BRILL *Multisensory Research (2017)* DOI:10.1163/22134808-00002565 brill.com/msr

Forty Years After *Hearing Lips and Seeing Voices*: the McGurk Effect Revisited

Agnès Alsius *, **Martin Paré and Kevin G. Munhall**

Psychology Department, Queen's University, Humphrey Hall, 62 Arch St., Kingston, Ontario, K7L 3N6 Canada

Received 4 October 2016; accepted 9 March 2017

Abstract

Since its discovery 40 years ago, the McGurk illusion has been usually cited as a prototypical paradigmatic case of multisensory binding in humans, and has been extensively used in speech perception studies as a proxy measure for audiovisual integration mechanisms. Despite the well-established practice of using the McGurk illusion as a tool for studying the mechanisms underlying audiovisual speech integration, the magnitude of the illusion varies enormously across studies. Furthermore, the processing of McGurk stimuli differs from congruent audiovisual processing at both phenomenological and neural levels. This questions the suitability of this illusion as a tool to quantify the necessary and sufficient conditions under which audiovisual integration occurs in natural conditions. In this paper, we review some of the practical and theoretical issues related to the use of the McGurk illusion as an experimental paradigm. We believe that, without a richer understanding of the mechanisms involved in the processing of the McGurk effect, experimenters should be really cautious when generalizing data generated by McGurk stimuli to matching audiovisual speech events.

Keywords

Speech perception, audiovisual integration, the McGurk effect

1. Introduction

Forty years ago, Harry McGurk and John MacDonald published *Hearing lips and seeing voices* (McGurk and MacDonald, 1976), a manuscript in which they described a remarkable audiovisual speech phenomenon that would come to be known as the McGurk illusion or the McGurk effect. The McGurk effect occurs when the visual signal of one phoneme is dubbed onto the acoustic signal of a different phoneme. With specific audiovisual pairs, the observers

* To whom correspondence should be addressed. E-mail: aalsius@gmail.com

 DOI:10.1163/22134808-00002565

do not notice intermodal conflict and often experience (i.e., hear) a phoneme that does not match the actual auditory signal. The discovery of the McGurk effect represented a milestone in the area of speech perception. Not only did it engender new studies investigating the properties of the illusion, but it also boosted an appreciation for the importance of audiovisual speech research.

The McGurk illusion effectively demonstrates that speech perception is not only an auditory process, but can involve the processing of phonetic components across modalities even when the auditory information is intact. For this reason, it has been widely cited as a paradigmatic case of multisensory integration across modalities (4901 Google Scholar citations of the original study, from 1976 to July 2016) and has been extensively used as a proxy measure for audiovisual speech integration. That is, the frequency with which an observer perceives this illusion under different experimental manipulations has been proposed to provide an index as to whether information has been integrated or if, on the contrary, observers perceive one specific sensory modality as dominant.

Using the McGurk illusion as a proxy for audiovisual speech integration is advantageous over other types of indices (e.g., speech in noise tasks, SPIN), as it allows researchers to present short, simple (non-semantic) linguistic stimuli a large number of times, in open or closed set tasks. However, despite the widespread acceptance among researchers of the McGurk illusion as a tool for studying the mechanisms underlying multisensory integration, the illusion varies enormously across studies and there is now accumulating evidence that the processing of McGurk stimuli differs from congruent audiovisual processing in many aspects, both at a phenomenological and neural level. Given that this phenomenon has traditionally been used to quantify the necessary and sufficient conditions under which audiovisual integration occurs, we believe a full understanding of the processes underlying the McGurk effect is required prior to extrapolating the results to naturally occurring (i.e., audiovisual matching) speech (see also Brancazio and Miller, 2005). In this paper, we review some of the practical and theoretical issues related to the use of the McGurk illusion as an experimental paradigm. We set out to ask what general conclusions could be drawn by looking at the published experimental literature for the last 40 years. Our aim to answer the following related questions was quite straight forward:

1. How strong is the effect in the normal adult population? This is important to know for developmental studies, studies of special populations, and in order to be able to understand the relative influence of aspects of vision on auditory speech perception.

2. How variable is the effect? This needs to be answered across individuals (the question of individual differences), within individuals (the question

of perceptual stability), across stimuli (the question of cue strength) and across experimental context.

3. Is the processing of incongruent illusory stimuli supported by the same integration mechanisms as those involved in the processing of audiovisually matching speech events? This is important to determine the suitability of this illusion as a proxy measure for audiovisual integration in natural conditions.

These are fundamental questions about the phenomenon that are essential to its basic understanding but also essential for the use of the McGurk effect as a tool to study multisensory processing in different populations and different tasks.

2. Assessing the Magnitude of the Illusion: Is a Systematic Meta-Analysis of the McGurk Literature Even Possible?

In order to answer the first two questions (i.e., assess the magnitude of the illusion and its variability) we attempted to aggregate the results reported in the literature to conduct a meta-analysis of published data. Out of the 4901 citations of the original study provided by Google Scholar, only 276 were experimental papers that (a) used the McGurk effect as a paradigm, and (b) were published in English peer-reviewed scientific journals. We then defined a modest set of inclusion criteria: (1) Participants must be healthy adults (18–35 years of age). (2) Participants and talkers must be native speakers of English (Note 1). (3) The stimulus pairing must be the classic fusion stimuli (acoustic 'ba'; i.e., A_{ba}, and visual 'ga'; i.e., V_{ga}). (4) The data have to be reported as the percentage of optimally fused response ('da' or 'tha'; note that a large numbers of studies describe the illusion as the emergence of a new percept) (Note 2). (5) Response alternatives should be open set or have an 'other' consonant response category. Finally, (6) the results should report means and variance estimates for the responses. Remarkably, the first five criteria reduced the number of studies from 276 to 21, showing the enormous heterogeneity in the paradigms using the illusion. This was one of the major factors that accounted for the impossibility to conduct a meta-analysis. The second factor was that out of the 21 studies, only two provide tables with means and standard deviations. It is not uncommon in systematic reviews to cast a net broadly and thus begin with a very large number of publications. The number gets reduced significantly when strict inclusion criteria for a review are defined. For example, Luckner *et al.* (2006) in a study of 40 years of literacy research in deaf education began with 964 articles but were left considering 22 papers that met their inclusion criteria. However, it is apparent that for an illusion that is so well known only a modest amount of actual experimental research has been

done using the McGurk effect. Further, much of the research that has been done does not permit any comparison between studies. We believe the impossibility of synthesizing information from published studies is illustrative of the state of the field. In the following section we summarize some of the factors that restrict our ability to systematically review the field and provide even preliminary answers to fundamental questions about the nature of the illusion.

3. Issues Related to Assessing the Magnitude and Variability of the Illusion

Previous literature using the McGurk effect has reported large variability in the incidence of the illusion, ranging from approximately 20 to 98% across studies (e.g., Nath and Beauchamp, 2012; Sekiyama and Tohkura, 1991). Indeed, some studies using the exact same stimuli have even shown different impact of the illusion across experiments (e.g., see Rosenblum and Saldaña, 1992; from 96% in Experiment 1 to 79 % in Experiment 2; or McGurk and MacDonald, 1976 and MacDonald and McGurk, 1978, for 98% and 64% of adults reporting 'da' responses for the $A_{ba\text{-}ba} + V_{ga\text{-}ga}$ stimulus pairing, respectively). Moreover, contrary to what was claimed in the original report (McGurk and MacDonald, 1976, p. 747), the effect is usually only observed on a percentage of trials over the course of an experiment (Brancazio, 2004; Brancazio and Miller, 2005; Massaro and Cohen, 1983). The large variability in the magnitude of the McGurk effect can be attributed to a number of factors. In the rest of this section, we summarize the main sources of variance in the illusion, namely, the definition of what constitutes an illusory percept, the quality of the auditory and visual stimuli, the specific audiovisual stimulus pairing used in the different studies, the different instructions, response structure and scoring methods, and the large inter-subject variability.

3.1. Definition and Quantification of the Effect

Despite numerous studies employing the McGurk effect as a paradigm, there is still no clear agreement among researchers on two critical aspects related to the illusion. Namely, (1) how to best define it and, (2) how to best validate the visual influence (i.e., what control or comparison group is used to demonstrate the illusion).

The classic, 'conservative' definition of the McGurk illusion requires that in the case of fusions "the information from the two modalities is transmitted into something new with an element not presented in either modality" and for combinations a "composite comprising relatively unmodified elements from each modality" be reported (McGurk and MacDonald, 1976).

However, this characterization excludes a broad range of responses to incongruent audiovisual speech stimuli. For this reason, many researchers have

opted to use a more flexible definition of the effect, including as illusory percepts those instances in which the visual information overrides the auditory component (e.g., reporting 'ga' in response to $A_{ba} + V_{ga}$ or 'va' in response to $A_{ba} + V_{va}$; e.g., Colin *et al.*, 2002; Rosenblum and Saldaña, 1992; Sams *et al.*, 1998) or even any instance in which the observer's response deviates from the auditory component of the audiovisual paring (e.g., Jordan *et al.*, 2000; Wilson *et al.*, 2016). This latter scoring method is the most liberal and the least theory-bound approach. Indeed, in a recent manuscript, Tiippana (2014) stressed the importance of defining the McGurk illusion as "a categorical change in auditory perception induced by incongruent visual speech, resulting in a single percept of hearing something other than what the voice is saying", because, as stated by the author, this definition includes all variants of the illusion.

While the exact definition of the illusion might seem trivial, it can have a large impact on the experimental results. In the more conservative definition of the effect, the McGurk illusion will be represented by the choice of a single response ('da' in case of $A_{ba} + V_{ga}$), whereas in the more flexible conception, the effect will be represented by the absence of a certain response choice (not /b/). This means that there are far more choices that indicate perception of the McGurk effect in the latter than the former. This may explain, in part, the large differences in reported magnitudes of the effect in different studies.

A second point, closely related to how the illusion is defined, involves the different methods researchers have used to validate the illusion. That is, the researcher must always consider the McGurk reports are the result of genuine audiovisual interactions or simply reflect cases of mistaken auditory and/or visual identities. For instance, in the classic McGurk pairing ($A_{ba} + V_{ga}$ = 'da'), if the auditory signal was compromised and participants sometimes misheard 'da' when presented with A_{ba} (i.e., auditory only condition), then one could not attribute the fused responses to the common information that is present in both modalities, but would rather have to account for these auditory errors. Unimodal baseline conditions are thus critical to ensure that illusory reports are explained by genuine integration mechanisms. Indeed, some researchers have used an error-adjusted McGurk effect to quantify their data. In this correction method, auditory misidentifications in the auditory baseline control condition are subtracted from the total number of illusory responses for each participant: (Illusory percepts = Fused percepts − Auditory misidentifications); Grant and Seitz's corrective method (1998); Colin *et al.*, 2002; Desai *et al.*, 2008). Instead of a unimodal auditory condition, when using a more liberal definition of the effect, some researchers have opted for an audiovisual congruent baseline (e.g., Jordan and Sergeant, 2000; Wilson *et al.*, 2016). This baseline shows the difference in auditory perception when supportive *versus* unsupportive cues are visually present.

Finally, as mentioned before, some researchers have considered visually-dominant responses as 'fused' percepts (i.e., $A_{ba} + V_{ga}$ = 'ga'). However, in these cases, it remains unclear if perceivers are truly integrating the two modalities or if they are perceiving some discrepancy between the two modalities and are simply reporting the modality with the least ambiguous signal (see Cienkowski and Carney, 2002; Tiippana, 2014). In order to reduce this bias, in experiments where visual-dominant responses are considered fused percepts, experimenters should instruct participants to report "what they heard" (rather than "what the talker said") and should include visual-only baselines as an independent index for assessing integration (Massaro, 1998).

3.2. Quality of the Auditory Information

Stimulus factors, such as the prominence of the auditory and visual signals, appear to have a large and difficult to assess impact on the magnitude of the McGurk effect. This issue is the most important challenge to understanding the McGurk effect and it is an issue that applies to all studies of audiovisual speech perception. The stimuli are complex, poorly controlled and minimally described in all experiments. In spite of this broad problem, some generalizations can be made about the stimuli used in experiments on the McGurk effect.

When considering the quality of the auditory information, data suggests that the McGurk fusion effect is strongest with a weak auditory consonant. It is not surprising that the acoustic stop consonant /b/ is often used because it is acoustically confusable with other voice stops. In fact, place of articulation is one of the weakest acoustic features in speech (Miller and Nicely, 1955) and the McGurk effect is largely an illusion of visual dominance of acoustic place of articulation cues. The effect can also be enhanced by weakening the auditory component, either by decreasing sound intensity (Colin *et al.*, 2002; though see Green *et al.*, 1988), by increasing acoustic noise (Alm *et al.*, 2009; Fixmer and Hawkins, 1998; Hardison, 1996; Jordan and Sergeant, 1998; Sekiyama and Tohkura, 1991), and by manipulating talker intelligibility (Fixmer and Hawkins, 1998) or acoustic cues (Green and Norrix, 1997). Green and Norrix (1997) explored how the three acoustic cues to place of articulation of the auditory component (i.e., release bursts, aspiration, and the voiced format transitions) contributed to the McGurk effect. They found that removing bursts and aspiration from the acoustic signal did not affect the magnitude of fusion responses but significantly decreased combination responses, an asymmetry likely explained by the largest impact of the removal of the cues in velar as compared to bilabial tokens. Increasing the acoustic energy of these cues had no impact on the magnitude of illusory percepts, suggesting that the McGurk effect for fusion stimuli does not occur simply because the release

bursts and aspiration are weak. Low-pass filtering the second and higher formant transitions, however, made the illusion significantly stronger, suggesting that the dynamic information contained within these formants could be critical for obtaining the McGurk effect.

3.3. Quality of the Visual Information

The quality of the image has also been manipulated by adding noise (Fixmer and Hawkins, 1998), by using spatial quantization (a reduction of the number of pixels in the image; Campbell and Massaro, 1997; MacDonald *et al.*, 2000) or spatial frequency filtering (Wilson *et al.*, 2016). The results of these studies show that the illusory effect decreases monotonically as the visual resolution decreases. Overall, these studies suggest that fine facial detail is not critical for the McGurk effect to occur (note, however, that fine facial detail has been shown to be important for some participants in SPIN tasks; see Alsius *et al.*, 2016). These results are consistent with eyetracking studies showing that the magnitude of the illusory percepts remains relatively unaltered when the face is presented in peripheral vision (Paré *et al.*, 2003).

In this line, other studies have shown that speech perception is reduced — but remains effective — across horizontal viewing angles (Jordan and Thomas, 2001), rotations in the picture plane (Jordan and Bevan, 1997; Massaro and Cohen, 1996), face size (Jordan and Sergeant, 1998), when removing the color of the talking face (Jordan *et al.*, 2000), or when introducing gross variations in the facial configuration (Hietanen *et al.*, 2001; though see Eskelund *et al.*, 2015). Studies occluding different parts of facial regions have found that presenting the lips, tongue and teeth alone is sufficient to elicit a significant proportion of McGurk illusions (Thomas and Jordan, 2004). The effect can also be elicited with computer-generated faces (Massaro and Cohen, 1990).

Overall, these studies suggest that the McGurk illusion is quite robust to stimulus degradation, and that the proportion of illusory percepts increases when the sensory degradation is auditory and it decreases (i.e., less McGurk) when it is visual.

3.4. Quality of the Talker

One puzzling aspect for audiovisual speech integration researchers using McGurk stimuli is that, even when the auditory and the visual channels are not degraded and are accurately dubbed, the illusion sometimes fails to occur for some stimuli. That is, some talkers produce utterances that lead to much weaker illusory percepts (across subjects) than others (Munhall *et al.*, 1996; Sekiyama, 1998). Talkers are selected by convenience from acquaintances or from members in a laboratory and filmed. The visual and acoustic variance in their productions is unknown and there is no simple metric to assess visual and auditory intelligibility of tokens and compare this to the population at

large. Basu Mallick *et al.* (2015) measured illusory responses to one particular stimulus pair ($A_{ba} + V_{ga}$) pronounced by 12 different talkers in a sample of 165 English-speaking adults. All these materials had been used in previously published studies. They found large differences in how frequently the illusion was reported across different talkers (17% to 58%). Interestingly, they showed that for individual stimuli, responses strongly deviated from normality, with 77% of participants almost never ($\leqslant$10%) or almost always ($\geqslant$90%) experiencing the illusion. Based on these results, the authors claim that the mean response frequency and the parametric statistical tests commonly run in the field to analyse McGurk data are invalid.

The particular characteristics that promote (or preclude) audiovisual integration from certain talkers, however, remain unknown. Potential possibilities that might explain talker variations in producing McGurk stimuli are clarity of articulation and speech rate. These factors are known to have a large impact on the boundaries of viseme groups, on the visual gain in speech in noise tasks, and on auditory intelligibility (Demorest and Bernstein, 1992; Gagné *et al.*, 1994). It has been shown, however, that the amount of integrated percepts for a particular talker cannot be predicted by the single-modality performance for that talker (Ver Hulst, 2006). In other words, a talker who is easily speechreadable will not necessarily produce stronger McGurk percepts.

Jiang and Bernstein (2011) carried out the most exhaustive study examining how physical stimulus characteristics account for the perceptual responses distributions of McGurk stimuli (i.e., fusion, combination, auditory dominant, visual dominant). They did this by extracting the perceptual response, and the auditory (i.e., acoustic line spectral pairs) and visual (kinematics from three-dimensional motions of retro-reflectors glued to the talkers faces) information from each individual token of a large set of stimuli. The stimuli consisted of the acoustic recordings /bA/ or /lA/ dubbed onto video stimuli /bA, dA, gA, vA, zA, lA, wA, thA/, articulated by four different talkers and edited with three different alignment methods (i.e., consonant-onset, vowel-onset, and minimum acoustic-to-phantom distance). The results showed that AV stimulus alignment was not a significant factor. They computed three general measures that could potentially account for responses distributions: mutual information (i.e., a measure of shared data structure information), the distance between the presented acoustic signal and the original acoustic signal recorded with the presented video, and the correspondence between the presented auditory and visual information. Regression analyses indicated that, across talkers, 52% of the variance in fusion responses was accounted for by the physical measures. They found that fusion responses increased with low correspondence and with a smaller minimum distance between the presented and the original auditory stimuli. Mutual information was not important for accounting for the four perceptual responses.

Further studies are required to investigate the specific talker's characteristics leading to stronger illusions, such optical flow analyses of visible articulation (lips and jaw movement, etc.), measures of the acoustic properties of the talker's utterances (e.g., formant values and transitions) as well as the temporal dynamics of the two channels. Without concerted efforts to calibrate our stimuli in this way, little progress can be made in understanding multisensory integration of speech. At the very least, we should encourage journals to take up the suggestion of the late Christian Benoit that a condition of acceptance for audiovisual speech publication should be that the stimuli be made available as supplementary material. This would allow the field to examine the multiple ways that stimuli can differ and also allow future analytic work to include the studies that have gone before.

3.5. Different Task Instructions, Task Structure, Response Structure and Scoring Methods

In addition to the complete lack of control of stimulus characteristics, other experimental factors add extraneous variance to the reported results. In the McGurk literature, one can find a variety of task instructions given to participants. Instructions can generally be comprised in two categories: "report what the talker said" (e.g., Hillock-Dunn *et al.*, 2016; White *et al.*, 2014) or "report what you heard" (e.g., Green *et al.*, 1991; Jordan and Thomas, 2001). While subtle, these differences in instructions could potentially have an impact on where participants focus attention, consequently impacting the incidence of the effect (Buchan and Munhall, 2011; Massaro, 1998; Summerfield and McGrath, 1984). For instance, instructing participants to report what they heard might potentially increase the perceptual weight of the auditory cue (or attenuate the weight of the visual cue) and thus lead to less illusory percepts (Colin *et al.*, 2005).

Task structure, or the order in which trials are presented, might also add an extraneous source of noise. Stimuli are often presented continuously with only short pauses between tokens to register participants'responses. This type of paradigm, while allowing a large number of responses to be obtained in a short time, overlooks the potential effects that priming, selective adaptation and/or visual recalibration might have on the categorization decision process. That is, previous studies have shown that the perception of the McGurk effect might cause a recalibration of the auditory boundaries between phonemes (Bertelson *et al.*, 2003), affecting subsequent auditory perception (i.e., an auditory /ba/ following the McGurk illusion $A_{ba} + V_{ga}$, is more often misperceived as 'da'; Lüttke *et al.*, 2016).

The response alternatives can also be a factor having a direct impact on the magnitude of the McGurk illusion. Studies using the McGurk illusion as a paradigm use two kinds of response instructions: Closed-set response tasks,

in which subjects are asked to choose from specific response options, or open-set response tasks, in which subjects can respond with any syllable. Studies comparing the impact of these tasks on the incidence of the McGurk effect have revealed that closed-set response tasks elicit substantially more illusory responses than open response tasks (by as much as 18% more of fusion responses, Basu Mallick *et al.*, 2015; see also Colin *et al.*, 2005; Massaro, 1998). The increased magnitude of McGurk responses in forced choice closed-set tasks could be explained by an attempt by subjects to equalize the frequency of use of each response alternative throughout the experiment (e.g., Erlebacher and Sekuler, 1971), or due to the introduction of choices that participants deemed unlikely. For instance, Rosenblum *et al.* (2000) used a forced choice procedure because they observed that pilot participants were reluctant to identifying 'bg' as an initial consonant. When presented with 'bg' as a possible response, the number of these reports significantly increased.

However, the primary drawback of the forced choice approach is that, in limited closed-set tasks (e.g., three-choice tasks), the perceptual utterance experienced by the listener might not correspond to any of the alternatives provided by the experimenter. That is, as mentioned above, for a particular stimulus pairing (e.g., $A_{ba} + V_{ga}$) participants could potentially experience a range of percepts (e.g., 'da', 'tha', 'la', 'ra', etc.). If the perceived utterance does not match any of the given alternatives (e.g., /da/, /ba/, /ga/), participant will have to use some sort of strategy to solve the task. For instance, some participants might respond at chance or others might try to 'accommodate' their percept to the most similar phoneme in the response set (e.g., /da/). In order to avoid this source of undesired experimental noise, experimenters should use a large number of alternative responses and include a category 'others'.

To sum up, the structure of the response task is a source a variability in the incidence of the McGurk illusion in different studies and thus, results from studies using open-set tasks (e.g., Ma *et al.*, 2009; McGurk and MacDonald, 1976; Ross *et al.*, 2007) and studies using closed-set tasks (e.g., Sekiyama *et al.*, 2014; Sumby and Pollack, 1954) cannot be directly compared.

3.6. Specific Audiovisual Stimulus Pairing

McGurk illusions have been reported to emerge from a range of audiovisual pairings. However, the magnitude of the illusion has been shown to be constrained by some characteristics of the audiovisual pairings.

Most studies employ nonsense syllables (Consonant-Vowels, CV or Vowel-Consonant-Vowels, VCV), but the effect has also been reported in isolated vowels (Summerfield and McGrath, 1984; Valkenier *et al.*, 2012), and in words (see the discussion of Brancazio, 2004; Dekle *et al.*, 1992; Easton and Basala, 1982). The McGurk effect for vowels (e.g., $A_a + V_e$), however, is smaller than for consonants (Massaro and Cohen, 1993; Summerfield and McGrath, 1984).

Moreover, some studies have reported differences in the magnitude of the McGurk effect as a function of the vowel context (though see Jordan and Bevan, 1997). Green *et al.* (1988) used the classic pairing $A_{b+vowel} + V_{g+vowel}$ and found that the frequency of illusory 'd' percepts decreased as the following syllable shifted from /i/ to /a/ /u/. The reduction of the McGurk effect in the /u/ context can be explained by a decrease in the visibility of the consonant V_g, as the lips are rounded and protruded during its production due to coarticulation. When considering the vowel contexts (i.e., /a/ and /i/), Green *et al.* (1991) found that $A_{ibi} + V_{igi}$ stimuli produced more 'd' percepts, whereas the $A_{aba} + V_{aga}$ produced more 'th' percepts. It is important to note here, however, that only one talker was used in this study. Large inter-talker differences in producing the illusion have been anecdotally found in our laboratory and reported elsewhere (Basu Mallick *et al.*, 2015), a finding that we addressed above. The fact that the frequency to which the illusion is perceived can be explained by the idiosyncrasies of a specific talker tempers Green *et al.*'s conclusions.

3.7. Large Inter-Subject Variability

Since its discovery, a number of studies have shown substantial individual variability in the susceptibility to the McGurk effect (e.g., Benoit *et al.*, 2010; McGurk and MacDonald, 1976; Tremblay *et al.*, 2007), with some individuals consistently reporting illusory percepts and others not experiencing the illusion at all (Basu Mallick *et al.*, 2015). Individual susceptibility to the illusion, however, is stable across time, as shown by high test-retest reliability of the illusion (Basu Mallick *et al.*, 2015; Strand *et al.*, 2014).

Predisposition to experience the illusion has been explained by general factors such as age (Burnham and Dodd, 2004; Cienkowsky and Carney, 2002; McGurk and MacDonald, 1976; Rosenblum *et al.*, 1997; with adults being more susceptible to the illusion than babies and children), gender (with females being more susceptible to the illusion than males, e.g., Aloufy *et al.*, 1996; though see Irwin *et al.*, 2006) and linguistic and cultural background (e.g., Sekiyama and Tokhura, 1991; though see Magnotti *et al.*, 2016). Note, however, that the outcomes of the studies exploring similar research questions (e.g., gender, cultural background) have not always been consistent, showing, again, the high complexity of the phenomenon. A reduced susceptibility to the illusion has also been found in a number of clinical conditions, including autism (e.g., Bebko *et al.*, 2013; Taylor *et al.*, 2010, though see Keane *et al.*, 2010), schizophrenia (e.g., De Gelder *et al.*, 2003; Surguladze *et al.*, 2001), Williams syndrome (e.g., Böhning *et al.*, 2002), learning disabilities (e.g., Boliek *et al.*, 2010; Norrix *et al.*, 2006), specific language impairement (e.g., Leybaert *et al.*, 2014), dyslexia (e.g., Bastien-Toniazzo *et al.*, 2009),

aphasia (e.g., Youse *et al.*, 2004) and Alzheimer's disease (e.g., Delbeuck *et al.*, 2007).

The case of individuals who are not sensitive to the McGurk effect, and thus consistently report the auditory cue, is puzzling. Already in the original study, McGurk and MacDonald (1976) reported that one participant (out of 54) consistently reported auditory responses. These participants have been generally overlooked by researchers and have been excluded from the experimental samples, either on the basis of study screening pre-tests (e.g., Bernstein *et al.*, 2008; Brancazio *et al.*, 2003; Hirvenkari *et al.*, 2010; Malfait *et al.*, 2014) or post-test results (e.g., Kislyuk *et al.*, 2008; Munhall *et al.*, 1996; Tiippana *et al.*, 2004; Van Wassenhove *et al.*, 2007). To estimate the proportion of healthy adults who do not experience the McGurk effect, we surveyed published studies that explicitly reported the number of participants who consistently report auditory responses (or reported marginal 'McGurk' responses). Our survey consists of 27 studies that included a total of 1044 participants (see Table 1). Out of this sample, 153 participants were categorized as non-McGurk perceivers, i.e., 14.6%. Across studies, that proportion varies from 1.8% (McGurk and MacDonald, 1976) to 67.7% (Gentilluci and Cattaneo, 2005), with a mean ($\pm$ SD) of 15 $\pm$ 13. Exploring why these participants process the audiovisual information differently than McGurk perceivers could enormously advance our understanding of the mechanisms at play in audiovisual speech integration.

Three possibilities should be considered when examining the points at which individual differences might disrupt the illusion: (1) superior sensitivity to detecting (the lack of) audiovisual correspondences; (2) a lower weighting of the visual cues (or underspecified extraction of sensory information) or higher weighting of the auditory cues; or (3) an inefficient combination of the two cues.

A pre-condition for the McGurk illusion is that the perceptual system has to (erroneously) process the auditory and visual sensory signals as belonging to the same external event. That is, in order for the audiovisual event to be merged the perceptual system must track a range of commonalities across heard and seen speech. These commonalities are reflected in the dynamic structure of both the auditory and the visual channels. It is still not known, however, which cues across heard and seen speech are tracked by the perceptual system to determine the audiovisual matching. The intelligibility of patients with facial paralysis — who have minimal facial muscle mobility and difficulty in closing the mouth and yet can develop understandable speech (i.e., Moebius syndrome) — decreases when the observers see the patient's face (in relation to the auditory only condition). The fact that a type of McGurk illusion occurs in talkers whose facial muscles are mostly paralyzed suggests that minimal cues are required for the system to detect crossmodal correspondences (Nelson

Table 1.
Reported number and proportion of healthy adults who do not experience the McGurk effect across 27 studies

Study authors	Publication year	Sample size	Non-perceivers	Proportion
McGurk and MacDonald	1976	54	1	0.018
Rosenblum and Saldaña	1992	51	2	0.039
Munhall *et al.*	1996	71	7	0.099
Sams *et al.*	1998	65	5	0.077
Fingelkurts *et al.*	2003	10	2	0.200
Paré *et al.*	2003	49	4	0.082
Desjardins and Werker	2004	8	1	0.125
Tiippana *et al.*	2004	17	3	0.176
Gentilluci and Cattaneo	2005	31	21	0.677
Saint Amour *et al.*	2007	12	1	0.083
Fingelkurts *et al.*	2007	9	2	0.222
Traunmüller *et al.*	2007	21	5	0.238
Van Wassenhove *et al.*[4]	2007	43	4	0.093
Kislyuk *et al.*[5]	2008	11	2	0.182
Andersen *et al.*[1]	2009	14	2	0.143
Beauchamp *et al.*	2010	16	4	0.250
Benoit *et al.*	2010	14	1	0.071
Wiersinga-Post *et al.*	2010	20	4	0.200
Bishop and Miller	2011	11	1	0.091
Tiippana *et al.*	2011	32	1	0.031
Nath and Beauchamp	2012	14	3	0.214
Stevenson *et al.*	2012	31	2	0.064
Basu Mallick *et al.*[1]	2015	275	60	0.218
Gurler *et al.*[1]	2015	40	6	0.150
Roa Romero *et al.*[3]	2015	25	6	0.240
Venezia *et al.*[2]	2016	34	1	0.029
Wilson *et al.*	2016	66	2	0.030
Total		1044	153	

Non-perceivers were identified as those reporting $<5\%$[1], $<15\%$[2], $<25\%$[3], $<40\%$[4], and $<50\%$[5] 'McGurk' responses.

and Hodge, 2000; Von Berg *et al.*, 2007). In this line, it has also been shown that exact temporal correspondence between the visual and auditory channel is not necessary for the illusion to occur (Jones and Jarick, 2006; Massaro and Cohen, 1993, 1996; Miller and D'Esposito, 2005; Munhall *et al.*, 1996; Soto-Faraco and Alsius, 2009; Van Wassenhove *et al.*, 2007). However, when the cross-modal incongruency is too large (as happens in dubbed movies) the perceptual system decreases the role of the visual input to block the integration mechanisms (Nahorna *et al.*, 2012).

Whereas it is often claimed that perceivers of McGurk stimuli are unaware of the discrepancy between the visual and the auditory channels (see Rosenblum and Saldaña, 1992), it is still possible that some are more sensitive to the cues indicating audiovisual correspondences and — correctly — detect that there is an audiovisual discrepancy when presented with such artificial stimuli. Supporting this hypothesis, Strand *et al.* (2014; see also Sakamoto *et al.*, 2012), presented participants with McGurk stimuli and asked them to both identify the syllables and to respond whether the auditory and visual signals were the same phoneme (congruent) or are different phonemes (incongruent). They found that susceptibility measures derived from identification tasks were (moderately) related to the ability to detect instances of audiovisual incongruity. Note, however, that the detection of audiovisual incongruity only explained a small part of the variance in (not) perceiving the McGurk fusions in their study.

Another possibility is that individuals who do not experience the illusion do process the McGurk stimuli as a unified event, but weigh the visual input less than McGurk perceivers. The idea that some participants rely more on the visual information and others on the auditory was already stated in 1985 by Seewald *et al.*, who claimed that there is an individual 'primary modality for speech perception'. Some contemporary models of multisensory integration suggest that the ultimate multisensory percept is a weighted average of sensory estimates (Schwartz, 2010). If the sensory estimate for the visual modality is low (due to individual characteristics, such as a reduced exposure to visual facial information across a lifetime for cultural reasons, see Sekiyama and Tokhura, 1991) then the auditory modality will contribute more to the final percept. Along these lines, the recent finding that highly skilled musicians, who possess superior auditory abilities, have significantly reduced sensitivity to the McGurk illusion suggest that some individuals can more strongly weight auditory cues over visual cues (Proverbio *et al.*, 2016).

A third possibility is that individuals who do not experience the McGurk illusion simply have poorer integration skills. In the SPIN literature, some researchers have hypothesized the existence of a specialized mechanism that would integrate auditory and visual speech information and would have its own source of variance. Support for differing efficiency in the operation of an integration mechanism was initially stated by Grant and Seitz (1998), who claimed that the large differences in individuals' AV recognition in SPIN can not solely be explained by differences in unimodal intelligibility levels. If the same 'integration' mechanism was at play in the processing of the McGurk illusion, it could potentially account for part of the variance observed in the illusory reports. Note, however, that if the low sensitivity to the illusion was explained by general poor integration mechanisms, one should observe a correlation between the McGurk illusion and SPIN performance, a relationship

that, as we describe below, remains to be thoroughly investigated. Lastly, it is also possible that it is not the integration mechanisms 'per se' what determines susceptibility to the illusion. Current evidence suggests that the McGurk effect is not only driven by perceptual process, but can also be modulated by cognitive processes, such as attention (Navarra *et al.*, 2010), expectation (Tuomainen *et al.*, 2005), awareness (Munhall *et al.*, 2009; Palmer and Ramsey, 2012; though see Soroker *et al.*, 1995) mental imagery (Berger and Ehrsson, 2013) or suggestion (Déry *et al.*, 2014; Lifshitz *et al.*, 2013). It is thus possible that susceptibility to the McGurk illusion is dependent on interindividual differences on the cognitive processes intervening at the binding stage.

Finally, it should be noted that rather than categorical, the effect could be a matter of degree. That is, it is possible that participants who reported auditory responses were indeed processing the visual speech information 'without being affected at a conscious verbalisable categorical level of processing' (MacDonald *et al.*, 2000). According to MacDonald *et al.*, 2000, if the illusion is indeed a matter of degree, there must be some sort of threshold that would make the illusion become explicit and cognitively transparent. Supporting this view is the evidence showing that an absence of reported illusory percepts does not necessarily mean that the auditory and visual information have not interacted at the implicit level (Brancazio, 2004; Brancazio and Miller, 2005; Gentilucci and Cattaneo, 2005; MacDonald *et al.*, 2000). MacDonald *et al.* (2000) applied various levels of spatial degradation filters (a mosaic transform) to McGurk videos. In addition to the judgments of the speech syllable perceived, participants were also asked to report the rate of auditory clarity for each stimulus. After the experiment, they divided participants into two groups: those who were highly sensitive to the illusion, and those who were weak McGurk perceivers. They found that those stimulus pairs that were more subject to illusory responses by the strong McGurk perceivers, corresponded to the tokens that the weak McGurk perceivers rated as less clear auditorily, despite the fact that they ultimately reported the auditory component. This suggests that some form of audiovisual interaction was taking place, even though the interaction did not reach the threshold for categorization.

Brancazio and Miller (2005) found that visual speaking rate influences phonetic judgments even when the McGurk effect does not occur. The authors presented an auditory continuum of voiced to voiceless consonants /bi-pi/ together with fast and slow V_{ti}. In this paradigm, the /b/-/p/ voicing boundary usually occurs at different points of the continuum depending on the rate of the visual stimuli (i.e., faster visual stimuli leading to the boundary at much shorter voice onset time). Critically, the authors compared conditions in which the McGurk effect continuum occurred (leading to a perceived '/di - ti/' continuum) to conditions in which the McGurk effect did not occur (participants heard it as '/bi - pi/'). They found that the rate of presentation of /t/ affected

the voicing boundary even when the McGurk illusion failed to occur. This suggests that phonetically relevant information is extracted from the visual signal even when the illusion does not occur.

Gentilucci and Cattaneo (2005) presented participants with McGurk stimuli and performed an acoustical analysis of the participants' spoken responses, for trials in which the illusion was perceived and trials in which participants responded the auditory component. The analyses showed that, even when the McGurk did not occur, participants' utterances were always influenced by the articulatory gestures of the speaker. This suggests that some phonetic features present in the visual signal were being processed.

Thus, the incidence of the McGurk effect may be underestimating the actual extent of interaction amongst the two modalities (see Brancazio and Miller, 2005), a finding that, by itself, questions the suitability of the illusion as a proxy measure for integration. In the next section we review some further issues related to the extended practice of generalizing McGurk data to audiovisual speech information.

4. Suitability of the Illusion as a Proxy Measure for Integration

Studies using the McGurk as an experimental tool to investigate the necessary and sufficient conditions under which audiovisual speech integration occurs, rely on the fundamental assumption that the processing of incongruent illusory stimuli is supported by the same integration mechanisms than those involved in the processing of naturally occurring (i.e., audiovisually matching) speech events.

In this section we want to consider evidence suggesting that the mechanisms involved in the processing of illusory audiovisual pairings differs from the mechanisms involved in the processing of congruent audiovisual speech events. In particular, the processing of McGurk stimuli — where the correspondence between the bottom-up sensory signals from the two modalities is imperfect-, could require the involvement of some additional mechanism to overcome the informational modality mismatch during integration (Green and Kuhl, 1991; Massaro and Cohen, 1983; Romero *et al.*, 2015).

There is now accumulating evidence that the integration of multisensory information does not entail the activation of a process that encompasses all the sensory attributes at once, but rather is more multifaceted than previously suggested (Eskelund *et al.*, 2011; Nahorna *et al.*, 2012; Soto-Faraco and Alsius, 2009). In this line, Soto-Faraco and Alsius (2009) showed that perceiver can gain access to individual sensory attributes of an illusory McGurk percept (i.e., integrated phonological percept and temporal relation between the two unimodal events). Furthermore, it has been found that in selective adaptation studies -where the repeated presentation of a sound that reduces reports of

sounds similar to the repeating one-, illusory percepts do not act as adaptors, but rather, perceivers adapt to the auditory component of the illusory pairing (Roberts and Summerfield, 1981; Saldaña and Rosenblum, 1994). The finding that selective speech adaptation mainly depends on the acoustic quality of the stimulus, and not the perceived stimulus, suggest that the observer has some access — even if unconsciously — to the features of the acoustic component of the illusion. Overall, thus, these studies suggest that the processing of McGurk stimuli is fragmentary and multistaged. The claim that the processing of McGurk stimuli might require additional processing mechanisms than perfectly matching audiovisual speech events is not trivial and should be considered cautiously by researchers in the field.

Even if representing an artificial situation, the McGurk illusion is currently commonly used in the field as a tool to explore how, where and when the perceptual system integrates visual and auditory signals. Yet the findings showing that different mechanisms might be at play in the processing of such unnatural stimuli, largely tempers some of the conclusions derived from these studies. For instance, in the past years it has been claimed that the integration of audiovisual speech can be modulated by top-down mechanisms, such as expectation (Tuomainen *et al.*, 2005), attention (Alsius *et al.*, 2009), suggestion (Déry *et al.*, 2014) or mental imagery (Berger and Ehrsson, 2013). However, it is possible that these top-down mechanisms could be only recruited to bind crossmodal sensory information in situations of perceptual conflict. Similarly, many clinical studies have reported atypical patterns of multisensory integration in a variety of clinical populations using McGurk stimuli. Yet, it remains unclear whether the integration deficits reported for these populations are limited to conditions in which the perceptual mechanism has to deal with intermodal discrepancies.

In this section we will review evidence that the McGurk illusion differs from audiovisual matching stimuli at a perceptual and neural level. Furthermore, we report evidence showing that the relationship between illusory percepts and the other classic proxy of audiovisual integration (namely, SPIN) is, at most, weak.

4.1. The Phenomenological Experience Arising From McGurk Stimuli Is not Equivalent to Its Audiovisual Congruent Counterpart

Participants often describe the phenomenological experience arising from a McGurk-type stimulus as being different from the experience derived from congruent audiovisual stimulus (Rosenblum and Saldaña, 1992; Soto-Faraco and Alsius, 2009). This subjective feeling might in fact be well-grounded as accumulating evidence shows that McGurk percepts are not as phonetically compelling as audiovisual congruent audiovisual syllable. Rosenblum and Saldaña (1992) asked participants to match the auditory syllable ‘va’ to either

an audiovisual consistent pairing (i.e., $A_{va} + V_{fa}$; same place of articulation) or an audiovisual discrepant syllable (i.e., $A_{ba} + V_{va}$; different place of articulation). They found that, even when the stimulus were all identified as 'va', participants were more likely to match the auditory 'va' to the audiovisual consistent pairing (see also Massaro and Ferguson, 1993). This suggests that, even when participants were not aware of the audiovisual discrepancy, they judged the discrepant audiovisual syllables as less compelling than the percepts from audiovisually consistent pairings. Similarly, Brancazio (2004) asked participants to identify the initial consonant of audiovisually congruent and incongruent stimuli and to report category goodness ratings of the percept. He found that McGurk percepts were rated as poor category exemplars as compared to percepts of audiovisually congruent tokens, and that ratings were even lower for stimuli in the incongruent condition when the McGurk effect did not occur (e.g., $A_{ba} + V_{da}$ perceived as /b/).

Furthermore, other studies have shown that participants require significantly more time to identify McGurk fusion responses than congruent audiovisual syllables (Beauchamp *et al.*, 2010; Brancazio, 2004; Green and Kuhl, 1991; Hessler *et al.*, 2013; Keane *et al.*, 2010; Massaro and Cohen, 1983; Nahorna *et al.*, 2012; Norrix *et al.*, 2006; Tiippana *et al.*, 2011). For instance Green and Kuhl (1991) reported longer reaction times for McGurk stimuli (e.g., $A_{ibi} + V_{igi}$, leading to 'idi') than for congruent audiovisual stimuli ($A_{ibi} + V_{ibi}$ syllable). The longer reaction times for mismatching stimuli generally suggests that perceptual decisions about incongruent stimuli are more difficult, an explanation that is consistent with the requirement of supplementary mechanisms to merge the discrepant information (Lüttke *et al.*, 2016). Note that, if that was the case, one should observe no differences in reaction time between McGurk and auditory trials in those instances in which the illusion fails to occur. It is also possible, however, that the McGurk percepts are simply poorer exemplars for that phoneme category, thus leading to an increased uncertainty about the response.

It has also been noted that McGurk pairings are more susceptible to manipulations in the spatial and temporal domain than their congruent counterparts. In the spatial domain, the audiovisual stimuli have been manipulated by placing the sources at different locations to create the ventriloquist illusion, in which the speaker's voice is visually captured at the location of the moving mouth. These studies have shown that the visual spatial capture is attenuated in the McGurk audiovisual pairs (i.e., less ventriloquist effect) relative to congruent audiovisual stimuli (Bishop and Miller, 2011; Jones and Munhall, 1997; Kanaya and Yokosawa, 2011). In the temporal domain, the visual and auditory pairings have been manipulated by introducing delays between the two sources. These studies have shown that the perceptual system can handle a relatively large temporal offset between auditory and visual speech signals

(with delays that vary from 40 to 100 ms with audio leading, up to 480 for video leading, e.g., Soto-Faraco and Alsius, 2009). It has also been found that McGurk pairs are more readily judged as asynchronous than congruent pairs (Van Wassenhove *et al.*, 2007). The different temporal profile for McGurk *vs* congruent tokens suggests that the perceptual system detects a decrease in audiovisual coherence in these types of artificial pairings.

Finally, it is also interesting to note that the McGurk illusion is also more susceptible to image degradation than congruent audiovisual speech stimuli. Jordan and Bevan (1997) found that whereas shifts in facial orientation did not substantially affect the beneficial effect of congruent visual speech stimuli on performance (relative to the unimodal auditory baseline), they do interfere with the McGurk illusion (more auditory responses when the face is presented away from the vertical). In other studies, Jordan and Sergeant found a similar pattern with reductions in the size of facial images (1998), and when manipulating the physical distance between the perceiver and the talker (2000). According to the authors, the difference between congruent and incongruent audiovisual stimuli may reflect the fact that little visual speech signal is sufficient to boost the auditory signal in congruent conditions, whereas the visual speech signal needs to be clear to influence the identification of incongruent auditory speech in McGurk stimuli (see also Rosenblum and Saldaña, 1996). These results indirectly support the existence of a preliminary binding stage (see Nahorna *et al.*, 2012), where the perceptual system quickly determines the coherence in the dynamics of the auditory and visual speech sources before categorization takes place.

Overall, these differences between congruent and incongruent McGurk percepts suggest that the processes involved in integrating incongruent stimuli are both in space and time different than the processes involved in integrating congruent stimuli. Indeed, fMRI and physiological (EEG and MEG) studies have shown that the McGurk illusion has a different neural signature than congruent audiovisual stimuli, with illusory stimuli possibly requiring additional neural processing.

4.2. The McGurk Illusion Has a Different Neural Signature Than Congruent Audiovisual Stimuli

A number of studies have contrasted incongruent McGurk speech *versus* congruent AV speech (e.g., Benoit *et al.*, 2010; Bernstein *et al.*, 2008; Irwin *et al.*, 2011; Jones and Callan, 2003; Olson *et al.*, 2002; Szycik *et al.*, 2012; Wiersinga-Post *et al.*, 2010). These studies have identified a network of cortical regions involved in the processing of McGurk stimuli, including temporal, frontal, insular and parietal areas.

Among these regions, one area has been identified by different studies as a key region involved in the processing of McGurk stimuli: the superior temporal sulcus (STS; Beauchamp *et al.*, 2010; Benoit *et al.*, 2010; Irwin *et al.*, 2011; Nath and Beauchamp, 2011, 2012; Szycik *et al.*, 2012). Some studies have contrasted BOLD activity for fused and non-fused responses to physically identical McGurk stimuli and have reported stronger activation for perceptually fused stimuli in the posterior part of the STS (pSTS; Benoit *et al.*, 2010; Miller and D'Esposito, 2005; Szycik *et al.*, 2012). Beauchamp *et al.* (2010) used transcranial magnetic stimulation (TMS) to temporally disrupt the left pSTS and found a significant decrease in the magnitude of reported McGurk percepts, but no interference with non-McGurk stimuli. Furthermore, Nath and Beauchamp (2012) exploited individual differences in susceptibility to McGurk fusion and found that BOLD response in pSTS positively correlated with the likelihood of perceiving the McGurk effect (see also Nath *et al.*, 2011). The exact role of the pSTS and other brain areas in merging incongruent AV sensory inputs, however, remains unclear.

Other studies, however, have not found pSTS activation selective for the processing of McGurk stimuli (Baum *et al.*, 2012; Bernstein *et al.*, 2008; Erickson *et al.*, 2014; Jones and Callan, 2003; Wiersinga-Post *et al.*, 2010). Baum *et al.* (2012) reported the case of a patient that showed robust illusory percepts despite having the left middle and posterior STS, as well as adjacent areas in the temporal and parietal lobes ablated (note, however, that the response amplitude to McGurk stimuli in the right STS in this patient was significantly greater than in healthy age-matched controls, possibly allowing compensation for the damaged left pSTS). In a recent study, Erickson *et al.* (2014), has also questioned an exclusive role of the pSTS in the perceptual shift that occurs in McGurk-like stimuli. Using the max criterion approach, which identifies the processing regions that respond more strongly to AV stimuli relative to both unimodal auditory and visual stimulation alone, they found that the left pSTS was recruited for congruent audiovisual speech, whereas the McGurk illusion activated the left pSTG. They suggest two mechanisms taking place at the left ST regions: Initially, the sensory cues are compared and integrated in the left STS. After that, the activation of pSTG reflects the creation of a merged percept arising from the conflicting sensory cues.

The multistage process of the McGurk illusion has been supported by recent studies examining the neural temporal signature of the phenomenon. Studies exploring the temporal signature of the McGurk stimuli processing have also found significant differences with the processing of congruent speech. Using electroenchephalograpy (EEG), Hessler *et al.* (2013) compared fused percepts (responses that deviated from the auditory component) to congruent audiovisual stimuli in an active oddball paradigm. They found that the McGurk stimuli

elicited a more negative waveform between 360 and 400 ms than congruent material. Their analyses subtracted the visual information from the signal and thus these differences cannot be explained by physical differences in the stimuli (i.e., the auditory component of the McGurk stimuli was the same as the auditory alone stimuli). The authors claim that a likely explanation for these different patterns of activity is found in the more difficult integration of McGurk-type stimuli, which requires the merging of incongruent information.

Neural synchrony has also been proposed as a mechanism involved in the McGurk illusion. In a magnetoencephalogram (MEG) study, Fingelkurts *et al.* (2003), compared brain oscillations in response to McGurk and congruent stimuli at two frequency bands (alpha (7–13 Hz) and beta (15–21 Hz)). They found that brain operation (Note 3) tended to be of a longer duration in response to the presentation of McGurk stimuli than to the presentation of congruent audiovisual stimuli. Furthermore, they compared the brain oscillations of participants who experienced the McGurk illusion ($n = 7$) with participants who did not ($n = 2$). They found that individuals who did not experience the McGurk effect had different integrative cortical networks of functional interactions compared to the subjects who perceived the McGurk effect. In particular, they observed negative values of operation synchrony in the non-McGurk individuals, meaning that the network of cortical sites responsible for integration (long connections between anterior and posterior brain areas and also between left and right hemisphere temporal sides) was actively unsynchronized. The authors suggest that subjects without the McGurk effect possibly process information from both modalities independently due to active process of unsynchronized brain operations.

In another MEG study, Keil *et al.* (2012) identified cortical responses underlying different perceptions of identical McGurk stimuli. They found that the perception of the McGurk illusion (defined as a response that matched neither the auditory nor the visual information) was preceded by increased prestimulus beta-band activity in parietal, frontal, and temporal areas. Furthermore, the disposition to fuse audiovisual speech information was enhanced as the left superior temporal gyrus (lSTG), which is considered as a key site of multimodal integration, was more strongly coupled to frontoparietal regions.

In a recent high-density electroencephalography (EEG) study, Romero *et al.* (2015) found that early, event-related responses (N1) to auditory onset were reduced during the perception of the McGurk illusion compared with congruent stimuli. Furthermore, they found a stronger post-stimulus suppression of beta band power (13–30 Hz) at short (0–500 ms) and long (500–800 ms) latencies during the perception of the McGurk illusion. Based on these results, the authors propose a three stage process of McGurk-like stimuli. First, the reduction of the N1 could be reflecting the impact of visual context on audiovisual speech integration, with stronger integration effects for incongruent McGurk

stimuli. Second, the early beta-band effects could reflect the detection of an audiovisual incongruency and the 'allocation of upcoming processing demands following the violation of the prediction based on the visual context'. Finally, the long-latency effect on beta-band power could be reflecting the formation of a subjectively congruent illusory percept that follows the detection of incongruence at earlier processing stages.

Overall, these studies suggest that different neural structures and mechanisms underlie the potentially multistage processing of the McGurk illusion.

4.3. There Is Mixed Evidence for the Correlation of the McGurk Effect With Speechreading (Visual-Only) Performance, and With AV Enhancement

The sources of variance for the individual differences in susceptibility to the McGurk effect have not been determined. Susceptibility to the illusion could be explained by individual differences in unimodal processing ability (e.g., speechreading ability), in the perceptual mechanism responsible for combining audiovisual speech information, or both (Grant and Seitz, 1998). Yet, the explicit link between McGurk susceptibility, speechreading abilities and other measures of audiovisual integration (e.g., performance in AV speech in noise tasks) has not been thoroughly explored, and the few studies investigating these critical relationships have yielded mixed results.

In an attempt to investigate the relationship between the McGurk illusion and the ability to extract visual speech information from the talker's face, Cienkowski and Carney (2002) ran correlation analyses between a sentence speechreading performance and optimally fused McGurk responses (e.g., 'da' resulting from $A_{ba} + V_{ga}$). According to the authors, finding that good speechreaders were also more susceptible to the McGurk effect would suggest that the successful integration of audiovisual speech information was associated with the successful processing of the visual input. The authors found that speechreading did not correlate with the amount of fused responses, suggesting that there is a fundamental difference between visual and audiovisual speech processing. It is important to note, however, that Cienkowski and Carney (2002) used speech stimuli of different complexity for the two audiovisual measures, and thus the lack of correlation could be interpreted as reflecting a distinction between phoneme perception and sentence perception, rather than reflecting a distinction between McGurk perception and speechreading ability (Grant and Seitz, 1998; Strand *et al.*, 2014).

Other studies have correlated susceptibility to the McGurk illusion to syllable speechreading ability. The results, however, are inconsistent across studies. Massaro *et al.* (1986) found a significant positive correlation ($r = 0.75$) between speechreading skill and the extent of the visual influence during audiovisual speech (i.e., the McGurk effect). Two recent studies found no correlation between speechreading of VCV syllables and the McGurk illusion,

when the effect was either defined as non-auditory response for incongruent audiovisual pairings (Wilson *et al.*, 2016, $r = 0.01$; see also Tremblay *et al.*, 2007, $r = -0.02$) or when it was defined as a percept different from the auditory or visual inputs (Strand *et al.*, 2014; $r = 0.14$). It is important to note, however, that Strand *et al.* did find moderate positive correlations when quantifying speechreading ability using more fine grained analysis of the information the participant was able to extract from the visual signal (i.e., the ability to identify the place of articulation of the utterance; $r = 0.32$).

Other studies point to a distinct use of visual information for speechreading and McGurk tasks. First, speechreading requires greater visual resolution than audiovisual speech (Lansing and McConkie, 2003; Wilson *et al.*, 2016). Second, speechreading tasks benefit from face familiarity with the talker (Yakel *et al.*, 2000), whereas the McGurk effect actually decreases when the talker is someone known by the observer (Walker *et al.*, 1995, but see Rosenblum *et al.*, 2000 on the effect of single *vs* multiple talkers). Further research should investigate why facial familiarity enhances visual speech salience in speechreading but not in McGurk stimuli.

Overall, these results suggest that, if present, the relationship between speechreading and the McGurk illusion is rather weak, and thus that the McGurk illusion depends also on other individual factors, which may include the ability to integrate cross-modal information. Interestingly, however, the relationship between the McGurk illusion and other measures of audiovisual integration, such as the audiovisual gain for speech perceived in noise has barely been studied.

Previous studies have shown that when the intelligibility of acoustic speech is impoverished by adding noise (SPIN), the concurrent presentation of corresponding visual speech cues improves comprehension dramatically (Cotton, 1935; Rosenblum *et al.*, 1996; Sumby and Pollack, 1954). Perception of audiovisual speech in noise has been shown to be super-additive (Calvert *et al.*, 2000; Ma *et al.*, 2009; Ross *et al.*, 2007; Sommers *et al.*, 2005; Wright *et al.*, 2003). That is, when the auditory signal is degraded with noise, performance in the audiovisual condition is greater than the linear sum of the unimodal (auditory only and visual only) performances. A recurrent finding in audiovisual SPIN studies is that the enhancement provided by the addition of the visual speech input (i.e., the degree of superadditivity) varies substantially between participants (MacLeod and Summerfield, 1990). Super-additive performance on audiovisual SPIN tasks results from the integration of congruent unimodal inputs — there is no intermodal conflict, and this can be considered a much more ecological way to measure audiovisual speech integration. One might assume that individuals that are more efficient in combining the auditory and visual speech information in SPIN will also be more susceptible to the McGurk illusion, that it is considered a measure of the strength

of AV integration. Yet, to our knowledge, only two studies have investigated this issue, and they present contradictory outcomes. Grant and Seitz (1998) tested hearing-impaired participants with nonsense syllable in noise, sentences in noise and McGurk task to obtain estimates of integration efficiency. They found that McGurk susceptibility positively correlated with AV benefit scores (AV-A/1-A) for consonant in noise ($r = 0.43$) and AV sentence in noise ($r = 0.46$) performance. Note, however, that, because the study tested hearing-impaired individuals as participants, the results might not generalize to hearing population. In fact, in a more recent study, Van Engen *et al.* (2016) measured normal-hearing participants' susceptibility to the McGurk illusion as well as their ability to identify sentences in noise across a range of signal-to-noise ratios and found no relationship between the two measures. This result critically questions the validity of the McGurk illusion as an index of natural audiovisual integration.

Finally, it is interesting to note that the McGurk effect has been shown to correlate with other audiovisual illusions (i.e., illusory flash effect) where a single visual flash is perceived as two flashes if it is accompanied by two closely successive sounds (Stevenson *et al.*, 2012; Tremblay *et al.*, 2007). The fact that the strength of two audio-visual illusions with distinct properties (speech *vs* nonspeech, visual dominance *vs* auditory dominance) is correlated at the individual level, suggests common individual characteristics contributing to the integration of multisensory material. Whether these characteristics are perceptual or cognitive (e.g., motivation, attention, etc.) remains to be elucidated.

5. Conclusion and Future Directions

We began this review questioning the use of the McGurk illusion as an ideal measure for audiovisual integration. We demonstrated that even basic features of the illusion such as the magnitude of the effect and its variability are almost impossible to know. Our review also highlighted some significant weaknesses in the stimuli and methods for evaluating the McGurk effect. Considering these weaknesses, we highly advise researchers to use SPIN tasks when possible in order to guarantee that the obtained results can be extrapolated to everyday speech processing. If using SPIN tasks is not possible, or in those studies exploring the illusion, we suggest that a few methodological requirements should become standard in all paradigms using the McGurk effect: (1) the illusion should generally be described as the ability of a visual stimulus to alter the perception of an auditory stimulus that is perfectly audible on its own (instead of the emergence of a new percept). The idea that only a classic fusion response (perception of 'da' for $A_{ba} + V_{ga}$) is a McGurk effect, ignores the

variability that exists in the perceptual experience of such stimuli. This definition necessarily implies that (2) the results are best reported as the percentage of responses that do not correspond to the auditory stimulus. Furthermore, if we want to establish how strong is the illusion (i.e., calculate the effect size) (3) researchers should systematically report means and variance estimates for the responses. Researchers should also make an effort to (4) have more stimulus control. Without some attempt to understand how much of a particular effect is limited to a specific set of recordings used in an experiment, we cannot separate the idiosyncrasies of a talker from the factors being manipulated in a study. This problem requires the development of methods to calibrate the stimuli being tested. Until that occurs, we should be using multiple talkers in all studies and journals should be archiving the stimuli. Finally, in terms of study design, we believe that (5) the response alternatives should have at least one openset choice such as 'other' or the responses should be completely openset. The restriction of responses to the three voiced stops in English will force subjects to provide a more homogeneous distribution of percepts than may be true. (6) Comparison stimuli or standards (such as testing auditory-only version of the dubbed stimulus or testing the congruent audiovisual stimuli) must be included in the stimulus set. The illusion, after all, is a measure of the degree to which the incongruent visual stimulus can alter perception. This alteration has to be assessed in reference to perception of the auditory stimuli without incongruent visual information.

Overall, the findings reported here challenge the well-established practice of using the McGurk illusion as a proxy measure for integration. Before extrapolating results to naturally occurring audiovisual speech events, experimenters have to ensure that their results are not restricted to this situation in which the perceptual mechanism is faced with incongruent crossmodal information.

Acknowledgements

The present work was made possible by a research grants from NSERC and CIHR.

Notes

1. Note here that we included studies that did not explicitly report the language spoken by participants/talkers but that were carried out in English-speaking countries (based on the Author's affiliation information).

2. The decision to include only studies reporting optimally fused responses was based on the still common practice in psychology — as well as in other areas citing the effect — of describing the illusion as the emergence

of a new percept (e.g., Peynircioğlu *et al.*, 2017). We believe that, in order for researchers to keep using this definition, it needs to be shown that the effect is robust even under this distinctive characterization.

3. Note that *brain operation* refers to the functional coupling (in terms of synchrony) of different brain areas. Operational synchrony (OS) was calculated by establishing the synchronization of rapid transition processes (RTPs, which are the markers of boundaries between quasistationary segments) in each MEG location.

References

Alm, M., Behne, D. M., Wang, Y. and Eg, R. (2009). Audio-visual identification of place of articulation and voicing in white and babble noise, *J. Acoust. Soc. Am.* **126**, 377–387.

Aloufy, S., Lapidot, M. and Myslobodskym, M. (1996). Differences in susceptibility to the "blending illusion" among native Hebrew and English speakers, *Brain Lang.* **53**, 51–57.

Alsius, A., Wayne, R., Paré, M. and Munhall, K. G. (2016). High visual resolution matters in audiovisual speech perception, but only for some, *Attent. Percept. Psychophys.* **78**, 1472–1487.

Andersen, T. S., Tiippana, K., Laarni, J., Kojo, I. and Sams, M. (2009). The role of visual spatial attention in audiovisual speech perception, *Speech Commun.* **51**, 184–193.

Bastien-Toniazzo, M., Stroumza, A. and Cavé, C. (2009). Audio-visual perception and integration in developmental dyslexia: an exploratory study using the McGurk effect, *Curr. Psychol. Lett.* **25**, 2–14.

Basu Mallick, D., Magnotti, J. F. and Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type, *Psychonom. Bull. Rev.* **22**, 1299–1307.

Baum, S. H., Martin, R. C., Hamilton, A. C. and Beauchamp, M. S. (2012). Multisensory speech perception without the left superior temporal sulcus, *NeuroImage* **62**, 1825–1832.

Beauchamp, M. S., Nath, A. R. and Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect, *J. Neurosci.* **30**, 2414–2417.

Bebko, J. M., Schroeder, J. H. and Weiss, J. A. (2013). The McGurk effect in children with autism and Asperger syndrome, *Autism Res.* **7**, 50–59.

Benoit, M. M., Raij, T., Lin, F. H., Jaaskelainen, I. P. and Stufflebeam, S. (2010). Primary and multisensory cortical activity is correlated with audiovisual percepts, *Hum. Brain Mapp.* **31**, 526–538.

Berger, C. C. and Ehrsson, H. H. (2013). Mental imagery changes multisensory perception, *Curr. Biol.* **23**, 1367–1372.

Bernstein, L. E., Auer, E. T., Jr, Wagner, M. and Ponton, C. W. (2008). Spatiotemporal dynamics of audiovisual speech processing, *NeuroImage* **39**, 423–435.

Bertelson, P., Vroomen, J. and De Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect, *Psychol. Sci.* **14**, 592–597.

Bishop, C. W. and Miller, L. M. (2011). Speech cues contribute to audiovisual spatial integration, *PLoS One* **6**, e24016. DOI:10.1371/journal.pone.0024016.

Böhning, M., Campbell, R. and Karmiloff-Smith, A. (2002). Audiovisual speech perception in Williams syndrome, *Neuropsychologia* **40**, 1396–1406.

Boliek, C., Keintz, C., Norrix, L. and Obrzut, J. (2010). Auditory-visual perception of speech in children with learning disabilities: the McGurk effect, *Can. J. Speech-Lang. Pathol. Audiol.* **34**, 124–131.

Brancazio, L. (2004). Lexical influences in audiovisual speech perception, *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 445–463.

Brancazio, L. and Miller, J. L. (2005). Use of visual information in speech perception: evidence for a visual rate effect both with and without a McGurk effect, *Percept. Psychophys.* **67**, 759–769.

Brancazio, L., Miller, J. L. and Paré, M. A. (2013). Visual influences on the internal structure of phonetic categories, *Percept. Psychophys.* **65**, 591–601.

Buchan, J. N. and Munhall, K. G. (2011). The influence of selective attention to auditory and visual speech on the integration of audiovisual speech information, *Perception* **40**, 1164–1182.

Burnham, D. and Dodd, B. (2004). Auditory-visual speech integration by pre-linguistic infants: perception of an emergent consonant in the McGurk effect, *Dev. Psychobiol.* **44**, 209–220.

Calvert, G. A., Campbell, R. and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex, *Curr. Biol.* **10**, 649–657.

Campbell, C. and Massaro, D. (1997). Perception of visible speech: influence of spatial quantization, *Perception* **26**, 129–146.

Cienkowski, K. M. and Carney, A. E. (2002). Auditory-visual speech perception and aging, *Ear Hear.* **23**, 439–449.

Colin, C., Radeau, M., Deltenre, P., Demolin, D. and Soquet, A. (2002). The role of sound intensity and stop-consonant voicing on McGurk fusions and combinations, *Eur. J. Cogn. Psychol.* **14**, 475–491.

Colin, C., Radeau, M. and Deltenre, P. (2005). Top-down and ottom-up modulation of audiovisual integration in speech, *Eur. J. Cogn. Psychol.* **17**, 541–560.

Cotton, J. C. (1935). Normal 'visual hearing', *Science* **82**, 592–593.

De Gelder, B., Vroomen, J., Annen, L., Masthof, E. and Hodiamont, P. (2003). Audio-visual integration in schizophrenia, *Schizophr. Res.* **59**, 211–218.

Dekle, D., Fowler, C. and Funnell, M. (1992). Auditory-visual integration in perception of real words, *Percept. Psychophys.* **51**, 355–362.

Delbeuck, X., Collette, F. and Van der Linden, M. (2007). Is Alzheimer's disease a disconnection syndrome? Evidence from a crossmodal audio-visual illusory experiment, *Neuropsychologia* **45**, 3315–3323.

Demorest, M. E. and Bernstein, L. E. (1992). Sources of variability of speechreading sentences: a generalizability analysis, *J. Speech Hear. Res.* **35**, 876–891.

Déry, C., Campbell, N. K. J., Lifshitz, M. and Raz, A. (2014). Suggestion overrides automatic audiovisual integration, *Consc. Cogn.* **24**, 33–37.

Desai, S., Stickney, G. and Zeng, F. G. (2008). Auditory-visual speech perception in normal-hearing and cochlear-implant listeners, *J. Acoust. Soc. Am.* **123**, 428–440.

Desjardins, R. N. and Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Dev. Psychobiol.* **45**, 187–203.

Easton, R. D. and Basala, M. (1982). Perceptual dominance during lipreading, *Percept. Psychophys.* **32**, 562–570.

Erickson, L. C., Zielinski, B. A., Zielinski, J. E., Liu, G., Turkeltaub, P. E., Leaver, A. M. and Rauschecker, J. P. (2014). Distinct cortical locations for integration of audiovisual speech and the McGurk effect, *Front. Psychol.* **5**, 534. DOI:10.3389/fpsyg.2014.00534.

Erlebacher, A. and Sekuler, R. (1971). Response frequency equalization: a bias model for psychophysics, *Percept. Psychophys.* **9**, 315–320.

Eskelund, K., Tuomainen, J. and Andersen, T. S. (2011). Multistage audiovisual integration of speech: dissociating identification and detection, *Exp. Brain Res.* **208**, 447–457.

Eskelund, K., MacDonald, E. N. and Andersen, T. S. (2015). Face configuration affects speech perception: evidence from a McGurk mismatch negativity study, *Neuropsychologia* **66**, 48–54.

Fingelkurts, A. A., Krause, C. M., Mottonen, R. and Sams, M. (2003). Cortical operational synchrony during audio-visual speech integration, *Brain Lang.* **85**, 97–312.

Fingelkurts, A. A., Fingelkurts, A. A. and Krause, C. M. (2007). Composition of brain oscillations and their functions in the maintenance of auditory, visual and audio-visual speech percepts: an exploratory study, *Cogn. Proc.* **8**, 183–199.

Fixmer, E. and Hawkins, S. (1998). The influence of quality of information on the McGurk effect, in: *Proceedings of AVSP'98, Terrigal, Sydney, Australia*, pp. 27–32.

Gagné, J. P., Masterson, V., Munhall, K. G., Bilida, N. and Querengesser, C. (1994). Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech, *J. Acad. Rehabil. Audiol.* **27**, 135–158.

Gentilucci, M. and Cattaneo, L. (2005). Automatic audiovisual integration in speech perception, *Exp. Brain Res.* **167**, 66–75.

Grant, K. W. and Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences, *J. Acoust. Soc. Am.* **104**, 2438–2450.

Green, K. P. and Kuhl, P. K. (1991). Integral processing of visual place and auditory voicing information during phonetic perception, *J. Exp. Psychol. Hum. Percept. Perform.* **17**, 278–288.

Green, K. and Norrix, L. (1997). Acoustic cues to place of articulation and the McGurk effect: the role of release bursts, aspiration and formant transitions, *J. Speech Lang. Hear. Res.* **40**, 646–665.

Green, K. P., Kuhl, P. K. and Meltzoff, A. N. (1988). Factors affecting the integration of auditory and visual information in speech: the effect of vowel environment, *J. Acoust. Soc. Am.* **84**, S155. DOI:10.1121/1.2025888.

Green, K., Kuhl, P., Meltzoff, A. and Stevens, E. (1991). Integrating speech information across talkers, gender, and sensory modality: female faces and male voices in the McGurk effect, *Percept. Psychophys.* **50**, 524–536.

Gurler, D., Doyle, N., Walker, E., Magnotti, J. and Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements, *Attent. Percept. Psychophys.* **77**, 1333–1341.

Hardison, D. M. (1996). Bimodal speech perception by native and nonnative speakers of English: factors influencing the McGurk effect, *Lang. Learn.* **46**, 3–73.

Hessler, D., Jonkers, R., Stowe, L. and Bastiaanse, R. (2013). The whole is more than the sum of its parts — audiovisual processing of phonemes investigated with ERPs, *Brain Lang.* **124**, 213–224.

Hietanen, J. K., Manninen, P., Sams, M. and Surakka, V. (2001). Does audiovisual speech perception use information about facial configuration? *Eur. J. Cogn. Psychol.* **13**, 395–407.

Hillock-Dunn, A., Grantham, D. W. and Wallace, M. T. (2016). The temporal binding window for audiovisual speech: children are like little adults, *Neuropsychologia* **88**, 74–82.

Hirvenkari, L., Jousmäki, V., Lamminmäki, S., Saarinen, V. M., Sams, M. E. and Hari, R. (2010). Gaze-direction-based MEG averaging during audiovisual speech perception, *Front. Hum. Neurosci.* **4**, 17.

Irwin, J. R., Whalen, D. H. and Fowler, C. A. (2006). A sex difference in visual influence on heard speech, *Percept. Psychophys.* **68**, 582–592.

Irwin, J. R., Mencl, W. E., Frost, S. J., Chen, H. and Fowler, C. (2011). Functional activation for imitation of seen and heard speech, *J. Neurolinguist.* **24**, 611–618.

Jiang, J. and Bernstein, L. E. (2011). Psychophysics of the McGurk and other audiovisual speech integration effects, *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 1193–1209.

Jones, J. and Callan, D. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect, *NeuroReport* **14**, 1129–1133.

Jones, J. and Jarick, M. (2006). Multisensory integration of speech signals: the relationship between space and time, *Exp. Brain Res.* **174**, 588–594.

Jones, J. A. and Munhall, K. (1997). The effects of separating auditory and visual sources on audiovisual integration of speech, *Can. Acoust.* **25**, 13–19.

Jordan, T. R. and Bevan, K. (1997). Seeing and hearing rotated faces: influences of facial orientation on visual and audio-visual speech recognition, *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 388–403.

Jordan, T. R. and Sergeant, P. (1998). Effects of facial image size on visual and audio visual speech recognition, in: *Hearing by Eye II. The Psychology of Speechreading and Audiovisual Speech*, R. Campbell, B. Dodd and D. Burnham (Eds), pp. 155–176. Psychology Press, London, UK.

Jordan, T. R. and Sergeant, P. C. (2000). Effects of distance on visual and audio-visual speech recognition, *Lang. Speech* **43**, 107–124.

Jordan, T. R. and Thomas, S. (2001). Effects of horizontal viewing angle on visual and audio-visual speech recognition, *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 1386–1403.

Jordan, T. R., McCotter, M. V. and Thomas, S. M. (2000). Visual and audiovisual speech perception with color and gray scale facial images, *Percept. Psychophys.* **62**, 1394–1404.

Kanaya, S. and Yokosawa, K. (2011). Perceptual congruency of audio-visual speech affects ventriloquism with bilateral visual stimuli, *Psychonom. Bull. Rev.* **18**, 123–128.

Keane, B. P., Rosenthal, O., Chun, N. H. and Shams, L. (2010). Audiovisual integration in high functioning adults with autism, *Res. Autism Spectr. Disord.* **4**, 276–289.

Keil, J., Müller, N., Ihssen, N. and Weisz, N. (2012). On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states, *Cereb. Cortex* **22**, 221–231.

Kislyuk, D. S., Möttönen, R. and Sams, M. (2008). Visual processing affects the neural basis of auditory discrimination, *J. Cogn. Neurosci.* **20**, 2175–2184.

Lansing, C. R. and McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences, *Percept. Psychophys.* **65**, 536–552.

Leybaert, J., Macchi, L., Huyse, A., Champoux, F., Bayard, C., Colin, C. and Berthommier, F. (2014). Atypical audio-visual speech perception and McGurk effects in children with specific language impairment, *Front. Psychol.* **5**, 422. DOI:10.3389/fpsyg.2014.00422.

Lifshitz, M., Aubert Bonn, N., Fischer, A., Kashem, I. F. and Raz, A. (2013). Using suggestion to modulate automatic processes: from Stroop to McGurk and beyond, *Cortex* **49**, 463–473.

Luckner, J. L., Sebald, A. M., Cooney, J., Young III, J. and Muir, S. G. (2006). An examination of the evidence based literacy research in deaf edducation, *Am. Ann. Deaf* **150**, 443–455.

Lüttke, C. S., Ekman, M., Van Gerven, M. A. and De Lange, F. P. (2016). McGurk illusion recalibrates subsequent auditory perception, *Sci. Rep.* **6**, 32891. DOI:10.1038/srep32891.

Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J. and Parra, L. C. (2009). Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space, *PLoS One* **4**, e4638. DOI:10.1371/journal.pone.0004638.

MacDonald, J. and McGurk, H. (1978). Visual influences on speech perception processes, *Percept. Psychophys.* **24**, 253–257.

MacDonald, J., Andersen, S. and Bachmann, T. (2000). Hearing by eye: how much spatial degradation can be tolerated? *Perception* **29**, 1155–1168.

MacLeod, A. and Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use, *Br. J. Audiol.* **24**, 29–43.

Magnotti, J. F., Mallick, D. B., Feng, G., Zhou, B., Zhou, W. and Beauchamp, M. S. (2016). Similar frequency of the McGurk effect in large samples of native Mandarin Chinese and American English speakers, *Exp. Brain Res.* **233**, 2581–2586.

Malfait, N., Fonlupt, P., Centelles, L., Nazarian, B., Brown, L. E. and Caclin, A. (2014). Different neural networks are involved in audiovisual speech perception depending on the context, *J. Cogn. Neurosci.* **26**, 1572–1586.

Massaro, D. W. (1998). *Perceiving Talking Faces: from Speech Perception to a Behavioural Principle*. MIT Press, Cambridge, MA, USA.

Massaro, D. W. and Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception, *J. Exp. Psychol. Hum. Percept. Perform.* **9**, 753–771.

Massaro, D. W. and Cohen, M. M. (1996). Perceiving speech from inverted faces, *Percept. Psychophys.* **58**, 1047–1065.

Massaro, D. W. and Ferguson, E. (1993). Cognitive style and perception: the relationship between category width and speech perception, categorization, and discrimination, *Am. J. Psychol.* **106**, 25–38.

Massaro, D. W., Thompson, L. A., Barron, B. and Laron, E. (1986). Developmental changes in visual and auditory contributions to speech perception, *J. Exp. Child Psychol.* **41**, 93–113.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **265**, 746–748.

Miller, L. M. and D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech, *J. Neurosci.* **25**, 5884–5893.

Miller, G. A. and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants, *J. Acoust. Soc. Am.* **27**, 338–352.

Munhall, K. G., Gribble, P., Sacco, L. and Ward, M. (1996). Temporal constraints on the McGurk effect, *Percept. Psychophys.* **58**, 351–362.

Munhall, K. G., Ten Hove, M., Brammer, M. and Paré, M. (2009). Audiovisual integration of speech in a bistable illusion, *Curr. Biol.* **19**, 1–5.

Nahorna, O., Berthommier, F. and Schwartz, J. L. (2012). Binding and unbinding the auditory and visual streams in the McGurk effect, *J. Acoust. Soc. Am.* **132**, 1061–1077.

Nath, A. R. and Beauchamp, M. S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech, *J. Neurosci.* **31**, 1704–1714.

Nath, A. R. and Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion, *Neuroimage* **59**, 781–787.

Navarra, J., Alsius, A., Soto-Faraco, S. and Spence, C. (2010). Assessing the role of attention in the audiovisual integration of speech, *Inf. Fusion* **11**, 4–11.

Nelson, M. A. and Hodge, M. M. (2000). Effects of facial paralysis and audiovisual information on stop place identification, *J. Speech Lang. Hear. Res.* **43**, 158–171.

Norrix, L. W., Plante, E. and Vance, R. (2006). Auditory-visual speech integration by adults with and without language learning disabilities, *J. Commun. Disord.* **39**, 22–36.

Olson, I. R., Gatenby, J. C. and Gore, J. C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex, *Brain Res.* **14**, 129–138.

Palmer, T. D. and Ramsey, A. K. (2012). The function of consciousness in multisensory integration, *Cognition* **125**, 353–364.

Paré, M., Richler, R. C., Ten Hove, M. and Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: the influence of ocular fixations on the McGurk effect, *Percept. Psychophys.* **65**, 533–567.

Peynircioğlu, Z. F., Brent, W., Tatz, J. R. and Wyatt, J. (2017). McGurk effect in gender identification: vision trumps audition in voice judgments, *J. Gen. Psychol.* **144**, 59–68.

Proverbio, A. M., Massetti, G., Rizzi, E. and Zani, A. (2016). Skilled musicians are not subject to the McGurk effect, *Sci. Rep.* **6**, 30423. DOI:10.1038/srep30423.

Roa Romero, Y., Senkowski, D. and Keil, J. (2015). Early and late beta band power reflects audiovisual perception in the McGurk illusion, *J. Neurophysiol.* **113**, 2342–2350.

Roberts, M. and Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory, *Percept. Psychophys.* **30**, 309–314.

Romero, Y., Senkowski, D. and Keil, J. (2015). Early and late beta band power reflects audiovisual perception in the McGurk illusion, *J. Neurophysiol.* **113**, 2342–2350.

Rosenblum, L. D. and Saldaña, H. M. (1992). Discrimination tests of visually-influenced syllables, *Percept. Psychophys.* **52**, 461–473.

Rosenblum, L. D. and Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception, *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 318–331.

Rosenblum, L. D., Johnson, J. A. and Saldaña, H. M. (1996). Visual kinematic information for embellishing speech in noise, *J. Speech Hear. Res.* **39**, 1159–1170.

Rosenblum, L. D., Schmuckler, M. A. and Johnson, J. A. (1997). The McGurk effect in infants, *Percept. Psychophys.* **59**, 347–357.

Rosenblum, L. D., Yakel, D. A. and Green, K. P. (2000). Face and mouth inversion effects on visual and audiovisual speech perception, *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 806–819.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C. and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments, *Cereb. Cortex* **17**, 1147–1153.

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W. and Foxe, J. (2007). Seeing voices: high-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion, *Neuropsychologia* **45**, 587–597.

Sakamoto, S., Mishima, H. and Suzuki, Y. (2012). Effect of consonance between features and voice impression on the McGurk effect, *Interdiscip. Inf. Sci.* **18**, 83–85.

Saldaña, A. G. and Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor, *J. Acoust. Soc. Am.* **95**, 3658–3661.

Sams, M., Manninen, P., Surakka, V., Helin, P. and Kättö, R. (1998). McGurk effect in Finnish syllables, isolated words, and words in sentences: effects of word meaning and sentence context, *Speech Commun.* **26**, 75–87.

Schwartz, J. L. (2010). A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject dependent, *J. Acoust. Soc. Am.* **127**, 1584–1594.

Seewald, R., Ross, M., Giolas, T. G. and Yonovitz, A. (1985). Primary modality for speech perception in children with normal and impaired hearing, *J. Speech Lang. Hear. Res.* **28**, 36–46.

Sekiyama, K. (1998). Face or voice? Determinant of compellingness to the McGurk effect, in: *Proceedings of AVSP'98, Terrigal, Sydney, Australia*, pp. 33–36.

Sekiyama, K. and Tohkura, Y. (1991). 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.

Sekiyama, K., Soshi, T. and Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: a heightened McGurk effect in older adults, *Front. Psychol.* **5**, 323. DOI:10.3389/fpsyg.2014.00323.

Sommers, M., Tye-Murray, N. and Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults, *Ear Hear.* **26**, 263–275.

Soroker, N., Calamaro, N. and Myslobodsky, M. S. (1995). Ventriloquist effect reinstates responsiveness to auditory stimuli in the 'ignored' space in patients with hemispatial neglect, *J. Clin. Exp. Neuropsychol.* **17**, 243–255.

Soto-Faraco, S. and Alsius, A. (2009). Deconstructing the McGurk–MacDonald illusion, *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 580–587.

Stevenson, R. A., Zemtsov, R. K. and Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions, *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 1517–1529.

Strand, J., Cooperman, A., Rowe, J. and Simenstad, A. (2014). Individual differences in susceptibility to the McGurk effect: links with lipreading and detecting audiovisual incongruity, *J. Speech Lang. Hear. Res.* **57**, 2322–2331.

Sumby, W. H. and Pollack, I. (1954). Visual contribution to speech intelligibility in noise, *J. Acoust. Soc. Am.* **26**, 212–215.

Summerfield, Q. and McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels, *Q. J. Exp. Psychol. A* **36**, 51–74.

Surguladze, S. A., Calvert, G. A., Brammer, M. J., Campbell, R., Bullmore, E. T., Giampietro, V. and David, A. S. (2001). Audio-visual speech perception in schizophrenia: an fMRI study, *Psychiat. Res.* **106**, 1–14.

Szycik, G. R., Stadler, J., Tempelmann, C. and Münte, T. F. (2012). Examining the McGurk illusion using high-field 7 Tesla functional MRI, *Front. Hum. Neurosci.* **6**, 95. DOI:10.3389/fnhum.2012.00095.

Taylor, N., Isaac, C. and Milne, E. (2010). A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children, *J. Autism Dev. Disord.* **40**, 1403–1411.

Thomas, S. M. and Jordan, T. R. (2004). Contributions of oral and extra-oral facial motion to visual and audiovisual speech perception, *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 873–888.

Tiippana, K. (2014). What is the McGurk effect? *Front. Psychol.* **5**, 725. DOI:10.3389/fpsyg.2014.00725.

Tiippana, K., Andersen, T. S. and Sams, M. (2004). Visual attention modulates audiovisual speech perception, *Eur. J. Cogn. Psychol.* **16**, 457–472.

Tiippana, K., Puharinen, H., Möttönen, R. and Sams, M. (2011). Sound location can influence audiovisual speech perception when spatial attention is manipulated, *See. Perceiv.* **24**, 67–90.

Traunmüller, H. and Öhrström, N. (2007). Audiovisual perception of openness and lip rounding in front vowels, *J. Phon.* **35**, 244–258.

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F. and Theoret, H. (2007). Speech and non-speech audio-visual illusions: a developmental study, *PLoS One* **2**, e742. DOI:10.1371/journal.pone.0000742.

Tuomainen, J., Andersen, T. S., Tiippana, K. and Sams, M. (2005). Audiovisual speech perception is special, *Cognition* **96**, B13–B22.

Valkenier, B., Duyne, J. Y., Andringa, T. C. and Baskent, D. (2012). Audiovisual perception of congruent and incongruent Dutch front vowels, *J. Speech Lang. Hear. Res.* **55**, 1788–1801.

Van Engen, K. J., Xie, Z. and Chandrasekaran, B. (2016). Audiovisual sentence recognition is not predicted by susceptibility to the McGurk effect, *Atten. Percept. Psychophys.* **79**, 396–403.

Van Wassenhove, V., Grant, K. W. and Poeppel, D. (2007). Temporal window of integration in bimodal speech, *Neuropsychologia* **45**, 598–607.

Venezia, J. H., Thurman, S. M., Matchin, W., George, S. E. and Hickok, G. (2016). Timing in audiovisual speech perception: a mini review and new psychophysical data, *Attent. Percept. Psychophys.* **78**, 583–601.

Ver Hulst, P. J. (2006). Visual and auditory factors facilitating multimodal speech perception, *Senior Honors Thesis*, Ohio State University, Columbus, OH, USA.

Von Berg, S., McColl, D. and Brancamp, T. (2007). Moebius syndrome: measures of observer intelligibility with versus without visual cues in bilateral facial paralysis, *Cleft Palate Craniofacial J.* **44**, 518–522.

Walker, S., Bruce, V. and O'Malley, C. (1995). Facial identity and facial speech processing: familiar faces and voices in the McGurk effect, *Percept. Psychophys.* **57**, 1124–1133.

White, T. P., Wigton, R. L., Joyce, D. W., Bobin, T., Ferragamo, C. and Wasim, N. (2014). Eluding the illusion? Schizophrenia, dopamine and the McGurk effect, *Front. Hum. Neurosci.* **8**, 565. DOI:10.3389/fnhum.2014.00565.

Wiersinga Post, E., Tomaskovic, S., Slabu, L., Renken, R., De Smit, F. and Duifhuis, H. (2010). Decreased BOLD responses in audiovisual processing, *NeuroReport* **21**, 1146–1151.

Wilson, A., Alsius, A., Paré, M. and Munhall, K. (2016). Spatial frequency requirements and gaze strategy in visual-only and audiovisual speech perception, *J. Speech Lang. Hear. Res.* **59**, 601–615.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J. and McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech, *Cereb. Cortex* **13**, 1034–1043.

Yakel, D. A., Rosenblum, L. D. and Fortier, M. A. (2000). Effects of talker variability on speechreading, *Percept. Psychophys.* **62**, 1405–1412.

Youse, K. M., Cienkowski, K. M. and Coelho, C. A. (2004). Auditory-visual speech perception in an adult with aphasia, *Brain Inj.* **18**, 825–834.