Phonological Processing in Deaf Children: When Lipreading and Cues Are Incongruent

J. Alegria
J. Lechat
Laboratoire de Psychologie Expérimentale, Université Libre de Bruxelles

Deaf children exposed to Cued Speech (CS), either before age two (early) or later at school (late), were presented with pseudowords with and without CS. The main goal was to establish the way in which lipreading and CS combine to produce unitary percepts, similar to audiovisual integration in speech perception, when participants are presented with synchronized but different lipreading and auditory information (the McGurk paradigm). In the present experiment, lips and cues were sometimes congruent and sometimes incongruent. It was expected that incongruent cues would force the perceptual system to adopt solutions according to the weight attributed to different sources of phonological information. With congruent cues, performance improved, with improvements greater in the early than the late group. With incongruent cues, performance decreased relative to lipreading only, indicating that cues were not ignored, and it was observed that the effect of incongruent cues increased when the visibility of the target phoneme decreased. The results are compatible with the notion that the perceptual system integrates cues and lipreading according to principles similar to those evoked to explain audiovisual integration.

It has been shown that deaf children possess phonological representations of speech that are primarily based on lipreading (Dodd, 1976, 1987; Dodd & Hermelin, 1977; Leybaert & Alegria, 1995). Lipreading alone, however, cannot sufficiently support speech development because visual information is far more ambiguous than auditory information. Because of this requirement, speech therapists and teachers for the deaf have attempted to reduce ambiguities in lipreading by way of added visual and tactile signals.

Cued Speech (CS) is a system of manual aids developed to help deaf children to understand speech by eliminating lipreading ambiguity (Cornett, 1967). In the French version of CS, the hand can adopt eight shapes and be placed at five positions around the mouth (see Figure 1). Hand shapes are intended to clarify consonants and hand positions to clarify vowels (see Alegria, Charlier, & Mattys, 1999, for a more detailed report). The main aim of this article is to explore the manner in which lips and hands combine to produce available phonological information for deaf children.

Several studies have demonstrated that CS substantially improves speech processing (Alegria et al., 1999; Charlier, Hage, Alegria, & Périer, 1990; Leybaert, Alegria, Hage, & Charlier, 1998; Nicholls & Ling, 1982). These data imply that lips and cues are integrated to permit the listener to discriminate between words and pseudowords. Precise speech processing contributes to the elaboration of phonological representation of words that the child can use in tasks involving those representations, as demonstrated, for example, by Charlier & Leybaert (2000) for rhyme judgments; LaSasso, Crain, & Leybaert (2003) for rhyme generation; Leybaert (2000) for spelling production; and Alegria (2003) for reading and spelling. Moreover, some of those studies have shown that the efficiency of CS was more pronounced in children exposed to it at home from an early age than in children exposed to CS later in their life and solely at school.
In this context, it is interesting to note the results reported by Leybaert & Lechat (2001) using the paradigm of immediate recall of series of words. They showed that similarity in cue hand placement in items to be recalled produced detrimental effects compared with control items. These results reveal the use of visually based phonology (CS hand position) by deaf persons in working memory in a similar manner as acoustically based phonology plays in the case of hearing persons. Interestingly, similarity effects in hand placement were not observed in hearing subjects who were experienced producers of CS. This suggests that the use of CS-based phonology in mental activities depends not only on mere knowledge of CS but also probably on the early integration of cues to the linguistic experience of the child.

Studies considering how cues and lipreading information combine are rare. Alegria et al. (1999) explored this question, examining the conditions in which lipreading and cues fail to combine. The rationale was that lips and cues must necessarily be integrated into a specific phoneme. For example, a bilabial articulatory mouth movement accompanied by hand shape 1 corresponds to the phoneme /p/ but neither the bilabial feature nor Hand Shape 1 alone has a phonetically unambiguous meaning.

The authors (Alegria et al., 1999) looked for errors they called CS errors, which suggested that the cues had been processed independently from lipreading, for example, perceiving the syllable /pa/ presented in a lipreading-plus-CS condition as /da/ or /3a/ presumably because /p/, /d/, and /3/ share the same cue (Hand Shape 1). Another case in which specific features of CS could create phonological interpretations that are not present in the auditory lip-read message is when articulated syllables do not have the canonical CV structure. For example the initial C in a CCV syllable has to be articulated using the neutral hand position that also corresponds to the vowels /a, o, 3/ (Figure 1). In that situation, a receiver can mistakenly perceive /ba–li/ or /bo–li/ instead of /bli/. Alegria et al. reported results showing a significant difference between canonical (CV-CV) and noncanonical structures (V-CCV, V-CVC, and VC-CV), with the former correctly perceived as bisyllabic and the latter sometimes perceived as tri-syllabic. This probably results from the fact that CV syllables translate into one CS unit (a hand shape executed at a hand position), and the others involve two CS units (including the neutral cue). The extra CS unit in noncanonical syllabic structures can be interpreted independently from the labial information.

### Integration of Lip Movements and Hand Cues

The present experiment used a design inspired by experiments in the domain of audiovisual integration in

![Figure 1](French version of Cued Speech: (a) neutral hand shape used to code isolated vowels and (b) neutral hand position used to code consonants other than onsets of CV syllables. (Alegria et al., 1999; Charlier et al., 1990; Charlier & Leybaert, 2000; Leybaert, 2000).]
speech perception. The classic work by McGurk and McDonald (1976) consisted of presenting the participants with synchronized but different lipreading and auditory information (e.g., lips saying /ga/ and ears receiving /ba/). These results were confirmed and expanded in several studies (e.g., Campbell, Dodd, & Burnham, 1998; Massaro, 1987, 1989). It is now widely agreed that speech perception is an audiovisual activity, and current research and theory concerns the manner in which these two sources of information are integrated to extract abstract phonological codes (Green, 1998; Liberman & Mattingly, 1985; Summerfield, 1987, 1991).

The basic feature of this experiment involved exposing the participants to a situation in which lips and cues were sometimes incongruent, for example, the mouthed syllable /va/ accompanied by Hand Shape 1 (/p, d, 5/). These two sources of information are incongruous in that the hand shape permits disambiguation of three specific labial features: (1) bilabial (“it is /p/” instead of /b/ or /m/, which are also bilabials), (2) alveolar (“it is /d/” instead of /t/, /s/, /z/, which share the same labial features), and (3) the lips protrusion characterizing /5/ (which is common to /5/ and /f/). The lip posture /va/ thus is not one of those concerned by Hand Shape 1.

The empirical question under consideration was, What information from lips and/or cues will be taken into consideration by the participant? And, if both were considered, what sort of compromise will the system produce? Experiments concerning the integration of auditory and visual information suggest that each domain has its own specificity. For example, lipreading gives information about place of articulation but not at all about voicing and nasality. Auditory information, on the contrary, is rather precise concerning these two features but less precise concerning the place. Moreover, the place of articulation feature can take values that are more or less salient. Bilability, for example, is the most salient place, and if it is present in the stimuli, it is always considered in the percept (a visual /ba/ plus an auditory /ga/ will produce a /bda/ or /bga/ percept).

Models of audiovisual speech perception try to conceptualize the notion of specificity and salience of each source of information (Green, 1998; Massaro, 1987, 1989; Summerfield, 1987, 1991), and the salience of lipreading information thus is considered here. It was hypothesized that because /r/ and /k/ are less salient than /v/ and /f/, they are more likely to allow cues to determine the response. Similarly, the vocalic context can modify the salience of a consonant, and thus it can be reasonably supposed that /v/ is easier to perceive in a /va/ than in a /vo/ context. Montgomery, Walden, & Prosek (1987) have shown that the consonantal context can modify lipreading of vowels. Highly visible consonants, like bilabials and labiodentals, reduce vowel intelligibility; vowels with highly visible features like the rounded /y/ tend to decrease the perception of consonants (Owens & Blazek, 1985).

Finally, the factor age of exposure to CS was considered in the experiment. Previous findings have indicated that children exposed earlier to CS (at home) are better at phonological processing than children exposed later and only at school. It was hypothesized that because the participants are more sensitive to cues and probably also to lipreading information, they might show greater sensitivity to the conflict between sources of information.

Method

Participants

The participants were 20 prelingually deaf children and adolescents. They all had normal intelligence (IQ was routinely evaluated in school with nonverbal tests), and they attended classes that corresponded to their age. None exhibited learning or language disabilities, in terms of reading or spelling, other than those typical in a school for deaf children. All children progressed adequately in an oral setting. The participants were split into two groups depending on the age at which they had been exposed to CS: early and late.

The early group had 10 participants (7 females). They had been exposed to CS at home before the age of 2 years. In all of these cases, the parents had adopted CS to communicate with their children and used it systematically. The mean age of the early group was 9 years with a range of 6 years 8 months to 11 years. They all “were profoundly deaf” (hearing loss >90 dB in the better ear across 0.5, 1, and 2 kHz). All used binaural hearing aids.
The late group consisted of 10 participants (3 females). Their mean age was 11 years 8 months, with a range of 9 years 4 months to 15 years 8 months. Among them, 9 were profoundly deaf. The 10th had a total hearing loss (>120 dB). All used binaural hearing aids. They had begun using CS between ages 2 years 11 months and 7 years 9 months at the end of preschool classes or at the beginning of primary school. With one exception, they used it exclusively at school as a means of communication in the classroom. The exception consisted of case in which the parents had learned CS but used it to communicate with their child in an unsystematic way. The mean duration of CS exposure was 6 years 10 months (from 5 years 5 months to 8 years 5 months). The participants’ ability to use CS was not systematically tested because all had learned CS and were exposed to it daily in the classroom at least during the previous 3 years. Their teachers and speech therapists informed us that they all exploited CS to improve speech perception.

Stimuli and Conditions
The stimuli were monosyllabic CV structures made with eight consonants: /k, g, s, z, f, v, k, r/ and four vowels /a, ʌ, ə, ɔ, ʊ/. These stimuli were videotaped using normal articulation by a speech therapist fluent in CS and well known by the participants, but they were presented without sound. Each syllable was presented in three conditions: lipreading alone, lipreading with congruent cues, and lipreading with incongruent cues (henceforth called lipreading, congruent, and incongruent conditions). The stimuli were grouped in two blocks of 48 trials: 18 lipreading, 18 congruent, and 12 incongruent. One third of the items in each condition tapped vowel identification, and the other two thirds tapped consonant identification.

The vowels examined /a/ and /ʌ/ were presented in favorable and unfavorable consonantal contexts: /s, z/ and /ʃ, ʒ/, respectively, that is, paired with consonants (vocalic contexts) in which they are more or less visible in lipreading. In the incongruous condition, the syllables with /a/ were accompanied by the cues corresponding to the consonant but executed at Place 3 (chin): /e, ə, ɔ/. Similarly, syllables with /ʌ/ were presented with the cues executed at Place 5 (neck): /e, ʌ, ɔ/. Places 3 (chin) and 5 (neck) were chosen for the incongruent condition because both target vowels, /a, ʌ/, possess a clearly visible visual feature: openness, which is absent in all of the vowels conveyed by Places 3 (chin) and 5 (neck).

The anterior consonants examined, /v/ and /f/, were presented in favorable and unfavorable vocalic contexts: /a, ʌ/ and /o, ʊ/, respectively. In the incongruous condition, the syllables with /v/ were accompanied by the cues corresponding to the vowel but executed with Hand Shape 1: /p, d, ʒ/. Similarly, syllables with /f/ were presented with the cues executed with Hand Shape 4: /b, n, w/. Hand Shapes 1 and 4 were chosen for the incongruent condition because both target consonants, /v/ and /f/, possess a clearly visible visual feature: labio-dentalness, which is absent in all of the consonants conveyed by these hand shapes.

The posterior consonants examined, /k/ and /r/, were chosen because they are intrinsically less visible than the anterior consonants /v/ and /f/. Posterior items, as was the case for anterior consonants, were presented in the same favorable and unfavorable vocalic contexts: /a, ʌ/ and /o, ʊ/, respectively. In the incongruous condition, the syllables with /k/ were accompanied by the cues corresponding to the vowel but executed with Hand Shape 1: /p, d, ʒ/. Similarly, syllables with /r/ were presented with the cues executed with Hand Shape 4: /b, n, w/. Hand Shapes 1 and 4 were chosen for the incongruent condition because both target consonants, /k/ and /r/, are poorly represented at lipreading level; all of the phonemes conveyed by Hand Shapes 1 and 4 possess a visual feature more visible than those of /r/ and /k/.

Procedure
The participants faced a TV screen on which series of meaningless CV syllables were presented. A trial began when the experimenter started the videotape. The face on the screen was slightly oriented toward the ground. Then, it rose and, when it reached the frontal position, pronounced the syllable. The face tilted down, and the tape was stopped. The raising of the head constituted the warning signal indicating that the stimulus was imminent. The stimulus was pronounced normally,
so it took less than 1 second. The participant gave a written response on a response sheet to each trial without time constraints. The participant’s task was to identify either the vowel or the consonant, never both. In vowel trials, the consonant was already written in the response sheet, and in consonant trials, the vowel was given. The next trial started as soon as the participant had written the response and looked toward the screen. Participants were encouraged to guess if they were uncertain. French spelling, as well as English, involves numerous inconsistencies at the phoneme-to-grapheme translation level. For example, the phoneme /a/ might be spelled “an,” “en,” “ant,” and so on. Prior to the first block, the experimenter showed the children a training list containing items different from those used during the experiment but having the same phonological structure.

Participants were informed that any phonologically acceptable response was adequate. In French, however, unlike English, there are no inconsistencies at the grapheme-to-phoneme translation level. This property allows us to evaluate the participant’s responses unambiguously on phonological grounds. The participants were tested either individually or in small groups of two in a quiet room. The experimenter was the second author, who is used to communicating with children.

Results

Effects of Cueing Conditions

The main goal of this experiment was to determine the effects of cues, either congruent or incongruent, on phonological processing as a function of the age at which children were exposed to CS. Figure 2 represents the mean percentage of errors per condition and group of participants. Vowels and consonants are presented separately.

The basic pattern of results is similar in both groups. As expected, congruent cues improved performance; incongruent cues reduced it. Two analyses of variance on the errors data, one for vowels and the other for consonants, were performed. The factors considered were Group (early vs. late) as a between-subjects factor and Condition (lipreading, congruent cues, and incongruent cues) and phonetic Context (favorable or unfavorable vocalic context for consonant identification and consonantal context for vowel identification) as within-subjects factors. In the analysis of consonants, the factor Place of Articulation (anterior /v, f/ and posterior /r, k/) also was included as a within-subjects factor.

Analysis of Vowels

The percentage of errors was considerably greater in the late than in the early group [i.e., the main effect of Group was significant, \( F(1, 18) = 19.61, p < .001 \)]. The pattern of results was similar in both groups, however, as shown by the absence of interaction between Group and Condition \( (F < 1) \). It was expected that modifying the visibility of vowels (with favorable or unfavorable contexts) by associating specific consonants to them would correlative modify the effects of cues, both congruent and incongruent. Table 1 represents the mean percentage of errors per Group and Condition separately for vowels presented in favorable and unfavorable consonantal contexts.

A slight effect of Context was seen in the early but not in the late group, suggesting that the positive effects of congruent cues as well as negative effects of incongruent cues were greater when vowels were presented in an unfavorable consonantal context. The critical interaction Context \( \times \) Condition \( \times \) Group failed to reach significance \( [F(2, 36) = 2.44, p = .10] \). Separate analysis per Group, however, showed that the
Context × Condition interaction was near the conventional significance level in the early group but not in the late group \(F(2, 18) = 3.18, p = .057\) and \(F(2, 18) = 1.22, p = .32\), respectively.

### Analysis of Consonants

Despite slight differences in terms of errors between late and early groups, neither the main effect of Group nor the Group × Condition interaction reached significance \(F(1, 18) = 3.65, p < .072\) and \(F(2, 36) = 2.48, p < .097\), respectively.

Two factors that could modify the visibility of consonants were examined: the vocalic Context associated with the consonant to be identified and the Place of Articulation of the consonant itself. As for the vowels, it was hypothesized that reducing the visibility would increase the effects of cues. Table 1 represents the mean percentage of errors per Group and Condition as a function of vocalic Context, favorable or unfavorable. Statistical analysis showed that the vocalic Context produced reliable effects on consonant identification that varied by Condition and Group. The critical interactions Context × Condition and Context × Condition Group were both significant \(F(2, 36) = 3.30, p = .041\) and \(F(2, 36) = 4.31, p = .021\), respectively. As expected, unfavorable vocalic contexts had a detrimental effect on consonant identification in the lipreading condition: The mean percentage of errors was 36.7% and 60.0% in favorable and unfavorable vocalic contexts, respectively, in the early group. The corresponding values in the late group were 55.8% and 70.8%. Inspection of relevant data in Table 1 suggests that the effects of Context were independent of Condition in the late group. Analysis of variance confirmed that the Context × Condition interaction was not significant in this group \(F < 1\). In the early group, on the contrary, the effect of cueing conditions did depend on the vocalic context. They were stronger in the unfavorable than in the favorable vocalic context. Analysis of variance showed a significant Context × Condition interaction \(F(2, 16) = 6.79, p = .006\) in the early group.

The second factor that might modify the visibility of consonants is the Place of Articulation, anterior or posterior. It was hypothesized that posterior consonants might demonstrate a greater effect of cueing because they are less visible than anterior consonants. Statistical analysis showed that the effects of Place of Articulation depended weakly on Conditions and Groups. Place of Articulation × Condition and Place of Articulation × Condition × Group interactions failed both to reach the conventional level of signification \(F(2, 36) = 5.09, p = .101,\) and \(F(2, 36) = 1.63, p = .209\), respectively. The lower third of Table 1 presents the mean percentage of errors per group and condition for anterior and posterior consonants separately. Inspection of Table 1 indicates that, as expected, congruent cues produced greater improvement of

<table>
<thead>
<tr>
<th>Context Vowels</th>
<th>Early Lipreading</th>
<th>Congruent Cues</th>
<th>Incongruent Cues</th>
<th>Late Lipreading</th>
<th>Congruent Cues</th>
<th>Incongruent Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable</td>
<td>13.3</td>
<td>8.3</td>
<td>32.5</td>
<td>43.3</td>
<td>43.3</td>
<td>65.0</td>
</tr>
<tr>
<td>Unfavorable</td>
<td>10.0</td>
<td>0.0</td>
<td>42.5</td>
<td>48.3</td>
<td>36.7</td>
<td>60.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Context Consonants</th>
<th>Early Favorable</th>
<th>Early Unfavorable</th>
<th>Late Favorable</th>
<th>Late Unfavorable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>36.7</td>
<td>19.2</td>
<td>80.0</td>
<td>55.8</td>
</tr>
<tr>
<td></td>
<td>60.0</td>
<td>17.5</td>
<td>96.2</td>
<td>70.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Articulation Place</th>
<th>Early Anterior</th>
<th>Early Posterior</th>
<th>Late Anterior</th>
<th>Late Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonants</td>
<td>50.0</td>
<td>22.5</td>
<td>81.2</td>
<td>55.0</td>
</tr>
<tr>
<td></td>
<td>46.7</td>
<td>14.2</td>
<td>95.0</td>
<td>71.7</td>
</tr>
</tbody>
</table>

Table 1: Mean percentage of errors per Group (early and late) and Condition (lipreading, congruent cues, and incongruent cues) in three cases: vowels presented in favorable and unfavorable consonantal contexts, consonants presented in favorable and unfavorable vocalic contexts, and anterior and posterior consonants.
performance in posterior than in anterior consonants in both groups of participants. Symmetrically incongruent cues had a detrimental effect only in the early group. Separate analysis of variance showed that the Place of Articulation x Condition interaction was clearly significant only in the early group [$F(2, 18) = 4.11, p = .034$, and $F(2, 18) = 2.76, p = .090$, for early and late groups, respectively].

Analysis of Errors

As defined in the Introduction, CS errors are substitutions based on the CS structure, for example, giving the response /vo/ for the stimulus /ko/ presented with congruent Hand Shape 2: /k, v, z/. Similarly, in the incongruent condition the response /do/ is considered a CS error if /ko/ was presented with the incongruent Hand Shape 1: /p, d, z/. CS errors were evaluated by comparing their frequency in congruent and in incongruent conditions to the corresponding lipreading-only condition, which acted as a control (i.e., manual cues were absent).

In congruent condition, the mean number of errors was small, especially in the early group (4.9% and 11.6% in early and late groups, respectively). Most of these errors were CS errors in the early group (94.2%) but not in the “late” group (37.0%). The mean percentage of CS errors was submitted to an analysis of variance considering Group (early and late), Condition (lipreading and congruent), and Context (favorable and unfavorable). The Group x Condition interaction was highly significant [$F(1, 18) = 18.54, p < .001$], but Context as well as Context x Group were not significant [$F(1, 18) = 1.60, p < .22$, and $F < 1$, respectively).

In the incongruent condition, errors were more numerous than in the congruent condition (17.1% and 18.0% for early and late groups, respectively). This permitted more detailed analysis. The mean percentage of errors that were CS errors for consonants accompanied by incongruous cues are presented in Table 2. As was the case in the congruent condition, the proportion of CS errors was greater in the early than in the late group (78.8% and 58.7%, respectively).

The manner in which visibility of the consonant to be identified affected the frequency of CS errors was examined by considering two factors: vocalic Context
An analysis of variance was performed with Group (early and late) as a between-subjects factor and Condition (lipreading and incongruent cues), Context (favorable and unfavorable), and Place of Articulation (anterior and posterior) as within-subjects factors. It was hypothesized that reducing the visibility would increase the frequency of CS errors. The analysis concerning the vocalic Context (see Table 2) showed that both the main effect of Context and the Condition x Context interaction were significant, indicating that unfavorable vocalic context tended to increase the frequency of CS errors \[F(1, 18) = 27.82, p < .001,\] and \[F(1, 18) = 6.00, p = .025,\] respectively. The triple interaction Context x Condition x Group was not significant \((F < 1)\).

The analysis concerning the Place of Articulation showed that the main effect, the interaction with Condition, and the triple interaction Context x Condition x Group \((F > 1\) in each case) were not significant.

Finally, the tendency to produce CS errors with vowels \((/a, \tilde{a}/)\) compared with consonants \((/v, f, k, r/)\) was examined in the incongruent condition. Table 2 shows that CS errors tended to be less numerous with vowels than with consonants (about 40% and 75%, respectively) in both groups of participants. An analysis of variance was done considering the mean percentage of CS errors per Category of Phonemes (vowels and consonants), Condition (lipreading and incongruent cues), and Group (early and late). Condition was highly significant, as was Condition x Group interaction \([F(1, 18) = 67.81, p < .001,\) and \(F(1, 18) = 4.23, p = .054,\) respectively]. The Category of Phonemes was also significant, but its interaction with Group was not \([F(1, 18) = 18.90, p < .001,\) and \(F(1, 18) = 1.16, p = .298,\) respectively].

Nature of Cued Speech Errors

CS errors have been defined as a substitution of the correct segment by one that shares the hand shape or the hand position of the cue. With incongruent cues, CS errors could have three values. For example, the syllable /ko/ presented in lip-read with the incongruent Hand Shape 1 could produce the CS errors /po/, /do/, and /30/. The question under consideration in this section is to examine how the participant chose between these alternatives. If the information given by lipreading was totally ignored, then the choice among the alternatives offered by cues would be random. If, on the contrary, lipreading information was taken into account, then the response would reflect a compromise between both sources of information (as in the McGurk effect). Figure 3 depicts the total number of the four possible types of error (three CS errors and the non-CS error category “other”) per phoneme and group of participants.

The results concerning the consonants are clear-cut. The dominant response was either /d/ and /n/ for Hand Shapes 1 and 4, respectively. This tendency was slightly stronger in the early than in the late group. This bias might reflect a tendency to choose the phoneme less marked at the lipreading level among those compatible with cues. For example, among the
phonemes related to Hand Shape 1, /p/ and /z/ produce more salient lipreading information than /d/.

To establish the reliability of these results, an analysis of variance with Group (early and late) as a between-subjects factor and Type of Error (marked and unmarked) as a within-subjects factor was performed. The analysis showed that Type of Error was highly significant \[ F(1, 18) = 38.12, p < .001 \] because the unmarked errors represented 84.4% of the total number of errors. Despite a tendency for the early group to produce more unmarked errors than the late group (86.8% and 69.3%, respectively), the Group factor did not reach the conventional significance level \[ F(1, 18) = 2.83, p = .110 \].

The pattern of errors obtained with vowels was unexpected. The most frequent CS error corresponded to the phoneme /e/ for the incongruent Hand Position 3. For Hand Position 5, no clear tendency was observed. A possible reason was that the spelling of French vowels is highly inconsistent. This explanation is considered in the Discussion section.

**Discussion**

The present experiment was designed to examine the factors that determine how lipreading and CS information combine to produce a unitary percept. To address this issue, we placed deaf youngsters in a situation in which cues were sometimes congruent and sometimes incongruent with lipreading. It was expected that with incongruent cues the speech-processing system would be forced to adopt phonological solutions that might reveal the weight it attributes to different sources of information.

The results obtained demonstrate that the addition of congruent cues to lipreading improved performance for consonants as well as for vowels in both groups of participants, early and late CS learners. These findings confirm previous results. They support the notion that deaf children do possess a processor responsible for extracting phonological information from the lipreading input and, when present, from CS (Alegria et al., 1999; Charlier et al., 1990; Nicholls & Ling, 1982).

When cues were incongruent, the frequency of errors increased in both groups of participants. It is important to specify that errors in the incongruent condition have been evaluated relative to the lipreading condition, that is, without cues. The increase of errors shows that cues are not ignored when they are incongruent with the lipreading.

It has been hypothesized that the speech-processing system integrates cues and lipreading according to principles identical to those evoked to explain audiovisual integration in the classic McGurk effect. Theories of audiovisual speech perception, despite not negligible differences concerning mechanisms and level at which integration occurs, admit that the contribution of each source depends on the reliability of the information it gives. For example, place of articulation is relatively poorly represented in the auditory signal compared to voicing and nasality. The contrary is true with lipreading, which gives basically no information about voicing and nasality while it is relatively unambiguous conveying some aspects of place of articulation (see Summerfield, 1987, for a clear discussion concerning the specific contribution of auditory and lipreading information). The contribution of lipreading to the final percept seems to depend directly on the ambiguity of the auditory information.

Similarly, in the present experiment the ambiguity of phonological features given by lipreading were manipulated. It was assumed that the contribution of cues to the final percept would be inversely proportional to the salience of features given by lipreading. It can be added that the effects of cues when they are congruent with lipreading must depend on the same factors. There is no reason to suppose that congruent cues contribute to speech perception, but incongruent cues do not. It is reasonable to consider that the mechanisms involved in both cases are the same.

The results concerning the effects of visibility of consonants are clear-cut. Cue effects, both positive and negative, increased when the consonants to be identified were presented in an unfavorable vocalic context (e.g., /v/ was more affected in /vo/ than in /va/ context) and when they were intrinsically difficult to perceive because they were articulated at a posterior place (/r, k/ as opposed to /v, f/). This general tendency was clearly present in the early group of children but not in the late group.
Vowel identification showed similar but less marked tendencies than consonant identification. The effect of cues tended to increase when vowels were presented in an unfavorable consonantal context (e.g., /a/ was less affected by cues in /sa/ than in /fa/). For vowels, the effects of unfavorable contexts were observed in the early but not in the late group of participants. It is tempting to conclude that lipreading and cues combine following principles similar to those observed in audiovisual speech perception. This notion is reconsidered after the examination of error data.

The next point of interest was CS substitution errors. Their presence constitutes additional evidence in favor of the notion that cues, even when incongruent, are indeed taken into consideration. The results showed a significant increase of CS errors when cues were presented relative to the lipreading-only condition (which played the role of control condition because no cues were presented).

Important differences between groups also were observed at this level. In the congruent condition, almost all (94%) of the errors on consonants in the early group were CS errors. The corresponding percentage reached only 37% in the late group. For vowels, the results were similar (100% and 15%, respectively). It must be remembered, however, that the early group basically did not produce errors in this case. When cues were incongruent, the frequency of CS errors showed a slight reduction in the early group but remained an extremely high proportion of total errors (79%). In the late group, these errors increased (59%).

The visibility factor played the expected role in the early but not in the late group. The former group produced more CS errors for posterior (less-visible) than for anterior (more visible) consonants (83% and 75%, respectively), as well as for consonants presented in unfavorable vocalic versus favorable contexts (87% and 70%, respectively).

These results indicate that the susceptibility of lipreading information to be “invaded” by cues depended on its reliability, at least in the early group. The more the phonetic value of lipreading was ambiguous and/or difficult to extract, the greater was the tendency to exploit cues. The contrast observed in this regard between early and late groups on two points; (the proportion of CS errors and the relationship between their frequency and the visibility conditions) strongly suggests that early CS learners were more careful analysts of the information presented in both lipreading and cues.

A final and important point about CS errors concerns the choice of a response among the three available on pure CS grounds. The case of consonants was clear: Participants systematically chose the alternative that was less marked from the point of view of lipreading. When the incongruent cues had the Hand Shape 1, corresponding to /p, d, ʒ/, the more frequent choice was /d/. Similarly with the Hand Shape 4, corresponding to /b, n, w/, the more frequent choice was /n/. The speech-processing system thus seems to operate on the assumption that “if the target had been /p/ or /ʒ/, this should be seen on the lips,” so the decision then is “it must be /d/.” The same processes led to choosing /n/ from among /b, n, w/. A similar mechanism has been proposed to explain some aspects of the classic McGurk effect. For example, an auditory /ba/ combined with a lipreading /ga/ never produces the response /ba/. Such a response is incompatible with lipreading information, which unambiguously shows that the bilabial feature is absent. The most frequent response, /da/, is a compromise between both sources of information.

The errors for vowels were inconsistent. A clear tendency for errors to be concentrated on the less-marked phoneme associated with the incongruent hand shape was not observed in either the early or the late group. The most probable reason for this is that all responses were written. This was an important source of artifacts that we failed to anticipate: Vowel orthography in French is very inconsistent, and word endings in the French orthographic system are rather complex. Despite the fact that in the present experiment children knew that they were dealing with pseudowords and that orthography was irrelevant, they apparently could not ignore orthographic information. For example, we observed responses like “ant” for /ã/, which is phonetically correct but includes unnecessary morphological information. The response “e” for /ə/ or /ã/ was frequent. It was difficult to determine what the subject was trying to spell, a “schwa” (“e muet” is very frequent in French) or the phoneme /e/. For all these reasons, it seems too speculative to discuss
these data further. In a future experiment, we shall consider a greater variety of vowels and not place them in final position using, for example, CV-CV structures with the first vowel as the critical item.

The results discussed so far are compatible with two extreme models. The correct one probably lies in between. One model supposes that both sources of speech information, lipreading and CS, are hierarchically processed, the former giving the core phonological information and the latter intervening optionally. That is, cues are taken into account when necessary to resolve lipreading ambiguities. According to this model, the participation of CS in speech processing takes place at a late, postperceptual stage. In the second model, the speech-processing device automatically integrates all of the available visually based phonological information, including CS and lipreading. The functioning of this device might be similar to the audiovisual speech-processing device postulated in normally hearing people (Green, 1998; Massaro, 1987; Summerfield, 1987; Summerfield & McGrath, 1984). Phonemes are defined as abstract units that allow the distinction between different morphemes, so that lipreading together with cues may logically elaborate phonemic contrasts. This formulation implies that the speech-processing device can combine exotic signals (cues) with natural phonological information (lipreading and available residual hearing). Fowler & Deckle (1991) showed that auditory and tactile lip information do combine to produce McGurk-like effects. Our findings, in contrast, indicated that such combinations are possible. Additional evidence (Breeuwer & Plomp, 1986; Green & Miller, 1985; Rosen, Fourcin, & Moore, 1981) demonstrating that nonspeech auditory signals also combine with lipreading to produce genuine phonetic percepts is also relevant in this context.

To return to our main question, that of the modality of lipreading and cues combination, an important problem for future research should be to understand better the exact nature of this integration, either as a conscious postperceptual mechanism or as an automatic compulsory one. Furthermore, this alternative must be considered in light of the early versus late learner contrast because its resolution could depend on the age of exposure to CS. The results of the present experiment showed important differences between both groups, but these differences were all compatible with the notion that the early group of children was more efficient exploiting cues than the late group. The difference might be simply a difference of degree and not of nature.

To conclude, it is important to underline some aspects of the present results that might have practical consequences for the education of deaf children. The notion that speech is multimodal and that it includes a natural visual dimension, lipreading, that allows an artificial visual signal, CS, to elaborate a phonological signal has important consequences. This means that CS is not just an artifact aimed at helping deaf children distinguish between words, but that it becomes a genuine part of speech processing and speech representation. The conditions necessary for CS to reach this perceptual and cognitive status are clearly related to the age of exposure to it. We recognize that the present results are not totally conclusive about the notion that early exposure is a condition sine qua non for the integration of cues in speech processing at a perceptual automatic level. As a matter of fact, this experiment showed that deaf children in the late group also exploited cues to improve speech perception, although to a lesser extent than children in the early group. The present results suggest that when the decision to adopt an oral education has been taken, intervention should take place as early as possible and involve the parents of the child to provide a complete and unambiguous speech model.

**Note**

1. French and English CS, as well as more than 50 other natural languages (Cornett, 1994), are based in exactly the same principle. Some differences do exist, however, that result from specific features of each phonological system. For example, English CS has four hand positions (French CS Hand Position 1, cheek-bone, is missing); in English CS but not in French, hand movements from one position to another are used to disambiguate diphthongs, which are frequent in the former but not in the latter language.

**References**


Received May 25, 2004; revisions received September 13, 2004; accepted September 14, 2004