Influence of the vocal effort on vowels.

Maria-Gabriella Di Benedetto* and Jean-Sylvain Liénard**.

*Dept. INFOCOM, Università degli Studi di Roma 'La Sapienza', via Eudossiana 18,
00184 Rome, Italy

**LIMSI-CNRS, BP 133, 91403 Orsay cedex, France

paper presented at the 127th meeting of the Acoustical Society of America, Cambridge, Massachusetts, June 1994.
Work supported by a CNR-CNRS grant.

Background

Objective

The present study aims at investigating the interaction between linguistic and non-linguistic information in the analysis of speech. In particular, this investigation is applied to vocal effort.
How do the acoustic cues which characterize vowels (for example) vary according to the degree of vocal effort which reflects the unconscious variations observable in normal life conditions?
By answering this question, it may be possible to explain a source of variations in automatic speech processing and consequently reduce speech variability.

Method

This study focused on the analysis of isolated French vowels, forming meaningful words, uttered by several speakers, according to different vocal efforts. The vocal efforts considered varied within a normal life dynamic range, from weak to strong. The recordings were made in low-constrained recording conditions.
The speech materials obtained were perceptually validated by running a perceptual test on vowel identity, speaker gender, and speaking style.
An acoustic analysis was carried out on these speech materials. The fundamental frequency, formants (F1, F2, and F3), and formant amplitudes (A1, A2, A3) were manually measured. On the basis of these measurements, a relation between the parameters values and vocal effort was searched.

Previous work

The role of the fundamental and the higher formants in the perception of speaker size, vocal effort, and vowel openness has been examined by Traunmüller (1985). On the basis of perceptual experiments in which synthetic stimuli were used, it was shown that whereas the phonetic quality of the vowel perceived remained constant, when the group of the lower formants (F0 and F1 in front vowels, and also F2 in back vowels) was moved upward the listeners perceived an increase in vocal effort, whether when both groups of formants were moved upward, the listeners perceived a decrease in speaker size. The articulatory dynamics of loud and normal speech were analyzed by Schulman (1989). Junqua (1993) analyzed the effect of the Lombard reflex on the basis of the acoustic analysis of the changes occurring during Lombard speech, and perceptual experiments. The results show that the effect is variable across speakers. In a recent paper, Granström and Nord (1992) analyzed the influence of the speaking style (defined as weak, normal, and strong) on long time average spectra. Results show that the average fundamental frequency was increased considerably in the loud version and that the relative level of the fundamental and the slope of the spectrum also varied significantly.

Data-base

Data-Base recording

The data-base CORENC consists of 12 French isolated vowels [i,e,ɛ,y,ø,œ,a,o,u,ɛ̃,ɑ̃,ɔ̃] uttered by 13 speakers (6 males and 7 females) according to three degrees of vocal effort, in two different sessions at six months interval. In French, these vowels correspond to lexical words. Consequently, they are easy to be specified by a non-phonetician speaker or listener. The vowel [ə] was not included in this analysis since it does not correspond to a lexical word.

The speech materials were recorded by means of an omnidirectional microphone. The distance between the speaker's mouth and the microphone was 30 centimeters. All the recordings were comparable in terms of sound level: the input level was maintained unchanged during both the recording and the subsequent processing. In order to obtain the three degrees of vocal effort, it was asked to repeat the vowels pronounced by the operator at a level appropriate to the distance between the speaker and the operator. Consequently, each series of vowels was either normal (1.5 meters between the speaker and the operator, denoted as N condition), or close (30 cm, denoted as C condition), or far (7 m, denoted as F condition). In the average, the dynamic interval between the two extreme conditions corresponds to 12 dB.

The total number of vowels recorded was 720. The subset of vowels used in the present experiment consisted of 612 tokens.

Data-Base evaluation

The data base evaluation was obtained by running an identification test on the 612 vowel tokens which were presented to the listener in a random order. The complete description of the evaluation test on the 720 tokens set is reported in Liénard and Di Benedetto (1992). Each token was presented once. The listener had to simultaneously determine the vowel identity, the speaker gender, and the vocal effort, or could also decide not to make a choice by checking a question mark box.

The speakers and the listeners belonged to separate groups and were not trained to the specific tasks of the experiment. Seven listeners participated in the evaluation test.

Results of this test show that the global identification error rate as regards vowel identity, speaker gender, and vocal effort was 9.6% (7% for the oral vowels and 15% for the nasal vowels), 4% (mainly due to one speaker), and 28%, respectively. The vowel identity score is quite high showing that the speakers in the recording phase uttered the vowels in a correct manner and that these vowels were correctly identified in the listening phase; this result validates the process of using isolated vowels and naive speakers in the experiment.

When the scores are rated according to the speaking style, the results are as follows:

- vowel identification score was relatively constant through the styles (C=4.4%, N=2.6%, F=2.6%), with a slightly higher error rate in the C condition;
- very few errors were observed on the speaker's gender (C=1.5%, N=1.7%, L=0.8%);
- the errors regarding the vocal effort were significantly lower in the F condition (F=3.6%), than in the C condition (C=10%). As expected, the higher error rate was observed for the N condition (N=14.4%) which is in between the other two conditions and consequently the ambiguity in this case is equivalent to the sum of the other two.

Measurements

The following parameters were estimated: fundamental frequency (F0), formant frequencies (F1, F2, F3), formant amplitudes (A1, A2, A3). The speech materials were analyzed using a spectrographic analysis program developed by J.L.Gauvain at LIMSI. These parameters were manually estimated by visual examination of the narrow-band spectra. Only the measurements in one frame were retained. The selected frame corresponded to the best representative of the token.

Results 1

F0 range: male and female speakers

It is well accepted that the amplitudes of the formants be measured in dB and the formant frequencies be measured in Bark, but which is the better measurement unit for F0 ? As far as F0 is concerned, is a musical scale based on logarithmic units better suited than a linear scale? Based on our measurements, it was concluded that a linear scale was more appropriate. When represented in logarithmic scale, the data showed a greater variability for male voices as compared to female voices and a difference between average male and female F0 varying according to the speaking style. On the contrary, when represented in linear units, the standard deviation for male and female F0 values was significantly similar, and the difference between the average values of F0 in the two groups was held constant. Since the F0 values were in a range which corresponded to a linear relation between hertz and Bark, it was equivalent to express the F0 values in hertz or in Bark. In the following, the F0 values will be expressed in Bark.

Fundamental frequency, formants and amplitudes variations

Averaged over all vowels and speakers, the formant frequencies exhibit significant variations with respect to vocal effort. In particular:

- F0 and F1 increase going from the C to the F condition at a rate of about 4 centibark/dB;
- F2 and F3 remain constant;
- A1, A2, and A3 all increase significantly from the C to the F condition. These parameters do not seem to vary with the same rate: A3 increases faster than A1 and A2, and A2 increases faster than A1 but slower than A3. Consequently, the interval between the A1 and A3 values decreases with vocal effort, revealing a tendency for the spectrum slope to decrease by reinforcing its higher frequencies.

Representation in the F1 vs. F2 plane

The representation in the F1 vs. F2 plane (F1 and F2 are expressed in centibark) show the F1 and F2 variations according to speaking style, for each vowel. According to a Student t-test on paired data, the F1 variations are all significant ($p=0.05$, or less), while the F2 variations are not. It can be observed that the F1 variations due to a change in speaking style vary from one vowel to the other, from a minimum obtained for the cardinal vowels [i,a,u] (minimum value of 20 centibark for [a]) to higher values obtained for the non extreme vowels (with a maximum of 76 centibark for [ø]). The F1 variations observed are sufficiently large to be perceptually relevant, as shown recently by Kewley-Port (1994).

Results 2

Spectral Center of Gravity

The categorical perceptual effect named Spectral Centre of Gravity (SCG) was found by Chistovich and her colleagues (Chistovich, Sheikin and Lublinskaya, 1979). These experimenters pointed out that if a two formant stimulus must be matched by a one formant stimulus, the matching criterion depends upon the distance between the location of the two formants. If the two formants are placed closer than 3.5 Bark approximately, the subjects match this stimulus with one formant located in a position corresponding to a weighted average of the two formants. In this case, the match is dependent upon the amplitudes of the formants. If the distance is greater than 3.5 Bark, the two formants are matched to one formant located at one of the two formants. In this case, insensitivity over a large range of amplitude variations is observed.

Formant differences

Averaged over all vowels and speakers, the differences between formant frequencies exhibit the following characteristics:

- since F1 and F0 vary concurrently, the F1-F0 difference, which is correlated to the degree of openness, remains approximately constant through the different styles.
- since F2 and F3 remain constant, the F3-F2 difference is constant through the different styles.
- since F2 is constant while F1 increases, the F2-F1 difference decreases from the C to the F condition.

Accordingly, the vowels were represented in the F1-F0 vs. F3-F2 plane. Results show that all vowels are well clustered and separate except in the [i,y] pair.

Spectral integration

Statistical tests were carried out on the analysis data. The results can be summarized as follows:

- the F3-F2 difference was shown not to be significantly stable through styles. It was also observed that the F3-F2 difference was lower than 3.5 Bark for [i,y,e,ɛ]. The center of gravity in this region was computed by taking into account the F2 and F3 frequencies and A2 and A3 amplitudes. This parameter proves to be very relevant. In fact, the center of gravity between F2 and F3 is significantly stable through styles for the vowels [i,y,e,ɛ], that is for those vowels for which spectral integration of F2 and F3 should occur according to the F3-F2 distance;
- the F2-F1 difference was shown not to be significantly stable through styles. It was also observed that the F2-F1 difference was lower than 3.5 Bark for [a,o]. In this case, the center of gravity between F1 and F2 is significantly stable through styles only for the vowel [a], which is the only low vowel of French.
- the F1-F0 difference was shown to be significantly stable through styles for the cardinal vowels [i,a,u]. It was also observed that the F1-F0 difference was lower

than 3.5 Bark for [i,y,u,o,ø,e]. Future work is planned to investigate the possible role of a center of gravity in this region.

Conclusions

Summary

The results of the analysis can be summarized as follows:

- a linear scale is better suited for representing F0 than a logarithmic scale. Through styles, the average difference in F0 values for male and female speakers was constant in average, and the standard deviation for male and female F0 values was very similar;
- from the C to the F condition there was an increase in F0 and F1 values of about 4 centibark/dB;
- from the C to the F condition F2 and F3 remained constant;
- from the C to the F condition A1, A2 and A3 increased;
- from the C to the F condition the global slope decreased slightly;
- the center of gravity of F2 and F3 was significantly stable for the vowels for which the relation F3-F2 < 3.5 Bark was verified;
- the center of gravity of F1 and F2 was significantly stable for the vowel [a].

Interpretation

The results of the present analysis show that various parameters vary concurrently when a vowel is pronounced by several speakers and according to different degrees of vocal efforts, from a C to a F condition. The behaviour of the following parameters was examined: F0, F1, F2, F3, A1, A2, A3.

In order to characterize in a complete way vowels uttered with different degrees of vocal effort, it may be essential to consider parameters which should be based on properties related to the entire spectral shape. In particular, the spectral center of gravity in the region of the first and second formants, and in the region of the second and third formants, remained significantly stable for those vowels for which spectral integration should occur, based on the Spectral Center of Gravity effect (difference of the two formants lower than 3.5 Bark, approximately).

Future investigation

The results of the present analysis show that it is appropriate to consider in a certain frequency region, a parameter which takes into account the spectral shape in that region, in its whole. A similar analysis should be carried out in the region of F1 and F0. In particular, the amplitude of the fundamental frequency A0 should be estimated and the center of gravity between F0 and F1 should be computed.

This last analysis may lead to the interesting hypothesis that it is appropriate to represent vowels by means of an integrated form of the spectrum.

References

Chistovich, L.A., Sheikin, R.L. & Lublinskaya, V.V. (1979). "Centres of gravity and spectral peaks as the determinants of vowel quality," in: B.Lindblom and S. Öhman, eds., Frontiers of Speech Communication Research, Academic Press, London, pp.143-157.

Granström, B. and Nord, L. (1992). "Neglected dimensions in speech synthesis," Speech Communication, vol.11, no.4&5, pp.459-462.

Junqua, J.C. (1993). "The Lombard reflex and its role on human listeners and automatic speech recognizers," J. Acoust. Soc. Am. 93 (1), pp.510-524.

Kewley-Port, D. and Watson, C.S. (1994). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. 95 (1), pp.485-496.

Liénard, J.S. and Di Benedetto, M.G. (1992). "Evaluation perceptive d'un corpus de voyelles Françaises émises isolément par plusieurs locuteurs selon diverses forces de voix," 19émes Journées d'Etude sur la Parole, Bruxelles, May 1992, pp.469-474.

Schulman, R. (1989). "Articulatory dynamics of loud and normal speech," J. Acoust. Soc. Am. 85 (1), pp.295-312.

Traunmüller, H. (1985). "The role of the fundamental and the higher formants in the perception of speaker size, vocal effort, and vowel openness," Proc. of the French-Swedish Seminar on Speech, Grenoble, GALF-EAFSR, B. Guérin and R. Carré eds, pp.209-219.

BACKGROUND

OBJECTIVE

- ♦ Analysis of the interaction between linguistic and non-linguistic information
    - ♦ Application to a non-linguistic factor: vocal effort
        - ♦ How do the acoustic cues vary according to the degree of vocal effort reflecting unconscious variations in normal-life conditions ?
            - ♦ Reduce speech variability by explaining a possible source of variations for automatic processing purposes

METHOD

- ♦ Analysis of isolated French vowels
    - • forming lexical units
    - • uttered by several speakers (males and females)
    - • according to several vocal efforts
    - • within a normal-life dynamic range : from weak to strong (12 dB)
    - • in low-constrained recording conditions

- ♦ Perceptual evaluation of the recorded material
    - • vowel identity
    - • speaker gender
    - • vocal effort

- ♦ Acoustic analysis
    - • manual measurement of F0, F1, F2, F3, A1, A2, A3
    - • search for relations between the above parameters and vocal effort

PREVIOUS WORK

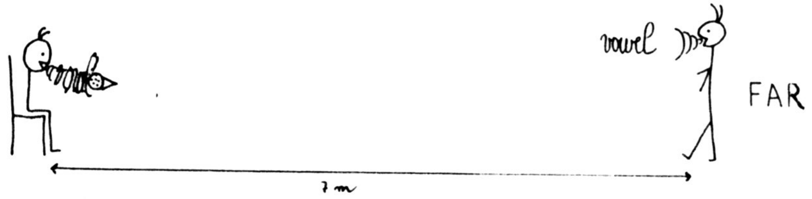
- ♦ Traunmüller (1985) - interaction between the fundamental and the higher formants
    - ♦ Schulman (1989) - articulatory dynamics of loud and normal speech
        - ♦ Junqua (1993) - Lombard reflex
            - ♦ Granström and Nord (1991) - influence of speaking style on long time average spectra

DATA-BASE

DATA-BASE RECORDING

- **speakers**: 7 **females** and 6 males
- 1 or 2 **repetitions**
- 12 French isolated vowels (9 orals and 3 nasals)
- 3 vocal efforts: Close, Normal, Far
- 720 vowel tokens

vowel
vowel
NORMAL
1.5 m

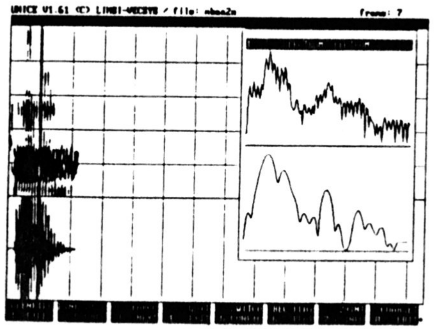


DATA-BASE EVALUATION

◆ Identification errors	vowels	9.6%	orals 7% nasals 15%	
	gender	4.0%	(mainly due to 1 speaker in C mode)	
	vocal effort	28.0%		
◆ Depending on style	vowels	C 4.4%	N 2.6%	F 2.6%
	gender	" 1.5%	" 1.8%	" 0.7%
	vocal effort	" 10.0%	" 14.4%	" 3.6%

MEASUREMENTS

- only for oral vowels
- **parameters** : F0, F1, F2, F3, A1, A2, A3
- manually estimated on one frame of the spectrogram


UNICE V1.61 (C) LIMSI-VECSYS / file: mbea2a

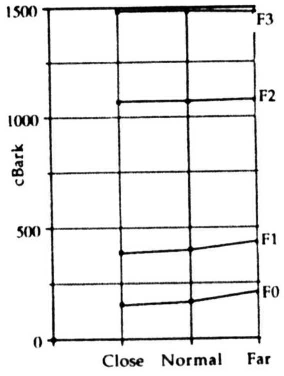
frame: 7

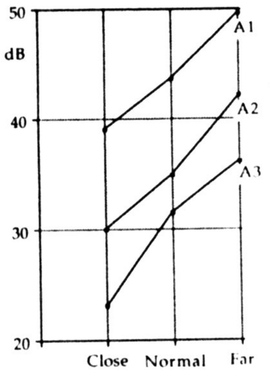
RESULTS 1

F0, FORMANTS AND AMPLITUDES VARIATIONS

How do the fundamental, formant frequencies and formant amplitudes vary in average?

- F0 increase
- F1 increase
- F2 increase
- F3 increase



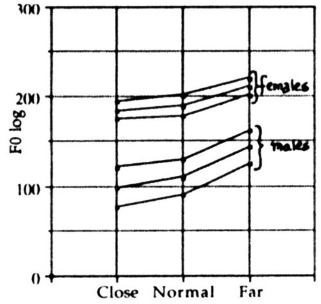


- all amplitudes increase
- A2 and A3 increase more than A1

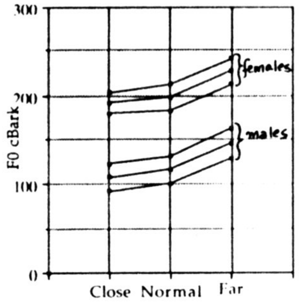
- **these** variations are significant according to a student t-test for paired samples (p = 0.01)
- each point represents the average of 153 measurements

F0 RANGE : MALE AND FEMALE SPEAKERS

Is a musical scale (log units) appropriate to represent F0 evolutions with respect to vocal effort?



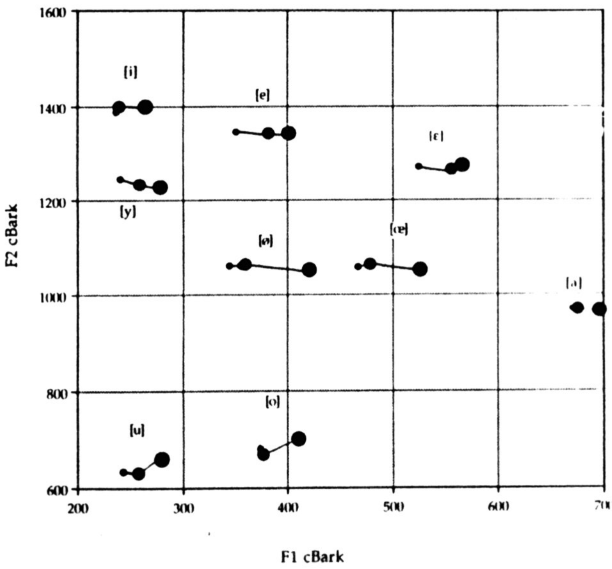
F0 log



F0 linear (Bark/Hertz)

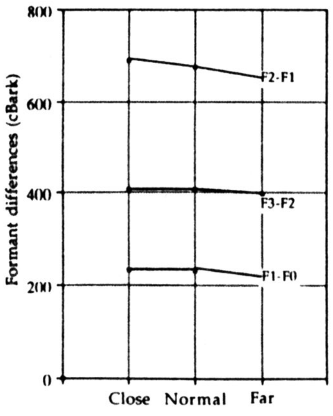
⇒ **linear scale better suited**

REPRESENTATION IN THE F1-F2 PLANE

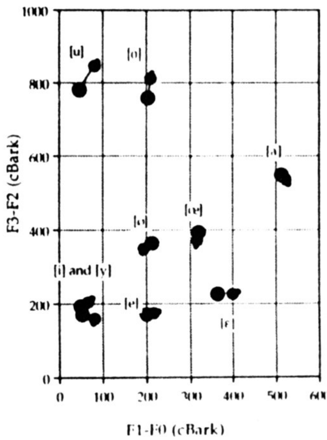


RESULTS 2

FORMANT DIFFERENCES



formant differences variations through styles



representation in a formant differences plane

SPECTRAL CENTER OF GRAVITY Chistovich, Sheikin, Lublinskaya (1979)

a short summary of Chistovich at al. experiment

- Experiment: two formant stimulus (F_A & F_B) matched by one formant stimulus (F_{SCG})

- Observations:
  - if F_A-F_B < 3.5 Bark then F_{SCG} is a weighted average of F_A & F_B **it depends on the amplitudes of F_A and F_B**
  - if F_A-F_B > 3.5 Bark then F_{SCG} corresponds to F_A or F_B **it is insensitive over a large range of amplitude variations**

SPECTRAL INTEGRATION

- F_2 and F_3 region
  - F_3-F_2 not significantly stable (statistically) through styles
  - F_3-F_2 < 3 Bark for [i,y,e,ɛ]
  - center of gravity of F_2 and F_3 (which takes into account F2 and F3 amplitudes) significantly stable through styles for [i,y,e,ɛ].

- F_2 and F_1 region
  - F_2-F_1 not significantly stable through styles
  - F_2-F_1 < 3 Bark for [a,o]
  - center of gravity of F_2 and F_1 (which takes into account F1 and F2 amplitudes) significantly stable through styles for [a] (the only French low vowel).

- F_1 and F_0 region
  - F_1-F_0 significantly stable through styles for [i,a,u]
  - F_1-F_0 < 3 Bark for [i,y,u,o,ø,e]
  - future work: possible center of gravity in the F0 and F1 region

CONCLUSIONS

SUMMARY

- linear scale better than log scale for F_0
- increase of F_1 and F_0 by 4 centibark/dB
- constancy of F_1 and F_2
- A_1, A_2, and A_3 increase
- global slope decreases slightly

INTERPRETATION

- various parameters vary concurrently
- parameters based on properties related to the entire spectral shape
- spectral center of gravity in the regions of F1 and F2, and of F2 and F3 remains significantly stable for those vowels for which spectral integration should occur

FUTURE INVESTIGATION

- analyze the amplitude of the fundamental frequency A_0
  - analyze the center of gravity in the region of F_1 and F_0
    - investigate the possibility of representing vowels by means of an integrated form of the spectrum