PERCEIVING ANTICIPATORY PHONETIC GESTURES IN FRENCH

Béatrice VAXELAIRE, Rudolph SOCK and Véronique HECKER
Institut de Phonétique de Strasbourg - Université Marc Bloch
Institut de la Communication Parlée de Grenoble, Université Stendhal-INPG
Laboratoire de Psychobiologie du Comportement Moteur et des Sports -Université Marc Bloch

ABSTRACT
Positing that listeners exploit early motoric cues and vocal tract shapes related to upcoming speech elements, this research examines the perceptual effects of anticipatory gestures and vocal tract configurations in the production of a French rounded vowel-like consonant: the so-called semivowel [Â].

1. INTRODUCTION

1This research, closely following that carried out for Swedish [1], presents results based on X-ray and acoustic data regarding three precise questions: (1) Do anticipatory labial and tongue gestures contribute to the auditory perception of the French rounded semivowel [Â]; (2) If such is the case, what are the perceptual effects and temporal extent of these gestures and the resulting vocal tract shapes; (3) How does speech rate affect the perceptual extension of these anticipatory gestures and vocal tract configurations? Previous findings, reported in two companion papers [2, 3], have shown timing constraints of anticipatory gestures in the production of French rounded vowels. Results obtained here will be compared with such findings on both the articulatory and auditory levels.

2. EXPERIMENT I: ANTICIPATORY COARTICULATION

2.1 Method
2.1.1. Speaker, speech sample and data acquisition
The speaker was an adult native male speaker of French (Alex) and had no speech or hearing defects. Of the two speakers, Alex showed more perceptible lip and jaw movements, accompanied with a clearer formant structure of the target semivowel. The speech sample consisted of the sentence "Les truites glacées" that provides the core sequence V1 + Cn + V2, where V1 is the unrounded vowel [e], Cn is the non-labial cluster [Ą] and V2 the rounded semivowel [Â].

This sentence was extracted from a corpus of 58 short sentences that embedded the target words, uttered by two speakers, at two speaking rates: normal and fast [4].

X-rays and a simultaneous audio recording of the speakers’ productions were obtained under medical care, with the help of a 35 mm camera, a video-camera, a stereo recorder and a highly directive microphone. The acoustic signal was recorded on track n° 1 of the stereo recorder, and each frame, appearing as a coding impulse in the form of a vertical bar, was recorded on track n° 2 of this recorder. The digitized acoustic signal was then visualized synchronous with the result of the processed frame-by-frame X-ray images (50 images per second), with the vertical bars appearing on an inferior line, below the digitized signal. An enlarged speech waveform display, with the corresponding spectrogramme, was made of the target utterance in both speech rates, to ensure that the [Â] was indeed a devoiced dorso-uvular constricture, without any formant structure that may, unwantedly, enhance labial anticipation.

With the help of a grid, measurement parameters for vocal tract configurations were determined on mid-sagittal profiles. Temporal events were detected on the audio signal and specific timing relations between these events allowed determining, in the VCV domain, acoustic durations that correspond to articulatory opening and closing gestures.

2.2. Timing of articulatory gestures
The articulators monitored for anticipatory behaviour were: upper lip protrusion, lower lip vertical displacement, lip opening, and tongue body vertical displacement. They had proven to be relevant in portraying anticipatory rounding behaviour in a previous study, reported in one of the companion papers [2].

The data show (Figure 1), in normal speech, that a certain amount of upper lip protrusion is already present in the configurations corresponding to the production of the unrounded vowel [e] (frames 798 - 803). However, remarkable onset of protrusion coincides with the beginning of apical contact (frame 802) and continues its forward movement throughout the obstructive phase of the apical consonant [t]. Maximum protrusion is attained at occlusion release (frame 806), thus amounting to a variation of 8 mm in the displacement of the upper lip, from the unrounded vowel configurations up to the end of the apical obstructive phase. A noticeable reduction in lip protrusion (5 mm) occurs during the [Â] configurations, and up to the early configurations of the semivowel [Â] (frames 807 - 811). Such a reduction in protrusion is very certainly due to the influence of the subsequent spread vowel [i]. Nevertheless, configurations corresponding to the semivowel still show rounding characteristics. The vertical displacement of the lower lip, contributing also to the rounding gesture, begins its upward trajectory (frames 802 - 805) during the obstructive phase of the apical consonant, to attain its maximum value well after release (frame 808), in the dorso-uvular [Â] configurations.

Lower lip displacement rate is 4 mm. In comparison with the behaviour of lip protrusion, lower lip vertical displacement also undergoes reduction in movement amplitude (5 mm) in the vicinities of the [Â] configurations, a reduction attributed, here again, to the influences from the spread vowel [i]. However, this lowering in the trajectory of the lower lip does not compromise the rounded nature of the semivowel. Lip opening is negatively correlated with upper lip protrusion. It reduces during the obstructive phase of the apical consonant and attains its minimum value well after occlusion release (frames...
3. EXPERIMENT II: PERCEPTUAL EFFECTS OF ANTICIPATORY COARTICULATION

3.1. Gating procedures and test tape elaboration

Gatings were made on the two digitally stored acoustic signals, one for the normal speech rate and the other for the fast rate. Decisions were made with reference to the vertical bars appearing below the digitized speech signal and representing X-ray images of the vocal tract. Initial gatings were carried out under auditory control to determine, in normal speech, the most extreme gating for vowel identification. Six gatings were made, as from this extreme gating, in 20 ms steps, towards the target semivowel. Thus the six gatings, located at different temporal distances from the acoustic onset of the semivowel were the following in the two speech rates: Gate 1 was at 100 ms; Gate 2, at 80 ms; Gate 3 at 60 ms; Gate 4, at 40 ms; Gate 5 at 20 ms; and Gate 6 corresponded to onset of the clear formant structure of the semivowel. Thus each “gated” sentence consisted of the sequence [le... plus an increasing amount of the acoustic information from the cluster [tA] preceding the rounded semivowel [A]. The test tape contained 12 gated out stimuli (6 in each speech rate) that were then randomized twice, one series in normal speech and the other in fast speech, resulting in two files containing 12 stimuli each. A dip, serving to alert the listeners as to the imminence of a stimulus, preceded each utterance by 1.4 seconds. The interstimulus interval was 4 seconds with a 10 second pause following each twelfth item. The test tape began with a training list of 12 items, then continued with the 24 stimuli.

3.2. Listener judgements

Thirteen adults, all native speakers of French, served as listeners for the perceptual experiment. They were all naive with regards to the purpose of the tests, and had normal speech and hearing. The tests were carried out in a sound-treated room at the Phonetics Institut of Strasbourg, designed for group listening experiments. The subjects were told that they were going to hear one of the following three truncated French sentences: (1) “Les traitres glacés”; (2) “Les trappes glacées”; (3) “Les truites glacées”. However, in this particular experiment, only the target sentence n° 3 was on the test tape, the other two only serving as distractors. Nevertheless, results from a similar experiment show that when distractor sentences were effectively included in the test tape, no significant change was observed in the results [5]. Moreover, subjects were told that sentence n° 1 was pronounced with a rather close-mid quality, quite frequent in the region where the experiment was carried out. The test tape was played simultaneously to all thirteen subjects. Score sheets were provided, and during the 4 second interval between test stimuli, the subjects had to perform two tasks for each of the 24 utterances: (1) mark with a cross which of the three “vowels” [i, e, Â] he or she believed to have been “gated out” of the particular test stimuli, and (2) check on a 5-point scale the level of confidence with which they were making each judgement, with “1” indicating very little confidence and “5” indicating relative certainty.

3.3. Perceptual effects of anticipatory gestures

The mean level of confidence rating, averaged over all 24 utterances by all 13 listeners (totaling 312 responses) was 2.1 with a standard deviation of 0.8, thus showing that overall judgments were made with a fairly good degree of confidence, in both speech rates. Moreover, the percentage of correct responses was highly correlated with confidence ratings (r=0.76 in normal speech and r=0.87 in fast speech), indicating that while being confident subjects were at the same time being highly performant, with regards to the identification task required of them (Figure 2). Also, subjects seemed to be quite severe with the use of confidence ratings on the subjective 5-point scale, as confidence points “2” and “3” revealed high correct identification scores in both speech rates, with the curve attaining ceiling values between points “3” and “5”. On a whole, subjects were more confident in the normal speaking condition. Further, as shown in Figure 3, confidence ratings drop off progressively as the gatings move away from the target semivowel, i.e. as the amount of available sensory information decreases (r=0.92 and r=0.95 in normal and fast speech respectively). Nevertheless, if the suitable model is that of a hyperbola in normal speech, with decrease in confidence ratings being less sensitive as gatings are temporally distant from the onset of the semivowel, it is rather of a linear regression in fast speech, as confidence ratings become quite sensitive to the distance from the vowel. When very unsure as to the identity of the gated out vowel, listeners largely used the vowel [e] as a “waste-basket” response, since the percentage of [e] responses remarkably exceeds that of the [a] responses at the most extreme gating (Gate 1). This may be explained in terms of proximity between [e] and [Â] in the acoustic space, with regards to [a].

Results given in Figure 4 are in line with what has previously been reported in the literature, as to listeners being able to guess the identity of vowels when those vowels had been gated-out of the acoustic signal [1, 3]. The percent correct identification is high very close to the semivowel and decreases as gatings moves away from this target semivowel (r=0.85 and r=0.94 in normal and fast speech respectively). Whilst correct identification decreases gradually as gatings
move away from the rounded semivowel in normal speech, percent correct identification is much more sensitive to the distance from the semivowel in fast speech, as identification scores are rather high close to the onset of the target semivowel, but fall of drastically as soon as the gating becomes more temporally distant. This could be explained with regards to the above mentioned abrupt changes in vocal tract configurations in fast speech. In normal speech, 70% of correct identification still occurs at the extreme gating, Gate 1, i.e. 100 ms before the acoustic onset of the rounded semivowel. It should be noted, however, that confidence ratings are quite low at this gating (1.4). Before this gating, identification was simply impossible for all listeners; 100 ms from the onset of the rounded semivowel, thus corresponds to a critical perceptual frontier for identification. From Gates 2 to 6, correct responses that are between 85% and 100% are backed by good confidence ratings. The perceptual effect of this anticipatory movement is reduced in fast speech, as 77% of correct identification, supported by a fairly good confidence rating (2) is only obtained at Gate 4, located at 40 ms before the semivowel. In fact the percentage of confidence rating does not change at Gate 5 (77% at 20 ms from the semivowel) but only when onset of a clear formant structure for the rounded semivowel has been attained (i.e. at Gate 6), where identification scores rise to 92% of correct responses. The perceptual breaking point is thus located between Gates 3 and 4, i.e. between 60 ms and 40 ms from the semivowel, in fast speech.

4. SENSORY-MOTOR RELATIONS

In summary, these results (identification scores and confidence ratings) show that configurations located before peak protrusion do not seem to have any direct anticipatory auditory effect (Figure 1, frames 798 - 806). In other terms, if the rounding gesture, anticipated well before the onset of the semivowel corresponds to visible anticipatory configurations, the robust components of the audible anticipatory labial gesture are located after peak protrusion, in both speech rates (as from frames 806-807 in Figure 1). Indeed, auditory breaking points — Gates 1 and 4 in normal and fast speech respectively — correspond to the onset (+) of the unrounding façade of the protrusion gesture (sic). However, these auditory breaking points coincide, more or less, with tongue tip release of the apical consonant and tongue body raising, in the hard palate region, for vowel constriction formation. Thus listeners, despite the relative unrounding gesture, seem to be hearing anticipatory vocal tract shapes, resulting especially from the combined effect of tongue tip release and tongue body constriction formation (see Figures 5 and 6).

5. CONCLUSIONS

Kinematic and acoustic results obtained recently in a similar experimental design [3] showed, for vowel [y], that the protrusion gesture became audible as from Peak Velocity (80 ms and 74 ms before the onset of the vowel in normal and fast speech respectively), when high identification scores (82% on average) are backed by strong confidence ratings in both speech rates. It was thus the kinematic span Peak Velocity-Maximum Protrusion that seemed to constitute the anticipatory acoustic efficient portion of the entire protrusion gesture (without even considering the efficient carryover façade during unrounding for the subsequent consonant). These differences in results may be attributed to several factors: 1) the nature of the two sounds, one being the highly rounded French vowel [y], and the other being a vowel-like consonant [Â], with a less stable and clear formant structure; 2) the nature of the subsequent sound: in the [y] experiment, this vowel was followed by the lateral constrictive [l] that did not seem to hinder the emergence of the protrusion gesture, whereas in the [Â] experiment, the subsequent spread vowel [i] did contribute largely to the partial and early unrounding of the semivowel; 3) while the obstruent interval was longer in the previous experiment (a string of five consonants), comprised of two occurrences of the fricative [s] — known to be quite favourable for extension of the perceptual effects of rounding [5] — the obstruent interval in this experiment only contained the cluster [tÂ], with the auditory inhibitory effect of the plosive [5] being quite remarkable.

To sum up, these results, together with previous findings, suggest the following: 1) extents of (visible) measured anticipatory gestures are longer than that of their efficient auditory portions [6]; 2) anticipatory gestures, contributing to the emergence of specific vocal tract shapes, are strongly vowel-dependent [1]; 3) the auditory anticipatory extent of a given gesture, like its articulatory anticipatory extent, is necessarily tied to the nature of the surrounding sounds and to the length of the preceding consonant string. More data is being analyzed for the second speaker, to test for eventual adjustments of listeners to speaker-specific strategies and vocal tract shapes. Future research will also be carried out on the different French rounded vowels and rounded vowel-like consonants [Â] and [w], in controlled segmental contexts, in order to obtain more readily comparable sensory-motor anticipatory extents.

NOTE

1. In loving memory of Christian Benoît.

REFERENCES

Figure 1. Frame-by-frame analysis in normal speech. [et%
µ]

Figure 2. Mean values of percept correct responses in relation to confidence level.

Figure 3. Mean values of confidence ratings at different gatings.

Figure 4. Mean values of correct responses at different gatings.

Figure 5. Audible anticipatory vocal tract shape.
Normal speech, frame 808.

Figure 6. Audible anticipatory vocal tract shape.
Fast speech.