk effect thus exhibit integration of auditory and visual speech information to a lesser extent compared to the healthy control group. Besides, the bipolar manic group was predicted to integrate auditory and visual speech information to a greater extent than their depressive phase counterparts should the auditory-visual speech integration be a process that occurs at the behavioural level requiring attentional resource allocation. The findings of this study did not support these main predictions that there was no significant difference between the control group and the bipolar group/ on the degree of auditory-visual speech integration. No group-based differences were also observed with the AO stimuli, either. On the other hand, lending partial support to the overall predictions, the control group performed overwhelmingly better than the disordered groups – as demonstrated in the analyses of both combined and separated bipolar groups' data – on the VO condition, virtually a lip-reading task. Further, the disordered subgroups did not differ from each other significantly on any of the measures. To sum up, the only significant result was the observation that the control group's VO processing was much more superior to that of the disordered groups but not on AO or AV stimuli as predicted. Although, seemingly, this presents a bit blurry picture – and surely warranting further scrutiny – these results present us with a number of possibilities for interpretation.

In normal integration process of auditory and visual speech information the information coming from these two modalities are integrated on a number of levels such as behavioural (i.e., phonetic vs phonological levels of integration; see Bernstein, Burnham, & Schwartz, 2002) or cortical (Campbell, 2007). Given the finding that the control group did not differ from the disordered groups with respect to the integration of auditory and visual speech information and

auditory-only information, but did on the visual-only information leaves us with the question as to why the visual-only (or lip-read) information acts differently in the disordered group in general.

Given the no difference between the groups with respect to auditory-visual speech integration, it seems that visual information is somehow integrated with the auditory information to a resultant percept whether the perceiver is disordered or not. This seems to suggest that what one needs to look at as to the source of the integration is not limited to behavioural level, thus calling for further scrutinising on multiple levels. That is, we need to look at both behavioural and cortical responses to the AV stimuli to understand how integration works in the bipolar individuals. There are suggestions and models in the literature that explain how the auditory-visual speech integration may be in the non-disordered population. Some studies claim that integration occurs at the cortical level, independent of other levels (Campbell, 2007), while some others argue in favour of a behavioural emphasis for the integration process, that is, phonetic (e.g., Burnham & Dodd, 2004) vs phonological in which the probabilistic values of auditory and visual speech inputs are weighted out such as Massaro's Fuzzy Logical Model of Perception in which both auditory and visual signals are evaluated on the basis of their probabilistic values, then integrated based on those values leading to a percept (e.g., Massaro, 1998). However, given the no difference between the disordered and control groups responses to the AV and AO stimuli, no known models seem to explain the difference between these groups over the VO stimuli.

So how do bipolar disordered perceivers, unlike people with other disorders such as schizophrenia (White et al., 2014) or Alzheimer's disease (Delbeuck et al., 2007) still integrate auditory and visual speech information whilst having difficulty perceiving visual-only speech

information (or lipreading)? On the basis of the data we have here, one can think of three broad possibilities as to how the auditory-visual speech integration occurs yet the visual speech information alone fails to be processed. One possibility is that, given that speech perception is primarily an auditory phenomenon, the visual speech information is carried somehow alongside the auditory information without any integrative process as some data suggests in auditory-dominant models (Altieri & Yang, 2016). One question that remains with respect to the perception of auditory-visual speech in the bipolar disorder context is how the visual speech information is integrated. It appears that, given the relative failure of visual-only speech processing in bipolar disordered individuals, further research is warranted to unearth the mechanism through which the visual speech information is eventually and evidently integrated. Such mechanisms are, as we have so far seen in auditory-visual speech perception research, are multi-faceted, thus calls for multi-level scrutiny, e.g., investigations at both behavioural and cortical levels. The finding that the non-disordered individuals treat visual-only speech information as a speech input whilst the bipolar individuals do not seem to (at least in relative terms), begs the question that whether, in fact, bipolar individuals have issues with the visual speech input per se. Given the no difference over the AV stimuli, the auditory dominance models (e.g., Altieri & Yang, 2016) may account for a possibility that visual speech information arriving with auditory speech information is somehow undermined but the resultant percept is still there as is the case with the non-disordered individuals. In fact, visual-only speech input is treated as speech in healthy population. Using the MRI imaging technique, Calvert and colleagues demonstrated that visual-only (lip-read) speech without any auditory input activated several areas of the cortex whose activation are normally associated with auditory speech input, particularly with superior temporal gyrus. Even more intriguingly, they also found that no

speech-related areas were active in response to faces performing non-speech movements (Calvert et al, 1997). Thus visual-only information is converted to an auditory code in healthy individuals and as far as our data seems to suggest, at behavioural level, this does not occur in the bipolar individuals. On the basis of our data and Calvert et al's findings, we may suggest (or speculate) that the way visual speech information is integrated in bipolar individuals is most likely different to the way it occurs in healthy individuals for different behavioural and/or cortical processes engaged. In order to understand how visual information is processed, we need both behavioural and cortical data to be obtained in real time, simultaneously in response to the same AV stimuli.


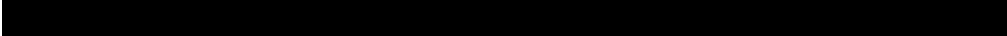

An understanding of how visual speech information is processed both with and without auditory input will provide us with a better understanding of how speech perceptual processes occur in this special population and thus will pave the way to develop finer criteria for both diagnostic and prognostic processes.

Unfortunately, we had limited time, resources thus a small sample size for reasons beyond our control. However, still, the data we present allows us to make the following prediction for subsequent studies: given the impoverished status of visual-only speech perception in the bipolar groups, yet still yielding an eventual AV speech integration, a compensatory mechanism with both behavioural and cortical levels, may be at work. In order to understand this mechanism, as we have just suggested, responses to McGurk stimuli at both behavioural and cortical levels must be obtained via known as well as by means of new type auditory-visual speech stimuli (Jerger, Damian, Tye-Murray & Abdi, 2014).



## References

- Altieri, N. & Yang, C-T. (2016). Parallel linear dynamic models can mimic the McGurk effect in clinical populations. *Journal of Computational Neuroscience*, 1-13.  
doi: 10.1007/s10827-016-0610-z
- Arnold, P., & Köpsel, A. (1996). Lip-reading, reading and memory of hearing and hearing-impaired children. *Scandinavian Audiology*, 25, 13–20.
- Bernstein, L. E., Burnham, D., & Schwartz, J.-L. (2002). Special session: Issues in audiovisual spoken language processing (when, where, and how?). In J.H.L. Hansen & B. Pellom (Eds.), *Proceedings of International Conference on Spoken Language Processing 2002*, 3, 1445-1448.
- Bovo, R., Ciorba, A., Prosser, S., & Martini, A. (2009). The McGurk phenomenon in Italian listeners. *Acta Otorhynolaryngologica Italica*, 29, 203-208.
- Burnham, D. (2003). Language specific speech perception and the onset of reading. *Reading and Writing*, 16(6), 573-609.

- Burnham, D. & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, 45, 204–220.
- Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D. & David, A.S. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276, 593-596.
- Campbell, R. (2007). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions B*, 363(1493), 1001-1010.
- Delbeuck, X., Collette, F., & Van, L. M. (2007). Is Alzheimer's disease a disconnection syndrome? Evidence from a cross-modal audio-visual illusory experiment. *Neuropsychologia*, 45, 14, 3315-3323.
- Deonarine, J.M., Dawber, E.J., & Munhall, K.G. (2012). Sumbly and Pollack revisited: The influence of live presentation on audiovisual speech perception. *The Journal of the Acoustical Society of America* 132, 2080 (2012); doi: <http://dx.doi.org/10.1121/1.4755680>
- Desjardins, R. N., Rogers, J., & Werker, J. F. (1997). “An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks.” *Journal of Experimental Child Psychology*, 66, 85-110.
- Desjardins, R. N. & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental psychobiology*, 45, 187-203.
- 
- 
- 

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Goodwin, G. & Sachs, G. (2010). *Fast facts: Bipolar Disorder* (2<sup>nd</sup> Ed). HEALTH Press: Abingdon, UK.

Jerger, S., Damian, M.F., Tye-Murray, N. & Abdi, H. (2014). Children use visual speech to compensate for non-intact auditory speech. *Journal of Experimental Child Psychology*, 126, 295-312.

Kuhl, P. K. & Meltzoff, A. N. (1982). "The bimodal perception of speech in infancy." *Science*, 21, 1138-1141.

McGurk H. & MacDonald J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748, doi: 10.1038/264746a0.

Magnotti, J.F., Mallick, D.B., Feng, G, Zhou, B., Wen, Z., & Beauchamp, M.S. (2015). Similar frequency of the McGurk effect in large samples of native Mandarin Chinese and American English speakers. *Experimental Brain Research*, 233, 2581-2586.

Massaro, D. W. (1998). *Perceiving talking faces: from speech perception to a behavioural principle*. The MIT Press: Cambridge, MA.

Meronen, A., Tiippana, K., Westerholm, J., & Ahonen, T. (2013). Audiovisual speech perception in children with developmental language disorder in degraded listening conditions. *Journal of Speech, Language, and Hearing Research*, 56, 211-221.

- Pearl, D., Yodashkin-Porat, D., Katz, N., Valevski, A., Aizenberg, D., Sigler, M., & Kikinzon, L. (2009). Differences in audio-visual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups. *Comprehensive Psychiatry*, *50*, 186-192, doi: 10.1016/j.comppsy.2008.06.004.
- Schelinski, S., Riedel, P., & von Kriegstein, K. (2014). Visual abilities are important for auditory-only speech recognition: Evidence from autism spectrum disorder. *Neuropsychologia*, *65*, 2, 1-11.
- Sekiyama, K. (1997). Audiovisual speech perception and its cross-language differences. *Japanese Journal of Psychonomic Science*, *15*, 122–127.
- Sekiyama, K. & Burnham, D. (2008). Impact of language on development of auditory–visual speech perception. *Developmental Science*, *11*, 306–320.
- Sekiyama, K. & Tohkura, Y. I. (1993). Cross-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*, 427–444.
- Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014). Brief report: arrested development of audiovisual speech perception in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*, 1470-1477.
- Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* *26*(2), 212–215.
- Surguladze, S. A., Calvert, G. A., Brammer, M. J., Campbell, R., Bullmore, E. T., Giampietro, V., & David, A. S. (2001). Audio-visual speech perception in schizophrenia: an fMRI study. *Psychiatry Research. Neuroimaging*, *106*, 1-14. doi. 10.1016/S0925-

4927(00)00081-0.

Szycik, G. R., Münte, T. F., Dillo, W., Mohammadi, B., Samii, A., & Dietrich, D. E. (2009).

Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study.

*Schizophrenia Research*, 110, 111-118.

White, T., L Wigton, R., W Joyce, D., Bobin, T., Ferragamo, C., Wasim, N., Lisk, S., & Shergill,

S. (2014). Eluding the illusion? Schizophrenia, dopamine and the McGurk effect.

*Frontiers in Human Neuroscience*, 8, 565. doi:10.3389/fnhum.2014.00565.