Perception of vowel nasalization in French

Article *in* Journal of Phonetics · September 1981

DOI: 10.1016/S0095-4470(19)30974-X

André-Pierre Benguerel
University of British Columbia - Vancouver

Journal of Phonetics (1981) 9, 309–321

Perception of vowel nasalization in French

André-Pierre Benguerel and André Lafargue
The Phonetics Laboratory, School of Audiology and Speech Sciences, University of British Columbia, Vancouver, British Columbia, Canada

Received 2nd July 1980

Abstract: Previous studies have shown that a nasalized vowel in a non-nasalized context may be produced with the velopharyngeal port closed for part of the duration of the vowel, and similarly, that an oral vowel in a nasalized context may be produced with the velopharyngeal port open for part of the duration of the vowel. Single cycles were spliced out of vowels of both types at times corresponding to various and known velar heights. They were used to construct the stimuli of two listening tests. Each test also contained non-manipulated reference vowels (oral and nasalized) as controls to evaluate the listeners' responses. Native speakers of French were asked to rate, on a 5-point scale, the degree of nasality of each stimulus. A definite relationship between the listeners' judgement of nasality and velar height was found. Results also suggest that a large velopharyngeal cross-section may be necessary for a French listener to perceive an open vowel such as [ɑ] as nasalized, whereas a smaller velopharyngeal cross-section may be sufficient to give a close vowel such as [ɛ] and [ɔ] a nasalized quality. It also appears that the listeners perform some kind of time integration of the nasality function in order to judge whether a vowel is nasalized or not.

Introduction

Any theory of speech production has to contend, at some point, with the transition from discrete phonological units or categories to continuously varying phonetic dimensions. Similarly, any theory of speech perception has to contend with the inverse transition, namely from the phonetic continuum to the discreteness of the phonological level. The use of a certain phonetic dimension in a given language may or may not reflect a corresponding contrast in the phonology of this language; for certain languages in fact, it may be the case in some instances, and it may not in other instances. In many languages, a given phonetic feature (e.g. vowel nasalization, in English) does not reflect a phonological contrast, even though, due to coarticulation for example, it is often present to some degree in the vicinity of certain segments. In other languages, this same phonetic feature may also mark a phonological contrast (e.g. vowel nasalization, in French); in these instances where it does, it is clearly not due only to coarticulatory influences; however it is also possible for the feature to be present to some degree, due to coarticulation for example, while the segment in question retains its phonologically unmarked categorization (non-nasalized in our example). It is essential, therefore, for the speaker–listener of such a language to differentiate, both in production and in perception, between two cases: (1) when the phonetic feature is simply a byproduct of the speech production process but is not meant to signal a phonological contrast, (2) when this same feature is meant to signal a phonological contrast.

0095–4470/81/030309+13 $02.00/0

If the phonetic feature is bipolar and can be assumed to vary monotonically between its two poles, one may expect the speaker–listener to have a threshold (or a "twilight" zone) below which the corresponding phonological feature has one value and above which it will have the other value.

The aims of this study are to investigate whether such a threshold can be determined experimentally, and whether it is speaker–listener independent and/or language independent. Identical labels are used to describe phonetic facts in different languages, but there is evidence that it is at best a similarity, not an identity: the fact that two languages label a phonetic feature with the same name does not imply that their respective speaker–listeners evaluate this feature in the same way. It is even probable that two speaker–listeners of the same language do not evaluate a given phonetic feature in exactly the same way. Ladefoged (1978, pp. 38–40) wrote:

> From an acoustic or a physiological point of view most phonological features are cover features, definable only in terms of complexes of phonetic parameters . . . when we are trying to compare languages or to give accurate phonetic descriptions of a single language, then we must learn to interpret phonological oppositions in terms of complexes of the real phonetic elements of the language.

One may add that Ladefoged's above statement most likely holds from the perceptual point of view as well.

In a previous study investigating velar movements in French (Benguerel *et al.*, 1976*a*), it was shown that a nasalized vowel (e.g. /ã/) in a non-nasal context (e.g. /tãt/) was produced with the velopharyngeal port closed for part of the duration of the vowel. By nasalized is meant "a configuration of the vocal tract in which there is an intermediate velopharyngeal opening and in which there is not an oral closure" (Peterson & Shoup, 1966, p. 60); by oral (or non-nasal) is meant "a configuration of the vocal tract in which there is a velopharyngeal closure" (ibid., p. 60). It was shown, similarly, that an oral vowel (e.g. /a/) in a nasal context (e.g. /an-/) was produced with the velopharyngeal port open for part of the duration of the vowel. It is important to observe that in the first instance (i.e. /tãt/), any native listener perceives a phonemically nasalized vowel, whereas in the second instance (i.e. /an-/), he perceives a phonemically oral vowel. In other words, phonemically oral vowels may be uttered with the velopharyngeal port partially open during their production, whereas phonemically nasalized vowels may be uttered with the velopharyngeal port partially closed. It thus seems natural to ask with how much velar opening an oral vowel can be produced and still be perceived as phonemically oral, and conversely with how little velar opening a nasalized vowel can be produced and still be perceived as phonemically nasalized. Since the velar position may change rapidly during the course of the vowel from closed to open, from open to closed, or from one position to the other and back, if one assumes that (perceived) nasalization is a function of velar opening, it seems reasonable to conclude that some kind of integration process over the duration of the vowel must be performed by the listener.

The aim of this study is to try to answer the above questions by constructing stimuli in which velar opening during production is known and varies systematically from one stimulus to the other, and by asking listeners to evaluate their degree of nasality. House & Stevens (1956) studied the perception of nasality by using stimuli produced on a line analog model with an adjustable degree of coupling between oral and nasal vocal tracts. The subjects they used, however, were speakers of American English, language in which nasalization of vowels is not phonemic. For the present experiment, it was necessary to use listeners of a language in which oral and nasalized vowels have a comparable status, and French was the language chosen.

Experimental procedure

Stimuli

The vowel stimuli used for the two perceptual tests to be described were of the following types:

I:	3 (reference) oral vowels:	ɑ, ɛ, ɔ
II:	3 (reference) nasalized vowels:	ɑ̃, ɛ̃, ɔ̃
III:	15 (test) nasalized vowels:	$\tilde{ɑ}_1$, $\tilde{ɑ}_2$, $\tilde{ɑ}_3$, $\tilde{ɑ}_4$, $\tilde{ɑ}_5$, $\tilde{ɛ}_1$, $\tilde{ɛ}_2$, $\tilde{ɛ}_3$, $\tilde{ɛ}_4$, $\tilde{ɛ}_5$, $\tilde{ɔ}_1$, $\tilde{ɔ}_2$, $\tilde{ɔ}_3$, $\tilde{ɔ}_4$, $\tilde{ɔ}_5$
IV:	9 (test) oral vowels:	$ɑ_1$, $ɑ_2$, $ɑ_3$, $ɛ_1$, $ɛ_2$, $ɛ_3$, $ɔ_1$, $ɔ_2$, $ɔ_3$.

Items of types I, II and III were used to construct Test 1 while items of types I, II and IV were used for Test 2. The 3 (reference) oral vowels were spliced out of the 3 words /tat/, /tɛt/, /tɔt/. The 3 (reference) nasalized vowels were spliced out of the 3 words /tɑt/, /tɛ̃t/, /tɔ̃t/. The 15 (test) nasalized vowels were constructed from 5 equally spaced single periods isolated in the vowel of each of the 3 words /ɑnɑ/, /ɛnɛ/, /ɔnɔ/ whereas the 9 (test) oral vowels were constructed from 3 equally spaced single periods isolated in the initial vowel of each of the 3 words /ɑnɑ/, /ɛnɛ/, /ɔnɔ/.

The words /tɑ̃t/, /tɛ̃t/, /tɔ̃t/ from which the stimuli of type III were constructed, were taken from the original recording of the experiments by Benguerel *et al.* (1976*a*, *b*). They had been originally produced in isolation by the first author, a native speaker of French from Lausanne, Switzerland. After digitization, these original recordings were observed on a computer scope screen. The experimenter was able to determine the fundamental frequency for each utterance by counting the number of samples in each cycle of the vowel.

A pilot experiment had shown that fundamental frequency, intensity and duration of the stimuli could influence the listeners' judgement by providing spurious identification cues. It was therefore decided to use stimuli where no variations of these parameters could be used as cues by the listeners. The /ɑnɑ/, /ɛnɛ/, /ɔnɔ/ utterances were found to have a lower fundamental frequency (130 Hz) than the /tɑ̃t/, /tɛ̃t/, /tɔ̃t/ utterances (143 Hz). It was decided to keep the original recordings of /tɑ̃t/, /tɛ̃t/, /tɔ̃t/ to construct the type III test stimuli and to re-record a series of /ɑnɑ/, /ɛnɛ/, /ɔnɔ/ utterances for the type IV test stimuli. Since the /tɑ̃t/, /tɛ̃t/, /tɔ̃t/ were too short to construct reference stimuli, the same speaker recorded the 9 utterances /tɑt/, /tɛt/, /tɔt/, /tɑ̃t/, /tɛ̃t/, /tɔ̃t/, /ɑnɑ/, /ɛnɛ/, /ɔnɔ/, from which the reference stimuli were obtained.

The utterances mentioned above were recorded on an audio tape using a Revox A 77 tape recorder and an Altec 681A LO microphone. A pure tone of 143 Hz, monitored with a calibrated frequency counter (Monsanto 110B), was presented via headphones to the speaker who attempted to phonate at the frequency he heard. The speaker also prolonged his /tɑt/, /tɛt/, /tɔt/, /tɑ̃t/, /tɛ̃t/, /tɔ̃t/ utterances.

Narrow band spectrograms of all the re-recorded utterances were made to determine the sequences with the steadiest fundamental frequency. Eventually, twelve words with a fundamental frequency between 142 and 146 Hz were selected to construct the test stimuli.

Vowel editing

The twelve selected utterances were digitized and stored on a PDP-12 computer, using a set of programs developed by Lloyd Rice at UCLA. The audio tape containing the selected utterances was played at half speed. The signal was sent to the computer via a Rockland LP-filter with a 48 db/octave attenuation slope and the cut-off frequency set at 6 kHz. The result of this operation was equivalent to the digitization of the utterance at a sample rate of 12,000 samples/s, when played at regular speed. The volume control of the tape recorder was adjusted for each utterance so that all the stimuli words were digitized at approximately the same peak intensity. The speech signal stored on the tape was then displayed on the oscilloscope screen. Single cycles, containing as much as possible the same number of samples, and taken at five different places in the /tɑ̃t/, /tɛ̃t/, /tɔ̃t/ utterances taken from the data of Benguerel *et al.* (1976*a*, *b*) and at three different places in the re-recorded /ɑnɑ/, /ɛnɛ/, /ɔnɔ/ utterances, were isolated and stored on digital tape. Intervals between the selected cycles were approximately the same. Edited cycles were always taken between two zero-crossing of the signal wave. Each of the single isolated cycles was then reproduced 40 times and stored on digital tape. The signals so produced were long enough (approximately 280 ms) to be identified by the listeners. Portions of vowels taken from the re-recorded /tɑt/, /tɛt/, /tɔt/, /tɑ̃t/, /tɛ̃t/, /tɔ̃t/ utterances, having a length of 40 cycles were isolated and stored on the computer tape. The envelope of the signals so produced had a rectangular shape. They were given a trapezoidal shape to produce a smooth onset and offset of 25 ms each.

The digitized signals were reconverted into analog signals and recorded on an audio tape via a low pass filter which was set at 6000 Hz to eliminate high frequency digital noise generated by the computer. Mingograms, including a speech power display, were made of the vowel stimuli obtained and differences in vowel intensity were found to be within ± 1 dB of each other.

Tests preparation

Two tests (Test 1, Test 2) were constructed. Both tests contained the six reference vowel stimuli ɑ, ɛ, ɔ, ɑ̃, ɛ̃, ɔ̃. This was done to evaluate the listeners' subjective judgement of the degree of nasality attributed to unmodified vowels and to help establish a base line for the evaluation of the listeners' responses to the vowels reconstructed from single cycles.

The items of Test 1 consisted of five presentations of each of the six reference vowel stimuli and of the fifteen reconstructed vowels from the original /tɑ̃t/, /tɛ̃t/, /tɔ̃t/ utterances. The test stimuli were recorded in randomized order on one track of a stereo tape. No two stimuli originating from the same utterance were juxtaposed. Ten "buffer" vowels, selected from the 105 test stimuli were recorded, five at the beginning, and five at the end of the tape.

For Test 2, the test items consisted of five presentations of each of the six reference vowel stimuli and of the nine reconstructed vowels from the re-recorded /ɑnɑ/, /ɛnɛ/, /ɔnɔ/ utterances. A total of 85 vowel stimuli were recorded.

In both tests a silence of 5 s was inserted between the end of one stimulus and the beginning of the next. In the 5 s preceding each stimulus, French numbers from 1 to 115 and from 1 to 85, for Test 1 and Test 2 respectively, were recorded at low intensity level on the other track of the stereo tape. Only items 6 to 110 (Test 1) and 6 to 80 (Test 2) were considered in the analysis.

Subjects

Twenty-seven native speakers of French were used as subjects. Their ages varied between 20 and 45. They had been in North America for periods ranging from a few days to eighteen years. Most regions of France, including St. Pierre and Miquelon (three subjects) were represented. Two subjects were born in Switzerland, another subject was born in Southern Belgium and one was born in Val d'Aosta in Northern Italy, and was bilingual. One of the

French-born subjects (aged 22) had been living in Canada for most of his life but the language spoken at home was French. All but one of the subjects had a working knowledge of English; eight subjects had no knowledge of phonetics, while ten had had phonetic training and fifteen were teachers of the French language.

The subjects reported no history of deafness or hearing loss. Two subjects who were concerned about a possible loss were given a thorough audiological examination. Their hearing was found to be well within limits of normal range.

Test procedure and equipment

The subject was seated in a quiet room with an ambiant noise of about 35 dBA. The test tape was played on a Revox A 77 tape recorder through a set of GSC headphones model TDH 39-300Z at a level corresponding to 70–80 dB SPL as measured through the same headphones with a B & K model 2203 sound level meter, a B & K model 4152 artificial ear and a 6 cm^3 coupler.

Subjects were instructed, in French, to assess the degree of nasality of each vowel stimulus on a 5-point scale. In addition, it was pointed out to the subjects that the tests did not contain any specific number of oral or nasalized vowels. They were encouraged to note down what they actually heard, and not to assume any particular distribution. For example, if they thought they heard every stimulus as nasalized they were instructed to check the appropriate column all the time.

The first five items were played to them. The tape was stopped and the subjects were asked if they had any questions. The tape was then reset at the first item and formal testing began. Fifteen subjects were presented with Test 1 first and the remaining twelve heard Test 2 first.

Results

Data analysis

In order to analyze the results of the perception tests, listeners' responses were assigned a number from 1 (oral) to 5 (nasalized) corresponding to the number of the column in which they had registered their answers. Listeners' responses were then de-randomized, sorted, and stored on a digital tape for further analysis. The mean and standard deviation of each listeners' responses to the last four presentations (m_4, s_4) of each stimulus were computed and listed, as in Table I. Also in each case and for each test, the average standard deviations ($\bar{s}_5$, $\bar{s}_4$) were computed. Table I is an example of the sorted data for subject number 10; it represents her responses to Test 1 and Test 2, sorted and analyzed as described above. Table I suggests that subject number 10 might have experienced some difficulty in assessing a degree of nasality to some of the stimuli the first time they were presented. It can also be noted that $\bar{s}_4$ is smaller than $\bar{s}_5$ in Test 1. The same observation was made for most of the subjects and in both tests. This suggests that the subjects' responses to the last four presentations of each stimulus are more representative of their judgement after adaptation to the test items than are their responses to all five presentations.

Normalization

Some listeners, consistently, did not use the extremes of the five-point scale while responding to presentations of the reference vowels. Therefore, their five/four response scores were distant from the extreme scores (1 or 5) for both oral and nasalized vowels. All the data were thus normalized by mapping the scores for the reference vowels onto 1 and 5 (oral and

Table I **Sorted data for Subject 10**

Test 1

Stimuli	Responses	m_5	s_5	m_4	s_4
ɑ	5 1 1 1 1	1.80	1.60	1.00	0.00
ɛ	1 1 1 1 1	1.00	0.00	1.00	0.00
ɔ	1 1 1 1 1	1.00	0.00	1.00	0.00
ɑ̃	5 2 4 4 5	4.00	1.10	3.75	1.09
ɛ̃	1 4 4 4 4	3.40	1.20	4.00	0.00
ɔ̃	5 5 5 5 5	5.00	0.00	5.00	0.00
ɑ̃$_1$	4 2 1 1 1	1.80	1.17	1.25	0.43
ɑ̃$_2$	5 1 1 1 1	1.80	1.60	1.00	0.00
ɑ̃$_3$	4 2 4 4 5	3.80	0.98	3.75	1.09
ɑ̃$_4$	5 5 5 4 5	4.80	0.40	4.75	0.43
ɑ̃$_5$	3 5 5 5 4	4.40	0.80	4.75	0.43
ɛ̃$_1$	5 1 1 1 1	1.80	1.60	1.00	0.00
ɛ̃$_2$	1 5 5 5 5	4.20	1.60	5.00	0.00
ɛ̃$_3$	5 5 5 5 5	5.00	0.00	5.00	0.00
ɛ̃$_4$	5 5 5 5 5	5.00	0.00	5.00	0.00
ɛ̃$_5$	1 1 1 1 1	1.00	0.00	1.00	0.00
ɔ̃$_1$	1 1 1 1 1	1.00	0.00	1.00	0.00
ɔ̃$_2$	4 4 1 1 1	2.20	1.47	1.75	1.30
ɔ̃$_3$	4 2 1 1 2	2.00	1.10	1.50	0.50
ɔ̃$_4$	4 5 5 4 2	4.00	1.10	4.00	1.22
ɔ̃$_5$	5 5 1 1 1	2.60	1.96	2.00	1.73
		$\bar{s}_5$	= 0.84	$\bar{s}_4$	= 0.39

Test 2

Stimuli	Responses	m_5	s_5	m_4	s_4
ɑ	2 2 2 1 4	2.20	0.98	2.25	1.09
ɛ	1 1 1 1 1	1.00	0.00	1.00	0.00
ɔ	1 1 1 1 1	1.00	0.00	1.00	0.00
ɑ̃	4 5 5 5 4	4.60	0.49	4.75	0.43
ɛ̃	5 5 5 4 5	4.80	0.40	4.75	0.43
ɔ̃	5 5 5 5 5	5.00	0.00	5.00	0.00
ɑ$_1$	4 4 4 1 2	3.00	1.26	2.75	1.30
ɑ$_2$	1 1 1 1 1	1.00	0.00	1.00	1.00
ɑ$_3$	2 4 4 1 5	3.20	1.47	3.50	1.50
ɛ$_1$	1 1 1 1 1	1.00	0.00	1.00	0.00
ɛ$_2$	1 1 1 1 1	1.00	0.00	1.00	0.00
ɛ$_3$	1 1 1 1 1	1.00	0.00	1.00	0.00
ɔ$_1$	1 1 1 1 1	1.00	0.00	1.00	0.00
ɔ$_2$	1 1 1 1 1	1.00	0.00	1.00	0.00
ɔ$_3$	1 1 1 1 1	1.00	0.00	1.00	0.00
		$\bar{s}_5$	= 0.31	$\bar{s}_4$	= 0.32

nasalized) and for the corresponding constructed stimuli by using the following transformation formula:

$$n = 4\frac{m-a}{b-a} + 1$$

where a is the subject's score for a given oral reference vowel, b is the subject's score for its nasalized counterpart, m is the subject's non-normalized score for one of its constructed counterparts and n is the subject's normalized score for this same constructed vowel. It can be seen that this formula maps the reference vowel scores onto 1 or 5, as desired since in these cases $m = a$ or b. It should be noted that the normalized score of a test vowel can be smaller than 1 or larger than 5. This indicates that the listener rates this test vowel as more oral, respectively more nasalized, than the corresponding reference vowel.

Criteria for selecting the best subjects

Each subject's responses were tabulated as in Tables I and II, and examined individually. Some of the listeners appeared to have performed more consistently than others. It was then decided to select the best subjects (according to some consistency criterion) and to compare their results with those of the entire population. Using the responses to the last four presentations, two criteria, based on ability to differentiate between oral and nasalized vowels were established. Criterion 1 required that a subject made the oral/nasalized distinction for the six pairs of reference vowels in both tests by having $m_4 < 3$ for all oral reference vowels and $m_4 > 3$ for all nasalized reference vowels. This subject's performance for the other stimuli could be expected to be adequate since he/she could use the reference vowels as guides in his/her judgement. Criterion 2 required that differentiation between oral and nasalized reference vowels be above a certain minimum established from the distribution of responses. Other criteria were considered but eventually rejected. Nine subjects were then grouped together and their scores as a group were compared with the scores of the entire population as a group. The non-normalized scores were averaged separately for the 9-subject group and for the 27-subject group and the standard deviations were found. The normalized scores of the 9-subject group were also averaged and their standard deviations computed. Since normalization of certain subject's scores resulted in some very large, very small or negative numbers, the averages and standard deviations of the normalized scores for the 27-subject group were not computed. Not surprisingly, the standard deviations of the 27-subject group were found to be larger than those of the 9-subject group. The results of the above computations are shown in Figs 1 and 2. Each figure summarizes the results of one test. On each graph the triangles and dash-dot line represents the average scores for the 27-subject group. The circles and broken line on the one hand and the squares and solid line on the other hand represent the average non-normalized and average normalized scores, respectively, for the 9-subject group.

Analysis of results

Whether, in the data presented in Figs 1 and 2, one considers the scores for the 9-subject group or for the 27-subject group, non-normalized or normalized (for the 9-subject group), the curves obtained have basically the same shape. The main difference is that the standard deviations, although not plotted, are always greater for the 27-subject group than for the 9-subject group. Average scores for the 9-subject group reach more extreme values than do those for the 27-subject group and as one should expect, the average scores for the normalized data also reach more extreme values than for the non-normalized ones, sometimes even greater than 5, e.g. for $\tilde{\varepsilon}_3$ and $\tilde{\varepsilon}_4$ on Fig. 1.

Table II Normalized data for Subject 10

Test 1

Stimuli	Normalized m_5	SD	Normalized m_4	SD
ɑ	1.00	0.00	1.00	0.00
ɛ	1.00	0.00	1.00	0.00
ɔ	1.00	0.00	1.00	0.00
ɑ̃	5.00	0.00	5.00	0.00
ɛ̃	5.00	0.00	5.00	0.00
ɔ̃	5.00	0.00	5.00	0.00
ɑ̃ 1	1.00	2.12	1.36	0.63
ɑ̃ 2	1.00	2.91	1.00	0.00
ɑ̃ 3	4.64	1.78	5.00	1.59
ɑ̃ 4	6.45	0.73	6.45	0.63
ɑ̃ 5	5.73	1.45	6.45	0.63
ɛ̃ 1	2.33	2.67	1.00	0.00
ɛ̃ 2	6.33	2.67	6.33	0.00
ɛ̃ 3	7.67	0.00	6.33	0.00
ɛ̃ 4	7.67	0.00	6.33	0.00
ɛ̃ 5	1.00	0.00	1.00	0.00
ɔ̃ 1	1.00	0.00	1.00	0.00
ɔ̃ 2	2.20	1.47	1.75	1.30
ɔ̃ 3	2.00	1.10	1.50	0.50
ɔ̃ 4	4.00	1.10	4.00	1.22
ɔ̃ 5	2.60	1.96	2.00	1.73

Test 2

Stimuli	Normalized m_5	SD	Normalized m_4	SD
ɑ	1.00	0.00	1.00	0.00
ɛ	1.00	0.00	1.00	0.00
ɔ	1.00	0.00	1.00	0.00
ɑ̃	5.00	0.00	5.00	0.00
ɛ̃	5.00	0.00	5.00	0.00
ɔ̃	5.00	0.00	5.00	0.00
ɑ 1	2.33	2.11	1.80	2.08
ɑ 2	− 1.00	0.00	− 1.00	0.00
ɑ 3	2.67	2.45	3.00	2.40
ɛ 1	1.00	0.00	1.00	0.00
ɛ 2	1.00	0.00	1.00	0.00
ɛ 3	1.00	0.00	1.00	0.00
ɔ 1	1.00	0.00	1.00	0.00
ɔ 2	1.00	0.00	1.00	0.00
ɔ 3	1.00	0.00	1.00	0.00

Since the average score curves were basically the same for the 27-subject group as for the 9-subject group, it was decided to use the curves representing the average scores of the whole population in analyzing the subjects' responses to individual vowel types. This was done in an attempt to obtain more general results and thus, give more weight to the

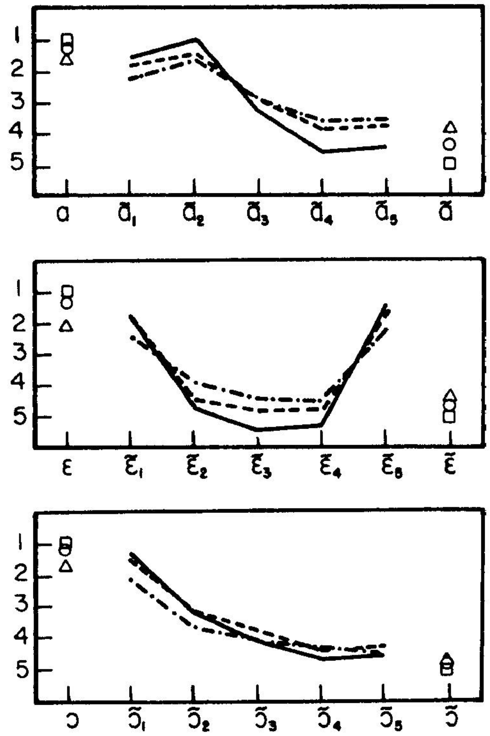

Figure 1 Average scores (last four responses) for Test 1 (see explanation in text).

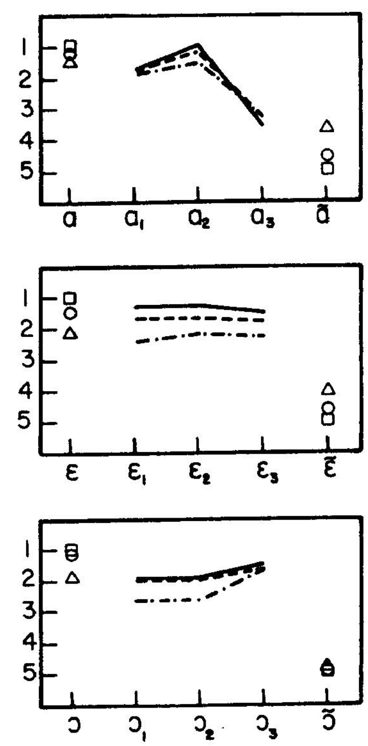

Figure 2 Average scores (last four responses) for Test 2 (see explanation in text).

observations. The same comments will hold for any curve, *mutatis mutandis*. If an average score is less than 3, it will be postulated that the listeners judged the vowels as "oral"; if an average score is more than 3, it will be postulated that the listeners judged the vowel as "nasalized", and if an average score is close to 3, it will be said that the listeners were uncertain about the quality of the vowel. Considering the subjects' average scores for the stimuli constructed from /tɑ̃t/ as displayed on Fig. 1, one may notice that the average scores for

$\tilde{a}_1$ and $\tilde{a}_2$ are less than 3 and that the average scores for $\tilde{a}_4$ and $\tilde{a}_5$ are more than 3. The average scores for $\tilde{a}_3$ are close to 3. The boundary between "oral" and "nasalized" appears to be around $\tilde{a}_3$. It can also be noticed that average scores for $\tilde{a}_2$ are usually less than those for $\tilde{a}_1$; this would suggest that subjects had a tendency to judge $\tilde{a}_2$ as more "oral" than $\tilde{a}_1$. These items were thoroughly rechecked to rule out a possible inversion of the two items in the preparation of the listening test tape. No explanation can be found to explain this observation.

For stimuli constructed from /tɛ̃t/, one notices (Fig. 1) steadily increasing average scores from $\tilde{\varepsilon}_1$ to $\tilde{\varepsilon}_4$, the boundary between "oral" and "nasalized" occurring between $\tilde{\varepsilon}_1$ and $\tilde{\varepsilon}_2$. Average scores for $\tilde{\varepsilon}_5$ are much smaller than those for $\tilde{\varepsilon}_4$ and would suggest that $\tilde{\varepsilon}_5$ had been judged as "oral" by the listeners. A discussion of this point will be presented in a later section.

For stimuli constructed from /tɔ̃t/, the graphs on Fig. 1 show an increase in average scores from $\tilde{ɔ}_1$ to $\tilde{ɔ}_5$, the boundary between "oral" and "nasalized" occurring between $\tilde{ɔ}_1$ and $\tilde{ɔ}_2$.

Considering the subjects' average scores to stimuli constructed from /ana/ displayed in Fig. 2, one notices that the average score for a_1 and a_2 are smaller than those for a_3. The average scores for a_1 and a_2 are less than 3 and the average scores for a_3 are slightly more than 3.

The graphs on Fig. 2 also show that the average scores for stimuli constructed from /ɛnɛ/ and /ɔnɔ/ are all less than 3, and that the average scores for $ɔ_3$ are slightly less than those for $ɔ_1$ and $ɔ_2$. So, the listeners judged stimuli constructed from /ana/, /ɛnɛ/ and /ɔnɔ/ as being mostly "oral", with the exception of $ɔ_3$ which was judged "in between". None of these stimuli were perceived as "nasalized" as was the case for stimuli constructed from /tãt/, /tɛ̃t/ and /tɔ̃t/, but average scores indicate that some stimuli were judged more "oral" than others.

Of the 9 subjects who satisfied criteria 1 and 2, 7 had been presented with Test 1 first, and 2 with Test 2 first. This was not judged significant since the probability of such an event is in this case

$$p = \binom{9}{7} \cdot \left(\frac{15}{27}\right)^7 \cdot \left(\frac{12}{27}\right)^2 = 11.62\%.$$

The possibility of sequential effects was examined. For the last four presentations of each vowel, a table was constructed of the rating given to the previous vowel presented. No influence could be found. A series of *t*-tests was then run to compare the score of each reference vowel in Test 1 with its score in Test 2. No significant difference was found.

The distribution of s_4's (standard deviation of the score of the last four presentations for one vowel and one subject) was then examined. Two *t*-tests showed that for Test 1 and Test 2, there was no significant difference between the scores of the reference vowels and those of the test vowels. The fact that the test vowels were made of manipulated speech whereas the reference vowels were made of re-recorded natural speech thus seems of no importance in the results.

Another series of *t*-tests was then run to compare the s_4's for each reference vowel with the distribution of the s_4's of all six reference vowels, for each test separately. The only significant difference was found in the 27-subject group, in Test 2, where the s_4's for ɔ̃ were statistically significantly different from the six reference vowels as a whole ($p \leqslant 0.01$). In fact, for 23 out of 27 subjects, s_4 for the vowel ɔ̃ was 0. This suggests that ɔ̃ was more easily identified than the other reference vowels in Test 2. Indeed, several listeners reported that they "recognized" more easily than the other stimuli. This could be due to the particular speaker used for the experiment. This possibility could be verified through a test using several speakers and observing whether there is any vowel–speaker interaction as far as perceived nasalization is concerned. Another possibility, to be discussed below, is that among different phonemically nasalized vowels, e.g. /ɛ̃/, /ɑ̃/, /ɔ̃/, some are perceived as more nasalized than others. In order to check this point, for each reference vowel and for each test, the distribution of the 27 subjects' m_4's was established. For each pair of reference vowels (i.e. ɛ–ɛ̃, ɑ–ɑ̃ and ɔ–ɔ̃) the difference between the lower quartile value for the nasalized vowel and the upper quartile for the oral vowel was computed and found to be as follows:

	Test 1	Test 2
ɑ – ɑ̃	1.75	1.75
ɛ – ɛ̃	1.50	1.25
ɔ – ɔ̃	2.00	2.75

Other intervals were considered but yielded exactly the same ranking of the three vowel pairs. The above table supports the hypothesis that (phonemically) different nasalized vowels may be assessed different degrees of nasalization.

Correlation between articulatory and perceptual data

Fiberscopic data pertaining to the test items of Test 1 was available (Benguerel *et al.*, 1976*a*). The limits of the time interval corresponding to each constructed stimulus were plotted on the graph *velar height vs time* for the utterance the stimulus had been extracted from. Because of the high token-to-token repeatability found by Benguerel *et al.* (1976*a*), a similar re-plotting was done for the test items of Test 2. An example is given in Fig. 3. A comparison between the listeners' judgements and the states of the velopharyngeal port for the different test vowels is given in Tables III and IV. The subjects' rating was taken as oral whenever the average score $\bar{m}_4$ was less than 3, as nasalized whenever $\bar{m}_4$ was more than 3, and "in-between" whenever $\bar{m}_4$ was 3 or very close to 3. Similarly, the velopharyngeal passage was considered closed whenever $h > h_0$ (cf. Benguerel *et al.*, 1976*a*), and open whenever $h < 0$. The high correlation between averaged subjects' judgement and velopharyngeal opening is quite obvious. Only 3 test vowels depart from the expected correlation, all three being manipulated versions of an /ɑ/. This may not be too surprising since it is a known fact (Fant, 1960, p. 139; Benguerel *et al.*, 1976*a*) that open (oral) vowels have generally a low velar height, with sometimes a slight velopharyngeal opening, despite the phonemically oral character of the vowel. It appears plausible that listeners may take this into account and require a greater degree of velopharyngeal opening to attribute nasality to a vowel along the dimension [ɑ–ɑ̃]. It is also worth mentioning that Ali *et al.* (1971) observed that listeners could predict the nasal quality of an upcoming consonant if it followed an /ɑ/. This finding is corroborated by Table IV.

Conclusion

The experiment described above demonstrated that for manipulated stimuli derived from natural nasalized vowels the degree of perceived nasalization increases monotonically with the degree of opening of the velopharyngeal port. In the manipulated stimuli, this (equivalent) degree of opening was constant since the stimuli were made of forty repetitions of the same period. In the natural stimuli however, the degree of opening of the velopharyngeal port varied widely through the course of the vowel, as illustrated in Fig. 3. The overall

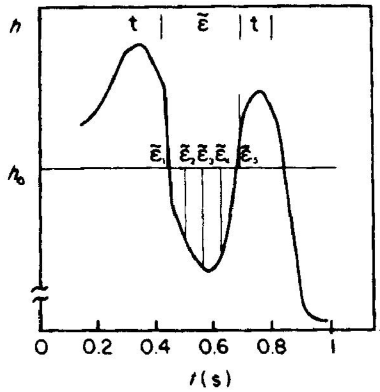

Figure 3 Time correspondence between the five ε̃ stimuli and the velar height curve of Benguerel *et al.* (1976*a*).

Table III **Comparison of the results of Test 1 with the articulatory data**

Stimulus	Subjects' judgement	Velopharyngeal port
ɑ̃ 1	oral	closed
ɑ̃ 2	oral	open
ɑ̃ 3	in-between	open
ɑ̃ 4	nasalized	open
ɑ̃ 5	nasalized	closed
ε̃ 1	oral	closed
ε̃ 2	nasalized	open
ε̃ 3	nasalized	open
ε̃ 4	nasalized	open
ε̃ 5	oral	closed
ɔ̃ 1	oral	closed
ɔ̃ 2	nasalized	open
ɔ̃ 3	nasalized	open
ɔ̃ 4	nasalized	open
ɔ̃ 5	nasalized	open

Table IV **Comparison of the results of Test 2 with the articulatory data**

Stimulus	Subjects' judgement	Velopharyngeal port
ɑ 1	oral	closed
ɑ 2	oral	closed
ɑ 3	in-between	open
ε 1	oral	closed
ε 2	oral	closed
ε 3	oral	closed
ɔ 1	oral	closed
ɔ 2	oral	closed
ɔ 3	oral	closed

nasalized quality perceived is thus the result of some averaging or of some integration, the detail of which needs to be determined quantitatively both in the dimension of velopharyngeal port opening and in the time dimension. There is also a definite possibility that other articulatory factors such as pharyngeal cavity modification may have to be taken into account. Two lines of attack of this question are possible.

One of them would use stimuli produced by a human speaker whose relevant articulatory parameters–in particular the area function of the vocal tract (both oral and nasal)–would be monitored and accurately measured during production. In addition to the difficulty of obtaining measurements sufficiently accurate for the purpose of the study, this method would also present, for the speaker, the difficulty of producing at different degrees of opening steady vowels in which the position of the velum would be held stably for a duration of a second or so.

The second method would involve the use of a vocal tract analog. With this method, the determination of the vocal tract area function is not the major problem since it is the input to the model. The difficulty is to produce nasalized vowels which are acceptable both spectrally and auditorily. The models considered so far have not met the criterion of sufficient auditory quality for nasalized vowels but hopefully some will soon. Such a model should ultimately also allow to investigate, as mentioned earlier, what makes the difference(s) between the phonetic realization of nasalization in different languages (in all of which it is contrastive) since we have at least informal evidence that the label "nasalization" covers a quality which is at least perceived, and possibly produced, differently by the speaker–listeners of each of these languages.

References

Ali, L., Gallagher, T., Goldstein, J. & Daniloff, R. (1971). Perception of Coarticulated Nasality. *Journal of the Acoustical Society of America*, **49**, 538–540.

Benguerel, A. -P., Hirose, H., Sawashima, M. & Ushijima, T. (1976*a*). Velar coarticulation in French: a fiberscopic study. *Journal of Phonetics*, **5**, 149–158.

Benguerel, A. -P., Hirose, H., Sawashima, M. & Ushijima, T. (1976*b*). Velar coarticulation in French: an electromyographic study. *Journal of Phonetics*, **5**, 159–167.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

House, A. S. & Stevens, K. N. (1966). Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders*, **21**, 218–232.

Ladefoged, P. (1978). Phonetic differences within and between languages. *UCLA Working Papers in Phonetics*, **41**, 32–40.

Peterson, G. E. & Shoup, J. E. (1966). A Physiological Theory of Phonetics. *Journal of Speech and Hearing Research*, **9**, 5–67.