Cross-Linguistic Variability in the McGurk Effect: A Computational Comparison of Polish and Turkish Native Speakers

Cengiz Acarturk (cengiz.acarturk@uj.edu.pl)[1] & Jakub Romanowski[1] & Melike Caglayan[1] & Ozgur Can Kayis[2,3] & Anna Maria Wilkosz[1,4] & Zehra Aleyna Yerelma[3] & Tugce Duger[3]

[1]Centre for Cognitive Science, Jagiellonian University, Poland
[2]Ludwig Maximilian University of Munich, Germany
[3]Department of Psychology, Hacettepe University, Turkiye
[4]Doctoral School in the Social Sciences, Jagiellonian University, Poland

Abstract

This study explores how native language influences audiovisual speech perception (the McGurk effect) by comparing Polish and Turkish speakers. Using a novel computational approach to quantify articulatory feature distances across diverse syllable pairs, the research reveals that Turkish speakers exhibit higher susceptibility to the effect, supporting the auditory intelligibility approach. However, the underlying audiovisual integration mechanism remains consistent across both languages. These findings suggest that while native linguistic structures modulate overall susceptibility, the articulatory drivers of speech perception are universal.

Keywords: McGurk effect; multisensory integration; Polish; Turkish; cross-linguistic speech perception

Introduction

Human communication is fundamentally based on multisensory processing (Stein & Meredith, 1993). In speech perception, multisensory processes integrate auditory signals and visual information for increasing speech intelligibility, thus the accuracy of speech perception (Sumby & Pollack, 1954; Binnie et al., 1974; Campbell & Dodd, 1987). Making the sensory inputs deliberately inconsistent leads to an illusion, namely the McGurk effect (McGurk & MacDonald, 1976). The McGurk effect has a perceptual outcome that emerges as an illusion: The listener perceives a syllable, not physically presented by auditory nor by visual sensory channel (e.g., perceiving /da/ when presented with auditory /ba/ and visual /ga/), demonstrating a tendency in the perceptual system towards generating a blended response. On the other hand, sensory modalities do not always result in integration (Colavita, 1974; Alais & Burr, 2004). When presented with conflicting inputs, the perceptual system may not integrate input into a blended response—as seen in the McGurk effect—it may instead experience cross-modal rivalry. This often leads to sensory dominance, where perception is reduced to its unimodal components and the observer relies entirely on either the visual or auditory signal (Tiippana, 2014). Accordingly, making two sensory inputs deliberately inconsistent, such as a phoneme and a viseme, may result in perception of either the phoneme or the viseme rather than experiencing the McGurk effect. A leading theoretical framework for this process is Massaro's Fuzzy Logical Model of Perception (FLMP). According to this model, auditory and visual features are processed independently and continuously before integrating them based on their relative reliability (Massaro, 1987; Cohen & Massaro, 1993). In the McGurk effect, the final percept is not a simple average, but a result of the system selecting the most probable interpretation that satisfies both inputs.

The McGurk effect is a robust phenomenon, and it has been employed to study neural mechanisms that contribute to audiovisual processing. For instance, functional magnetic resonance imaging (fMRI) studies have employed the McGurk stimuli to examine audiovisual integration, identifying the superior temporal sulcus (STS) as a likely cortical location of convergence (Calvert et al., 2000; Hickok & Poeppel, 2007; Nath & Beauchamp, 2012). Beauchamp et al. (2010) demonstrated that temporarily disrupting the activation of STS via Transcranial Magnetic Stimulation (TMS) alters the interaction between audiovisual inputs and reduces the McGurk effect, thus weakening multisensory integration without significantly affecting unimodal speech perception, and also going beyond correlation and reporting causal evidence. The interpretation of STS—as a specialized cortical area involved in visual-auditory speech processing—is consistent with the previous fMRI localization of multisensory responses in the STS (Calvert et al., 2000). It has also been supported by further studies that reported a decrease in the occurrence of the McGurk effect under focal disruption or inhibition of STS activity through brain stimulation (e.g., Marques et al., 2014).

The specific circumstances relevant to the McGurk effect, and more generally multisensory integration have received intense research interest for the past several decades. One is the individual perceptual traits related to the observation of the effect among individuals. For instance, Mallick et al. (2015) reported that the effect emerged in participants according to an all-or-nothing principle, meaning that individuals either experienced the illusion continuously or hardly experienced it at all. In their study, illusion rates spanned the entire spectrum from 0% to 100%, but they found that most participants (77%) either experienced the illusion very consistently (in more than 90% of trials) or very infrequently (in less than 10% of trials). They also demonstrated that the individual differences showed high test-retest reliability indicating a stable perceptual trait rather than a temporary fluctuation. These findings suggest the McGurk effect can be conceived as an individual trait.

To explain why some individuals consistently perceive the illusion while others do not, Magnotti & Beauchamp (2017) proposed the Causal Inference Model (CIM) of multisensory

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

speech perception, suggesting that the perceptual system must first evaluate the likelihood that the auditory and visual signals originated from the same causal source. According to this framework, individual differences arise from how different observers weigh this likelihood. Individuals with a high priority for integration are more likely to conclude the signals have a common source, resulting in a McGurk percept, whereas those with a low priority treat the signals as independent, leading to sensory dominance or rivalry. Further studies examining the relationship between personal traits and the McGurk effect have not revealed clear findings. For instance, gender has been shown not to be the primary determinant of multisensory speech integration. Mallick et al.'s (2015) large-scale data showed that participant or speaker gender had no main effect on the frequency of observation of the McGurk effect, under standard full-video viewing conditions. Nevertheless, Irwin et al. (2006) reported gender-related differences under specific circumstances of increased perceptual load. The study indicates the role of gender on multisensory integration, yet only under specific circumstances of perceptual overload.

Another individual trait, which was studied in relation to the McGurk effect is ageing. Although initial studies revealed no connection between the occurrence of the McGurk effect and age (e.g., Burnham & Dodd, 1996), recent studies have reported significant relationships. According to Hirst et al. (2018), the occurrence of the McGurk effect increases with age in childhood and reaches a level similar to that in adults at ages 10-12. Other researchers proposed that the McGurk effect further increases throughout life at later stages, a finding interpreted as a consequence of increased reliance on visual cues in relation to age-related hearing loss (Chládková et al., 2021). Consequently, there exists accumulated evidence that changes in physical conditions, thus the changes in reliance on environmental sensory cues, have an impact on the occurrence of the effect. An equally important aspect of the study of the McGurk effect, which has also been subject to intense research interest, is the linguistic context.

The Cross-Linguistic Context

To understand whether the McGurk effect is a universal perceptual phenomenon or one shaped by language-specific phonetic and cultural characteristics, research has increasingly focused on how the effect is observed across diverse linguistic contexts. Following the reports about the presence of the effect in English-speaking populations, Sekiyama and Tohkura (1991) demonstrated significant cross-linguistic variability, showing that Japanese-speaking participants are notably less susceptible to the effect than English speakers. This disparity has led to the development of several key hypotheses to explain why certain languages appear more resistant to visual influence. They formulated the auditory intelligibility hypothesis, suggesting that the effect is modulated by the clarity, thus reliability of the auditory signal. In relatively phonologically simple languages like Japanese, which has a limited set of vowels and unambiguous syllable structures, the auditory signal is often sufficient for accurate perception. Consequently, Japanese speakers rely on visual cues only when the auditory information is degraded or presented in a non-native language. Sekiyama & Tohkura (1993) elaborated on the initial proposal of the auditory intelligibility hypothesis, reporting that when participants were presented with stimuli in languages other than their native language, they relied more on visual information, thereby increasing the McGurk effect. Moreover, they proposed cultural differences as a likely reason for Japanese participants' reduced use of visual information. This proposal was established on their observation that Western cultures (e.g., Americans) tend to maintain direct eye contact and look at a speaker's face more frequently than Japanese speakers, for whom direct face-to-face gaze can be considered less common in social interactions. More recent studies expanded the findings in Japanese to divergent user groups, such as profoundly deafened children with cochlear implants (e.g., Tona et al., 2015).

Another factor influencing the cross-linguistic differences in the McGurk effect is the variation in prosodic and tonal structures between languages. It has been suggested that speakers of tonal languages rely more heavily on auditory information to distinguish meaning and are therefore less influenced by visual cues. Burnham and Lau (1998) supported this hypothesis by demonstrating that Cantonese-speaking participants exhibit a reduced McGurk effect compared to Japanese, and Japanese speakers compared to English speakers, relying significantly more on auditory information. This finding has further been observed in development stages, showing that the degree of visual influence increased over age in English-speaking but not Japanese-speaking children (Sekiyama & Burnham, 2008). However, the claim that the effect relies on variation in prosodic and tonal structures has been challenged by research in other languages, suggesting that linguistic complexity is not a universal factor in audiovisual speech integration. For example, studies comparing English and Mandarin Chinese have yielded conflicting results. Chen and Hazan (2009) reported that Mandarin speakers utilized visual information at levels comparable to English speakers. This was further corroborated by Magnotti et al. (2025), which reported no significant difference between Mandarin and English native speakers, suggesting that the effect of tone may be syllable-specific rather than a global linguistic trait.

The lack of consensus in current cross-linguistic research highlights the need to examine a broader range of languages with diverse phonological systems. The present study reports findings from Polish and Turkish native speakers by presenting a computational approach for evaluating audiovisual stimuli, and experimental findings from Polish and Turkish native speakers, thus a comparative analysis between the two languages.

Polish is characterized by high phonological complexity, consisting of numerous sibilants and complex consonant clusters. In a study using a wide array of syllable pairs, Majewski (2008) reported that inducing the McGurk effect in

Polish listeners is difficult, with participants selecting the auditory stimulus in over 94% of trials. This suggests a strong auditory dominance in Polish speech perception. In contrast, Turkish—an agglutinative language with shallow orthography-to-phonology mapping structure—has shown a high magnitude of McGurk susceptibility. Erdener (2015) reported that visual information carries greater importance for Turkish speakers than has been observed in many other languages. Moreover, Turkish speakers show even higher McGurk perception rates when presented with non-native stimuli, aligning with the auditory intelligibility approach, which suggests that integration increases as native auditory reliability decreases. By comparing these two understudied languages in the context of the McGurk effect, the present study aims to identify the specific articulatory drivers of these divergent integration patterns.

To move beyond descriptive comparisons, we employ a novel computational approach using the PanPhon (Mortensen et al., 2016) package to analyze the feature distance between auditory and visual stimuli. By quantifying differences in articulatory features (e.g., place, manner, and voicing), we investigate whether certain features are more likely to lead to McGurk effect than others. We compare performance across native speakers of Polish and Turkish to determine if these feature-level drivers are universal or language-specific. Accordingly, the study focuses on an empirical, a computational, and a theoretical research question: (1) Do Polish and Turkish native speakers differ in their overall susceptibility to the McGurk effect? Based on previous findings, we expect Turkish speakers to show significantly higher rates of the illusion than Polish speakers. (2) Which articulatory features (calculated via PanPhon phoneme distance) are relatively stronger predictors of the McGurk effect? We aim to identify whether certain phonetic categories consistently facilitate or inhibit the blended responses. (3) To what extent does the auditory intelligibility hypothesis explain the differences between these two languages? We will assess if the reported auditory and visual dominance in the two languages can be linked to the inherent phonological reliability of their respective systems.

Method

Stimuli and Experiment Design

The audiovisual stimuli were obtained from raw video recordings of four native Polish and Turkish speakers (two females) producing isolated syllables. A multi-step preprocessing pipeline was used to achieve temporal consistency. First, the onset times of each syllable were manually identified in Audacity, producing accurate recording onset values. To standardize temporal alignment across videos, these onsets were automatically adjusted such that each syllable began at 3000 ms of its corresponding clip. Audio and video streams were then split to different files. Lastly, stimulus-response pairings were automatically generated by Python scripts. This procedure produced a uniform corpus of temporally aligned audiovisual stimuli ready for presentation.

The final stimulus set consisted of short audiovisual recordings of one male native Polish speaker and one male native Turkish speaker, each containing the production of a single consonant–vowel (CV) syllable formed by combining a target consonant onset with the vowel /a/ (e.g., pa, ta, ka, ma, na, and so on, resulting in 125 syllables for Polish, and 77 syllables for Turkish).

Following previous work, three phonemes (ba, ma, and pa) were selected as the auditory stimuli, as these consonant sounds have been reported to reliably produce perceptual effects associated with the McGurk effect (McGurk & MacDonald, 1976; Sekiyama, 1994; Matchin, Groulx, & Hickok, 2014). In each trial, participants were presented with a single video recording. After the presentation of the video recording, three response options were presented: the first target syllable (the phoneme), the second target syllable (the viseme), and an open-response option in case the participant wants to submit an alternative syllable. To minimize systematic biases, the presentation of the stimuli and choice buttons were randomized across trials. Stimuli were presented in full screen at a resolution of 1040 × 585 pixels, with each video recording lasting six seconds. A practice session (nine trials) was presented to the participants before the experiment session, which consisted of 116 trials for Polish and 68 trials for Turkish.

Participants and Procedure

A total of 34 Polish native-speaker participants (22 females) and 34 Turkish native-speaker participants (20 females) took part in the study. Ages of the Polish participants ranged from 19 to 33 years (M = 23.1, SD ≈ 3.5). Ages of the Turkish participants ranged from 18 to 62 years (M = 33.88, SD ≈ 14.55). All participants reported Polish and Turkish, correspondingly, as their native language, which was a prerequisite for inclusion in the study.

Stimuli presentation and response collection were programmed within a web-based interface using HTML, CSS, and JavaScript integrated with JATOS for data management. All data (including demographics, item-level responses, and reaction times) were submitted anonymously and securely to the server. Ethics permission for the study was granted by the institutional ethics review board.

Results

Analysis Methodology

As the first step of the analysis, we labeled the stimuli (phonemes and visemes) by the syllables in the standard orthography of the corresponding language, using their equivalent sounds written in International Phonetic Alphabet (IPA) symbols. Since the stimuli consisted of a consonant followed by the same open central unrounded /a/ vowel, both in Turkish and Polish, the analysis was focused on the features of the consonant. We utilized PanPhon, a Python package that consists of a database of segments written in

IPA and their equivalent articulatory and phonological features, as well as methods for operations on them. The PanPhon database consists of over 5000 entries (IPA strings), each entry consisting of an IPA symbol represented by a vector of 21 articulatory features[1], and each feature having a ternary value of 1, 0, or -1.[2] The operations consisted of methods such as translating the IPA strings to corresponding features and comparing the feature vectors. We use the `feature_edit_distance()` method to calculate the distance between the phonemes in the stimuli, assuming a larger distance between two phonemes being an indicator of a larger difference between them. This method calculates the minimum string edit cost to get from the source to the target string, assuming each feature to be of equal weight but counting the difference between non-applicable features and set values to be less than the difference between two applicable values. The calculation results in a single numerical value for each phoneme-viseme combination in the stimuli, separately for the two languages of the experiment. For example, the classic pairing of auditory /ba/ with visual /ga/, often perceived as /da/, corresponds to a relatively high feature distance in this framework, which fits with the idea that larger mismatches increase the chance of fusion. We also used the package to translate each viseme to a list of features to examine which ones most frequently coincide with the McGurk effect.

McGurk Effect in Relation to Feature Vector Distances

Using the PanPhon library, we calculated an unweighted feature-based phoneme distance for every viseme-phoneme pair present in both the Polish and the Turkish datasets. We excluded two Polish phonemes ("nga" and "gia") and one Turkish phoneme ("ğa"), as they do not occur in isolation, nor in word beginnings. The resulting phoneme distance values for Polish ranged from 0 (indicating no viseme-phoneme discrepancy) to 0.438, with a mean value of M = 0.174 (SD = 0.125). For Turkish, the distances ranged from 0 to 0.354 with M = 0.151 (SD = 0.112). We juxtaposed these values with the mean McGurk rate (probability calculated as the percentage of McGurk trials within all trials) obtained from the participants. The mean McGurk rate for Polish was M = 26.8% (SD = 19.4%) and for Turkish, M = 38.2% (SD = 27.6%). Separate analyses for each language confirmed that feature distance was a significant predictor of McGurk susceptibility for both groups, though the effect was nearly twice as large in the Turkish sample ($\beta = 1.91$, $p < .001$) compared to the Polish sample ($\beta = 0.93$, $p = .006$). A combined Generalized Linear Mixed Model (GLMM) revealed a significant main effect of language ($\chi^2(1) = 10.09$, $p = .001$) and a highly significant interaction between distance and language ($\chi^2(1) = 44.43$, $p < .001$), with Turkish participants exhibiting higher overall susceptibility to the McGurk effect and increases in distance than Polish participants. A follow-up correlation analysis, however, showed a strong positive association between distance and McGurk effect for both Polish ($\rho = 0.739$, 95% CI [0.615, 0.829], $p < 0.001$), and Turkish ($\rho = 0.731$, 95% CI [0.599, 0.817], $p < 0.001$). The results are presented in Figure 1.

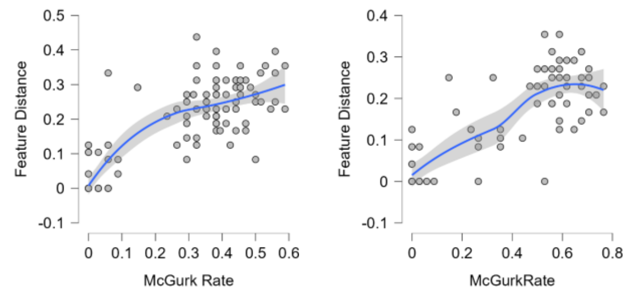

Figure 1: Feature-based phoneme distance - McGurk scatterplot for Polish (left) and Turkish (right).

This pattern of results was preserved after removing 0-distance stimuli, in other words the videos in which the phoneme and the viseme were the same. To sum up, Turkish stimuli elicited a higher ratio of the McGurk effect, even though the stimuli exhibited a smaller mean phoneme distance than in Polish while the general statistical observations remain similar. These findings indicate a statistically significant positive relationship between phoneme distance and susceptibility of the participants to experience the McGurk effect. Since Turkish and Polish are vastly different languages from differing families, the observed relationship has the potential to generalize to other languages as well. Below, we proceed with the analysis by taking a broader look at the distribution of percepts, including phoneme-dominant and viseme-dominant cases, alongside the McGurk percepts.

Distribution of Percepts across the Stimuli

Table 1 shows the percentage of the participants, who reported their percepts for each stimuli video recording. Based on the reported percepts, we classified each trial as a McGurk trial, a viseme-dominant trial, or a phoneme-dominant trial. Zero-distance stimuli were not included in the analysis. A close inspection of the results suggest that the Turkish native-speaker participants not only displayed a higher susceptibility to the McGurk effect, but they also seemed to rely on the phonetic aspect of the stimuli less compared to Polish native-speaker participants.

[1] syl: syllabic, son: sonorant, cons: consonantal, cont: continuant, delrel: delayed release, lat: lateral, nas: nasal, strid: strident, voi: voice, sg: spread glottis, cg: constricted glottis, ant: anterior, cor: coronal, distr: distributed, lab: labial, hi: high (vowel/consonant, not tone), lo: low (vowel/consonant, not tone), back: back, round: round, velaric: velaric airstream mechanism (click), tense: tense, long: long (Mortensen et al. 2016).

[2] The value 1 corresponds to the presence of a feature, -1 to its absence, and 0 to its non-applicability for a given type of sound.

Table 1: Responses across the stimuli (top 20). (P. Dom: Phoneme Dominance; V. Dom: Viseme Dominance).

Polish					Turkish				
Viseme	IPA	McGurk	V.Dom.	P.Dom.	Viseme	IPA	McGurk	V.Dom.	P.Dom.
sza	/ʃa/	0.48	0.29	0.23	ga	/ga/	0.69	0.19	0.13
hia	/xia/	0.47	0.27	0.25	sa	/sa/	0.66	0.20	0.15
ja	/ja/	0.46	0.29	0.24	ya	/ja/	0.64	0.21	0.16
zia	/ʑa/	0.46	0.21	0.33	ca	/dʒa/	0.62	0.18	0.21
la	/la/	0.44	0.29	0.26	za	/za/	0.61	0.19	0.21
sa	/sa/	0.44	0.31	0.26	ka	/ka/	0.59	0.30	0.11
ka	/ka/	0.44	0.29	0.28	ra	/ɾa/	0.59	0.23	0.19
ra	/ra/	0.43	0.31	0.25	ta	/ta/	0.57	0.28	0.15
cza	/t͡ʃa/	0.42	0.28	0.29	ja	/ʒa/	0.55	0.28	0.17
dża	/d͡ʒa/	0.42	0.24	0.34	şa	/ʃa/	0.55	0.22	0.24
ha	/xa/	0.42	0.28	0.31	ha	/ha/	0.53	0.33	0.14
rza	/ʒa/	0.41	0.32	0.26	ça	/tʃa/	0.50	0.27	0.23
wa	/va/	0.41	0.25	0.35	da	/da/	0.50	0.38	0.12
dza	/d͡za/	0.41	0.33	0.27	na	/na/	0.50	0.36	0.14
za	/za/	0.40	0.34	0.25	la	/la/	0.47	0.41	0.12
dzia	/d͡ʑa/	0.39	0.32	0.30	fa	/fa/	0.41	0.51	0.08
kia	/kʲa/	0.39	0.33	0.29	va	/va/	0.39	0.52	0.09
ga	/ɡa/	0.37	0.33	0.29	pa	/pa/	0.12	0.49	0.40
nia	/ɲa/	0.37	0.32	0.32	ba	/ba/	0.01	0.63	0.35
sia	/ɕa/	0.34	0.30	0.35	ma	/ma/	0.00	0.50	0.50

With respect to the likelihood of eliciting the McGurk effect, the Polish visemes exhibit a graded, monotonic pattern, with McGurk rates decreasing systematically across visemes. In contrast, the Turkish visemes show a more abrupt distribution, with all visemes except /pa/, /ba/, and /ma/, eliciting McGurk responses of at least 39%. Despite these cross-linguistic differences, there is some overlap in viseme ranking across languages. For instance, /ʃa/, /ja/, /ka/, /sa/, and /dʒa/ are all ranked highly in both Polish and Turkish. Regarding viseme dominance, /ba/ ranks the highest in both languages, with /ma/ also eliciting a substantial number of viseme-dominant responses. Similarly, /ma/, /pa/, and /ba/ elicited the highest number of phoneme-dominant responses, with /wa/ also ranking highly in Polish. It should be noted, however, that because only three sounds (/pa/, /ba/, and /ma/) were used to generate the video recordings (for keeping the duration of the experiment manageable), the resulting number of trials per viseme or phoneme is uneven. Consequently, these results should be interpreted as exploratory rather than statistically conclusive. This is particularly important given the limited sample size, uneven distribution of stimuli across conditions, and demographic variability between groups.

Viseme Features Predicting the McGurk Effect

The stimuli consisted of video recordings that voiced the three selected phonemes (/ma/, /ba/, /pa/) combined with all possible visemes in both languages. This was due to the practical limitation to keep the stimuli size feasible. Removing this constraint would lead to 10,000 video combinations including all possible combinations of phonemes and visemes. These three phonemes share many characteristics, most evidently, being both labial and plosive and, as such, being highly visually distinct from each other. Nevertheless, the selection of the three phonemes was an operational assumption needed for preparing the stimuli set, and this assumption prevents us scaling up the findings about the relationship between phoneme-specific features and McGurk percepts. On the other hand, the specific viseme features corresponding to more frequent McGurk percepts worths examining across the two languages.

To explore which articulatory features of the viseme-corresponding phonemes (henceforth, visemes) were most closely associated with audiovisual integration, we converted each viseme into feature vectors representing articulatory properties, using the PanPhon library. For each individual feature, we then calculated the mean rate of McGurk effect occurrence across all viseme-phoneme pairs in which that feature was present in the viseme. This calculation revealed an aggregated measure of how strongly each articulatory characteristic correlated with the McGurk effect. Among a total of 21 articulatory features specified by the PanPhon library, the top seven features observed, which drove the McGurk percepts in Turkish, were observed among the top 14 features in the Polish data. These features are shown in Table 2.

Table 2: The articulatory features and corresponding average McGurk effect percentage rates for Polish and Turkish, consisting of 14 features (out of 37).

Feature	Polish	Turkish
back (present)	0.275	0.539
high (present)	0.287	0.475
anterior (absent)	0.298	0.444
labial (absent)	0.296	0.440
voice (absent)	0.281	0.436
sonorant (absent)	0.277	0.417
distributed (absent)	0.282	0.409
coronal (present)	0.290	0.393
continuant (present)	0.302	0.380
delayed release (present)	0.295	0.385
strident (present)	0.301	0.379
nasal (absent)	0.277	0.396
consonantal (present)	0.272	0.400
round (absent)	0.274	0.394

The findings show that across both languages, McGurk effect rates were higher for features associated with limited visual salience, such as non-labial, posterior (back, anterior absent). In Polish, features reflecting fricative and affricate manner of articulation (continuant, strident, delayed release) also exhibited McGurk rates, consistent with the reduced visibility of such constrictions on the lips. Conversely, lower McGurk rates in both languages were observed for labial and rounded visemes, which provide clear and unambiguous visual cues to the observer. Taken together, these results indicate that visually informative phonemes mitigate the McGurk effect, whereas articulations with limited visual distinctiveness promote the effect in both languages. The observation that less salient labial features of the viseme, dissimilar to the accompanying phonemes, coincide more often with the McGurk effect further supports the *distance claim* in the present study. In other words, when the visual component lacks clear, distinctive cues, such as reduced lip rounding, incomplete closure, or articulations produced further back in the vocal tract, the perceptual system appears more prone to integrating conflicting auditory and visual signals into an integrated, McGurk percept. As all phonemes used were labial and plosive, any mouth movements of the viseme that could be seen to match, though, would result in participants responding in line with the phoneme rather than creating the McGurk percept.

In summary, the primary conclusion drawn from the analysis of the results is that feature-based distance constitutes a strong candidate predictor of the likelihood for eliciting the McGurk effect. Although Polish and Turkish differ in several linguistic dimensions, the strength of the effect, when measured by correlation, was comparable across the two languages. The juxtaposition of these typologically distinct languages, therefore, makes it feasible to hypothesize that this correlation may hold for other languages, as well.

Discussion and Conclusion

The present study investigated the cross-linguistic variability of the McGurk effect by comparing Polish and Turkish through an empirical and computational perspective. Our first and third research questions addressed the overall susceptibility differences between the two languages and the extent to which they align with the auditory intelligibility approach. The results provide support for our hypothesis that Turkish-speaking participants exhibited a significantly higher susceptibility to the McGurk effect compared to Polish-speaking participants. This difference provides supporting evidence for the auditory intelligibility approach. One factor that should be considered when interpreting the difference between groups is age. The Turkish sample had a higher mean age and a wider age range than the Polish sample. Previous work suggests that reliance on visual speech cues may increase with age (Hirst et al., 2018; Sekiyama et al., 2014; Rosemann et al., 2021; Diaz & Yalcinbas, 2021), suggesting that the higher McGurk rates in the Turkish group may partly reflect age-related differences rather than language alone. Specifically, in Polish, the high phonological complexity appears to have provided greater reliance on auditory signal, leading to a higher frequency of phoneme-dominant responses. Conversely, the phonetic structure of Turkish was more permeable to visual influence, leading to a higher frequency of McGurk responses. This finding is in line with Stacey et al. (2020), which reported that adding noise to the auditory signal significantly increases McGurk responses, while blurring the visual signal decreases them. The second research question aimed to identify the articulatory features that most effectively predicted the occurrence of the McGurk effect. By utilizing the PanPhon package to quantify feature-based phoneme distances, the present study aligns with the principles of the FLMP. If, as Massaro (1987) suggests, the perceptual system evaluates articulatory features independently, then the specific distance between auditory and visual features (such as place of articulation or voicing) should be a primary predictor of whether an individual experiences a blended percept or exhibits sensory dominance. Although phoneme distance is strongly related to McGurk responses in the present data, it should not be treated as a full explanation. Specifically, phoneme distance can be conceived as a measure of mismatch between modalities, rather than a complete account of the process. While Polish and Turkish differ significantly in their overall susceptibility, the strength of the correlation between the distance and the likelihood of the McGurk effect was comparable across both languages. This finding suggests that while the basis of integration might be set by the language's phonological complexity, the underlying mechanism of integration—based on articulatory features such as "back," "high," and "anterior" —might be a universal across human perceptual systems.

Our study has several limitations that should be considered when interpreting the results. First, the sample size is modest, which limits generalization. Second, the participant groups differ in age, with the Turkish sample showing a higher mean age and wider variability, which may have influenced the observed differences in McGurk susceptibility. Third, the stimulus set was restricted to a limited number of phonemes (/ba/, /ma/, /pa/), due to practical constraints, which limits the range of conclusions about feature effects. Studies with a wider variety of vowels are needed to examine whether the findings hold across a larger stimulus set. Another limitation concerns the use of equal weighting across articulatory features in the distance metric. Future work could assign weights to features based on perceptual relevance.

In conclusion, the McGurk effect is not a uniform phenomenon, but a dynamic, audiovisual process shaped by the characteristics of a language's phonological system. While Turkish native speakers are more visually driven and Polish native speakers more auditory-dominant, our findings indicate that both groups utilize similar articulatory feature-weighting mechanisms. Taken together, the results show that feature-based distance provides a useful way to connect phonological representations with observed perceptual responses in audiovisual speech.

Acknowledgments

This research has been supported by Jagiellonian University Strategic Programme Excellence Initiative Priority Research Area DigiWorld (ID.UJ) under the project title "Cognitive Aspects of Interaction and Communication in Natural and Artificial Agents".

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257-262

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *The Journal of Neuroscience,* 30(7), 2414–2417. https://doi.org/10.1523/JNEUROSCI.4865-09.2010

Binnie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech and Hearing Research, 17*(4), 619-630.

Burnham, D., Dodd, B. (1996). Auditory-visual speech perception as a direct process: The McGurk effect in infants and across languages. In Stork, D.G., Hennecke, M.E. (eds.) S*peechreading by humans and machines*. NATO ASI Series, vol 150. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-13015-5_7

Burnham, D., & Lau, S. (1998). The effect of tonal information on auditory reliance in the McGurk effect. In Proceedings of Audio-Visual Speech Processing Conference, AVSP'98 (pp. 37-42)

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current biology : CB, 10*(11), 649–657. https://doi.org/10.1016/s0960-9822(00)00513-3

Campbell, R., & Dodd, B. (1980). Hearing by eye. *The Quarterly Journal of Experimental Psychology, 32*(1), 85–99. https://doi.org/10.1080/00335558008248235

Chen, Y., & Hazan, V. (2009). Developmental factors and the non-native speaker effect in auditory-visual speech perception. *The Journal of the Acoustical Society of America, 126*(2), 858-865. https://doi.org/10.1121/1.3158823

Chládková, K., Podlipský, V. J., Nudga, N., & Šimáčková, Š. (2021). The McGurk effect in the time of pandemic: Age-dependent adaptation to an environmental loss of visual speech cues. *Psychonomic Bulletin & Review, 28*(3), 992–1002. https://doi.org/10.3758/s13423-020-01852-2

Cohen, M. M., & Massaro, D. W. (1993). Modeling coarticulation in synthetic visual speech. In *Models and techniques in computer animation* (pp. 139-156). Tokyo: Springer Japan.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics, 16*(2), 409-412

Diaz, M., & Yalcinbas, E. (2021). The neural bases of multimodal sensory integration in older adults. *International Journal of Behavioral Development*, 45, 409 - 417. https://doi.org/10.1177/0165025420979362

Erdener, D. (2015). Türkçede McGurk İllüzyonu [The McGurk Illusion in Turkish]. *Türk Psikoloji Dergisi, 30*(76), 19-27.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. Nature Reviews Neuroscience, 8(5), 393-402

Hirst, R. J., Stacey, J. E., Cragg, L., Stacey, P. C., & Allen, H. A. (2018). The threshold for the McGurk effect in audio-visual noise decreases with development. *Scientific Reports, 8*(1), 12372. https://doi.org/10.1038/s41598-018-30798-8

Irwin, J. R., Whalen, D. H., & Fowler, C. A. (2006). A sex difference in visual influence on heard speech. *Perception & Psychophysics, 68*(4), 582–592. https://doi.org/10.3758/bf03208760

Magnotti, J. F., & Beauchamp, M. S. (2017). A causal inference model explains perception of the McGurk effect and other incongruent audiovisual speech. *PLoS Computational Biology, 13*(2), e1005229. https://doi.org/10.1371/journal.pcbi.1005229

Magnotti, J. F., Basu Mallick, D., Feng, G., Zhou, B., Zhou, W., & Beauchamp, M. S. (2025). The McGurk effect is similar in native Mandarin Chinese and American English speakers. *Frontiers in Psychology*, 16, 1531566. https://doi.org/10.3389/fpsyg.2025.1531566

Majewski, W. (2008). McGurk effect in Polish listeners. *Archives of Acoustics, 33*(4), 447-454

Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review, 22*(5), 1299–1307. https://doi.org/10.3758/s13423-015-0817-4

Marques, L. M., Lapenta, O. M., Merabet, L. B., Bolognini, N., & Boggio, P. S. (2014). Tuning and disrupting the brain-modulating the McGurk illusion with electrical stimulation. *Frontiers in Human Neuroscience*, 8, 533. https://doi.org/10.3389/fnhum.2014.0053

Massaro, D. W., & Simpson, J. A. (1987). Speech perception by ear and eye: A paradigm for psychological inquiry. Hillsdale, NJ: Erlbaum.

Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI. *Journal of Cognitive Neuroscience, 26*(3), 606-620

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

Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., & Levin, L. (2016). Panphon: A resource for mapping IPA segments to articulatory feature vectors. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 3475-3484).

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage, 59*(1), 781–787. https://doi.org/10.1016/j.neuroimage.2011.07.024

Rosemann, S., Gieseler, A., Tahden, M., Colonius, H., & Thiel, C. (2021). Treatment of age-related hearing loss alters audiovisual integration and resting-state functional connectivity: A randomized controlled pilot trial. *eNeuro*, 8. https://doi.org/10.1523/eneuro.0258-21.2021

Sekiyama, K. (1994). Differences in auditory-visual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility. *Journal of the Acoustical Society of Japan (E), 15*(3), 143-158. https://doi.org/10.1250/ast.15.143

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science, 11*(2), 306-320. https://doi.org/10.1111/j.1467-7687.2008.00677.x

Sekiyama, K., Soshi, T., & Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: a heightened McGurk effect in older adults. *Frontiers in Psychology*, 5. https://doi.org/10.3389/fpsyg.2014.00323

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America, 90*(4 Pt 1), 1797–1805. https://doi.org/10.1121/1.401660

Sekiyama, K., & Tohkura, Y. I. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics, 21*(4), 427-444. https://doi.org/10.1016/S0095-4470(19)30229-3

Stacey, J. E., Howard, C. J., Mitra, S., & Stacey, P. C. (2020). Audio-visual integration in noise: Influence of auditory and visual stimulus degradation on eye movements and perception of the McGurk effect. *Attention, Perception, & Psychophysics, 82*(7), 3544-3557.

Stein, B. E., & Meredith, M. A. (1993). The merging of the senses. MIT Press

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212-215.

Tiippana K. (2014). What is the McGurk effect? *Frontiers in Psychology*, 5, 725. https://doi.org/10.3389/fpsyg.2014.00725

Tona, R., Naito, Y., Moroto, S., Yamamoto, R., Fujiwara, K., Yamazaki, H., Shinohara, S. & Kikuchi, M. (2015). Audio–visual integration during speech perception in prelingually deafened Japanese children revealed by the McGurk effect. *International Journal of Pediatric Otorhinolaryngology, 79*(12), 2072-2078. https://doi.org/10.1016/j.ijporl.2015.09.016