Audiovisual perception of openness and lip rounding in front vowels

Hartmut Traunmüller and Niklas Öhrström
Department of Linguistics, Stockholm University

Postal address:

Lingvistik
Stockholms universitet
S-106 91 Stockholm

Corresponding author: Hartmut Traunmüller, hartmut@ling.su.se

Abstract

Swedish nonsense syllables /gig/, /gyg/, /geg/ and /gøg/, produced by four speakers, were video-recorded and presented to male and female subjects in auditory, visual and audiovisual mode and also in cross-dubbed audiovisual form with incongruent cues to vowel openness, roundedness, or both. With audiovisual stimuli, subjects perceived openness nearly always by ear. Most subjects perceived roundedness by eye rather than by ear although the auditory conditions were optimal and the sensation was an auditory one. This resulted in fused percepts such as when an acoustic /geg/ dubbed onto an optic /gyg/ was predominantly perceived as /gøg/. Since the acoustic cues to openness are prominent, while those to roundedness are less reliable, this lends support to the "information reliability hypothesis" in multisensory perception: The perception of a feature is dominated by the modality that provides the more reliable information. A mostly male minority relied less on vision. The between-gender difference was significant. Presence of lip rounding (a visibly marked feature) was noticed more easily than its absence. The influence of optic information was not fully explicable on the basis of the subjects' success rates in lipreading compared with auditory perception. It was highest in stimuli produced by a speaker who smiled.

Keywords: audiovisual speech perception, audiovisual integration, vowel perception, lip rounding, lipreading, McGurk effect, information reliability

1. Introduction

It has been known for a long time that in speech communication, optic information for speech is not only relied upon by the deaf, but also to some extent by people with normal hearing. Sumby and Pollack (1954) investigated how the visibility of a speaker's face influenced the intelligibility of English words produced by the speaker as a function of the auditory signal-to-noise (S/N) ratio. Their results indicated that visual cues contribute substantially to speech comprehension at low S/N ratios. These results were later confirmed by Erber (1969). Amcoff (1970) investigated the ability of normal hearing subjects to visually discriminate between different vowels and consonants of Swedish. These and similar investigations (Risberg & Agelfors, 1978) showed very good visual perception of labial features such as presence or absence of lip rounding in vowels and presence or absence of bilabial or labiodental closure in consonants. However, the contribution of the visual sense to speech perception was generally assumed to be of importance only in a noisy environment and in people with a hearing loss. Most theoretical models of speech perception treated the phenomenon as a purely auditory process.

The topic of audiovisual integration in speech perception was found to be even more extensive when McGurk and MacDonald (1976) published their seminal study of audiovisual perception of speech stimuli with conflicting cues. The stimuli used by McGurk and MacDonald (1976) were repeated syllables with stop consonants and a following open vowel (e.g., [baba], [gaga], [papa] and [kaka]) that in the crucial cases had been dubbed on visual stimuli with different consonants. With such stimuli they made an unexpected observation: An auditory [baba] presented in synchrony with a visual [gaga] evoked the percept of [dada]. In addition to "fusions" analogous to this example, in which the percept neither agrees with the acoustic nor with the optic stimulus, McGurk and MacDonald also observed "combinations" such as when an auditory [gaga] presented together with a visual [baba] evoked the percept of [gabga] or [bagba]. These observations demonstrated that ordinary speech perception is a bimodal process in which information from the auditory and visual modality is integrated in an interesting fashion. The influence of the optic signal varies with the choice of vowel context (Green, 1996).

Subjects exposed to stimuli with conflicting audiovisual cues typically report "hearing" also the features conveyed by optic information. In a pioneering magnetoencephalographic study, Sams et al. (1991) actually concluded that visual information from lip movements modifies activity in the human auditory cortex. For a summary of more recent research with a wider perspective on the cortical mechanisms involved see Möttönen (2004). Interactions in which a percept is altered by information received in a different modality have also been observed in other cases than audiovisual speech perception. Shams et al. (2000) have, e.g., shown that visual perception (of flashing frequency) can be altered by sound, even when the visual stimulus lacks ambiguity.

Audiovisual integration appears to be very robust since it persists even when the visual and auditory components are from speakers different in sex (Green, Kuhl & Melzoff, 1991). Furthermore, Hietanen, Manninen, Sams and Surakka (2001) investigated audiovisual integration with different facial configurations (e.g., inverted mouth and eyes), and they found that the McGurk-effect remained strong as long as the configuration was symmetric. Another proof of the robustness of audiovisual integration is due to Rosenblum and Saldaña (1996) who showed that audiovisual integration does not require detailed information about the speaker's face. In their identification task with incongruent audiovisual stimuli, the visual information about the speaker's face was solely present in form of point-lights

attached to the face. Their results showed a strong visual influence in perceiving consonantal place features in spite of the paucity of the visual information.

Although audiovisual integration is characterized by robustness in many ways, it is subject to variation between languages or cultures and also between genders and individuals. Sekiyama and Tohkura (1993) carried out a cross-language study in which Anglo-American and Japanese listeners were presented incongruent audiovisual stimuli from both Anglo-American and Japanese speakers. The Japanese listeners proved to be much less influenced by the visual signal than the Anglo-American listeners when presented stimuli from Japanese speakers. Nevertheless, the visual contribution to perception among the Japanese listeners was much more important in noise (Sekiyama & Tohkura, 1991) and when the speaker was a foreigner. Hayashi and Sekiyama (1998) made a similar comparison between Japanese and Chinese listeners with incongruent audiovisual stimuli from Japanese and Chinese speakers. Both groups of listeners showed little susceptibility to the visual signal but, again, for the Japanese listeners, the visual contribution was more important when the speaker was a foreigner. With Chinese second language speakers of Japanese, Sekiyama (1997) reported the initially low susceptibility to visual information to increase with the length of stay in Japan.

In addition to cultural and experiential differences, there also appears to be a gender difference in susceptibility to visual input: Johnson, Hicks, Goldberg and Myslobodsky (1988) showed that women perform better than men in lipreading tasks. This also suggests women to be more susceptible to visual input in audiovisual identification tasks with incongruent cues. This suggestion was confirmed by Aloufy, Lapidot and Myslobodsky (1996) with speakers of English, but the between-gender difference was much smaller among speakers of Hebrew. Recently, Irwin, Whalen and Fowler (in press) failed to observe a significant between-gender difference in Anglo-American listeners when stimuli of the type used by McGurk and MacDonald were presented visually alone or in incongruently dubbed fashion, while they did observe such a difference when shorter incongruent audiovisual stimuli were presented. They also reported subjects to be more susceptible to static visual stimuli than to dynamic ones.

While it has been observed that the strength of the McGurk effect varies as a function of vowel context and has even been investigated with mismatched vowels (Green & Gerdman, 1995), audiovisual perception of vowels as such has to our knowledge not yet been studied in experiments analogous to those of McGurk and MacDonald (1976). However, Summerfield and McGrath (1984) carried out experiments in which manipulated auditory [bVd] syllables were presented to English subjects together with visual [i], [ɑ] and [u] faces. They observed that the phoneme boundaries in a two-dimensional auditory vowel space described by F1 and F2' in Bark units (where F2' is a function of F2, F3 and F4) were moved a bit closer towards the position of the vowels presented visually. These results showed that visual input biases the auditory perception of vowels. Johnson, Strand and d'Imperio (1999) showed, investigating the [ʊ]-[ʌ] boundary of American English, that vowel perception is biased not only by visual input but also, to a minor extent, by cognitive factors.

Lisker and Rossi (1992) did an investigation with phonetically trained subjects who had to tell whether French and non-French vowels presented to them in auditory, visual, and congruent as well as incongruent audiovisual mode were rounded or unrounded. Among the subjects of Lisker and Rossi, the probability of hearing a vowel as rounded increased on average by roughly 30% when the image showed lip rounding. Since the subjects were not asked about perceived vowel identity but about perceived roundedness, these substantial results allow no conclusions about and no immediate comparison with audiovisual integration in the perception of other features.

The present study is aimed at the question whether fusions analogous to those observed by McGurk and MacDonald (1976) appear also in vowel perception and how the information on roundedness and openness from the two modalities is fused in these cases. Languages in which lip rounding is an independent distinctive feature are suited to test these questions. In Swedish, lip rounding is distinctive among front vowels, as in most Germanic and the Finno-Ugric languages, as well as in French. In Swedish, except for the variety spoken in Finland, there are two qualitatively different types of rounding, which can be easily distinguished from each other as well as from unrounded lips in lipreading (Amcoff, 1970; Traunmüller, 1979). The auditory cues to lip rounding appear to be less reliable than the visual cues, and in Swedish /y/ vs. /i/ they are probably also less reliable than those in French. Therefore, it could be expected that perceivers are more heavily influenced by vision in the perception of this feature and perhaps less so in the perception of openness, for which F1 nearly always provides a prominent relational acoustic cue. This would be in line with Robert-Ribes, Schwartz, Lalluache & Escudier (1998), who concluded on the basis of audiovisual perception experiments in varied levels of noise that vision serves a complementary function to audition in that it conveys rounding better than height and backness, while audition conveys height better than backness and rounding. We shall also keep an eye on between-subject differences, which may be correlated with gender.

Most investigations within this field were performed with speech produced by a single speaker. Such studies leave it completely unclear to which population of speakers the observed results can be generalized. Different speakers have been reported to be easier or more difficult to lip-read (Gagné, Masterson, Munhall, Bilida and Querengesser, 1994; Kricos, 1996) and this is likely to affect audiovisual integration as well. We are going to use two representatives of each sex, which appears to be the minimum required in order to justify the assumption of some external validity across speakers.

1.1. Swedish vowels

Phonemically, nine short and nine long vowels can be distinguished in standard Swedish, but it is also possible to defend the position that there are just nine vowel phonemes and to consider the quantity distinction as a suprasegmental one, which affects not only the duration of vowels but, in a contrastive way, also that of consonants that follow. Among the long vowel phonemes, which are listed in Table 1, three are unrounded and six are rounded (produced with protruded lips). Although the epithet "spread" is often used as a synonym of "unrounded", active spreading of the lips occurs only exceptionally in contrastive contexts. Among the rounded vowels, three are "out-rounded" (just rounded), and three are "in-rounded", i.e., labialized in addition to being rounded. While the protrusion of the lips is the same in both kinds of rounded vowels, the vertical distance between the lips is smaller in in-rounded vowels although the jaw is lower (Traunmüller, 1979). Thereby, the visibility of the teeth is reduced. The distance between the corners of the mouth opening is also smaller. The distinction between in- and out-rounding is only maintained among the long vowels.

INSERT TABLE 1 ABOUT HERE

In open syllables, the close-mid vowels /eː/, /øː/ and /oː/ are normally diphthongized to become more open towards their end, while the close vowels /iː/, /yː/, /ʉː/ and /uː/ tend to become first even more close and then more schwa-like towards their very end (Eklund & Traunmüller, 1997). For the vowels /ɛː/ and /øː/, the more open allophones [æː] and [œː] are used in certain contexts, generally before /r/, and by some speakers more generally. While only three degrees of openness are phonemically distinctive, F1 is distinctly higher in

[æː] and [ɒː] as compared with [ɛː], which requires a forth degree of openness to be distinguished in allophones. In the present investigation, only four vowels were to be intended by the speakers: /iː/, /eː/, /yː/ and /øː/, but all the long vowels were allowed as response alternatives. Short vowels did not occur.

2. Method

2.1. Subjects

The speakers were two men (29 and 45 years of age) and two women (21 and 29 years). They were students and researchers at the Department of Linguistics, Stockholm University. All four were speakers of the regional variety of standard Swedish.

In the perception experiment, 10 men (16, 20, 22, 22, 26, 26, 27, 28, 30, 49 years) and 11 women, approximately matched in age (18, 20, 21, 22, 23, 26, 26, 28, 28, 30, 48 years), volunteered as listeners. They reported normal hearing and had normal or corrected-to-normal vision. All were phonetically naïve native speakers of Swedish.

2.2. Speech material

The choice of the speech material was based on the following considerations: (1) The vowels should differ by independent distinctive features whose visual reflection is sufficiently clear to be perceived by lipreading. Within the Swedish vowel system, lip rounding, labialization and openness all fulfill these requirements – only backness does not. (2) The influence of context on the realization of the distinctive vowel features chosen should be minimal. (3) The stimuli should all be phonotactically allowed non-words. Based on these considerations, the long front vowels /iː/, /yː/, /eː/ and /øː/ were chosen, and they were embedded within a /g_g/ frame to produce the phonotactically possible nonsense syllables /giːg/, /gyːg/, /geːg/ and /gøːg/. Due to a diachronic process in which /g/ before front vowels has developed into /j/, there are few words in contemporary Swedish in which an initial [g] is followed by any of the chosen front vowels. Velar consonants affect the visibility of vowel features hardly at all, since they do not require any particular position of the lips and the jaw. Labial consonants would impair vowel recognition substantially, as can be seen in Amcoff's (1970) data.

The speakers' faces were recorded using a video camera Panasonic NV-DS11. The camera and each face were at equal heights and the distance between them was 2.5 meters. The acoustic signal was recorded using a microphone (AKG CK 93) at a distance of 70 cm from the speaker's mouth.

The recorded utterances were subsequently edited and dubbed using Adobe Premiere 6.0. Each visual stimulus was synchronized, based on the burst of the first [g], with each one of the auditory stimuli from the same speaker. In the perception experiment, each visual and each auditory stimulus was also presented alone. In this way, 24 final stimuli were obtained from each speaker. A demonstration is accessible on the Web (Öhrström & Traunmüller, 2004).

During the experiment, all the recorded and the manipulated stimuli from all speakers were pooled and each stimulus was presented twice in random order. Thus, the perception test consisted of 192 stimuli in total. These were presented in 24 blocks of eight stimuli each, using Microsoft Windows Media Player. The interval between successive stimuli was 4 s and 10 s between blocks.

Figure 1 shows for each speaker the frequency values of the formants F1 and F2, measured at 30%, 50% and 70% of the duration of each vowel. While these data show the usual pattern of formant movement in the realizations of the vowel /e:/, the realizations of /ø:/ were very open from their beginning in at least two of the speakers, as can be seen in the high values obtained for F1 in this vowel already at the 30%-point. It can also be seen that F1 and F2 had almost the same values in the [i:] as in the [y:] of each speaker. In order to distinguish these vowels, F3 and F4 have to be considered. For each speaker, both formants had higher values in [i:] than in [y:]. The data suggest that the length of the vocal tract of female speaker S was above average, which agrees with her slightly long-necked appearance.

INSERT FIGURE 1 ABOUT HERE

2.3. Experimental procedure

The subjects participated one by one. They wore headphones AKG K135 and were seated with their faces at 60 cm from a computer screen. The height of the speakers' faces was roughly 17 cm on screen. The subjects were instructed in written and in spoken form to visually focus on the speaker's mouth while listening and to write down which vowel they had heard and to write down the vowel perceived by lipreading when the stimulus was purely visual. They used response sheets and were allowed to choose any of the 9 ordinary Swedish letters that represent vowels. Only one answer per stimulus was allowed. Prior to the experimental blocks, one training block with 8 stimuli was run. The subjects were supervised during the whole session to make sure they were focused on the screen all the time. The entire session lasted about 20 min.

3. Results

A preliminary analysis of the results suggested that the listeners did not all agree in their behavior. In order to see whether different groups need to be distinguished, stepwise regression analyses were performed for the individual results of each listener, using the independent factors "auditory openness", "auditory roundedness", "visual openness" and "visual roundedness".

The result showed that "auditory openness" explained most of the variance for all 21 listeners. For the majority (16 'Sharp Eyes'), "visual roundedness" explained next to most, i.e., it explained more than "auditory roundedness" did. For a minority (5 'Sharp Ears'), "auditory roundedness" explained more of the variance than "visual roundedness" did. It is of interest to note that the Sharp Ears' group included 4 of the 10 male listeners (40%) but only 1 of the 11 female listeners (9%).

There were just 9 cases (among 1344 responses) in which vowel openness in stimuli with conflicting cues was perceived in accordance with the optic signal. Table 2 shows to the left for each listener in how many cases of audiovisual stimuli with conflicting cues to roundedness (64 stimulus presentations) roundedness was perceived in accordance with the nominal acoustic or optic stimulus. There were no response omissions. Among the Sharp Eyes, > 50% of the rounding responses were visual.

INSERT TABLE 2 ABOUT HERE

It is evident from Table 2 that the susceptibility to the optic signal was, on average, higher among female than among male subjects. There was no male among the six subjects who relied most heavily on optic information. The difference between the genders turned out to be clearly significant ($p < 0.01$), but there was also substantial between-gender overlap in behavior.

There was also a significant ($p < 0.01$) between-gender difference in the error rate obtained in lipreading, when no auditory signal was presented. This error rate was 35% among men and 25% among women. As for the roundedness feature, the error rate was 4.7% among men and only 1.1% among women (difference significant at $p < 0.05$). Women performed better than men also in the perception of openness (error rates 25% and 32% respectively, difference significant at $p < 0.05$). Unexpectedly, there was no significant correlation between the error rates or, equivalently, between the success rates in visual perception of openness and roundedness ($p \approx 0.4$).

As can be grasped from Table 2, the reliance on optic cues in audiovisual perception of roundedness was only weakly correlated with the same listeners' success rate in visual perception of roundedness ($r = 0.38$, $p \approx 0.09$). This was not essentially affected by removing the results for three cases in which the stimuli were misclassified by half the subjects or more. These stimuli can be regarded as optically misproduced, but they were not eliminated from the following analyses. The reliance on optic cues in audiovisual perception of roundedness was also weakly but negatively correlated with the listeners' success rates in auditory perception of roundedness ($r = -0.40$, $p \approx 0.07$). Considering the success rates in auditory and in visual perception of roundedness as two independent variables resulted in a moderately successful model with $r = 0.60$ (36% of the variance accounted for, $p < 0.02$). A simpler model, in which the difference between the visual and the auditory success rate of each subject in the single-mode conditions was taken as the only independent variable (rightmost column in Table 2), did almost equally well ($r = 0.60$). It showed the reliance on optic information as compared with the acoustic to be significantly correlated ($p < 0.005$) with the difference between the visual and the auditory success rates of the subject.

A comparison of the data on roundedness errors in single mode (Table 2, "Difference") shows that there were fewer errors in visual than in auditory perception: 2.7% as compared with 6.7%. The difference was significant at $p < 0.002$ in a paired samples t-test.

Table 3 shows the results obtained with each one of the four speakers. The results can be seen to be biased in a speaker specific way. The unrounded vowels of long-necked speaker S were often auditorily perceived as rounded. In the visual condition, the vowels were often perceived as more open, except those of speaker J, which were perceived as less open than intended. In audiovisual conditions, listeners were most sensitive to visual input from speaker J, who always had a smile in her face. The data suggest a correlation with the attractiveness of a speaker's look, but general validity of this observation cannot be asserted on this basis.

INSERT TABLE 3 ABOUT HERE

The distinction between the two listener groups that are kept apart in the analyses presented in the following is based on the individual performance that showed itself in the preliminary analysis and not on gender. The pooled results of each of the two groups of subjects are shown in form of confusion matrices in Table 4. In these confusion matrices, six conditions of stimulus presentation are distinguished: (1) image without sound (lipreading); (2) sound without image; (3) sound plus matching image; (4) sound plus image discrepant in

openness; (5) sound plus image discrepant in roundedness; and (6) sound plus image discrepant in both openness and roundedness.

INSERT TABLE 4 ABOUT HERE

When the stimuli were presented in visual mode alone (1), auditory mode alone (2), and audiovisual mode (3), the overall error rates obtained were 28%, 8% and 0.2% for Sharp Eyes, and 34%, 3% and 0.6% for Sharp Ears. The reduction in error rate that can be seen for auditory presentation compared with visual presentation and for audiovisual compared with auditory presentation was significant ($p < 0.001$) in each case. While the errors in lipreading were dominated by incorrectly perceived openness, the errors obtained in the auditory mode can all be considered as errors in perceived lip rounding. In the cases in which an /øː/ had been perceived as an /ɛː/, the /øː/ appears to have been realized as a more open allophone [œː]. The error rates for feature perception are listed in Table 5.

INSERT TABLE 5 ABOUT HERE

When the auditory stimuli were presented with conflicting cues such as an image discrepant in openness (4), in roundedness (5) and in both openness and roundedness (6), we cannot speak of "error rates" but we can calculate the percentage of cases in which the responses were in agreement with the optic rather than with the acoustic stimulus. These percentages were 0.2%, 78% and 78% for Sharp Eyes, and 0.6%, 20% and 26% for Sharp Ears. The results obtained with conflicting cues in both roundedness and openness are shown in Table 6. We have here to distinguish between 'auditory responses' in agreement with the acoustic signal, 'visual responses' in agreement with the optic signal and two types of fused responses in which one feature agreed with the acoustic and the other feature with the optic signal. However, all cases of fused responses turned out to be cases in which visual roundedness had been fused with auditory openness.

INSERT TABLE 6 ABOUT HERE

The pattern of confusions in roundedness was asymmetric, in particular among Sharp Eyes (see Table 7). These rarely identified a vowel as unrounded when the lips could be seen as rounded, but identifications with rounded vowels were not equally rare when the lips remained visibly unrounded. This asymmetry was significant in lipreading ($p < 0.01$), as well as in audiovisual perception with discrepant cues to roundedness ($p < 0.001$) and to roundedness as well as openness ($p < 0.001$). In all these conditions, the asymmetry failed to attain significance among Sharp Ears.

INSERT TABLE 7 ABOUT HERE

Using the numerical values for roundedness (0 or 1) and openness (0 for [i y], 1 for [e ø o], 2 for [ɛ ɒ]), the mean values of perceived roundedness and openness were computed for each stimulus in each one of the two groups of listeners. These data were used as dependent variables in stepwise linear regression analyses in which the interaction factors (visual openness * visual roundedness and auditory openness * auditory roundedness) were also considered as independent variables. The result is shown in Table 8. The weights listed there represent the slopes of the regression lines between the dependent variables and the independent variable listed on the row in question. These weights remain the same if the analysis is done on the basis of the individual results. The high values of r^2 obtained show that the mean results can be predicted very well on the basis of these independent variables

although the correlation cannot be assumed to be strictly linear. As can be seen here again, both groups relied on audition in openness perception, with no significant contribution of visual cues. However, in roundedness perception, Sharp Eyes relied on vision, and the contribution of auditory roundedness even failed to attain independent significance. The other group relied mainly on audition, but did not fully neglect visual cues to roundedness, which could also be seen in Table 2.

INSERT TABLE 8 ABOUT HERE

4. Discussion

The present investigation has, firstly, shown that normal hearing listeners benefit substantially from optic information even under ideal auditory conditions, at least with *some* speakers, and that auditory perception of Swedish vowels can be said to be deficient. While the error rate was very low in the audiovisual condition (only two errors in all, equal to 0.3%), it was as high as 8% among Sharp Eyes and 3% among Sharp Ears in the ideal auditory-only condition (46 errors in all). A similar error rate of 7% can be calculated for the Swedish vowels [i:] [y:] [e:] and [ø:] pronounced in isolation by six male and six female speakers in the study by Eklund & Traunmüller (1997), where similarly ideal auditory conditions prevailed. Differences in the distribution of the errors between the two studies probably reflect speaker-specific particularities such as observed between the four speakers used in the present investigation. In both studies, the vast majority of vowel confusions concerned the feature of roundedness. This allows us to conclude that the auditory cues for this distinctive feature are less reliable than those for the degree of openness.

The investigation has, secondly, shown that audiovisual integration of the kind that results in a new, fused percept that neither agrees with the acoustic nor with the optic stimulus occurs not only in consonants, but also in vowels. This happened particularly often in the Sharp Eyes' group. These perceived, for instance, an auditory [y:] presented in synchrony with a visual [e:] as an /i:/ in 107 of 128 cases. This can be considered as analogous to the fusions observed in consonant perception by McGurk & MacDonald (1976) when an auditory [ba] was presented in synchrony with a visual [ga] and subjects perceived a /da/. In this case, the visual absence of labial closure appears to have prevented listeners from perceiving /ba/. In the present experiment, the visual absence of lip rounding appears to have prevented listeners from perceiving /y:/. We can see that the short duration of the acoustic manifestation that is characteristic of stop consonants is not a prerequisite for audiovisual fusions to occur.

An auditory [e:] synchronized with a visual [y:] was mostly perceived as an /ø:/ (in 127 of 128 cases by the Sharp Eyes). In this case, the visual presence of lip rounding appears to have prevented listeners from perceiving an /e:/. This effect was stronger than the effect of visible absence of lip rounding, exemplified by the before mentioned case. Also in the other cases, listeners appeared to notice the incompatibility of visibly rounded lips with the presence of an unrounded vowel more easily than the incompatibility of a neutral (unrounded) shape with the presence of a rounded vowel (see Table 7), although even this kind of mistake did not occur very often in lipreading. This can be understood on the basis of the hypothesis that listeners attach a lower weight to cues that are merely indicative of the *absence* of marked features. When an unrounded vowel is produced, the speaker's lips normally just remain in their neutral state. Unrounded vowels are, in this sense, unmarked for roundedness. Perceptually salient properties that provide reliable cues for the presence of marked features would, according to this hypothesis, seldom be ignored. The absence of such a property does not provide an equally reliable cue for the absence of the

corresponding feature, since natural speech is often degraded, which entails a loss of such properties. Therefore, it is rational to attach a lower weight to their absence.

When an auditory [eː] is synchronized with a visual [yː], conditions are reminiscent of those that prevail when an auditory [ga] is synchronized with a visual [ba]. In both cases, an optic stimulus that shows a clearly visible labial feature accompanies an acoustic stimulus that lacks this feature. Auditory [ga] synchronized with visual [ba] would probably be perceived as a labiovelar /g͡ba/ by listeners familiar with labiovelar stops, but McGurk & MacDonald's subjects were not. Instead, they perceived consonant clusters such as /gba/ or /bga/ in which the auditory and the visual place features appear in serial combinations of familiar phones. In the present experiment, serial combinations analogous to these were not expected, since all the possible combinations of distinctive features result in vowel phonemes that exist in the subjects' language. Therefore, instances of perceived serial combinations were not investigated. Nevertheless, some of the subjects informally reported having heard non-standard diphthongs in some cases when exposed to stimuli with conflicting cues. This remains to be further investigated.

While visible cues contributed to the perception of roundedness among all subjects, their contribution exceeded that of auditory cues in the more susceptible majority group. It is especially striking that the contribution of "auditory roundedness", aside from its interaction with auditory openness, even failed to attain independent significance within this group. This makes the present results concerning the perception of roundedness in vowels appear much more extreme than those reported by Lisker & Rossi (1992). Our subjects also made significantly fewer roundedness errors in visual than in auditory single mode perception. Since most auditory errors were due to one of our four speakers (see Table 3), the conclusion that the roundedness feature should be considered as primarily conveyed by the optic rather than by the acoustic speech signal remains tentative. However, our data provide no support for the opposite hypothesis.

The paucity of /ɒː/ responses to [œː] stimuli in lipreading, which contrasts with the abundance of confusions between these vowels in Amcoff's (1970) data, can only be understood as due to priming by the set of vowels the subjects perceived in the auditory and audiovisual conditions. However, the paucity of confusions with /ʉ/, /u/ and /o/ agrees with Amcoff's data and must be primarily due to the fact that these are in-rounded (labialized). The absence of labialization in [yː] and [øː] can easily be seen.

Our results agree with those obtained previously with consonants in that they demonstrate a dominant role of visual input in the perception of labial features. While these are the most easily visible features, visibility alone does not suffice as an explanation, since openness is also easily visible. However, under audiovisual conditions, the perceptual weight of the visible cues to openness was found to be insignificant and close to zero among all listeners, even among those who were most susceptible to optic input. The auditory cues to this feature were evidently much more salient than the visual. For the perception of this feature, the contribution of the optic signal may become important only when the S/N ratio in the acoustic channel is unfavorable. The importance of optic cues is known to increase when the acoustic S/N ratio decreases (Sumby & Pollack, 1954; McLeod & Summerfield, 1987; Robert-Ribes et al., 1998).

The results obtained in the present and in previous experiments on audiovisual speech perception lend support to the "information reliability hypothesis" in multisensory perception: Perception is dominated by the modality that provides the more reliable information (Schwartz et al., 1998; Wada, Kitagawa & Noguchi, 2003; Andersen 2005). This hypothesis is compatible with the observations in audiovisual speech perception if it is

applied separately to the cues for each distinctive feature. In the present experiment, audition provided the most reliable information on openness, while vision provided the most reliable information on roundedness. The acoustic cues to labiality are very weak, in particular for the distinction between Swedish [y:] and [i:], while the position of F1 in relation to its neutral position, about which f0 and the higher formants are indirectly informative, is an acoustically strong cue to openness, which is also resistant to noise (Robert-Ribes et al., 1998).

The reliability of visual cues to vowel openness is lowered by the fact that the openness of the mouth varies not only as a function of vowel quality. In ordinary situations, listeners have a chance of tuning in to the voice and the face of a speaker and can so perform an appropriate linguistic demodulation of the acoustic as well as of the optic speech signal despite this variation. Due to sufficient intrinsic cues provided by f0 and the formants above the second one, subjects had a good chance of tuning in to the acoustic signal even in this experiment with mixed presentation of speakers, but there may not have been sufficient intrinsic information in the optic signal in order to factor out the contributions of vocal effort and speaker specific behavior to the visible openness of the mouth. Such uncertainty may explain the lack of a significant correlation between the subjects' success rates in visual perception of openness and roundedness. The recognition of lip rounding and labial closure is more robust against paralinguistic variation since the visible cues for these features are bounded by saturation effects, and their recognition does not require distinguishing intermediate degrees. The weight of visual cues to openness could have turned out larger if the speakers had been presented in blocked fashion.

The fact that auditory realizations of an /øː/ were often perceived as an /ɛː/ when combined with a visual [eː] or even with an [iː] is reflected in the significant auditory roundedness-openness interaction (Table 8) and it highlights the weakness of visual cues for the perception of openness. The /ɛː/ responses can only be understood on the basis of acoustic cues: the frequency position of F1 in a Swedish [øː] is, typically, higher than in [eː], and the F2 of [øː] is close to that of an [ɛː] (Eklund & Traunmüller, 1997). The acoustic data from the stimuli used in the present experiment agree with this, except that some of the realizations of /øː/ were remarkably open from their beginning and more like [œː] throughout. The high frequency of occurrence of /ɛː/ responses to auditory [œː] stimuli presented together with visually unrounded vowels follows naturally from the information reliability hypothesis applied at the level of cues.

Audiovisual integration can be described by Bayesian and fuzzy logical models, as exposed by Massaro and Stork (1998). In this approach, the probability for the presence of phonemic segments (or strings of segments) is established within each modality before integration. Although this procedure has been shown to be successful, it involves a degradation of the available information. The highly reliable information on labiality features that is only available in the visual modality is degraded by the less reliable information offered in the same modality on other features, such as the degree of openness in vowels. Similarly, the highly reliable auditory information on openness is degraded by the less reliable information offered in the same modality on labiality features. This drawback would be avoided if intermodal integration was assumed to take auditory and visible *cues* as its input. Similar arguments against late integration have been voiced by Schwartz et al. (1998) and Andersen (2005).

The information reliability hypothesis and the mentioned models, even if assumed to operate on the basis of cues, fall short of explaining why listeners report *hearing* vowels as rounded, when this perception can only have been based on optic input. However, this is in agreement with a previously suggested "modality appropriateness hypothesis" (see Welch

& Warren, 1980), which applies to perception in general and according to which the modality that is more attuned to the perceptual task at hand will dominate in cross-modal interactions and capture the sensation. Since audition is, after all, most appropriate for perceiving spoken messages, it captures the phonetic information received by the visual sense so that it appears to come from the auditory sense. This also shows us that sensory integration in normal speech perception cannot be fully described as an amodal process subordinated to auditory and visual perception, but it requires a key role to be ascribed to the auditory modality.

Since lip rounding is not an independent distinctive feature within the vowel system of English, a result equivalent to the present one could not have been obtained by Summerfield and McGrath (1984), and the very clear difference in susceptibility to visual input observed here between openness and roundedness is not reflected in their data. The lower relative weight attached by the subjects of Lisker and Rossi (1992) to visible lip rounding is likely to have been due to the larger acoustic difference between unrounded and rounded front vowels of French as compared with that between the unrounded and out-rounded vowels of Swedish. This is particularly true for the pair transcribed in both languages as /i:/ vs. /y:/, for which the difference in F2 is substantial in French, at least according to the data by Lisker and Rossi (1992), while it is only marginal and very unreliable in Swedish. It shall also be noted that Lisker and Rossi reported the pooled results of all their subjects without respect to their individual susceptibility to optic cues although they reported this to have varied greatly across subjects, which agrees with the present observations.

The fact that the minority of listeners who did not rely very much on optic information in the present study included mainly male listeners is in accord with the reported greater susceptibility of female listeners to visual input and their greater proficiency in lipreading (Johnson et al., 1988), which the present results confirm. This is likely to be due to the gender difference in gaze behavior. Women tend to look at a speaker's face more frequently and for longer periods of time than men (Argyle & Ingham, 1972). This tendency has been shown to be present already at an age of 4 months (Leeb & Rejskind, 2004). It is reasonable to assume that this tends to make women not only more attentive to visual information in speech perception but also more skillful in assimilating such information, while men may become more skillful in assimilating the auditory information.

While the gender difference in gaze behavior may be due to nature or culture (Leeb & Rejskind's results do not allow rejecting any of these hypotheses), clear cases of culturally conditioned differences are also known. In some cultures, gaze at a speaker's face is avoided, in particular when the speaker is of a higher status. Gaze behavior may also be affected by socialization experiences with the mother during the first year of life, so that less gaze may be expected from those who have been carried a lot on the back during infancy (Argyle & Cook, 1976). Such factors may explain why Japanese men as well as women are less susceptible to visual input (Sekiyama & Tohkura, 1993; Hayashi & Sekiyama, 1998). Sekiyama (1997) considered also a linguistic explanation for the ethnic differences observed. Gaze behavior would explain the gender difference observed elsewhere as well. The greater prominence of effects of visual input shown by Japanese subjects when listening to a foreigner and by Chinese subjects who had been in Japan for a longer period (Sekiyama, 1997) requires a different explanation.

We have seen that women are overrepresented among the better lip readers as well as among those who rely more on optic cues in audiovisual speech perception. Although this suggests a correlation, an individual perceiver's susceptibility to optical cues in audiovisual perception could not be predicted satisfactorily on the basis of success rate in lipreading

alone. A similar observation was made by Sekiyama (1997) concerning the distinction between labial and non-labial consonants. However, when the individual success rates in auditory perception were taken into account in addition to those in lipreading, 36% of the variance was explained in the present experiment. This result still allows for the possibility that the attention subjects pay to visual information as compared with the auditory may be a free parameter in audiovisual integration. Listeners may, e.g., attach a higher weight to visual input from a smiling speaker than from a speaker with a more austere look, as our data tentatively suggest.

Acknowledgments

This paper is an elaboration of a contribution to the Swedish Phonetics Conference Fonetik 2004. Its preparation was supported by grant 421-2004-2345 from the Swedish Research Council. We are grateful to Doug Whalen and to three anonymous reviewers for stimulating comments.

References

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

Amcoff, S. (1970). *Visuell perception av talljud och avläsestöd för hörselskadade* [Visual perception of speech sounds and speech reading support for the hard of hearing]. *Rep. 7*, Lärarhögskolan, Pedagogiska institutionen, Uppsala.

Andersen, T. S. (2005). *Model-based assessment of factors influencing categorical audiovisual perception.* (Doctoral dissertation, Helsinki University of Technology). http://lib.tkk.fi/Diss/2005/isbn9512275481/

Argyle, M. & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Argyle, M. & Ingham, R. (1972). Gaze, mutual gaze, and proximity. *Semiotica, 6*, 32-49.

Eklund, I. & Traunmüller, H. (1997). Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica, 54*, 1-21.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12*, 423-425.

Gagné, J. P., Masterson, V., Munhall, K. G., Bilida, N. & Querengesser, C. (1994). Across talker variablility in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech. *Journal of the Academy of Rehabilitative Audiology, 27*, 135-158.

Green, K. P. (1996). The use of auditory and visual information in phonetic perception. In D. G. Stork & M. E. Hennecke, (Eds.), *Speech Reading by Humans and Machines: Models, Systems, and Applications* (pp. 55-77). Berlin: Springer.

Green, K. P. & Gerdeman, A. (1995). Cross-modal discrepancies in coarticulation and the integration of speech information: the McGurk effect with mismatched vowels. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1409-1426.

Green, K. P., Kuhl, P. K. & Meltzoff, A. N. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk

effect. *Perception & Psychophysics, 50*, 524-536.

Hayashi, T. & Sekiyama, K. (1998). Native-foreign language effect in the McGurk effect: a test with Chinese and Japanese. In the Proceedings of the *AVSP'98* (pp. 61-66). Terrigal, Australia.

Hietanen, J. K., Manninen, P., Sams, M. & Surakka, V. (2001). Does audiovisual speech perception use information about facial configuration? *European Journal of Cognitive Psychology, 13*, 395-407.

Irwin, J. R., Whalen, D. H. & Fowler, C. A. (in press). A sex difference in visual influence on heard speech. *Perception & Psychophysics.*

Johnson, F. M., Hicks, L., Goldberg, T. & Myslobodsky, M. (1988). Sex differences in lipreading. *Bulletin of the Psychonomic Society, 26*, 106-108.

Johnson, K., Strand, E. A. & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*, 359-384.

Kricos, P. B. (1996). Differences in visual intelligibility across talkers. In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading by Humans and Machines: Models, Systems, and Applications* (pp. 43-53). Berlin: Springer.

Leeb, R. T. & Rejskind, F. G. (2004). Here's looking at you, kid! A longitudinal study of perceived gender differences in mutual gaze behavior in young infants. *Sex Roles, 50*, 1-14.

Lisker, L. & Rossi, M. (1992). Auditory and visual cueing of the [± rounded] feature of vowels. *Language and Speech, 35*, 391-417.

Massaro, D. W. & Stork, D. G. (1998). Speech recognition and sensory integration. *American Scientist, 86*, 236-244.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

McLeod, A. & Summerfield, A. Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*, 131-141.

Möttönen, R. (2004). *Cortical Mechanisms of Seeing and Hearing Speech*. (Doctoral dissertation, University of Technology, Helsinki). http://lib.hut.fi/Diss/2004/isbn9512274272/

Öhrström, N. & Traunmüller, H. (2004). *Audiovisuell perception av vokaler / Audiovisuelle Wahrnehmung von Vokalen / Perception audiovisuelle de voyelles* (Demonstration): http://www.ling.su.se/fon/fonetik_2004/saima_0.html

Risberg, A. & Agelfors, E. (1978). Information extraction and information processing in speech-reading. In *STL-QPSR 2-3/1978*, pp. 62-82. Department of Speech Transmission and Musical Acoustics, Royal Institute of Technology, Stockholm.

Robert-Ribes, J., Schwartz, J.-L., Lallouache, T. & Escudier, P. (1998). Complementarity and synergy in bimodal speech: Auditory, visual and audio-visual identification of French oral vowels in noise. *Journal of the Acoustical Society of America, 103*, 3677-3689.

Rosenblum, L. D. & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 318-331.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S-T., Simola, J (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127*, 141-145.

Schwartz, J.-L., Robert-Ribes, J. & Escudier, P. (1998). Ten years after Summerfield: A taxonomy of models for audio-visual fusion in speech perception. In R. Campbell (Ed.), *Hearing by Eye: The Psychology of Lipreading* (pp. 3-51). Hove, U.K.: Erlbaum.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics, 59*, 73-80.

Sekiyama, K. & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America, 90*, 1797-1805.

Sekiyama, K. & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics, 21*, 427-444.

Shams, L., Kamitani, Y. & Shimojo, S. (2000). What you see is what you hear. *Nature, 408*, 788.

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

Summerfield, A. Q. & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology, A 36*, 51-74.

Traunmüller, H. (1979). Lippenrundung bei schwedischen Vokalen [Lip rounding in Swedish vowels]. *Phonetica, 36*, 44-56.

Wada, Y., Kitagawa, N. & Noguchi, K. (2003). Audio-visual integration in temporal perception. *International Journal of Psychophysiology, 50*, 117-124.

Watkins, K. E., Strafella, A. P. & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologica, 41*, 989-994.

Welch, R. B. & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638-667.

Table 1

Distinctive features and dimensions of the long vowel phonemes of Swedish. Openness values within parentheses refer to allophones.

Phoneme	/iː/	/eː/	/ɛː/	/yː/	/øː/	/ɒː/	/oː/	/uː/	/ʉː/
Letter	i	e	ä	y	ö	a	å	o	u
Open	0	1	2 (3)	0	1 (2)	3	1	0	0
Back	0	0	0	0	0	1	1	1	0
Rounded	0	0	0	1	1	1	1	1	1
Labialized	0	0	0	0	0	0	1	1	1

Since IPA offers no symbols for labialized vowels, the symbol "ʉ" is here used for representing a front vowel. (This phoneme is actually realized as a mid vowel [ʉː] in the variety of Swedish spoken in Finland, where in-rounding is absent.)

Table 2

Audiovisual perception of roundedness compared with performance in single mode (listening only or lipreading only) for each subject. In the auditory mode, all errors involved roundedness.

		Response to audiovisual stimuli with incongruent cues to roundedness		Roundedness errors in single mode (n = 64)		
Listener	Sex	Auditory	Visual	Auditory	Visual	Difference
Sharp	m	55	9	0	0	0
Ears	m	55	9	1	2	-1
	m	46	18	1	1	0
	m	43	21	2	4	-2
	f	41	23	1	0	1
Sharp	f	30	34	3	2	1
Eyes	f	30	34	1	1	0
	m	30	34	2	0	2
	m	24	40	5	2	3
	m	22	42	1	3	-2
	m	19	45	0	0	0
	m	19	45	3	2	1
	f	15	49	2	0	2
	f	15	49	4	0	4
	m	15	49	4	1	3
	f	8	56	3	0	3
	f	5	59	5	1	4
	f	4	60	3	0	3
	f	2	62	1	0	1
	f	2	62	1	0	1
	f	1	63	2	0	2

Table 3

Audiovisual perception of roundedness compared with perception in single mode (listening only or lipreading only) for each speaker. '-': perceived as less rounded or less open; '+': perceived as more rounded or more open.

		Responses to audiovisual stimuli with incongruent cues to roundedness		Roundedness errors in single mode				Openness errors in single mode			
Speaker	Sex	Auditory	Visual	Auditory		Visual		Auditory		Visual	
				-	+	-	+	-	+	-	+
P	m	163	173	2	1	1	1	0	1	9	42
R	m	145	191	0	1	0	1	0	0	5	62
S	f	100	236	0	28	2	11	0	1	7	34
J	f	69	267	9	4	1	4	0	4	30	8

Table 4
Confusion matrices for six different conditions. Columns: Auditory stimulus presented, visual stimulus presented; perceived vowel quality for 6 male and 10 female 'Sharp Eyes' and for 1 female and 4 male 'Sharp Ears'.

		Sharp Eyes							Sharp Ears					
Aud.	Vis.	i	y	e	ø	ɛ	ɒ	o	i	y	e	ø	ɛ	ɒ
					Image only (lipreading)									
*	i	74	3	43	7	1			23	1	15	1		
*	y		77		51				1	25		14		
*	e	21	1	102	1	3			6	3	31			
*	ø		10	1	115		1	1		10	1	27		2
					Sound only									
i	*	117	11						37	3				
y	*	4	124						2	38				
e	*			108	20						40			
ø	*				122	6						40		
					Sound plus matching image									
i	i	127	1						39	1				
y	y		128							40				
e	e			128							40			
ø	ø				128							40		
					Sound plus image discrepant in openness only									
i	e	128							39	1				
y	ø		128							40				
e	i			127	1						40			
ø	y				128							40		
					Sound plus image discrepant in roundedness only									
i	y	7	120		1				31	9				
y	i	99	28	1					8	32				
e	ø			19	109						31	9		
ø	e			8	59	61						34	6	
					Sound plus image discrepant in both openness and roundedness									
i	ø	16	108		4				28	12				
y	e	107	17	4					14	26				
e	y			1	127						26	14		
ø	i			6	79	43						39	1	

Table 5
Average error rate (in percent) for stimuli presented in visual mode alone, in auditory mode alone, and in audiovisual mode without conflicting cues. All subjects pooled, n = 672 in each condition. The error rate listed for "openness" within parentheses considers only the distinction of /i:/ and /y:/ from any more open vowels and vice versa.

	Roundedness	Openness	Backness	Labiality
Visual mode	3.0	28.3 (27.3)	0.6	0.1
Auditory mode	6.8	0.9 (0.0)	0.0	0.0
Audiovisual mode	1.1	0.0 (0.0)	0.0	0.0

Table 6
Response percentages for stimuli with fully conflicting cues (/i:/&/ø:/, /y:/&/e:/, /e:/&/y:/, /ø:/&/i:/).

	Sharp Eyes n = 512	Sharp Ears n = 160
Auditory response	22	75
Visual response	2	0
Auditory openness fused with visual roundedness	76	25
Auditory roundedness fused with visual openness	0	0

Table 7
Confusion matrix for roundedness in lipreading and in audiovisual perception with conflicting auditory cues; results from the Sharp Eyes' group only. Rows: Visual roundedness; Columns: perceived roundedness.

	Lipreading		Conflicting auditory roundedness		Conflicting auditory roundedness and openness	
	Unrounded	Rounded	Unrounded	Rounded	Unrounded	Rounded
Visually unrounded	244	12	169	87	160	96
Visually rounded	1	255	26	230	17	239

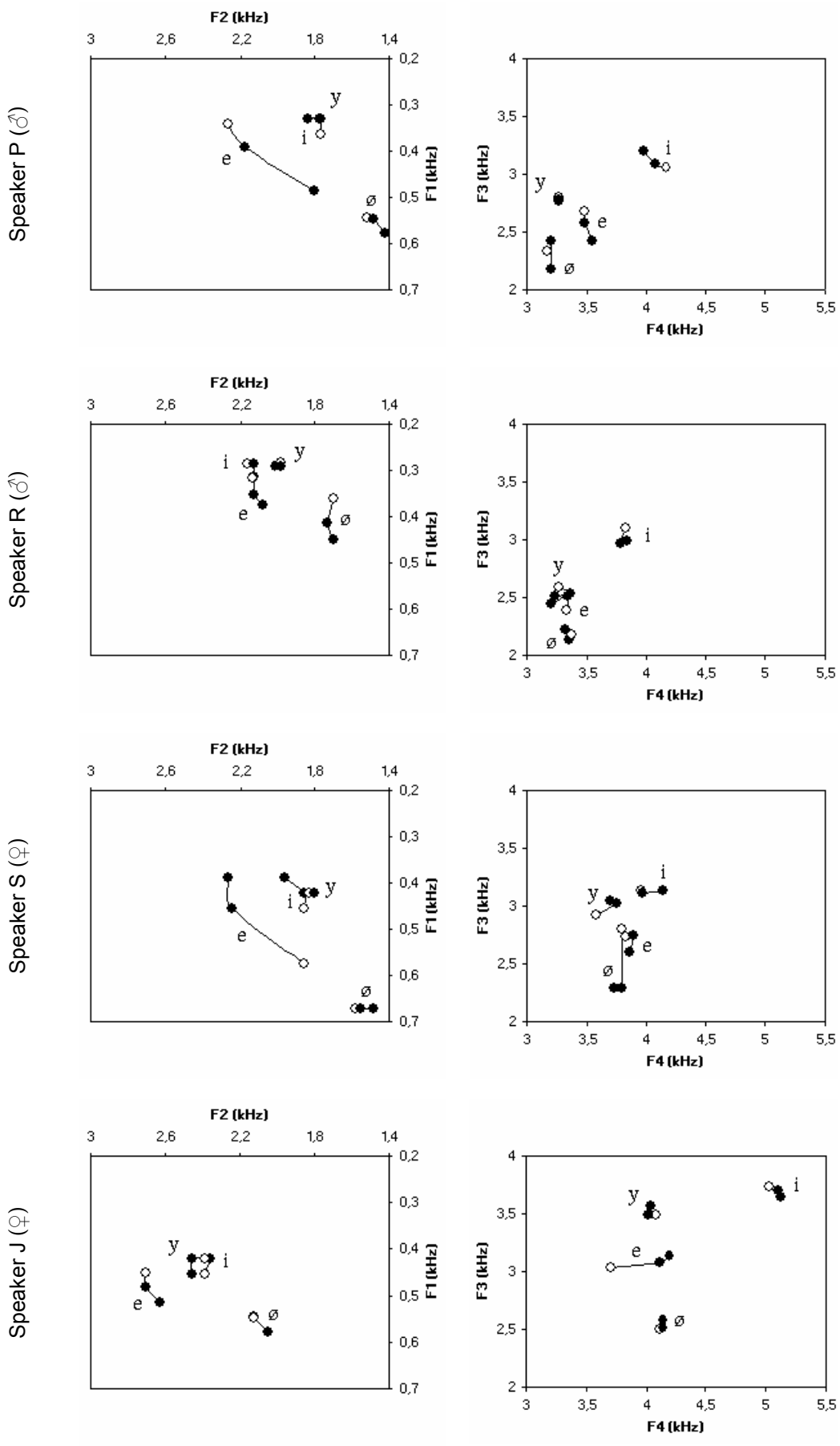
Table 8
Results of stepwise linear regression analyses of the roundedness (Round) and openness (Open) of each stimulus as perceived by each group of listeners: Constant (intercept) and unnormalized weights (w) of significant independent variables. The symbol "I" stands for "Interaction", "ns" and "-" both for "not significant", "-" in cases that appear to be a priori irrelevant. Last row: Variance of the group mean data explained (r^2).

Regression equation: Dep. var. = $C + O_A w_{OA} + O_V w_{OV} + R_A w_{RA} + R_V w_{RV} + I_A w_{IA} + I_V w_{IV}$

	Sharp Eyes n = 1536		Sharp Ears n = 480	
	Round	Open	Round	Open
Constant (C)	0.18	0.01	0.03	0.00
Auditory openness (O_A)	-	0.99	-	1.02
Visual openness (O_V)	-	ns	-	ns
Auditory roundedness (R_A)	ns	-	0.77	-
Visual roundedness (R_V)	0.77	-	0.22	-
Auditory openness*roundedness (I_A)	0.27	0.20	ns	ns
Visual openness*roundedness (I_V)	ns	ns	ns	ns
r^2	.926	.965	.972	.995

FIGURE CAPTION

Fig. 1. Formant frequencies of each speaker measured at 30% of the duration of the vowel (open symbols), at 50% and at 70% (filled symbols), shown for each vowel in the frame [g_g]. Left plots: F1 and F2 with traditionally oriented axes; right plots: F3 vs. F4.


Speaker P (♂)
Speaker R (♂)
Speaker S (♀)
Speaker J (♀)
F2 (kHz)
F1 (kHz)
F3 (kHz)
F4 (kHz)
i
y
e
ø