The impact of visual information in speech perception for individuals with hearing loss: a mini review

Ahyeon Choi[1], Hayoon Kim[1], Mina Jo[1], Subeen Kim[1], Haesun Joung[1], Inyong Choi[1,2] and Kyogu Lee[1,3,4]*

[1]Music and Audio Research Group, Department of Intelligence and Information, Seoul National University, Seoul, Republic of Korea, [2]Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA, United States, [3]Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, Republic of Korea, [4]Artificial Intelligence Institute, Seoul National University, Seoul, Republic of Korea

This review examines how visual information enhances speech perception in individuals with hearing loss, focusing on the impact of age, linguistic stimuli, and specific hearing loss factors on the effectiveness of audiovisual (AV) integration. While existing studies offer varied and sometimes conflicting findings regarding the use of visual cues, our analysis shows that these key factors can distinctly shape AV speech perception outcomes. For instance, younger individuals and those who receive early intervention tend to benefit more from visual cues, particularly when linguistic complexity is lower. Additionally, languages with dense phoneme spaces demonstrate a higher dependency on visual information, underscoring the importance of tailoring rehabilitation strategies to specific linguistic contexts. By considering these influences, we highlight areas where understanding is still developing and suggest how personalized rehabilitation strategies and supportive systems could be tailored to better meet individual needs. Furthermore, this review brings attention to important aspects that warrant further investigation, aiming to refine theoretical models and contribute to more effective, customized approaches to hearing rehabilitation.

KEYWORDS

speech perception, hearing loss, cochlear implant, visual information, cross-modal plasticity, multisensory integration, age-related variation, linguistic level

1 Introduction

Speech perception in individuals with hearing loss, presents a complex and multifaceted challenge. These individuals often rely on visual cues to compensate for their auditory limitations in everyday situations (Moradi et al., 2017; Tyler et al., 1997; Kaiser et al., 2003; Moody-Antonio et al., 2005; Kanekama et al., 2010). The seamless integration of auditory and visual information is particularly crucial for effective speech perception.

Previous research related to audio-visual speech perception has reported that visual cues such as lip movements, facial expressions, and gestures play a decisive role in understanding spoken language for those with hearing loss (Mitchel et al., 2023; Zhang et al., 2021; Regenbogen et al., 2012; Peelle and Sommers, 2015; Munhall et al., 2004). However, not all individuals with hearing loss benefit equally from visual information. The dependency on visual cues can significantly vary depending on factors such as the degree of hearing loss, age, and experience with communication aids like cochlear implants (CI) (Altieri and Hudock, 2014; Iler Kirk et al., 2007; Kanekama et al., 2010).

Contrary to the majority of studies that report benefits from visual information, some research shows no correlation between hearing loss and visual dependency (Lewis et al., 2018; Tye-Murray et al., 2007; Blackburn et al., 2019; Stevenson et al., 2017; Pralus et al., 2021). These studies may have been influenced by specific factors such as the subjects' aging (Tye-Murray et al., 2007), the experimental environment (Lewis et al., 2018; Blackburn et al., 2019), and the characteristics of the language stimuli (Stevenson et al., 2017; Pralus et al., 2021). Indeed, various factors, including age, experimental design, and hearing loss-related factors can contribute to the inconsistency in research outcomes (Bergeson et al., 2010).

To address these factors, this review aims to compare existing studies to gain comprehensive insights into how individuals with hearing impairments process and integrate visual cues in speech perception. Specifically, we conducted a thorough literature search using relevant keywords, such as "audiovisual" AND "speech perception" AND ("hearing impairment" OR "auditory disorder" OR "hearing defect"). The search was performed across Google Scholar and Scopus databases without restrictions on publication dates to ensure a comprehensive collection of relevant literature.

Through reviewing the extensive pool of studies collected, we identified key factors that contribute to varying outcomes in audio-visual speech perception research. These factors were categorized based on age, linguistic stimuli, and hearing loss-related factors including onset, intervention timing, and duration. Age was further classified into categories such as infants, children, adolescents, adults, and the elderly to assess the influence of developmental and aging processes. The stimuli were analyzed at different linguistic levels, including phonemes, syllables, words, and sentences, with experimental results organized into tables for detailed analysis. Additionally, cases of hearing loss were categorized based on onset, intervention timing, and duration, with interpretations of experimental results informed by these characteristics.

By evaluating various experiments from multiple perspectives according to these criteria, this review aims to provide a broad understanding of how individuals with hearing impairments process and integrate visual cues in speech perception. Ultimately, the findings from this review are expected to identify research gaps and offer valuable insights that could inform future studies in audio-visual speech perception for individuals with hearing loss.

2 Age-related variations in individuals with hearing loss

The first criterion we focused on is the age of individuals with hearing loss. Previous research has shown diverse age groups among participants, raising the possibility that visual information may or may not have an impact depending on age. Given the dynamic nature of language development, especially in early childhood, participant age becomes a crucial factor in integrating various sources of information. Therefore, we categorized participants into age groups encompassing infants, children, adolescents, adults, and seniors, exploring potential age-related changes in audiovisual speech perception.

Studies on infants and young children with hearing loss suggest that they become more adept at using visual information as they grow (Tona et al., 2015; Iler Kirk et al., 2007; Oryadi-Zanjani et al., 2017; Taitelbaum-Swead and Fostick, 2017). Literature frequently identifies the age of around 6 as a critical period for the initiation of sophisticated visual information utilization based on the onset of the McGurk effect and the beginning of audiovisual integration in the superior temporal sulcus (STS) (Tona et al., 2015; Iler Kirk et al., 2007; Oryadi-Zanjani et al., 2017).

Another study highlights 3.25 years as a significant age for beginning to integrate auditory and visual cues for speech perception (Holt et al., 2011). This difference can be attributed to the developmental stages of sensory integration, where basic multisensory integration starts around 3-4 years and becomes more refined by the age of 6. These developmental milestones indicate a gradual enhancement in utilizing visual information for speech perception. Compared to typically developing children, those with hearing loss are observed to rely more significantly on visual information during these early stages (Lalonde and McCreery, 2020; Oryadi-Zanjani et al., 2015, 2017; Taitelbaum-Swead and Fostick, 2017; Yamamoto et al., 2017), underscoring the role of early auditory information loss in fostering increased dependence on visual cues.

Despite the abundance of studies on infants and young children with hearing loss, research focusing specifically on adolescents is relatively scarce. While some studies include a range from infancy to adolescence, indicating that those with hearing loss may benefit more from visual information than their hearing counterparts (Tyler et al., 1997; Lamoré et al., 1998), there is a notable gap in studies dedicated exclusively to the adolescent group. This lack of research on adolescents with hearing loss suggests a need for more targeted investigation in this age group to better understand their unique challenges and benefits in utilizing visual information for speech perception.

Studies on individuals beyond the completion of language development, i.e., adults and seniors, do not pinpoint a specific age as crucial. Instead, they often analyze the influence of aging by comparing younger and older adult groups. While some studies report that older participants utilize visual information more as they age (Taitelbaum-Swead and Fostick, 2017; Puschmann et al., 2019; Hällgren et al., 2001), others suggest that the benefit of visual information remains consistent across age groups (Lasfargues-Delannoy et al., 2021; Hay-McCutcheon et al., 2005; Rigo and Lieberman, 1989). Studies specifically targeting seniors present conflicting results on how aging influences audiovisual integration, with some indicating potential benefits (Puschmann et al., 2019) and others suggesting no significant impact (Brooks et al., 2018; Tye-Murray et al., 2007). This variability underscores the importance of considering aging as a factor, dependent on experimental design and participant characteristics.

However, investigations into the neuromodulatory effects of hearing aid use and auditory training on audiovisual integration in seniors have revealed that these interventions can strengthen functional connectivity in the STS, similar to patterns observed during the developmental stages of audiovisual processing (Yu et al., 2017). While these findings are based on experiments with only two participants, they underscore the need for further research with larger sample sizes. These results highlight the necessity of considering the extent of hearing aid use and auditory

TABLE 1 Summary of visual salience per linguistic stimuli.

Stimuli	Subgroup	Description	Visual saliency	Reference
Phoneme	Labial consonants	/p/, /b/, /m/	High visual saliency due to distinct lip closure and release	(Lamoré et al., 1998; Tsao, 2019)
	Non-labial consonants	/t/, /d/, /k/, …	Less distinct than labials but still detectable through articulatory movements	(Tye-Murray et al., 2007)
	Vowels	/a/, /e/, /i/, /o/, /u/	Sustained vowels providing moderate to high visual saliency depending on the articulatory features	(Busby et al., 1984; Lei et al., 2008)
Syllables	–	CVC, VCV, CV, VC	High visual saliency, particularly for syllables with labial consonants	(Suess et al., 2022; Leybaert and LaSasso, 2010; Kishon-Rabin et al., 1997; Sato et al., 2020)
Words	–	Common or nonsense words	Moderate visual saliency, influenced by the recognizability of phoneme elements, particularly labial phonemes	(Puschmann et al., 2019; Taitelbaum-Swead and Fostick, 2017; Mantokoudis et al., 2013; Tye-Murray et al., 2016; Tsao, 2019)
Sentences	–	Simple declarative, complex sentences	Low to moderate visual saliency; effectiveness of visual information diminishes with increased linguistic complexity	(Moberly et al., 2020; Moody-Antonio et al., 2005; Hennesy et al., 2022; Tye-Murray et al., 2016; Tsao, 2019)

training when assessing the impact of aging on audiovisual speech perception.

3 Linguistic level differences in individuals with hearing loss

The stimuli used in speech perception experiments are highly diverse, and changes in the linguistic level of the stimuli can significantly impact the utilization of visual information. Therefore, we categorized the experiments based on linguistic levels, such as phonemes, syllables, words, sentences, and suprasegmental elements. We then performed a comprehensive literature review, synthesizing findings from various studies to identify and describe trends in visual information utilization according to the linguistic level of the stimuli (Table 1). This approach allowed us to qualitatively assess and summarize how different linguistic levels influence the reliance on visual information in speech perception.

Phonemes, as a foundational linguistic element, is particularly influenced by visual information, leading to notable improvements in speech perception. Researches consistently demonstrate that visual information significantly enhances phoneme perception for individuals with hearing loss (Tona et al., 2015; Huyse et al., 2013; Rouger et al., 2008). Both vowel identification (Busby et al., 1984) and consonant identification (Tye-Murray et al., 2007) experiments performed better in the audiovisual than audio-only or visual-only conditions. Researchers differ in their claims about whether vowels or phonemes are generally more visually salient to further aid audiovisual integration. Some studies argue that vowels benefit more from visual cues due to sustained articulatory features and clearer formant structure (Lei et al., 2008). In contrast, others assert consonants are more salient due to distinct visual cues (Van Soeren, 2023; Moradi et al., 2017). This discrepancy may arise from the specific phonemes selected for the experiments. In any case, studies consistently agree that labial consonants (e.g., /p/, /b/, /m/) are the most visually distinctive stimuli compared to vowels and coronal phonemes due to distinct articulatory features

such as lip closure and release, which are easily detectable by lip-reading (Lamoré et al., 1998; Tsao, 2019). In the same vein, Rouger et al. (2008) reported that French-speaking CI users placed greater reliance on visual information for labial phonemes, particularly under conditions of auditory uncertainty.

At the syllable level, individuals with hearing loss also benefit from audiovisual (AV) integration, facilitating speech perception (Leybaert and LaSasso, 2010; Kishon-Rabin et al., 1997; Sato et al., 2020). Studies reveal that labial syllables containing labial consonants provide greater visual cues, enhancing perceptual accuracy. For instance, a study on auditory-visual integration with hearing loss focusing on consonant recognition training highlighted overall improvements across all consonants, with significant gains observed in labial consonants within nonsense syllables (Grant et al., 1998). These findings align with those of Tona et al. (2015) and Yamamoto et al. (2017), who observed that Japanese children with cochlear implants (CIs) showed improved recognition of labial syllables (/ba/, /pa/) due to the distinct visual information provided by lip movements, compared to non-labial syllables (/da/, /ga/, /ka/, /ta/). Similarly, Suess et al. (2022) reported that bilabial syllables contribute more to the recognition of hearing-impaired, indicating that the informativeness of each viseme category varies subtly.

The influence of visual information on word recognition presents mixed results, though it generally has a positive impact. Tye-Murray et al. (2007) suggests that the recognizability of a word can be predicted based on the independent recognition of its phoneme elements as transmitted components. Bernstein et al. (2000) examined how certain phonological features influence word recognition visually, pointing out that words with round and coronal phonological traits were more effectively transmitted than those with continuant or voicing traits. This suggests that labial phonemes play a crucial role in transmitting speech and lexical content in an audiovisual context. Accordingly, studies emphasize the positive impact of visual information on word recognition (Puschmann et al., 2019; Taitelbaum-Swead and Fostick, 2017; Mantokoudis et al., 2013). However, Altieri and Hudock (2014) have argued that audiovisual integration does

not provide significant benefits over auditory-only input in word recognition, challenging the general assumption of the advantages of visual information.

Sentence recognition demonstrates a consistently lower visual information effectiveness than stimuli of smaller units. Some studies show increased reliance on visual enhancement during sentence recognition for individuals with severe hearing impairment (Moberly et al., 2020; Moody-Antonio et al., 2005; Hennesy et al., 2022). However, others suggest that the hypothesis that hearing impairment enhances lipreading ability is supported at the word level but not at the sentence level (Tye-Murray et al., 2007). Similarly, Tsao (2019) highlighted audiovisual benefits at the word level, but it remained unclear whether these benefits extend to larger linguistic units, such as sentences. Words and sentences involve additional cognitive processes like lexical access and syntactic parsing, which can either enhance or diminish the role of visual information depending on the context (Jackendoff, 2002). For instance, in sentence-level stimuli, visual cues related to prosody and facial expressions become more relevant, especially for conveying emotions or stress patterns, which are crucial for understanding the overall message (Buchan et al., 2008).

Apart from a trend in segmental units such as phoneme, syllable, word, and sentence, results indicate that for suprasegmental elements, such as intonation and stress within utterance components, visual information does not provide substantial aid (Stevenson et al., 2017; Tsao, 2019). According to Stevenson et al. (2017), CI users derive meaningful visual assistance in phonemic perception and, to a certain extent, for word and sentence recognition tasks. Yet, this assistance is not extended to suprasegmental information. Furthermore, a study on Chinese users (Tsao, 2019) exploring the enhancement of speech perception through AV integration does not demonstrate a discernible contribution of visual elements to tone differentiation.

Overall, smaller linguistic units consistently show benefits from visual information for individuals with hearing impairment, who are strongly affected by labial phonological features. However, as linguistic units increase in complexity, the influence of visual information becomes less pronounced. For smaller units like phonemes and syllables, listeners detect speech by assigning the signal more directly to a phonetic category. In contrast, sentence recognition requires listeners to access phonetic and lexical representations and make corresponding decisions.

4 Hearing loss factors: onset, intervention timing, and cross-modal plasticity

In the preceding two chapters, we reviewed how factors such as age and linguistic level can lead to varying outcomes in the utilization of visual information for speech perception among individuals with hearing loss. Another crucial factor to consider is the characteristics related to hearing loss itself. Specifically, we aim to compare the utilization of visual information based on the onset of hearing loss and the timing of cochlear implant (CI) and hearing aid (HA) interventions.

The timing of hearing loss onset and the point at which auditory devices are worn can significantly influence how individuals perceive speech, moderating the extent to which visual information is utilized (Tyler et al., 1997; Tona et al., 2015; Colmenero et al., 2004; Bavelier et al., 2006; Stevens and Neville, 2006). Additionally, since the onset of hearing loss and the timing of intervention can affect cross-modal plasticity (Kral and Sharma, 2023; Buckley and Tobey, 2011), it is important to examine how these factors impact the utilization of visual information. Therefore, this chapter aims to review (1) the onset and intervention timing of hearing loss, and (2) cross-modal plasticity on the utilization of visual information for speech perception among individuals with hearing loss.

4.1 Three cases based on the onset and intervention timing of hearing loss

To better understand the impact of intervention timing on speech perception outcomes in individuals with hearing loss, we have created a schematic representation of hearing loss onset and intervention timing across pre-lingual and post-lingual periods. Figure 1 presents three case categories based on these factors. The pre-lingual period, typically occurring before age 6, encompasses the critical period for language development. Case 1 involves pre-lingual hearing loss with corresponding pre-lingual intervention. Case 2 represents post-lingual hearing loss with post-lingual intervention, further divided into Case 2.1 (shorter duration before intervention) and Case 2.2 (longer duration). Case 3 illustrates pre-lingual hearing loss with delayed intervention after language acquisition.

The onset of hearing loss, whether it occurs before or after language acquisition, plays a crucial role in shaping the development of cognitive and perceptual abilities. A study by AuerJr and Bernstein (2007) highlights that individuals with early-onset hearing loss exhibit superior speech-reading abilities compared to those with normal hearing. This finding suggests that these individuals may develop compensatory strategies that enhance their capacity to interpret speech through visual cues, such as lip movements and facial expressions.

Intervention timing is closely related to the duration of hearing loss, with early intervention corresponding to a "shorter duration" (e.g., Case 1, 2.1) and delayed intervention to a longer duration (e.g., Case 2.2, 3). According to research by Gilley et al. (2010), early intervention leads to better audiovisual (AV) integration abilities, even when cochlear implants are worn for the same duration. This suggests that a shorter duration of hearing loss before intervention results in better cognitive outcomes in AV speech processing.

Moreover, in the case of early intervention, particularly for early-onset hearing loss (e.g., Case 1), Jerger et al. (2020) compared children with pre-lingual hearing loss who began using hearing aids around age 2 to their normal-hearing peers. The study found that speech detection and attention significantly improved when visual information was added to auditory-only stimuli. Building on the McGurk effect, Tona et al. (2015) demonstrated that children who received cochlear implants before age 3 utilized visual information to supplement auditory information. Similarly, Iler Kirk et al. (2007) suggested that children with cochlear implants
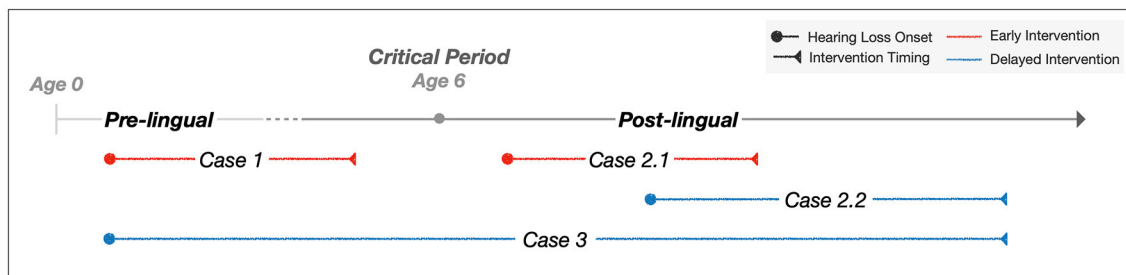
FIGURE 1
Schematic representation of hearing loss onset and intervention timing across pre-lingual and post-lingual periods.

could significantly enhance word recognition through audiovisual (AV) integration facilitated by lip reading.

Comparing intervention timings for individuals with the same onset of hearing loss reveals significant differences in outcomes. Both cases involve early-onset hearing loss, with Case 1 receiving early intervention and Case 3 undergoing intervention later in life. Case 3 involves individuals who experienced pre-lingual deafness but received intervention only after acquiring some language skills, indicating a longer duration of hearing loss. Although such cases are relatively rare and have received limited research attention, Moody-Antonio et al. (2005) provide evidence that some adults with congenital hearing impairment show significant improvements in speech perception after receiving intervention. This suggests that even without early auditory language experience, individuals can integrate auditory information from the intervention with visual language cues, demonstrating potential for enhanced speech perception despite missing the 'critical period' for early intervention.

However, differences in the use of visual information due to the timing of hearing loss and intervention between Case 1 and Case 3 are noteworthy. Research indicates that earlier intervention generally leads to better outcomes compared to later intervention. For instance, Tyler et al. (1997) found higher scores in tasks like vowel recognition and word comprehension when intervention occurred before age 5. Kral et al. (2019) suggest that intervening before age 3 is most effective for language acquisition. Additionally, Kanekama et al. (2010) highlight that intervention within 5 years of onset is particularly beneficial.

Comparing cases with the same onset and intervention timing but differing durations highlights the impact of the duration of hearing loss on language recognition outcomes. Case 2.1 experienced intervention after a relatively short duration of hearing loss, while Case 2.2 had a longer duration before intervention. The decision to categorize Case 2.1 and Case 2.2 based on duration stems from the limitations in existing research. In the study by Baskent and Bazo (2011), the wide age range of participants complicates the ability to isolate the effects of duration due to cognitive decline and age-related factors. Additionally, many studies do not provide specific details about the duration of hearing loss, making it difficult to draw clear conclusions about the effects of different durations within the post-lingual hearing loss group.

Bernhard et al. (2021) suggest that language recognition, assessed through sentence and single-word identification, shows a negative correlation with the duration of hearing loss before cochlear implantation. Lee et al. (2021) confirmed that the duration of hearing loss significantly negatively impacts both audio-only and audio-visual speech recognition. These findings indicate that even with similar onset and intervention timing, a shorter duration of hearing loss (Case 2.1) is more advantageous for language recognition compared to a longer duration (Case 2.2).

In this section, we have analyzed interventions broadly, including both cochlear implants (CI) and hearing aids (HA). However, further detailed comparative analysis of each type of assistive device is warranted. Additionally, it would be valuable to explore not only the duration of hearing loss but also the impact of the length of time using hearing aids or cochlear implants on training outcomes. Furthermore, to address the research gaps identified in the three cases discussed, there is a need for targeted additional studies for each case to better understand their specific contexts and outcomes.

4.2 Cross-modal plasticity with hearing loss patient

Cross-modal plasticity is commonly observed in individuals with hearing impairments and continues to be a factor even after cochlear implantation (Kral and Sharma, 2023; Lomber et al., 2010; Stropahl and Debener, 2017; Bavelier and Hirshorn, 2010; Campbell and Sharma, 2016). when individuals with hearing impairments are deprived of auditory input, much of the auditory cortex is commandeered by visual processing to compensate for the lack of auditory input. This cross-modal compensatory reorganization, where increased emphasis is placed on processing visual information, is a crucial aspect of adaptation in those with hearing impairments and has been pointed out as an important factor in the individual differences in their audio visual speech perception.

Previous studies have often shown that cross-modality may disrupt the integrative functions of the auditory cortex, potentially hindering speech perception adaptation in CI users (Liang et al., 2017). However, plasticity does not refer to a fixed or unchanging property, but rather to the ability to change for adaptation. In other words, the reorganization of the auditory cortex, which has been largely occupied to compensate for adaptation and deprivation before cochlear implant (CI) implantation, can change again after implantation. For example, a study on cross-modal plasticity

before and after implantation due to visual speech revealed that cross-modal activation in the auditory regions after implantation is associated with better speech understanding, and that the simultaneous development of activation in the auditory cortex by both visual and auditory speech provides adaptive benefits (Anderson et al., 2017). Another study demonstrated that the imbalanced cross-modal activity in the auditory cortex observed in individuals with hearing impairments returned to normal levels comparable to those of the control group after cochlear implant (CI) implantation. Additionally, the activation of Broca's area, which showed lower activity during speechreading compared to that of normal individuals, increased after CI implantation. Broca's area is known as a brain region that contributes to speech production. This suggests that the recovery of auditory input through CI can plastically adjust the imbalance in the auditory cortex (Rouger et al., 2012).

There is also research suggesting that the use of visual cues varies depending on the proficiency of CI users. For example, Kelly N. Jahn observed that proficient CI users process visual cues differently than non-proficient users, supported by brain imaging studies showing distinct patterns of brain activation in response to visual cues (Jahn et al., 2017; Doucet et al., 2006; Sharma et al., 2015). However, the differences in visual cue utilization patterns suggest a complex relationship with cross-modality, but the exact correlation remains unclear. This is a topic that will require further exploration and research in the future.

The divergence in findings regarding the changes and adaptiveness of cross-modality may be attributed to various factors such as the timing of hearing loss onset–whether it is prelingual or postlingual, the duration of hearing impairment-how long one has experienced hearlong loss. Although previous studies have categorized participants based on the onset of hearing loss, they have observed cross-modality in both groups. Variables like age, language, and the nature of the tasks employed in the studies may also contribute to these varied results. To unravel this complexity, future research must refine the control of these variables. While there is extensive literature on the cross-modality of CI users, particularly concerning the clinical impact of CI on speech perception outcomes, more in-depth studies are needed on the long-term evolution of cross-modality after cochlear implantation and its influence on the integrated perception of speech. Additionally, with advancements in cochlear implant technology, the time required to achieve adaptive and balanced cross-modality may decrease. This remains a promising topic for further exploration and should be actively discussed in future research.

5 Discussion

This review has explored the intricate dynamics of AV speech perception in individuals with hearing loss, with a particular focus on the influences of age, stimuli diversity, and hearing loss factors. Our qualitative analysis reveals that these factors significantly modulate the reliance on and the efficacy of visual information in supplementing auditory cues for speech perception. Building on these findings, we discuss the research gaps related to each factor,

propose directions for future studies, and highlight the implications and potential applications of our results.

5.1 Age-related variation

Based on the review of literature across different age groups, it is evident that early life stages of language and cognitive development, as well as the later stages of aging, are particularly critical periods that can significantly influence the use of visual information in speech perception. Specifically regarding aging, considering the variability in results due to task differences and sensory loss (Brooks et al., 2018), additional research that controls for task and age-related characteristics of participants in a consistent manner across age groups is warranted. Furthermore, given the lack of literature on adolescents, recruiting participants from all age groups–infants, children, adolescents, adults, and the elderly–and measuring AV speech perception using the same experimental design would provide clearer insights into age-related effects. Additionally, considering the neuromodulatory effects of aging on sensory integration in speech perception (Yu et al., 2017), there is a need for age-tailored rehabilitation strategies for individuals with hearing loss.

5.2 Linguistic diversity

The variability in response to different linguistic stimuli highlights the complexity of AV speech perception. Labial consonants, for example, are more readily enhanced by visual cues, pointing to the significant role of visually salient speech features in improving perception. However, the effectiveness of AV integration in speech perception is influenced by language-specific viseme configurations, with languages having denser phoneme spaces (e.g., English, French) showing a higher visual dependency. Investigating the efficacy of such tailored interventions across diverse linguistic contexts remains an essential avenue for future research. Also, there is a need for further comparative analysis across linguistic levels, particularly examining the relationship between phoneme recognition and higher linguistic units, as discussed in studies such as Lamoré et al. (1998) to understand how visual information contributes to speech perception as linguistic complexity increases. Moreover, this finding calls for a nuanced approach to auditory rehabilitation, suggesting tailored AV training programs emphasizing visually salient features to enhance speech perception for hearing loss such as Schumann and Ross (2022), or even further, strengthening the connection between these phonetic and lexical representations (Tye-Murray et al., 2016).

5.3 Hearing loss factors

This review has highlighted the critical influence of the onset of hearing loss and the timing of intervention on the utilization of visual information in speech perception. Early intervention, particularly in cases of pre-lingual hearing loss, significantly enhances speech perception by optimizing the use of visual cues.

In contrast, later interventions, whether in pre-lingual or post-lingual cases, show varied outcomes influenced by factors such as cognitive decline and the duration of hearing loss. A key factor in these processes is cross-modal plasticity, where the auditory cortex is reorganized to process visual information in response to auditory deprivation. Although this reorganization can initially support speech perception through enhanced visual reliance, its persistence post-cochlear implantation may disrupt the balance between auditory and visual inputs. Therefore, understanding the long-term effects of cross-modal plasticity, particularly how it evolves after CI implantation, is crucial. Future research should investigate the impact of the duration of hearing loss before intervention and how individual differences in cross-modal plasticity influence speech perception outcomes. Additionally, the development of adaptive technologies and tailored rehabilitation strategies that consider these factors could significantly improve audiovisual integration for CI users, ultimately enhancing their speech perception abilities in diverse listening environments.

5.4 General discussion

Moving forward, future research should prioritize the practical applications of visual information in hearing rehabilitation. Developing guidelines for personalized rehabilitation plans that consider age, type of stimuli, and specific characteristics of hearing loss will be crucial. One promising approach involves the integration of AI-assisted training systems, which can customize learning paths based on the user's individual characteristics, such as age, degree of hearing loss, and the use of assistive devices. These systems can provide real-time feedback and adaptive learning by analyzing how effectively a user utilizes visual cues during speech perception tasks. For instance, if a user struggles with specific phonemes or words, the AI can detect this and adjust the training regimen, offering additional practice or emphasizing visual cues more heavily to optimize learning outcomes.

Moreover, AI systems can support the learning of various linguistic and phonetic features, tailoring exercises to the user's linguistic background. For example, in languages with dense phoneme spaces like English, the system could focus more on tasks that rely heavily on visual cues, whereas in vowel-centric languages, it could emphasize the visual recognition of vowels. Additionally, these systems can strengthen the integration of visual and auditory information by using training methods based on phenomena like the McGurk effect, helping users more effectively combine these sensory inputs. The long-term tracking and analysis of a user's progress would allow for continuous refinement of personalized learning strategies, which is particularly beneficial for those receiving later-stage interventions.

By focusing on these areas, we can create more effective, tailored interventions that fully leverage both auditory and visual cues. Such advancements will not only improve speech perception and communication abilities but also significantly enhance the quality of life for those affected by hearing loss. Through this review and discussions, we have gained a deeper understanding of

how visual information influences speech perception in individuals with hearing loss, particularly in relation to the critical roles of age, linguistic diversity, and the timing of hearing loss. By integrating these variables into future studies on audiovisual speech perception, we can refine theoretical models, bridge gaps in current research, and enhance the predictive power of future investigations. The ongoing exploration of these factors is essential for advancing our knowledge of sensory integration processes and developing more holistic and personalized approaches to hearing rehabilitation.

Author contributions

AC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing. HK: Conceptualization, Investigation, Visualization, Writing – original draft, Writing – review & editing. MJ: Conceptualization, Investigation, Writing – original draft. SK: Conceptualization, Investigation, Visualization, Writing – original draft. HJ: Data curation, Investigation, Writing – review & editing. IC: Writing – review & editing. KL: Project administration, Supervision, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Altieri, N., and Hudock, D. (2014). Hearing impairment and audiovisual speech integration ability: a case study report. *Front. Psychol.* 5:678. doi: 10.3389/fpsyg.2014.00678

Anderson, C. A., Wiggins, I. M., Kitterick, P. T., and Hartley, D. E. (2017). Adaptive benefit of cross-modal plasticity following cochlear implantation in deaf adults. *Proc. Nat. Acad. Sci.* 114, 10256–10261. doi: 10.1073/pnas.1704785114

AuerJr, E. T., and Bernstein, L. E. (2007). Enhanced visual speech perception in individuals with early-onset hearing impairment. *J. Speech Lang. Hear. Res.* 50, 1157–1165. doi: 10.1044/1092-4388(2007/080

Baskent, D., and Bazo, D. (2011). Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment. *Ear Hear.* 32, 582–592. doi: 10.1097/AUD.0b013e31820fca23

Bavelier, D., Dye, M. W., and Hauser, P. C. (2006). Do deaf individuals see better? *Trends Cogn. Sci.* 10, 512–518. doi: 10.1016/j.tics.2006.09.006

Bavelier, D., and Hirshorn, E. A. (2010). I see where you're hearing: how cross-modal plasticity may exploit homologous brain structures. *Nat. Neurosci.* 13, 1309–1311. doi: 10.1038/nn1110-1309

Bergeson, T. R., Houston, D. M., and Miyamoto, R. T. (2010). Effects of congenital hearing loss and cochlear implantation on audiovisual speech perception in infants and children. *Restor. Neurol. Neurosci.* 28, 157–165. doi: 10.3233/RNN-2010-0522

Bernhard, N., Gauger, U., Romo Ventura, E., Uecker, F. C., Olze, H., Knopke, S., et al. (2021). Duration of deafness impacts auditory performance after cochlear implantation: a meta-analysis. *Laryngoscope Investigat. Otolaryngol.* 6, 291–301. doi: 10.1002/lio2.528

Bernstein, L., Tucker, P., and Demorest, M. (2000). Speech perception without hearing. *Percept. Psychophys.* 32, 233–252. doi: 10.3758/BF03205546

Blackburn, C. L., Kitterick, P. T., Jones, G., Sumner, C. J., and Stacey, P. C. (2019). Visual speech benefit in clear and degraded speech depends on the auditory intelligibility of the talker and the number of background talkers. *Trends Hear.* 23:2331216519837866. doi: 10.1177/2331216519837866

Brooks, C. J., Chan, Y. M., Anderson, A. J., and McKendrick, A. M. (2018). Audiovisual temporal perception in aging: The role of multisensory integration and age-related sensory loss. *Front. Hum. Neurosci.* 12:192. doi: 10.3389/fnhum.2018.00192

Buchan, J., Pare, M., and Munhall, K. (2008). The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain Res.* 1242, 162–171. doi: 10.1016/j.brainres.2008.06.083

Buckley, K. A., and Tobey, E. A. (2011). Cross-modal plasticity and speech perception in pre-and postlingually deaf cochlear implant users. *Ear Hear.* 32, 2–15. doi: 10.1097/AUD.0b013e3181e8534c

Busby, P. A., Tong, Y. C., and Clark, G. M. (1984). Underlying dimensions and individual differences in auditory, visual, and auditory-visual vowel perception by hearing–impaired children. *J. Acoust. Soc. Am.* 75, 1858–1865. doi: 10.1121/1.390987

Campbell, J., and Sharma, A. (2016). Visual cross-modal re-organization in children with cochlear implants. *PLoS ONE* 11:e0147793. doi: 10.1371/journal.pone.0147793

Colmenero, J., Catena, A., Fuentes, L., and Ramos, M. (2004). Mechanisms of visuospatial attention in deafness. *Eur. J. Cognit. Psychol.* 16, 791–805. doi: 10.1080/09541440340000312

Doucet, M., Bergeron, F., Lassonde, M., Ferron, P., and Lepore, F. (2006). Cross-modal reorganization and speech perception in cochlear implant users. *Brain* 129, 3376–3383. doi: 10.1093/brain/awl264

Gilley, P., Sharma, A., Mitchell, T., and Dorman, M. (2010). The influence of a sensitive period for auditory-visual integration in children with cochlear implants. *Restor. Neurol. Neurosci.* 28, 207–218. doi: 10.3233/RNN-2010-0525

Grant, K., Walden, B., and Seitz, P. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *J. Acoust. Soc. Am.* 103, 2677–2672. doi: 10.1121/1.422788

Hällgren, M., Larsby, B., Lyxell, B., and Arlinger, S. (2001). Evaluation of a cognitive test battery in young and elderly normal-hearing and hearing-impaired persons. *J. Am. Acad. Audiol.* 12, 357–370. doi: 10.1055/s-0042-1745620

Hay-McCutcheon, M. J., Pisoni, D. B., and Kirk, K. I. (2005). Audiovisual speech perception in elderly cochlear implant recipients. *Laryngoscope* 115, 1887–1894. doi: 10.1097/01.mlg.0000173197.94769.ba

Hennesy, T., Cardon, G., Campbell, J., Glick, H., Bell-Souder, D., and Sharma, A. (2022). Cross-modal reorganization from both visual and somatosensory modalities in cochlear implanted children and its relationship to speech perception. *Otol. Neurotol.* 43, e872–e879. doi: 10.1097/MAO.0000000000003619

Holt, R. F., Kirk, K. I., and Hay-McCutcheon, M. (2011). Assessing multimodal spoken word-in-sentence recognition in children with normal hearing and children with cochlear implants. *J. Speech Lang. Hear. Res.* 54, 632–657. doi: 10.1044/1092-4388(2010/09-0148)

Huyse, A., Berthommier, F., and Leybaert, J. (2013). Degradation of labial information modifies audiovisual speech perception in cochlear-implanted children. *Ear Hear.* 34, 110–121. doi: 10.1097/AUD.0b013e3182670993

Iler Kirk, K., Hay-McCutcheon, M. J., Frush Holt, R., Gao, S., Qi, R., and Gerlain, B. L. (2007). Audiovisual spoken word recognition by children with cochlear implants. *Audiol. Med.* 5, 250–261. doi: 10.1080/16513860701673892

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jahn, K. N., Stevenson, R. A., and Wallace, M. T. (2017). Visual temporal acuity is related to auditory speech perception abilities in cochlear implant users. *Ear Hear.* 38:236. doi: 10.1097/AUD.0000000000000379

Jerger, S., Damian, M., Karl, C., and Abdi, H. (2020). Detection and attention for auditory, visual, and audiovisual speech in children with hearing loss. *Ear Hear.* 41, 508–520. doi: 10.1097/AUD.0000000000000798

Kaiser, A. R., Kirk, K. I., Lachs, L., and Pisoni, D. B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *J. Speech Lang. Hear. Res.* 46, 390–404. doi: 10.1044/1092-4388(2003/032)

Kanekama, Y., Hisanaga, S., Sekiyama, K., Kodama, N., Samejima, Y., Yamada, T., et al. (2010). "Long-term cochlear implant users have resistance to noise, but short-term users don't," in *Auditory-Visual Speech Processing 2010* (Makuhari: International Speech Communication Association (ISCA)), 1–4.

Kishon-Rabin, L., Haras, N., and Bergman, M. (1997). Multisensory speech perception of young children with profound hearing loss. *J Speech Lang. Hear. Res.* 40, 1135–1150. doi: 10.1044/jslhr.4005.1135

Kral, A., Dorman, M. F., and Wilson, B. S. (2019). Neuronal development of hearing and language: cochlear implants and critical periods. *Annu. Rev. Neurosci.* 42, 47–65. doi: 10.1146/annurev-neuro-080317-061513

Kral, A., and Sharma, A. (2023). Crossmodal plasticity in hearing loss. *Trends Neurosci.* 46, 377–393. doi: 10.1016/j.tins.2023.02.004

Lalonde, K., and McCreery, R. W. (2020). Audiovisual enhancement of speech perception in noise in school-age children who are hard-of-hearing. *Ear Hear.* 41:705. doi: 10.1097/AUD.0000000000000830

Lamoré, P. J., Huiskamp, T. M., van Son, N. J., Bosnian, A. J., and Smoorenburg, G. F. (1998). Auditory, visual and audiovisual perception of segmental speech features by severely hearing-impaired children. *Audiology* 37, 396–419. doi: 10.3109/00206099809072992

Lasfargues-Delannoy, A., Strelnikov, K., Deguine, O., Marx, M., and Barone, P. (2021). Supra-normal skills in processing of visuo-auditory prosodic information by cochlear-implanted deaf patients. *Hear. Res.* 410:108330. doi: 10.1016/j.heares.2021.108330

Lee, H. J., Lee, J. M., Choi, J. Y., and Jung, J. (2021). The effects of preoperative audiovisual speech perception on the audiologic outcomes of cochlear implantation in patients with postlingual deafness. *Audiol. Neurotol.* 26, 149–156. doi: 10.1159/000509969

Lei, J., Fang, J., Wang, W., and Mei, Y. (2008). The visual-auditory effect on hearing-handicapped students in lip-reading chinese phonetic identification. *Psychol. Sci.* 31, 312–314.

Lewis, D. E., Smith, N. A., Spalding, J. L., and Valente, D. L. (2018). Looking behavior and audiovisual speech understanding in children with normal hearing and children with mild bilateral or unilateral hearing loss. *Ear Hear.* 39:783. doi: 10.1097/AUD.0000000000000534

Leybaert, J., and LaSasso, C. J. (2010). Cued speech for enhancing speech perception and first language development of children with cochlear implants. *Trends Amplif.* 14, 96–112. doi: 10.1177/1084713810375567

Leybaert, J., and LaSasso, C. J. (2010). Cued speech for enhancing speech perception and first language development of children with cochlear implants. *Trends. Amplif.* 14, 96–112. doi: 10.1177/1084713810375567

Liang, M., Chen, Y., Zhao, F., Zhang, J., Liu, J., Zhang, X., et al. (2017). Visual processing recruits the auditory cortices in prelingually deaf children and influences cochlear implant outcomes. *Otol. Neurotol.* 38, 1104–1111. doi: 10.1097/MAO.0000000000001494

Lomber, S. G., Meredith, M. A., and Kral, A. (2010). Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. *Nat. Neurosci.* 13, 1421–1427. doi: 10.1038/nn.2653

Mantokoudis, G., Dähler, C., Dubach, P., Kompis, M., Caversaccio, M. D., and Senn, P. (2013). Internet video telephony allows speech reading by deaf individuals and improves speech perception by cochlear implant users. *PLoS ONE* 8:e54770. doi: 10.1371/journal.pone.0054770

Mitchel, A. D., Lusk, L. G., Wellington, I., and Mook, A. T. (2023). Segmenting speech by mouth: The role of oral prosodic cues for visual speech segmentation. *Lang. Speech* 66, 819–832. doi: 10.1177/00238309221137607

Moberly, A. C., Vasil, K. J., and Ray, C. (2020). Visual reliance during speech recognition in cochlear implant users and candidates. *J. Am. Acad. Audiol.* 31, 030–039. doi: 10.3766/jaaa.18049

Moody-Antonio, S., Takayanagi, S., Masuda, A., Auer Jr, E. T., Fisher, L., and Bernstein, L. E. (2005). Improved speech perception in adult congenitally deafened cochlear implant recipients. *Otol. Neurotol.* 26, 649–654. doi: 10.1097/01.mao.0000178124.13118.76

Moradi, S., Lidestam, B., Danielsson, H., Ng, E. H. N., and Rönnberg, J. (2017). Visual cues contribute differentially to audiovisual perception of consonants and vowels in improving recognition and reducing cognitive demands in listeners with hearing impairment using hearing aids. *J. Speech, Lang. Hear. Res.* 60, 2687–2703. doi: 10.1044/2016_JSLHR-H-16-0160

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., and Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychol. Sci.* 15, 133–137. doi: 10.1111/j.0963-7214.2004.01502010.x

Oryadi-Zanjani, M. M., Vahab, M., Bazrafkan, M., and Haghjoo, A. (2015). Audiovisual spoken word recognition as a clinical criterion for sensory aids efficiency in persian-language children with hearing loss. *Int. J. Pediatr. Otorhinolaryngol.* 79, 2424–2427. doi: 10.1016/j.ijporl.2015.11.004

Oryadi-Zanjani, M. M., Vahab, M., Rahimi, Z., and Mayahi, A. (2017). Audiovisual sentence repetition as a clinical criterion for auditory development in persian-language children with hearing loss. *Int. J. Pediatr. Otorhinolaryngol.* 93, 167–171. doi: 10.1016/j.ijporl.2016.12.009

Peelle, J. E., and Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex* 68, 169–181. doi: 10.1016/j.cortex.2015.03.006

Pralus, A., Hermann, R., Cholvy, F., Aguera, P.-E., Moulin, A., Barone, P., et al. (2021). Rapid assessment of non-verbal auditory perception in normal-hearing participants and cochlear implant users. *J. Clin. Med.* 10:2093. doi: 10.3390/jcm10102093

Puschmann, S., Daeglau, M., Stropahl, M., Mirkovic, B., Rosemann, S., Thiel, C. M., et al. (2019). Hearing-impaired listeners show increased audiovisual benefit when listening to speech in noise. *Neuroimage* 196, 261–268. doi: 10.1016/j.neuroimage.2019.04.017

Regenbogen, C., Schneider, D. A., Gur, R. E., Schneider, F., Habel, U., and Kellermann, T. (2012). Multimodal human communication–targeting facial expressions, speech content and prosody. *Neuroimage* 60, 2346–2356. doi: 10.1016/j.neuroimage.2012.02.043

Rigo, T. G., and Lieberman, D. A. (1989). Nonverbal sensitivity of normal-hearing and hearing-impaired older adults. *Ear Hear.* 10, 184–189. doi: 10.1097/00003446-198906000-00008

Rouger, J., Fraysse, B., Deguine, O., and Barone, P. (2008). Mcgurk effects in cochlear-implanted deaf subjects. *Brain Res.* 1188, 87–99. doi: 10.1016/j.brainres.2007.10.049

Rouger, J., Lagleyre, S., Démonet, J.-F., Fraysse, B., Deguine, O., and Barone, P. (2012). Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients. *Hum. Brain Mapp.* 33, 1929–1940. doi: 10.1002/hbm.21331

Sato, T., Yabushita, T., Sakamoto, S., Katori, Y., and Kawase, T. (2020). In-home auditory training using audiovisual stimuli on a tablet computer: Feasibility and preliminary results. *Auris Nasus Larynx* 47, 348–352. doi: 10.1016/j.anl.2019.09.006

Schumann, A., and Ross, B. (2022). Adaptive syllable training improves phoneme identification in older listeners with and without hearing loss. *Audiol. Res.* 12, 653–673. doi: 10.3390/audiolres12060063

Sharma, A., Campbell, J., and Cardon, G. (2015). Developmental and cross-modal plasticity in deafness: evidence from the P1 and N1 event related potentials in cochlear implanted children. *Int. J. Psychophysiol.* 95, 135–144. doi: 10.1016/j.ijpsycho.2014.04.007

Stevens, C., and Neville, H. (2006). Neuroplasticity as a double-edged sword: deaf enhancements and dyslexic deficits in motion processing. *J. Cogn. Neurosci.* 18, 701–714. doi: 10.1162/jocn.2006.18.5.701

Stevenson, R. A., Sheffield, S. W., Butera, I. M., Gifford, R. H., and Wallace, M. T. (2017). Multisensory integration in cochlear implant recipients. *Ear Hear.* 38, 521–538. doi: 10.1097/AUD.0000000000000435

Stropahl, M., and Debener, S. (2017). Auditory cross-modal reorganization in cochlear implant users indicates audio-visual integration. *NeuroImage: Clini.* 16, 514–523. doi: 10.1016/j.nicl.2017.09.001

Suess, N., Hauswald, A., Zehentner, V., Depireux, J., Herzog, G., Rösch, S., et al. (2022). Influence of linguistic properties and hearing impairment on visual speech perception skills in the german language. *PLoS ONE* 17:e275585. doi: 10.1371/journal.pone.0275585

Taitelbaum-Swead, R., and Fostick, L. (2017). Audio-visual speech perception in noise: Implanted children and young adults versus normal hearing peers. *Int. J. Pediatr. Otorhinolaryngol.* 92, 146–150. doi: 10.1016/j.ijporl.2016.11.022

Tona, R., Naito, Y., Moroto, S., Yamamoto, R., Fujiwara, K., Yamazaki, H., et al. (2015). Audio-visual integration during speech perception in prelingually deafened japanese children revealed by the mcgurk effect. *Int. J. Pediatr. Otorhinolaryngol.* 79, 2072–2078. doi: 10.1016/j.ijporl.2015.09.016

Tsao, Y.-W. (2019). *Audiovisual Speech Perception in People With Hearing Loss Across Languages: A Systematic Review of English and Mandarin.* Tucson, AZ: The University of Arizona. Available at: http://hdl.handle.net/10150/633080

Tye-Murray, N., Sommers, M. S., and Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear Hear.* 28, 656–668. doi: 10.1097/AUD.0b013e31812f7185

Tye-Murray, N., Spehar, B., Sommers, M., and Barcroft, J. (2016). Auditory training with frequent communication partners. *J. Speech Lang. Hear. Res.* 59:871. doi: 10.1044/2016_JSLHR-H-15-0171

Tyler, R. S., Fryauf-Bertschy, H., Kelsay, D. M., Gantz, B. J., Woodworth, G. P., Parkinson, A., et al. (1997). Speech perception by prelingually deaf children using cochlear implants. *Otolaryngol.-Head Neck Surg.* 117, 180–187. doi: 10.1016/S0194-5998(97)70172-4

Van Soeren, D. (2023). The role of word recognition factors and lexical stress in the distribution of consonants in spanish, english and dutch. *J. Linguist.* 59, 149–177. doi: 10.1017/S0022226722000081

Yamamoto, R., Naito, Y., Tona, R., Moroto, S., Tamaya, R., Fujiwara, K., et al. (2017). Audio-visual speech perception in prelingually deafened japanese children following sequential bilateral cochlear implantation. *Int. J. Pediatr. Otorhinolaryngol.* 102, 160–168. doi: 10.1016/j.ijporl.2017.09.022

Yu, L., Rao, A., Zhang, Y., Burton, P. C., Rishiq, D., and Abrams, H. (2017). Neuromodulatory effects of auditory training and hearing aid use on audiovisual speech perception in elderly individuals. *Front. Aging Neurosci.* 9:30. doi: 10.3389/fnagi.2017.00030

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., and Vigliocco, G. (2021). More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proc. Royal Soc. B.* 288:20210500. doi: 10.1098/rspb.2021.0500