A preliminary study on speech perception and production by blind people

Fabrice Hirsch[1+2], Rudolph Sock[2], Béatrice Vaxelaire[2], Camille Fauth[2], Fayssal Bouarourou[2], Marion Béchet[2], Melissa Barkat-Defradas[1]

[1]Université Paul Valéry - Montpellier III, Praxiling UMR 5267, CNRS, et Institut de Phonétique de Strasbourg – IPS & U.R. 1339 Linguistique, Langues et Parole – LilPa, E.R. Parole et Cognition

[2]Université de Strasbourg, Institut de Phonétique de Strasbourg – IPS & U.R. 1339 Linguistique, Langues et Parole – LilPa, E.R. Parole et Cognition

fabrice.hirsch@univ-montp3.fr

Abstract. *The present research deals with speech production and perception by blind people in [VCV] and [CVC] sequences. The aim of this work is double: first, we want to observe if blind people are capable of perceiving potential anticipatory effects of a rounded vowel in [V1CV2]. This would be due to compensatory auditory readjustments related to perturbation of the visual canal. The second goal of this research evokes timing particularities and formant structures in the production of blind subjects, particularities which may result from the fact that congenitally blind subjects have not benefited from face-to-face communication cues, known to ontogenetically optimize replication of visible gestures during speech. Results show that blind subjects perceive rounded vowels earlier than control sighted subjects. Concerning the temporal dimension, we notice differences in timing gestures between non-sighted and sighted groups.*

1. Introduction

Speech production and perception has been a topic of much interest for phoneticians in the last ten years (Cavé *et al.*, 2010; Hirsch *et al.*, 2010; Sato *et al.*, 2010; Menard *et al.*, 2009, Moos *et al.*, 2008, *e.g.*.). Concerning speech perception, Moos *et al.* (2008), for instance, observed that non-sighted listeners are able to perceive ultra-fast speech, whereas unimpaired people cannot understand such sentences. Moreover, Menard *et al.* (2009) showed that congenital blind people have more accurate auditory discrimination abilities than sighted adult speakers.

Concerning speech production, the same research (Menard, 2009) reveals differences in the formant structure of vowels, *i.e.* in spatial organization, when data from blind and sighted people are compared. In other words, the two groups would not produce exactly the same gestures to carry out vowels. This could be explained by the fact that non-sighted people do not have a visual model during speech acquisition; hence they would develop other articulatory patterns.

Consequently, and according to the literature, studying speech perception and production by congenital blind people presents at least two interests: first, it allows focalising on specificities of speech perception and production by blind people; secondly, analysing consequences of speech acquisition without any visual input could contribute to obtain more information about speech acquisition by sighted people.

The present research deals with speech perception and production by blind people. The first part of this article treats auditory effects of anticipatory lip movements in [V1CV2] sequences, where V1 is vowel [i], C is the fricative [s] and V2 a rounded vowel [Vlab]. The aim of the investigation is to analyse the behaviour of congenital blind subjects and sighted people when they have to perceive potential anticipatory effects of the rounded vowel. The *hypothesis* here is that blind subjects would perceive a protruded vowel earlier than unimpaired subjects, and in a more robust manner (following a specific confidence threshold) in [V1CV2] sequences. This would be so, due to compensatory auditory readjustments related to perturbation of the visual canal.

The aim of the speech production part of the current study is to highlight timing particularities in the production of congenital blind subjects. More precisely, we wanted to know if eventual differences in timing perception by non-sighted people would be accompanied by specificities in timing gestures. Such findings would help verify if lack of visual clues (if it is at all the reason for differences between blind and sighted people) has consequences on the temporal level.

Concerning this point, several hypotheses could be made for blind subjects: 1. If differences are observed during the perception task, we should notice specificities in the speech production task also; in this case, we would obtain a group effect between blind people and control groups; 2. For blind people, the only auditory information available would be sufficient to acquire articulatory and acoustic patterns similar to those of sighted people. If this hypothesis is correct, results for the two groups would be similar.

2. Method

2.1. Speech perception tests in [VsVlab] sequences

Material and acoustic measures for [VsVlab] sequences

French sentences, like “C’est [isVlab] ça” were recorded in an anechoic room. [Vlab] was either [y], [ø], [u] or [o]. Each sentence was pronounced three times by a speaker. This corpus allows testing lip aperture and protrusion effects on early perception of the rounded vowel. The first three formant of [i] and Vlab were measured at the middle of the vowel. Furthermore, the inferior limit of the frication noise (ILFN) was tracked every 10 ms, from onset of the clear formant structure of the rounded vowel [Vlab], backwards into the obstruent interval [s], up to the [i] in order to analyse, from a phenomenological perspective, acoustic cues that may be allowing precocious perception of the upcoming rounded vowel. Notice that the frication noise tends to decrease in the vicinities of a labialised vowel.

Auditory tests

The gating paradigm was used, sentences being truncated every 10 ms from onset of the clear formant structure of the rounded vowel [Vlab], backwards into the obstruent interval [s], up to the [i].

20 subjects participated in the speech perception test: 10 sighted subjects and 10 blind people. They had to carry out two tasks: 1) identify the truncated vowel; 2) attribute a confidence score (cs) to their answers, in a subjective range going from 1 (not sure or guessing) to 5 (quite confident).

2.2. Speech production tests

Speaker and material

5 blind and 5 sighted people were recorded pronouncing [CVC] sequences inserted in a French carrier sentence. C was either [p], [t] or [k] and V was [i], [a] or [u]. The carrier sentence was the following: "C'est la [CVp] à Bordeaux" ([s ɛ l a CVp a b ɔ r d o]).

Temporal analyses

As shown in Figure 1, the following temporal measures (see Sock, 1998) were obtained using the speech editor Praat:

- Acoustic silence: this measure goes from the end of the preceding vowel (offset of a clear formant structure) to the beginning of the burst-release;
- Voice Onset Time (VOT): this interval goes from the beginning of the burst-release up to onset of the clear formant structure of the vowel;
- Vowel duration: this measure is taken as the interval between onset and offset of the formant structure of the vowel, flanked by the two consonants;
- Voice Termination Time which is measured as the distance between offset of the vowel's formant structure to the last regular voicing buzz within the second consonantal closure;
- Consonantal duration of C2 ([p]) which corresponds to the interval between offset of the formant structure of the target vowel to onset of the subsequent vowel.

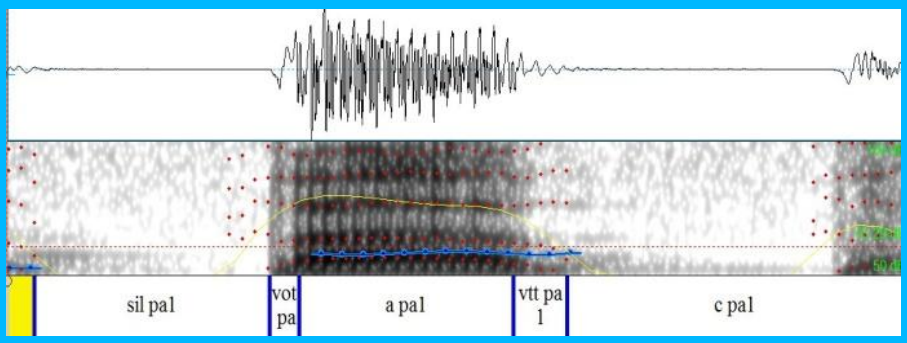

Figure 1: Example of measures taken for a sequence [pa].

To limit effects due to signal elasticity, measures of all parameters were normalized by calculating the percentage of time taken by a given parameter within the [CVp] domain.

3. Results

3.1. Speech perception tests

Results given here are part of a recent investigation (Hirsch *et al.*, 2010). As shown in Figure 2, congenitally blind people tend to perceive the rounded vowel before the control group. Indeed, the vowel [y] starts to be perceived by 70% of the blind people group at 110 ms from onset of the clear formant structure of the rounded vowel (cs= 3), whereas the control group identifies the same rounded vowel only at 90 ms before its onset (nc=2).

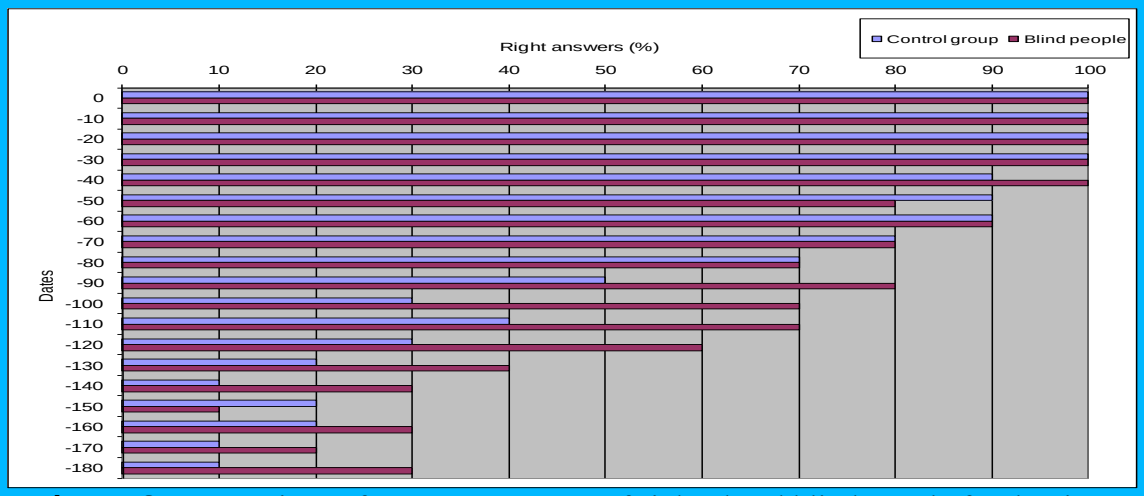

Figure 2. Comparison of correct responses of sighted and blind people for the [isy] sequence. The *x* axis shows correct percent responses, and the *y* axis gives truncation dates.

The date where control subjects start to perceive the vowel [y] coincides with a special event on the spectrogram (Figure 3). Indeed, we can observe that the inferior limit of ILFN for the [s] decreases in two steps: first, it slightly diminishes just after the [i]. Secondly, a more remarkable inflexion of the ILFN is observed up to the formant structure of the labialised vowel. More precisely, ILFN was at 3864 Hz at the first truncation point after the [i] and at 3473 Hz at 90 ms for the [y]; at this date, this measure drops to 1337 Hz, at 10 ms from onset of the clear formant structure of the [y]. We can notice here that control subjects starts to perceive the rounded vowel at the moment where ILFN begins to fall drastically, whereas blind people perceive the [y] as from the first slight declination of the ILFN.

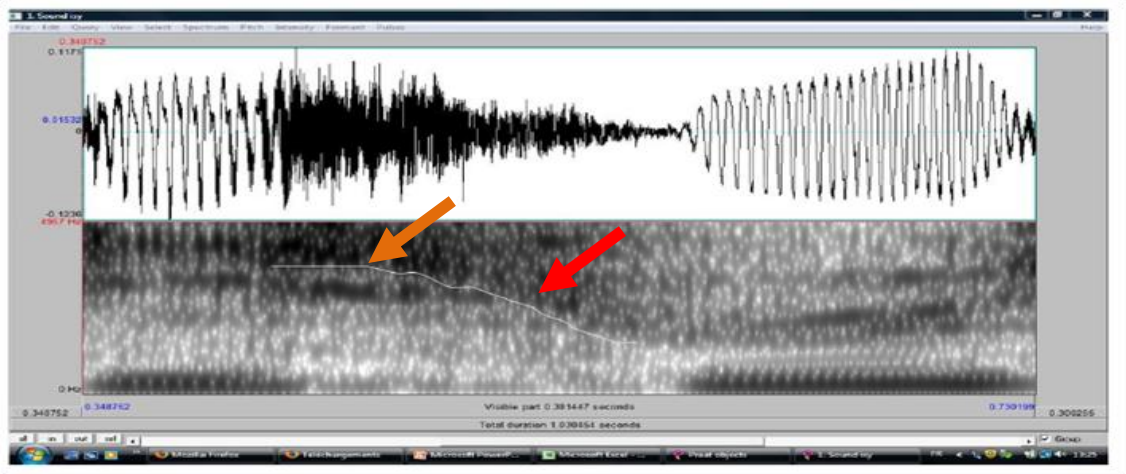

Figure 3. Acoustic signal and spectrogram for the [isy] sequence. The inferior limit of the frication noise falls in two times: first, a slight decrease (orange arrow) and, secondly, a more remarkable inflexion (red arrow)

Concerning perception of vowel [ø], results are similar, since non-sighted people perceive this vowel earlier than the control group (Figure 3): 70% of the blind people group recognize the last rounded vowel cited (cl= 2) at 30 ms from onset of the clear formant structure of the rounded vowel, while control subjects do not recognize the rounded vowel before its acoustic onset.

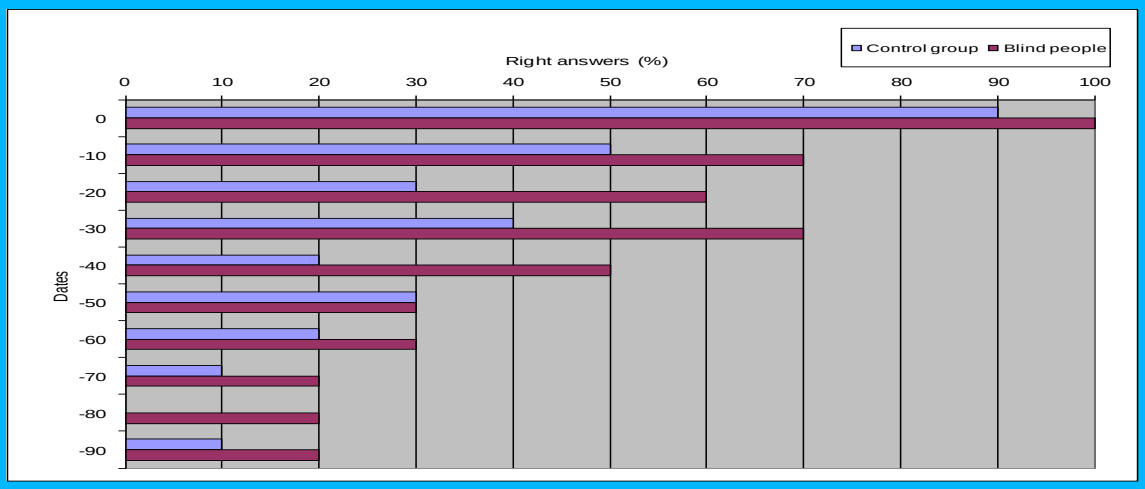

Figure 4. Comparison of correct responses of sighted and blind people for the [isø] sequence. The *x* axis shows correct percent responses, and the *y* axis gives truncation dates.

Contrary to the [isy] sequence, no noticeable inflexion was visible in the ILFN. Indeed, it slightly diminishes from [i] to to [Vlab], since it was measured at 4027 Hz at the first date after the [i] and at 3277 Hz just before onset the [ø].

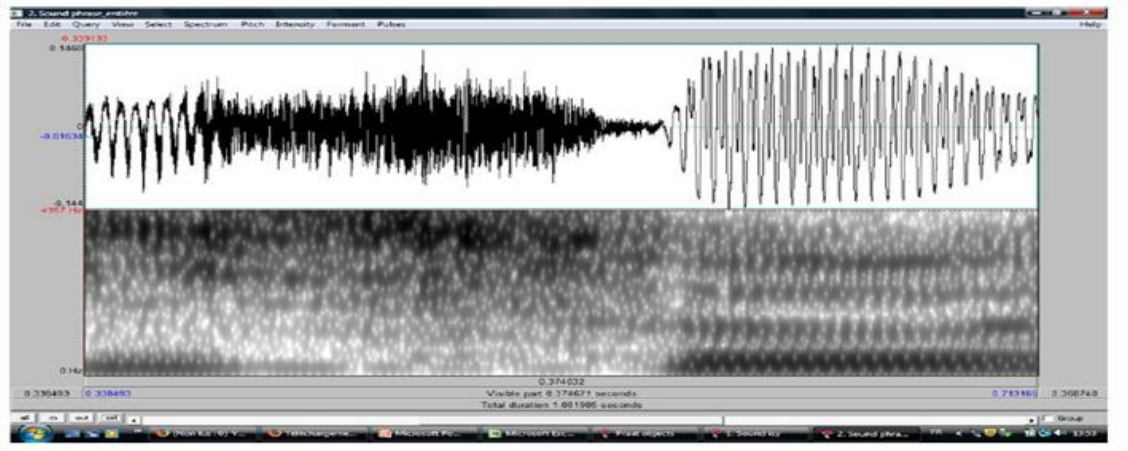

Figure 5. The acoustic signal and spectrogram for the [isø] sequence. The inferior limit of the frication noise slightly decreases from vowel [i] to vowel [ø].

To summarise, [y] is perceived earlier than [ø]in the obstruent interval, by both control and blind subjects. We obtain comparable results for the backward round vowels tested: [u] is identified before [o]. When data for the control group and the group of blind subjects are compared, we observe that blind subjects perceive rounded vowels earlier than control sighted subjects.

If differences temporal between non-sighted and sighted people have been observed in speech perception, it is hypothesised that related timing differences may also appear in the speech production domain. This issue is addressed below.

3.2. Speech production tests

Analyses of variance (ANOVA) were carried out and indeed reveals group effects for the temporal data ($p<0.006819$).

Results for temporal organisation of [CVC] sequences show differences between blind people and the control group. Figure 6 shows the percentage of parameters Acoustic Silence (for C1), Voice Onset Time, Vowel duration, Voice Termination Time and Consonantal duration of C2 in a [tap] sequence. We can notice here that the temporal organisation of these cues is different for the two groups.

Indeed, silence for C1 constitutes 34% (SD = 2%) of the [CVC] sequence for the control group, while the same interval is quantified at 27% (SD = 4%) of [tap] for the blind people.

Voice Onset Time was measured at 6% (SD = 1%) for the sighted subjects, and at 12% (SD = 2%) for the non-sighted group.

Vowel [a] was evaluated at 28% (SD = 2%) of the sequence for the control subjects and at 20% (SD = 4%) of the same sequence by blind people.

Voice termination Time constitutes 7% (SD = 2%) of [tap] for the sighted people; the same parameter being at 13% (SD = 3%) for the non-sighted group.

Concerning the second consonant of the sequence, it was quantified at 25% (SD = 1%) of the sequence of the control group, and at 28% (SD = 2%) of the blind people.

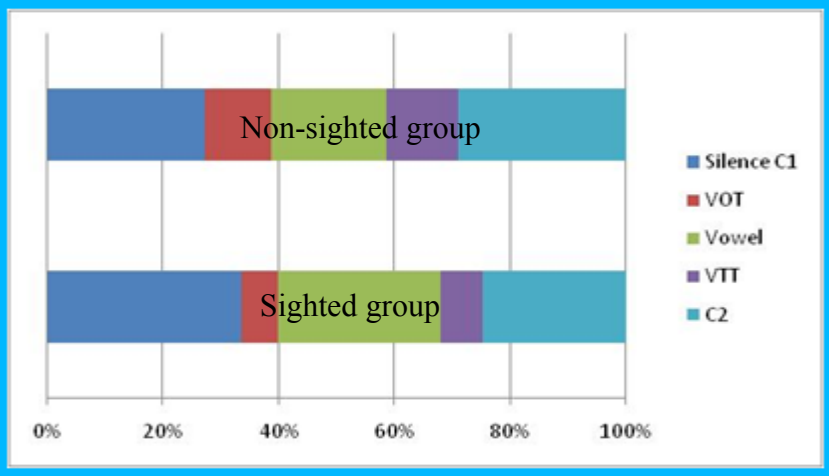

Figure 6. *Average values (in %) for Acoustic Silence of C1, VOT, Vowel Duration, Voice Termination Time and C2 for the non-sighted and the sighted groups in [tap] sequences.*

4. Discussion and conclusion

4.1. Speech perception tests

Acoustic and perceptual studies of [isVlab] sequences reveal two phenomena, one for vowels which are highly protruded and another one for less protruded vowels. For the first category ([y] and [u]), the inferior limit of the frication noise slightly decreases from offset of vowel [i], and then shows a much more steep inflexion. The date of the inflexion corresponds more or less with the moment where sighted people start to perceive the rounded vowel. This result could be compared with those of Sock *et al*. (2011), presented at the present International Seminar on Speech Production, who notice that the date where listeners perceive the rounded vowel [y] coincides with the acceleration peak of the lips gestures, at the articulatory level. Concerning the less protrudes vowel ([o] and [ø]), the ILFN decreases slowly during the [s]; no inflexion was observed. This result could be explained by the fact that a diminution of protrusion would provoke modifications in kinematic parameters. In other words, the acceleration peak would arrive later; consequently, it would not provoke any inflexion.

At the beginning of this work, it was hypothesized that blind subjects would perceive a protruded vowel earlier than unimpaired subjects, and in a more robust manner (following a specific confidence threshold) in [V1sV2] sequences. Results confirm this initial hypothesis, since each rounded vowel was identified earlier by blind people compared to control subjects. According to the literature, results could be explained by the fact that non-sighted people develop auditory compensation allowing them to perceive more precise changes in the inferior limit of the frication noise.

4.2. Temporal organization

Analyses of the temporal organisation of the speech of blind people and of that of the control group reveal differences in the way timing is implemented in the production

of CVC sequences. Differences between the two groups could be explained by the fact that blind people, who acquire speech largely through the auditory, imitate more or less precisely what they hear. So, differences could emerge in acquisition of timing, which seems to persist in the speech of blind people. More precisely, jaw and tongue gestures are tightly coupled in speech production (Vaxelaire *et al*., 2010). It could be posited that jaw opening, which should be acquired according to visual cues, is not controlled in the same way by blind people and control subjects, thus provoking differences in temporal organisation between the two groups. Articulatory data would allow us to confirm or infirm this hypothesis.

References

Cavé, C. Sato, M., Ménard, L. and Brasseur, A. Interactions audio-tactiles et perception de la parole : comparaisons entre sujets aveugles et voyants. Proceedings of the XXVIIIth Journées d'Etudes sur la Parole, Mons, CD-Rom, 2010.

Hirsch, F. Dreyfus, H. Sock, R. Vaxelaire, B. Fauth, C. Bouarourou, F. and Béchet, M. Etude préliminaire de la perception précoce des voyelles labialisées par des auditeurs déficients visuels. Proceedings of the XXVIIIth Journées d'Etudes sur la Parole, Mons, CD-Rom, 2010.

Menard, L. Dupont, S. Baum, S.R and Aubin, J. Production and perception of French vowels, by congenitally blind adults and sighted adults. Journal of Acoustical Society of America, volume 126 (3), 1406-1414, 2009.

Moos, A. Hertrich, I. Dietrich S. Trouvain, J. and Ackermann H. Perception of ultra-fast speech by a blind listener – Does he use his visual system? Proceedings of the International Speech Production Seminar, Strasbourg, 297-300, 2008.

Sato, M. Cavé, C. Ménard, L. and Brasseur, A. Auditory-tactile speech perception in congenitally blind and sighted adults. Neuropsychologia, 48(12), 3683-3686, 2010.

Sock, R. Vaxelaire, B. Hirsch, F. Ferbach-Hecker, V. Roy, JP. Bouarourou, F., Béchet, M. and Fauth, C. (2011) Presenting the Anticipatory Perception based on Events (APE) hypothesis. Proceedings of the International Speech Production Seminar, Montreal, 2011.

Sock, R. *Organisation Temporelle en Production de la Parole. Emergence de Catégories Sensori-Motrices Phonétiques*. Thèse de Doctorat d'Etat, Grenoble III, 1998.

Vaxelaire B., Fauth C., Hirsch F., Bouarourou F., Sock R. (2010) Anticipatory and sustained coupling in jaw and tongue gestures in VCV sequences. An X-ray study. Poster presented during the International Summerschool "Cognitive and Physical Models of Speech Production,Speech Perception and Production-Perception Interaction", 2010.