Auditory and visual information in speech perception: A developmental perspective

**Key words:** speech perception, children, older adults, audiovisual integration, audiovisual perception, developmental speech perception
Abstract

This study investigates the development of audiovisual speech perception from age 4 to 80, analyzing the contribution of modality, context, and special features of specific language being tested. Data of 77 participants in five age groups is presented in the study. Speech stimuli were introduced via auditory, visual, and audiovisual modalities. Monosyllabic meaningful and nonsense words were included in a signal-to-noise ratio of 0dB. Speech perception accuracy in audiovisual and auditory modalities by age resulted in an inverse U shape, with lowest performance at ages 4-5 and 65-80. In the visual modality, a clear difference was shown between performance of children (ages 4-5 and 8-9) and adults (age 20 and above). The findings of the current study have important implications for strategic planning in rehabilitation programs for child and adult speakers of different languages with hearing difficulties.
**Introduction**

Speech perception changes across the life span. It develops and matures from infancy through adulthood (Burnham, Earnshaw, & Clark, 1991). This development may be attributed to improvements in psychoacoustic abilities (Schneider, Trehub, Morrongiello, & Thorpe, 1986; Trehub, Schneider, Morrongiello, & Thorpe, 1988), amount of exposure to language (Hazan & Barrett, 2000), and improvements in speech production abilities as a result of the close relationship between perception and production (Kishon-Rabin, Taitelbaum-Sweed, & Segal, 2009). Later in life, speech perception deteriorates. Changes in speech perception in older adults may be attributed to decrease in peripheral auditory thresholds, decline in functioning of central systems, and impaired cognitive abilities (Helfer & Freyman, 2008; Rajan & Cainer, 2008).

**Audiovisual integration across life**

Although the auditory modality is the primary modality for speech perception, the visual modality also helps in infant language development (Bergeson, Houston, & Miyamoto, 2010) and enhances speech perception accuracy for children and adults (Chen & Hazan, 2009; Cienkowski & Carney, 2002). Visual articulators—primarily the lips, teeth, and tongue—help convey important aspects of speech, including information about the place of articulation of consonants and vowels. Adding lip reading to audition enables the normal hearing listener to perceive speech more accurately, even when the acoustic information is undistorted (Summerfield, 1992). For example, visually impaired infants with normal hearing are often delayed in acquiring phonetic distinctions that are easy to see but difficult to hear (such as /b/ and /d/).
In conditions of noise or hearing impairment, the visual modality plays a complementary role. It enables normal hearing individuals to understand speech in conditions with a poor signal to noise ratio, compared to the auditory modality alone. Accordingly, normal hearing listeners use visual speech to improve speech perception when its quality is degraded by adverse listening conditions such as background noise (MacLeod & Summerfield, 1987; Sumby & Pollack, 1954). Indeed, behavioral studies have shown that infants, from a very early age, link what they hear and what they see. Some researchers found that very young infants are aware of the association between lip movements and speech sounds. Soon after birth, they already decode language across modalities (Dodd, 1979; Kuhl & Meltzoff, 1982). This ability develops and matures during childhood (Wightman, Kistler, & Brungart, 2006; Massaro, Thompson, Barron, & Laren, 1986). While studies have verifiably demonstrated the changes in audiovisual integration in children and adults over time, data is still lacking regarding the development of audiovisual integration across the life span. Therefore, it is important to test the developmental course of audio visual speech perception in different age groups.

**Audiovisual research paradigms**

The best known phenomenon for demonstrating multimodal speech integration is known as the McGurk Effect (McGurk & MacDonald, 1976). In this effect, individuals hear one syllable while seeing another one, which results in the perception of a third stimulus. The classic demonstration of this effect utilizes place of articulation, whereby the individuals see /ba/, hear /ga/, and perceive /da/. Researchers showed this effect in infants as young as four or five months (Burnham & Dodd, 2004; Rosenblum, Schmuckler, & Johnson, 1997). Studies that tested the
McGurk Effect in children found positive association between age and the amount of influence of visual speech on audiovisual speech perception. At a young age, children do not benefit from visual information; indeed, the contribution of this type of information increases with age (Chen & Hazan, 2009; Desjardins, Rogers, & Werker, 1997; Sekiyama & Burnham, 2008). Most of the studies on the McGurk Effect found that the development in audiovisual perception occurs during early childhood, toward the first years of school (Chen & Hazan, 2009). Young adults aged 18-35 were found to perform similarly to older normal hearing adults aged 65-74 in audiovisual integration using the McGurk Paradigm (Cienkowski & Carney, 2002).

Other paradigms have also been used to test audiovisual integration (Ross et al., 2011; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Tye-Murray, Spehar, Myerson, Sommers, & Hale, 2011). Calculating the contribution of each modality to multisensory integration enables the evaluation of each modality's contribution to speech perception. Ross et al. (2011) tested developmental course in audiovisual integration in children aged 5 to 14 years. They found that younger children benefited substantially less from the addition of visual speech than did older children. In comparing audiovisual integration of young adults to that of older adults, young adults demonstrated a higher accuracy rate in perceiving speech in audiovisual and visual conditions, as compared to older adults, whereas no difference was found when the specific contribution of auditory and visual information was calculated (Sommers, Tye-Murray, & Spehar, 2005; Tye-Murray, Sommers, & Spehar, 2007; Tye-Murray, Sommers, Spehar, Myerson, & Hale, 2010; Tye-Murray et al., 2011). As can be seen, using different paradigms while focusing on auditory, visual, and audiovisual perception provides significant data; clearly, however, further studies are needed.
**Context effects on audiovisual integration**

Another factor that may affect audiovisual speech perception (besides age) is the context in which perception takes place. The perception of spoken signals improves when introduced with increased contextual information (phonological, lexical, and grammatical). For instance, listeners better utilize visual information that is presented along with a meaningful utterance than they do for a nonsensical one. In addition, when the recognition of semantically normal and anomalous sentences accompanied by four types of auditory masking, was studied, the semantic context was beneficial across all types of maskers, both via auditory and audiovisual modalities (Van Engen, Phelps, Smiljanic, & Chandrasekaran, 2014). Studies on auditory speech perception show that the use of context increases gradually with age (Nittroer & Boothroyd, 1990). However, it would seem that the developmental effect of context usage on audiovisual speech perception has not been studied enough.

**Audiovisual integration in different languages**

When testing speech perception, the listener's native language should also be taken into consideration. Studies have examined whether audiovisual integration is language-specific or universal. Earlier studies testing the McGurk effect with synthetic stimuli showed similar processing for native speakers of English, Spanish, Japanese, and Dutch (Massaro, Tsuzaki, Cohen, & Heredia, 1993; Massaro, Cohen, & Smeele, 1995), although differences derived from the characteristics of the language were observed. These differences can be attributed to different phonological structures among languages or different cultural habits. Indeed, a weaker McGurk effect was found in Japanese and Chinese than English. This difference was explained
as being culturally influenced, as Japanese and Chinese individuals tend to avoid eye contact and therefore miss some visual information (Sekiyama, & Tohkura, 1993; Sekiyama, 1997).

In terms of developmental benchmarks, previous studies suggest that both English and Japanese children are less influenced by visual modality when under the age of six than those who have reached age six (Sekiyama & Burnham, 2008). Between the ages of six to eight years, an increase in the effect of the visual modality has been noted only among English-speaking children. Researchers have explained these differences as relating to acoustic-phonetic issues. Since Japanese has less vowels and phonological contrasts, the acoustic-phonetic information can more easily be retrieved from the auditory channel than in English, enabling better auditory perception in Japanese listeners. This is also the case with Hebrew. Indeed, speech reading performance of phonological contrasts in Hebrew was found to develop even after age 12 (Kishon-Rabin & Henkin, 2000), while English speaking children were found to achieve adult-level speech reading performance of these contrasts already at the age of seven (Hnath-Chisolm, Laipply, & Boothroyd, 1998). To date, no study has tested the development of audiovisual integration among Hebrew speakers. Therefore, the effect of the acoustic-phonetic characteristics of Hebrew on audiovisual integration is still not clear. We anticipate that visual performance in Hebrew will bear greater resemblance to Japanese, with later maturation occurring, than English.

The current study

The aim of the current study was to test audiovisual speech perception across the life span. The existing data in this field is mainly based on comparisons of
younger and older children, or younger and older adults. In this study, we aimed to draw a broad developmental picture by using five age groups, with participants aged 4 to 80 years. Previous studies involving adult participants have used different test material (such as nonsense words, meaningful words, and sentences), however, when a wide age range is involved (as in our study), age-related differences in experience and linguistic redundancy might interfere with audiovisual integration. Therefore, in the current study we chose to use monosyllabic meaningful and nonsense words, which serve as stimuli that mainly include acoustical information and entail minimal linguistic redundancy.

The literature regarding audiovisual speech perception across the life span reveals various currently unaddressed issues, including the developmental course, the specific contribution of auditory and visual information, the effect of context, and language characteristics. Therefore, in the present study we aimed to test audiovisual speech perception across the life span, while (1) employing a wide range of age groups; (2) using a research paradigm that enables calculation of each modality's contribution to audiovisual integration; (3) studying the developmental effect of context on audiovisual integration; and (4) examining audiovisual speech perception in Hebrew, a language with comparatively limited number of vowels and relatively late development of speech reading, in order to further understand its specific/universal nature.
Method

Participants

Seventy-seven participants were included in the study. They were divided into five groups based on age: age four to five (n=15); age eight to nine (n=17); age 20-30 (n=15); age 40-55 (n=15); and age 65-80 (n=15). All participants met the inclusion criteria of having: (1) Hebrew as a native language; (2) good hearing thresholds (all children and the 20-30 year old group had pure-tone air-conduction thresholds less than 15 dB HL bilaterally at octave frequencies from 250 - 4,000 Hz while all adults (40-55 and 65-80 years old) were screened for age-normal hearing at 500, 1,000, 2,000, and 4,000 Hz) according to Lebo and Redell (1972); (3) normal or corrected visual ability; (4) normal speech and language abilities (based on parental report) with no articulation problems (based on articulation tests) for the children; and (5) no reported cognitive or neurological problems.

Speech perception tests

In order to avoid age-related linguistic experience effects, we used monosyllabic meaningful and nonsense words, which include mainly acoustical information and minimal linguistic redundancy.

Meaningful Words

Monosyllabic meaningful Hebrew AB lists (based on Boothroyd, 1968) were used in the present study. This test included 12 lists (narrated in a film, as described in the sub-section on Apparatus below), each consisting of ten monosyllabic words. Each list contained ten syllables in a consonant-vowel-consonant (CVC) pattern in which the 19 consonants in the Hebrew language appeared either at the initial or the final position, and each of the five Hebrew vowels (/a/, /e/, /i/, /o/, /u/) appeared twice.
Nonsense Words

This test was used in order to calculate auditory and visual enhancement as well as the contribution of lexical context provided by meaningful vs. nonsense words (k factor, Boothroyd & Nittrouer, 1988; Nittrouer & Boothroyd, 1990). It resembles the structure of the meaningful words test and also includes 12 lists of ten monosyllabic CVC syllables. In this test, the syllables were nonsensical for Hebrew speakers but contained some phonological redundancy, in accord with Hebrew linguistic rules (for example, the consonants /b/ and /p/ never appear in the final position).

Apparatus

A female native Hebrew speaker with intelligible articulation and clear facial movements was filmed and recorded. The speaker looked directly into the camera, starting and ending each utterance with a natural face/closed mouth position. The speaker was recorded against a bright background in a quiet, well-lit recording studio. Her face appeared in full on the entire screen. The audiovisual recordings were digitized using Apple Final Cut Pro X software with 64-bit resolution.

The words were recorded in a studio using a SONTRONICS TCS-6 microphone and Samplitude classic 8.1 recording software. They were edited using the Sound Forge program, which digitized (16-bit) at a sampling rate of 44kHz. Word level was normalized using the overall RMS. White noise generated by the Sound Forge program was added to the normalized words in an signal-to-noise ratio (SNR) of 0. The noise was added to words in the audiovisual and auditory conditions, while in the visual condition the words were presented in quiet.
The words were presented using the Winamp Media Player 5.7 software, via S-Tech supra aural headphones in 70 dBSPL, as measured by a TA 1350A Sound Level Meter. Participant responses were recorded using a SONY ICD-PX312 recording device placed in close proximity to the participants. An inter-acoustic AD229B audiometer was used to screen and measure hearing levels.

Procedure

The study was approved by the Institution ethics committee and was conducted in accordance with Good Clinical Practice (GCP) guidelines. All participants received a full explanation about the study and signed an informed consent. For participants under age 18, parental informed consent was obtained and assent to participate was obtained from the children. All potential candidates were screened for hearing levels prior to participation in the study.

The study included six different conditions: two context conditions (meaningful and nonsense words) and three modality conditions (Auditory, Visual, or Audiovisual). Two lists (20 words) were presented for each condition. The order of the words within each list as well as the order of words and nonsense words were randomly intermixed across participants. Each participant was tested under all six conditions.

Participants were required to repeat each word immediately after hearing it or after viewing the video, and their responses were recorded. Two independent raters
transcribed these responses. When the raters disagreed (this occurred in less than 1% of the words), a transcription of a third rater was used.

The study was not supported by any sponsor or funding agency.
Results

1. Accuracy rate in perceiving meaningful words: Modality and age group

Table 1 presents the mean and SD for the accuracy of perceiving meaningful words in each modality and age. A two-way repeated measures analysis of variance (ANOVA) was conducted, with modality (auditory, visual, and audiovisual) as a within-subjects variable, and age group (4-5, 8-9, 20-30, 40-55, and 65+) as a between-subjects variable. This analysis revealed significant main effects for modality ($F_{(2,144)}=1126.614$, $p=.000$, Partial Eta Squared=.940) and age group ($F_{(4,72)}=10.797$, $p=.000$, Partial Eta Squared=.375). Main effects were followed by pairwise comparisons using the Least Significant Difference test. As can be seen in Table 1, a higher accuracy rate was found in the audiovisual modality (mean=82.969, SE=1.251) than in the auditory (mean=69.612, SE=1.662, LSD=13.357, $p=.000$), and the lowest accuracy rate was found in the visual modality (mean=11.059, SE=.893, LSD=71.910, $p=.000$ for visual vs. audiovisual, LSD=-58.553, $p=.000$ for visual vs. auditory).

Overall, the results of participants aged 20-30 (mean=62.778, SE=2.078) do not differ from participants aged 40-55 (LSD=4.556, $p=.125$), but were higher than those of the other age groups (age 4-5: LSD=18.333, $p=.000$; age 8-9: LSD=8.268, $p=.005$; age 65+: LSD=10.000, $p=.001$). Results of participants aged 40-55 (mean=58.222, SE=2.078) were also similar to those of participants aged 8-9 (mean=54.510, SE=1.952, LSD=3.712, $p=.197$) and to those of the oldest (65+) age group (mean=52.778, SE=2.078, LSD=5.444, $p=.068$); yet they were higher than the results of 4-5 year olds (LSD=13.778, $p=.000$). Participants aged 4-5 (mean=44.444,
SE=2.078) had the lowest accuracy rate as compared with other age groups (age 8-9: LSD=10.065, p=.001; age 65+: LSD=8.333, p=.006).

A modality X age group interaction was also found ($F_{(8,144)}=2.157$, $p=.034$, Partial Eta Squared=.107). Differences between age groups had a similar pattern for both the audiovisual and auditory inputs presented in noise, with an increase in accuracy rate from ages 4-5 to 20-30 years, and then a decrease among age 65+. The visual input presented in quiet, however, showed a different pattern, with a gradual increase in the accuracy rate of perceiving meaningful words demonstrated from ages 4-5 to 65+, with a small peak in accuracy at ages 20-30. The accuracy rate for each age group, as well as post-hoc analysis, is presented in Table 1.

![Insert table 1 around here](image)

2. **Auditory and visual enhancement to perceiving words: Type of word and age group**

In order to assess the enhancement of auditory ($AE$) and visual ($VE$) information for speech perception of meaningful and nonsense words, we followed the formula of Sommers et al. (2005). Accurate raw data for meaningful words is presented in Table 1 and for nonsense words in Table 2. The performance in each of these modalities was subtracted from the performance in the audiovisual ($AV$) condition, as follows:

\[(a) \quad AE = AV-V/I-V\]

\[(b) \quad VE = AV-A/I-A\]
A two-way repeated measures ANOVA was calculated for auditory and visual enhancement, with type of word (meaningful, nonsense) as a within-subjects variable and age group (4-5, 8-9, 20-30, 40-55, and 65+) as a between-subjects variable.

Figure 1a presents auditory enhancement by type of word and age group. Main effects for type of word ($F_{(1,72)}=21.567$, $p=.000$, Partial Eta Squared=.230) and age group ($F_{(4,72)}=4.239$, $p=.004$, Partial Eta Squared=.191) were found for auditory enhancement, but no interaction emerged ($F_{(4,72)}=1.95$, $p=.940$, Partial Eta Squared=.011). Auditory enhancement for meaningful words (mean=.802, SE=.018) was higher than for nonsense words (mean=.684, SE=.021). Figure 1a shows that the pattern of age group differences in overall auditory enhancement was similar to that of perceiving words in the auditory modality, with increase in auditory enhancement apparent between the ages of 4-5 and 20-30, followed by a decrease toward age 65+.

Similar auditory enhancement was found for age groups 4-5 (mean=.656, SE=.034) and 65+ (mean=.686, SE=.034, LSD=.030, $p=.541$), and for age groups 8-9 (mean=.792, SE=.032), 20-30 (mean=.818, SE=.034, LSD=.026, $p=.584$), and 40-55 (mean=.762, SE=.034, LSD=.030, LSD=.056, p=.248, respectively), with the former age groups demonstrating lower enhancement than the latter (4-5 vs. 8-9: LSD=.136, $p=.005$; 4-5 vs. 20-30: LSD=.162, $p=.001$; 4-5 vs. 40-55: LSD=.106, $p=.031$; 65+ vs. 8-9: LSD=.107, $p=.025$; 65+ vs. 20-30: LSD=.133, $p=.007$; 65+ vs. 40-55: LSD=.176, $p=.016$).

Figure 1b presents visual enhancement by type of word and age group. Main effects for type of word ($F_{(1,72)}=4.551$, $p=.036$, Partial Eta Squared=.059) and age
group \((F(4,72)=4.717, p=.002, \text{Partial Eta Squared}=.208)\) were found for visual enhancement. As was shown with the auditory enhancement, visual enhancement for meaningful words (mean=.506, \(SE=.031\)) was higher than for nonsense words (mean=.422, \(SE=.024\)). Post-hoc analysis using the Least Significant Difference test for age groups revealed that participants aged 4-5 demonstrated overall lower visual enhancement (mean=.320, \(SE=.044\)) than all other age groups (8-9: mean=.467, \(SE=.041\), LSD=-.148, \(p=.016\); 20-30: mean=.495, \(SE=.044\), LSD=-.175, \(p=.006\); 40-55: mean=.582, \(SE=.044\), LSD=-.263, \(p=.000\); 65+: mean=.457, \(SE=.044\), LSD=-.137, \(p=.029\)), which were found to be similar to each other.

A type of word \(\times\) age group interaction \((F(4,72)=4.247, p=.014, \text{Partial Eta Squared}=.910)\) revealed that in age group 4-5, no difference exists between meaningful and nonsense words in terms of visual contribution (LSD=.103, \(p=.920\)), although significant differences between these types of words was found for all other age groups (8-9: LSD=2.191, \(p=.025\); 20-30: LSD=2.077, \(p=.030\); 40-55: LSD=2.611, \(p=.012\); 65+: LSD=2.179, \(p=.025\), see figure 1b).

---

3. Context effects in perceiving words: Modality and age group

We tested the effects of context on perceiving words in different modalities and age groups using Boothroyd and Nittrouer’s (1988) and Nittrouer and Boothroyd’s (1990) formulas for calculating the contribution of lexical context provided by meaningful vs. nonsense words \((k\text{-factor})\), and the amount of phonemes needed for word recognition \((j\text{-factor})\). These contributions were calculated using the following formulas:
$k = \log \left( \frac{1-p_c}{1-p_i} \right)$

where $p_c$ is the probability of recognition in context and $p_i$ is the probability of recognition absent any context (Boothroyd & Nittrouer, 1988; Nittrouer & Boothroyd, 1990). The contribution of lexical context was measured by comparing recognition of real word ($p_c$) versus nonsense syllables ($p_i$) as follows:

$$j = \log \frac{p_w}{p_p}$$

where $p_w$ is the probability of recognition of the entire word and $p_p$ is the probability of recognition of each of its parts (Boothroyd & Nittrouer, 1988; Nittrouer & Boothroyd, 1990).

Due to low accuracy rate for visual modality, which can cause bias in calculating the $j$-factor (Boothroyd & Nittrouer, 1988), this rate was calculated only for auditory and audiovisual modalities.

Two separate repeated measures ANOVAs were carried out on $k$- and $j$-factors, with modality as a within-subjects variable and age group as a between-subjects variable. For the $k$-factor, a significant main effect was found only for age group ($F(4,65)=2.854$, $p=.043$, Partial Eta Squared=.127), with higher $k$-factor for participants aged 65+ (mean=1.388, SE=.059) than for those of the other age groups (4-5: mean=1.157, SE=.059, LSD=.232, $p=.007$; 8-9: mean=1.227, SE=.059, LSD=.215, $p=.048$; 20-30: mean=1.173, SE=.069, LSD=.215, $p=.021$; 40-55: mean= 1.272, SE=.061, LSD=.217, $p=.017$). No effect was found for modality ($F(2,130)=1.584$, $p=.209$, Partial Eta Squared=.024), and no modality X age group interaction was found ($F(8,130)=.807$, $p=.597$, Partial Eta Squared=.047).

We found a larger $j$-factor for auditory (mean=2.909, SE=.071) than for audiovisual modality (mean=2.716, SE=.080, $F(1,58)=4.155$, $p=.046$, Partial Eta
Squared=.067). No main effect was found for age group ($F_{(4,58)}=.552$, $p=.698$, Partial Eta Squared=.037), and no modality X age group interaction emerged ($F_{(4,58)}=.510$, $p=.728$, Partial Eta Squared=.034).
**Discussion**

The current study aimed to detect patterns of speech perception performance by participants in different age groups, using different modalities—auditory and visual—and the contribution of context. The scientific literature shows that audiovisual perception is superior to each modality separately, especially with regard to the visual modality (Ross et al., 2011; Sommers et al., 2005). Differences between modalities were also found in the current study, however with a unique pattern for each modality across age groups. Results for the visual input presented in quiet showed positive association with age, with similar performance in age groups 4-5 and 8-9, lower than that of age 20-30 and above. In contrast, the results for the test of auditory input presented in noise showed an inverse U pattern, with lowest and comparable performance of age groups 4-5 and 65+ than participants aged 8-55 years.

Audiovisual integration results were similar to the auditory modality in its inverse U-shape pattern, with similar and lower performance in age groups 4-5 and 65+, as compared to age groups 8-9, 20-30, and 40-55. This finding reflects mature audiovisual performance already at age 8-9 in terms of both accuracy and enhancement. Most of the previous studies that used the McGurk Effect show similar results of audiovisual integration maturation around ages 6-8 years (Massaro et al., 1986), however, some studies show later audiovisual maturation, around age 14 (Brandwein et al., 2011; Ross et al., 2011). These latter studies applied different paradigms and stimuli, such as words in negative SNRs or tones. Findings of a later maturation age in these studies might result from more difficult conditions for young children than the ones used in the present study. A close look at the data from Ross et al. (2011) shows that under better SNR conditions (no noise and SNR of -3), comparable performance of age groups 5-7 and adults was found. In the present study,
we chose to use SNR 0 in order to be able to present the same stimuli to different age groups, avoiding both floor and ceiling effects in all modalities; different SNR levels might not be appropriate for such large age range.

It should be noted that the similarity in auditory performance found in the 4-5 and 65+ age groups reflects different mechanisms. At age 4-5, the auditory system has not yet matured (Boothroyd, 1997; Paus et al., 1999) due to underdeveloped psychoacoustic abilities (such as frequency discrimination and temporal processing). In addition, production abilities are not yet fully developed and are in reciprocal relationships with perception (Kishon-Rabin et al., 2002; Kishon-Rabin et al., 2009). These factors lead to difficulty in perceiving speech in noise (presented in SNR 0). The 65+ age group, in contrast, probably suffers from deterioration of the auditory peripheral system (although they were screened for age-normal hearing). This is in addition to some decrease in central and cognitive system functions (Anderson, White-Schwoch, Parbery-Clark, & Kraus, 2013; Fostick, Ben-Artzi, & Babkoff, 2013) that make it difficult to perceive speech in noise even at SNR 0. These results imply that the auditory modality, as measured in the current paradigm, matures earlier than the visual one, but deteriorates faster in older age. This finding supports other evidence in the literature that the visual modality develops slower and later in life than the auditory one (Desjardins et al., 1997; Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000; Sekiyama & Burnham, 2008). Future research will need to confirm whether or not this pattern of results within the visual modality is sustained when visual stimuli are distorted.

Both auditory and visual enhancement for meaningful words was higher than for nonsense words. This was expected, since for meaningful words the participants’ lexicon is also available as an informative source for speech perception, in addition to
auditory and visual information. Nonsense words, on the other hand, lack the additional information of the lexicon, and judgments in this condition are based only on perceiving the auditory signal and the visemes (visual phonemes). This also explains higher accuracy rate for meaningful words compared to nonsense words and shows the importance of linguistic redundancy in speech perception, and its usage via both auditory and visual modalities. This finding was reported previously (Boothroyd & Nittrouer, 1988; Nittrouer & Boothroyd, 1990; Van Engen et al., 2014); however, to the best of our knowledge, this is the first study to test developmental visual and auditory enhancement. Auditory enhancement demonstrated a similar pattern by age as the accuracy data. However, visual enhancement exhibited a different pattern. Visual enhancement was similar across the different age groups, with the exception of being lower in the 4-5 age group. For this group, moreover, no difference was found in the visual enhancement between meaningful and nonsense words. This lack of difference in visual enhancement between meaningful and nonsense words, along with the low performance of the younger group might suggest that visual enhancement, i.e., using visual information when auditory information is also present, requires more experience and sophistication than merely using the auditory information, which might be more intuitive for speech perception. At the young age of 4-5 years, the linguistic experience is rather scarce; therefore, the usage of visual information is low, and meaningful words demonstrate no apparent advantage.

Speech perception depends on prior knowledge and expectations, as well as on the ability to use it correctly. Using context while perceiving speech takes place across modalities. In the current study, we tested the amount of context used in each modality, and developmentally, by age. The effect of context on audiovisual integration among different age groups was tested using Boothroyd and Nittrouer’s
(1988) formulas to calculate context use in the perception of nonsense vs. meaningful words (k-factor) and the amount of word fragments (phonemes) required to perceive whole words (j-factor) (see also Most & Adi-Bensaid, 2001). K-factor was found to be higher for the age 65+ group in all modalities. Previous studies report that older adults indeed use more context than younger ones, as a result of their greater experience and as a compensation for lack of auditory acuity (Dubno, Ahlstrom, & Horwitz, 2000; Pichora-Fuller, Schneider, & Daneman, 1995). However, no difference was found between modalities in the amount to which lexical redundancy contributes to speech perception (k-factor). This finding might be due to the use of monosyllabic words, which do not carry much redundancy. J-factor, on the other hand, was found to be lower in the audiovisual modality. As can be expected, fewer phonemes are needed for identifying words when information from both auditory and visual modalities is available. Although the auditory channel provides most of the information needed for speech perception, the visual channel complements it and contributes information about phonological contrasts that are masked in the auditory channel. For example, acoustic cues for perceiving place of articulation are ambiguous when noise is present (Nittrouer, 2005). These cues are available visually via lip reading (Hnath-Chisolm et al., 1998; Kishon-Rabin & Henkin, 2000).

No age effect was found for j-factor, which implies that the use of contextual information in the word level is available already from age 4-5, as was reported previously (Nittrouer & Boothroyd, 1990). The results regarding j-factor were available only for audiovisual and auditory modalities; this is due to the low accuracy rate in the visual channel, especially for nonsense words, that disabled the j-factor calculation for this condition (Boothroyd & Nittrouer, 1988). We can assume that
more phonemes are needed for speech perception in the visual channel than in the auditory one; it would be interesting to investigate this issue in future studies.

Audiovisual integration also seems to differ between languages (Sekiyama, 1997; Sekiyama & Burnham, 2008). The data from our findings show that audiovisual integration, using the current paradigm in the Hebrew language, matures before eight years of age. In line with our hypothesis, these results are different from the findings regarding English speakers, for whom such maturation takes place at a later age (Ross et al., 2011); this may be attributed to aspects unique to the Hebrew language, such as limited phonemic inventory of vowels and consonants. Indeed, Hebrew has only five vowels and 19 consonants, as compared to 12-14 vowels and 24 consonants in English (Boothroyd, 1986). This difference enables better distinction between vowels, and less phonological contrasts, in Hebrew. Moreover, acoustic differences also exist between these languages, such as differences in categorical perception of voicing plosives and in the relative weighting of some acoustic cues (Kishon-Rabin, Rotshtein, & Taitelbaum, 2002; Taitelbaum-Swead, Hildesheimer, & Kishon-Rabin, 2003). In addition, there is a larger gap between voiced and voiceless plosives than in English, and the voiced ones carry more acoustic information (voicing leads) (Raphael, Faber, Most, Kollia & Milstein, 1995). Support for the fact that Hebrew speakers use the acoustic information in the voicing lead comes from studies of post-lingually deaf adults who were cochlear-implanted. These participants produce shorter voicing leads before the implantation, which are restored afterwards, resembling Hebrew norms for longer voicing leads (Kishon-Rabin, Taitelbaum, Tobin, & Hildesheimer, 1999). The availability of auditory information important for speech perception in Hebrew can explain the findings of younger maturation age amongst Hebrew speakers and the relatively low enhancement of visual information in the
Hebrew-speaking population. Indeed, similar findings emerged in studies regarding other languages that have a relatively limited amount of vowels and phonological contrasts, such as Japanese (Sekiyama & Burnham, 2008).

The major finding of this study is that, as expected, the auditory channel, which matures early in life (for most paradigms) is the most important for speech perception, especially when testing monosyllabic words in Hebrew. However, despite the current language and paradigm, the visual channel does seem important: when it is available (audiovisual condition), speech perception is significantly more accurate than in its absence (auditory condition), and less acoustic information is needed for identifying the presented words (a smaller j-factor). The visual channel also contributes to better speech perception in older age, when the auditory channel starts to deteriorate.

Clinically, the present study suggests that different developmental rates of auditory and visual modalities should be taken into account when rehabilitating children and adults with hearing loss. The results of this study suggest that older adult rehabilitation could be more efficient, with emphasis on visual strategies in speech perception. The contribution of visual information to speech perception also demonstrates the importance of also exposing young children with hearing aids or cochlear implants to audio-visual training, especially under adverse listening conditions, in order to improve their audiovisual integration. Some of the results in the current study are explained by the differences that exist between languages. These differences must be taken into account when adapting rehabilitation methods developed for one language and applying them to another.

Acknowledgements
The authors would like to thank Lital Raifen, Lilach Sabi, Inbal Shemer and Or Barel for collecting data for this study.

**Declaration of interest**

The authors have no declaration of interest to report.

**REFERENCES**


Evans, A. C. (1999). Structural maturation of neural pathways in children and

adults listen to and remember speech in noise. *The Journal of the Acoustical
Society of America, 97*, 593–608.

Rajan, R., & Cainer, K. E. (2008). Ageing without hearing loss or cognitive
impairment causes a decrease in speech intelligibility only in informational

Intermediate values of voice onset time. In F. Bell-Berti, & L. J. Raphael (Eds)
Producing Speech: Contemporary Issues for Katherine Safford Harris AIP, New
York.


Ross, L. A, Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe,
J. J. (2011). The development of multisensory speech perception continues into
the late childhood years. *The European Journal of Neuroscience, 33*, 2329–2337.

you see what I am saying? Exploring visual enhancement of speech
comprehension in noisy environments. *Cerebral Cortex, 17*, 1147–1153.


Figure 1a: Mean auditory enhancement for meaningful (solid line) and nonsense (dashed line) words in five age groups.
Figure 1b: Mean visual enhancement scores for meaningful (solid line) and nonsense (dashed line) words in five age groups.

Table 1. Means (SD) of meaningful word accuracy for different age groups in all modalities and post-hoc comparisons of age groups in each modality.
<table>
<thead>
<tr>
<th></th>
<th>4-5 years</th>
<th>8-9 years</th>
<th>20-30 years</th>
<th>40-55 years</th>
<th>65-80 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audiovisual</td>
<td>72.333 (8.837)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>86.177 (9.606)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>90.333 (5.815)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>87.333 (7.527)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>78.667 (18.561)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Auditory</td>
<td>58.333 (14.960)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>72.059 (12.254)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>79.000 (11.832)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>74.000 (12.421)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>64.667 (20.042)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Visual</td>
<td>2.667 (3.200)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5.294 (5.145)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>19.000 (11.052)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>13.333 (7.480)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>15.000 (9.820)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Table 2. Means (SD) of nonsense words accuracy for different age groups in all modalities. The data was used to calculate auditory and visual enhancement and the use of context (j and k factors).

- <sup>a</sup>4-5=65+ (p>.05), <8-9, 20-30, 40-55 (p<.01);
- <sup>b</sup>8-9=20-30=40-55 (p>.05), >4-5, 65+ (p<.01);
- <sup>c</sup>4-5=8-9 (p>.05), <20-30, 40-55, 65+ (p<.01);
- <sup>d</sup>20-30=40-55=65+ (p>.05), >4-5, 8-9 (p<.01)